BBPH: Using Progressive Hedging Within Branch and Bound to Solve Multi-Stage Stochastic Mixed Integer Programs

David L. Woodruff and Jason Barnett

University of California Davis

Jean-Paul Watson

Sandia National Laboratories

Berlin, January 2019

The paper (and talk) in one slide

- Progressive hedging (PH) is an iterative, scenario decomposition method for solving multi-stage stochastic programs (Rockefellar and Wets).
- PH alone is not guaranteed to convergenge for stochastic MIPs.
- We motivate and describe a provably convergent branch and bound algorithm that uses PH within each (outer) node.
- Computational experiments show that for some difficult problem instances BBPH can find improved solutions within a few branches (but that's not really the main thing).

A few B&B algorithms

- Shabbir Ahmed. A scenario decomposition algorithm for 0-1 stochastic programs.
  Technical report, ISYE, Georgia Tech, 2013
- Claus C Carøe and Rüdiger Schultz. Dual decomposition in stochastic integer programming.
  Operations Research Letters, 24(1):37–45, 1999
- Laureano F. Escudero, Araceli Garín, María Merino, and Gloria Pérez. Bfc-msmip: an exact branch-and-fix coordination approach for solving multistage stochastic mixed 0–1 problems.
  TOP, 17(1):96–122, 2009

Our Paper

J Barnett, JP Watson, DL Woodruff *Operations Research Letters* 45 (1), 2017, 34-39

Bounds

- Lower bound with roughly the same effort as a PH iteration
  D. Gade, G. Hackebeil, S.M. Ryan, J.P. Watson, R.J.B. Wets, and D.L. Woodruff. Obtaining lower bounds from the progressive hedging algorithm for stochastic mixed-integer programs. *Mathetmatical Programming Series B*, 157:47–67, 2016
- (not an integer relaxation)
- Upper bound, e.g., by computing the expected value of a scenario solution

- "Side effect:"
  Ge Guo, Gabriel Hackebeil, Sarah M Ryan, Jean-Paul Watson, and David L Woodruff. Integration of progressive hedging and dual decomposition in stochastic integer programs. *Operations Research Letters*, 43(3):311–316, 2015

Motivation

- "Unsolvable by PH" examples in the paper
- Theorem in the paper
- But mainly it puts you in the midst of an exact algorithm when you use PH

Sketch of PHBB

Let u be the non-leaf decisions and y the leaf-node decisions.

1. **Initialize:** Set $\bar{z} = \infty$ and $\underline{z}_0 = -\infty$. Set $\mathcal{L} \leftarrow \{(\text{SMIP})\}$.
2. **Choose node:** Select a node $N^i \in \mathcal{L}$. If $\mathcal{L} = \emptyset$, goto Step 5.
3. **Calculate bounds:** Remove N^i from $\mathcal{L}$. Run PH on the selected outer B&B node N^i. After each iteration of PH,
   1. Obtain $\hat{x} = (\hat{u}, \bar{y})$, calculate the corresponding objective value m; set $\bar{z} \leftarrow \min\{m, \bar{z}\}$. Remove nodes $N^i \in \mathcal{L} \cup \mathcal{N}$ with $\underline{z}_i > \bar{z}$.
   2. Compute bound $D(w_\nu)$. If $\bar{z} - D(w_\nu) < \epsilon$, terminate PH and return to step 2.
4. **Branch:** Select a non-fixed variable $u^1(i)$, and create subnodes N^{i0} and N^{i1} corresponding to the branches $u^1(i) \leq \bar{u}^1(i)$ and $u^1(i) > \bar{u}^1(i)$ respectively. If $u^1(i)$ are fixed $\forall i$, continue with $u^2(i)$ (and so on). If N^{i0} and N^{i1} are fully real-valued problems, add them to $\mathcal{N}$, else add them to $\mathcal{L}$. Return to step 2.
5. **Choose terminal node:** Select a node $N^i \in \mathcal{N}$ and continue to Step 6. If $\mathcal{N} = \emptyset$, terminate BBPH with the following: if $\bar{z} = \infty$, the problem is infeasible, otherwise our solution is $\hat{x}$ with objective value $\bar{z}$.
6. **Solve terminal node:** Remove N^i from $\mathcal{N}$. Run PH on N^i with

Three Potential Levels of Parallelism

1. "Outer" BBPH nodes
2. PH by scenario
3. BB nodes within the scenario MIP solves

Conclusions

- An implementation of PH and some test instances are available at `pyomo.org`
- BBPH was released with the paper.
- BBPH puts PH in an exact algorithm.

Mulistage Stochastic Formulation

(for people who already know most of the notation)

$$\min_x f_1\left(x^1\right) + \mathbb{E}\sum_{t=2}^{T} f_t\left(x^t; \vec{x}^{\,t-1}, \vec{\xi}^{\,t}\right) \tag{1}$$

$$\text{subject to } x(\xi) \in Y_\xi, \;\; \xi \in \Xi, \tag{2}$$

Multi-stage formulation with a tree

Followed by Sketch of PH Algorithm

Let $\mathcal{G}_t$ be the set of all scenario tree nodes for stage t and let $\mathcal{G}_t(\xi)$ be the node at time t for a particular scenario, ξ. For a particular node $\mathcal{D}$ let $\mathcal{D}^{-1}$ be the set of scenarios that define the node.
In the presence of a scenario tree, non-anticipativity must be enforced at each non-leaf node, so using the discrete scenario tree notation, problem (1) becomes

$$\min_{x,\hat{x}} \sum_{\xi \in \Xi} \pi_\xi \left[f_1(x^1(\xi)) + \sum_{t=2}^{T} f_t \left(x^t(\xi); \vec{x}^{\,t-1}, \vec{\xi}^{\,t} \right) \right] \tag{3}$$

$$\text{s.t. } x(\xi) \in Y_\xi, \quad \xi \in \Xi \tag{4}$$

$$x^t(\xi) - \hat{x}^t(\mathcal{D}) = 0, \quad t = 1, \ldots, T-1, \ \mathcal{D} \in \mathcal{G}_t, \ \xi \in \mathcal{D}^{-1} \tag{5}$$

1. **Initialization:** Let $\nu \leftarrow 0$ and $w_\nu(\mathcal{G}_t(\xi)) \leftarrow 0$, $\forall \xi \in \Xi$, $t = 1, \dots, T$. Compute for each $\xi \in \Xi$:

$$x_{\nu+1}(\xi) \in \arg\min_{x \in X(\xi)} \sum_{\xi \in \Xi} \pi_\xi \left[f_1(x^1(\xi)) + \sum_{t=2}^{T} f_t\left(x^t(\xi); \vec{x}^{\,t-1}, \vec{\xi}^{\,t}\right) \right].$$

2. **Iteration Update:** $\nu \leftarrow \nu + 1$.
3. **Aggregation**: For each $t = 1, \dots, T-1$ and each $\mathcal{D} \in \mathcal{G}_t$:

$$\bar{x}^t_\nu(\mathcal{D}) \leftarrow \left(\sum_{\hat{\xi} \in \mathcal{D}^{-1}} \pi_{\hat{\xi}} x^t_\nu(\mathcal{G}_t(\xi)) \right) \Big/ \left(\sum_{\hat{\xi} \in \mathcal{D}^{-1}} \pi_{\hat{\xi}} \right).$$

4. **Weight Update:** For each $t = 1, \dots, T-1$ and each $\xi \in \Xi$:

$$w_\nu(\mathcal{G}_t(\xi)) \leftarrow w_{\nu-1}(\mathcal{G}_t(\xi)) + \rho[x^t_\nu(\mathcal{G}(\xi)) - \bar{x}_\nu(\mathcal{G}^t(\xi))].$$

5. **Decomposition:** For each $\xi \in \Xi$: assign $x_{\nu+1}(\xi) \in \arg\min_{x \in X(\xi)}$

$$f_1(x^1(\xi)) + \sum_{t=2}^{T} f_t\left(x^t(\xi); \vec{x}^{\,t-1}, \vec{\xi}^{\,t}\right) + \sum_{t=1}^{T-1} \left[\langle w^t_\nu(\xi), x^t \rangle + \frac{\rho}{2} \| x^t - \bar{x}^t_\nu(\xi) \|^2 \right].$$

6. **Termination criterion:** If the solutions at the tree nodes are equal (up to a given tolerance ϵ) or the maximum iteration count is reached, stop. Otherwise, return to step 2.

Test results for network flow minimization problems; times are wall-clock seconds.

	EF	PH		BBPH		PH Bundles	
Instance	Objective	Time	Obj.	Time	Obj.	Time	Obj.
1ef50	158653	2579	166848	18161	162163	31891	162946
2ef50	151060	1172	156211	15330	156211	9486	154065
3ef50	161466	2843	167871	33390	165733	30228	166174
4ef50	153854	2296	157229	18150	157229	8349	157697
5ef50	150401	1190	155002	8556	152686	7841	156109

Test results for three stage network flow with a branching factor of ten for the second stage and three for the third; times are wall-clock seconds.

	EF		PH		BBPH	
Instance	Time	Objective	Time	Objective	Time	Objective
1ef10	10,046	160,964	1,934	166,189	18,606	163,647
2ef10	10,045	156,637	2,065	160,849	13,767	160,052
3ef10	3,424	157,025	1,748	166,406	13,581	160,032
4ef10	2,327	170,067	1,428	191,796	13,786	176,127
5ef10	4,295	161,840	2,432	169,287	14,911	168,542