Progressive Hedging Innovations for a Class of Stochastic Recourse Allocation Problems

by Jean-Paul Watson, David L. Woodruff, and David R. Strip

Bjorn Berg
Yingying Wang

November 18, 2010

Outline

- Lagrangian Method and Subgradient Algorithm
- Introduction to Progressive Hedging
- Progressive Hedging Algorithm
- Example
- Algorithm Improvements
  - Computing effective ρ values
  - Accelerating convergence
  - Termination criteria
  - Detecting cyclic behavior

The Lagrangian Method

$$\begin{aligned} minimize \quad & f(x, y) \\ s.t. \quad & g(x, y) = c \end{aligned}$$

Introduce a new variable λ called Lagrange multiplier.
Define the Lagrange function as:

$$\Lambda(x, y, \lambda) = f(x, y) + \lambda(g(x, y) - c)$$

Solve

$$\nabla_{x,y,\lambda}\Lambda(x, y, \lambda) = 0$$

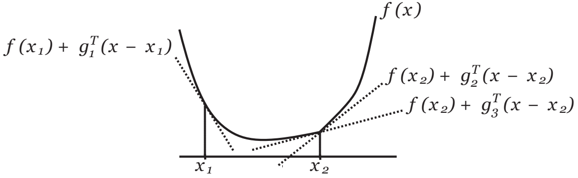
For N variables and M constraints,

$$L(x_1, x_2, ..., x_N, \lambda_1, \lambda_2, ...\lambda_M) = f(x_1, x_2, ..., x_N) + \sum_{k=1}^{M} \lambda_k[g_k(x_1, x_2, ..., x_N) - c_k]$$

Subgradient Algorithm

g is a subgradient of f (not necessarily convex) at x if

$$f(y) \geq f(x) + g^T(y - x) \quad \forall y$$



g_2,g_3 are subgradients at x_2, g_1 is a subgradient at x_1.

At each iteration one takes a step from the present point in the descending direction.

Progressive Hedging

- A scenario-based (horizontal) decomposition technique for solving stochastic programs
- Initially used for solving large-scale stochastic linear programs [Rochafellar and Wets, 1991]
  - Possesses convergence properties when decision variables are continuous
- Shown to be very efficient heuristic for stochastic mixed-integer programs
  - Robust to number of stages and number of variables in each stage
  - Lacks provable convergence and optimality gaps
- The basic idea is to relax the non-anticipitivity constraints and penalize the disagreement between scenario decision variables in the objective function

$$\begin{aligned} minimize \;\; & cx + \sum_{s \in S} Pr(s)(f_s \cdot y_s) \;\; (EF) \\ s.t. \;\; & (x, y_s) \in \mathcal{Q} \;\; \forall s \in S \\ & x_s = x \;\; \forall s \in S \end{aligned}$$

Progressive Hedging Algorithm

1. Solve each scenario problem
2. An implementable solution is obtained by averaging over all scenarios
3. Solve each scenario problem with an addition objective term penalizing the lack of implementability using a sub-gradient estimator for the non-anticipitivity constraints and a squared proximal term
4. If convergence is not sufficient, return to step 2
5. Post-process (e.g. round) to obtain a fully admissible and implementable solution

[Watson, Woodruff, and Hart. PySP: Modeling and Solving Stochastic Programs in Python. *Working Paper.* 2010.]

Progressive Hedging Algorithm

1. $k = 0$.
2. For all $s \in S$, $x_s^{(k)} = argmin_x(cx + f_s \cdot y_s) : (x, y_s) \in \mathcal{Q}$.
3. $\bar{x}^{(k)} = \sum_{s \in S} Pr(s) x_s^{(k)}$.
4. $w_s^{(k)} = \rho(x_s^{(k)} - \bar{x}^{(k)})$.
5. $k = k + 1$.
6. For all $s \in S$,

$$x_s^{(k)} = argmin_{x,y_s}(cx + w_s^{(k-1)}x + \rho/2||x - \bar{x}^{(k)}||^2 + f_s \cdot y_s) : (x, y_s) \in \mathcal{Q}$$

7. $\bar{x}^{(k)} = \sum_{s \in S} Pr(s) x_s^{(k)}$.
8. For all $s \in S$, $w_s^{(k)} = w_s^{(k-1)} + \rho(x_s^{(k)} - \bar{x}^{(k)})$.
9. $g^{(k)} = \sum_{s \in S} Pr(s)||x_s^{(k)} - \bar{x}^{(k)}||$.
10. If $g^{(k)} < \epsilon$, then go to Step 5.Otherwise, terminate.

Example

We can invest $\$10,000$ in either of two investments, A or B.
We would like a return of $\$25,000$.
The two scenarios are considered equally likely:
Scenario 1: A returns the initial investment while B returns 3 times the initial investment.
Scenario 2: A returns 4 times the initial investment and B returns twice the initial investment.
Objective: Squares any return less than $\$25,000$.

$$\begin{aligned} min \quad & 0.5(y_1^2 + y_2^2) \\ s.t. \quad & x_A + x_B \leq 10 \\ & x_A + 3x_B + y_1 \geq 25 \\ & 4x_A + 2x_B + y_2 \geq 25 \\ & x_A, x_B, y_1, y_2 \geq 0 \end{aligned}$$

Example(cont.)

Iteration 0: Set $\rho = 2$. Begin with $w^0 = 0$, $x_1^0 = (x_{1A}^0, x_{1B}^0) = (0, 10)^T$ and $x_2^0 = (x_{2A}^0, x_{2B}^0) = (10, 0)^T$. The initial value of $\bar{x}^0 = (5, 5)^T$.

Iteration 1:

Step 1.

Subproblem 1.

$$\begin{aligned} min \quad & 0.5[y_1^2 + (x_{1A}^1 - 5)^2 + (x_{1B}^1 - 5)^2] \\ s.t. \quad & x_{1A}^1 + x_{1B}^1 \leq 10 \\ & x_{1A}^1 + 3x_{1B}^1 + y_1 \geq 25 \\ & x_{1A}^1, x_{1B}^1, y_1 \geq 0 \end{aligned}$$

Solution:
$(x_{1A}^1, x_{1A}^1, y_1) = (10/3, 20/3, 5/3)$.

Subproblem 2.

$$\begin{aligned} min \quad & 0.5[y_2^2 + (x_{2A}^1 - 5)^2 + (x_{2B}^1 - 5)^2] \\ s.t. \quad & x_{2A}^1 + x_{2B}^1 \leq 10 \\ & 4x_{2A}^1 + 2x_{2B}^1 + y_2 \geq 25 \\ & x_{2A}^1, x_{2B}^1, y_2 \geq 0 \end{aligned}$$

Solution:
$(x_{2A}^1, x_{2A}^1, y_2) = (5, 5, 0)$.

We have $\bar{x}^1 = (25/6, 35/6)^T$.

Step 2. The new multiplier:
$w^1 = (w_{1A}^1, w_{1B}^1, w_{2A}^1, w_{2B}^1)^T = (-5/3, 5/3, 5/3, -5/3)^T$.

Example(cont.)

Iteration 0: Set $\rho = 2$. Begin with $w^0 = 0$, $x_1^0 = (x_{1A}^0, x_{1B}^0) = (0, 10)^T$ and $x_2^0 = (x_{2A}^0, x_{2B}^0) = (10, 0)^T$. The initial value of $\bar{x}^0 = (5, 5)^T$.

Iteration 1:

Step 1.

Subproblem 1.

$$\begin{aligned} min \quad & 0.5[y_1^2 + (x_{1A}^1 - 5)^2 + (x_{1B}^1 - 5)^2] \\ s.t. \quad & x_{1A}^1 + x_{1B}^1 \le 10 \\ & x_{1A}^1 + 3x_{1B}^1 + y_1 \ge 25 \\ & x_{1A}^1, x_{1B}^1, y_1 \ge 0 \end{aligned}$$

Solution:
$(x_{1A}^1, x_{1A}^1, y_1) = (10/3, 20/3, 5/3)$.

Subproblem 2.

$$\begin{aligned} min \quad & 0.5[y_2^2 + (x_{2A}^1 - 5)^2 + (x_{2B}^1 - 5)^2] \\ s.t. \quad & x_{2A}^1 + x_{2B}^1 \le 10 \\ & 4x_{2A}^1 + 2x_{2B}^1 + y_2 \ge 25 \\ & x_{2A}^1, x_{2B}^1, y_2 \ge 0 \end{aligned}$$

Solution:
$(x_{2A}^1, x_{2A}^1, y_2) = (5, 5, 0)$.

We have $\bar{x}^1 = (25/6, 35/6)^T$.

Step 2. The new multiplier:
$w^1 = (w_{1A}^1, w_{1B}^1, w_{2A}^1, w_{2B}^1)^T = (-5/3, 5/3, 5/3, -5/3)^T$.

Example(cont.)

Iteration 0: Set $\rho = 2$. Begin with $w^0 = 0$, $x_1^0 = (x_{1A}^0, x_{1B}^0) = (0, 10)^T$ and $x_2^0 = (x_{2A}^0, x_{2B}^0) = (10, 0)^T$. The initial value of $\bar{x}^0 = (5, 5)^T$.

Iteration 1:

Step 1.

Subproblem 1.

$$\begin{aligned} min \quad & 0.5[y_1^2 + (x_{1A}^1 - 5)^2 + (x_{1B}^1 - 5)^2] \\ s.t. \quad & x_{1A}^1 + x_{1B}^1 \leq 10 \\ & x_{1A}^1 + 3x_{1B}^1 + y_1 \geq 25 \\ & x_{1A}^1, x_{1B}^1, y_1 \geq 0 \end{aligned}$$

Solution:
$(x_{1A}^1, x_{1A}^1, y_1) = (10/3, 20/3, 5/3)$.

Subproblem 2.

$$\begin{aligned} min \quad & 0.5[y_2^2 + (x_{2A}^1 - 5)^2 + (x_{2B}^1 - 5)^2] \\ s.t. \quad & x_{2A}^1 + x_{2B}^1 \leq 10 \\ & 4x_{2A}^1 + 2x_{2B}^1 + y_2 \geq 25 \\ & x_{2A}^1, x_{2B}^1, y_2 \geq 0 \end{aligned}$$

Solution:
$(x_{2A}^1, x_{2A}^1, y_2) = (5, 5, 0)$.

We have $\bar{x}^1 = (25/6, 35/6)^T$.

Step 2. The new multiplier:
$w^1 = (w_{1A}^1, w_{1B}^1, w_{2A}^1, w_{2B}^1)^T = (-5/3, 5/3, 5/3, -5/3)^T$.

Example(cont.)

Iteration 2: *Step 1.*

Subproblem 1.

$$\begin{aligned} min \quad & y_1^2 - (5/3)(x_{1A}^2 - 25/6) \\ & + (5/3)(x_{1B}^2 - 35/6) \\ & + (x_{1A}^2 - 25/6)^2 \\ & + (x_{1B}^2 - 35/6)^2 \\ s.t. \quad & x_{1A}^2 + x_{1B}^2 \leq 10 \\ & x_{1A}^2 + 3x_{1B}^2 + y_1 \geq 25 \\ & x_{1A}^2, x_{1B}^2, y_1 \geq 0 \end{aligned}$$

Solution:
$(x_{1A}^2, x_{1A}^2, y_1) = (10/3, 20/3, 5/3)$.

Subproblem 2.

$$\begin{aligned} min \quad & y_2^2 + (5/3)(x_{2A}^2 - 25/6) \\ & - (5/3)(x_{2B}^2 - 35/6) \\ & + (x_{2A}^2 - 25/6)^2 \\ & + (x_{2B}^2 - 35/6)^2 \\ s.t. \quad & x_{2A}^2 + x_{2B}^2 \leq 10 \\ & 4x_{2A}^2 + 2x_{2B}^2 + y_2 \geq 25 \\ & x_{2A}^2, x_{2B}^2, y_2 \geq 0 \end{aligned}$$

Solution:
$(x_{2A}^2, x_{2A}^2, y_2) = (10/3, 20/3, 0)$.

We have $\bar{x}^2 = (10/3, 20/3)^T$.

Step 2. The new multiplier: $w^2 = w^1$.

Example(cont.)

k	$\hat{x}_A^k$	$\hat{x}_B^k$	$\rho_{1A}^k = -\rho_{2A}^k$	$\rho_{1B}^k = -\rho_{2B}^k$	x_{1A}^k	x_{1B}^k	x_{2A}^k	x_{2B}^k
0	5.0	5.0	0.0	0.0	3.33	6.67	5.0	5.0
1	4.17	5.83	-1.67	1.67	3.33	6.67	3.33	6.67
2	3.33	6.67	-1.67	1.67	3.06	6.94	2.50	7.50
3	2.78	7.22	-1.11	1.11	2.78	7.22	2.41	7.59
4	2.59	7.41	-0.74	0.74	2.65	7.35	2.41	7.59
5	2.53	7.47	-0.49	0.49	2.59	7.41	2.43	7.57
6	2.50	7.50	-0.33	0.33	2.56	7.44	2.45	7.55
7	2.50	7.50	-0.22	0.22	2.54	7.46	2.46	7.54
8	2.50	7.50	-0.15	0.15	2.53	7.48	2.48	7.52
9	2.50	7.50	-0.10	0.10	2.52	7.48	2.48	7.52
10	2.50	7.50	-0.07	0.07	2.51	7.49	2.49	7.51
11	2.50	7.50	-0.04	0.04	2.51	7.49	2.49	7.51
12	2.50	7.50	-0.03	0.03	2.50	7.50	2.50	7.50

Figure: PHA iterations

Effective ρ value computation

If the the magnitude of the cost vector, c, is high and the decision variables, x, are limited to small values, using a small ρ will result in slow convergence because the penalty term will be small.

General Heuristic

- The $\rho(i)$ used for variable i should be relative to the unit cost, $c(i)$, in magnitude.

Numerical Heuristic

- After iteration 0, for each variable x_i,

$$\rho(i) = \frac{c(i)}{(x^{max} - x^{min} + 1)}$$

in the discrete case, and

$$\rho(i) = \frac{c(i)}{\max((\sum_{s \in S} Pr(s)|x_s^{(0)} - \bar{x}^{(0)}), 1)}$$

in the continuous case

Convergence accelerators

- Fix decision variables once their value has stabilized over the last $\mu|S|$ iterations
- Slamming: fix decision variables that have *sufficiently* converged
- Solve a variation of the EF problem where the converged variables are fixed
- Aggressive variable fixing

$$x(i) = \max_{s \in S} x_s(i)$$

where $c(i) \times \max_{s \in S} x_s(i)$ is minimal

Termination criteria

Original termination criteria:

$$g^{(k)} = \sum_{s \in S} Pr(s) ||x_s^{(k)} - \bar{x}^{(k)}|| < \epsilon$$

or

$$td = (\sum_{i,s:\bar{x}(i)>0} \frac{|x_s(i) - \bar{x}(i)|}{\bar{x}(i)})/|S|$$

Problem: Small differences in $||x_s - x_{s'}||$ does not necessarily imply small differences in the costs $||cx_s - cx_{s'}||$.
Additional termination condition:
Let $x^{max}(i) = max_{s \in S}(x_s^{(k)}(i))$, $x^{min}(i) = min_{s \in S}(x_s^{(k)}(i))$.

$$qd = (cx^{max}/cx^{min}) * 100$$

Terminate if qd drops below a parameterized threshold value λ_q, e.g., $\lambda_q = 1\%$.

Detecting cyclic behavior

For all types of stochastic mixed-integer programs, there is a risk of non-convergence of the PH algorithm.

Cycle detection

- Repeated occurrence of $w_s(i)$ vectors.
- z_s: integer hash weight for each scenario $s \in \mathcal{S}$.
- Hash value $h(i) = \sum_{s \in \mathcal{S}} z_s w_s(i)$.
- Equal $h(i)$ implies potential cycle.

Avoidance mechanism

- Fix $x(i) = max_{s \in S} x_s(i)$.

Thank you!