Stochastic optimization
L-shaped method

Fabian Bastin
fabian.bastin@umontreal.ca
Université de Montréal – CIRRELT – IVADO

Two-stage linear program with recourse

Summarize our formulations.

$$\min_{x \in \mathbb{R}^n_+ \mid Ax=b} \left\{ c^T x + E_\xi \left[\min_{y \in \mathbb{R}^p_+} \{ q(\xi)^T y \mid Wy = h(\xi) - T(\xi)x \} \right] \right\}$$

$$\min_{x \in \mathbb{R}^n_+ \mid Ax=b} \left\{ c^T x + E_\xi \left[v(\xi, h(\xi) - T(\xi)x) \right] \right\}$$

$$\min_{x \in \mathbb{R}^n_+ \mid Ax=b} \left\{ c^T x + E_\xi \left[Q(x, \xi) \right] \right\}$$

$$\min_{x \in \mathbb{R}^n_+} \left\{ c^T x + \mathcal{Q}(x) \mid Ax = b \right\}$$

We have studied its properties. But how to solve it?

Notations: reminder

- First-stage feasible set:

$$K_1 = \{x \in \mathbb{R}^n_+ \mid Ax = b\}.$$

- Second-stage feasible set:

$$K_2^s = \{x \mid \mathcal{Q}(x) < \infty\}.$$

We can therefore rewrite the problem as

$$\min_x \{c^T x + \mathcal{Q}(x) \mid x \in K_1 \cap K_2^s\}.$$

Deterministic equivalent

Assume now that the support of ξ is finite, and that we have S scenarios. The probability of scenario s is p_s.

The deterministic equivalent is

$$\min_{x \in X, y_1, \ldots, y_S} c^T x + p_1 q^T y_1 + p_2 q^T y_2 + \ldots p_S q^T y_S$$

s.t.

$$
\begin{aligned}
Ax & & & & & = b \\
T_1 x & + W y_1 & & & & = h_1 \\
T_2 x & & + W y_2 & & & = h_2 \\
\vdots & & & \ddots & & \\
T_S x & & & & + W y_S & = h_S \\
\end{aligned}
$$

$$x \in X, \ y_1 \in Y, \ y_2 \in Y, \ldots, y_S \in Y.$$

Properties of the deterministic equivalent

- $y_s \stackrel{def}{=} y(\omega_s)$ is the recourse action to take if the scenario s occurs.
- We obtain a linear program,. . . but very large:
    - $n + pS$ variables;
    - $m1 + mS$ constraints.
- The constraints has a very specific structure (staircase).
- Idea: use decomposition techniques.

Decomposition methods

General principle: the nonlinear term (the recourse function) implies to compute the solution of all the second-stage recourse linear programs.

Is is possible to avoid repeated evaluations of all the concerned functions in order to compute these solutions?

Idea: build a master program in x, and evaluate the objective function exactly (including all the scenarios) as a subproblem only.

Block structure

Primal:

$$\begin{pmatrix} A & & & & \\ T_1 & W & & & \\ T_2 & & W & & \\ \vdots & & & \ddots & \\ T_S & & & & W \end{pmatrix}$$

Dual:

$$\begin{pmatrix} A^T & T_1^T & T_2^T & \cdots & T_S^T \\ & W^T & & & \\ & & W^T & & \\ & & & \ddots & \\ & & & & W^T \end{pmatrix}$$

The figures have give the name to the method: L-shaped.

Main ideas

Assume for now that Ξ contains one element only. The problem becomes

$$
\min_{x,y} c^T x + q^T y
$$
$$
\text{s.t. } Ax = b,
$$
$$
Tx + Wy = h,
$$
$$
x \geq 0, \ y \geq 0.
$$

The problem can be reformulated as

$$
\min_{x} c^T x + f(x)
$$
$$
\text{s.t. } Ax = b, x \geq 0,
$$

with

$$
f(x) \stackrel{def}{=} \min_{y} \{q^T y \mid Wy = h - Tx, \ y \geq 0\}.
$$

Problem reformulation

Assume moreover that the problem is feasible and $\{x \mid Ax = b, \ x \geq 0\}$ is bounded. This implies that

$$\inf\{f(x) \mid Ax = b, \ Tx + Wy = h, \ x \geq 0\}$$

is finite and $f(x)$ can be uniformly bounded from below on K_1, say by θ.

If we explicitly know $f(x)$, we can write

$$
\begin{aligned}
\min_x \ & c^T x + f(x) \\
\text{t.q. } & Ax = b, \\
& \theta - f(x) \geq 0, \\
& x \geq 0.
\end{aligned}
$$

First ideas

Try to construct a sequence of new (additional) linear constraints in order to define a feasible $\mathcal{B}_1$ of vectors $(x_1, \ldots, x_n, \theta)^T$ in $\mathbb{R}^{n+1}$, such that, with

$$\mathcal{B}_0 = \{(x^T, \theta^T) \mid Ax = b, \ x \geq 0, \ \theta \in \mathbb{R}\},$$

the problem

$$\min_{(x,\theta) \in \mathcal{B}_0 \cap \mathcal{B}_1} c^T x + \theta$$

gives a (first-stage) solution of the initial problem and an estimation θ of the second-stage objective value.

Dual decomposition method

Step 1 Let θ_0 be a lower bound for

$$\min_y\{q^T y \mid Ax = b,\ Tx + Wy = h,\ x \geq 0,\ y \geq 0\},$$

and let $(\hat{x}, \hat{\theta})$ be a solution of

$$\min_{x,\theta}\{c^T x + \theta \mid Ax = b,\ \theta \geq \theta_0,\ x \geq 0\}.$$

Set also $\mathcal{B}_1 := \{\mathbb{R}^n \times \{\theta\} \mid \theta \geq \theta_0\}$.

Step 2 Evaluate

$$f(\hat{x}) = \min_y\{q^T y \mid Wy = h - T\hat{x},\ y \geq 0\}$$
$$= \max_\pi\{(h - T\hat{x})^T \pi \mid W^T \pi \leq q\}. \text{ (duality)}$$

We have to consider two cases.

Dual decomposition method (cont'd)

Step 2 (a) If $f(\hat{x}) = +\infty$, $\hat{x}$ is not feasible with all the second-stage constraints, there is no y satisfying the linear system.

Farkas' lemma

Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Then, exactly one of the following two statements is true:

- $\exists\, x \in \mathbb{R}^n$ such that $Ax = b$ and $x \geq 0$;
- $\exists\, u \in \mathbb{R}^m$ such that $u^T A \geq 0$ and $u^T b < 0$;

Therefore, if y is not feasible, $\exists\, u$ such that

$$W^T u \geq 0 \text{ et } (h - T\hat{x})^T u < 0.$$

or, by setting $\sigma = -u$,

$$W^T \sigma \leq 0 \text{ et } (h - T\hat{x})^T \sigma > 0.$$

Dual decomposition method (cont'd)

Now, given x feasible (for the first and second stage), there exists $y \geq 0$ s.t. $Wy = h - Tx$. For such a y, we have

$$\sigma^T(h - Tx) = \sigma^T Wy \leq 0.$$

as $W^T \sigma \leq 0$ and $y \geq 0$. Therefore

$$\sigma^T h \leq \sigma^T Tx.$$

This inequality should be valid for any feasible x (therefore excluding $\hat{x}$), and we will enforce it by adding it to the first-stage linear program: feasibility cut

$$\sigma^T(h - Tx) \leq 0.$$

Redefine $\mathcal{B}_1 := \mathcal{B}_1 \cap \{(x^T, \theta) \,|\, \sigma^T(h - Tx) \leq 0\}$.

Dual decomposition method (cont'd)

Step 2 (b) If $f(\hat{x})$ is finite, given $\hat{x}$, we have a primal optimal solution $\hat{y}$ and a dual optimal solution $\hat{\pi}$ with

$$f(\hat{x}) = (h - T\hat{x})^T \hat{\pi}$$

by the strong duality property.

Moreover, for any x, we can write

$$
\begin{aligned}
f(x) &= \sup_{\pi} \left\{ (h - Tx)^T \pi \mid W^T \pi \leq q \right\} \text{ (second stage dual)} \\
&\geq (h - Tx)^T \hat{\pi} \\
&= \hat{\pi}^T (h - Tx).
\end{aligned}
$$

Since we want θ to represent $f(x)$, we would like

$$\theta \geq \hat{\pi}^T (h - Tx).$$

Dual decomposition method (cont'd)

$\theta \geq \hat{\pi}^T(h - Tx)$ is violated at $(\hat{x}, \hat{\theta})^T$ iff $\hat{\pi}^T(h - T\hat{x}) > \hat{\theta}$. In this case we introduce the optimality cut, rejecting the suboptimal solution $(\hat{x}^T, \hat{\theta})^T$:

$$\theta \geq \hat{\pi}^T(h - Tx),$$

and we redefine $\mathcal{B}_1 := \mathcal{B}_1 \cap \{(x^T, \theta) \,|\, \theta \geq \hat{\pi}^T(h - Tx)\}$.

If $f(\hat{x}) \leq \hat{\theta}$, stop: $\hat{x}$ is a first-stage solution. Otherwise, go to Step 3

Step 3 Solve the updated problem

$$\min\{c^T x + \theta \,|\, (x^T, \theta) \in \mathcal{B}_0 \cap \mathcal{B}_1\}.$$

Return to Step 2 with the new solution $(\hat{x}, \hat{\theta})$.

Convergence

Stopping rule: $f(\hat{x}) \leq \hat{\theta}$.
Since

$$\hat{\theta} = \max_{1 \leq i \leq k} (h - T\hat{x})^T \pi^i \leq \max_{1 \leq i \leq r} (h - T\hat{x})^T \pi^i$$
$$= f(\hat{x}) \leq \hat{\theta},$$

the algorithm stops for $\hat{\hat{\theta}}$ optimal.

Proposition

If the two-stage stochastic program is solvable and
$\{x \mid Ax = b, \ x \geq 0\}$ is bounded, the dual decomposition method
gives an optimal solution after a finite number of iterations.

S scenarios

If Ξ is finite, but is not a singleton?

We can introduce feasibility and optimality cuts for all the recourse functions

$$Q_s(x) = \min \left\{ (q_s^T y_s \,|\, W y_s = h_s - T_s x, \; y_s \geq 0 \right\},$$

$s = 1, \ldots, S$. We obtain a *multi-cut* dual decomposition method

Alternatively, we can combine cuts corresponding to blocks $s = 1, \ldots, S$ with their respective probabilities.

Single-cut algorithm

For S, we can proceed as before, by remplacing $\mathcal{Q}(x)$ by θ: singe-cut. This leads to the master program

$$\min c^T x + \theta$$
$$\text{t.q. } Ax = b,$$
$$\theta \geq \mathcal{Q}(x),$$
$$x \geq 0, \ \theta \in \mathbb{R}.$$

This program has $n + 1$ variables only!

Master program

Denote by (x^*, θ^*) the optimal solution; we aim to have the constraint active on θ:

$$\theta^* = \mathcal{Q}(x^*).$$

How to do? Use of cuts.

A cut is an additional linear constraint, restraining the feasible space of the master program, but preserving the optimal solution of the original problem.

As in the simple example, two types of cuts have to be considered:

- feasibility cut;
- optimality cut.

Feasibility cuts

Solving the master problem can produce a solution $(\hat{x}, \hat{\theta})$ such that $\mathcal{Q}(\hat{x}) = +\infty$, i.e. $\hat{x} \notin K_2$.

The feasibility cuts aim to keep the solutions of the master program in $K_1 \cap K_2$.

Recall that from Farkas lemma, we have

$$\{y \in \mathbb{R}_+^p \mid Wy = h - T\hat{x}\} = \emptyset$$
$$\Rightarrow \exists \sigma \in \mathbb{R}^m \text{ t.q. } W^T\sigma \le 0 \text{ et } (h - T\hat{x})^T\sigma > 0.$$

If our problem is well defined, we should have $K_1 \cap K_2 \ne \emptyset$. For any feasible x, i.e. $x \in K_1 \cap K_2$, we know that $\exists y \ge 0$ such that $Wy = h - Tx$.

Feasibility cuts (cont'd)

We therefore aim to add the inequality

$$\sigma^T(h(\xi_s) - T(\xi_s)x) \le 0,$$

where s is a scenario for which $\hat{x}$ is not feasible.

How to find σ? Basic ideas:

- we have detected that the primal is not feasible as the dual is unbounded;
- therefore, there is some feasible direction along which we can increase the dual value as much as we want;
- this direction is the ray σ that we are looking for.

Lot of technical details in Birge & Louveaux (Chapter 5). We will try to simplify and keep the main practical ideas.

Feasibility cuts (cont'd)

Similar to the two-phase method in linear programming. The first phase looks for a feasible solution by solving an artificial program. We do the same scenario by scenario, but what we look for is an infeasible scenario!

For $s = 1, \ldots, S$, solve the linear program

$$\min w = e^T v^+ + e^T v^-$$
$$\text{s.t. } Wy + Iv^+ - Iv^- = h_s - T_s x^\nu$$
$$y, v^+, v^- \geq 0,$$

where $e^T = (1, \ldots, 1)$, until, for some s', $w_* > 0$.

As $v^+, v^- \geq 0$, $w_* \geq 0$;
- if $w^* = 0$, $v_*^+ = v_*^- = 0$ and $Wy^* = h_s - T_s x^\nu$; in other terms the second-stage problem is feasible for x^ν;
- if $w^* \neq 0$, $v_*^+ \neq 0$ or $v_*^- \neq 0$, so $Wy^* \neq h_s - T_s x^\nu$, and we cannot find some y such that $Wy = h_s - T_s x^\nu$.

Feasibility cuts (cont'd)

The dual of the previous problem is

$$\max_{\sigma} (h_s - T_s x^{\nu})^T \sigma$$
$$\text{s.t. } W^T \sigma \leq 0$$
$$\sigma \leq e$$
$$-\sigma \leq e$$

or, equivalently,

$$\max_{\sigma} (h_s - T_s x^{\nu})^T \sigma$$
$$\text{s.t. } W^T \sigma \leq 0$$
$$\|\sigma\|_{\infty} \leq 1$$

By duality, if $w_* \neq 0$, $(h_s - T_s x^{\nu})^T \sigma_* > 0$. Moreover, as $W^T \sigma_* \leq 0$, σ_* satisfies the conditions we were looking for!

Example 4.2 B&L, p. 174

Consider the problem

$$\min_x 3x_1 + 2x_2 + E_\xi[\min_y -15y_1 - 12y_2]$$

$$\text{s.t. } 3y_1 + 2y_2 \le x_1$$
$$2y_1 + 5y_2 \le x_2$$
$$0.8\xi_1 \le y_1 \le \xi_1$$
$$0.8\xi_2 \le y_2 \le \xi_2$$
$$x, y \ge 0$$

The scenarios are

$$\xi = \begin{cases} (4,6), \ p_1 = 0.25 \\ (4,8), \ p_2 = 0.25 \\ (6,6), \ p_3 = 0.25 \\ (6,8), \ p_4 = 0.25 \end{cases}$$

Deterministic equivalent

Interpretation: we face two investment decisions x_1 and x_2, that must allow to satisfy at least 80% of the demand.

Deterministic equivalent:

$$\min_{x \geq 0, y \geq 0} 3x_1 + 2x_2 - \sum_{s=1}^{4} \frac{1}{4} \left(15y_1^{(s)} + 12y_2^{(s)} \right)$$

$$\text{s.t. } 3y_1^{(s)} + 2y_2^{(s)} \leq x_1, \ s = 1, \ldots, 4$$

$$2y_1^{(s)} + 5y_2^{(s)} \leq x_2, \ s = 1, \ldots, 4$$

$$0.8\xi_1^{(s)} \leq y_1^{(s)} \leq \xi_1^{(s)}, \ s = 1, \ldots, 4$$

$$0.8\xi_2^{(s)} \leq y_2^{(s)} \leq \xi_2^{(s)}, \ s = 1, \ldots, 4$$

$$x, y \geq 0$$

where

$$\xi^{(1)} = (4, 6), \ \xi^{(2)} = (4, 8), \ \xi^{(3)} = (6, 6), \ \xi^{(4)} = (6, 8)$$

Deterministic equivalent

It is possible to solve the deterministic equivalent.

We obtain the solution

$$(x_1^*, x_2^*) = (27.2, 41.6)$$
$$\left(y_1^{(1)*}, y_2^{(1)*}\right) = (4, 6)$$
$$\left(y_1^{(2)*}, y_2^{(2)*}\right) = (4, 6.72)$$
$$\left(y_1^{(3)*}, y_2^{(3)*}\right) = (5.0667, 6)$$
$$\left(y_1^{(4)*}, y_2^{(4)*}\right) = (4.8, 6.4)$$

for an expected optimal value of 22.44.
Is it possible to limit the expansion?

Example B&L: master program

We now build the master program

$$\min_{x,\theta} 3x_1 + 2x_2 + \theta \tag{1}$$

$$\text{s.t. } x \geq 0. \tag{2}$$

Since there is currently no constraint on θ, we simply ignore it. The solution is then $x^* = (0, 0)$, but then, no scenario is feasible!

In order to keep the discussion short, we will consider the scenario $s = 4$, with $\xi_4 = (6, 8)$ (but in general, it is impossible to automatically choose the most convenient scenario).

Feasibility cut

In order to determine the feasibility cut, we have to solve the program

$$\min_{y,v} w = v_1^- + v_2^- + v_3^+ + v_4^- + v_5^+ + v_6^-$$

$$\text{s.t. } 3y_1 + 2y_2 + v_1^+ - v_1^- = 0$$
$$2y_1 + 5y_2 + v_2^+ - v_2^- = 0$$
$$y_1 + v_3^+ - v_3^- = 4.8$$
$$y_1 + v_4^+ - v_4^- = 6.0$$
$$y_2 + v_5^+ - v_5^- = 6.4$$
$$y_2 + v_6^+ - v_6^- = 8.0$$
$$y \geq 0, \ v \geq 0.$$

The optimal value is $w^* = 11.2$, associated to $v_3^+ = 4.6$ and $v_5^+ = 6.4$, and the dual vector (obtained with glpk)

$$\sigma^\nu = \begin{pmatrix} -1 & -1 & 1 & 0 & 1 & 0 \end{pmatrix}$$

Feasibility cut (cont'd)

We define the feasibility cut

$$D_1 x \geq d_1,$$

with

$$D_1 = (\sigma^\nu)^T T = \begin{pmatrix} -1 & -1 & 1 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 0 & -1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 \end{pmatrix}$$

and

$$d_1 = (\sigma^\nu)^T h_1 = \begin{pmatrix} -1 & -1 & 1 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 4.8 \\ 6.0 \\ 6.4 \\ 8.0 \end{pmatrix} = 11.2$$

Feasibility cut (cont'd)

We obtain the cut

$$x_1 + x_2 \geq 11.2$$

which excludes $\hat{x} = (0, \ 0)$.

Birge & Louveaux reports the cut

$$3x_1 + x_2 \geq 123.2$$

Can we justify this difference?

The dual solution is not unique, and we can also choose

$$\sigma^\nu = \begin{pmatrix} -3/11 & -1/11 & 1 & 0 & 1 & 0 \end{pmatrix}^T$$

d_1 does not change, but D_1 become $(3/11 \ 1/11)$. This solution makes the first two dual constraints active. In other terms, this solution corresponds to a different dual basis.

Dual

$$\max_{\sigma^\nu} 4.8\sigma_3^\nu + 6.0\sigma_4^\nu + 6.4\sigma_5^\nu + 8.0\sigma_6^\nu$$

$$\text{s.t. } 3\sigma_1^\nu + 2\sigma_2^\nu + \sigma_3^\nu + \sigma_4^\nu \leq 0 \quad (y_1)$$

$$2\sigma_1^\nu + 5\sigma_2^\nu + \sigma_5^\nu + \sigma_6^\nu \leq 0 \quad (y_2)$$

$$\sigma_1^\nu \leq 0 \quad (v_1^+), \ -\sigma_1^\nu \leq 1 \quad (v_1^-)$$

$$\sigma_2^\nu \leq 0 \quad (v_2^+), \ -\sigma_2^\nu \leq 1 \quad (v_2^-)$$

$$\sigma_3^\nu \leq 1 \quad (v_3^+), \ -\sigma_3^\nu \leq 0 \quad (v_3^-)$$

$$\sigma_4^\nu \leq 0 \quad (v_4^+), \ -\sigma_4^\nu \leq 1 \quad (v_4^-)$$

$$\sigma_5^\nu \leq 1 \quad (v_5^+), \ -\sigma_5^\nu \leq 0 \quad (v_5^-)$$

$$\sigma_6^\nu \leq 0 \quad (v_6^+), \ -\sigma_6^\nu \leq 1 \quad (v_6^-)$$

Dual (cont'd)

By the complementarity conditions, $v_3^+ > 0$ and $v_5^+ > 0$ imply that $\sigma_3^\nu = 1$ and $\sigma_5^\nu = 1$. The strong duality requires that $\sigma_4^\nu = \sigma_6^\nu = 0$ in order to have the dual objective value 11.2.

The optimal solutions can however differ on σ_1^ν and σ_2^ν.

The solution

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

sets two bound constraints active for σ_1^ν and σ_2^ν while the solution obtained by Birge and Louveaux make these constraints inactive, but the first two dual constraints are now active.

Feasibility cuts: B&L results

Use of CPLEX or Gurobi produce the B&L cut, and the first-stage solution becomes $x^1 = (41.067, 0)^T$.

If we continue as in B&L, a second feasibility cut is generated: $x_2 \geq 22.4$. The first-stage solution becomes $x^1 = (33.6, 22.4)^T$.

A third feasibility cut is generated: $x_2 \geq 41.6$. The first-stage decision is now $x^1 = (27.2, 41.6)^T \in \mathcal{K}_1 \cap \mathcal{K}_2$.

Another choice for the initial realization would have produced different cuts.

On the problem formulation

The production of feasibility cuts can be inefficient.

With $\xi_1 = 6$ and $\xi_2 = 8$, it is easy to see that we must have $y_1 \geq 4.8$ and $y_2 \geq 6.4$.

From the first two second-stage constraints, this implies $x_1 \geq 27.2$ and $x_2 \geq 41.6$.

We could therefore consider the following initial program

$$\min 3x_1 + 2x_2 + \mathcal{Q}(x)$$
$$\text{s.t. } x_1 \geq 27.2,$$
$$x_2 \geq 41.6.$$

The recourse is now relatively complete!

Optimality cuts

The problem

$$\min c^T x + \mathcal{Q}(x)$$
$$\text{s.t. } x \in K_1 \cap K_2,$$

is equivalent to

$$\min c^T x + \theta$$
$$\text{s.t. } \mathcal{Q}(x) \leq \theta,$$
$$x \in K_1 \cap K_2,$$

The constraint $\theta \geq \mathcal{Q}(x)$ cannot be directly used, and will be expressed by means of optimality cuts.

Master problem

Initially, the only constraint on θ is $\theta \in \mathbb{R}$, and we obtain the infimum of the objective with $\theta = -\infty$.

As long as $\theta = -\infty$, we remove it from the objective function. In other terms, we solve the first-stage problem only.

Temporarily ignore this constraint. We obtain an optimal solution $(\hat{x}, \theta)^T$, with θ set to $-\infty$ during the first iteration of the method.

The structure of the deterministic equivalent is exploited by minimizing each second-stage function separately, and the decomposition is based on the primal problem. Therefore we speak of primal decomposition, or Benders decomposition.

Optimality cut

Consider the expected second-stage cost

$$\mathcal{Q}(x) = \sum_{j=1}^{S} p_j Q(x, \xi_j),$$

with

$$Q(x, \xi_j) = \min\{q(\xi_j)^T y \mid Wy = h(\xi_j) - T(\xi_j)x, \ y \geq 0\}.$$

We know that $Q(x, \xi_j)$ is convex and piecewise linear, and since Ξ is finite, $\mathcal{Q}$ is also convex and piecewise linear.

Optimality cut (cont'd)

If the second-stage problem is feasible for a given ξ, we can consider its dual

$$\max\{\pi^T(h(\xi) - T(\xi)x) \,|\, \pi^T W \leq q(\xi)^T\}.$$

- If $q(\xi)$ is fixed, the dual is feasible or unfeasible for all x and ξ, as x and ξ do not enter the constraints.
- If q depends on the realization ξ, given ξ, if the dual is feasible for some x, it is feasible for any value of x.
- From the duality properties, the primal is unbounded implies that the dual is not feasible. Assume that we know that our problem is bounded.

Subproblems

Consider the master problem solution $(\hat{x}, \hat{\theta})$. We have

$$\mathcal{Q}(\hat{x}) = \sum_j p_j Q(\hat{x}, \xi_j) = \sum_j p_j q(\xi_j)^T y_j,$$

where y_j is the second-stage optimal solution associated to ξ_j.

From the duality principle, we also have

$$\sum_j p_j q(\xi_j)^T y_j = \sum_j p_j (\hat{\pi}_j)^T (h(\xi_j) - T(\xi_j)\hat{x}),$$

where $\hat{\pi}_j$ is the dual optimal solution associated to ξ_j.

The constraints in the dual problem are $\pi^T W \leq q(\xi_j)$, and are independent of x.

Subproblems (cont'd)

For any x in $\mathcal{K}_1 \cap \mathcal{K}_2$, we have

$$\mathcal{Q}(x) = \max_\pi \sum_j p_j \pi_j^T (h(\xi_j) - T(\xi_j)x)$$
$$\geq \sum_j p_j (\hat{\pi}_j)^T (h(\xi_j) - T(\xi_j)x),$$

The constraint $\theta \geq \mathcal{Q}(x)$ implies

$$\theta \geq \sum_j p_j (\hat{\pi}_j)^T (h(\xi_j) - T(\xi_j)x) = \alpha + \beta^T x,$$

or

$$-\beta^T x + \theta \geq \alpha,$$

with $\alpha = \sum_j p_j (\hat{\pi}_j)^T h(\xi_j)$ and $\beta = -\sum_j T(\xi_j)^T p_j (\hat{\pi}_j)$.

Master problem

Taking account of feasibility and optimality cuts, the master problem can be formulated as:

$$
\begin{aligned}
&\min\ c^T x + \theta \\
&\text{s.t.}\ Ax = b, \\
&\qquad D_\ell^T x \geq d_\ell,\ \ell = 1, \ldots, K, \qquad \text{(feasibility cuts)} \\
&\qquad -\beta_\ell^T x + \theta \geq \alpha_\ell,\ \ell = 1, \ldots, L, \qquad \text{(optimality cuts)} \\
&\qquad x \geq 0,\ \theta \in \mathbb{R}.
\end{aligned}
$$

Let (x^ν, θ^ν) be an optimal solution, with ν the iteration index. If no optimality cut is present, θ^ν is set to $-\infty$ and is not considered in the computation of x^ν.

L-shaped algorithm

Step 0. Initialization: $\nu := 0$, $K := 0$, $L := 0$, $\theta^0 := -\infty$.

Step 1. Solve the master problem in order to get x^0. Stop if the program is not feasible ($K_1 = \emptyset$).

Step 2. Generate a feasibility cut. For $s = 1, \ldots, S$, solve the linear problem

$$\min w = e^T v^+ + e^T v^-$$
$$\text{s.t. } Wy + Iv^+ - Iv^- = h_s - T_s x^\nu$$
$$y, v^+, v^- \geq 0,$$

where $e^T = (1, \ldots, 1)$, until we find a strictly positive optimal value for some scenario *s*.

L-shaped algorithm (cont'd)

Step 2. (continuation) Let σ^ν be the optimal solution of the dual program, and define

$$D_{K+1} = (\sigma^\nu)^T T_s,$$
$$d_{K+1} = (\sigma^\nu)^T h_s,$$

in order to generate a feasibility cut. Increment K and go to Step 4.

If for all s, $w = 0$, go to Step 3.

L-shaped algorithm (cont'd)

Step 3. Compute $\mathcal{Q}(x^\nu)$, by solving each second-stage problem: for $s = 1, \ldots, S$, solve

$$\min \ q_s^T y$$
$$\text{t.q. } Wy = h_s - T_s x^\nu,$$
$$y \geq 0.$$

Stop if $\theta^\nu \geq \mathcal{Q}(x^\nu)$: x^ν is an optimal solution.
Otherwise, increment *L* and build the optimality cut $-\beta_L^T x + \theta \geq \alpha_L$.

Step 4. Increment ν and solve the master program to get x^ν. Stop if the program is not feasible. Otherwise, go to step 2.

Remarks

A pseudo-code is also available in Kall and Wallace.

Theorem
If ξ is a random vector with finite moments of order 2, the L-shaped algorithm converges in a finite number of iterations towards an optimal solution if it exists, or establishes that the two-stage (linear, with fixed recourse) stochastic program is not feasible.

Subgradients and optimality cuts

We can also interpret the optimality cuts in terms of subgradients. Recall that $\mathcal{Q}(x)$ is convex and piecewise linear, so not differentiable everywhere.

We can however show that the optimality cuts are support hyperplans of $\mathcal{Q}(x)$, by means of subgradients, and the feasibility cuts adequately define the feasible points in K_2.

Subgradients

Definition (Subgradient)

A vector $\eta \in \mathbb{R}^n$ is a subgradient of a convex function f at x iff

$$f(y) \geq f(x) + \eta^T(y - x), \ \forall y \in \mathbb{R}^n.$$

Definition (Subdifferential)

The set of all subgradients of f at x is called the subdifferenntial of f at x and is denoted by $\partial f(x)$.

The graph of the (linear) function $h(y) = f(x) + \eta^T(y - x)$ is a support hyperplan for the convex set
$epi(f) = \{(x, \mu) : x \in \mathbb{R}^n, \mu \in \mathbb{R}, \mu \geq f(x)\} \subseteq \mathbb{R}^{n+1}$, at
$(x, f(x))$.

Subgradients (cont'd)

Recall that, by duality,

$$Q(\hat{x}, \xi_s) = \max_{\pi} \left\{ \pi^T (h(\xi_s) - T(\xi_s)x) \mid \pi^T W \le q(\xi_s)^T \right\}.$$

Therefore, if $x \in \mathcal{K}_1 \cap \mathcal{K}_2$, $\exists \hat{\pi}$ such that

$$Q(\hat{x}, \xi_s) = \hat{\pi}^T h(\xi_s) - \hat{\pi}^T T(\xi_s)\hat{x}.$$

With respect to the dual variables,

$$Q(x, \xi_s) \ge \hat{\pi}^T h(\xi_s) - \hat{\pi}^T T(\xi_s)x, \ \forall x \in K_2(\xi_s)$$

and thus

$$Q(x, \xi_s) \ge Q(\hat{x}, \xi_s) - \hat{\pi}^T T(\xi_s)(x - \hat{x}), \ \forall x \in K_2(\xi_s).$$

In other words, $-T(\xi_s)^T \hat{\pi} \in \partial Q(\hat{x}, \xi_s)$.

Similarly, by taking the expectation, $-\sum_s p_s T(\xi_s)^T \hat{\pi}_s \in \partial \mathcal{Q}(x)$.

Optimality cuts: example

We consider the problem

$$\min_x c^T x + E_\xi[Q(x, \xi)],$$
$$\text{s.t. } 0 \leq x \leq 10.$$

Assume that $c = 0$, so that the program can be reduced to the second stage, with a second-stage problem defined as

$$\min y^+ + y^- \text{ t.q. } y^+ - y^- = \xi - x, \ y^+ \geq 0, \ y^- \geq 0.$$

It is possible to analytically compute $Q(x, \xi)$, that can be expressed as:

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

$\mathcal{Q}(x)$

What about $\mathcal{Q}(x)$?

Assume that the three realizations have the same probability.

A previous analysis showed that we have to examine 4 cases:

1. $x \leq 1$: $\mathcal{Q}(x) = 7/3 - x$;
2. $1 \leq x \leq 2$: $\mathcal{Q}(x) = 5/3 - x/3$;
3. $2 \leq x \leq 4$: $\mathcal{Q}(x) = x/3 + 1/3$;
4. $4 \leq x$: $\mathcal{Q}(x) = x - 7/3$;

Application of the *L*-shaped algorithm

We assume that the starting point is $x^1 = 0$. The sequence of iterates for the *L*-shaped method is as follows.

Iteration 1 x^1 is not optimal; create the cut $\theta \geq 7/3 - x$.
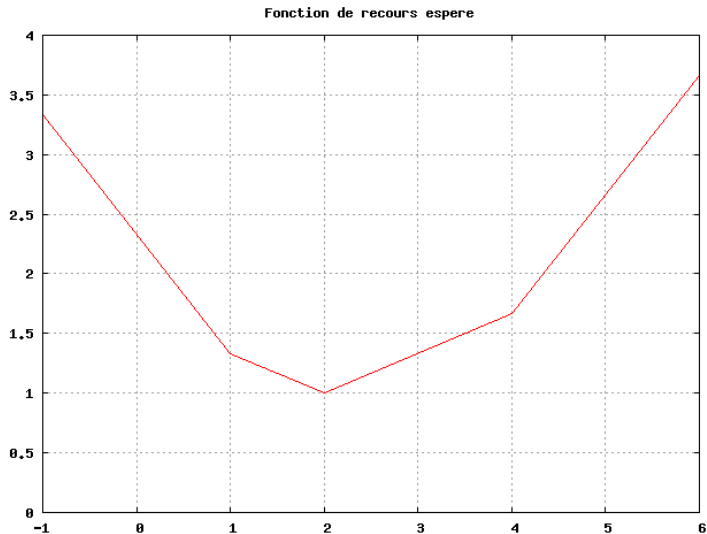
Iteration 2 $x^2 = 10$, $\theta^2 = -23/3$ are not optimal; create the cut $\theta \geq x - 7/3$.

Iteration 3 $x^3 = 7/3$, $\theta^3 = 0$ are not optimal; create the cut $\theta \geq (x + 1)/3$.
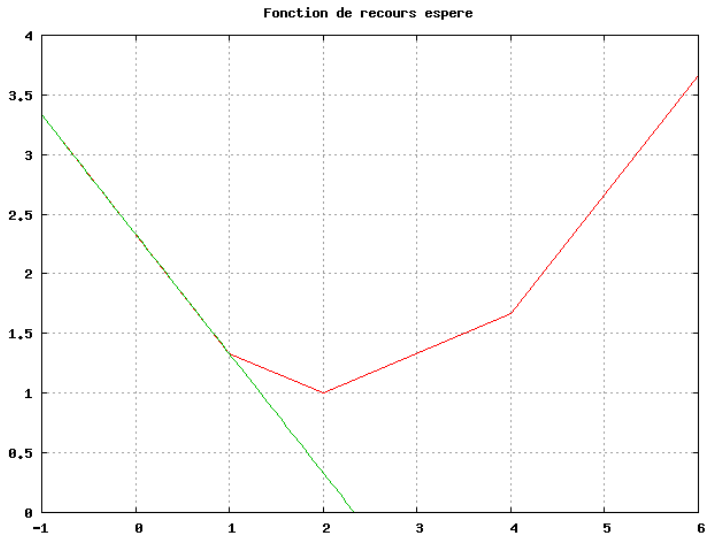
Iteration 4 $x^4 = 1.5$, $\theta^4 = 2.5/3$ are not optimal; create the cut $\theta \geq (5 - x)/3$.

Iteration 5 $x^5 = 2$, $\theta^5 = 1$, which is the optimal solution.
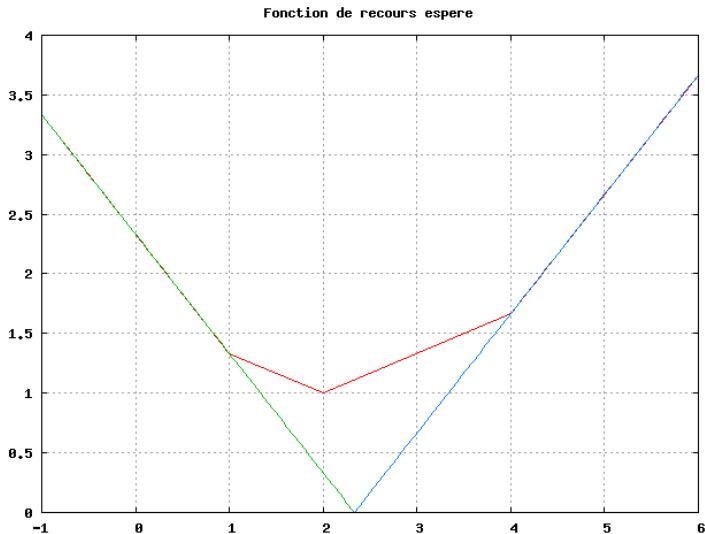
$\mathcal{Q}(x)$, no cut

Fonction de recours espere

$\mathcal{Q}(x)$, addition of the cut $\theta \geq 7/3 - x$

Fonction de recours espere

$\mathcal{Q}(x)$, addition of the cut $\theta \geq x - 7/3$

Fonction de recours espere

$\mathcal{Q}(x)$, addition of the cut $\theta \geq (x+1)/3$

Fonction de recours espere

$\mathcal{Q}(x)$, addition of the cut $\theta \geq (5 - x)/3$

Fonction de recours espere
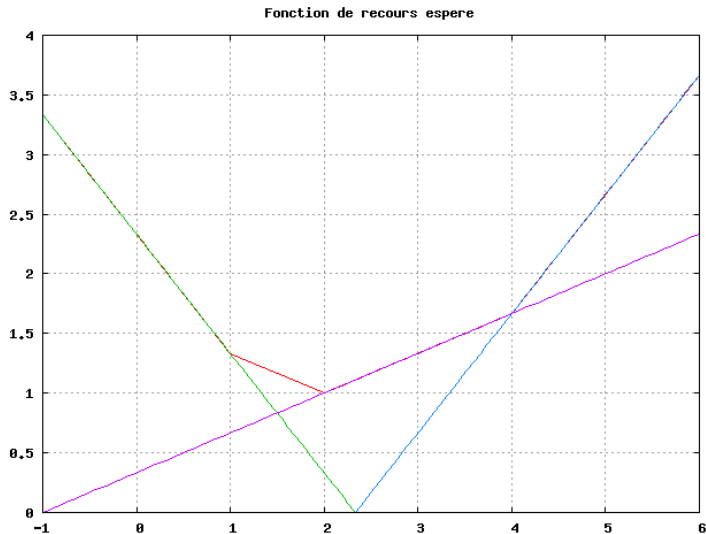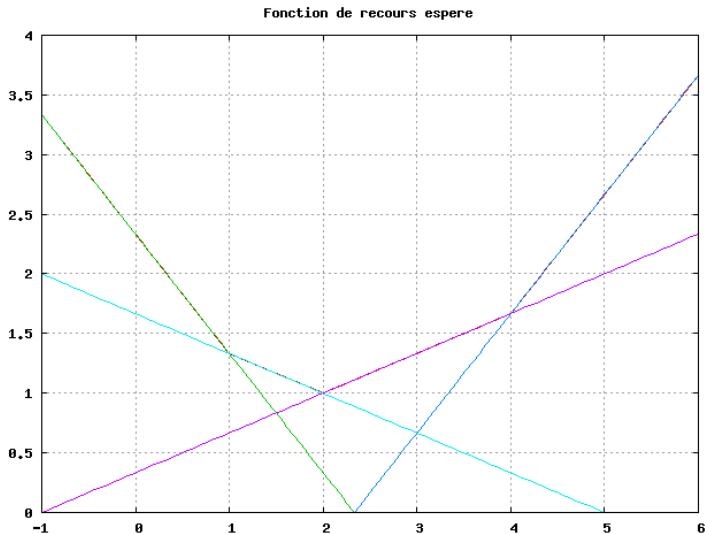
Multi-cut version

- Why do we build one cut only? Idea: approximate each $Q(x, \xi_s)$ by an auxiliary variable θ_s, $s \in S$.

- We new introduce the cuts

$$\theta_s \geq p_s(\hat{\pi}_s)^T (h(\xi_s) - T(\xi_s)x) = \alpha_s + \beta_s^T x, \ s = 1 \in S.$$

  In other terms, we keep $|S|$ cuts instead of aggregating them. If it is not possible to build these $|S|$ cuts, the solution is optimal.

- Which method works the best? The multi-cut version offers a more detailed information, so in general, we can expect less major iterations, but as more cuts are added, the first-stage program is more expensive to solve. As a rule of thumb, the multi-cut approach is more efficient when the number of realizations S is not significantly greater than the number of first-stage constraints.

Reminder: subgradients

Recall that, by duality,

$$Q(\hat{x}, \xi_s) = \max_{\pi}\{\pi^T(h(\xi_s) - T(\xi_s)\hat{x}) \,|\, \pi^T W \leq q(\xi_s)^T)\}.$$

Therefore, $\exists\, \hat{\pi}$ t.q. $Q(\hat{x}, \xi_s) = \hat{\pi}^T h(\xi_s) - \hat{\pi}^T T(\xi_s)\hat{x}$. With respect to the dual variables, we have

$$Q(x, \xi_s) \geq \hat{\pi}^T h(\xi_s) - \hat{\pi}^T T(\xi_s)x, \ \forall x \in K_2(\xi_s)$$

and therefore

$$Q(x, \xi_s) \geq Q(\hat{x}, \xi_s) - \hat{\pi}^T T(\xi_s)(x - \hat{x}), \ \forall x \in K_2(\xi_s).$$

In other terms, $-T(\xi_s)^T\hat{\pi} \in \partial Q(\hat{x}, \xi_s)$.

Similarly, by taking the expectation, $-\sum_s p_s T(\xi_s)^T\hat{\pi}_s \in \partial \mathcal{Q}(x)$.

Multicut *L*-shaped method

Step 0 With θ_s^0 a lower bound for $Q(x, \xi_s)$, define

$$\mathcal{B}_0 = \{\mathbb{R}_n^+ \times \{\theta_1, \theta_2, \ldots, \theta_{|S|}\} \,|\, Ax = b\},$$
$$\mathcal{B}_1 = \{\mathbb{R}_n^+ \times \{\theta_1, \theta_2, \ldots, \theta_{|S|}\} \,|\, \theta_s \geq \theta_s^0, \ \forall s \in S\}.$$

Set $\nu = 0$.

Step 1 Increment ν. Solve the master problem

$$\min \left\{ c^T x + \sum_{s \in S} \theta_s \,|\, (x, \theta_1, \theta_2, \ldots, \theta_{|S|}) \in \mathcal{B}_0 \cap \mathcal{B}_1 \right\}.$$

Record the solution $(x^\nu, \theta_1^\nu, \theta_2^\nu, \ldots, \theta_{|S|}^\nu)$.

Step 2 Evaluate $\mathcal{Q}(x^\nu) = \sum_{s \in S} p_s Q(x^\nu, \xi_s)$.
$\mathcal{Q}(x^\nu) < \infty$?

Multicut *L*-shaped method: step 2

- If $\mathcal{Q}(x^\nu) = \infty$, there is a realization ξ_k for which $Q(x^\nu, \xi_k) = \infty$, we add a feasibility cut and set

$$\mathcal{B}_1 := \mathcal{B}_1 \cap \{(x, \theta_1, \theta_2, \ldots, \theta_{|S|}) \,|\, \sigma^T(h(\xi_k) - T(\xi_k)x) \geq 0\}.$$

  Note: this inequality is the same than in the simple-cut version.
  Return to Step 1.

- If $\mathcal{Q}(x^\nu) < \infty$, we are able to solve all the second-stage problems, for all the scenarios, and obtain the corresponding dual solutions $\hat{\pi}_s$. We get the subgradients

$$u = -T_s^T \hat{\pi}_s \in \partial Q(x^\nu, \xi_s).$$

Multicut L-shaped method: step 2 - $\mathcal{Q}(x^\nu) < \infty$

- If $p_s Q(x^\nu, \xi_s) \leq \theta_s, \forall s \in S$, stop: x^ν is optimal.
- If $p_s Q(x^\nu, \xi_s) > \theta_s$,

  $\mathcal{B}_1 := \mathcal{B}_1 \cap \{(x, \theta_1, \theta_2, \ldots, \theta_{|S|}) \mid \theta_s \geq p_s Q(x^\nu, \xi_s) + p_s u^T (x - x^\nu)\}.$

  Return to Step 1.

Hybrid approaches

- Partition the scenarios in C "clusters" $\mathcal{S}_1$, $\mathcal{S}_1$, ..., $\mathcal{S}_C$, and define
$$\mathcal{Q}_{[\mathcal{S}_k]}(x) = \sum_{s \in \mathcal{S}_k} p_s Q(x, \xi_s).$$

  Thus
  $$\mathcal{Q}(x) = \sum_{k=1}^{C} \mathcal{Q}_{[\mathcal{S}_k]}(x)$$
  $$- \sum_{s \in \mathcal{S}_k} p_s (T(\xi_s))^T \hat{\pi}_s \in \partial \mathcal{Q}_{[\mathcal{S}_k]}(x).$$

- We can therefore do the same: approximate $\mathcal{Q}_{[\mathcal{S}_k]}(x)$ with the subgradient inequality.

Algorithm

Step 0 Let θ_k^0 be a lower bound for $\mathcal{Q}_{[\mathcal{S}_k]}(x)$. Set

$$\mathcal{B}_0 = \{\mathbb{R}_n^+ \times \{\theta_1, \theta_2, \ldots, \theta_C\} \mid Ax = b\},$$
$$\mathcal{B}_1 = \{\mathbb{R}_n^+ \times \{\theta_1, \theta_2, \ldots, \theta_C\} \mid \theta_k \geq \theta_k^0\}.$$

Set $\nu = 0$.

Step 1 Increment ν and solve the master problem

$$\min c^T x + \sum_{k=1}^C \theta_k \text{ t.q. } (x, \theta_1, \theta_2, \ldots, \theta_C) \in \mathcal{B}_0 \cap \mathcal{B}_1.$$

Let $(x^\nu, \theta_1^\nu, \theta_2^\nu, \ldots, \theta_C^\nu)$ be the found solution.

Algorithm - Step 2

- Evaluate $\mathcal{Q}(x^{\nu}) = \sum_{s \in S} p_s Q(x^{\nu}, \xi_s)$.

- If $\mathcal{Q}(x^{\nu}) = \infty$, there is some ξ_k such that $Q(x^{\nu}, \xi_k) = \infty$. We add a feasibility cut:

$$\mathcal{B}_1 := \mathcal{B}_1 \cap \left\{ (x, \theta) \,|\, \sigma^T(h(\xi_k) - T(\xi_k)x) \leq 0 \right\}.$$

  This inequality has no term in θ_s; it is the same inequality than in the standard L-Shaped method.
  Go back to Step 1.

- If $\mathcal{Q}(x^{\nu}) < \infty$, the S LP's corresponding to the S scenarios can be solved, with the dual solutions π_s^{ν}, and

$$u_k := - \sum_{s \in \mathcal{S}_k} p_s (T(\xi_s))^T \hat{\pi}_s \in \partial \mathcal{Q}_{[\mathcal{S}_k]}(x).$$

  If $\mathcal{Q}_{[\mathcal{S}_k]}(x) \leq \theta_k$, $k = 1, \ldots, C$, stop: x^{ν} is optimal.

For each k such that $\mathcal{Q}_{[\mathcal{S}_k]}(x) > \theta_k$, add an optimality cut:

$$\mathcal{B}_1 := \mathcal{B}_1 \cap \left\{ (x, \theta_1, \theta_2, \ldots, \theta_C) \mid \theta_k \geq \mathcal{Q}_{[\mathcal{S}_k]}(x^\nu) + u_k^T (x - x^\nu) \right\}.$$

Example

Assume $n_1 = 1$ (dimension of x), $m_1 = 0$ (no first-stage constraint), $m_2 = 3$ (three second-stage constraints), $n_2 = 6$ (dimension of y), and

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

Assume moreover $S = 2$, and that the two realizations of ξ have the same probability. These realizations are

$$\xi^1 = (q^1, h^1, T^1)^T,$$
$$\xi^2 = (q^2, h^2, T^2)^T.$$

Example (cont'd)

Assume

$$q^1 = (1, 0, 0, 0, 0, 0)^T, \ q^2 = (3/2, 0, 2/7, 1, 0, 0)^T,$$
$$h^1 = (-1, 2, 7)^T, \ h^2 = (0, 2, 7)^T,$$
$$T^1 = T^2 = (1, 0, 0)^T.$$

For the first value ξ^1, we have thus the second-stage program

$$\min_y y_1$$
$$\text{s.t. } y_1 - y_2 - y_3 - y_4 = -1 - x,$$
$$y_2 + y_5 = 2,$$
$$y_3 + y_6 = 7,$$
$$y \geq 0.$$

Example (cont'd)

It is easy to see that, for ξ^1,

$$Q_1(x) = \begin{cases} -x - 1 & \text{if } x \leq -1, \\ 0 & \text{if } x \geq -1. \end{cases}$$

For the second value of $Q(x, \xi)$, we can show that

$$Q_2(x) = \begin{cases} -1.5x & \text{if } x \leq 0, \\ 0 & \text{if } 0 \leq x \leq 2, \\ \frac{2}{7}(x - 2) & \text{if } 2 \leq x \leq 9, \\ x - 7 & \text{if } x \geq 9. \end{cases}$$

Assume also that $-20 \leq x \leq 20$ and $c = 0$.

What are the (expected) properties that we have for $Q_1(x)$ and $Q_2(x)$?

What about the recourse?

Example: *L*-shaped

Start from $x = -2$. We have the following sequence of iterates

Iteration 1 $x^1 = -2$, θ^1 is omitted and we add the cut

$$\theta \geq -0.5 - 1.25x.$$

Iteration 2 $x^2 = 20$, $\theta^2 = -25.5$, new cut:

$$\theta \geq 0.5x - 3.5.$$

Iteration 3 $x^3 = 12/7$, $\theta^3 = -37/14$, new cut:

$$\theta \geq= 0.$$

Iteration 4 $x^4 \in [-2/5, 7]$, θ^4. If x^4 is chose as any value in $[0, 2]$, stop.

Example: multicut *L*-shaped

The multicut approach would give the following sequence.

Iteration 1 $x^1 = -2$, θ^1 and θ^2 are omitted; we add the cuts
$\theta_1 \geq -0.5 - 0.5x$, $\theta_2 \geq -3/4x$. Note:
$\theta_1 \geq \frac{1}{2}(-x - 1)$ and $\theta_1 \geq \frac{1}{2}(-1.5x)$

Iteration 2 $x^2 = 20$, $\theta_1^2 = -10.5$, $\theta_2^2 = -15$; new cuts: $\theta_1 \geq 0$,
$\theta_2 \geq 0.5x - 3.5$.

Iteration 3 $x^3 = 2.8$, $\theta_1^3 = 0$, $\theta_2^3 = -2.1$; new cut:
$\theta_2^3 \geq 1/7(x - 2)$.

Iteration 4 $x^4 = 0.32$, $\theta_1^4 = 0$, $\theta_2^4 = -0.24$; new cut: $\theta_2 \geq 0$.

Iteration 5 $x^5 = 0$, $\theta_1^5 = \theta_2^5 = 0$. Stop.

This example shows the the multicut approach can sometimes
take more iterations than the single cut version.

Regularization

The *L*-shaped method suffers from "stability" issues,

- especially during the first iterations when we do not have a good model of $\mathcal{Q}(x)$;
- especially when we start from a good estimate of the solution.

It is therefore useful to regularize the method. Various techniques exist.

Some references on the stabilization of the Benders decomposition:

- Marsten, Hogan, Blankenship (1975);
- Lemaréchal (1975,...);
- Kiwiel (1983,...);
- Neame, Boland, and Ralph (1998).

Regularization of the *L*-Shaped method

- **Penalization of the step length**: A. Ruszczyński, *A Regularized Decomposition for Minimizing a Sum of Polyhedral Functions*, Mathematical Programming 35, pp. 309–333, 1986.

- **Approach in $\|\cdot\|_\infty$**: J. T. Linderoth and S. J. Wright, *Implementing a Decomposition Algorithm for Stochastic Programming on a Computational Grid*, Computational Opimization and Applications 24, pp. 207–250, 2003.

Penalization of the step length

At iteration k, we solve the program

$$\min_x c^T x + \sum_{j \in C} \theta_j + \frac{1}{2\rho} \|x - x_k\|^2.$$

- Large ρ: as the L-shaped.
- Small ρ: we do not move a lot.

We speak of regularized decomposition method.

Trust-region approach

The principle is also to penalize the step length. At iteration k, we require $\|x - x_k\|_k \leq \Delta_k$, where $\| \cdot \|_k$ designs an usual norm (typically 2 or ∞).

- Big Δ_k: as the *L*-shaped.
- Small Δ_k: we do not move a lot.

We speak of the trust-region regularization.

In both case, we see the introduction of nonlinear programming concepts.

Summary

Let $m(x)$ be the "model" used as first-stage problem.

- Multicut L-shaped

$$m(x) = \min_x \left\{ c^T x + \sum_{k=1}^{C} \theta_k \,|\, (x, \theta_1, \theta_2, \ldots, \theta_C) \in \mathcal{B}_0 \cap \mathcal{B}_1 \right\}.$$

- Trust-region

$$m(x) = \min_x \left\{ c^T x + \sum_{k=1}^{C} \theta_k \,|\, (x, \theta_1, \theta_2, \ldots, \theta_C) \in \mathcal{B}_0 \cap \mathcal{B}_1, \right.$$
$$\left. \|x - x^l\|_\infty \leq \Delta_l \right\}.$$

- Regularized decomposition

$$m(x) = \min_x \left\{ c^T x + \sum_{k=1}^{C} \theta_k + (1/2\rho_l)\|x - a^l\|_2^2 \right.$$
$$\left. \text{s.t. } (x, \theta_1, \theta_2, \ldots, \theta_C) \in \mathcal{B}_0 \cap \mathcal{B}_1, \right\}.$$

Regularized decomposition method

Reference: Birge and Louveaux, Section 5.2. Multicut version.

Step 0   Set $\rho = 1$, $r = \nu = 0$, $s(k) = 0$ for all $k = 1, \ldots, S$. Select a feasible solution a^1.

Step 1   Set $\nu := \nu + 1$, and solve the regularized master program

$$\min c^T x + \sum_{k=1}^{S} \theta_k + \frac{1}{2\rho} \|x - a^\nu\|^2,$$

$$\text{s.t. } Ax = b,$$
$$D_\ell^T x \geq d_\ell, \ \ell = 1, \ldots, r,$$
$$-\beta_{\ell(k)}^T x + \theta_k \geq \alpha_{\ell(k)}, \ \ell(k) = 1, \ldots, s(k),$$
$$x \geq 0.$$

Regularized decomposition method (cont'd)

Step 1 (cont'd) Let (x^ν, θ^ν) be an optimal solution of this
  program, 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 If a feasibility cut is generated, set $a^{\nu+1} = a^\nu$
  (unfeasible null step), and return to Step 1.

Step 3 For $k = 1, \ldots, S$, solve the linear subproblem

$$\min_y q_k^T y$$
$$\text{s.t. } Wy = h_k - T_k x^\nu,$$
$$y \geq 0.$$

Compute $\mathcal{Q}_k(x^\nu)$.

Regularized decomposition method (cont'd)

Step 3 (cont'd) If

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

add an optimality cut of the form

$$-\beta_{\ell(k)}^T x + \theta_k \geq \alpha_{\ell(k)}.$$

Set $s(k) = s(k) + 1$.

Step 4 If for all k,

$$\theta_k^\nu \geq p_k (\pi_k^\nu)^T (h_k - T_k x^\nu),$$

set $a^{\nu+1} := x^\nu$ (exact serious step). Return to Step 1.

Regularized decomposition method (cont'd)

Step 5 If
$$c^T x^\nu + \mathcal{Q}(x^\nu) \le c^T a^\nu + \mathcal{Q}(a^\nu),$$

set $a^{\nu+1} := x^\nu$ (serious approximative step).
Otherwise, set $a^{\nu+1} := a^\nu$.
Return to Step 1.

Immediate extension: allow ρ to vary.

The implementation of the regularized *L*-shaped raises various practical questions. An efficient implementation is proposed in Ruszczyński (1986).

Example: regularized *L*-shaped

Consider again Birge and Louveaux, Section 5.1, Exercice 5, page 200.

Take $a^1 = -0.5$ as starting point. This point corresponds to the solution of the expected value problem. i.e the problem obtained by replacing ξ by $E[\xi]$. The second stage program becomes

$$
\min \frac{5}{4}y_1 + \frac{1}{7}y_3 + \frac{1}{2}y_4
$$
$$
\text{s.t. } y_1 - y_2 - y_3 - y_4 = -\frac{1}{2} - x,
$$
$$
y_2 + y_5 = 2,
$$
$$
y_3 + y_6 = 7,
$$
$$
y \geq 0.
$$

Exercise: check that $a^1 = -0.5$ is indeed the optimal value of this program (using the KKT/primal-dual conditions).

Example: regularized *L*-shaped (cont'd)

For $a^1 = -0.5$, the regularized master problem has also for solution $x^1 = -0.5$, and the second stage program of the expected-value problem has the solution $(0, 0, 0, 0, 2, 7)$, while $Q_1(a^1) = 0$, $Q_2(a^1) = 3/4$, so $\mathcal{Q}(a^1) = 3/8$. The regularized approach leads to the following iterations.

Iteration 1 The cuts $\theta_1 \geq 0$, $\theta_2 \geq -3/4x$ are added. $a^2 := a^1$.

Iteration 2 The regularized master program is

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

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

Example: regularized *L*-shaped (cont'd)

Iteration 3 The regularized master problem is

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

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

Theorem
*If the original problem has a solution, the algorithm stops after a
finite number of iterations. Otherwise, it generates a sequence
of feasible points $\{a^\nu\}$ such that $\mathcal{Q}(a^\nu) \to -\infty$, as $\nu \to \infty$.*

Proof.
See Birge and Louveaux, Section 5.2.   □

Regularization using trust-region methods

Basic philosophy of trust-region methods: replace the original problem with some locally valid approximation easier to handle.

<table>
<tr><td align="center">Original problem</td><td align="center">Model</td></tr>
<tr><td>

$\min\limits_{x} c^T x + \mathcal{Q}(x)$

s.t. $Ax = b$

$\quad x \geq 0.$

</td><td>

$\min\limits_{x,\theta} m(x, \theta) := c^T x + \theta$

s.t. $Ax = b$

$\quad$ feasibility cuts

$\quad$ optimality cuts

$\quad x \geq 0.$

</td></tr>
</table>

Regularization using trust-region methods (cont'd)

- The model $m(\cdot)$ is valid locally only.
- Constrain the candidate iterate to stay in a trust region $\mathcal{B}_k$ centered at the current solution x_k:

$$\mathcal{B}_k = \{x \mid \|x - x_k\|_k \le \Delta_k\},$$

  where Δ_k is the trust-region radius, and $\|\cdot\|_k$ is the norm used at iteration k.

- In the context a linear programming, it is common to use ∞-norm as

$$\|x - x_k\|_\infty \le \Delta_k \text{ iff } -\Delta_k \boldsymbol{e} \le x - x_k \le \Delta_k \boldsymbol{e}.$$

Regularization using trust-region methods (cont'd)

- The trust-region subproblem at iteration k is now

$$\min_{x,\theta} m(x) := c^T x + \theta$$
$$\text{s.t. } Ax = b$$
$$\text{feasibility cuts}$$
$$\text{optimality cuts}$$
$$\|x - x_k\|_\infty \le \Delta_k$$
$$x \ge 0.$$

- We can extend the model to multi cuts or scenarios partition in clusters.
- Only move if the model does perform well.

Regularization using trust-region methods (cont'd)

- Let x^* be the first-stage solution.
- Set $0 < \eta_1 \le \eta_2 < 1$. For instance, $\eta_1 = \eta_2 = 10^{-4}$.
- Compare the actual reduction to the prediction reduction:

$$\rho = \frac{c^T(x_k - x^*) + \mathcal{Q}(x_k) - \mathcal{Q}(x^*)}{m(x_k) - m(x^*)}$$

- Unsuccessful iteration: if $\rho < \eta_1$, set $x_{k+1} = x_k$, $\Delta_{k+1} \le \Delta_k$, add cuts. Enforce $\Delta_{k+1} < \Delta_k$ if Δ_k has been kept the same for several consecutive unsuccessful iterations or if $\rho\Delta_k$ it too large.
- Successful iteration: if $\rho \ge \eta_1$, set $x_{k+1} = x^*$.
- Very successful iteration: if $\rho \ge \eta_2$, set $\Delta_{k+1} = \min\{\Delta_{\max}, \beta\Delta_k\}$, with $\beta > 1$.

Regularization using trust-region methods: notes

Additional details and considerations in Linderoth and Wright.

- Allow deletion of inactive cuts.
- Investigate parallelization of the method.
- Do not distinguish successful and very successful iterations: $\eta_1 = \eta_2 \in (0, 1/2)$.