Stochastic Optimization

Multistage Stochastic Programs: Solution Methods

Alireza Ghaffari-Hadigheh

Azarbaijan Shahid Madani University (ASMU)

hadigheha@azaruniv.edu

Fall 2017

Overview

Introduction

- Most practical decision problems involve a sequence of decisions that react to outcomes that evolve over time.
- The multistage stochastic linear program with fixed recourse

$$\begin{aligned} \min z = {} & c^1x^1 + \mathrm{E}_{\xi^2}[\min c^2(\omega)x^2(\omega^2) + \cdots + \mathrm{E}_{\xi^H}[\min c^H(\omega)x^H(\omega^H)]\ldots] \\ \text{s. t.} \quad & W^1x^1 = h^1\,, \\ & T^1(\omega^2)x^1 + W^2x^2(\omega^2) = h^2(\omega)\,, \\ & \quad \cdots \vdots \\ & T^{H-1}(\omega^H)x^{H-1}(\omega^{H-1}) + W^Hx^H(\omega^H) = h^H(\omega)\,, \\ & x^1 \geq 0\,; \quad x^t(\omega^t) \geq 0\,, \qquad t = 2,\ldots,H\,; \end{aligned} \tag{4.1}$$

- $c^1 \in \mathbb{R}^{n_1}$, $h^1 \in \mathbb{R}^{m_1}$, are known vectors.
- $\xi^t(\omega)^T = (c^t(\omega)^T, h^t(\omega)^T, T_{1.}^{t-1}(\omega), \ldots, T_{m_t.}^{t-1})$ is a random N_t -vector defined on (Ω, Σ_t, P) for all $t = 2, \ldots, H$,
- Each W^t is a known $m_t \times n_t$ matrix.
- The decisions x depend on the history up to time t , which we indicate by ω^t .
- We suppose that Ξ^t is the support of ξ_t

Deterministic equivalent form

- Deterministic equivalent form in terms of a dynamic program.
- If the stages are 1 to H , define states as $x^t(\omega^t)$.
- The only interaction between periods is through this realization.
- Terminal conditions

$$\begin{aligned} Q^H(x^{H-1},\xi^H(\omega)) = \min c^H(\omega)x^H(\omega) & \\ \text{s. t. } W^H x^H(\omega) = h^H(\omega) - T^{H-1}(\omega)x^{H-1} , & \\ x^H(\omega) \geq 0 . & \end{aligned} \tag{4.2}$$

- Solutions for other stages can be obtained with a backward recursion $\mathcal{Q}^{t+1}(x^t) = E_{\xi^{t+1}}[Q^{t+1}(x^t, \xi^{t+1}(\omega))]$.

$$\begin{aligned} Q^t(x^{t-1},\xi^t(\omega)) = \min c^t(\omega)x^t(\omega) + \mathcal{Q}^{t+1}(x^t) & \\ \text{s. t. } W^t x^t(\omega) = h^t(\omega) - T^{t-1}(\omega)x^{t-1} , & \\ x^t(\omega) \geq 0 , & \end{aligned} \tag{4.3}$$

Final Goal

- Other state information in terms of the realizations of the random parameters up to time t should be included if the distribution of ξ^t is not independent of the past outcomes.
- The value we seek is:

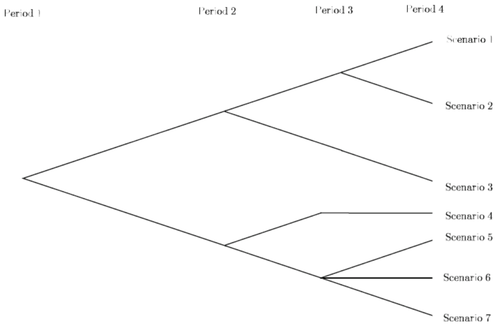
$$\begin{aligned} \min z &= c^1 x^1 + \mathcal{Q}(x^1) \\ \text{s. t. } W^1 x^1 &= h^1 \, , \\ x^1 &\geq 0 \, , \end{aligned} \tag{4.4}$$

- feasibility sets: $K^t = \{x^t | \mathcal{Q}^{t+1}(x^t) < \infty\}$

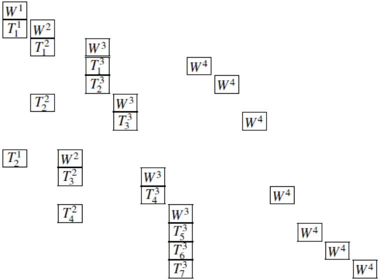
Theorem

The sets K^t and functions $\mathcal{Q}^{t+1}(x^t)$ are convex for $t = 1, \ldots, H-1$ and, if Ξ^t is finite for $t = 1, \ldots, H-1$, then K^t and $\mathcal{Q}^{t+1}(x^t)$ are polyhedral.

- We may describe the feasibility sets K^t in terms of intersections of feasibility sets for each outcome if we have finite second moments for ξ^t in each period.
- This result is also true when we have a finite number of possible realizations of the future outcomes.
- In this case, the set of possible future sequences of outcomes are called scenarios.
- The description of scenarios is often made on a tree.



- The deterministic equivalent program to (4.1) with a finite number of scenarios is still a linear program.
- It has the structural form



- Subscripts indicate different scenario realizations for the T^t matrices.
- A difficulty is that, these problems become extremely large as the number of stages increases, even if only a few realizations are allowed in each stage.

Block separable recourse

Definition

A multistage stochastic linear program (4.1) has block separable recourse if for all periods $t = 1, \ldots, H$ and all ω, the decision vectors, $x^t(\omega)$, can be written as $x^t(\omega) = (w^t(\omega), y^t(\omega))$ where w^t represents aggregate level decisions and y^t represents detailed level decisions. The constraints also follow these partitions:

1. The stage t objective contribution is $c^t x^t(\omega) = r^t w^t(\omega) + q^t y^t(\omega)$.
2. The constraint matrix W^t is block diagonal:
$$W^t = \begin{pmatrix} W^t & 0 \\ 0 & T^t \end{pmatrix} \quad (1)$$
3. The other components of the constraints are random but we assume that for each realization of ω , $T^t(\omega)$ and $h^t(\omega)$ can be written:
$$T^t(\omega) = \begin{pmatrix} R^t(\omega) & 0 \\ S^t(\omega) & 0 \end{pmatrix} \ \textit{and} \ \ ht(\omega) = \begin{pmatrix} b^t(\omega) \\ d^t(\omega) \end{pmatrix} \quad (2)$$
where the zero components of T^t correspond to the detailed level variables.

- With block separable recourse, $Q^t(x^{t-1}, \xi^t(\omega))$ is rewritten as the sum of two quantities,

$$Q^t_w(w^{t-1}, \xi^t(\omega)) + Q^t_y(w^{t-1}, \xi^t(\omega)),$$

where we need not include the y^{t-1} terms in x^{t-1} ,

$$\begin{aligned} Q^t_w(w^{t-1}, \xi^t(\omega)) = \min\, & r^t(\omega) w^t(\omega) + \mathcal{Q}^{t+1}(x^t) \\ \text{s. t. } & W^t w^t(\omega) = b^t(\omega) - R^{t-1}(\omega) w^{t-1} , \\ & w^t(\omega) \geq 0 , \end{aligned} \tag{4.11}$$

and

$$\begin{aligned} Q^t_y(w^{t-1}, \xi^t(\omega)) = \min\, & q^t(\omega) y^t(\omega) \\ \text{s. t. } & T^t y^t(\omega) = d^t(\omega) - S^{t-1}(\omega) w^{t-1} , \\ & y^t(\omega) \geq 0 . \end{aligned} \tag{4.12}$$

- The multistage stochastic linear program with a finite number of possible future scenarios has a deterministic equivalent linear program.
- The extensive form is not readily accessible to manipulations such as the factorizations for extreme or interior point methods.
- Generally, some special structure is required for efficient solution in the general case and require exponential effort in the horizon H for provably tight approximations with high probability.

Approximation approaches

1. value function approximation: replacing Q^t with some simplified representation, such as an outer or inner linearization;
2. Constraint relaxation and dualization: relaxing constraints into a Lagrangian or looking at dual forms that may not be implementable but may give bounds or guidelines for implementable policies;
3. policy restriction: restricting the set of alternative actions to a simplified form that allows for efficient computation;
4. time, state, and path aggregation or scenario generation and reduction: starting with a large set of possibilities and then combining (or selecting) them to form more tractable representations;
5. Monte Carlo methods: sampling to obtain smaller, more tractable representations.

Nested Decomposition Procedures

- Nested decomposition procedures were proposed for deterministic models
- These approaches are essentially inner linearizations that treat all previous periods as subproblems to a current period master problem.
- The previous periods generate columns that can be used by the current period master problem.
- A difficulty with these primal nested decomposition or inner linearization methods is that the set of inputs may be fundamentally different for different last period realizations.
- Because the number of last period realizations is the total number of scenarios in the problem, these procedures are not well adapted to the bunching procedures.
- Some success has been achieved by applying inner linearization to the dual, which is again outer linearization of the primal problem.
- The general primal approach is to use an outer linearization built on the two-stage L -shaped method.

The basic idea

- The basic idea of the nested L -shaped or Benders decomposition method is to place cuts on $Q^{t+1}(x^t)$ and to add other cuts to achieve an x^t that has a feasible completion in all descendant scenarios.
- The cuts represent successive linear approximations of Q^{t+1} .
- Due to the polyhedral structure of Q^{t+1} , this process converges to an optimal solution in a finite number of steps.
- In general, for every stage $t = 1, \ldots, H-1$ and each scenario at that stage, $k = 1, \ldots, \mathcal{K}^t$, we have the following master problem, which generates cuts to stage $t-1$ and proposals for stage $t+1$
- $a(k)$ is the ancestor scenario of k at stage $t-1$, $x_{a(k)}^{t-1}$ is the current solution from that scenario, and where for $t = 1$, we interpret $b = h^1 - T^0 x^0$ as initial conditions of the problem

$$\min\ (c_k^t)^T x_k^t + \theta_k^t \tag{1.1}$$

$$\text{s. t. } W^t x_k^t = h_k^t - T_k^{t-1} x_{a(k)}^{t-1}\ , \tag{1.2}$$

$$D_{k,j}^t x_k^t \geq d_{k,j}^t\ , \qquad j = 1, \ldots, r_k^t\ , \tag{1.3}$$

$$E_{k,j}^t x_k^t + \theta_k^t \geq e_{k,j}^t\ , \qquad j = 1, \ldots, s_k^t\ , \tag{1.4}$$

$$x_k^t \geq 0\ , \tag{1.5}$$

Nested L -Shaped Method for Multistage Stochastic Linear Programs

Refer to the stage H problem in which θ_k^H and constraints (1.3) and (1.4) are not present.
To designate the period and scenario of the problem in (1.1)-(1.5), we also denote this subproblem, NLDS(t, k) .

Step 0 Set $t = 1$, $k = 1$, $r_k^t = s_k^t = 0$, add the constraint $\theta_k^t = 0$ to (1.1)-(1.5) for all t and k , and let **DIR = FORE** . Go to Step 1.

Step 1 Solve the current problem, NLDS(t, k) . If infeasible and $t = 1$, then stop; problem is infeasible. If infeasible and $t > 1$, then let $r_{a(k)}^{t-1} = r_{a(k)+1}^{t-1}$ and let DIR = BACK .

Step 1(continued) Let the infeasibility condition be obtained by a dual basic solution, $\pi_k^t, \rho_k^t \geq 0$, such that $(\pi_k^t)^T W^t + (\rho_k^t)^T D_k^t \leq 0$ but $(\pi_k^t)^T (h_k^t - T_k^{t-1} x_{a(k)}^{t-1}) + (\rho_k^t)^T d_k^t > 0$. Let $D_{a(k),r_{a(k)}^{t-1}}^{t-1} = (\pi_k^t)^T T_k^{t-1}, d_{a(k),r_{a(k)}^{t-1}}^{t-1} = \pi_k^t h_k^t + (\rho_k^t)^T d_k^t$. Let $t = t - 1$, $k = a(k)$ and return to Step 1. If feasible, update the values of x_k^t, θ_k^t, and store the value of the complementary basic dual multipliers on constraints (1.2)-(1.4) as $(\pi_k^t, \rho_k^t, \sigma_k^t)$, respectively. If $k < \mathcal{K}^t$, let $k = k + 1$, and return to Step 1. Otherwise, ($k = \mathcal{K}^t$), if $t = 1$, set $\mathbf{DIR = FORE}$; if $\mathbf{DIR = FORE}$ and $t < H$, let $t = t + 1$ and return. If $t = H$, let $DIR = BACK$. Go to Step 2.

Step 2 If $t = 1$, let $t = t + 1$, $k = 1$ and go to Step 1. Otherwise, for all scenarios $j = 1, \dots, \mathcal{K}^{t-1}$ at $t - 1$, compute

$$E_j^{t-1} = \sum_{k \in \mathcal{D}^t(j)} \frac{p_k^t}{p_j^{t-1}} (\pi_k^t)^T T_k^{t-1}$$

and

$$e_j^{t-1} = \sum_{k \in \mathcal{D}^t(j)} \frac{p_k^t}{p_j^{t-1}} [(\pi_k^t)^T h_k^t + \sum_{i=1}^{r_k^t} (\rho_{ki}^t)^T d_{ki}^t + \sum_{i=1}^{s_k^t} (\sigma_{ki}^t)^T e_{ki}^t].$$

The current conditional expected value of all scenario problems in $\mathcal{D}^t(j)$ is $\overline{\theta}_j^{t-1} = e_j^{t-1} - E_j^{t-1} x_j^{t-1}$.

Nested L -Shaped Method for Multistage Stochastic Linear Programs

Step 2 (Continued) If the constraint $\theta_j^{t-1} = 0$ appears in NLDS$(t-1, j)$, then remove it, let $s_j^{t-1} = 1$, and add a constraint (1.4) with E_j^{t-1} and e_j^{t-1} to NLDS$(t-1, j)$. If $\overline{\theta}_j^{t-1} > \theta_j^{t-1}$, then let $s_j^{t-1} = s_{j+1}^{t-1}$ and add a constraint (1.4) with E_j^{t-1} and e_j^{t-1} to NLDS$(t-1, j)$. If $t = 2$ and no constraints are added to NLDS(1) ($j = \mathcal{K}^1 = 1$), then stop with x_1^1 optimal. Otherwise, let $t = t - 1, k = 1$. Ift $= 1$, let **DIR = FORE** . Go to Step 1.

Theorem

If all Ξ^t are finite and all x^t have finite upper bounds, then the nested L -shaped method converges finitely to an optimal solution of (4.1)

Example

- Suppose we are planning production of air conditioners over a three month period.
- In each month, we can produce 200 air conditioners at a cost of $100 each.
- We may also use overtime workers to produce additional air conditioners if demand is heavy, but the cost is then $300 per unit.
- We have a one-month lead time with our customers, so that we know that in Month 1, we should meet a demand of 100.
- Orders for Months 2 and 3 are, however, random, depending heavily on relatively unpredictable weather patterns.
- We assume this gives an equal likelihood in each month of generating orders for 100 or 300 units.

- We can store units from one month for delivery in a subsequent month, but we assume a cost of $50 per unit per month for storage.
- We assume also that all demand must be met.
- Our overall objective is to minimize the expected cost of meeting demand over the next three months.
- We assume that the season ends at that point and that we have no salvage value or disposal cost for any leftover items. This resolves the end-of-horizon problem here.

Variable:

- x_k^t : the regular-time production in scenario k at month t ,
- y_k^t : the number of units stored from scenario k at month t ,
- w_k^t : the overtime production in scenario k at month t ,
- d_k^t : the demand for month t under scenario k .

The Model

$$
\begin{aligned}
\min\ & x^1 + 3.0w^1 + 0.5y^1 + \sum_{k=1}^{2} p_k^2 (x_k^2 + 3.0w_k^2 + 0.5y_k^2) \\
& + \sum_{k=1}^{4} p_k^3 (x_k^3 + 3.0w_k^3) \\
\text{s. t.}\quad & x^1 \le 2\,, \\
& x^1 + w^1 - y^1 = 1\,, \\
& y^1 + x_k^2 + w_k^2 - y_k^2 = d_k^2\,, \\
& x_k^2 \le 2\,, \quad k = 1,2\,, \\
& y_{a(k)}^2 + x_k^3 + w_k^3 - y_k^3 = d_k^3\,, \\
& x_k^3 \le 2\,, \quad k = 1,\ldots,4\,, \\
& x_k^t, y_k^t, w_k^t \ge 0\,, \quad k = 1,\ldots,\mathcal{K}^t\,, \quad t = 1,2,3\,,
\end{aligned} \tag{1.7}
$$

where $a(k) = 1$, if $k = 1,2$ at period 3, $a(k) = 2$ if $k = 3,4$ at period 3, $p_k^2 = 0.5$, $k = 1,2$, $p_k^3 = 0.25$, $k = 1,\ldots,4$, $d_1^2 = 1$, $d_2^2 = 3$, and $d^3 = (1,3,1,3)^T$.

Step 0. All subproblems NLDS(t, k) have the explicit $\theta_k^t = 0$ constraint.
DIR = FORE .

Iteration 1:

Step 1. Here $t = 1$, $k = 1$. The subproblem $NLDS(1,1)$ is:

$$\begin{aligned} \min\ & x^1 + 3w^1 + 0.5y^1 + \theta^1 \\ \text{s. t.}\ & x^1 \le 2\,, \\ & x^1 + w^1 - y^1 = 1\,, \\ & x^1, w^1, y^1 \ge 0\,, \\ & \theta^1 = 0\,, \end{aligned} \tag{1.8}$$

which has the solution $x^1 = 1$; other variables are zero.

Step 1. Now, $t = 2$, $k = 1$, and $NLDS(2,1)$ is

$$\begin{aligned} \min\ & x_1^2 + 3w_1^2 + 0.5y_1^2 + \theta_1^2 \\ \text{s. t.}\ & x_1^2 \le 2\,, \\ & x_1^2 + w_1^2 - y_1^2 = 1\,, \\ & x_1^2, w_1^2, y_1^2 \ge 0\,, \\ & \theta_1^2 = 0\,, \end{aligned} \tag{1.9}$$

which has the solution, $x_1^2 = 1$; other variables are zero.

Step 1. Here, $t=2$, $k=2$, and $NLDS(2,2)$ is

$$\begin{aligned} \min\ & x_2^2+3w_2^2+0.5y_2^2+\theta_2^2 \\ \text{s. t.}\ & x_2^2 \le 2\,, \\ & x_2^2+w_2^2-y_2^2=3\,, \\ & x_2^2,w_2^2,y_2^2 \ge 0\,, \\ & \theta_2^2=0\,, \end{aligned} \tag{1.10}$$

which has the solution, $x_2^2=2$, $w_2^2=1$; other variables are zero.

Step 1. Next, $t=3$, $k=1$. $NLDS(3,1)$ is

$$\begin{aligned} \min\ & x_1^3+3w_1^3+0.5y_1^3+\theta_1^3 \\ \text{s. t.}\ & x_1^3 \le 2\,, \\ & x_1^3+w_1^3-y_1^3=1\,, \\ & x_1^3,w_1^3,y_1^3 \ge 0\,, \\ & \theta_1^3=0\,, \end{aligned} \tag{1.11}$$

which has the solution, $x_1^3=1$; other primal variables are zero. The complementary basic dual solution is $\pi_1^3=(0,1)^T$.

Step 1. Next, $t = 3$, $k = 2$. $NLDS(3,2)$ has the same form as $NLDS(3,1)$, except we replace the second constraint with $x_2^3 + w_2^3 - y_2^3 = 3$. It has the solution, $x_2^3 = 2$, $w_2^3 = 1$; other primal variables are zero. The complementary basic dual solution is $\pi_2^3 = (-2,3)^T$.

Step 1. For $t = 3$, $k = 3$, we have the same subproblem and solution as $t = 3$, $k = 1$, so $x_3^3 = 1$; other primal variables are zero. The complementary basic dual solution is $\pi_3^3 = (0,1)^T$.

Step 1. For $t = 3$, $k = 4$, we have the same subproblem and solution as $t = 3$, $k = 2$, $x_4^3 = 2$, $w_4^3 = 1$; other primal variables are zero. The complementary basic dual solution is $\pi_4^3 = (-2,3)^T$. Now, $DIR = BACK$, and we go to Step 2.

Iteration 2:

Step 2. For scenario $j=1$ and $t-1=2$, we have

$$
\begin{aligned}
E_{11}^2 &= \left(\frac{0.25}{0.5}\right)(\pi_1^3 T_1^2 + \pi_2^3 T_2^2) \\
&= (0.5)\,(0\ 1)\begin{pmatrix}0 & 0 & 0\\ 0 & 0 & 1\end{pmatrix} + (0.5)\,(-2\ 3)\begin{pmatrix}0 & 0 & 0\\ 0 & 0 & 1\end{pmatrix} \\
&= (0\ 0\ 2) \qquad (1.12)
\end{aligned}
$$

and

$$
\begin{aligned}
e_{11}^2 &= \left(\frac{0.25}{0.5}\right)(\pi_1^3 h_1^3 + \pi_2^3 h_2^3) \\
&= (0.5)\,(0\ 1)\begin{pmatrix}2\\ 1\end{pmatrix} + (0.5)\,(-2\ 3)\begin{pmatrix}2\\ 3\end{pmatrix} \\
&= 3\,, \qquad (1.13)
\end{aligned}
$$

which yields the constraint, $2y_1^2 + \theta_1^2 \geq 3$, to add to $NLDS(2,1)$.

For scenario $j=2$ at $t-1=2$, we have the same, $E_{21}^2 = (0\ 0\ 2)$, $e_{21}^2 = 3$. Now $t=2$ and $k=1$.

Step 1. $NLDS(2,1)$ is now:

$$\begin{aligned} \min\ & x_1^2 + 3w_1^2 + 0.5y_1^2 + \theta_1^2 \\ \text{s. t.}\ & x_1^2 \le 2\,, \\ & x_1^2 + w_1^2 - y_1^2 = 1\,, \\ & 2y_1^2 + \theta_1^2 \ge 3\,, \\ & x_1^2, w_1^2, y_1^2 \ge 0\,, \end{aligned} \tag{1.14}$$

which has an optimal basic feasible solution, $x_1^2 = 2$, $y_1^2 = 1$, $\theta_1^2 = 1$, $w_1^2 = 0$, with complementary dual values, $\pi_1^2 = (-0.5, 1.5)^T$, $\sigma_{11}^2 = 1$.

Step 1. $NLDS(2,2)$ has the same form as (1.14) except that the demand constraint is $x_2^2 + w_2^2 - y_2^2 = 3$. The optimal basic feasible solution found to this problem is $x_2^2 = 2$, $w_2^2 = 1$, $\theta_2^2 = 3$, $y_2^2 = 0$, with complementary dual values, $\pi_2^2 = (-2, 3)^T$, $\sigma_{11}^2 = 1$. We continue in $DIR = BACK$ to Step 2.

Step 2. For scenario $t-1=1$, we have

$$\begin{aligned} E_1^1 &= (0.5)(\pi_1^2 T_1^2 + \pi_2^2 T_2^2) \\ &= (0.5)\begin{pmatrix}-0.5 & 1.5\end{pmatrix}\begin{pmatrix}0 & 0 & 0\\ 0 & 0 & 1\end{pmatrix} + (0.5)\begin{pmatrix}-2 & 3\end{pmatrix}\begin{pmatrix}0 & 0 & 0\\ 0 & 0 & 1\end{pmatrix} \\ &= \begin{pmatrix}0 & 0 & 2.25\end{pmatrix} \end{aligned} \tag{1.15}$$

and

$$\begin{aligned} e_1^1 &= (0.5)(\pi_1^2 h_1^2 + \pi_2^2 h_2^2) + (0.5)(\sigma_{11}^2 e_{11}^2 + \sigma_{21}^2 e_{21}^2) \\ &= (0.5)\begin{pmatrix}-0.5 & 1.5\end{pmatrix}\begin{pmatrix}2\\ 1\end{pmatrix} + (0.5)\begin{pmatrix}-2 & 3\end{pmatrix}\begin{pmatrix}2\\ 3\end{pmatrix} + (0.5)((1)(3)+(1)3) \\ &= (0.5)(0.5+5+6) = 5.75, \end{aligned} \tag{1.16}$$

which yields the constraint, $2.25y^1 + \theta^1 \geq 5.75$, to add to $NLDS(1)$.

Step 1. $NLDS(1)$ is now:

$$\begin{aligned} \min\ & x^1 + 3w^1 + 0.5y^1 + \theta^1 \\ \text{s. t.}\quad & x^1 \le 2\,, \\ & x^1 + w^1 - y^1 = 1\,, \\ & 2.25y^1 + \theta^1 \ge 5.75\,, \\ & x^1, w^1, y^1 \ge 0\,, \end{aligned} \tag{1.17}$$

with optimal basis feasible solution, $x^1 = 2$, $y^2 = 1$, $w^1 = 0$, $\theta^1 = 3.5$. $DIR = FORE$.

This procedure continues through six total iterations to solve the problem. At the last iteration, we obtain $\bar{\theta}^1 = 3.75 = \theta^1$, so no new cuts are generated for Period 1. We stop with a current solution as optimal, $x^{1*} = 2$, $y^{1*} = 1$, $z^* = 2.5 + 3.75 = 6.25$. In Exercise 2, we ask the reader to generate each of the cuts.

Block Separability and Special Structure

Lagrangian-Based Methods for Multiple Stages