BENDERS

The "L-SHAPED" METHOD

©Dennis Bricker, 2001
Dept of Industrial Engineering
The University of Iowa

DECOMPOSITION

Consider the **Two-stage stochastic LP with recourse**:

$$\text{Minimize } cx + \sum_{k=1}^{K} p_k Q_k(x)$$
$$\text{subject to } x \in X$$

where, *for example*, the **feasible set of first-stage decisions** is defined by

$$X = \left\{ x \in R^n : Ax = b, x \geq 0 \right\}$$

Here k indexes the finitely-many possible realizations of a random vector ξ, with p_k the probability of realization k.

The first-stage variables x are to be selected *before* ξ is observed.

Then the set of **second-stage decision variables** y_k are to be selected, after x has been selected ***and*** the k[th] realization of ξ is observed.

The **cost of the second stage** when scenario k occurs is

$$Q_k(x) = \text{Minimum } \left\{ q_k y : W_k y = h_k - T_k x,\ y \geq 0 \right\}$$

That is, y is a ***recourse*** which must be chosen so as to satisfy some linear constraints in the least costly way.

Note that, in general,

- the coefficient matrices T and W,
- the right-hand-side vector h, and
- the second-stage cost vector q

are *all* random.

We assume that recourse is *complete*,

i.e., for any choice of x and realization ξ, the set

$$Y_k \equiv \left\{ y : W_k y = h_k - T_k x,\ y \geq 0 \right\} \neq \emptyset$$

(This may require the introduction of artificial variables with large costs.)

The objective is to minimize the **expected total costs** of first and second stages.

The *deterministic equivalent LP* is a large-scale problem which *simultaneously* selects

- the first-stage variables x and
- the second-stage variables y_k for *every* realization k

$$\text{P: Find } Z = \text{minimum } cx + \sum_{k=1}^{K} p_k q_k y_k$$

subject to

$$T_k x + W y_k = h_k, k = 1, \ldots K;$$
$$x \in X$$
$$y_k \geq 0, k = 1, \ldots K$$

This can be an extremely large LP, with $K \times n_2$ variables and $K \times m_2$ constraints.

Benders' Decomposition

Benders' partitioning--known also in stochastic programming as the **"L-Shaped Method"**--

achieves *separability* of the second stage decisions, that is,

a separate LP is solved for each of the K scenarios.

Benders' partitioning was introduced by J.F. Benders for solving mixed-integer LP problems, that is, LP problems where some of the variables are restricted to integers:

Benders, J. F. (1962). "Partitioning Procedures for Solving Mixed-Variables Programming Problems." *Numerische Mathematik* **4**: 238-252.

The equivalent "L-Shaped Method" was later introduced independently by van Slyke & Wets for solving stochastic LP with Recourse (**SLPwR**) problems:

Van Slyke, R. M. and R. J.-B. Wets (1969). "L-Shaped linear programs with applications to optimal control and stochastic programming." *SIAM Journal of Applied Mathematics* **17**: 638-663.

In both cases, the central idea is to ***partition*** the variables into 2 sets

(MIP: *integer* and *continuous* variables) or

(SLPwR: *1st stage* & *2nd stage* variables)

and to ***project*** the problem onto the first set of variables.

- **MIP**: $\underset{x \in X}{Minimize}\ cx + V(x)$

  where *V(x)* = optimal value of continuous LP after integer variables x have been fixed.

- **SLPwR**: $\underset{x \in X}{Minimize}\ cx + Q(x)$

  where *Q(x)* = minimal expected cost of second stage when first-stage variables x have been fixed.

Given a first-stage decision x_0, define a function $Q_k(x_0)$ equal to the optimum of the second stage for each scenario k=1, …K:

$$Q_k(x_0) = \min q_k y_k$$

subject to

$$Wy_k = h_k - T_k x_0$$
$$y_k \ge 0$$

Then $P(x_0) = cx_0 + \sum_{k=1}^{k} p_k Q_k(x_0)$ provides us with an ***upper*** bound on the optimal value Z.

If, as before, we introduce the variables x_k for each scenario k, together with the *nonanticipativity constraints*, we obtain the second-stage problem for scenario k,

$$\text{Minimize } q_k y_k$$
subject to
$$T_k x_k + W y_k = h_k,$$
$$x_k = x_0,$$
$$y_k \ge 0$$

whose *linear programming dual* is the linear program

$$\text{Maximize } h_k \pi_k + x_0 \lambda_k$$
subject to:
$$\pi_k T_k + \lambda_k I = 0$$
$$\pi_k W \le q_k$$

It is not necessary to introduce the variables x_k, but it is done in anticipation of later defining a cross-decomposition algorithm, which is a hybrid of Benders' decomposition and Lagrangian relaxation.

We can eliminate λ_k (the dual variables for the constraint $x_k = x_0$) by using the equality constraint to obtain $\lambda_k = -\pi_k T_k$ and

$$Q_k(x_0) = \text{Max } (h_k - T_k x_0)\pi_k$$
subject to
$$\pi_k W \le q_k$$

The original problem now reduces to

$$Z = \underset{x_0 \in X}{\text{Minimize}}\ cx_0 + \sum_{k=1}^{K} p_k Q_k(x_0)$$

Denote by $\Pi_k = \{\pi_k : W^T \pi_k \le q_k\}$ the polyhedral feasible region of the second-stage problem for scenario k.

Denote by $\hat{\pi}_k^i$ the i[th] extreme point of Π_k, $i = 1, 2, \ldots I_k$.

By enumerating the large (but *finite*) number of extreme points of Π_k, we can write

$$Q_k(x_0) = \max_{i=1,\ldots I_k} \left\{\hat{\pi}_k^i (h_k - T_k x_0)\right\} = \max_{i=1,\ldots I_k} \left\{\hat{\lambda}_k^i x_0 + \hat{\alpha}_k^i\right\}$$

where $\lambda_k^i = -\hat{\pi}_k^i T_k$ and $\hat{\alpha}_k^i = \hat{\pi}_k^i h_k$.

(Note that this demonstrates that $Q_k(x_0)$ is a piecewise–linear convex function.)

Benders' *(Complete) Master Problem* then uses this representation of $Q_k(x_0)$ to provide an alternate method for evaluating *Z*, namely

$$Z = \text{Min } cx_0 + \sum_{k=1}^{K} p_k \theta_k$$

subject to $x_0 \in X$, and

$$\theta_k \ge \hat{\lambda}_k^i x_0 + \hat{\alpha}_k^i, \text{ i=1, ...I; k=1, ...K}$$

"Multi-Cut" Version

While it is possible in *principle* to solve the problem using Benders' *Complete Master Problem*,
in *practice* the magnitude of the number of dual extreme points makes it prohibitively expensive.

However, if a *subset* of the dual extreme points of Π_k are available, e.g., $\hat{\pi}_k^i$, $i = 1, \ldots M_k$ where $M_k < I_k$, then we obtain an *underestimate* of $Q_k(x_0)$, which we denote by

$$\underline{Q}_k(x_0) = \max_{i=1,\ldots M_k} \left\{\hat{\lambda}_k^i x_0 + \hat{\alpha}_k^i\right\}$$

Thus, by making use of dual information obtained after M evaluations of $Q_k(x_0)$, we obtain a *Partial Master Problem*,

$$\Phi_M = \text{Min } cx_0 + \sum_{k=1}^{K} p_k \theta_k$$

subject to $x_0 \in X$, and

$$\theta_k \ge \hat{\lambda}_k^i x_0 + \hat{\alpha}_k^i, \text{ i=1, ...M; k=1, ...K}$$

which provides a ***lower bound*** on the solution of *Z*.

Benders' algorithm solves the current *Partial Master Problem*, obtaining

- x_0 (a "trial solution") and
- an underestimate $\sum_{k=1}^{K} p_k \underline{Q}_k(x_0)$ of the associated expected second-stage cost.

The *actual* expected second-stage cost, i.e., $\sum_{k=1}^{K} p_k Q_k(x_0)$, is then evaluated by solving the second-stage problem for each scenario. Additional constraints are added to the Partial Master Problem to complete the iteration.

At each iteration of Benders' algorithm, then,

- the *subproblem* solution

  $$P(x_0) = cx_0 + \sum_{k=1}^{K} p_k Q_k(x_0)$$

  provides an ***upper*** bound for Z, and
- the *Partial Master Solution*

  $$\Phi_M = \underline{P}(x_0) = cx_0 + \sum_{k=1}^{K} p_k \underline{Q}_k(x_0)$$

  provides a ***lower*** bound for Z.

Benders' Algorithm-- "Uni-cut" Version

In the uni-cut version, at each iteration *i* the K constraints

$$\theta_k \geq \hat{\lambda}_k^i x_0 + \hat{\alpha}_k^i,\ \text{k=1,...K}$$

are aggregated before adding them to the **Partial Master Problem**:

$$Z = \text{Min } cx_0 + \theta$$

$$\text{subject to } x_0 \in X, \text{ and}$$

$$\theta \geq \sum_{k=1}^{K} p_k \left[\hat{\lambda}_k^i x_0 + \hat{\alpha}_k^i\right],\ \text{i=1, ...I}$$

"Uni-Cut" Version

Generally, *more iterations* are required, but there are *fewer cuts* (& less computation) in each Partial Master Problem.

Benders' algorithm is as follows:

Step 0. Select an arbitrary $x_0 \in X$. Initialize the upper bound $\overline{Z} = +\infty$ and lower bound $\underline{Z} = -\infty$.

Note: This allows the user to make use of knowledge about his/her problem by using an initial *"guess"* at the solution. Another alternative is to solve the *Expected–Value LP problem* to obtain the initial x_0:

$$\text{Minimize } cx$$

$$\text{subject to } Ax = b,$$

$$\left[\sum_{k=1}^{K} p_k T_k\right] x + Wy = \sum_{k=1}^{K} p_k h_k$$

$$x \geq 0$$

Step 1a. Solve the primal subproblems to evaluate $Q_k(x_0)$ and the optimal dual variables π_k, k=1,...K and compute $P(x_0)$.

1b. For each scenario, generate an *optimality cut.*

1c. *Uni-cut version:* Aggregate the K optimality cuts and add to Benders' master problem.

Multi-cut version: Add each of the K optimality cuts to Benders' master problem.

1d. Update the upper bound, $\overline{Z} = \min\left\{\overline{Z}, P(x_0)\right\}$.

1e. If $\overline{Z} - \underline{Z} \leq \varepsilon$, **STOP**; else continue to *Step 2*.

Step 2a. Solve the *Partial Master Problem* to obtain

- an optimal x_0, and
- an underestimate $\underline{P}(x_0) = cx_0 + \sum_{k=1}^{K} p_k \underline{Q}_k(x_0)$ of the expected cost $P(x_0)$.

2b. Update the *lower* bound, $\underline{Z} = \max\left\{\underline{Z}, \underline{P}(x_0)\right\}$.

2c. If $\overline{Z} - \underline{Z} \leq \varepsilon$, STOP; else return to *Step 1a*.

At each iteration, the number of constraints (and therefore the size of the basis) of the Partial Master Problem increases, adding to the computational burden.

Furthermore, because constraints have been added, the solution of each partial master problem is generally infeasible in the partial master problem which follows.

For these reasons,

it is preferable to solve the *dual* of the partial master problem, *which is formed by appending a column to the dual of the previous partial master problem,*

so that the solution of the dual of the *previous* Partial Master Problem may serve as an initial basic feasible solution for the Partial Master Problem which follows.

If $X = \{x : Ax = b, x \geq 0\}$, the linear programming dual of Benders' Partial Master problem is

$$\Phi_M = Max\ \ bu + \sum_{k=1}^{K}\sum_{i=1}^{M} \hat{\alpha}_k^i v_k^i$$

$$\text{subject to } A^T u - \sum_{k=1}^{K}\sum_{i=1}^{M} \hat{\lambda}_k^i v_k^i = c$$

$$\sum_{i=1}^{M} v_k^i = p_k,\quad k = 1,\ldots K$$

$$v_k^i \geq 0,\quad i = 1,\ldots M; k = 1,\ldots K$$

(The dual variable u is

unrestricted in sign if X is defined by $Ax=b$, but

nonnegative if $Ax \geq b$, and

nonpositive if $Ax \leq b$.)

It can be shown that, in fact,

this dual of Benders' Master Problem is identical to the Master Problem of *Dantzig-Wolfe* decomposition applied to the original large-scale deterministic equivalent LP!