Toward Scalable Stochastic Unit Commitment

David L. Woodruff

University of California Davis

Paris, June 2014

Outline

A Teaser

Let's begin with a few computational results. I will leave out a lot of details.

Carrion and Arroyo

- We use Carrion and Arroyo as a staring point for the formulation
- No network constraints
- Validated against the Alstom test models

WECC-240 Family

- We use *WECC-240-r1* (85 generators) for purposes of parameter tuning and analysis.
- Then fix the PH configuration and examine performance on the out-of-sample and more realistic *WECC-240-r2* and *WECC-240-r3* cases.
- We analyze scalability to the larger *WECC-240-r2-x2* (170 generators) and *WECC-240-r2-x4* (340 generators) cases.
- Using modest scale parallelism.
- (These are harder to solve than ISO-NE instances of similar size.)

Extensive form

Solution quality statistics for the extensive form of the *WECC-240-r1* instance, given 2 hours of run time.

Scenarios	Obj Value	MIP LB	Gap %	Run Time (s)
3	64279.18	63708.67	0.89	7291
5	62857.52	62052.75	1.26	7309
10	61873.01	60769.78	1.77	7444
25	61496.24	59900.40	2.59	7739
50	61911.74	59432.08	4.01	8279
100	62388.85	3500.70	94.39	9379

Larger Instances using PH

Solve time (in seconds) and solution quality statistics for PH executing on 50-scenario instances.

Instance	Convergence	Obj. Value	PH L.B.	Time
		Red Sky Results		
WECC-240-r2-x2	0.0 (in 22 iters)	117794.429	116538.868	741
WECC-240-r2-x4	0.0 (in 19 iters)	232189.338	228992.984	1421

Stages and Scenarios

- Use $t \in 1, \ldots, T$ to index stages
- Random variable, which may be vector valued, $\boldsymbol{\xi}_t$,
- The symbol $\overrightarrow{\xi}^t$ to refer to the realized values of all random variables up to and including stage t.
- A full realization of the uncertainty, i.e.,

$$\overrightarrow{\xi}^T = \left(\xi^t,\ t = 2, \ldots, T\right)$$

is a *scenario*.

Abstract Formulations

- Use x^t to represent the part of the decision vector that corresponds to stage t.
- (Aside: For two-stage, many authors use x and y instead).
- Use $\vec{x}^{\,t}$ for $1 \leq t \leq T$ to represent the decisions for all stages up to, and including, stage t.

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

Two-stage Linear

$$\begin{array}{ll} \min_x & c^T x + E[Q(x, \boldsymbol{\xi})] \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array}$$

where $Q(x, \xi)$ is the optimal value of the second-stage problem

$$\begin{array}{ll} \min_y & q(\xi)^T y \\ \text{subject to} & T(\xi)x + W(\xi)y = h(\xi) \\ & y \geq 0 \end{array}$$

In this special two-stage case, most authors do not use a stage superscript for ξ because it appears only in the second stage.

Scenario Trees

- $\boldsymbol{\xi} = \{\boldsymbol{\xi}^t\}_{t=1}^T$ is defined on a discrete probability space.
- Scenario probability π_ξ.
- We organize realizations, ξ, into a tree with the property that scenarios with the same realization up to stage t share a node at that stage. so, $\vec{\xi}^t$ refers also to a node.
- Let $\mathcal{G}_t$ be the scenario tree nodes for stage t and let $\mathcal{G}_t(\xi)$ be the node at time t for scenario ξ.
- For a particular node $\mathcal{D}$ let $\mathcal{D}^{-1}$ be the set of scenarios that define the node.

Three-stage NewsVendor

- Basic idea: Uncertain Supply
- An order is placed (and paid for) and then some quantity is delivered that is a fraction of the quantity ordered
- At the time of delivery additional inventory can be acquired at a cost that was not exactly known when the first-stage order was placed.
- Finally, demand is realized.

- The quantity delivered in response to the first-stage order and the second-stage order cost are revealed before the second-stage order decision is made.
- The demand quantity is revealed before the third-stage auxiliary variable for sales quantity is computed. There are three stages of decisions: first-stage order quantity, second-stage order quantity, third-stage quantity sold.
- The final "decision" is quite easy to make since if the price is positive it is optimal to supply the demand if there is enough inventory, or the entire inventory if not.

3-Stage Data

Suppose we use the following symbols for the data and decision variables:

- p: selling price per item (data)
- c: stage one cost per item (data)
- ξ_1^2: fraction of stage one order delivered (uncertain data)
- ξ_2^2: cost per item ordered in the second stage (uncertain data)
- ξ^3: demand (uncertain data)
- q^1: stage one order quantity (decision variable)
- q^2: stage two order quantity (decision variable)
- q^3: quantity delivered to customer (auxiliary decision variable)

Functions

$$f_1\left(q^1\right) := c\, q^1$$

$$f_2\left(q^2;\, \vec{q}^{\,1},\, \vec{\xi}^{\,2}\right) := \xi^2 q^2$$

$$f_3\left(q^3;\, \vec{q}^{\,2},\, \vec{\xi}^{\,3}\right) := -p\ \min\left\{\xi^1 q^1 + q^2, \xi^3\right\}$$

although we might prefer to write

$$f_3(q^3;\, \vec{q}^{\,2},\, \vec{\xi}^{\,3}) := -p\, q^3 \text{ s.t. } q^3 \leq \xi^1 q^1 + q^2 \text{ and } q^3 \leq \xi^3$$

Data Values

Suppose, for the sake of illustration, that we have estimated the following data:

$$p = 100$$

$$c = 70$$

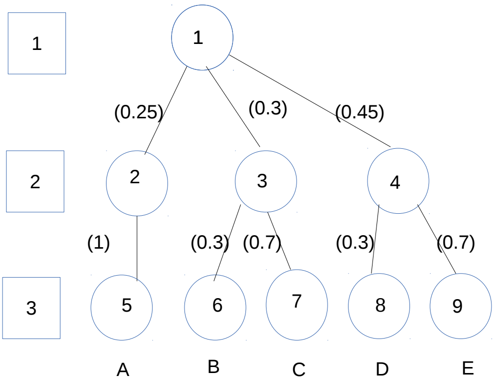
$$\xi^2 = \begin{cases} (1, 80) & \text{with probability } 0.25 \\ (0.6, 85) & \text{with probability } 0.30 \\ (0.6, 95) & \text{with probability } 0.45 \end{cases}$$

ξ^3:

If $\xi^2 = (1, 80)$ then $\xi^3 = 250$ with probability 1
Else:

$$\xi^3 = \begin{cases} 300 & \text{with probability } 0.3 \\ 400 & \text{with probability } 0.7 \end{cases}$$

Sample Tree



Tree Notation examples

the set of nodes at stage two, $\mathcal{G}_2$, is $\{2, 3, 4\}$;

the stage three node for scenario B, $\mathcal{G}_3(B)$, is 6;

if we consider node $\mathcal{D} = 3$, then the scenario set $\mathcal{D}^{-1}$ is $\{B, C\}$;

the probability of scenario B, π_B, is $(0.3)(0.3) = 0.09$;

for example.

Scenario Tree Formulations

Non-anticipativity must be enforced at each non-leaf node:

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

$$\pi_\xi x^t(\xi) - \pi_\xi \hat{x}^t(\mathcal{D}) = 0, \;\; t = 1, \ldots, T-1, \; \mathcal{D} \in \mathcal{G}_t, \; \xi \in \mathcal{D}^{-1} \tag{3}$$

Progressive Hedging

1: **Initialization**: Let $\nu \leftarrow 0$ and $w^{\nu}(\xi) \leftarrow 0, \ \forall \xi \in \Xi, \ t = 1, \ldots, T$. Compute for each $\xi \in \Xi$:

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

2: **Iteration Update**: $\nu \leftarrow \nu + 1$

3: **Aggregation**: Compute for each $t = 1, \ldots, T-1$ and each $\mathcal{D} \in \mathcal{G}_t$:

$$\bar{x}^{(\nu)}(\mathcal{D}) \leftarrow \sum_{\xi \in \mathcal{D}^{-1}} \pi_\xi x^{t,(\nu)}(\xi) / \sum_{\xi \in \mathcal{D}^{-1}} \pi_\xi$$

4: **Price Update**: Compute for each $t = 1, \ldots, T-1$ and each $\xi \in \Xi$

$$w^{(t,\nu)}(\xi) \leftarrow w^{(t,\nu-1)}(\xi) + \rho \left[x^{t,(\nu)}(\xi) - \bar{x}^{\nu}(\mathcal{G}_t(\xi))\right]$$

5: **Decomposition**: Compute for each $\xi \in \Xi$

$$\begin{aligned} x^{(\nu+1)}(\xi) &\in \arg\min\nolimits_x f_1(x^1) \\ &+ \textstyle\sum_{t=2}^{T} f_t\left(x^t; \ \vec{x}^{\,t-1}(\xi), \ \vec{\xi}^{\,t}\right) \\ &+ \textstyle\sum_{t=1}^{T-1} \left[w^{(t,\nu)}(\xi)^\top x^t + \frac{\rho}{2}\|x^t - \bar{x}^{(\nu)}(\mathcal{G}_t(\xi))\|^2\right] \end{aligned}$$

6: **Termination**: If a criterion is met, Stop. Otherwise goto step 2

Bounds Paper

"Obtaining Lower Bounds from the Progressive Hedging Algorithm for Stochastic Mixed-Integer Programs"
Dinakar Gade, Gabriel Hackebeil, Sarah M. Ryan, Jean-Paul Watson, Roger J-B Wets, and David L. Woodruff

Two-Stage Stochastic MIP

$$\min \; c^\top x + \mathbb{E}[f(x,\tilde{\xi})] \tag{4}$$
$$\text{s.t. } Ax \geq b, \tag{5}$$

where $\tilde{\xi}$ is a random vector defined on a probability space $(\Xi, \mathcal{A}, \mathcal{P})$ and for a particular realization ξ of $\tilde{\xi}$, $f(x,\xi)$ is defined as:

$$f(x,\xi) = \min \; g(\xi)^\top y \tag{6}$$
$$\text{s.t. } Wy \geq r(\xi) - T(\xi)x. \tag{7}$$

$x \in \mathbb{Z}_+^{p_1} \times \mathbb{R}^{n_1 - p_1}$	First-stage variables
$y(\xi) \in \mathbb{Z}_+^{p_2} \times \mathbb{R}^{n_2 - p_2}$	Second-stage variables
$c \in \mathbb{Q}^{n_1}$, $g(\xi) \in \mathbb{Q}^{n_2}$	First- and second-stage costs
$A \in \mathbb{Q}^{m_1 \times n_1}$, $b \in \mathbb{Q}^{m_2}$	First-stage constraint coefficients and right-hand-sides
$W \in \mathbb{Q}^{m_2 \times n_2}$	Recourse matrix
$T(\xi) \in \mathbb{Q}^{m_2 \times n_1}$, $r(\xi) \in \mathbb{Q}^{m_2}$	Technology matrix and right-hand-sides

Scenario Formulation of Two-Stage SMIP

$$\min \sum_{\xi \in \Xi} p_\xi \left[c^\top x(\xi) + g(\xi)^\top y(\xi) \right] \tag{8}$$

$$\text{s.t. } x(\xi) - \hat{x} = 0, \quad \xi \in \Xi \tag{9}$$

$$Ax(\xi) \geq b, \quad \xi \in \Xi \tag{10}$$

$$Wy(\xi) \geq r(\xi) - T(\xi)x, \quad \xi \in \Xi \tag{11}$$

$$x(\xi) \in \mathbb{Z}_+^{p_1} \times \mathbb{R}^{n_1 - p_1} \quad \xi \in \Xi \tag{12}$$

$$y(\xi) \in \mathbb{Z}_+^{p_2} \times \mathbb{R}^{n_2 - p_2} \quad \xi \in \Xi \tag{13}$$

$$\hat{x} \in \mathbb{Z}_+^{p_1} \times \mathbb{R}^{n_1 - p_1} \tag{14}$$

p_ξ	Probability of scenario ξ
$\hat{x}$	Variable to model non-anticipativity
(9)	Non-anticipativity constraints
$X(\Xi)$	Feasible set for scenario ξ defined by (10)-(13)

New Lower Bound

Proposition 1: The price system $w(\xi)$ defines implicit lower bounds

Let z^* be the optimal objective function value of the stochastic program. Let $w(\xi) \in \mathbb{R}^n$ be such that $\sum_{\xi \in \Xi} p_\xi w(\xi) = 0$ (component-wise). Let

$$D_\xi(w(\xi)) := \min_{(x(\xi),y(\xi)) \in X(\xi)} \left(c^\top x(\xi) + g(\xi)^\top y(\xi) + w(\xi)^\top x(\xi) \right).$$

Then $D(w) := \sum_{\xi \in \Xi} p_\xi D_\xi(w(\xi)) \le z^*$.

- PH weights satisfy $\sum_{\xi \in \Xi} p_\xi w^\nu(\xi) = 0$ for every ν.
- Every so often, use the current weights to compute $D(w)$.

New Lower Bound

Proposition 1: The price system $w(\xi)$ defines implicit lower bounds

Let z^* be the optimal objective function value of the stochastic program. Let $w(\xi) \in \mathbb{R}^n$ be such that $\sum_{\xi \in \Xi} p_\xi w(\xi) = 0$ (component-wise). Let

$$D_\xi(w(\xi)) := \min_{(x(\xi), y(\xi)) \in X(\xi)} \left(c^\top x(\xi) + g(\xi)^\top y(\xi) + w(\xi)^\top x(\xi) \right).$$

Then $D(w) := \sum_{\xi \in \Xi} p_\xi D_\xi(w(\xi)) \leq z^*$.

- PH weights satisfy $\sum_{\xi \in \Xi} p_\xi w^\nu(\xi) = 0$ for every ν.
- Every so often, use the current weights to compute $D(w)$.

New Lower Bound

Proposition 1: The price system $w(\xi)$ defines implicit lower bounds

Let z^* be the optimal objective function value of the stochastic program. Let $w(\xi) \in \mathbb{R}^n$ be such that $\sum_{\xi \in \Xi} p_\xi w(\xi) = 0$ (component-wise). Let

$$D_\xi(w(\xi)) := \min_{(x(\xi),y(\xi)) \in X(\xi)} \left(c^\top x(\xi) + g(\xi)^\top y(\xi) + w(\xi)^\top x(\xi) \right).$$

Then $D(w) := \sum_{\xi \in \Xi} p_\xi D_\xi(w(\xi)) \leq z^*$.

- PH weights satisfy $\sum_{\xi \in \Xi} p_\xi w^\nu(\xi) = 0$ for every ν.
- Every so often, use the current weights to compute $D(w)$.

New Lower Bound

Proposition 1: The price system $w(\xi)$ defines implicit lower bounds

Let z^* be the optimal objective function value of the stochastic program. Let $w(\xi) \in \mathbb{R}^n$ be such that $\sum_{\xi \in \Xi} p_\xi w(\xi) = 0$ (component-wise). Let

$$D_\xi(w(\xi)) := \min_{(x(\xi),y(\xi)) \in X(\xi)} \left(c^\top x(\xi) + g(\xi)^\top y(\xi) + w(\xi)^\top x(\xi) \right).$$

Then $D(w) := \sum_{\xi \in \Xi} p_\xi D_\xi(w(\xi)) \leq z^*$.

- PH weights satisfy $\sum_{\xi \in \Xi} p_\xi w^\nu(\xi) = 0$ for every ν.
- Every so often, use the current weights to compute $D(w)$.

How Tight is this Bound?

Ordinary Lagrangian by dualizing non-anticipativity:

$$L(x,\lambda) := \sum_{\xi\in\Xi} p_\xi(c^\top x(\xi) + g(\xi)^\top y(\xi) + \lambda(\xi)^\top x(\xi) - \lambda(\xi)^\top \hat{x}),\ x(\xi), y(\xi)) \in X(\xi), \forall \xi \in \Xi$$

$$\text{Primal: } F(\lambda) = \min_x L(x,\lambda) \quad (15)$$

$$\text{Dual: } z_{LD} := \sup_\lambda F(\lambda) \quad (16)$$

Theorem: [Carøe & Schultz 99]

$$z_{LD} = \min \sum_{\xi\in\Xi} p_\xi[c^\top x(\xi) + g(\xi)^\top y(\xi)] \quad (17)$$

$$\text{s.t. } x(\xi), y(\xi) \in \text{clconv}(X(\xi)), \xi \in \Xi$$

$$x(\xi) - \hat{x} = 0, \xi \in \Xi$$

How Tight is this Bound?

Ordinary Lagrangian by dualizing non-anticipativity:

$$L(x,\lambda) := \sum_{\xi\in\Xi} p_\xi (c^\top x(\xi) + g(\xi)^\top y(\xi) + \lambda(\xi)^\top x(\xi) - \lambda(\xi)^\top \hat{x}),\ x(\xi), y(\xi)) \in X(\xi), \forall \xi \in \Xi$$

$$\text{Primal: } F(\lambda) = \min_x L(x,\lambda) \quad (15)$$

$$\text{Dual: } z_{LD} := \sup_\lambda F(\lambda) \quad (16)$$

Theorem: [Carøe & Schultz 99]

$$z_{LD} = \min \sum_{\xi\in\Xi} p_\xi [c^\top x(\xi) + g(\xi)^\top y(\xi)] \quad (17)$$

$$\text{s.t.} \quad x(\xi), y(\xi) \in \text{clconv}(X(\xi)), \xi \in \Xi$$

$$x(\xi) - \hat{x} = 0, \xi \in \Xi$$

How Tight is this Bound?

Ordinary Lagrangian by dualizing non-anticipativity:

$$L(x,\lambda) := \sum_{\xi\in\Xi} p_\xi (c^\top x(\xi) + g(\xi)^\top y(\xi) + \lambda(\xi)^\top x(\xi) - \lambda(\xi)^\top \hat{x}),\ x(\xi), y(\xi)) \in X(\xi), \forall \xi \in \Xi$$

$$\text{Primal: } F(\lambda) = \min_x L(x,\lambda) \tag{15}$$

$$\text{Dual: } z_{LD} := \sup_\lambda F(\lambda) \tag{16}$$

Theorem: [Carøe & Schultz 99]

$$\begin{aligned} z_{LD} = \min \quad & \sum_{\xi\in\Xi} p_\xi [c^\top x(\xi) + g(\xi)^\top y(\xi)] \\ \text{s.t.} \quad & x(\xi), y(\xi) \in \operatorname{clconv}(X(\xi)), \xi \in \Xi \\ & x(\xi) - \hat{x} = 0, \xi \in \Xi \end{aligned} \tag{17}$$

Can Be As Tight as Possible!

Proposition 2: Best PH bound equals best bound from dual decomposition

Suppose PH is applied to (17). Then in the limit, one obtains a solution $(\hat{x}^*, w^*(\xi))$, *where* $\hat{x}^*$ *solves the primal L.R. and* $\{w^*(\xi), \xi \in \Xi\}$ *solves the dual L.R. Moreover, in the limit, the PH lower bound is equal to* z_{LD}.

- Only the duality gap remains.
- PH can be interpreted as a primal-dual algorithm.
- Sequences of primal solutions $\{\hat{x}^\nu\}_{\nu=1}^\infty$ and dual solutions $\{\{w^\nu(\xi)\}_{\nu=1}^\infty, \xi \in \Xi\}$ converge to a saddle point of the ordinary Lagrangian.

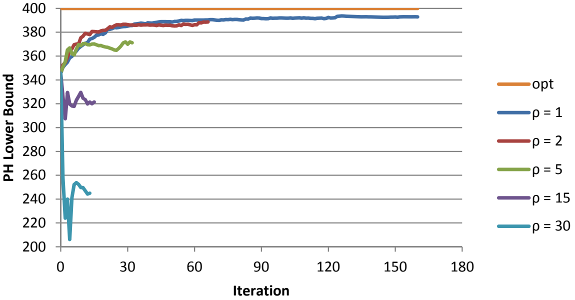
Can Be As Tight as Possible!

Proposition 2: Best PH bound equals best bound from dual decomposition

Suppose PH is applied to (17). Then in the limit, one obtains a solution $(\hat{x}^*, w^*(\xi))$*, where* $\hat{x}^*$ *solves the primal L.R. and* $\{w^*(\xi), \xi \in \Xi\}$ *solves the dual L.R. Moreover, in the limit, the PH lower bound is equal to* z_{LD}*.*

- Only the duality gap remains.
- PH can be interpreted as a primal-dual algorithm.
- Sequences of primal solutions $\{\hat{x}^\nu\}_{\nu=1}^\infty$ and dual solutions $\{\{w^\nu(\xi)\}_{\nu=1}^\infty, \xi \in \Xi\}$ converge to a saddle point of the ordinary Lagrangian.

Example (5 generator UC)



PH Implementation Issues

"Progressive hedging innovations for a class of stochastic mixed-integer resource allocation problems," JP Watson, DL Woodruff, Computational Management Science 8 (4), pp. 355-370

- There are some research topics?

Termination

- Look at scaled, average deviation from the average
- Look at lower bound?
- Backtracking?
- Repair?

Primary and Auxiliary variables

If you have variables that are strictly a function of other variables, then do not require non-anticipativity for the auxiliary variables.

Parallel Processing

- Trivial, but not trivially scalable for MIPs
- Bundling can help with variance
- Asynchronous parallel has sampling issues?

Bundling Scenarios

- Combine scenarios into bundles so that every scenario is in exactly one bundle.

- In steps 1 and 5, solve the EF for each bundle.
- Let $\mathcal{B}(s)$ represent the scenario index bundle that contains the scenario index s.
- E.g., $\mathcal{S} = \{1, 2, 3, 4, 5\}$. One might choose to form bundles $\{1, 3, 4\}$ and $\{2, 5\}$, and $\mathcal{B}(2) = \{2, 5\}$, etc.

Fixing and Slamming

- The following acceleration methods are designed for one-sided constraints, such as when the problem for each scenario is to minimize

$$c \cdot x$$

  subject to

$$Ax \geq b$$

  with $x \geq 0$ where the elements of vectors c and b and the matrix A are all non-negative.

- Fix variables that seem to have converged.
- Slamming — fix low cost variables that have not converged.
- Slamming can be used to force termination.

Heuristics and Sub-problem solvers

- The mipgap can be steadily decreased to a final mip gap (linearly in the convergence metric?)
- Heuristic sub-problem solvers can also be used with increasing effort and can use $\hat{x}$ for guidance?

Setting ρ

- Techniques for selecting ρ in proportion to the unit-cost of the associated decision variable and an alternative, mathematically-based heuristic approach.
- An effective ρ value should have something to do with $\delta f(x)/\delta x$?
- The best ρ value for a given problem need *not* be fixed at a constant value i.e., the introduction of per-iteration ρ_i may in fact be more appropriate for some problems?
- Scenario or iteration dependent schemes?

It is easy for UC

For a given thermal generator g, we compute the production cost $\overline{p}_g$ associated with the average power output level. We then introduce a global scaling factor α, and compute generator-specific values $\rho_g = \alpha \overline{p}_g$.

General Motivation

- Consider a scalar x for which non-anticipativity must be enforced.
- So only a scalar w weight multiplier is required.
- Suppose x is constrained to be an integer taking on small values and that at optimality $w = w^*$ is quite large.
- So if ρ is small, this situation will result in many iterations required for convergence of PH because at each iteration, w can grow only by the product of two small quantities.
- Set ρ so updates move fairly quickly to a "good" value w^* of the weight w.
- For practical reasons, we want the magnitude of w to approach from below in order to minimize oscillation or thrashing.

A Simple Formula

- Consider a single decision variable x with corresponding cost coefficient c (scenarios s).
- After iteration zero of PH completes, we have an estimate of the optimal value for x, which is $\bar{x}^{(0)}$.
- If we set a value of ρ that will result in $w = c$, then the proximal term

$$\rho/2 \left\| x_s^{(k-1)} - \bar{x}^{(k-1)} \right\|^2$$

will force the solution to be $\bar{x}^{(0)}$ in the subsequent PH iteration.

- The value of w is updated by

$$w_s^{(k)} := w_s^{(k-1)} + \rho \left(x_s^{(k-1)} - \bar{x}^{(k-1)} \right)$$

so the value of ρ for a given scenario s resulting in $w = c$ is

$$\rho_s := \frac{c}{\left| x_s - \bar{x}^{(0)} \right|}$$

SEP

- If w elements approach their ultimate value from below that mitigates thrashing or cycling.
- so we use a bound on the denominator and drop the dependence on s. After PH iteration 0, for each variable x we define $x^{\max} = \max_{s \in \mathcal{S}} x_s^{(0)}$ and $x^{min} = \min_{s \in \mathcal{S}} x_s^{(0)}$. Since $(x^{\max} - x^{\min} + 1) > |x_s - \bar{x}|$ we use

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

  for integer variable i.
- Continuous variables can change gradually without large discrete jumps so instead use the following formula to determine $\rho(i)$ for the continuous variables:

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

- Parameter free (but therefore, beatable)

Cost Proportional ρ

- Set $\rho(i)$ equal to a multiplier $k > 0$ of the element unit cost $c(i)$.
- Better: somehow estimate $\delta f(x)/\delta x$ and make ρ a fraction of that?

Generating Scenarios: Overview

- Most forecasts are given as a single (vector) point
- To generate scenarios we need either:
  1. Good probabilistic forecasts, or
  2. Analysis of the error distributions of point forecasts.
- A little notation:
  - The value (perhaps vector) of interest ℓ
  - Leading indicator (if there is one) w
  - Forecast function (if there is one) $\ell(w)$

Probabilistic Forecasts

- One way: fit a function $\ell(w)$, then find a way to generate w forecasts (or values) with known probabilities
- Another way: have multiple forecast functions and assign probabilities to each.

Analysis of Error Distributions

- You need, of course, a history of forecasts (or leading indicators) and corresponding observations.
- It makes sense to group "similar conditions" thereby creating error distributions that are conditional on the grouping.
- If you are fitting your own forecast function, you can also segment the data based on forecast error characteristics and fit forecast that are conditional on the error category.

Similar Conditions

E.g.,

- Forecast level quantiles
- Derivative patterns
- Weather quantiles

Scenarios from Error Distributions

- Using cutting points of the distribution to get skeleton points for the scenarios from the center of the error distribution between the cutting points.
- The probability of the skeleton points is trivial to compute from the cutting points.
- Scenarios are formed by connecting skeleton points, if necessary.

Conclusion

- Due to renewables, uncertainty may need to be handled explicity, rather than via reserves.
- To use Stochastic Programming, one needs algorithms and scenario generation methods
- We have proposed both
- Research continues
- We are estimating the potential for energy savings using ISO-NE as a test platform