"A Cross-Decomposition Algorithm for Two-Stage Stochastic Linear Programming with Recourse"

Abstract: We consider a paradigm of linear optimization in the face of uncertainty, in which (first-stage) decisions must be made before the uncertainty is resolved, and then recourse (second-stage decisions) is available to compensate. When a finite set of scenarios can be identified and their probability estimated, and the objective is to minimize the sum of the first-stage cost and the expected value of the second-stage cost, a (generally large) deterministic equivalent LP problem can be constructed. Benders' (primal) decomposition and Lagrangian (dual) decomposition each yields a family of smaller subproblems, one for each scenario, and a coordinating "master" problem. Cross-decomposition is a hybrid primal-dual iterative approach which eliminates the master problems and uses the primal and dual subproblems to provide both upper and lower bounds on the optimal expected cost at each iteration. A small example illustrates the computation.

EXAMPLE

- A farmer raises wheat, corn, and sugar beets on 500 acres of land. Before the planting season he wants to decide how much land to devote to each crop.
- At least 200 tons of wheat and 240 tons of corn are needed for cattle feed, which can be purchased from a wholesaler if not raised on the farm.
- Any grain in excess of the cattle feed requirement can be sold at $170 and $150 per ton of wheat and corn, respectively.
- The wholesaler sells the grain for 40% more (namely $238 and $210 per ton, respectively.)
- Up to 6000 tons of sugar beets can be sold for $36 per ton; any additional amounts can be sold for $10/ton.

DATA

	Wheat	Corn	Sugar Beets
Average Yield	2.5 T/Acre	3 T/Acre	20 T/Acre
Planting cost	$150/Acre	$230/Acre	$260/Acre
Selling price	$170/T	$150/T	$36/T first 6000T
			$10/T otherwise
Purchase price	$238/T	$210/T	
Minimum Rqmt	200T	240T	

DECISION VARIABLES

We distinguish between two types of decisions:

First stage (before growing season):

x_1 = acres of land planted in wheat

x_2 = acres of land planted in corn

x_3 = acres of land planted in beets

Second stage (after harvest):

w_1 = tons of wheat sold

w_2 = tons of corn sold

w_3 = tons of beets sold at $36/T

w_4 = tons of beets sold at $10/T

y_1 = tons of wheat purchased

y_2 = tons of corn purchased

LINEAR PROGRAMMING MODEL

Minimize $150x_1 + 230x_2 + 260x_3 + 238y_1 - 170w_1 + 210y_2 - 150w_2 - 36w_3 - 10w_4$

subject to

$$x_1 + x_2 + x_3 \le 500$$
$$2.5x_1 + y_1 - w_1 \ge 200$$
$$3x_2 + y_2 - w_2 \ge 240$$
$$w_3 + w_4 \le 20x_3$$
$$w_3 \le 6000$$
$$x_i \ge 0,\ \text{i=1,2,3};\ y_i \ge 0,\ \text{i=1,2};\ w_i \ge 0,\ \text{i=1,2,3,4}$$

OPTIMAL SOLUTION

Profit = $118,600

	Wheat	*Corn*	*Sugar Beets*
Plant	120 Acres	80 Acres	300 Acres
Yield	300T	240T	6000T
Sales	100T	--	6000T
Purchase	--	--	--

In actuality, crop yields are uncertain, depending upon weather conditions during the growing season.

Three scenarios have been identified

- "good" (20% higher than average)
- "fair" (average)
- "bad" (20% below average),

each equally likely:

Scenario k	*Wheat yield (tons/acre)*	*Corn yield (tons/acre)*	*Beet yield (tons/acre)*
1. Good	3	3.6	24
2. Fair	2.5	3	20
3. Bad	2	2.4	16

Scenario #1: "Good" Yield: Optimal Profit = $167,667

	Wheat	*Corn*	*Sugar Beets*
Plant	183.333 Acres	66.67 Acres	250 Acres
Yield	550T	240T	6000T
Sales	350T	--	6000T
Purchase	--	--	--

Scenario #3: "Bad" Yield: Optimal Profit = $59,950

	Wheat	*Corn*	*Sugar Beets*
Plant	100 Acres	25 Acres	375 Acres
Yield	200T	60T	6000T
Sales	--	--	6000T
Purchase	--	--	--

If a *perfect forecast* was available, then, the expected profit would be

$$\tfrac{1}{3}\times\$167,667+\tfrac{1}{3}\times\$118,600+\tfrac{1}{3}\times\$59,950=\$115,406$$

The stochastic decision problem is to optimize the first-stage cost plus the ***expected*** second-stage costs:

$$\text{Minimize } 150x_1+230x_2+260x_3+\tfrac{1}{3}\sum_{k=1}^{3}Q_k\left(x\right)$$

subject to $x_1+x_2+x_3\le 500$

$x_j\ge 0$, j=1,2,3

where

$Q_k\left(x\right)$ = second-stage costs in scenario k, *if* first-stage decisions x have been implemented

$$Q_1\left(x\right)=\text{Minimum } 170w_1+150w_2+36w_3+10w_4-238y_1-210y_2$$
$$\text{s.t. } y_1-w_1\ge 200-3x_1$$
$$y_2-w_2\ge 240-3.6x_2$$
$$w_3+w_4\le 24x_3$$
$$y_1\ge 0,\ y_2\ge 0,\ w_1\ge 0, w_2\ge 0,\ 0\le w_3\le 6000, w_4\ge 0$$

$$Q_2\left(x\right)=\text{Minimum } 170w_1+150w_2+36w_3+10w_4-238y_1-210y_2$$
$$\text{s.t. } y_1-w_1\ge 200-2.5x_1$$
$$y_2-w_2\ge 240-3x_2$$
$$w_3+w_4\le 20x_3$$
$$y_1\ge 0,\ y_2\ge 0,\ w_1\ge 0, w_2\ge 0,\ 0\le w_3\le 6000, w_4\ge 0$$

$$Q_3\left(x\right)=\text{Minimum } 170w_1+150w_2+36w_3+10w_4-238y_1-210y_2$$
$$\text{s.t. } y_1-w_1\ge 200-2x_1$$
$$y_2-w_2\ge 240-2.4x_2$$
$$w_3+w_4\le 16x_3$$
$$y_1\ge 0,\ y_2\ge 0,\ w_1\ge 0, w_2\ge 0,\ 0\le w_3\le 6000, w_4\ge 0$$

TWO-STAGE LINEAR PROGRAMMING WITH RECOURSE

$$\text{Minimize } z=cx+E\left[\min q\left(\omega\right)y\left(\omega\right)\right]$$

subject to

$$Ax=b$$
$$T\left(\omega\right)x+Wy\left(\omega\right)=h\left(\omega\right),$$
$$x\ge 0,\ y\left(\omega\right)\ge 0$$

where

x = first-stage decision

and

$y\left(\omega\right)$ = second-stage decision after random event ω is observed

which must satisfy the ***second-stage constraints***

$T\left(\omega\right)x+Wy\left(\omega\right)=h\left(\omega\right)$,

where $q\left(\omega\right)$, $T\left(\omega\right)$ & $h\left(\omega\right)$ are random variables

DETERMINISTIC EQUIVALENT PROBLEM

Assume a finite number of scenarios.

For each scenario k, define a set of second-stage variables, y^k, and arrays T_k, q_k, and h_k

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

$$\text{Minimize } cx+\sum_{k=1}^{K}p_kQ_k\left(x\right)$$

subject to $x\in X$

where the cost of the second stage is

$$Q_k\left(x\right)=\text{Minimum }\left\{q_ky:Wy=h_k-T_kx,\ y\ge 0\right\}$$

Consider the *deterministic LP* derived from the 2-stage stochastic LP:

$$Z=\min cx+\sum_{k=1}^{K}p_kq_ky^k$$

subject to

$$T_kx+Wy^k=h_k, k=1,\ldots K;$$
$$x\in X$$
$$y^k\ge 0, k=1,\ldots K$$

where the feasible set of first-stage decisions is defined by

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

EXAMPLE:

Second stage decisions:

For each scenario k (k=1,2,3), define a set of decision variables:

- w_1^k = tons of wheat sold
- w_2^k = tons of corn sold
- w_3^k = tons of beets sold at $36/T
- w_4^k = tons of beets sold at $10/T
- y_1^k = tons of wheat purchased
- y_2^k = tons of corn purchased

DETERMINISTIC EQUIVALENT LP:

$$\text{Minimize } 150x_1+230x_2+260x_3+\tfrac{1}{3}\left(238y_1^1-170w_1^1+210y_2^1-150w_2^1-36w_3^1-10w_4^1\right)$$
$$+\tfrac{1}{3}\left(238y_1^2-170w_1^2+210y_2^2-150w_2^2-36w_3^2-10w_4^2\right)$$
$$+\tfrac{1}{3}\left(238y_1^3-170w_1^3+210y_2^3-150w_2^3-36w_3^3-10w_4^3\right)$$

subject to

$x_1+x_2+x_3\le 500$

Scenario 1	*Scenario 2*	*Scenario 3*
$3x_1+y_1^1-w_1^1\ge 200$	$2.5x_1+y_1^2-w_1^2\ge 200$	$2x_1+y_1^3-w_1^3\ge 200$
$3.6x_2+y_2^1-w_2^1\ge 240$	$3x_2+y_2^2-w_2^2\ge 240$	$2.4x_2+y_2^3-w_2^3\ge 240$
$24x_3-w_3^1-w_4^1\ge 0$	$20x_3-w_3^2-w_4^2\ge 0$	$16x_3-w_3^3-w_4^3\ge 0$
$w_3^1\le 6000$	$w_3^2\le 6000$	$w_3^3\le 6000$

$x_i\ge 0$, i=1,2,3;

$y_i^k\ge 0$, i=1,2 & k=1,2,3;

$w_i^k\ge 0$, i=1,2,3,4 & k=1,2,3

Thus, all possible second-stage decisions are made simultaneously, in a single large LP.

Optimal Solution: Expected profit= $108,390

		Wheat	*Corn*	*Sugar Beets*
First stage	Plant:	170 Acres	80 Acres	250 Acres
k=1	Yield	510 T	288 T	6000 T
"Good yield"	Sales	310 T	48 T	6000 T
	Purchase	--	--	--
k=2	Yield	425 T	240 T	5000 T
"Fair yield"	Sales	225 T	--	5000 T
	Purchase	--	--	
k=3	Yield	340 T	192 T	4000 T
"Bad yield"	Sales	140 T	--	4000 T
	Purchase	--	48 T	--

- Using the original solution (where *expected values of yields* were assumed, i.e., planting 120 acres of wheat, 80 acres of corn, & 300 acres of beets) his expected profit would be $107,240 (which is $1,150 less than the optimal expected value) .
    - The *Expected Value of Perfect Information* is $115,406 - $108,390 = $7016

LAGRANGIAN DECOMPOSITION: "SPLITTING" FIRST-STAGE VARIABLES

For each scenario k, define a first-stage decision x^k which must equal the original first-stage decision (which we now denote by x^0). We can then write the equivalent LP:

$$Z = \min cx_0 + \sum_{k=1}^{K} p_k q_k y^k$$

subject to

$$x^0 \in X$$

In order to separate the LP by scenario, we need to "relax" the constraints

$$x^0 = x^k, k = 1, \ldots K;$$

LAGRANGIAN RELAXATION

Given a family of Lagrangian multiplier vectors λ_k, k=1,…K, we define the relaxation:

$$D(\lambda) = \min cx^0 + \sum_{k=1}^{K} p_k q_k y^k + \sum_{k=1}^{K} \lambda_k \left(x^k - x^0\right)$$

subject to $x^0 \in X$

$$T_k x^k + W y^k = h_k, k = 1, \ldots K;$$

$$x^k \geq 0, k = 1, \ldots K;\ y^k \geq 0, \text{k=1,2,...K}$$

That is,

$$D(\lambda) = \min \left(c - \sum_{k=1}^{K} \lambda_k \right) x^0 + \sum_{k=1}^{K} \left[\lambda_k x^k + p_k q_k y^k \right]$$

subject to the above constraints.

This is motivated by the fact that the problem then separates into K+1 subproblems:

$$D(\lambda) = D_0\left(\lambda_1, \ldots \lambda_K\right) + \sum_{k=1}^{K} D_k\left(\lambda_k\right)$$

where

$$D_0(\lambda) = \min \left(c - \sum_{k=1}^{K} \lambda_k \right) x^0$$

subject to $x^0 \in X$

and, for each k=1, …K:

$$D_k(\lambda) = \min \lambda_k x^k + p_k q_k y^k$$

subject to $T_k x^k + W y^k = h_k$

$$x^k \geq 0;\ y^k \geq 0$$

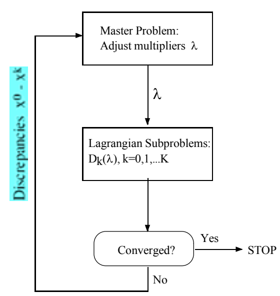
Dual Subproblem 0 for 1st Stage	Dual Subproblem for Scenario k, k=1, …K
$\min \left(c - \sum_{k=1}^{K} \lambda_k \right) x^0$ subject to $x^0 \in X$	Min $\lambda_k x^k + p_k q_k y^k$ subject to $T_k x^k + W y^k = h_k$, $x^k \geq 0, y^k \geq 0$

The value $D(\lambda) = D_0\left(\lambda_1, \ldots \lambda_K\right) + \sum_{k=1}^{K} D_k\left(\lambda_k\right)$ provides a *lower bound* on the optimal cost Z.

The *Lagrangian dual* problem is to select the multipliers which will produce the *tightest* such lower bound:

$$\widehat{D} = \max_{\lambda} D(\lambda)$$

Note: In the linear case, $\hat{D} = Z$ and there is no "duality gap".



BENDERS' DECOMPOSITION

Benders' partitioning (commonly known in stochastic programming as the "L-Shaped Method") achieves separability of the second stage decisions, but in a different manner. Given a first-stage decision x^0, solve for each scenario k=1, …K the second-stage LP:

$$P_k\left(x^0\right) = \min q_k y^k$$

subject to $W y^k = h_k - T_k x^0,\ y^k \geq 0$

Then $P\left(x^0\right) = cx^0 + \sum_{k=1}^{k} p_k P_k\left(x^0\right)$ provides us with an *upper* bound on the optimal cost Z, i.e.,

$$D(\lambda) \leq Z \leq P\left(x^0\right)$$

Furthermore, solving each LP provides us with a vector λ_k of dual variables corresponding to the constraints $x^0 = x^k$.

If π_k is the dual solution of the LP

$$P_k\left(x^0\right) = \min q_k y^k$$

subject to $W y^k = h_k - T_k x^0,\ y^k \geq 0$

then $\lambda_k = -T_k^{\mathrm{T}} \pi_k$

An aside: Computing λ_k:
The dual of

$$\text{Min } q_k y^k$$
subject to
$$T_k x^k + W y^k = h_k,$$
$$x^k = x^0,$$
$$x^k \ge 0$$

is the LP

$$\text{Max } h_k \pi_k + x^0 \lambda_k$$
subject to:
$$T_k^T \pi_k + I \lambda_k = 0$$
$$W^T \pi_k \le q_k$$

If we eliminate λ_k using the equality constraint, we obtain $\lambda_k = -T_k^{\mathrm{T}} \pi_k$ and the dual LP

$$\text{Max } \left(h_k - T_k x^0\right) \pi_k$$
subject to
$$W^T \pi_k \le q_k$$

The original problem now is seen to be equivalent to

$$\text{Min } cx^0 + \sum_{k=1}^{K} p_k P_k\left(x^0\right)$$
subject to $x^0 \in X$

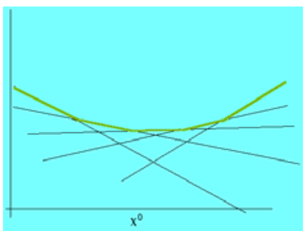
By making use of dual information obtained after M evaluations of $P_k\left(x^0\right)$, Benders' procedure forms an approximation (a convex piecewise-linear function) of $P_k\left(x^0\right)$:

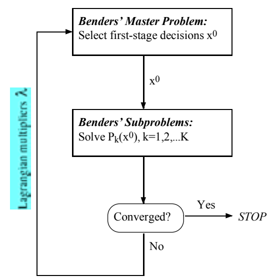
$$P_k\left(x^0\right) \ge \max_{i=1,\ldots M} \left\{\alpha_k^i x^0 + \beta_k^i\right\}$$

so that the original problem reduces (with introduction of new variables θ_k)to

$$\text{Min } cx^0 + \sum_{k=1}^{K} p_k \theta_k$$
subject to $x^0 \in X$
and
$$\theta_k \ge \alpha_k^i x^0 + \beta_k^i,\ \text{i=1, ...M; k=1, ...K}$$

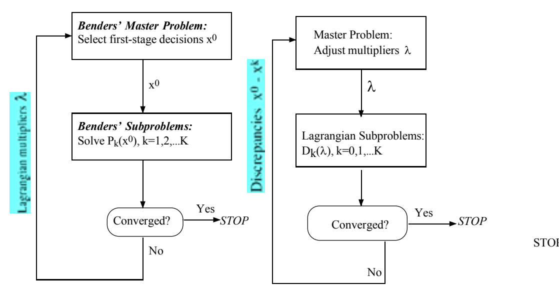
That is, we have approximated $P_k\left(x^0\right)$ by the maximum of a finite number of linear functions, i.e., by a *piecewise-linear convex function*:





In either the Lagrangian relaxation approach or Benders' decomposition, the burden of the computation lies in the respective master problems: searching for the optimal λ in the case of Lagrangian relaxation, & searching for the optimal x^0 in the case of Benders' decomposition.

The subproblems, being LPs separable by scenario, are easily solved in comparison.



Cross-Decomposition

Cross-decomposition is a hybrid of Benders' decomposition and Lagrangian relaxation, in which the subproblem of each algorithm serves the purpose of the master problem of the other.
That is, Benders' subproblem receives the first-stage decisions x^0 from the Dual subproblem D_0 rather than from the Benders' master problem.

Primal Subproblem for Scenario k	Information exchange	Dual Subproblem 0 for 1st Stage
Min $p_k q_k y^k$ subject to $Wy^k = \left(h_k - T_k x^0\right)$ $y^k \ge 0$ *(x^0 fixed)*	$\lambda = [\lambda_1, \ldots \lambda_L]$ → ← x_0	Min $\left(c - \sum_{k=1}^{K} \lambda_k\right) x^0$ subject to $x^0 \in X$

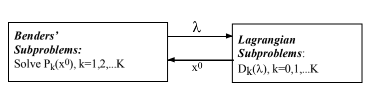
Likewise, the Dual subproblem D_0 receives the necessary multipliers λ from the Benders' subproblem, rather than from the Dual master problem.

Cross-Decomposition

Note that the algorithm can be "streamlined"-- only one of the dual subproblems $D_0(\lambda)$ needs to be solved at each iteration, except when the termination criterion

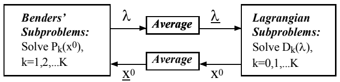
$$P\left(x^0\right) - D\left(\lambda\right) \le \varepsilon$$

is to be tested.



Mean Value Cross Decomposition

Convergence is improved if the mean of all previously generated Lagrangian multipliers and first-stage decisions are sent to the Lagrangian and Benders' subproblems, respectively.

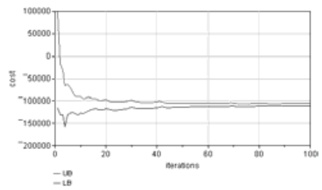


Example

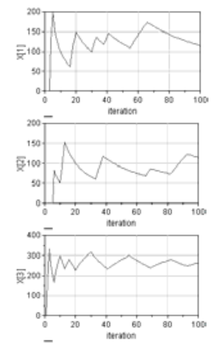
The cross-decomposition algorithm described above was implemented in the APL language (APL+WIN 3.0). First, the mean of all prior primal & dual solutions was used at each iteration. The result after 100 iterations was as follows:

```
Total cost: ¯106456.94, found at iteration #72

Best lower bound: ¯110752.17
Gap= 4295.23, or 4.03%

Stage One Variables:
 i       X[i]

 1      159.72
 2       83.33
 3      250.00
 4        6.94
```


The plot of upper & lower bounds at each iteration :



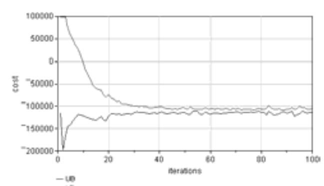
The mean values of first-stage variables used in the primal subproblems at each iteration.



As an alternative, exponential smoothing (with smoothing factor 10%) was used for both primal and dual solutions. After 100 iterations, the following was the best solution found:

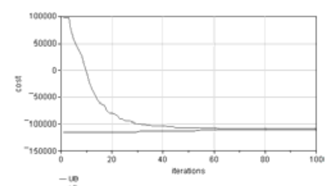
```
Total cost: ¯108210.7881, found at iteration #68

Best lower bound: ¯111187.0364
Gap= 2976.24833, or 2.750417387%

Stage One Variables:
 i       X[i]

 1      166.70
 2       81.90
 3      250.87
 4        0.14
```

This solution is very nearly optimal. (Optimal solution is –$108390.)



Benders' & Lagrangian subproblems at each iteration

Best upper & lower bounds at each iteration



Research Issues

1. Given that the number of scenarios is *extremely* large (or probability distributions are continuous and not discrete), how does one do "sampling" of scenarios in the cross-decomposition algorithm?
2. How can the cross-decomposition algorithm be extended to multi-(i.e., greater than 2) stages?
3. Given uncertainty in the parameters of the probability distributions describing future scenarios, perhaps it is not appropriate to continue iterations until the duality gap between upper & lower bounds is nearly zero-- can we determine an appropriate gap between upper & lower bound for a termination criterion for the cross-decomposition algorithm?
4. Case of integer first-stage decisions.
   - The Lagrangian subproblems $D_k(\lambda)$ for scenarios k=1,…K are now mixed-integer LP problems, which are substantially more difficult to solve.
   - The computational savings obtained by solving only the Lagrangian subproblem $D_0(\lambda)$ and not the Lagrangian subproblems $D_k(\lambda)$ for scenarios k=1,…K at every iteration become more significant!
   - The Lagrangian subproblems $D_k(\lambda)$ for scenarios k=1,…K may occasionally be solved, in order to test the duality gap as a termination criterion. How can information about the dual variables gathered from Benders' subproblems be accumulated in order to construct a Benders' master problem for each individual $D_k(\lambda)$?

References

A comprehensive textbook:

Birge, R. and F. Louveaux (1997). *Introduction to stochastic programming*. New York, Springer-Verlag.

Illustration of use of "variable splitting":

Guignard, M. and S. Kim (1987). "Lagrangean decomposition: a model yielding stronger Lagrangean bounds." *Mathematical Programming* **39**: 215-228.

Jornsten, K. and M. Nasberg (1986). "A new Lagrangian relaxation approach to the generalized assignment problem." *European Journal of Operational Research* **27**: 313-323.

Development of the Cross-Decomposition algorithm:

Van Roy, T. J. (1983). "Cross decomposition for mixed integer programming." *Mathematical Programming* **25**: 46-63.

Holmberg, K. (1990). "On the convergence of cross decomposition." *Mathematical Programming* **47**(2): 269-296.

Holmberg, K. (1997). "Mean value cross decomposition applied to integer programming problems." *European Journal of Operational Research* **97**(1): 124-138.