Stochastic Optimization

Solution Methods

Alireza Ghaffari-Hadigheh

Azarbaijan Shahid Madani University (ASMU)

hadigheha@azaruniv.edu

Fall 2017

Overview

1. The L -Shaped Method
   - Basic Idea
   - L *-Shaped Algorithm*
   - Optimality Cuts
   - Feasibility cuts
   - The multicut version
2. Regularized Decomposition
   - The Regularized Decomposition Algorithm
3. The Piecewise Quadratic Form of the L *-shaped Methods*
   - Assumptions
   - Examples
   - An Algorithm
4. Bunching and Other Efficiencies
   - Full decomposability
   - Bunching

Basic Idea

- Basic two-stage stochastic linear program

$$\begin{aligned} \min\ z &= c^T x + \mathcal{Q}(x) \\ s.t. \quad & Ax = b \\ & x \geq 0, \end{aligned} \tag{1}$$

- $\mathcal{Q}(x) = E_\xi Q(x, \xi(\omega))$
- $Q(x, \xi(\omega)) = \min_y \{q(\omega)^T y | Wy = h(\omega) - T(\omega)x, y \geq 0\}$.
- The nonlinear objective term involves a solution of all second-stage recourse linear programs, we want to avoid numerous function evaluations for it.
- The basic idea: To approximate the nonlinear term in the objective.
- A master problem in x,
- Evaluate the recourse function exactly as a subproblem.

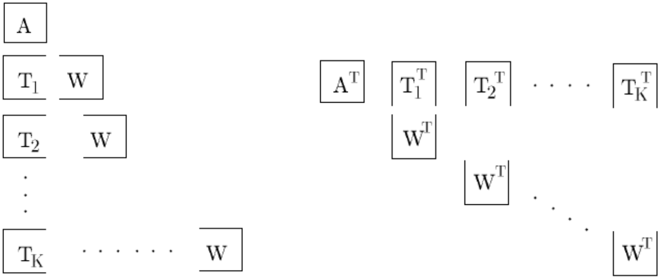
Assumption

The random vector ξ has finite support.
$k = 1, \ldots, K$ index its possible realizations
p_k are their probabilities.

The deterministic equivalent program

Associate one set of second-stage decisions, say, y_k , to each realization ξ , i.e., to each realization of q_k, h_k , and T_k .
Extensive form (EF)

$$
\begin{array}{lcll}
\min\ z & = & c^T x + \sum_{k=1}^{K} p_k q_k^T y_k & \\
s.t. & & Ax = b & \\
 & & T_k x + W y_k = h_k, & k = 1, \ldots, K \\
 & & x \geq 0, y_k \geq 0, & k = 1, \ldots, K
\end{array} \tag{2}
$$



- This picture has given rise to the name.
- Taking the dual of the extensive form, one obtains a dual block-angular structure.
- Exploit this dual structure by performing a Dantzig-Wolfe [1960] decomposition (inner linearization) of the dual or a Benders [1962] decomposition (outer linearization) of the primal.

L -Shaped Algorithm

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

Step 1 Set $\nu = \nu + 1$. Solve

$$\min \quad z = c^T x + \theta \tag{3}$$

$$\text{s.t.} \quad Ax = b,$$

$$D_\ell x \geq d_\ell, \quad \ell = 1, \ldots, r, \tag{4}$$

$$E_\ell x + \theta \geq e_\ell, \quad \ell = 1, \ldots, s, \tag{5}$$

$$x \geq 0, \theta \in \mathbb{R}.$$

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

Step 2 Check if $x \in K_2$ If not, add at least one cut (4) and return to Step 1. Otherwise, go to Step 3.

L -Shaped Algorithm

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

$$\begin{aligned} \min \quad & w = q_k^T y \\ s.t. \quad & Wy = h_k - T_k x^\nu, \\ & y \geq 0 \end{aligned} \tag{6}$$

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

$$E_{s+1} = \sum_{k=1}^{k} p_k \cdot (\pi_k^\nu)^T T_k. \tag{7}$$

$$e_{s+1} = \sum_{k=1}^{K} p_k . (\pi_k^\nu)^T h_k. \tag{8}$$

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 (5), and return to Step 1.

- The method consists of solving an approximation of (4) by using an outer linearization of Q .
- This approximation is program (3)-(5). It is called the master program.
- It consists of finding a proposal x , sent to the second stage.
- Two types of constraints are sequentially added:
  - (i) feasibility cuts (4) determining $\{x|Q(x) < +\infty\}$
  - (ii) optimality cuts (5), which are linear approximations to Q on its domain of finiteness.

Optimality cuts

Example

$$
\begin{aligned}
z \quad & = \min 100x_1 + 150x_2 + E_\xi(q_1 y_1 + q_2 y_2) \\
s.t. \quad & x_1 + x_2 \le 120, \\
& 6y_1 + 10y_2 \le 60x_1, \\
& 8y_1 + 5y_2 \le 80x_2, \\
& y_1 \le d_1, y_2 \le d_2, \\
& x_1 \ge 40, x_2 \ge 20, y_1, y_2 \ge 0
\end{aligned}
$$

$\xi^T = (d_1, d_2, q_1, q_2)$ takes on the values $(500, 100, -24, -28)$ with probability 0.4 and $(300, 300, -28, -32)$ with probability 0.6
The second stage is always feasible ($y = (0,0)^T$ is always feasible as $x \ge 0$ and $d \ge 0$).
Thus $x \in K_2$ is always true and Step 2 can be omitted.

Solution

Iteration 1:

Step 1 Ignoring θ , the master program is $z = \min\{100x_1 + 150x_2 | x_1 + x_2 \leq 120, x_1 \geq 40, x_2 \geq 20\}$ with solution $x^1 = (40, 20)^T$ and $\theta^1 = -\infty$.

Step 3

- For $\xi = \xi_1$, solve the program $w = \min\{-24y_1 - 28y_2 | 6y_1 + 10y_2 \leq 2400$, $8y_1 + 5y_2 \leq 1600, 0 \leq y_1 \leq 500, 0 \leq y_2 \leq 100\}$. The solution is $w_1 = -6100$, $y^T = (137.5, 100), \pi_1^T = (0, -3, 0, -13)$.
- For $\xi = \xi_2$, solve the program $w = \min\{-28y_1 - 32y_2 | 6y_1 + 10y_2 \leq 2400$, $8y_1 + 5y_2 \leq 1600, 0 \leq y_1 \leq 300, 0 \leq y_2 \leq 300\}$. The solution is $w^2 = -8384$, $y^T = (80, 192), \pi_2^T = (-2.32, -1.76, 0, 0)$.

- Using $h_1 = (0, 0, 500, 100)^T$ and $h_2 = (0, 0, 300, 300)^T$ in (8), $e_1 = 0.4.\pi_1^T.h_1 + 0.6.\pi_2^T.h_2 = 0.4.(-1300) + 0.6.(0) = -520.$
- The matrix T is identical in the two scenarios.

$$\begin{bmatrix} -60 & 0 \\ 0 & -80 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Thus, (7) gives $E_1 = 0.4 \,.\, \pi_1^T T + 0.6 \,.\, \pi_2^T T$
$= 0.4(0, 240) + 0.6(139.2, 140.8) = (83.52, 180.48).$

- Finally, as $x^1 = (40, 20)^T$,
$w^1 = -520 - (83.52, 180.48).x^1 = -7470.4.$
Thus, $w_1 = -7470.4 > \theta^1 = -\infty$, add the cut

$$83.52x_1 + 180.48x_2 + \theta \geq -520.$$

Iteration 2:

Step 1 . Solve $z = \min\{100x_1 + 150x_2 + \theta | x_1 + x_2 \leq 120,$
$x_1 \geq 40, x_2 \geq 20, 83.52x_1 + 180.48x_2 + \theta \geq -520\}$
with solution
$z = -2299.2, x^2 = (40, 80)^T, \theta^2 = -18299.2$.

Step 3
- For $\xi = \xi_1$ the program
$w = \min\{-24y_1 - 28y_2 | 6y_1 + 10y_2 \leq 2400,$
$8y_1 + 5y_2 \leq 6400, 0 \leq y_1 \leq 500, 0 \leq y_2 \leq 100\}$
has solution $w_1 = -9600,$
$y^T = (400, 0), \pi_1^T = (-4, 0, 0, 0)^T$.
- For $\xi = \xi_2$ the program
$w = \min\{-28y_1 - 32y_2 | 6y_1 + 10y_2 \leq 2400,$
$8y_1 + 5y_2 \leq 6400, 0 \leq y_1 \leq 300, 0 \leq y_2 \leq 300\}$
has solution: $w_2 = -10320, y^T = (300, 60), \pi_2^T =$
$(-3.2, 0, -8.8, 0)$.

- Apply (7) and (8),

$$e_1 = 0.4.(0) + 0.6.(-2640) = -1584.$$
$$E_1 = 0.4 . (240, 0) + 0.6 . (192, 0) = (211.2, 0).$$

- As $w_2 = -1584 - 211.2 \times 40 = -10032 > -18299.2$, add the cut $211.2x_1 + \theta \geq -1584$.

Iteration 3:

Step 1 Master program has solution $z = -1039.375$, $x^3 = (66.828, 53.172)^T$, $\theta^3 = -15697.994$.

Step 3 Add the cut $115.2x_1 + 96x_2 + \theta \geq -2104$.

Iteration 4:

Step 1 Master program has solution $z = -889.5$, $x^4 = (40, 33.75)^T$, $\theta^4 = -9952$.

Step 3 The second-stage program for $\xi = \xi_2$ has multiple solutions. Selecting one of them, add the cut $133.44x_1 + 130.56x_2 + \theta \geq 0$

Iteration 5:

Step 1 Solve first stage program
$z = \min\{100x_1 + 150x_2 + \theta | x_1 + x_2 \leq 120, x_1 \geq 55, x_2 \geq 25, 83.52x_1 + 180.48x_2 + \theta \geq -520, 211.2x_1 + \theta \geq -1584, 115.2x_1 + 96x_2 + \theta \geq -2104, 133.44x_1 + 130.56x_2 + \theta \geq 0\}$. It has solution $z = -855.833$, $x^5 = (46.667, 36.25)^T$, $\theta^5 = -10960$.

Step 3
- For $\xi = \xi_1$ the program $w = \min\{-24y_1 - 28y_2 | 6y_1 + 10y_2 \leq 2800, 8y_1 + 5y_2 \leq 2900, 0 \leq y_1 \leq 500, 0 \leq y_2 \leq 100\}$ has the solution $w_1 = -10000$, $y^T = (300, 100)$, $\pi_1^T = (0, -3, 0, -13)$.
- For $\xi = \xi_2$ the program $w = \min\{-28y_1 - 32y_2 | 6y_1 + 10y_2 \leq 2800, 8y_1 + 5y_2 \leq 2900, 0 \leq y_1 \leq 300, 0 \leq y_2 \leq 300\}$ has the solution $w_2 = -11600$, $y^T = (300, 100)$, $\pi_2^T = (-2.32, -1.76, 0, 0)$.

- Apply formulaes (7) and (8) to obtain

$$\begin{aligned} e_5 &= 0.4 \times (-1300) + 0.6 \times (0) = -520, \\ E_5 &= 0.4 \cdot (0, 240) + 0.6 \cdot (139.2, 140.8) = (83.52, 180.48). \end{aligned}$$

As $w_5 = -520 - (83.52, 180.48) \cdot x^5 = -10960 = \theta^5$, stop. $x^5 = (46.667, 36.25)^T$ is the optimal solution.

▶ This example is small, it is easy to write down the extensive form and solve it with an LP-solver to check whether $(46.667, 36.25)^T$ is the optimal solution.

- The second-stage program for $\xi = \xi_2$ at Iteration 4 has multiple solutions. An alternative cut is $165.12x_1 + 46.08x_2 + \theta \geq -1584$.

Example

$$\begin{aligned} z &= \min E_{\xi}(y_1 + y_2) \\ s.t. & \quad 0 \leq x \leq 10, \\ & \quad y_1 - y_2 = \xi - x, \\ & \quad y_1, y_2 \geq 0, \end{aligned}$$

- ξ takes the values 1 , 2 and 4 with probability $\frac{1}{3}$ each.
- $h = \xi$, $T = [1]$ and x are all scalars.
- Step 2 can be omitted.

Iteration 1.
Take $x^1 = 0$ as starting point.

Step 3

- For $\xi = \xi_1$, solve the program $w = \min\{y_1 + y_2 | y_1 - y_2 = 1, y_1, y_2 \geq 0\}$. The solution is $w_1 = 1$, $y^T = (1, 0)$, $\pi_1 = (1)$.
- For $\xi = \xi_2$, solve the program $w = \min\{y_1 + y_2 | y_1 - y_2 = 2, y_1, y_2 \geq 0\}$. The solution is $w_2 = 2$, $y^T = (2, 0)$, $\pi_2 = (1)$.
- For $\xi = \xi_3$, solve the program $w = \min\{y_1 + y_2 | y_1 - y_2 = 4, y_1, y_2 \geq 0\}$. The solution is $w_3 = 4$, $y^T = (4, 0)$, $\pi_3 = (1)$.
- Using $h_k = \xi_k$, one obtains $e_1 = \frac{1}{3} \cdot 1 \cdot (1 + 2 + 4) = \frac{7}{3}$. Formula (7) gives $E_1 = \frac{1}{3} \cdot 1 \cdot (1 + 1 + 1) = 1$. Finally, as $x^1 = (0)$, $w^1 = \frac{7}{3} > -\infty$; add the cut, $\theta \geq \frac{7}{3} - x$.

Iteration 2:

Step 1 $x^2 = 10$,

Step 3 . x^2 is not optimal; add the cut $\theta \geq x - \frac{7}{3}$

Iteration 3:

Step 1 $x^3 = \frac{7}{3}$,

Step 3 . x^3 is not optimal; add the cut $\theta \geq \frac{x-1}{3}$

Iteration 4:

Step 1 $x^4 = 1.5$,

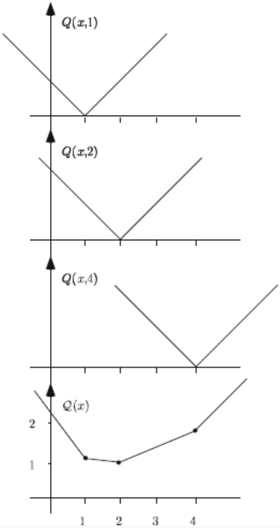
Step 3 . x^4 is not optimal; add the cut $\theta \geq \frac{5-x}{3}$

Iteration 5:

Step 1 $x^5 = 2$,

Step 3 . x^5 is optimal.

- These cuts are supporting hyperplanes of $\mathcal{Q}(x)$.



- $\mathcal{Q}(x) = E_\xi Q(x,\xi) = \sum_{k=1}^{K} p_k Q(x,\xi_k)$,
- $Q(x,\xi) = \min\{y_1 + y_2 | y_1 - y_2 = \xi - x, y_1, y_2 \geq 0\}$.
- If $x \leq \xi$, the second-stage optimal solution is $y^T = (\xi - x, 0)$ and $y^T = (0, x - \xi)$ if $x \geq \xi$.

$$Q(x,\xi) = \begin{cases} \xi - x & if \quad x \leq \xi, \\ x - \xi & if \quad x \geq \xi. \end{cases}$$

- Consider Iteration 1. $x^1 = 0$ is the starting point.
- Step 3 obtains the cut $\theta \geq \frac{7}{3} - x$.
- For $x = x^1$, $Q(x,1) = 1, Q(x,2) = 2, Q(x,4) = 4$ and $Q(x) = \frac{7}{3}$.
- Around $x = x^1$, $Q(x,1) = 1 - x, Q(x,2) = 2 - x, Q(x,4) = 4 - x$ and $Q(x) = \frac{7}{3} - x$.
- Around $x = x^1$ is simply $0 \leq x \leq 1$.
- This can be seen from the construction of $Q(x,1)$ where $Q(x,1)$ changes when $x = 1$.
- In general, such a range can be obtained by linear programming sensitivity analysis around the second stage optimal solutions.
- We conclude that $\mathcal{Q}(x) = \frac{7}{3} - x$ within $0 \leq x \leq 1$.
- The optimality cut at the end of Iteration 1 is $\theta \geq \frac{7}{3} - x$

- Step 2 of the L-shaped method consists of determining whether a first-stage decision $x \in K_1$ is also second stage feasible, i.e. $x \in K_2$.

Step 2 For $k = 1, \ldots, K$ solve the linear program

$$\begin{aligned} \min w' &= e^T \nu^+ + e^T \nu^- && (9) \\ s.t. \quad & Wy + I\nu^+ - I\nu^- = h_k - T_k x^\nu, && (10) \\ & y \geq 0, \nu^+ \geq 0, \nu^- \geq 0, \end{aligned}$$

- $e^T = (1, \ldots, 1)$, until, for some k , the optimal value $w' > 0$
- σ^ν be the associated simplex multipliers
- Define

$$\begin{aligned} D_{r+1} &= (\sigma^\nu)^T T_k && (11) \\ d_{r+1} &= (\sigma^\nu)^T h_k && (12) \end{aligned}$$

- Set $r = r + 1$, add to the constraint set (4), and return to Step 1. If for all k , $w' = 0$, go to Step 3.

Example

$$\begin{aligned} \min \quad & 3x_1 + 2x_2 - E_\xi(15y_1 + 12y_2) \\ s.t. \quad & 3y_1 + 2y_2 \le x_1, \\ & 2y_1 + 5y_2 \le x_2, \\ & 0.8\xi_1 \le y_1 \le \xi_1, \\ & 0.8\xi_2 \le y_2 \le \xi_2, \\ & x, y \ge 0, a.s., \end{aligned}$$

- $\xi_1 = 4$ or 6 and $\xi_2 = 4$ or 8 , independently, with probability $\frac{1}{2}$ each and $\xi = (\xi_1, \xi_2)^T$.

To keep the discussion short, assume the first considered realization of ξ is $(6, 8)^T$.

Starting from an initial solution $x^1 = (0,0)^T$, Program (9)-(10) reads as follows

$$\begin{aligned} w' &= \min \nu_1^+ + \nu_1^- + \nu_2^+ + \nu_2^- + \nu_3^+ + \nu_3^- \\ &\quad + \nu_4^+ + \nu_4^- + \nu_5^+ + \nu_5^- + \nu_6^+ + \nu_6^- \\ s.t. &\quad \nu_1^+ - \nu_1^- + 3y_1 + 2y_2 \leq 0, \\ &\quad \nu_2^+ - \nu_2^- + 2y_1 + 5y_2 \leq 0, \\ &\quad \nu_3^+ - \nu_3^- + y_1 \geq 4.8, \\ &\quad \nu_4^+ - \nu_4^- + y_2 \geq 6.4, \\ &\quad \nu_5^+ - \nu_5^- + y_1 \leq 6, \\ &\quad \nu_6^+ - \nu_6^- + y_2 \leq 8, \\ &\quad \nu^+, \nu^-, y \geq 0 \end{aligned}$$

- The optimal solution is $w' = 11.2$ with non-zero variables $\nu_3^+ = 4.8$ and $\nu_4^+ = 6.4$.
- The dual variables are $\sigma^1 = (-3/11, -1/11, 1, 1, 0, 0)$.
- $h = (0, 0, 4.8, 6.4, 6, 8)^T$ and T consists of the two columns $(-1, 0, 0, 0, 0, 0)^T$ and $(0, -1, 0, 0, 0, 0)^T$.
- Thus, $D_1 = (-0.273, -0.091, 1, 1, 0, 0) \cdot T = (0.273, 0.091)$, and $d_1 = (-0.273, -0.091, 1, 1, 0, 0) \cdot h = 11.2$, creating the feasibility cut $\frac{3}{11}x_1 + \frac{1}{11}x_2 \geq 11.2$ or $3x_1 + x_2 \geq 123.2$.
- The first-stage solution is then $x^2 = (41.067, 0)^T$.
- A second feasibility cut is $x_2 \geq 22.4$.
- The first-stage solution becomes $x^3 = (33.6, 22.4)^T$.
- A third feasibility cut $x^2 \geq 41.6$ is generated.
- The first-stage solution is: $x^4 = (27.2, 41.6)^T$, which yields feasible second-stage decisions.

In some cases, Step 2 can be simplified.

- When the second stage is always feasible. The stochastic program is then said to have complete recourse.
- When it is possible to derive some constraints that have to be satisfied to guarantee second-stage feasibility. These constraints are sometimes called induced constraints. They can be obtained from a good understanding of the model.
- When Step 2 is not required for all $k = 1, \dots, K$, but for one h_k.

Theorem

When ξ is a finite random variable, the L -shaped algorithm finitely converges to an optimal solution when it exists or proves the infeasibility of Problem

$$\begin{aligned} \min \quad & c^T x + \mathcal{Q}(x) \\ s.t. \quad & x \in K_1 \cap K_2. \end{aligned}$$

- 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 aggregated in (11) and (12) to generate one cut (5).
- In the multicut version, one cut per realization in the second stage is placed.
- Adding multiple cuts at each iteration corresponds to including several columns in the master program of an inner linearization algorithm.

The Multicut L -Shaped Algorithm

Step 0 . Set $r = \nu = 0$ and $s_k = 0$ for all $k = 1, \dots, K$.

Step 1 Set $\nu = \nu + 1$. Solve the linear program (13)-(16):

$$\begin{aligned} \min z \quad &= \quad c^T x + \sum_{k=1}^{K} \theta_k && (13)\\ s.t. \quad & \quad Ax = b, && (14)\\ & \quad D_\ell x \geq d_\ell, \ell = 1, \dots, r, && (15)\\ & \quad E_{\ell(k)} x + \theta_k \geq e_{\ell(k)}, \ell(k) = 1, \dots, s_k, && (16)\\ & \quad x \geq 0, k = 1, \dots, K, \end{aligned}$$

Let $(x^\nu, \theta_1^\nu, \dots, \theta_K^\nu)$ be an optimal solution of (13)-(16). If no constraint (16) is presented for some k , θ_k^ν is set equal to $-\infty$ and is not considered in the computation of x^ν.

Step 2 As before.

Step 3 For $k = 1, \dots, K$ solve the linear program (10).

Let π_k^ν be the simplex multipliers associated with the optimal solution of problem k . If

$$\theta_k^\nu < p_k(\pi_k^\nu)^T(h_k - T_k x^\nu), \tag{17}$$

define

$$E_{s_k+1} = p_k(\pi_k^\nu)^T T_k, \tag{18}$$
$$e_{s_k+1} = p_k(\pi_k^\nu)^T h_k, \tag{19}$$

and set $s_k = s_k + 1$. If (17) does not hold for any $k = 1, \dots, K$, stop; x^ν is an optimal solution. Otherwise, return to Step 1.

illustration on Example in Page 16. Starting from $x^1 = 0$,

- Iteration 1: x^1 is not optimal, add 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: $x^2 = 10$, $\theta_1^2 = -3, \theta_2^2 = -\frac{8}{3}, \theta_3^2 = -2$ is not optimal; add 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: $x^3 = 2, \theta_1^3 = \frac{1}{3}, \theta_2^3 = 0, \theta_3^3 = \frac{2}{3}$ is the optimal solution.

- Regularized decomposition is a method that combines a multicut approach for the representation of the second-stage value function with the inclusion in the objective of a quadratic regularizing term.
- This additional term is included to avoid two classical drawbacks of the cutting plane methods.
  - Initial iterations are often inefficient.
  - Iterations may become degenerate at the end of the process.

The Regularized Decomposition Algorithm

Step 0 Set $r = \nu = 0, s_k = 0$ for all $k = 1, \ldots, K$. Select a^1 a feasible solution.

Step 1 Set $\nu = \nu + 1$. Solve the regularized master program

$$\begin{aligned} \min \quad & c^T x + \sum_{k=1}^{k} \theta_k + \frac{1}{2}\|x - a^\nu\|^2 \qquad (20) \\ s.t. \quad & Ax = b, \\ & D_\ell x \geq d_\ell, \ell = 1, \ldots, r, \\ & E_{\ell(k)} x + \theta_k \geq e_{\ell(k)}, \ell(k) = 1, \ldots, s_k, k = 1, \ldots, K, \\ & x \geq 0. \end{aligned}$$

Let (x^ν, θ^ν) be an optimal solution to (20) where $(\theta^\nu)^T = (\theta_1^\nu, \ldots, \theta_K^\nu)^T$. If $s_k = 0$ for some k , θ_k^ν is ignored in the computation. If $c^T x^\nu + e^T \theta^\nu = c^T a^\nu + \mathcal{Q}(a^\nu)$, stop; a^ν is optimal.

Step 2 As before, if a feasibility cut (4) is generated, set $a^{\nu+1} = a^{\nu}$ (null infeasible step), and go to Step 1.

Step 3 For $k = 1, \dots, K$, solve the linear subproblem (10). Compute $\mathcal{Q}_k(x^{\nu})$. If (17) holds, add an optimality cut (16) using formulas (18) and (19). Set $s_k = s_k + 1$.

Step 4 If (17) does not hold for any k , then $a^{\nu+1} = x^{\nu}$ (exact serious step); go to Step 1.

Step 5 . If $c^T x^{\nu} + \mathcal{Q}(x^{\nu}) \leq c^T a^{\nu} + \mathcal{Q}(a^{\nu})$, then $a^{\nu+1} = x^{\nu}$ (approximate serious step); go to Step 1. Else, $a^{\nu+1} = a^{\nu}$ (null feasible step), go to Step 1.

When a serious step is made, the value $\mathcal{Q}(a^{\nu+1})$ should be memorized, so that no extra computation is needed in Step 1 for the test of optimality.

Example

- Consider Exercise 1 of Section 5.1d.
- Take $a^1 = -0.5$ as a starting point. It corresponds to the solution of the problems with $\xi = \bar{\xi}$ with probability 1.
- We have $Q(a^1) = \frac{3}{8}$.

Iteration 1: Cuts $\theta_1 \geq 0$, $\theta_2 \geq -\frac{3}{4}x$ are added. Let $a^2 = a^1$.
Iteration 2: The regularized master is

$$\begin{aligned} \min \quad & \theta_1 + \theta_2 + \frac{1}{2}(x + 0.5)^2 \\ s.t. \quad & \theta_1 \geq 0, \theta_2 \geq -\frac{3}{4}x, \end{aligned}$$

with solution $x^2 = 0.25$: $\theta_1 = 0, \theta_2 = -\frac{3}{16}$. A cut $\theta_2 \geq 0$ is added. As $\mathcal{Q}(0.25) = 0 < \mathcal{Q}(a^1), a^3 = 0.25$ (approximate serious step 1).

Iteration 3: The regularized master is

$$\begin{aligned} \min \quad & \theta_1 + \theta_2 + \frac{1}{2}(x - 0.25)^2 \\ s.t. \quad & \theta_1 \geq 0, \theta_2 \geq -\frac{3}{4}x, \theta_2 \geq 0, \end{aligned}$$

with solution $x^3 = 0.25$, $\theta_1 = 0$, $\theta_2 = 0$. Because $\theta^\nu = \mathcal{Q}(a^\nu)$, a solution is found.

Two-stage quadratic stochastic programs

$$\begin{aligned}
\min \quad & z(x) = c^T x + \frac{1}{2} x^T C x \\
& \qquad + E_\xi[\min[q^T(\omega)y(\omega) + \frac{1}{2} y^T(\omega)D(\omega)y(\omega)]] \\
s.t. \quad & Ax = b, \qquad T(\omega)x + Wy(\omega) = h(\omega), \\
& x \geq 0, \qquad y(\omega) \geq 0
\end{aligned} \tag{21}$$

- c, C, A, b , and W are fixed matrices of size $n_1 \times 1$, $n_1 \times n_1$, $m_1 \times n_1$, $m_1 \times 1$, and $m_2 \times n_2$, respectively
- q, D, T, and h are random matrices of size $n_2 \times 1$, $n_2 \times n_2$, $m_2 \times n_1$, and $m_2 \times 1$.
- The random vector ξ is obtained by piecing together the random components of q, D, T , and h .

Assumption 1

The random vector ξ has a discrete distribution.

Assumption 2

The matrix C is positive semi-definite and the matrices $D(\omega)$ are positive semi-definite for all ω . The matrix W has full row rank.

- The first assumption guarantees the existence of a finite decomposition of the second-stage feasibility set K_2 .
- The second assumption guarantees that the recourse functions are convex and well-defined.

Recourse function for a given $\xi(\omega)$

$$\begin{aligned} Q(x,\xi(\omega)) \quad = \quad & \min\{q^T(\omega)y(\omega) + \frac{1}{2}y^T(\omega)D(\omega)y(\omega)| \\ & T(\omega)x + Wy(\omega) = h(\omega), y(\omega) \geq 0\}, \end{aligned} \tag{22}$$

Example

$$
\begin{array}{ll}
\min & z(x) = 2x_1 + 3x_2 + E_\xi \min -6.5y_1 - 7y_2 + \frac{y_1^2}{2} + y_1y_2 + \frac{y_2^2}{2} \\
s.t. & 3x_1 + 2x_2 \le 15, y_1 \le x_1, y_2 \le x_2 \\
& x_1 + 2x_2 \le 8, y_1 \le \xi_1, y_2 \le \xi_2 \\
& x_1 + x_2 \ge 0, x_1, x_2 \ge 0, y_1, y_2 \ge 0.
\end{array}
$$

- This problem consists of finding some product mix (x_1, x_2) that satisfies some first-stage technology requirements.
- In the second stage, sales cannot exceed the first-stage production and the random demand.
- In the second stage, the objective is quadratic convex because the prices are decreasing with sales.
- We might also consider financial problems where minimizing quadratic penalties on deviations from a mean value leads to efficient portfolios.

- ξ_1 can take the three values 2, 4, and 6 with probability $\frac{1}{3}$,
- ξ_2 can take the values 1, 3, and 5 with probability $\frac{1}{3}$,
- ξ_1 and ξ_2 are independent of each other.
- For very small values of x_1 and x_2 , it always is optimal in the second stage to sell the production, $y_1 = x_1$ and $y_2 = x_2$. $0 \leq x_1 \leq 2$ and $0 \leq x_2 \leq 1$, $y_1 = x_1, y_2 = x_2$ is the optimal solution of the second stage for all ξ .
- If needed, the reader may check this using the Karush-Kuhn-Tucker conditions.
- $Q(x, \xi) = -6.5x_1 - 7x_2 + \frac{x_1^2}{2} + x_1x_2 + \frac{x_2^2}{2}$ for all ξ and $\mathcal{Q}(x) = -6.5x_1 - 7x_2 + \frac{x_1^2}{2} + x_1x_2 + \frac{x_2^2}{2}$.
- Here, the cell is $\{(x_1, x_2) | 0 \leq x_1 \leq 2, 0 \leq x_2 \leq 1\}$. Within that cell, $\mathcal{Q}(x)$ is quadratic.

Definition

A finite closed convex complex $\mathcal{K}$ is a finite collection of closed convex sets, called the cells of $\mathcal{K}$, such that the intersection of two distinct cells has an empty interior.

Definition

A piecewise convex program is a convex program of the form $\inf\{z(x)|x \in S\}$ where f is a convex function on $\mathbb{R}^n$ and S is a closed convex subset of the effective domain of f with nonempty interior.
The region where f is finite is called the effective domain of f ($dom\ f$).

Assumption

Let $\mathcal{K}$ be a finite closed convex complex such that

(a) the n -dimensional cells of $\mathcal{K}$ cover S ,

(b) either f is identically $-\infty$ or for each cell C_ν of the complex there exists a convex function $z_\nu(x)$ defined on S and continuously differentiable on an open set containing C_ν which satisfies

- $z(x) = z_\nu(x) \forall x \in C_\nu$,
- $\nabla z_\nu(x) \in \partial z(x) \forall x \in C_\nu$.

Definition

A piecewise quadratic function is a piecewise convex function where on each cell C_ν the function z_ν is a quadratic form.

Initialization Let $S_1 = S$, $x^0 \in S$, $\nu = 1$.

Step 1 Obtain C_ν , a cell of the decomposition of S containing $x^{\nu-1}$. Let $z_\nu(.)$ be the quadratic form on C_ν .

Step 2 Let $x^\nu \in argmin\{z_\nu(x)|x \in S_\nu\}$ and $w_\nu \in argmin\{z_\nu(x)|x \in C_\nu\}$. If $w_\nu$is the limiting point of a ray on which $z_\nu(x)$ is decreasing to $-\infty$, the original PQP is unbounded and the algorithm terminates.

Step 3 If

$$\nabla^T z_\nu(w^\nu)(x^\nu - w^\nu) = 0, \tag{23}$$

then stop; w^ν is an optimal solution.

Step 4 Let $S_{\nu+1} = S_\nu \cap \{x|\nabla^T z_\nu(w^\nu)x \leq \nabla^T z_\nu(w^\nu)w^\nu\}$. Let $\nu = \nu + 1$; go to Step 1.

- One big issue in the efficient implementation of the L -shaped method is in Step 3.
- The second-stage program (6) has to be solved K times to obtain the optimal multipliers, π_k^ν .
- For a given x^ν and a given realization k , let B be the optimal basis of the second stage.
- It is well-known from linear programming that B is a square submatrix of W such that $(\pi_k^\nu)^T = q_{k,B}^T B^{-1}, q^T - (\pi_k^\nu)^T W \geq 0, B^{-1}(h_k - T_k x^\nu) \geq 0$, where $q_{k,B}$ denotes the restriction of q_k to the selection of columns that define B .
- Important savings can be obtained in Step 3 when the same basis B is optimal for several realizations of k .
- This is especially the case when q is deterministic.
- Then, two different realizations that share the same basis also share the same multipliers π_k^ν .

assumptions

- q is deterministic.
- Define the set of possible right-hand sides in the second stage.

$$\tau = \{t | t = h_k - T_k x^\nu \text{ for some } k = 1, \ldots, K\} \tag{24}$$

- Let B be a square submatrix and $\pi^T = q_B^T B^{-1}$.
- B satisfies the optimality criterion $q^T - \pi^T W \geq 0$.
- Define a bunch as

$$Bu = \{t \in \tau | B^{-1} t \geq 0\} \tag{25}$$

  the set of possible right-hand sides that satisfy the feasibility condition.
- Thus, π is an optimal dual multiplier for all $t \in Bu$.
- By virtue of Step 2 of the L -shaped method, only feasible first-stage $x^\nu \in K_2$ are considered.
- By construction, $\tau \subseteq \text{pos}\, W = \{t | t = Wy, y \geq 0\}$.

Full decomposability

- Full decomposition of $\text{pos}\, W$ into component bases.
- Can only be done for small problems or problems with a well-defined structure.

Farming example: The second stage

$$\begin{aligned} Q(x,\xi) \quad &= \quad \min 238y_1 - 170y_2 + 210y_3 - 150y_4 - 36y_5 - 10y_6 \\ s.t. \quad & \quad y_1 - y_2 - w_1 = 200 - \xi_1 x_1, \\ & \quad y_3 - y_4 - w_2 = 240 - \xi_2 x_2, \\ & \quad y_5 + y_6 + w_3 = \xi_3 x_3, \\ & \quad y_5 + w_4 = 6000, \\ & \quad y, w \geq 0, \end{aligned}$$

w_1 to w_4 are slack variables.

- This second stage has complete recourse, so $\text{pos}\, W = \mathbb{R}^4$.
- The matrix

$$w = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

- Theoretically, $\binom{10}{4} = 210$ bases could be found.
- w_1, w_2, and w_3 are never in the basis, as they are always dominated by y_2, y_4 , and y_6, respectively.
- y_5 is always in the basis.
- y_1 or y_2 and y_3 or y_4 are always basic.

- not only is a full decomposition of pos W available, but an immediate analytical expression for the multipliers is also obtained.

$$\pi_1(\xi) = \begin{cases} 238 & \text{if } \xi_1 x_1 < 200, \\ -170 & \text{otherwise} \end{cases}$$

$$\pi_2(\xi) = \begin{cases} 210 & \text{if } \xi_2 x_2 < 240, \\ -150 & \text{otherwise} \end{cases}$$

$$\pi_3(\xi) = \begin{cases} -36 & \text{if } \xi_3 x_3 < 6000, \\ 0 & \text{otherwise} \end{cases}$$

$$\pi_4(\xi) = \begin{cases} 10 & \text{if } \xi_3 x_3 > 6000, \\ 0 & \text{otherwise} \end{cases}$$

- The decomposition is thus (1,3,5,6) , (1,3,5,10) , (1,4,5,6) , (1,4,5,10) , (2,3,5,6) , (2,3,5,10) , (2,4,5,6) , (2,4,5,10),
- The four variables in a basis are described by their indices (the index is $6 + j$ for the j -th slack variable).

- the set of possible right-hand sides in the second stage: $\tau = \{t | t = h_k - T_k x \text{ for some } k = 1, \ldots, K\}$
- Consider some k . Denote $t_k = h_k - T_k x$.
- Arbitrarily be $k = 1$, or if available, a value of k such that $h_k - T_k x = \bar{t}$, the expectation of all $t_k \in \tau$.
- Let B_1 be the corresponding optimal basis and $\pi(1)$ the corresponding vector of simplex multipliers.
- Then, $Bu(1) = \{t \in \tau | B_1^{-1} t \geq 0\}$. Let $\tau_1 = \tau \backslash Bu(1)$.
- Repeat the same operations.
- Some element of τ_1 is chosen.
- The corresponding optimal basis B_2 and its associated vector of multipliers $\pi(2)$ are formed .
- Then, $Bu(2) = \{t \in \tau 1 | B_2^{-1} t \geq 0\}$ and $\tau_2 = \tau_1 \backslash Bu(2)$.
- The process is repeated until all $t_k \in \tau$ are in one of b total bunches.

- Then, (7) and (8) are replaced by

$$E_{s+1} = \sum_{\ell=1}^{b} \pi(\ell)^T \sum_{t_k \in Bu(\ell)} p_k T_k \quad (26)$$

$$e_{s+1} = \sum_{\ell=1}^{b} \pi(\ell)^T \sum_{t_k \in Bu(\ell)} p_k h_k \quad (27)$$

- This procedure still has some drawbacks.
    - The same $t_k \in \tau$ may be checked many times against different bases.
    - A new optimization is restarted every time a new bunch is considered.
- Some savings can be obtained in organizing the work in such a way that the optimal basis in the next bunch is obtained by performing only one (or a few) dual simplex iterations from the previous one.

Example

Consider the following second stage:

$$\begin{aligned} \max \quad & 6y_1 + 5y_2 + 4y_3 + 3y_4 \\ s.t. \quad & 2y_1 + y_2 + y_3 \leq \xi_1, \\ & y_2 + y_3 + y_4 \leq \xi_2, \\ & y_1 + y_3 \leq x_1, \\ & 2y_2 + y_4 \leq x_2 \end{aligned}$$

- $\xi_1 \in \{4, 5, 6, 7, 8\}$ with equal probability 0.2 each
- $\xi_2 \in \{2, 3, 4, 5, 6\}$ with equal probability 0.2 each

- Theoretically $\binom{8}{4} = 70$ different possible bases.
- In view of the possible realizations of ξ , at most 25 different bases can be optimal.
- t^1 to t^{25}: the possible right-hand sides

$$t^1 = \begin{pmatrix} 4 \\ 2 \\ x_1 \\ x_2 \end{pmatrix}, t^2 = \begin{pmatrix} 4 \\ 3 \\ x_1 \\ x_2 \end{pmatrix}, \ldots, t^{25} = \begin{pmatrix} 8 \\ 6 \\ x_1 \\ x_2 \end{pmatrix}$$

- Consider the case where $x_1 = 3.1$ and $x_2 = 4.1$.
- Start from $\xi = \bar{\xi} = (6, 4)^T$.
- Represent a basis again by the variable indices with $4 + j$ the index of the j th slack.
- The optimal basis is $B_1 = \{1, 4, 7, 8\}$ with $y_1 = 3, y_4 = 4, w_3 = 0.1, w_4 = 0.1$, the values of the basic variables.

- The optimal dictionary associated with B_1

$$\begin{aligned} z &= 3\xi_1 + 3\xi_2 - y_2 - 2y_3 - 3w_1 - 3w_2, \\ y_1 &= \frac{1}{2}\xi_1 - \frac{1}{2}y_2 - \frac{1}{2}y_3 - \frac{1}{2}w_1, \\ y_4 &= \xi_2 - y_2 - y_3 - w_2, \\ w_3 &= 3.1 - \frac{1}{2}\xi_1 + \frac{1}{2}y_2 - \frac{1}{2}y_3 + \frac{1}{2}w_1, \\ w_4 &= 4.1 - \xi_2 - y_2 + y_3 + w_2. \end{aligned}$$

- This basis is optimal and feasible as long as $\frac{\xi_1}{2} \leq 3.1$ and $\xi_2 \leq 4.1$, which in view of the possible values of ξ amounts to $\xi_1 \leq 6$ and $\xi_2 \leq 4$, so that $Bu(1) = \{t^1, t^2, t^3, t^6, t^7, t^8, t^{11}, t^{12}, t^{13}\}$.

- Neighboring bases can be obtained by considering either $\xi_1 \geq 7$ or $\xi_2 \geq 5$.
- Let us start with $\xi_2 \geq 5$.
- This means that w_4 becomes negative and a dual simplex pivot is required in Row 4.
- This means that w_4 leaves the basis, and, according to the usual dual simplex rule, y_3 enters the basis.
- The new basis is $B_2 = \{1, 3, 4, 7\}$

$$
\begin{aligned}
z &= 3\xi_1 + \xi_2 + 8.2 - 3y_2 - 3w_1 - w_2 - 2w_4, \\
y_1 &= \frac{\xi_1}{2} - \frac{\xi_2}{2} + 2.05 - y_2 - \frac{w_1}{2} + \frac{w_2}{2} - \frac{w_4}{2}, \\
y_3 &= \xi_2 - 4.1 + y_2 - w_2 + w - 4, \\
y_4 &= 4.1 - 2y_2 - w - 4, \\
w_3 &= 5.15 - \frac{\xi_1}{2} - \frac{\xi_2}{2} + \frac{w_1}{2} + \frac{w_2}{2} - \frac{w_4}{2}.
\end{aligned}
$$

- The condition $\xi_1 - \xi_2 + 4.1 \geq 0$ always holds.
- This basis is optimal as long as $\xi_2 \geq 5$ and $\xi_1 + \xi_2 \leq 10$,
- So that $Bu(2) = \{t^4, t^5, t^9\}$.
- Neighboring bases are B_1 when $\xi_2 \leq 4$ and B_3 obtained when $w_3 < 0$, i.e., $\xi_1 + \xi_2 \geq 11$.
- This basis corresponds to w_3 leaving the basis and w_2 entering the basis.

$$
\begin{array}{ll}
B_1 = \{1,4,7,8\} & Bu(1) = \{t^1, t^2, t^3, t^6, t^7, t^8, t^{11}, t^{12}, t^{13}\} \\
B_2 = \{1,3,4,7\} & Bu(2) = \{t^4, t^5, t^9\} \\
B_3 = \{1,3,4,6\} & Bu(3) = \{t^{10}, t^{14}, t^{15}\} \\
B_4 = \{1,4,5,6\} & Bu(4) = \{t^{19}, t^{20}, t^{24}, t^{25}\} \\
B_5 = \{1,2,4,5\} & Bu(5) = \{t^{18}, t^{22}, t^{23}\} \\
B_6 = \{1,2,4,8\} & Bu(6) = \{t^{16}, t^{17}, t^{21}\} \\
B_7 = \{1,2,5,8\} & Bu(7) = \varnothing.
\end{array}
$$

Several paths are possible, as one may have chosen B_6 instead of B_2 as a second basis.