Winter 2008 CS567 Stochastic Linear/Integer Programming
Guest Lecturer: Xu, Huan

Class 3: Algorithms for Two-Stage Linear Recourse Problem

Computation in stochastic programs with recourse has focus on two-stage problems with finite numbers of realizations, for example, the farmer's problem introduced in class 1. As we saw in the capacity expansion model, this problem can also represent multiple stage of decisions with block separable recourse and it provides a foundation for multistage methods. Hence, the two-stage problem is our primary model for computation.

The general model is to choose some initial decision that minimizes current costs plus the expected value of future recourse actions. Notice that, with a finite number of second -stage realizations and all linear functions, we can always form the full deterministic equivalent linear program (i.e., the extensive form). However, with many realizations, this form of the problems becomes quite large. Methods that ignore the special structure of the stochastic linear program become inefficient.

1 The L-Shaped Algorithm

The main idea of the L-Shaped algorithm is to approximate the non-linear term in the objective (i.e., the recourse function). Since the nonlinear objective term involves a solution of all second-stage recourse linear programs, we want to avoid numerous function evaluations for it. We therefore use that term to build a master problem in x, but we only evaluate the recourse function exactly as a subproblem.

Consider the following Extensive Form of a stochastic program with K realizations, and p_k be the probability that the k^{th} realization happens:

$$\begin{aligned}
\text{minimize:} \quad & c^\top x + \sum_{k=1}^{K} p_k q_k^\top y_k \\
\text{such that:} \quad & Ax = b; \\
& T_k x + W y_k = h_k, \quad k = 1, \cdots, K; \\
& x \geq 0, \quad y_k \geq 0, \quad k = 1, \cdots, K.
\end{aligned}$$

The block structure of this extensive form has a L-Shape, which gives the name *L-Shaped method* for the following algorithm. Basically, this structure easily leads to a Benders decomposition or equivalently a Dantzig-Wolfe decomposition of its dual. This method has been extended in stochastic programming to take care of feasibility questions and is known as L-shaped method. It proceeds as follows:

L-Shaped Algorithm

Step 0: Set $r = s = \nu = 0$.

Step 1: Set $\nu = \nu + 1$. Solve the following LP:

$$\begin{aligned} \text{minimize:} \quad & c^\top x + \theta & (1)\\ \text{such that:} \quad & Ax = b; \\ & D_l x \geq d_l, \quad l = 1, \cdots, r; & (2)\\ & E_l x + \theta \geq e_l, \quad l = 1, \cdots, s; & (3)\\ & x \geq 0, \theta \in \mathbb{R}. \end{aligned}$$

Let (x^ν, θ^ν) be an optimal solution. If no constraint (3) is present, θ^ν is set equal to $-\infty$ and is not considered in the computation of x^ν.

Step 2: For $k = 1, \cdots, K$ solve the following LP:

$$\begin{aligned} \text{Minimize:} \quad & w' \triangleq^\top v^+ + e^\top v^- \\ \text{Subject to:} \quad & Wy + Iv^+ - Iv^- = j_k - T_k x^v, \\ & y \geq 0, \; v^+ \geq 0, \; v^- \geq 0, \end{aligned}$$

where $e^\top = (1, \cdots, 1)$ and I is the identity matrix, until for some k the optimal value $w' > 0$. In this case, let σ^ν be the associated simplex multipliers and define

$$D_{r+1} := (\sigma^\nu)^\top T_k$$

and

$$d_{r+1} := (\sigma^\nu)^\top h_k$$

to generate a constraint (called a *feasibility cut*) of type (2). Set $r := r + 1$, add to the constraint set (2), and return to Step 1. If for all k, $w' = 0$, go to Step 3.

Step 3: For $k = 1, \cdots, K$ solve the linear program:

$$\begin{aligned} \text{Minimize:} \quad & w = q_k^\top y \\ \text{Subject to:} \quad & Wy = h_k - T_k x^\nu, & (4)\\ & y \geq 0. \end{aligned}$$

Let π_k^ν be the simplex multipliers associated with the optimal solution of Problem k of type (4). Define

$$E_{s+1} := \sum_{k=1}^{K} p_k (\pi_k^\nu)^\top T_k$$

and

$$e_{s+1} := \sum_{k=1}^{K} p_k (\pi_k^\nu)^\top h_k.$$

Let $w^\nu = e_{s+1} - E_{s+1} x^\nu$. If $\theta^\nu \geq w^\nu$, stop; x^ν is an optimal solution. Otherwise, set $s := s + 1$, add to the constraint set (3) and return to Step 1.

This method approximated $\mathcal{Q}$ using an outer linearization. Two types of constraints are sequentially added :(i) feasibility cuts (2) determining $\{x | \mathcal{Q}(x) < +\infty\}$ and (ii) *optimality cuts* (3), which are linear approximations to $\mathcal{Q}$ on its domain of finiteness. We use the following example to illustrate the optimality cuts: Let

$$Q(x, \xi) = \begin{cases} \xi - x & \text{if } x \leq \xi, \\ x - \xi & \text{otherwise,} \end{cases}$$

and let $\boldsymbol{\xi}$ take on the values 1, 2 and 4 each with probability 1/3. Assume $c = 0$ and $0 \leq x \leq 10$. Assume the starting point is $x^1 = 0$. The sequence of iterations for the L-shaped method is as follows:

Iteration 1: x^1 is not optimal; send the cut

$$\theta \geq 7/3 - x.$$

Iteration 2: $x^2 = 10$, $\theta^2 = -23/4$ is not optimal; send the cut

$$\theta \geq x - 7/3.$$

Iteration 3: $x^3 = 7/3$, $\theta^3 = 0$ is not optimal; send the cut

$$\theta \geq x/3 + 1/3.$$

Iteration 4: $x^4 = 1.5$, $\theta^4 = 5/6$ is not optimal; send the cut

$$\theta \geq 5/3 - x/3.$$

Iteration 5: $x^5 = 2$, $\theta^5 = 1$ is the optimal solution.

2 Proof of Convergence: Optimality Cut

We next prove that the constraint of type (3) defined in Step 3 are supporting hyperplanes of $\mathcal{Q}(x)$, and the algorithm will converge to an optimal solution, provided the constraint (2) adequately define feasible set.

First notice the original problem

$$\begin{aligned} \text{minimize:} \quad & c^\top x + \mathcal{Q}(x) \\ \text{subject to:} \quad & x \in \mathcal{K} \end{aligned}$$

is equivalent to

$$\begin{aligned} \text{minimize:} \quad & c^\top x + \theta \\ \text{subject to:} \quad & x \in \mathcal{K} \\ & \mathcal{Q}(x) \leq \theta. \end{aligned}$$

In step 3, problem (4) is solved repeatedly for each $k = 1, \cdots K$, yielding optimal simplex multipliers π_k^ν, $k = 1, \cdots, K$. It followed from duality of LP that for each K

$$Q(x^\nu, \xi_k) = (\pi_k^\nu)^\top (h_k - T_k x^\nu).$$

Moreover, by convexity of $Q(x, \xi_k)$, it follows from the subgradient inequality that

$$Q(x, \xi_k) = (\pi_k^\nu)^\top h_k - (\pi_k^\nu)^\top T_k x.$$

Taking the expectation of these two realization we obtain

$$\mathcal{Q}(x^\nu) = \mathbb{E}(\boldsymbol{\pi}^\nu)^\top (\mathbf{h} - \mathbf{T} x^\nu) = \sum_{k=1}^{K} p_k (\pi_k^\nu)^\top (h_k - T_k x^\nu)$$

and

$$\mathcal{Q}(x) \geq \mathbb{E}(\boldsymbol{\pi}^\nu)^\top (\mathbf{h} - \mathbf{T} x) = \sum_{k=1}^{K} p_k (\pi_k^\nu)^\top h_k - \left(\sum_{k=1}^{K} p_k (\pi_k^\nu)^\top T_k \right) x.$$

By $\theta \geq \mathcal{Q}(x)$ it follows that a pair (x, θ) is feasible only if $\theta \geq \mathbb{E}(\boldsymbol{\pi}^\nu)^\top(\mathbf{h} - \mathbf{T}x)$ which corresponds to (3).

Hence all constraints (3) are supporting linear function of $\mathcal{Q}(x)$. Hence if $(x^\nu, \theta\nu)$ is optimal. then $\mathcal{Q}(x^\nu = \theta^\nu)$ because θ is unrestricted except for $\theta \geq \mathcal{Q}(x)$. This happens when $\theta^\nu = \mathbb{E}(\boldsymbol{\pi}^\nu)^\top(\mathbf{h} - \mathbf{T}x^\nu)$, which justifies the termination criterion in Step 3.

This means that at each iteration, either $\theta^\nu \geq \mathcal{Q}(x^\nu)$ implying termination or $\theta^\nu < \mathcal{Q}(x^\nu)$. In the latter case, none of the already defined optimality cuts adequately impose $\theta \geq \mathcal{Q}(x)$, so a new set of multipliers π_k^ν will be defined at x^ν to generate an appropriate constraint. The finite convergence of the algorithm follows from the fact that there is only a finite number of different combinations of the K multipliers π_k, because each corresponds to one of the finitely many different bases.

3 Proof of Convergence: Feasibility Cut

We now prove that Step 2 generate at most a finite number of constraints to guarantee x being feasible. x being feasible is to say

$$x \in \{x | \text{for } k = 1, \cdots, K, \exists y \geq 0 \,\text{s.t}\, Wy = h_k - T_k x\}.$$

This is equivalent to

$$h_k - T_k x \in \text{range}(W),^1 \quad k = 1, \cdots, K.$$

Here $range(W) \triangleq \{t | \exists y \geq 0, \text{ s.t. } t = Wy\}$. In Step 2, a subproblem is solved that tests whether $h_k - T_k x^\nu$ belongs to range(W) or not. If not, it means that for some k, $h_k - T_k x^\nu \notin$ range(W). Then, there must be a hyperplance separating $h_k - T_k x^\nu$ and range(W).

We claim that the simplex multipliers σ^ν gives the hyperplane. By duality, w' being strictly positive is the same as $(\sigma^\nu)^\top(h_k - T_k x^\nu) > 0$. Also we have $(\sigma^\nu)^\top W \leq 0$ because σ^ν is an optimal simplex multiplier, and, at the optimum the reduced costs associated with y must be non-negative.

Therefore, σ^ν gives the separating hyperplane, i.e., a necessary condition for x being feasible is that $(\sigma^\nu)^\top(h_k - T_k x) \leq 0$. There is at most a finite number of such constraints, because there are only a finite number of optimal based to the problem solved in Step 2.

In conclusion, we have the following theorem:

Theorem 1. *When $\boldsymbol{\xi}$ is a finite random variable, the L-shaped algorithm finitely converges to an optimal solution when it exists, or prove its infeasibility.*

We illustrate the feasibility cuts using the following example:

$$\begin{aligned} \text{Minimize:} \quad & 3x_1 + 2x_2 = \mathbb{E}_{\boldsymbol{\xi}}(15y_1 + 12y_2) \\ \text{Subject to:} \quad & 3y_1 + 2y_2 \leq x_1, \\ & 2y_1 + 5y_2 \leq x_2, \\ & 0.8\boldsymbol{\xi}_1 \leq y_1 \leq \boldsymbol{\xi}_1, \\ & 0.8\boldsymbol{\xi}_2 \leq y_2 \leq \boldsymbol{\xi}_2, \\ & x, y \geq 0, \end{aligned}$$

with $\boldsymbol{\xi}_1 = 4$ or 6 and $\boldsymbol{\xi}_2 = 4$ or 8 independently, with probability 1/2 each and $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)^\top$.

[1] That why we need fixed recourse

Assume the first considered realization of $\boldsymbol{\xi}$ is $(6,\, 8)^\top$. Starting from an initial solution $x^0 = (0,\, 0)^\top$, a first feasibility cut $3x_1 + x_2 \geq 123.2$ is generated. The first-stage solution is then $x^1 = (41.067, 0)^\top$. A second feasibility cut is $x_2 \geq 22.4$. The first-stage solution becomes $x^2 = (33.6, 22.4)^\top$. A third feasibility cut $x_2 \geq 41.6$ is generated. The first stage solution is

$$x^3 = (27.2,\, 41.6)^\top,$$

which yields feasible second-stage decisions.

4 Enhancement: Feasibility

Consider the previous example. A simple look at the problem reveals that for feasibility when $\xi_1 = 6$ and $\xi_2 = 8$, it is at least necessary to have $y_1 \geq 4.8$ and $y_2 \geq 6.4$, which in turn implies $x_1 \geq 27.2$ and $x_2 \geq 41.6$. Adding them in the original problem, we can consider the following program as an initial problem:

$$\begin{aligned} \text{Minimize:} \quad & 3x_1 + 2x_2 + \mathcal{Q}(x) \\ \text{Subject to:} \quad & x_1 \geq 27.2, \\ & x_2 \geq 41.6, \end{aligned}$$

which immediately appears to be feasible. Such situations frequently occur in practice. And in this section we discuss how to enhance the computational performance of the L-shaped algorithm by exploiting such properties.

For the following cases, the feasibility check process (i.e., Step 2) can be simplified.

A first case is when the second stage is always feasible, e.g. the farmer's problem. The stochastic program is then said to have complete recourse. To be more specific, if $\text{range}(W) in \mathbb{R}^{m_2}$, it is said to have *complete recourse*. If $Ax = b$, $x \geq 0$ implies $h - Tx \in \text{range}(W)$, it is said to have *relatively complete recourse*.

The second case is when it is possible to derive some constraints (often called *induced constraints*) that have to be satisfied to guarantee second-stage feasibility. For example, the previous problem.

A third case is when step 2 is not required for all $k = 1, \cdots, K$, but for one single h_k. Assume T is deterministic. Also assume that W is such that $t \in \text{range}(W)$ for all $t \geq 0$. Let $a_I = \min_k\{h_{ik}\}$ to be the componentwise minimum of h. Also assume there exists on realization h_l such that $a = h_l$. Then the second stage constraint is equivalent to $W_y = a - Tx$, $y \geq 0$.

5 Enhancement: Multicut Version

In Step 3 of the L-shaped method. all K realizations of the second-stage program are optimized to obtain their optimal simplex multipliers. These multipliers are then aggregated to generate on cut. In multicut version, one cur per realization in the second stage is placed. **L-Shaped Algorithm**

Step 0: Set $r = \nu = 0$ and $s(k) = 0$ for all $k = 1, \cdots, K$

Step 1: Set $\nu = \nu + 1$. Solve the following LP:

$$
\begin{aligned}
\text{minimize:} \quad & c^\top x + \theta \\
\text{such that:} \quad & Ax = b; \\
& D_l x \geq d_l, \quad l = 1, \cdots, r; \\
& E_{l(k)} x + \theta \geq e_{l(k)}, \quad l(k) = 1, \cdots, s(k); \\
& x \geq 0, \theta \in \mathbb{R}.
\end{aligned} \tag{5}
$$

Let $(x^\nu, \theta_1^\nu, \cdots, \theta_K^\nu)$ be an optimal solution. If no constraint (5) is present, θ^ν is set equal to $-\infty$ and is not considered in the computation of x^ν.

Step 2: As before.

Step 3: For $k = 1, \cdots, K$ solve the linear program:

$$
\begin{aligned}
\text{Minimize:} \quad & w = q_k^\top y \\
\text{Subject to:} \quad & Wy = h_k - T_k x^\nu, \\
& y \geq 0.
\end{aligned} \tag{6}
$$

Let π_k^ν be the simplex multipliers associated with the optimal solution of Problem k of type (6). If

$$\theta_k^\nu < p_k (\pi_k^\nu)^\top (h_k - T_k x^\nu), \tag{7}$$

define

$$E_{s(k)+1} := p_k (\pi_k^\nu)^\top T_k,$$

$$e_{s(k)+1} := p_k (\pi_k^\nu)^\top h_k,$$

and set $s(k) = s(k) + 1$. If (7) does not hold for any k stop; x^ν is an optimal solution. Otherwise, return to Step 1.

We illustrate this algorithm by the first example. Starting from $x^1 = 0$, the sequence of iterations is as follows:

Iteration 1: x^1 is not optimal, send the cuts

$$\theta_1 \geq \frac{1-x}{3};\ \theta_2 \geq \frac{2-x}{3};\ \theta_3 \geq \frac{4-x}{3}.$$

Iteration 2: $x^2 = 10, \theta_1^2 = -3, \theta_2^2 = -8/3, \theta_3^2 = -2$ is not optimal; send the cuts

$$\theta_1 \geq \frac{x-1}{3};\ \theta_2 \geq \frac{x-2}{3};\ \theta_3 \geq \frac{x-4}{3}.$$

Iteration 3: $x^3 = 2, \theta_1^3 = 1/3, \theta_2^3 = 0, \theta_3^3 = 2/3$ is the optimal solution.