

OPTIMAL SCENARIO TREE REDUCTIONS FOR THE STOCHASTIC UNIT COMMITMENT PROBLEM

Ali Koc

Soumyadip Ghosh

IBM T.J. Watson Research Center
P.O. Box 218, 1101 Kitchawan Rd.
Yorktown Heights NY 10598, USA

ABSTRACT

Scenario tree reductions of multi-period stochastic processes have been used as an important technique in obtaining good approximate solutions of multi-period convex stochastic programs. The scenario reduction step is aimed often at optimal approximation of the underlying stochastic process. We provide a new fast computationally cheap scenario tree reduction procedure and describe its approximation capabilities. Our context is the stochastic Unit Commitment Problem, the stochastic version of a problem that is at the heart of many modern energy markets. Its solution determines wholesale contracts between energy producers and energy consumers a day before actual transactions. We show that the new technique performs better than earlier prescriptions in obtaining approximations to the original program. However, these techniques of approximating only the underlying distributions without attention to the cost functions may produce weaker approximations of the optimal solution value; we provide a couple of illustrations to this point.

1 INTRODUCTION

Optimization under uncertainty has been a key focus of the operations research community in the last few decades (Sahinidis 2004). Many real-world applications require decisions to be made under uncertain information about the future. The canonical form of such a stochastic programming problem is given by

$$\min_{x \in X} F(x) \triangleq \min_{x \in X} \mathbb{E}_P f(\omega, x) \triangleq \min_{x \in X} \int_{\Omega} f(\omega, x) P(d\omega), \quad (1)$$

where $X \subset \mathbb{R}^n$ is a given nonempty convex closed set, Ω a closed subset of $\mathbb{R}^s$ and $\mathcal{B}$ the Borel σ -field relative to Ω , the function f from $\Omega \times \mathbb{R}^n$ to the extended reals $\overline{\mathbb{R}}$ is measurable with respect to ω and lower semi-continuous and convex with respect to x , and P is a fixed probability measure on $(\Omega, \mathcal{B})$. Amongst the various stochastic optimization models that fit into this formulation, the two-stage and the more general multi-stage stochastic programming problems with recourse are the most well-known; in these models, the X is the feasible set of first-stage decisions and $f(\omega, x)$ evaluates the optimized decisions under the realization ω of all the multi-stage uncertainties.

This article is concerned mainly with multi-stage stochastic optimization programs. Multi-stage programs of the form (1) are usually large-scale problems where P may not be known precisely, and where f is given implicitly as an iterated optimal value that depends on the decisions made in the earlier stages as each stage's uncertainty resolves itself. The most practical approaches to solving these problems assume some form of finiteness of the underlying distribution P .

Our interest lies in applications of this program in the scheduling of large-scale energy production at regional- and national-grid levels. In these applications (refer Section 4), the stochastic process of uncertainty inherent in the problem is modeled by a form that has continuous marginal distributions in each decision epoch. Thus, the central focus of our work is in approximating the solution to (1) when P

is modeled with continuous distributions. The approximation typically requires that the problem (1) under the original distribution P be approximated by a computationally tractable problem by replacing P with a distribution $\tilde{P}$ with finite, manageable number of probability atoms.

Finite sets of sample paths of multi-stage stochastic programs can be compactly represented by a temporal network structure called a *scenario tree*. These tree-networks are derived by combining sample paths that share sub-paths starting with the root node (first-stage), and the sample paths then are set as branches of the same common sub-path trunk. The associated problem (1) thus eliminates variables that were common between paths that shared a common ancestral sub-path, which not only makes the formulation more compact but also automatically incorporates non-anticipativity. The latter is a key requirement for multi-stage stochastic programs that all decisions made up to any stage under a sample path cannot depend on the remaining sample path stages, i.e. anticipate the future.

The approximating distribution $\tilde{P}$ is to be chosen such that the identified approximate optimal solution of (1) for $\tilde{P}$ forms a good approximation of the solution under P . Moreover the approximation should be stable, in that the optimality gap is continuous with respect to the underlying probability distribution P . How can this be ensured? The approaches studied in the literature use the following bound on the approximation error (let $\tilde{F}(x)$ be defined analogous to $F(x)$ in (1) for the approximation $\tilde{P}$):

$$\begin{aligned} F(\operatorname{argmin}_x \tilde{F}(x)) - F(\operatorname{argmin}_x F(x)) &\leq 2 \sup_x |F(x) - \tilde{F}(x)| \\ &\leq 2 \bar{L}_1 d(P, \tilde{P}), \end{aligned} \quad (2)$$

where $\bar{L}_1$ is a uniform Lipschitz bound on the cost functions $f(\cdot, x)$ on X , and $d(\cdot, \cdot)$ is a distributional metric. (Refer Pflug (2001) for a generalization of (2) to functions that are uniformly Lipschitz only under a transformation.) The scenario-tree reduction can be optimized by identifying an approximation distribution $\tilde{P}$ that minimizes the *Kantorovich* or *Wasserstein* metric $d(P, \tilde{P})$ on the set of distributions:

$$\begin{aligned} d(P, \tilde{P}) &= \inf \{ \mathbb{E}(|X - \tilde{X}|) : \text{where joint distribution of } (X, \tilde{X}) \\ &\quad \text{is arbitrary, but with marginals } X \sim P, \tilde{X} \sim \tilde{P} \}. \end{aligned} \quad (3)$$

Thus, model (1) remains stable to perturbations to its underlying probability distribution as long as the Kantorovich distance between the original distribution and the perturbed distribution is kept under control.

Various scenario-tree approximation approaches have been prescribed in the literature, amongst which the approach taken by Dupačová, Gröve-Kuska, and Römisch (2003) and Heitsch and Römisch (2003) (see also Gröve-Kuska and Römisch 2005, Chpt. 30) seems popular. They start by describing approximations for finite distributions with large support by an optimally chosen (in a certain precise sense) distribution over a small subset of the support. For this case, the infimization problem (3) can be written down as a transportation problem whose unit transportation costs are defined by a metric on the sets of support points. (These distributional metrics are thus also termed *transportation* metrics.) Moreover, the problem of finding the optimal approximation $\tilde{P}^*$ that minimizes the distance $d(P, \tilde{P})$ can be simplified to a set-covering problem when the support of $\tilde{P}$ is restricted to be a (small) fixed subset of the support of P itself. Dupačová, Gröve-Kuska, and Römisch (2003) use this insight to construct a recursive procedure to choose the best subset of a preset size for $\tilde{P}$ from the support of P . The recursion solves a large set-covering problem at each stage.

Dupačová, Gröve-Kuska, and Römisch (2003) suggest that for processes P with continuous marginals the approximation determination pre-processing step would first generate a large set of sample-paths and then the optimal scenario-tree reduction from that set be chosen as the representative finite-support approximation. Pflug (2001) suggest an alternative where the sample-generation step is skipped and the continuous stochastic process is directly approximated by a discrete process with appropriately chosen parameters.

Section 2 provides a detailed formulation of a generic T -stage stochastic program and describes the scenario-tree reduction approach to solving these problems. In Section 2.1, we provide an alternative

approach to this task of identifying a good approximation $\tilde{P}$ of P . Our primary motivation arises from a practical consideration. Observe that the computational complexity of the multi-stage program (1) is driven not by the size of the support of $\tilde{P}$ (i.e. the number of distinct sample paths in the support) but by the number of nodes in the scenario tree implied by $\tilde{P}$. Thus, the optimal approximation question is better posed by imposing a constraint on the number of unique support points the scenario-tree can have in each stage, as opposed to constraining the total number of unique scenarios in the approximation $\tilde{P}$. Our approach decouples the various stages and solves small individual-stage combinatorial problems to obtain each stage's approximations, and thus does not require the computationally intensive multi-dimensional set-cover based recursion of Dupačová, Gröve-Kuska, and Römisch (2003). Given a fixed computational budget that implies a certain overall number of nodes in the scenario tree, one could in a straightforward manner search for the best approximation possible by imposing a combination of individual stage node limits that satisfies this overall bound. This approach holds the potential of being a much faster and better pre-processing procedure for the main program (1).

Section 4 describes the stochastic unit commitment problem, a multi-stage formulation that is of high interest in Energy applications. The problem is solved to determine the schedule of generators for the production of power to meet demand over a typical period of a whole day. A deterministic version of this problem is central to the proper functioning of many current regional and national electricity-grids. The uncertainty in the stochastic formulation arises from the incorporation of intermittent renewable generators in the planning cycle. The numerical experiments described and results observed in Section 5 illustrate the advantages of our scenario tree approximation approach for the stochastic unit commitment problem.

Note that the application of the upper bound (2) in determining a good approximation of P by $\tilde{P}$ detaches the distribution approximation problem from the underlying cost functions $f(\omega, x)$ and allows the user to concentrate solely on the approximation procedure. In Section 3, we provide two examples that illustrate why depending on (2) to ignore the underlying cost functions will provide only sub-optimal approximations. Incorporating the cost functions directly into the distribution approximation scheme can thus gain better approximations; exactly how this is to be done remains an interesting, open and active area of interest.

To summarize, our contributions are:

- We provide a new approach to scenario-tree generation for cases where P has continuous marginals in each decision stage. This works by optimally (in a certain precise sense) reducing a large scenario-tree that is sampled from P . This scenario-tree pre-processing method for the main program (1) needs to solve only a set of small combinatorial problems for each stage, and thus can free up more of the computational budget for solving the main program as compared to existing pre-processing procedures. (Section 2.1).
- Numerical experiments with realistic instances of the stochastic unit commitment problem that arises in energy grid management illustrate the efficacy of this approach (Sections 4 and 5).
- Additionally, we provide examples where reducing the problem of approximating the optimal value of the stochastic multi-stage program to that of approximating the underlying distribution P may limit the power of the approximation (Section 3).

2 SCENARIO TREES AND MULTI-STAGE STOCHASTIC OPTIMIZATION

The approximation algorithm presented in this article reduces a large yet finite scenario-tree to a small manageable tree, and hence the rest of this section deals in distributions P with a finite albeit large support. Let $\Xi = \{p_i, \xi_i\}_{i=1}^S$ be a T -dimensional random vector with a discrete probability distribution P with finite support, where $\xi_i \in \mathbb{R}^T$, $P(\Xi = \xi_i) = p_i$, $p_i \geq 0$, $i = 1, \dots, S$, $\sum_{i=1}^S p_i = 1$. Let $\tilde{\Xi} = \{q_j, \tilde{\xi}_j\}_{j=1}^{\tilde{S}}$ be a random vector with the reduced distribution $\tilde{P}$, where $\tilde{P}(\tilde{\Xi} = \tilde{\xi}_j) = q_j$, $q_j \geq 0$, $j = 1, \dots, \tilde{S}$, $\sum_{j=1}^{\tilde{S}} q_j = 1$. Let Ξ'

$(\tilde{\Xi}^t)$ denote the t th marginal distribution of Ξ ($\tilde{\Xi}$), i.e.,

$$Pr(\Xi^t = \xi_{i'}) = \sum_{i=1}^S Pr(\Xi = \xi_i, \xi_i^t = \xi_{i'}) = \rho_{i'}, \quad i' \in I^t, \quad t = 1, \dots, T,$$

and similarly for $\tilde{\Xi}^t$. The set I^t ($\tilde{I}^t$) is the support of Ξ^t ($\tilde{\Xi}^t$) with cardinality m^t ($\tilde{m}^t$).

2.1 Lower Bound on the Minimum Kantorovich Distance

The Kantorovich distance between Ξ and $\tilde{\Xi}$ reduces to the optimal value of a linear transportation problem (Dupačová, Gröve-Kuska, and Römisch 2003) (we use the stochastic processes Ξ , $\tilde{\Xi}$ and their distributions P , $\tilde{P}$ interchangeably in the sequel):

$$d(P, \tilde{P}) = \inf \left\{ \sum_{i=1}^S \sum_{j=1}^{\tilde{S}} \eta_{ij} c(\xi_i, \tilde{\xi}_j) : \eta_{ij} \geq 0, \sum_{i=1}^S \eta_{ij} = q_j, \sum_{j=1}^{\tilde{S}} \eta_{ij} = p_i, \forall i, \forall j \right\}, \quad (4)$$

where $c(\xi_i, \tilde{\xi}_j) = \sum_{t=1}^T |\xi_i^t - \tilde{\xi}_j^t|$. (Dupačová, Gröve-Kuska, and Römisch 2003) show that the optimal solution to the minimal-distance problem

$$\min_{\tilde{P} \mid |\text{support}(\tilde{\Xi})| = \tilde{S}} d(P, \tilde{P})$$

over the set of all reduced scenario trees $\tilde{P}$ with support cardinality of $\tilde{S}$ is given by the distribution $\tilde{P}^*$ that assigns the probability weights $q_j^* = \sum_{i \in I(j)} p_i$, where

$$I(j) = \left\{ i = 1, \dots, S \mid j \in \underset{k=1, \dots, \tilde{S}}{\operatorname{argminn}} \sum_{t=1}^T |\xi_i^t - \tilde{\xi}_k^t| \right\}.$$

(The singleton subset *argminn* of the set *argmin* is obtained by, without loss of generality, picking the smallest element in lexicographic order.) We develop a lower bound on the minimum Kantorovich distance between Ξ^t and an arbitrary distribution $\tilde{\Xi}^t$ with a restriction on the number of distinct realizations at each decision epoch to be at most $\tilde{m}^t$, $t = 1, \dots, T$. Note that this is distinct from the restriction placed by the existing scenario-tree reduction methods on the number of total scenarios $\tilde{S}$ in the reduction $\tilde{\Xi}$.

Proposition 1 The Kantorovich distance $d(P, \tilde{P})$ obeys the lower bound

$$\min_{\tilde{P} \mid |\tilde{I}^t| = \tilde{m}^t} d(P, \tilde{P}) \geq \sum_{t=1}^T M^t(\tilde{m}^t),$$

where the $M^t(\tilde{m}^t)$ are the optimal solutions to the $\tilde{m}^t$ -weighed-median clustering of the (finite) support I^t of the t th marginal Ξ^t weighed by the p_i , i.e.

$$M^t(\tilde{m}^t) \triangleq \min_{\xi_1^t, \dots, \xi_{\tilde{m}^t}^t} \sum_{i=1}^S \min_{j=1, \dots, \tilde{m}^t} p_i |\xi_i^t - \xi_j^t|.$$

Proof. The dual of the transportation problem embedded in the definition of $d(P, \tilde{P})$ can be used to establish, via weak duality, the first of the bounds (refer Thm 2 in Dupačová, Gröve-Kuska, and Römisch (2003) for details):

$$\begin{aligned} d(P, \tilde{P}) &\geq \sum_{i=1}^S p_i \min_{j=1, \dots, \tilde{S}} \sum_{t=1}^T |\xi_i^t - \tilde{\xi}_j^t| \\ &\geq \sum_{i=1}^S p_i \sum_{t=1}^T \min_{j=1, \dots, \tilde{S}} |\xi_i^t - \tilde{\xi}_j^t| = \sum_{t=1}^T \sum_{i=1}^S \min_{j=1, \dots, \tilde{S}} p_i |\xi_i^t - \tilde{\xi}_j^t|. \end{aligned} \quad (5)$$

The second inequality is obtained by changing the order of summation and minimization, and the bound on the cardinality ($\tilde{S} \leq \prod_{t=1}^T \tilde{m}^t$). For the minimum Kantorovich distance distribution, we use arbitrary realizations $\tilde{\xi}_j$ for the reduced distribution $\tilde{\Xi}$, with the restriction that the cardinality of the support $\tilde{I}^t$ of $\tilde{\Xi}^t$ is $\tilde{m}^t$, $t = 1, \dots, T$:

$$\begin{aligned} \min_{\tilde{P} \mid |\tilde{I}^t| = \tilde{m}^t} d(P, \tilde{P}) &\geq \min_{\substack{\tilde{\xi}_1^1, \dots, \tilde{\xi}_{\tilde{S}}^1 \\ \dots \\ \tilde{\xi}_1^T, \dots, \tilde{\xi}_{\tilde{S}}^T}} \sum_{t=1}^T \sum_{i=1}^S \min_{j=1, \dots, \tilde{S}} p_i |\xi_i^t - \tilde{\xi}_j^t| \\ &\geq \sum_{t=1}^T \min_{\tilde{\xi}_1^t, \dots, \tilde{\xi}_{\tilde{m}^t}^t} \sum_{i=1}^S \min_{j=1, \dots, \tilde{m}^t} p_i |\xi_i^t - \tilde{\xi}_j^t| \\ &= \sum_{t=1}^T M^t(\tilde{m}^t). \end{aligned} \quad (6)$$

where the $\tilde{m}^t$ -medians M^t are as defined in the statement of the proposition. $\square$

Proposition 1 applies to minimizing approximation discrepancy between the original scenario set Ξ and all scenario-trees $\tilde{\Xi}$ that have marginal supports of each stage $\tilde{I}^t$ restricted to a specified set of values $\tilde{m}^t$. If we allow $\tilde{\Xi}$ to have as many realizations as Ξ , then the reduced distribution closest to Ξ trivially becomes Ξ itself. In this case, $\tilde{m}^t$ becomes equal to m^t for $t = 1, \dots, T$, and thus $M^t(\tilde{m}^t)$ becomes equal to zero. Thus, $\tilde{m}^t$ is a control parameter on the flexibility of choosing the reduced distribution with as many realizations as possible.

A scenario-tree $\tilde{\Xi}^*$ can be constructed to match the lower bound given by Proposition 1. Algorithm 1 provides a simple construction procedure. The summation in Step 7 involves many scenarios ($\xi_{i^1}, \dots, \xi_{i^T}$) that do not exist in the original scenario tree P ; the probability associated with them is zero. This also implies that some scenarios ($j^1, \dots, j^T$) constructed by the convolution in the “for all” step may be assigned a probability of zero. The support set $\tilde{\Xi}$ can be pruned to eliminate these if computationally convenient.

Proposition 2 The scenario-tree constructed by the Algorithm 1 matches the lower bound on approximation discrepancy given by Proposition 1.

Proof. We will show in particular that the lower bound form of (6) is achieved by the construction. The distance between the original Ξ and the $\tilde{\Xi}$ constructed by Algorithm 1. Given $i \in \{1, \dots, S\}$, and $t \in \{1, \dots, T\}$, we let $j^t(i) \in \tilde{I}^{*t}$ denote the smallest index such that $|\xi_i^t - \xi_{j^t(i)}^t| \leq |\xi_i^t - \xi_{j'}^t|$, $\forall j' \in \tilde{I}^{*t}$. Let $j(i) \in \{i, \dots, \tilde{S}\}$ such that $\tilde{\xi}_{j(i)} = (\tilde{\xi}_{j^1(i)}, \dots, \tilde{\xi}_{j^T(i)})$. By construction of the $\tilde{\Xi}$, such $j(i)$ exists. Then,

$$\min_{j=1, \dots, \tilde{S}} \sum_{t=1}^T p_i |\xi_i^t - \tilde{\xi}_j^t| = \sum_{t=1}^T p_i |\xi_i^t - \tilde{\xi}_{j(i)}^t| = \sum_{t=1}^T p_i |\xi_i^t - \tilde{\xi}_{j^t(i)}^t|. \text{ Thus,}$$

$$d(P, \tilde{P}) = \sum_{i=1}^S \min_{j=1, \dots, \tilde{S}} \sum_{t=1}^T p_i |\xi_i^t - \tilde{\xi}_j^t| = \sum_{i=1}^S \sum_{t=1}^T p_i |\xi_i^t - \tilde{\xi}_{j^t(i)}^t| = \sum_{t=1}^T M^t(\tilde{m}^t).$$

$\square$

The scenario-tree constructed by Algorithm 1 thus is the optimal solution to the minimization problem stated in Proposition 1. The constraint placed on the approximation scenario-trees is different from that placed by the methods studied by Dupačová, Gröve-Kuska, and Römisch (2003), that of limiting the maximum number of scenarios in the approximation. However, note that Algorithm 1 decouples the T stages of the problem into individual $\tilde{m}^t$ -median optimization problems. This procedure thus solves combinatorial optimization problems that are potentially smaller than those required by the recursive

Algorithm 1 Construct $\tilde{\Xi}$ from Ξ

Input: $\Xi = \{\xi_i\}_{i=1}^S$, $\xi_i \in \mathbb{R}^T$, with marginal distributions $\Xi^t = \{\xi_{i^t}\}_{i^t \in I^t}$, $t = 1, \dots, T$.
 $\tilde{m}^t \in \mathbb{N}$, $t = 1, \dots, T$.

Output: $\tilde{\Xi} = \{\tilde{\xi}_j\}_{j=1}^{\tilde{S}}$, $\tilde{\xi}_j \in \mathbb{R}^T$, with marginal distributions $\tilde{\Xi}^t = \{\tilde{\xi}_{j^t}\}_{j^t \in \tilde{I}^t}$, $|\tilde{I}^t| = \tilde{m}^t$, $t = 1, \dots, T$.

STEPS:

1. $\tilde{I}^{*t} \leftarrow \underset{(\xi_1, \dots, \xi_{\tilde{m}^t}) \in \mathbb{R}^{\tilde{m}^t}}{\operatorname{argminn}} \sum_{i=1}^S \min_{k=1, \dots, \tilde{m}^t} p_i |\xi_i^t - \tilde{\xi}_k^t|$, $t = 1, \dots, T$. (i.e., $\tilde{I}^{*t}$ is the set of $\tilde{m}^t$ -medians of Ξ^t .)
2. $I^t(j^t) \leftarrow \left\{ i^t \in I^t \mid j^t \in \underset{k^t \in \tilde{I}^{*t}}{\operatorname{argminn}} \sum_{t=1}^T |\xi_{i^t} - \tilde{\xi}_{k^t}| \right\}$, $j^t \in \tilde{I}^{*t}$, $t = 1, \dots, T$. (i.e., $I^t(j^t)$ is the set of points in I^t that are closest to point j^t of the t th stage of reduced tree $\tilde{\Xi}$.)
3. $k \leftarrow 0$,
- for all** $(j^1, \dots, j^T) \in \tilde{I}^{*1} \times \dots \times \tilde{I}^{*T}$ **do**
 4. $j \leftarrow j + 1$,
 5. $\tilde{\xi}_j = (\tilde{\xi}_{j^1}, \dots, \tilde{\xi}_{j^T})$,
 6. $\operatorname{support}(\tilde{\Xi}) \leftarrow \operatorname{support}(\tilde{\Xi}) \cup \{\tilde{\xi}_j\}$,
 7. $Pr(\tilde{\Xi} = \tilde{\xi}_j) = q_j = \sum_{i^1 \in I^1(j^1)} \dots \sum_{i^T \in I^T(j^T)} Pr(\Xi = (\xi_{i^1}, \dots, \xi_{i^T}))$,
- end for**
8. $\tilde{S} \leftarrow j$.

procedures of Dupačová, Gröve-Kuska, and Römisich (2003). This is true if the support of the marginals I^t has a significantly lower cardinality than the overall scenario tree's cardinality S . In this case, our procedure stands to gain significant performance improvements over the recursive methods where the outer recursions require one to consider set-cover or k -median clustering problems with the full cardinality S of the original scenario-set of P . For scenario trees where the support of the marginals grow with the stage index t , one can define the cardinality of $\tilde{m}^t$ to grow in a similar fashion, thus hopefully keeping the performance improvements intact. So, the procedure given here can be expected to quickly obtain a good scenario reduction without potentially requiring a significant fraction of the computation budget just for this pre-processing step to the original multi-stage program (1).

2.2 Further Improvements in Scenario-Tree Reduction

Algorithm 1 produces scenario reductions $\tilde{P}$ that match the lower bound in the inequality established by Proposition 1. Algorithm 1 allows us to generate a scenario tree such that the reduced tree has $\tilde{m}^t$ number of realizations at each time period. We now impose an additional constraint on top of the per-marginal support restrictions studied in the previous section. This constraint can take the form of an upper bound on either the total scenarios generated by the scheme, or the sum of the support points over all marginals. The former is the setup used in current methods (Dupačová, Gröve-Kuska, and Römisich 2003) where the size of the reduction desired is typically given by specifying only the total number of scenarios $\tilde{S} (\leq \prod_{t=1}^T \tilde{m}^t)$, not by the number of realizations at each time period. The latter constraint on the sum of all marginal support points arises naturally from the practical consideration of multi-stage problems that are being solved using methods such as Bender's decomposition or the L-shaped method, which scale exactly with this total of all marginal support points. Thus, under the first constraint, one matches the popular methods in the literature to produce approximations that minimize the Kantorovich distance $d(P, \tilde{P})$ amongst all distributions $\tilde{P}$ that have a maximum support size $\tilde{S}$, and under the second constraint one obtains the best approximation for a given computational budget to solve the resulting approximation problem (1).

The search for the best breakdown of $\tilde{S}$ into $\{\tilde{m}^t\}$ under either constraint can be accomplished by a secondary pre-processing step that is formulated as the following optimization problem. Let the binary

decision variable z_η^t take value 1 if the scenario tree has η scenarios at time period t , and 0 otherwise. Consider the following minimization problem:

$$\begin{aligned} \min_z \quad & \sum_{t=1}^T \sum_{\eta=1}^n M^t(\eta) z_\eta^t \\ \text{s.t.} \quad & \sum_{t=1}^T \sum_{\eta=1}^n \log(\eta) z_\eta^t \leq \log(\tilde{S}), \quad \text{or} \end{aligned} \quad (7a)$$

$$\sum_{t=1}^T \sum_{\eta=1}^n \eta z_\eta^t \leq N, \quad (7b)$$

$$\sum_{\eta=1}^n z_\eta^t = 1, \quad t = 1, \dots, T,$$

$$z_\eta^t \in \{0, 1\}, \quad t = 1, \dots, T, \quad \eta = 1, \dots, \tilde{S}.$$

This optimization problem gives the parameters $(\tilde{m}^t)$ of the closest scenario tree when at most N marginal support points are desired overall when form (7b) is used. When constraint form (7a) is active, the optimal set $\{\tilde{m}^t\}$ chosen has at most $\tilde{S}$ scenarios in the reduced tree $\tilde{P}$. To see this, note that this is the linearized form of the following constraint under a transformation by the increasing function $\log$:

$$\prod_{t=1}^T \left(\sum_{\eta=1}^n \eta z_\eta^t \right) \leq \tilde{S} \Leftrightarrow \sum_{t=1}^T \log \left(\sum_{\eta=1}^n \eta z_\eta^t \right) \leq \log(\tilde{S}) \Leftrightarrow \sum_{t=1}^T \sum_{\eta=1}^n \log(\eta) z_\eta^t \leq \log(\tilde{S}).$$

The binary nature of the z variables and the constraint $\sum_{\eta=1}^n z_\eta^t = 1, \quad t = 1, \dots, T$ is used.

The straightforward method to solving the integer program (7a) or (7b) would pre-compute the various $\tilde{m}^t$ –median problems for $\tilde{m}^t = 1, \dots, \tilde{S}, t = 1, \dots, T$ and then solve the program itself. Though this seems a task as daunting as the combinatorial optimization problems arising out of the existing scenario-tree reduction procedures, our observations from the experiments show that in practice these problems solve quickly. We note in closing that the function $M^t(\eta)$ represents the residual deviation in an η –median problem, and hence is non-increasing in η . This property can be exploited by either knapsack-like formulation, especially by (7b) which is exactly a knapsack problem, to construct speedy heuristics for quick approximations.

3 LIMITATIONS OF SCENARIO-TREE REDUCTION APPROACHES

The procedures described in this article approximate the optimal solution of original multi-stage stochastic program (1) by solving the same problem for an approximate distribution $\tilde{P}$ in place of the original continuous distribution P and appeals to the upper bound in (3) to control the quality of the approximation. The following two examples show that this can lead to approximation inaccuracies that can be easily improved upon.

Example 1 (Newsvendor) Consider the single-stage stochastic newsvendor problem as described in Pflug (2001). The newsvendor has to decide the number x of newspapers to order for a price of α that she can sell for a profit of β . If the demand realized is ω , then the scenario-dependent payoff can be written down as $f(\omega, x) = \alpha[x - \omega]^+ + \beta[\omega - x]^+$. The optimal solution to this problem can be computed explicitly as $x^* = P^{-1}(\beta/(\alpha + \beta))$. Pflug (2001) choose a symmetrical 3–point discrete approximation $\tilde{P}$ for P when P is a Gaussian distribution, and show that the Kantorovich distribution metric minimization procedure yields an optimal reduction $\tilde{P}^*$ that has a tiny but positive approximation error which can be computed exactly. However, a 3–point symmetric approximation for the Gaussian distribution P can be easily constructed such that it has *zero* error: if $\beta < \alpha$ then let the left-most point of the approximation coincide with the $(\beta/(\alpha + \beta))$ th quantile of the Gaussian distribution and place a mass of $(\beta/(\alpha + \beta))$ or higher on it. This approximation will then yield results that are *exactly optimal*.

In this first example, the underlying distribution is continuous and the objective functions is well-behaved; indeed the newsvendor problem is one of the foundational problems of stochastic optimization. Yet, obtaining scenario reductions by minimizing distributional distance measures alone without heed to the actual objective function leads to (slightly) poorer estimation of the actual optimal solution.

Example 2 (Facility Location) Consider a simple facility-location problem within a hypercube in $\mathbb{R}^d$. Suppose that the uncertain demand can be realized in any one of the 2^d corners. Let $D \triangleq 2^d$. Each realization of demand can be coded as a dyadic sequence $(s_1 s_2 \dots s_D)$ where $s_k = 1$ if the k th corner has one unit of demand and zero otherwise. Further, let the original distribution P consist of four realizations with equal probability: $\{\omega^1 = (010101 \dots 01), \omega^2 = (101010 \dots 10), \omega^3 = (0101 \dots 010000 \dots 00)$ and $\omega^4 = (1010 \dots 100000 \dots 00)\}$ where last two scenarios have zeros in the last $D/2 = 2^{d-1}$ entries. The scenario pairs have the following distances c_{ij} as calculated by the Kantorovich metric: $c_{12} = D$, $c_{13} = c_{24} = D/4$, $c_{14} = c_{23} = 3D/4$ and $c_{23} = D/2$. The optimal 2-scenario reduction $\tilde{P}^*$ under the Kantorovich metric of the 4-scenario P would then consist of $\{\omega^1, \omega^2\}$.

Consider the optimization objective function $f(\omega, x) = \sum_D s_i |x - C_i|^2$, i.e. the optimal median location problem, where the C_i are the corners (D in number). Then, under both scenarios chosen in the 2-scenario reduction $\tilde{P}^*$ the optimal solution will be the center of the hypercube. However, under each of the two scenarios omitted, the optimal facility placement will be in a face of the hypercube. Thus the “optimal” reduction obtained by minimizing the distribution distance metric will be significantly poorer than a $\tilde{P}$ choice of say $\{s^1, s^3\}$.

The objective function is well-behaved in this case too. This second example illustrates the additional complexity introduced by the combinatorial nature of the multi-stage programs. We do not give a constructive method to consider objective function structure in creating scenario reductions and obtaining the alternate approximations; indeed, this is an area of active interest and investigation.

4 THE STOCHASTIC UNIT COMMITMENT PROBLEM

Unit commitment is the problem of finding an optimal schedule and corresponding generation amounts for a set of generators over a twenty-four-hour horizon so that total cost of generation and transmission is minimized, and a set of constraints, such as demand requirement, upper and lower limits of generation, minimum up/down time limits, ramp up/down constraints, transmission constraints are satisfied. Unit commitment (UC) adequately covers all the major aspects at the core of planning and operational decisions faced by independent system operators (ISOs), regional transmission organizations (RTOs), and utility companies. Hence it has received a good deal of attention in the industry.

The academic literature on the unit commitment problem dates back to the 1960s. An integer programming approach proposed by Dillon, Edwin, Kochs, and Taud (1978) was one of the earliest optimization-based approaches to the unit commitment problem. The authors address the unit commitment problem of hydro-thermal systems with reserve requirements. It was one of the earliest papers that can solve real life problems with 20 units. The authors developed two sets of valid inequalities that were globally valid to the problem and used these inequalities in the branch-and-bound algorithm. The dynamic programming approach developed by Jr., H. D. Powell, and Rayburn (1987) was one of the earliest successful dynamic programming algorithms. The algorithm featured a classification of units so as to reduce the number of states. The authors addressed the problem at San Diego Gas & Electric System with 30 generators. Expert systems, fuzzy logic, meta-heuristic algorithms, and ant colony systems are among the other approaches that have been applied to the unit commitment problem. Surveys by Sheble and Fahd (1994) and by Padhy (2004) review the academic literature on the unit commitment problem, and the book by Wood and Wollenburg (1996) addresses several operational and planning problems in the energy industry, including the unit commitment problem.

First a synopsis of the notation and the decision variables involved:

Indices and sets:

$i \in I$ generators
 $t \in \{1, \dots, T\}$ time periods

Data:

$S_i(\cdot)$ startup cost function of unit i
 $H_i(\cdot)$ shutdown cost function of unit i
 $f_i(\cdot)$ generation cost function of unit i
 (Q_i, q_i) maximum and minimum amounts for units i 's offer
 (R_i, r_i) ramp-up and ramp-down amounts for unit i
 (L_i, l_i) minimum up and down times for unit i
 d_t load at time period t

Decision variables:

g_{it} generation provided by unit i at time period t
 s_{it} binary variable indicating if unit i started at time period t
 u_{it} binary variable indicating if unit i is on at time period t
 $\bar{u}_{it}$ vector of u_{it} variables. $\bar{u}_{it} \equiv (u_{i\tau})$, $\tau = t - l_i + 1, \dots, t$. For $\tau < 0$, $u_{i\tau}$ is set to the on/off state of unit i at time period τ .
 $\underline{u}_{it}$ vector of u_{it} variables. $\underline{u}_{it} \equiv (u_{i\tau})$, $\tau = t - L_i + 1, \dots, t$. For $\tau < 0$, $u_{i\tau}$ is set to the on/off state of unit i at time period τ .

We use the multi-stage stochastic programming approach to address the uncertainty and intermittency of the wind power generators. We assume that the uncertainty evolves as a discrete time continuous stochastic process, which is represented within the unit-commitment formulation with a finite probability distribution approximation. The original stochastic process model may often be derived from statistical models of wind predictions, which typically involve various stat-based prediction models such as time-series models etc.

The finite space stochastic approximation is represented as a *scenario tree*. The information structure is presented as a rooted scenario tree where the nodes n ($n \in N$) in level t ($t \in T$) of the scenario tree constitute the states of the world that can be distinguished by the information available up to time period t (Singh, Philpott, and Wood 2009). The set of *leaf nodes*, N_L ($N_L \subset N$), contains the nodes without any successor. The root node is the node without any predecessor. In general, $n(\tau) \in N$ represents the τ^{th} predecessor of node n . The level of the root node is zero, and in general the level of a node, t_n , is defined such that $n(t_n)$ is the root node. The tree has a depth of T , and all leaf nodes have a level of T . By convention, $n(0)$ is the node n itself, and $n(\tau)$, $\tau > t_n$, is an empty set.

The root node occurs with probability one, and the state of the world corresponding to node $n \in N$ occurs with probability π_n , given that its predecessor, $n(1)$, has occurred; and p_n denotes the associated unconditional probability, i.e., $p_n = \prod_{\tau=0}^{t_n-1} \pi_{n(\tau)}$. There is a one-to-one matching between the leaf nodes of the scenario tree and the scenarios. Given a leaf node, $n \in N_L$, a T -tuple $[n(T), n(T-1), \dots, n(1), n]$ represents a scenario with probability of occurrence equal to p_n . Two scenarios sharing the same state of the world at time periods $1, \dots, \tau$, for some $\tau < T$, have to observe the same set of decision variables in the optimization model, in order to make sure that the model does not cheat by foreseeing (anticipating) the future.

The full stochastic unit commitment formulation is given below in (8):

Formulation:

$$\min_{\mathbf{g}, \mathbf{u}, \mathbf{s}} \sum_{n \in N} p_n \sum_{i \in I} S_i(\bar{u}_{in}) + H_i(\underline{u}_{in}) + f_i(g_{in}) \quad (8a)$$

$$\text{s.t. } \sum_{i \in I} g_{in} \geq d_n, \quad n \in N, \quad (8b)$$

$$q_i u_{in} \leq g_{in} \leq Q_i u_{in}, \quad i \in I, n \in N \quad (8c)$$

$$s_{in} \geq u_{in} - u_{i,n(1)}, \quad i \in I, n \in N \quad (8d)$$

$$g_{in} - g_{i,n(1)} \leq R_i, \quad i \in I, n \in N \quad (8e)$$

$$g_{i,n(1)} - g_{in} \leq r_i, \quad i \in I, n \in N \quad (8f)$$

$$\sum_{\tau=0}^{L_i-1} s_{i,n(\tau)} \leq u_{in}, \quad i \in I, n \in N \quad (8g)$$

$$\sum_{\tau=1}^{l_i-1} s_{i,n(\tau)} \leq 1 - u_{i,n(l_i)}, \quad i \in I, n : n(l_i) \in N \quad (8h)$$

$$\sum_{\tau=1}^{l_i-1-p} s_{i,n(\tau)} \leq 1 - u_{i,n(l_i-p)}, \quad i \in I, n \in N_L, p \in \{1, \dots, l_i - 1\} \quad (8i)$$

$$u_{in} \in \{0, 1\}, s_{in} \in \{0, 1\}, g_{in} \geq 0 \quad (8j)$$

Objective function (8a) is the total cost of generation summed over all nodes and is equivalent to the expected cost of generation over all time periods. Cost of generation includes startup cost, shutdown cost, and fuel cost. Startup (shutdown) cost can be a step function that takes its highest step value when the unit has been down (up) for a certain number of time periods, or it can be a piecewise function that increases with the number of time periods the unit has been down (up). Fuel cost may be a nonlinear function of generation level, which may not be convex. Constraint (8b) requires total generation be greater than the load. Constraint (8c) has two functions: If the unit is down, it forces the generation to be zero; and, if the unit is up, it arranges the generation level to be between the upper and lower limits of the generator. Constraint (8d) links the start-up variables to up/down variables. Constraints (8e) and (8f) handle ramp-up and ramp-down limits. Constraint (8g) is for minimum up time and constraints (8h) and (8i) are for minimum down time requirement. Finally constraint (8j) puts binary and non-negativity requirements.

5 EXPERIMENTS

We consider a simple wind model with four time periods. Wind speed at each time period follows a Gaussian random variable with a mean of 10 miles per hour and a standard deviation of 2.5 miles per hour. The distribution of wind at each stage is assumed to be independent of previous stages; this is a significant departure from reality, but will be sufficient to illustrate our point. We compare three approaches to generate a scenario tree, the results of which are presented in Figure 5. The first approach samples three values from the Gaussian random variable for each stage and constructs a scenario tree with all the combinations of these, i.e. with 81 scenarios (data marked “3-sample” in Figure 5). The second approach does similarly by sampling five scenarios per time period with a total of 625 scenarios (data marked “5-sample” in Figure 5). And the third approach uses the scenario tree reduction technique developed in this article. It first generates hundred samples from the Gaussian random variable at each stage, and then reduces it to a scenario-tree that has three points per stage. It generates a scenario tree for four time periods using these three points with a total of 81 scenarios (data marked “3-s-r” in Figure 5).

For each method of scenario tree generation, we solve the stochastic unit commitment problem using the corresponding tree. We repeatedly solve fifty of these problems for each scenario-generation method. The figure provides a box-plot of the observed optimal objective function approximation values for the

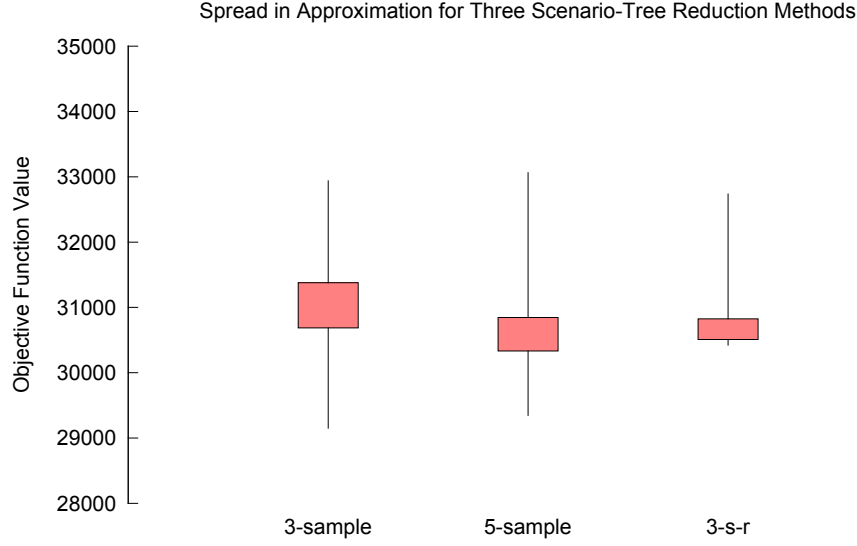


Figure 1: Comparison of the optimal value approximations for three scenario tree generation methods from 50 runs. The standard deviation of the three samples sets are: “3-sample” at 691.6032, “5-sample” at 513.4892, and “3-s-r” at 316.5449. The boxes are of one standard-deviation’s length, centered around the mean of the approximation values.

three methods. The boxes are of length equal to the observed standard deviation (these values are given in the caption), centered around the observed mean approximation objective value from the 50 runs. The lines extend to the min and max observed approximation value. The key take-away from the plots in Figure 5 is the variability of the approximation for the optimal solution under the various scenario-tree selection procedures. The “5-sample” method has better variance than the “3-sample” solutions, as can be expected because of the additional scenarios used. The plot, along with the variance values also show significantly better convergence of the approach outlined in this paper. The “3-s-r” method has a standard deviation that is almost half the standard deviation as the plain sampling of “3-sample” and also outstrips the performance of the “5-sample” method by almost as much. Recall that the “3-s-r” method is actually reducing a scenario tree with $100^3 = 1000000$ scenarios to $3^3 = 81$ scenarios, whereas the first approach uses only 81 of the first scenarios and the second approach uses 625 scenarios.

We note that generating samples and performing scenario reduction takes negligible amount of time (in the order of seconds). The main time-consuming step is solving the stochastic unit commitment problem. Thus, the results are biased towards “5-sample” since we spent much more time to solve the problem with five samples at each epoch than we do for the problems with three samples as in the cases of “3-sample” and “3-s-r.” However, the main conclusion here is to show the effectiveness of “3-s-r”; thus, we do not worry about this bias since the experiments of “3-s-r” takes the least amount of time and performs the best. These results clearly illustrate the importance of a scenario-reduction step in determining a probability approximation with a small support set that works well and is more robust to perturbations.

REFERENCES

- Dillon, T., K. Edwin, H. Kochs, and R. Taud. 1978. “Integer programming approach to the problem of optimal unit commitment with probabilistic reserve determination”. *IEEE Transactions Power Systems* 97:2154–2166.
- Dupačová, D., N. Gröve-Kuska, and W. Römisch. 2003. “Scenario reduction in stochastic programming: An approach using probability metrics”. *Mathematical Programming* 95:493–511.

- Gröve-Kuska, N., and W. Römisch. 2005. "Stochastic unit commitment". *Applications of Stochastic Programming*:633–653.
- Heitsch, H., and W. Römisch. 2003. "Scenario reduction algorithms in stochastic programming". *Computational Optimization Applications* 24:187–206.
- Jr., W. L. S., J. H. D. Powell, and J. C. Rayburn. 1987. "Dynamic programming approach to unit commitment". *IEEE Transactions Power Systems* 2:339–347.
- Padhy, N. 2004. "Unit commitment - A bibliographical survey". *IEEE Transactions Power Systems* 19:1196–2005.
- Pflug, G. 2001. "Scenario tree generation for multiperiod financial optimization by optimal discretization". *Mathematical Programming* 89:251–271. 10.1007/PL00011398.
- Sahinidis, N. V. 2004. "Optimization under uncertainty". *Computers & Chemical Engineering*:071–983.
- Sheble, G. B., and G. N. Fahd. 1994. "Unit commitment - Literature synopsis". *IEEE Transactions Power Systems* 9:128–135.
- Singh, K. J., A. B. Philpott, and R. K. Wood. 2009. "Dantzig-wolfe decomposition for solving multistage stochastic capacity-planning problems". *Operations Research* 57 (5): 1271–1286.
- Wood, A. J., and B. F. Wollenburg. 1996. *Power Generation Operation and Control*. New York: John Wiley and Sons.

AUTHOR BIOGRAPHIES

ALI KOC is a postdoctoral researcher in the Business Analytics and Mathematical Sciences Department at IBM Watson Research Center. He received his PhD degree from the University of Texas at Austin in 2010, and M.S. and B.S degrees from Bilkent University in 2005 and 2002. He worked on several projects including the ones at the San Diego Supercomputing Center in 2007, at Freescale Semiconductor Inc in 2008, and at the South Texas Project Nuclear Operating Company in 2009. He joined the Business Analytics and Mathematical Sciences Department in March 2010. He is currently performing research on optimization of electric power generation and distribution under uncertainty. He received 2011 INFORMS Young Researcher Council Award, 2009 INFORMS Future Academician Colloquium Award, and 2009 University of Texas at Austin Graduate School Continuing Fellowship Award, among others. His research interests are in the areas of stochastic modeling and optimization, algorithms for high-performance computing, power systems optimization, and production planning and control. He is a member of INFORMS.

SOUMYADIP GHOSH is a Research Staff Member of the Mathematical Sciences Division at IBM T.J. Watson Research Center, Yorktown Heights, NY. His simulation research interests include theory and practice in particular on simulation-based optimization techniques and input dependence modeling and risk modeling. His other interests lie in the fields of smarter energy management, supply-chain analysis, and queueing theory based scheduling of large-scale production systems. He can be contacted at ghoshs@us.ibm.com and his web-page is <https://researcher.ibm.com/researcher/view.php?person=us-ghoshs>.