Partially Adaptive Stochastic Optimization for Electric Power Generation Expansion Planning

Jikai Zou* Shabbir Ahmed* Andy Sun*

July 2016

Abstract

Electric Power Generation Expansion Planning (GEP) is the problem of determining an optimal construction and generation plan of both new and existing electric power plants to meet future electricity demand. We consider a stochastic optimization approach for this capacity expansion problem under demand and fuel price uncertainty. In a two-stage stochastic optimization model for GEP, the capacity expansion plan for the entire planning horizon is decided prior to the uncertainty realized and hence allows no adaptivity to uncertainty evolution over time. In comparison, a multi-stage stochastic optimization model allows full adaptivity to the uncertainty evolution, but is extremely difficult to solve. To reconcile the trade-off between adaptivity and tractability, we propose a partially adaptive stochastic mixed integer optimization model in which the capacity expansion plan is fully adaptive to the uncertainty evolution up to a certain period and follows the two-stage approach thereafter. Any solution to the partially adaptive model is feasible to the multi-stage model, and we provide analytical bounds on the quality of such a solution. We propose an efficient algorithm that solves a sequence of partially adaptive models, to recursively construct an approximate solution to the multi-stage problem. We identify sufficient conditions under which this algorithm recovers an optimal solution to the multi-stage problem. Finally, we conduct extensive test of our algorithm on a realistic GEP problem. Experiments show that, within a reasonable computation time limit, the proposed algorithm produces a significantly better solution than solving the multi-stage model directly.

Keywords: generation expansion planning, multi-stage stochastic optimization, approximation algorithm

1 Introduction

Generation expansion planning (GEP) is the problem of determining an optimal construction and generation plan over a finite planning horizon of both existing and new generation power plants to meet future electricity demand, while satisfying operational, economic, and regulatory constraints. The objective of GEP is to minimize the total investment cost and generation cost. Investment cost depends on the number of newly built generators over the planning horizon, and generation cost reflects the cost incurred at the operation level. GEP is considered as a major problem in power system planning. It is challenging due to its large scale, long-term horizon, and nonlinear and discrete nature. A major challenge in GEP, as well as in more general capacity expansion problems, is to deal with uncertainty in future demand, as well as various other uncertainties such as technological breakthroughs, cost structures, etc. The first stochastic capacity expansion

*H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA. E-mail: jikai.zou@gatech.edu, sahmed@isye.gatech.edu, andy.sun@isye.gatech.edu

planning model with demand uncertainty dates back to Manne (1961), and has been followed by extensive work in the area (see e.g., Erlenkotter 1967; Giglio 1970; Freidenfelds 1980; Davis et al. 1987; Bean et al. 1992). These early works assume simplified underlying stochastic processes for the uncertainties to obtain analytical solutions and are typically restricted to capacity expansion with a single resource.

Stochastic optimization approaches for capacity expansion problems utilize scenario trees to model uncertainty. Both two-stage and multi-stage stochastic optimization models have been proposed. In a typical two-stage stochastic GEP model, the first-stage decisions are the capacity expansion decisions over the *entire* planning horizon, which are made before any uncertainty is realized; then the operational level decisions for generation production are made as the second stage decision, which are fully adaptive to uncertainty realizations. For example, Jin et al. (2011) propose such a two-stage stochastic GEP model and further consider both risk-neutral and risk-averse objectives, and apply a random sampling based method to approximately solve the problem. Bloom (1983); Bloom et al. (1984) investigate in a two-stage GEP model with probabilistic reliability constraints and solve it using generalized Benders' decomposition. The second stage subproblems are nonlinear because of the system reliability constraints and are solved using a procedure called probabilistic simulation. Bienstock and Shapiro (1988) propose another type of two-stage stochastic GEP model which decomposes the entire planning horizon into two stages. The authors consider demand and fuel price uncertainty, and solve the problem using a Benders' decomposition type algorithm.

Multi-stage stochastic optimization models allow the capacity expansion decisions to be fully adaptive to uncertainty realizations. For instance, Berman et al. (1994) consider a scenario-based multi-stage stochastic optimization model for capacity expansion of a single technology. Chen et al. (2002) extend this model to multiple technologies. In both models, the capacity expansion decisions are assumed to be continuous variables. Ahmed and Sahinidis (2003) consider a multi-stage stochastic formulation, where the capacity expansion decisions are binary variables. The authors propose an LP-relaxation-based approximation algorithm for this problem and prove its asymptotic optimality as the planning horizon goes to infinity. Ahmed et al. (2003) further exploit the special structure of the stochastic lot-sizing problem and develop a branch-and-bound algorithm to obtain global optimal solutions. Singh et al. (2009) propose a column-generation approach for solving such problems. Huang and Ahmed (2009) show that multi-stage capacity expansion models can have significant advantages in terms of total expected costs over two-stage models. Ryan et al. (2011) and Wallace and Fleten (2003) present comprehensive surveys on stochastic modeling in planning and operation of electric power systems. In a broader domain of multi-stage stochastic optimization, extensive research has been done in term of how to allocate "stages" (Grinold 1986; Nielsen and Zenios 1996; Dempster et al. 2000; Dupačová et al. 2009), but most of them assume continuous variables and focus on financial applications. It is fair to say that, despite the intense research efforts, multi-stage stochastic expansion models remain extremely challenging to solve for large-scale cases within a reasonable amount of computation time.

In this paper, we develop a unifying framework and an efficient algorithm for solving the stochastic capacity expansion problem. The key contributions of the paper are summarized here.

1. We propose a so-called *partially adaptive* (PA) stochastic model for the general capacity expansion planning problem and apply it to the generation expansion problem. The partially adaptive stochastic model allows the capacity expansion decision to be fully adaptive to uncertainty up to a certain set of prescribed nodes in the scenario tree, and then restricts the expansion decisions to a two-stage structure thereafter. In this way, the PA model provides a natural generalization for the two-stage and multi-stage models. A related idea is used to approximate an infinite horizon stochastic program by a

finite-horizon model through aggregating all future decisions after a finite time period (Grinold 1986). The two approaches have very different motivations. More importantly, the PA model proposed in our paper allows more flexibility in constructing the 'compressed' scenario tree. Also, to the best of our knowledge, partially adaptive models have not been considered for the stochastic capacity expansion problem.

2. We provide theoretical analysis of the performance of the PA model. In particular, we prove upper and lower bounds for the performance difference between a PA model and full multi-stage model. The bounds are related to the cost parameters as well as the structure of the compressed scenario tree. Such a performance analysis for the PA model appears to be new.

3. We further propose a new efficient approximation algorithm (Algorithm 1) to solve the PA model. The algorithm recursively solves a sequence of smaller PA models and constructs a feasible solution to the multi-stage stochastic optimization model. We identify analytical conditions on problem parameters under which the proposed algorithm recovers an optimal multi-stage solution. We also conduct extensive computational experiments on a large scale 10-year GEP problem with real-world demand and fuel price data. The results demonstrate that the proposed algorithm can solve the GEP problem within a reasonable computation time limit and yields a significantly better solution than directly solving the multi-stage model.

The paper is organized as follows. Section 2 presents the PA model. In Section 3, we provide theoretical analysis of the performance of the PA model. In Section 4, we propose the new approximation algorithm. Section 5 presents computational results. Finally, we provide some concluding remarks in Section 6.

2 Model development

In this section, we develop a partially adaptive stochastic mixed integer optimization model for the GEP problem. We first introduce the scenario tree representation of uncertainty in Section 2.1. Then, in Section 2.2, we propose a detailed PA model for the GEP problem. In Sections 2.3, we present a PA model for the general capacity expansion problem, and discuss its connections to two-stage and multi-stage models.

2.1 Scenario tree representation of uncertainty

We consider two main sources of uncertainty in the GEP model, namely, the uncertainty of future demand and fuel prices, mainly natural gas prices, which are historically very volatile (Jin et al. (2011)). A scenario tree approach is adopted to model the evolution of these uncertainties over a multi-year finite planning horizon. In the following, we introduce some standard notations for the scenario tree model (see e.g. Ruszczynski and Shapiro (2003)).

Let $\mathcal{T}$ be a scenario tree with T levels. Each level corresponds to a year in the planning horizon of the GEP model. A node $n \in \mathcal{T}$ represents a specific joint realization of uncertain parameters $\xi_n = (d_n, g_n)$, where d_n is the demand and g_n is the fuel price. The nodes are indexed in the order of positive integers, with the root node indexed as node 1. Let t_n denote the level of node n, and let $\mathcal{S}_t$ be the set of all nodes in level t for each $t = 1, \ldots, T$. Each node $n \in \mathcal{T} \setminus \{1\}$ has a unique parent $a(n)$. The subtree with root n and ending at period τ ($t_n \leq \tau \leq T$) is denoted as $\mathcal{T}(n, \tau)$. Denote $\mathcal{T}(n, T)$ as $\mathcal{T}(n)$ for simplicity. The set of children nodes of n is denoted as $\mathcal{C}(n)$. There is a (unconditional) probability p_n associated with each node n,

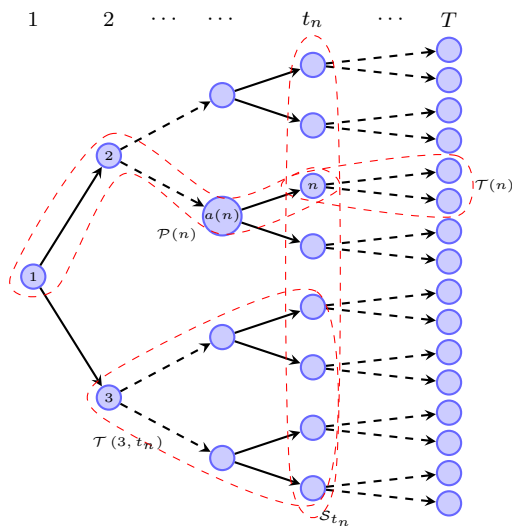

Figure 1: An example of scenario tree.

corresponding to the realization of uncertain parameters at this node. The sum of probabilities of all nodes in each level equals 1; that is, $\sum_{n \in \mathcal{S}_t} p_n = 1$ for all t. The sum of probabilities of the nodes with the same parent equals the probability of their parent node; that is, $\sum_{m \in \mathcal{C}(n)} p_m = p_n$ for all $n \in \mathcal{T}$. The unique path from the root to a node n is denoted by $\mathcal{P}(n)$. A path from the root node to a leaf node is called a *scenario*. We say two scenarios s and s' are *indistinguishable* at period t if the paths corresponding to s and s' pass through the same set of nodes up to period t. Let $\Omega = \{\mathcal{P}(n) : n \in \mathcal{S}_T\}$ be the set of all possible scenarios. Figure 1 shows an illustration of a general scenario tree.

2.2 Partially adaptive model for GEP problem

Now we can introduce the PA model for the GEP problem. The basic structure of the model is adapted from the two-stage model of Jin et al. (2011). As discussed in the introduction, the PA-GEP model imposes a multi-stage structure for the capacity expansion decision from period 1 up to period μ, i.e. the expansion decisions in these periods need to satisfy *non-anticipativity* constraints, and then from period $\mu + 1$ to T, the expansion decision follows a two-stage structure, i.e. they are made without the knowledge of uncertainty realization after period μ. Note that the operational level decisions of generation production have a multi-stage structure throughout the planning horizon. In the following, we present a detailed PA-GEP model with *critical time* μ.

Parameters

T	Number of time periods in planning horizon.
$\mathcal{I}$	Set of available technologies to expand capacity.
$\mathcal{K}_t$	Set of sub-periods in each planning period t.
c_{it}	Unit construction cost of a type i generator at period t. (\$/MW)
m_i^{max}	Maximum capacity of a type i generator. (MW)
n_i^{max}	Maximum rating of output of a type i generator. (MW)
u_i^{max}	Maximum number of type i generators that can be built over the planning horizon.

u_i^0	Number of pre-existing type i generators.
b_{ink}	Unit generation cost for a type i generator in sub-period k at node n. (\$/MWh)
d_{nk}	Hourly demand in sub-period k at node n. (MW)
h_{nk}	Number of hours in sub-period k at node n.
q	Unit penalty cost for unmet demand. (\$/MW)
r	Annual interest rate.
μ	Critical time of the model.
Variables	
x_{in}	Number of type i generators built at node n.
v_{ink}	Hourly generation of a type i generator in sub-period k at node n. (MW)
w_{nk}	Unmet hourly demand in sub-period k at node n. (MW)

Formulation

$$\min \sum_{n\in\mathcal{T}} \left(\sum_{i\in\mathcal{I}} \frac{p_n c_{it_n} m_i^{max}}{(1+r)^{t_n-1}} x_{in} + \sum_{i\in\mathcal{I}} \sum_{k\in\mathcal{K}_{t_n}} \frac{p_n h_{nk} b_{ink}}{(1+r)^{t_n-1}} v_{ink} + \sum_{k\in\mathcal{K}_{t_n}} \frac{p_n h_{nk} q}{(1+r)^{t_n-1}} w_{nk} \right) \tag{2.1a}$$

$$\text{s.t.} \sum_{m\in\mathcal{P}(n)} x_{im} + u_i^0 \geq \frac{1}{n_i^{max}} v_{ink}, \quad \forall\, i\in\mathcal{I}, k\in\mathcal{K}_{t_n}, n\in\mathcal{T} \tag{2.1b}$$

$$\sum_{m\in\mathcal{P}(n)} x_{im} \leq u_i^{max}, \quad \forall\, i\in\mathcal{I}, n\in\mathcal{S}_T \tag{2.1c}$$

$$\sum_{i\in\mathcal{I}} v_{ink} + w_{nk} = d_{nk}, \quad \forall\, k\in\mathcal{K}_{t_n}, n\in\mathcal{T} \tag{2.1d}$$

$$x_{in_1} = x_{in_2}, \quad \forall\, i\in\mathcal{I}, n_1, n_2\in\mathcal{T}(m)\cap\mathcal{S}_t, m\in\mathcal{S}_\mu, t>\mu \tag{2.1e}$$

$$x_{in}\in\mathbb{Z}_+, v_{ink}, w_{nk}\in\mathbb{R}_+, \quad \forall\, i\in\mathcal{I}, k\in\mathcal{K}_{t_n}, n\in\mathcal{T}. \tag{2.1f}$$

In the PA-GEP model (2.1), the objective function consists of investment, generation, and penalty costs, all of which are discounted to the beginning of the planning horizon. Constraints (2.1b) indicate that the output by each type of generators during any sub-period cannot exceed the aggregate output rating of all available (both pre-existing and newly built) generators of that type. Constraints (2.1c) require the total number of each type of generators built in all possible scenarios to be within the quantity limitation. Furthermore, in constraints (2.1d), we enforce the equality between the demand and the sum of generation output and unmet demand. Note that renewable generation such as wind and solar power can be easily incorporated in the above expansion model, e.g. by modeling wind and solar availability as random negative demand on the scenario tree.

The key constraint of the PA-GEP model is (2.1e), which imposes a two-stage model after period μ at each possible outcome in that period. As a result, for any possible realization at period μ, there is only one capacity expansion decision in each period subsequently. In other words, future capacity expansion decisions are made at period μ without knowing further uncertainty realizations. Note that non-anticipativity constraints, which impose the requirement that period t decision for two scenarios that are indistinguishable at stage t must be identical, are implicitly embedded in the nodal formulation due to the scenario tree structure.

2.3 Partially adaptive model for general capacity expansion planning problem

The model (2.1) can be written in a more abstract and general way to obtain a PA model for a general stochastic capacity expansion planning problem. All the analysis of the performance of the PA model in

Section 3 is done for the following generic model.

$$
\begin{aligned}
[\Pi_{\text{PA}}(\mu)] \quad \min \quad & \sum_{n\in\mathcal{T}} p_n \left(\mathbf{a}_n^\top \mathbf{x}_n + \sum_{k\in\mathcal{K}_{t_n}} \mathbf{b}_{nk}^\top \mathbf{y}_{nk} \right) && \text{(2.2a)}\\
\text{s.t.} \quad & \sum_{m\in\mathcal{P}(n)} \mathbf{x}_m \geq \mathbf{A}_{nk}\mathbf{y}_{nk}, \quad \forall\, k\in\mathcal{K}_{t_n}, n\in\mathcal{T} && \text{(2.2b)}\\
& \sum_{m\in\mathcal{P}(n)} \mathbf{x}_m \leq \mathbf{u}, \quad \forall\, n\in\mathcal{S}_T && \text{(2.2c)}\\
& \mathbf{B}_{nk}\mathbf{y}_{nk} = \mathbf{d}_{nk}, \quad \forall\, k\in\mathcal{K}_{t_n}, n\in\mathcal{T} && \text{(2.2d)}\\
& \mathbf{x}_{n_1} = \mathbf{x}_{n_2}, \quad \forall\, n_1, n_2 \in \mathcal{T}(m)\cap\mathcal{S}_t, m\in\mathcal{S}_\mu, t>\mu && \text{(2.2e)}\\
& \mathbf{x}_n \in \mathbb{Z}_+^I,\ \mathbf{y}_{nk}\in\mathbb{R}_+^J, \quad \forall\, k\in\mathcal{K}_{t_n}, n\in\mathcal{T}. && \text{(2.2f)}
\end{aligned}
$$

Here, the vector $\mathbf{x}_n$ corresponds to the investment decisions (x_{in}) in all types of generators at node n, thus $I = |\mathcal{I}|$; $\mathbf{y}_{nk}$ corresponds to the operation level decisions in sub-period k at node n, i.e., the generation output (v_{ink}) and the amount of unmet demand (w_{nk}), thus $J = I + 1$. The parameters $\mathbf{a}_n$ and $\mathbf{b}_{nk}$ correspond to the objective coefficients of $\mathbf{x}_n$ and $\mathbf{y}_{nk}$, respectively. The matrices $\mathbf{A}_{nk}$ and $\mathbf{B}_{nk}$ correspond to the coefficient matrix on the right-hand-side in constraints (2.1b) and coefficient matrix on the left-hand-side in constraints (2.1d), respectively. Constraints (2.2e) correspond to (2.1e). Finally, the parameter vectors $\mathbf{u}$ and $\mathbf{d}_{nk}$ correspond to the right-hand-side vectors in constraints (2.1c) and (2.1d), respectively. Note that $\mathbf{d}_{nk}$ has dimension of 1 in model (2.1), since electricity is the only type of output.

As a unifying framework, the proposed PA model $\Pi_{\text{PA}}(\mu)$ generalizes the existing two-stage and multi-stage models in the capacity planning literature (see e.g., Ahmed and Sahinidis 2003; Singh et al. 2009; Jin et al. 2011). Figure 2 illustrates the decision structures of the PA, multi-stage, and two-stage models. In particular, the left network in Figure 2 represents the decision structure of the multi-stage model, where the decision maker has a specific expansion plan for each node in the scenario tree. The middle network corresponds to a PA model, where each node n has an expansion plan $\mathbf{x}_n$ up to time period μ. After that, the subtree $\mathcal{T}(n)$ is "compressed" into a chain, where the decision maker only has one expansion plan for all the nodes in each following time period. The two-stage model Π_{TS} on the right has the most simple expansion plan, where each planning period has only one decision throughout the planning horizon.

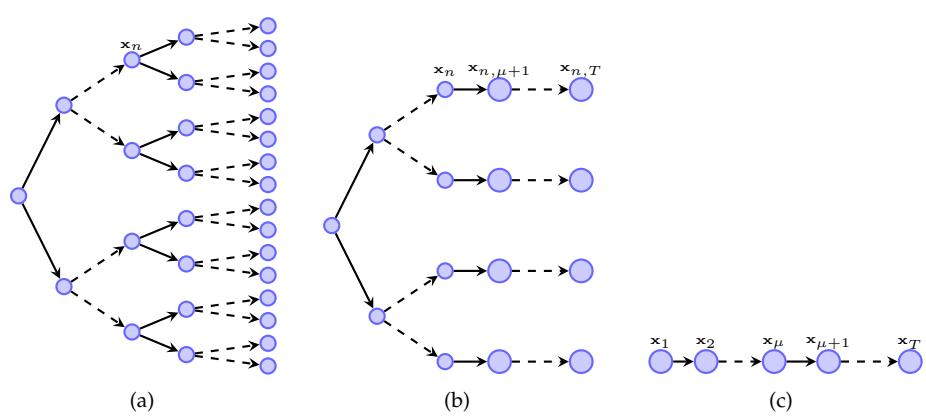

Figure 2: Capacity expansion decisions in $\mathbf{\Pi}_{\text{MS}}$ (a), $\mathbf{\Pi}_{\text{PA}}(\boldsymbol{\mu})$ (b) and $\mathbf{\Pi}_{\text{TS}}$ (c)
In a multi-stage model (a), $\mu = T$; in a PA model (b), $\mu = t_n$; in a two-stage model (c), $\mu = 1$.

For later use, we also provide explicit formulations for the two-stage and multi-stage capacity expansion models below.

$$[\Pi_{\text{TS}}] \quad \min \quad \sum_{t=1}^{T} \bar{\mathbf{a}}_t \mathbf{x}_t + \sum_{n \in \mathcal{T}} p_n \left(\sum_{k \in \mathcal{K}_{t_n}} \mathbf{b}_{nk}^\top \mathbf{y}_{nk} \right) \tag{2.3a}$$

$$\text{s.t.} \quad \sum_{s=1}^{t_n} \mathbf{x}_s \geq \mathbf{A}_{nk} \mathbf{y}_{nk}, \quad \forall\, k \in \mathcal{K}_{t_n}, n \in \mathcal{T} \tag{2.3b}$$

$$\sum_{s=1}^{T} \mathbf{x}_s \leq \mathbf{u} \tag{2.3c}$$

$$\mathbf{B}_{nk} \mathbf{y}_{nk} = \mathbf{d}_{nk}, \quad \forall\, k \in \mathcal{K}_{t_n}, n \in \mathcal{T} \tag{2.3d}$$

$$\mathbf{x}_t \in \mathbb{Z}_+^I,\ \mathbf{y}_{nk} \in \mathbb{R}_+^J, \quad \forall\, t = 1, \ldots, T, k \in \mathcal{K}_{t_n}, n \in \mathcal{T}. \tag{2.3e}$$

$$[\Pi_{\text{MS}}] \quad \min \quad \left\{ \sum_{n \in \mathcal{T}} p_n \left(\mathbf{a}_n^\top \mathbf{x}_n + \sum_{k \in \mathcal{K}_{t_n}} \mathbf{b}_{nk}^\top \mathbf{y}_{nk} \right) \ :\ \text{(2.2b), (2.2c), (2.2d), (2.2f).} \right\} \tag{2.4}$$

In Π_{TS}, $\bar{\mathbf{a}}_t$ is the average cost across the period t, i.e., $\bar{\mathbf{a}}_t = \sum_{n \in \mathcal{S}_t} p_n \mathbf{a}_n$. Note that in a two-stage formulation, other choices of investment cost, such as the highest cost in each stage may also be used.

3 Performance of $\Pi_{\text{PA}}(\mu)$

In this section, we investigate the performance of the PA model $\Pi_{\text{PA}}(\mu)$, namely, the difference between optimal values of the PA model and the multi-stage model Π_{MS}. Let $v^{\text{PA}}(\mu)$ and v^{MS} be the optimal value of $\Pi_{\text{PA}}(\mu)$ and Π_{MS}, respectively. Note that an optimal solution to $\Pi_{\text{PA}}(\mu)$ is a feasible solution to the multi-stage model Π_{MS} with value $v^{\text{PA}}(\mu)$. Thus, we define $\text{Gap}(\mu) := v^{\text{PA}}(\mu) - v^{\text{MS}}$ as the performance gap between the two models. It is clear that $\text{Gap}(\mu) \geq 0$, for all $1 \leq \mu \leq T$, and $\text{Gap}(T) = 0$, since the optimal solution to the PA model is a feasible solution to the multi-stage model. Also, $\text{Gap}(1)$ is the gap between two-stage and multi-stage models, which is studied in Huang and Ahmed (2009). Therefore, the following analysis generalizes the bounds therein.

Notice that both $\Pi_{\text{PA}}(\mu)$ and Π_{MS} are integer programs, thus they both become difficult to solve when the planning horizon is long. Moreover, as long as $\mu < T$, Π_{MS} will have significantly more integer variables than $\Pi_{\text{PA}}(\mu)$ since the number of nodes in a scenario tree grows exponentially in the number of planing periods. Thus the multi-stage model becomes much more difficult to solve. If $\text{Gap}(\mu)$ is small for modest μ, then $\Pi_{\text{PA}}(\mu)$ can provide a good, easier-to-compute approximation to Π_{MS} with guaranteed performance. For this reason, we provide analytical lower and upper bounds for $\text{Gap}(\mu)$, using instance data and the optimal LP-relaxation solutions of these two models.

A brief outline of our approach is summarized as follows. Motivated by an important substructure of these models, we decompose the original problem into subproblems, each of which corresponds to a single type of expansion technology. We solve the LP relaxations of the original PA and multi-stage models with multiple types of technologies. Then we use their optimal solutions and input data to bound the gap for single-technology subproblems. Finally we aggregate these bounds for subproblems to obtain both upper and lower bounds on $\text{Gap}(\mu)$.

In the following, we derive an upper bound in detail, a lower bound can be derived in a similar fashion

as explained in subsection 3.4. Main results are summarized in Theorems 3 and 4 in Section 3.4. All the proofs can be found in the Online Supplement.

3.1 Decomposition reformulation

We follow a similar approach in Huang and Ahmed (2009) to decompose the capacity expansion planning model by separating capacity expansion decisions from operation decisions. For simplicity, we let $\mathbf{x}$ and $\mathbf{y}$ denote vectors $\{\mathbf{x}_n\}_{n\in\mathcal{T}}$ and $\{\{\mathbf{y}_{nk}\}_{k\in\mathcal{K}_{t_n}}\}_{n\in\mathcal{T}}$. The multi-stage model Π_{MS} can be decomposed as follows.

$$v^{\text{MS}} = \min\left\{\sum_{n\in\mathcal{T}}\sum_{k\in\mathcal{K}_{t_n}} p_n\mathbf{b}_{nk}^\top\mathbf{y}_{nk} + \sum_{i\in\mathcal{I}} V_i(\mathbf{y}) \;:\; \mathbf{B}_{nk}\mathbf{y}_{nk} = \mathbf{d}_{nk}, \mathbf{y}_{nk}\in\mathbb{R}_+^J, \forall k\in\mathcal{K}_{t_n}, n\in\mathcal{T}\right\}, \tag{3.1}$$

where for each $i\in\mathcal{I}$,

$$V_i(\mathbf{y}) = \min\left\{\sum_{n\in\mathcal{T}} p_n a_{in}x_{in} \;:\; [\mathbf{A}_{nk}\mathbf{y}_{nk}]_i \le \sum_{m\in\mathcal{P}(n)} x_{im} \le u_i, x_{in}\in\mathbb{Z}_+, \forall k\in\mathcal{K}_{t_n}, n\in\mathcal{T}\right\}. \tag{3.2}$$

The above reformulation allows problem (3.2) to be solved for each type of technology individually, if operation decision $\mathbf{y}$ is given. Let $\{\mathbf{x}^{\text{MLP}}, \mathbf{y}^{\text{MLP}}\}$ be an optimal solution to the linear relaxation of the multi-stage model Π_{MS}, and let $\delta_{in} = \max_{k\in\mathcal{K}_{t_n}}\{[\mathbf{A}_{nk}\mathbf{y}_{nk}^{\text{MLP}}]_i\}$ for all $i\in\mathcal{I}$ and $n\in\mathcal{T}$. Since $\mathbf{y}^{\text{MLP}}$ is optimal in the LP relaxation of Π_{MS} but not necessarily in $\Pi_{\text{PA}}(\mu)$, and $\Pi_{\text{PA}}(\mu)$ is an integer program, we have

$$v^{\text{PA}}(\mu) \le \sum_{n\in\mathcal{T}}\sum_{k\in\mathcal{K}_{t_n}} p_n\mathbf{b}_{nk}^\top\mathbf{y}_{nk}^{\text{MLP}} + \sum_{i\in\mathcal{I}} o_i^{\text{P}}(\mu), \quad v^{\text{MS}} \ge \sum_{n\in\mathcal{T}}\sum_{k\in\mathcal{K}_{t_n}} p_n\mathbf{b}_{nk}^\top\mathbf{y}_{nk}^{\text{MLP}} + \sum_{i\in\mathcal{I}} v_i^{\text{M}},$$

where

$$o_i^{\text{P}}(\mu) = \min \quad \sum_{n\in\mathcal{T}} p_n a_{in}x_{in} \tag{3.3a}$$

$$\text{s.t.} \quad \sum_{m\in\mathcal{P}(n)} x_{im} \ge \delta_{in}, \quad \forall\, n\in\mathcal{T} \tag{3.3b}$$

$$\sum_{m\in\mathcal{P}(n)} x_{im} \le u_i, \quad \forall\, n\in\mathcal{S}_T \tag{3.3c}$$

$$x_{in_1} = x_{in_2}, \quad \forall\, n_1, n_2\in\mathcal{T}(m)\cap\mathcal{S}_t, m\in S_\mu, t>\mu \tag{3.3d}$$

$$x_{in}\in\mathbb{Z}_+, \quad \forall\, n\in\mathcal{T}, \tag{3.3e}$$

and

$$v_i^{\text{M}} = \min \quad \left\{\sum_{n\in\mathcal{T}} p_n a_{in}x_{in} \;:\; \text{s.t. (3.3b)} - \text{(3.3c)},\; x_{in}\in\mathbb{R}_+, \quad \forall\, n\in\mathcal{T}.\right\} \tag{3.4}$$

Note that $o_i^{\text{P}}(\mu)$ is the optimal investment cost of a partially adaptive, single-technology capacity expansion planning problem with fixed operation decisions $\mathbf{y}^{\text{MLP}}$. Let $v_i^{\text{P}}(\mu)$ denote the optimal value of its LP relaxation. v_i^{M} is the optimal investment cost of the LP relaxation of a multi-stage, single-technology problem with fixed operation decisions $\mathbf{y}^{\text{MLP}}$. Moreover, $\mathbf{u}$ is an integral vector, thus $\delta_{in}\le\lceil\delta_{in}\rceil\le u_i$ for all $i\in\mathcal{I}$ and $n\in\mathcal{T}$. Since x_{in} takes nonnegative integer values in (3.3), the problem remains the same if the

right-hand-side of constraints (3.3b) is rounded up to $\lceil \delta_{in} \rceil$.

Because of the decomposition reformulation, we can focus on single-technology problems, and bound the gap between the PA model and the multi-stage model by

$$\text{Gap}(\mu) = v^{\text{PA}}(\mu) - v^{\text{MS}} \leq \sum_{i \in \mathcal{I}} (o_i^{\text{P}}(\mu) - v_i^{\text{M}}). \tag{3.5}$$

3.2 Some useful results for single-technology problems

Reformulation of (3.3):

In problem (3.3), constraints (3.3d) enforce two-stage approach after period μ at each possible outcome in that period. This is equivalent to combining some nodes into a single node after period μ, but still maintaining a tree structure in the capacity expansion decisions, as illustrated in Figure 2. Therefore, we can formulate a equivalent multi-stage model on the "compressed" tree as follows.

$$o_i^{\text{P}}(\mu) = \min \quad \sum_{n \in \hat{\mathcal{T}}} \hat{p}_n \hat{a}_{in} x_{in} \tag{3.6a}$$

$$\text{s.t.} \quad \sum_{m \in \mathcal{P}(n)} x_{im} \geq \lceil \hat{\delta}_{in} \rceil, \quad \forall\, n \in \hat{\mathcal{T}} \tag{3.6b}$$

$$\sum_{m \in \mathcal{P}(n)} x_{im} \leq u_i, \quad \forall\, n \in \mathcal{S}_T \tag{3.6c}$$

$$x_{in} \in \mathbb{Z}_+, \quad \forall\, n \in \hat{\mathcal{T}}, \tag{3.6d}$$

where $\hat{\mathcal{T}}$ is the compressed tree, for any node $n \in \hat{\mathcal{T}}$ such that $t_n \leq \mu$, we have $\hat{p}_n = p_n$, $\hat{a}_{in} = a_{in}$, $\hat{\delta}_{in} = \delta_{in}$. For any node $n \in \hat{\mathcal{T}}$ such that $t_n > \mu$, let $U_n \subset \mathcal{T}$ consist the nodes that are "compressed" to node n by constraints (3.3d), and $\hat{p}_n = \sum_{m \in U_n} p_m$, $\hat{a}_{in} = \sum_{m \in U_n} p_m a_{im}$, $\hat{\delta}_{in} = \max\{\delta_{im} : m \in U_n\}$.

Totally unimodularity of (3.6):

The following result shows that the feasible region of problem (3.6) is an integral polytope as long as the right-hand-side vector is integral. The boundedness follows directly from upper bound constraints (3.6c) and nonnegativity of x_{in}. As a result, the LP relaxation of problem (3.6) admits integer optimal solutions.

Proposition 1. *In problem* (3.6)*, the left-hand-side coefficient matrix is totally unimodular.*

Redundancy of constraints (3.6c):

Since $\{\delta_{in}, i \in \mathcal{I}, n \in \mathcal{T}\}$ are defined by an optimal solution $\{\mathbf{x}^{\text{MLP}}, \mathbf{y}^{\text{MLP}}\}$ to the LP relaxation of multi-stage model, we can further simplify problem (3.6) by removing the redundant constraints (3.6c).

Proposition 2. *If $\hat{p}_n \hat{a}_{in} > 0$ for all $n \in \hat{\mathcal{T}}$ and $i \in \mathcal{I}$, then in the LP relaxation of problem* (3.6)*, constraints* (3.6c) *are redundant.*

As a remark, the result in Proposition 2 also applies to the LP relaxation of problem (3.3), as well as problem (3.4).

3.3 An upper bound on $o_i^{\mathrm{P}}(\mu) - v_i^{\mathrm{M}}$

There are two main steps in obtaining an upper bound on $o_i^{\mathrm{P}}(\mu) - v_i^{\mathrm{M}}$.

- *Step 1.* We show that $o_i^{\mathrm{P}}(\mu) \leq v_i^{\mathrm{P}}(\mu) + C$, where C is some constant that is dependent on the input data and $\{\delta_{in}\}_{n\in\mathcal{T}}$.
- *Step 2.* We derive upper and lower bounds for $v_i^{\mathrm{P}}(\mu)$ and v_i^{M}.

Then an upper bound for $o_i^{\mathrm{P}}(\mu) - v_i^{\mathrm{M}}$ can be expressed as "upper bound of $v_i^{\mathrm{P}}(\mu) -$ lower bound of $v_i^{\mathrm{M}} + C$".

Step 1

With previous results for the single-technology problems, we have the following result.

Proposition 3. *$o_i^P(\mu) \leq v_i^P(\mu) + a_{i1} \cdot \lambda_i$, where $\lambda_i = \max_{n\in\mathcal{T}}\{\lceil \delta_{in} \rceil - \delta_{in}\}$. If $\{\delta_{in}\}_{n\in\mathcal{T}}$ are all integers, the inequality is tight.*

Step 2

Before presenting lower and upper bounds for $v_i^{\mathrm{P}}(\mu)$ and v_i^{M}, we define some useful parameter and show their relations. For simplicity, we omit the index i in this step.

Let us define

$$\delta^{(\mu-)} := \sum_{n\in S_{\mu-1}} p_n \max_{m\in\mathcal{P}(n)}\{\delta_m\}, \quad \delta^{(\mu)} := \sum_{n\in S_{\mu}} p_n \max_{m\in\mathcal{P}(n)\cup\mathcal{T}(n)}\{\delta_m\};$$

$$\underline{a}_{\mu-} := \min_{n\in\mathcal{T}:t_n\leq\mu}\{a_n\}, \quad \underline{a}_{\mu+} := \min_{n\in\mathcal{T}:t_n\geq\mu}\{a_n\}, \quad a_* = \min_{n\in\mathcal{T}}\{a_n\};$$

$$\bar{a}_{\mu-} := \max_{n\in\mathcal{T}:t_n\leq\mu}\{a_n\}, \quad \bar{a}_{\mu+} := \max_{n\in\mathcal{T}:t_n\geq\mu}\{a_n\}, \quad a^* = \max_{n\in\mathcal{T}}\{a_n\}.$$

If $\mu = 1$, we have $\delta^{(1-)} = 0$, $\delta^{(1)} = \max_{n\in\mathcal{T}}\{\delta_n\}$, and $\underline{a}_{1-} = \bar{a}_{1-} = a_1$, $\underline{a}_{1+} = a_*$, $\bar{a}_{1+} = a^*$; if $\mu = T$, then $\delta^{(T)} = \sum_{n\in\mathcal{S}_T} p_n \max_{m\in\mathcal{P}(n)}\{\delta_m\}$, and $\underline{a}_{T-} = a_*$, $\bar{a}_{T-} = a^*$. It is easy to see that as μ increases from 1 to T, $\delta^{(\mu-)}$, $\underline{a}_{\mu+}$ and $\bar{a}_{\mu-}$ are monotone increasing while $\delta^{(\mu)}$, $\underline{a}_{\mu-}$ and $\bar{a}_{\mu+}$ are monotone decreasing. One can treat $\{\delta_n\}_{n\in\mathcal{T}}$ as the "demand" in the single-technology problem, then $\delta^{(1)}$ is the largest demand across the entire scenario tree, and $\delta^{(T)}$ is the average of the maximum demand in each scenario (path). The following proposition reveals the relation between these δ's.

Proposition 4. *For any $\mu \in \{1, \ldots, T\}$, the following relation holds, $\delta^{(\mu-)} \leq \delta^{(T)} \leq \delta^{(\mu)} \leq \delta^{(1)}$.*

Now we derive lower and upper bounds for $v^{\mathrm{P}}(\mu)$, and the bounds for v^{M} can be attained by setting $\mu = T$. We briefly discuss the idea of finding these bounds:

- *Lower bound:* starting from an optimal solution to the LP relaxation of (3.3), relax the coefficients in the objective funtion to reach a lower bound for $v^{\mathrm{P}}(\mu)$;
- *Upper bound:* construct a feasible solution to the LP relaxation of (3.3), and the objective function value given by this solution yields an upper bound for $v^{\mathrm{P}}(\mu)$.

Specifically, we have the following theorem.

Theorem 1. $(\underline{a}_{\mu-} - a_*)\delta^{(\mu-)} + a_*\delta^{(\mu)} \leq v^P(\mu) \leq (\bar{a}_{\mu-} - \bar{a}_{\mu+})\delta^{(\mu-)} + \bar{a}_{\mu+}\delta^{(\mu)}$.

Setting $\mu = T$ and applying Proposition 4, we immediately have the following corollary.

Corollary 1. $a_* \delta^{(T)} \leq v^M \leq a^* \delta^{(T)}$.

Suppose that the cost parameters a_n are nearly constant, that is, $a^* \approx a_* \approx \underline{a}_{\mu-} \approx \underline{a}_{\mu+} \approx \bar{a}_{\mu-} \approx \bar{a}_{\mu+} \approx a$, then we have $v^{\mathrm{P}}(\mu) \approx a\delta^{(\mu)}$ and $v^{\mathrm{M}} \approx a\delta^{(T)}$. As μ increases from 1 to T, this approximate value of $v^{\mathrm{P}}(\mu) - v^{\mathrm{M}}$ decreases from $a\left(\delta^{(1)} - \delta^{(T)}\right)$ to 0. Combining results in these two steps, we obtain an upper bound of $o^{\mathrm{P}}(\mu) - v^{\mathrm{M}}$, summarized in the following theorem.

Theorem 2. *$o^P(\mu) - v^M \leq \left[(\bar{a}_{\mu-} - \bar{a}_{\mu+})\delta^{(\mu-)} + \bar{a}_{\mu+}\delta^{(\mu)} - a_*\delta^{(T)}\right] + a_1 \cdot \lambda_i$, where $\lambda_i = \max_{n\in\mathcal{T}}\{\lceil \delta_{in} \rceil - \delta_{in}\}$.*

3.4 Upper and lower bounds for Gap(μ)

Upper bound:

As discussed before, aggregating the bound in Theorem 2 for each type of capacity expansion technology yields an upper bound for Gap(μ).

Theorem 3. *Let $\mathbf{y}^{MLP}$ be the operation level decisions in an optimal solution to the linear relaxation of multi-stage model Π_{MS}. For each type of expansion technology $i \in \mathcal{I}$, let $\delta_{in} = \max_{k\in\mathcal{K}_{t_n}}\{[\mathbf{A}_{nk}\mathbf{y}^{MLP}_{nk}]_i\}$, and $\lambda_i = \max_{n\in\mathcal{T}}\{\lceil \delta_{in} \rceil - \delta_{in}\}$. We further define*

$$a_{i,*} = \min_{n\in\mathcal{T}}\{a_{in}\}, \quad \bar{a}_{i,\mu-} = \max_{n\in\mathcal{T}:t_n\leq\mu}\{a_{in}\}, \quad \bar{a}_{i,\mu+} = \max_{n\in\mathcal{T}:t_n\geq\mu}\{a_{in}\},$$
$$\delta_i^{(\mu-)} = \sum_{n\in\mathcal{S}_{\mu-1}} p_n \max_{m\in\mathcal{P}(n)}\{\delta_{im}\}, \quad \delta_i^{(\tau)} = \sum_{n\in\mathcal{S}_\tau} p_n \max_{m\in\mathcal{P}(n)\cup\mathcal{T}(n)}\{\delta_{im}\},$$

for $\tau = \mu, T$. Then for a capacity expansion planning problem with multiple technologies,

$$Gap(\mu) \leq \sum_{i\in\mathcal{I}}\left[(\bar{a}_{i,\mu-} - \bar{a}_{i,\mu+})\delta_i^{(\mu-)} + \bar{a}_{i,\mu+}\delta_i^{(\mu)} - a_{i,*}\delta_i^{(T)}\right] + \sum_{i\in\mathcal{I}} a_{i1}\cdot\lambda_i. \tag{3.7}$$

Lower bound:

Now suppose we start the entire analysis with an optimal solution $\{\mathbf{x}^{\mathrm{PLP}}, \mathbf{y}^{\mathrm{PLP}}\}$ to the LP relaxation of PA model $\Pi_{\mathrm{PA}}(\mu)$, and let $\gamma_{in} = \max_{k\in\mathcal{K}_{t_n}}\{[\mathbf{A}_{nk}\mathbf{y}^{\mathrm{PLP}}_{nk}]_i\}$ for all $i \in \mathcal{I}$ and $n \in \mathcal{T}$. Then we have

$$v^{\mathrm{PA}}(\mu) \geq \sum_{n\in\mathcal{T}}\sum_{k\in\mathcal{K}_{t_n}} p_n \mathbf{b}_{nk}^\top \mathbf{y}^{\mathrm{PLP}}_{nk} + \sum_{i\in\mathcal{I}} v_i^{\mathrm{P}}(\mu), \quad v^{\mathrm{MS}} \leq \sum_{n\in\mathcal{T}}\sum_{k\in\mathcal{K}_{t_n}} p_n \mathbf{b}_{nk}^\top \mathbf{y}^{\mathrm{PLP}}_{nk} + \sum_{i\in\mathcal{I}} o_i^{\mathrm{M}},$$

where $v_i^{\mathrm{P}}(\mu)$ and o_i^{M} are defined similarly as (3.3) and (3.4) with δ_{in} substituted by γ_{in}. Applying similar techniques as in Theorem 3, we can obtain

$$v^{\mathrm{PA}}(\mu) - v^{\mathrm{MS}} \geq \sum_{i\in\mathcal{I}}(v_i^{\mathrm{P}}(\mu) - o_i^{\mathrm{M}}) \geq \sum_{i\in\mathcal{I}}[v_i^{\mathrm{P}}(\mu) - v_i^{\mathrm{M}} - a_{i1}\cdot\eta_i], \tag{3.8}$$

where $\eta_i = \max_{n\in\mathcal{T}}\{\lceil \gamma_{in} \rceil - \gamma_{in}\}$. Putting Theorem 1, Corollary 1, and (3.8) together, the following result follows immediately.

Theorem 4. *Let $\mathbf{y}^{PLP}$ be the operation level decisions in an optimal solution to the linear relaxation of PA model $\Pi_{PA}(\mu)$. For each type of generator $i \in \mathcal{I}$, let $\gamma_{in} = \max_{k\in\mathcal{K}_{t_n}}\{[\mathbf{A}_{nk}\mathbf{y}^{PLP}_{nk}]_i\}$, and $\eta_i = \max_{n\in\mathcal{T}}\{\lceil \gamma_{in} \rceil - \gamma_{in}\}$. We*

further define

$$a_i^* = \max_{n\in\mathcal{T}}\{a_{in}\}, \quad a_{i,*} = \min_{n\in\mathcal{T}}\{a_{in}\}, \quad \underline{a}_{i,\mu-} = \max_{n\in\mathcal{T}:t_n\le\mu}\{a_{in}\},$$
$$\gamma_i^{(\mu-)} = \sum_{n\in\mathcal{S}_{\mu-1}} p_n \max_{m\in\mathcal{P}(n)}\{\gamma_{im}\}, \quad \gamma_i^{(\tau)} = \sum_{n\in\mathcal{S}_\tau} p_n \max_{m\in\mathcal{P}(n)\cup\mathcal{T}(n)}\{\gamma_{im}\},$$

for $\tau = \mu, T$. Then for a capacity expansion planning problem with multiple technologies,

$$Gap(\mu) \ge \sum_{i\in\mathcal{I}}\left[(\underline{a}_{i,\mu-} - a_{i,*})\gamma_i^{(\mu-)} + a_{i,*}\gamma_i^{(\mu)} - a_i^*\gamma_i^{(T)}\right] - \sum_{i\in\mathcal{I}} a_{i1}\cdot\eta_i. \tag{3.9}$$

The bounds in Theorem 3 and 4 are dependent on the input data, in particular the investment cost $\mathbf{a}$, and optimal solutions to the LP relaxation of Π_{MS} and $\Pi_{\text{PA}}(\mu)$. Note that these bounds could be weak when investment costs in the first planning period are very large and λ_i, η_i are close to 1. For general cost structures, however, decision makers can choose a value of μ within $[1, T]$ and calculate these bounds. If these bounds indicate that the current PA model with parameter μ is a good enough approximation to the multi-stage model, then we can just solve the PA model without bearing addition computational effort. If the bounds are still not appealing, a larger value of μ should be considered. The following small example demonstrate the possibility that the bounds in Theorem 3 and Theorem 4 could be useful in assessing the critical values for μ.

Example 1. Consider a single technology problem. The underlying scenario tree has three stages and each node has two branches with equal probability. Let $\mathcal{S}_1 = \{1\}$, $\mathcal{S}_2 = \{2,3\}$, $\mathcal{S}_3 = \{4,5,6,7\}$. Assume investment cost $a_1 = 10$, $a_2 = \cdots = a_7 = 8$, and unit generation cost is 1. Let the demand be $d_1 = 1$, $d_2 = 3$, $d_3 = 5$, $d_4 = 4$, $d_5 = 5$, $d_6 = 5$, $d_7 = 6$. We assume that one unit of demand requires one unit of generator. Thus solving a multistage problem yields an optimal cost of 71, and solving a PA model with $\mu = 2$ yields an optimal cost of 75, thus $Gap(2) = 4$. According to (3.7) and (3.9), we have $4 \le Gap(2) \le 6$, the lower bound is actually tight.

4 An approximation algorithm for solving Π_{MS}

In this section, we first propose an approximation algorithm based on PA, which recursively traverses the scenario tree and computes a solution to the multi-stage stochastic optimization model. Then, we identify sufficient conditions under which the obtained solution is indeed optimal for the multi-stage model.

4.1 Algorithm description

The key idea of the proposed algorithm is to traverse the scenario tree $\mathcal{T}$ and at each node $n \in \mathcal{T}$ solve a PA model on the subtree $\mathcal{T}(n)$, then use the solutions of these PA models to synthesize a multi-stage solution. To be more precise, let us first introduce some notations. We denote the PA model (2.2) formulated on the subtree $\mathcal{T}(n)$ as $\Pi_{\text{PA}}(\mu, n)$ and $\Pi_{\text{PA}}(\mu)$ always means $\Pi_{\text{PA}}(\mu, 1)$. Also denote the optimal expansion and generation decision of $\Pi_{\text{PA}}(\mu, n)$ as $\mathbf{z}^{\mu,n} := \{\mathbf{x}_m^{\mu,n}, \{\mathbf{y}_{mk}^{\mu,n}\}_{k\in\mathcal{K}_m}\}_{m\in\mathcal{T}(n)}$. A *topological ordering* $\supset(\mathcal{T})$ on the nodes of a scenario tree $\mathcal{T}$ is a linear ordering such that for every edge (u, v) from parent node u to child node v, u comes before v in the ordering $\sigma(\mathcal{T})$. An important implication is that, when the nodes of the tree

are traversed in a topological ordering, a node n is always visited *after* all its ancestor nodes on the path $\mathcal{P}(n)$ from the root node to n are visited.

The proposed algorithm visits each node $n \in \mathcal{T}$ in the order $\sigma(\mathcal{T})$ and solves the PA model $\Pi_{\text{PA}}(\mu, n)$ on the subtree $\mathcal{T}(n)$ rooted at node n. In this process, the root node's optimal solution $\mathbf{z}_n^{\mu,n}$ of $\Pi_{\text{PA}}(\mu, n)$ is recorded. Since $\sigma(\mathcal{T})$ is a topological ordering, at node n, all its ancestors k on the path $\mathcal{P}(n)$ have already been traversed by the algorithm, we can use the obtained generation capacity solutions $\{\mathbf{x}_m^{\mu,m}\}_{m \in \mathcal{P}(n)}$ of these ancestors to compute the initial installed generation capacity for the $\Pi_{\text{PA}}(\mu, n)$ problem (cf. constraints (2.2b) in model (2.2)). After the entire tree $\mathcal{T}$ is traversed in this way, the algorithm outputs a solution $\{\mathbf{z}_n^{\mu,n}\}_{n=1}^N$, in which the expansion decision has a fully adaptive structure over the entire planning horizon. This procedure is summarized in Algorithm 1.

Algorithm 1

Input: A topological ordering $\sigma(\mathcal{T}) = \{1, 2, \ldots, N\}$ of all the nodes in $\mathcal{T}$, a critical time μ, and the initial installed capacity $u_0 = 0$.

Solve the PA model $\Pi_{\text{PA}}(\mu, 1)$ at root node 1. Denote the optimal solution as $\{\mathbf{x}^{\mu,1}, \mathbf{y}^{\mu,1}\}$.
Update $\mathbf{x}_1 \leftarrow \mathbf{x}_1^{\mu,1}$ and $\mathbf{y}_1 \leftarrow \mathbf{y}_1^{\mu,1}$.
for $n = 2, \ldots, N$ **do**
  Solve $\Pi_{\text{PA}}(\mu, n)$ on $\mathcal{T}(n)$ with initial installed capacity computed from $\mathbf{x}_k$ for $k \in \mathcal{P}(a(n))$. Denote its optimal solution as $\{\mathbf{x}^{\mu,n}, \mathbf{y}^{\mu,n}\}$.
  Update $\mathbf{x}_n \leftarrow \mathbf{x}_n^{\mu,n}$ and $\mathbf{y}_n \leftarrow \mathbf{y}_n^{\mu,n}$.
end for
return $\{\mathbf{x}_n, \mathbf{y}_n\}_{n=1}^N$.

We would like to remark that the algorithm can use different critical times μ_n at different node n. The algorithm can also be terminated before visiting all the nodes in the scenario tree — the resulting solution is always a feasible multi-stage solution. Also note that there is flexibility in choosing the topological ordering $\sigma(\mathcal{T})$. We will look into this issue in the computation section.

4.2 Optimality of Algorithm 1 for a special GEP problem

In this subsection, we present a sufficient condition under which Algorithm 1 recovers an optimal solution for a multi-stage generation expansion planning problem. The proof is given in Online Supplement.

Theorem 5. *For a multi-stage generation expansion planning problem* (2.4)*, if the instance satisfies the following conditions:*

i) the unit investment costs of each type of generators are stationary, i.e., in (2.1a)*,* $c_{it_n} = c_i$ *for all* $i \in \mathcal{I}$ *and* $n \in \mathcal{T}$*;*
ii) all generators share the same unit generation cost, i.e., in (2.1a)*,* $b_{ink} = b_{nk}$ *for all* $k \in \mathcal{K}_{t_n}$ *and* $n \in \mathcal{T}$*;*
iii) demand must be satisfied by generation, i.e., no penalty is allowed;

then the solution returned by Algorithm 1 is optimal to the multi-stage problem.

In Theorem 5, the exactness of Algorithm 1 is established on the assumption that all generators share the same generation cost, which may seem restrictive. However, if all generators considered have similar generation costs, i.e., $b_{ink} \in [\bar{b}_{nk} - \epsilon, \bar{b}_{nk} + \epsilon]$ where ϵ is small for all $i \in \mathcal{I}$, one can use Algorithm 1 to obtain a multistage solution to the problem with generation cost $\{\bar{b}_{nk}\}_{n,k}$, and that solution will have an objective

function value of at most $v^{\mathrm{MS}} + \sum_{n \in \mathcal{T}} \sum_{k \in \mathcal{K}_{t_n}} \frac{p_n d_{nk} \epsilon}{(1+r)^{t_n - 1}}$, where v^{MS} is the optimal value for the multistage problem with original generation costs. Thus, this result suggests that the algorithm produces nearly optimal solutions for a larger class of problems than proved in the theorem.

5 Computational Experiments

In this section, we present extensive computational experiments to evaluate the proposed partially adaptive stochastic model and Algorithm 1.

5.1 Experiment data and setup

All the data in the experiments are obtained from the real world data collected by Jin et al. (2011). In particular, the authors of Jin et al. (2011) collected hourly electricity demand data from years 1991 to 2007 for the Midwest region from the Midcontinent Independent System Operator (MISO), the natural gas price data from years 1970 to 2006 for the same region from the Energy Information Administration (EIA), and generation build cost data suggested by the Joint Coordinated System Planning Report 2008 (JCSP). The GEP model has the year 2008 as the reference year and a 10-year planning horizon.

The uncertainty in the GEP problem comes from two sources: the natural gas price and the electricity demand. Jin et al. (2011) verified that both of these stochastic processes can be reasonably modeled as geometric Brownian motions with high temporal correlation. Hourly electricity demand is aggregated into three types sub-periods: peak-load, medium-load, and low-load, according to the load duration curve of the reference year. The following years' load duration curves are assumed to share the same structure as the reference year's with incremental demand growth. There are six types of generators available for capacity expansion, namely Base Load, Combined Cycle (CC), Combined Turbine (CT), Nuclear, Wind, and Integrated Gasification Combined Cycle (IGCC). Among these six types of generators, both CC and CT power plants are fueled by natural gas, which are subject to price uncertainty, and IGCC power plants are fueled by coal, whose price is usually quite stable and thus assumed known.

The 10-year scenario tree is generated by applying a nonlinear programming approach introduced in Høyland and Wallace (2001). In particular, the discrete samples and their associated probability structure for the scenario tree are constructed so that the approximate distribution matches as well as possible the desired statistical properties of the underlying continuous random variables, i.e. electricity demand and fuel prices. The number of branches at each node in the scenario tree is chosen to be 3 to balance the size of the tree and the accuracy of approximation (Jin et al. (2011)). For our case, the resulting tree has in total $3^9 = 19,683$ scenarios. The total expected generator investment cost and operation cost are discounted at an annual rate of 8%.

When solving any mixed integer problem, the relative MIP optimality gap is set to 5×10^{-3}. Also, we impose a time limit of 5 hours (18000 sec) on solving any of the two-stage, multi-stage, and PA models. The same limit is applied to the total computation time of Algorithm 1 as well. Algorithm 1 is implemented in Python 2.7.8 with Gurobi Python interface and Gurobi 5.5.0 as the MIP solver. All numerical experiments are conducted on a Macbook Pro with 8G RAM and a 2.3 GHz Intel Core i5 processor.

5.2 Performance of PA model

In the first set of experiments, we solve the PA model $\Pi_{\text{PA}}(\mu)$ for different planning horizons T and different critical times μ. The experiments aim to answer two questions: (1) What is the value of the multi-stage model comparing to the two-stage model? (2) How fast can the PA model be solved and how well does the PA model approximate the multi-stage model as μ increase? Recall again, for a fixed T, $\mu = 1$ corresponds to the two-stage model, and $\mu = T$ corresponds to the multi-stage model. Results are presented in Table 2.

Table 2: Solving PA models on GEP instances

T	μ	MIPObj (million \$)	OptGap (%)	$v^{\text{PA}}/v^{\text{MS}} - 1$ (%)	Time (sec.)
3	1	[3065.34, 3079.21]	0.45	[43.34, 44.39]	0.03
	2	[2426.14, 2435.36]	0.39	[13.45, 14.20]	0.06
	3	[2132.58, 2138.57]	0.28	0	0.08
4	1	[4598.61, 4598.61]	0.00	[43.70, 44.42]	0.17
	2	[4046.02, 4066.33]	0.50	[26.43, 27.70]	2.27
	3	[3436.09, 3453.31]	0.50	[7.37, 8.45]	5.45
	4	[3184.19, 3200.11]	0.50	0	4.11
5	1	[6225.63, 6250.71]	0.40	[42.92, 44.57]	0.26
	2	[5667.38, 5685.59]	0.32	[30.11, 31.50]	26.46
	3	[5029.97, 5072.60]	0.85	[15.47, 17.32]	18004.46
	5	[4323.61, 4355.93]	0.75	0	18002.52
6	1	[8018.62, 8055.41]	0.40	[42.41, 44.99]	2.73
	2	[7360.44, 7397.24]	0.50	[30.72, 33.15]	16.56
	4	[6203.99, 6267.00]	1.02	[10.18, 12.80]	18003.85
	6	[5555.77, 5630.84]	1.35	0	18003.95
7	1	[9889.73, 9927.02]	0.38	[41.02, 43.57]	9.72
	2	[9199.12, 9236.99]	0.41	[31.17, 33.59]	13.86
	4	[8033.43, 8097.94]	0.80	[14.55, 17.11]	18010.35
	7	[6914.58, 7013.13]	1.43	0	18002.95
8	1	[11825.95, 11884.75]	0.50	[39.41, 42.37]	86.82
	2	[11135.28, 11189.00]	0.48	[31.27, 34.04]	309.86
	5	[9401.26, 9510.03]	1.16	[10.83, 13.92]	18000.91
	8	[8347.79, 8482.69]	1.62	0	18000.30
9	1	[13717.37, 13735.28]	0.13	[35.56, 40.14]	343.97
	2	[13110.29, 13174.39]	0.49	[29.56, 34.42]	1261.38
	4	[11849.82, 11941.19]	0.77	[17.11, 21.84]	18001.53
	7	[10367.49, 10509.44]	1.37	[2.46, 7.23]	18017.75
	9	[9801.03, 10118.72]	3.24	0	18007.46
10	1	[15606.96, 15680.86]	0.47	[31.42, 38.82]	3730.84
	2	[14998.67, 15087.53]	0.59	[26.30, 33.57]	18002.23
	5	[13225.33, 13385.83]	1.21	[11.36, 18.50]	18020.00
	8	[11793.87, 12127.18]	2.83	[0.00, 7.36]	18004.19
	10	[11295.61, 11875.77]	5.14	0	18083.68

Columns 1 and 2 in Table 2 show the length of the planning horizon of each instance and the value of μ, respectively. Column 3 presents two numbers: the left one is the best lower bound on the optimal cost of the $\Pi_{\text{PA}}(\mu)$ model found within the time limit, denoted as $v_L(\mu)$, and the right one is the cost of the best feasible solution for $\Pi_{\text{PA}}(\mu)$, denoted as $v_H(\mu)$. Column 4 computes the gap between these

two numbers, namely $(v_H(\mu) - v_L(\mu))/v_L(\mu)$, which measures the quality of the solution of the $\Pi_{\text{PA}}(\mu)$ model. Column 5 computes lower and upper bounds on the gap between the $\Pi_{\text{PA}}(\mu)$ model and the multi-stage model. In particular, the left number in Column 5 is $(v_L(\mu) - v_H(T))/v_H(T)$, and the right one is $(v_H(\mu) - v_L(T))/v_L(T)$. For example, when $T = 6$, $\mu = 4$, the lower bound is $(6203.99 - 5630.84)/5630.84 \times 100\% = 10.18\%$, and the upper bound is $(6267.00 - 5555.77)/5555.77 \times 100\% = 12.80\%$. Column 6 reports the wall clock time of the MIP solver.

Table 2 provides answers to the above two questions. First, there is a significant value in solving the multi-stage model. In particular, comparing to the two-stage model, the multi-stage model reduces the total expected cost by more than 30% for all test instances (see Column 5). However, as Column 6 shows, the multi-stage models for larger instances are difficult to solve within the time limit. This implies that it may be worthwhile to approximate the multi-stage model by a PA model, which leads to the answer to the second question. Second, as μ increases, the performance of the $\Pi_{\text{PA}}(\mu)$ model improves quite significantly, especially for instances with larger horizon length. For example, for the $T = 9$ instance, solving $\mu = 4$ obtains a gap between PA and multi-stage models in the range of 17.11-21.84%, and solving $\mu = 7$ further shrinks the gap to no more than 7.23%. For the $T = 10$ instance, the $\Pi_{\text{PA}}(\mu)$ model with $\mu = 5$ obtains a gap no more than 18.50% higher than the multi-stage model. Furthermore, for all the test instances, we observe that the PA model with a mid-range value of μ already decreases the gap of the two-stage model ($\mu = 1$) by more than 50%. This suggests the benefit of solving PA models with small μ values. Indeed, from Column 6, we can see the PA models with large μ including the multi-stage model are computationally challenging to solve. Thus, recursively traversing the tree by solving PA models with small μ may provide a better multi-stage solution within a reasonable time limit. This is demonstrated by the next set of experiments.

5.3 Performance of Algorithm 1

This second set of experiments evaluate Algorithm 1 in the following way. On a T-year scenario tree $\mathcal{T}$, Algorithm 1 solves all the PA models $\Pi_{\text{PA}}(\mu, n)$ with $\mu = 2$ for the subtree $\mathcal{T}(n)$, for each node n up to level T_0 in the tree $\mathcal{T}$, where $T_0 < T$. Denote the optimal solution of $\Pi_{\text{PA}}(\mu, n)$ as $\mathbf{z}^{\mu,n} := \{\mathbf{x}_m^{\mu,n}, \{\mathbf{y}_{mk}^{\mu,n}\}_{k \in \mathcal{K}_m}\}_{m \in \mathcal{T}(n)}$. The final solution $\{\mathbf{z}_n\}_{n \in \mathcal{T}}$ generated by Algorithm 1 is obtained as follows: $\mathbf{z}_n := \mathbf{z}_n^{\mu,n}$, for each node n up to level T_0, i.e. $t_n \leq T_0$, and $\mathbf{z}_n := \mathbf{z}_n^{\mu,m}$ for each node n with $t_n \geq T_0 + 1$, where $m = \mathcal{P}(n) \cap \mathcal{S}_{T_0}$. That is, the output solution $\mathbf{z}$ has a multi-stage expansion plan up to period $T_0 + 1$, and if $T_0 \leq T - 2$, $\mathbf{z}$ has a two-stage expansion plan from level $T_0 + 2$ to the end of the planning horizon T. In other words, $\mathbf{z}$ has the same decision structure in capacity expansion plan as the solution of a PA model $\Pi_{\text{PA}}(\mu, 1)$ for $\mu = T_0 + 1$, solved at root node 1.

Compared to directly solving the $\Pi_{\text{PA}}(T_0 + 1, 1)$ model, Algorithm 1 solves a sequence of much smaller problems of $\Pi_{\text{PA}}(2, n)$, therefore, may significantly save computation time. Also, the $\Pi_{\text{PA}}(T_0 + 1, 1)$ model may not be solvable to optimality within the time limit, whereas $\Pi_{\text{PA}}(2, n)$ can usually be solved to high precision quickly. In fact, we have already solved the $\Pi_{\text{PA}}(T_0 + 1, 1)$ model for various T and T_0 in the previous experiments. Most of these models for $T \geq 5$ and $T_0 \geq 2$ cannot be solved within the time limit of 5 hours as shown in Table 2.

Of course, we also need to evaluate how good the resulting solution $\mathbf{z}$ is compared to an optimal (or the best found) solution of $\Pi_{\text{PA}}(T_0 + 1, 1)$. Table 2 Column 3 presents the best lower bounds on the optimal costs of $\Pi_{\text{PA}}(T_0 + 1, 1)$ and the best feasible solution found. Using the notation in the previous subsection, these costs are denoted as $v_L(T_0 + 1)$ and $v_H(T_0 + 1)$, respectively. The expected cost of the $\mathbf{z}$ solution is given as $v(\mathbf{z}) := \sum_{n \in \mathcal{T}} p_n(\mathbf{a}_n^\top \mathbf{x}_n + \sum_{k \in \mathcal{K}_{t_n}} \mathbf{b}_{nk}^\top \mathbf{y}_{nk})$, where $\mathbf{x}$ and $\mathbf{y}$ are the capacity expansion and

generation production components of $\mathbf{z}$ (cf. objective function (2.2a)). For each T and T_0, define OptGap $:= (v(\mathbf{z}) - v_L(T_0+1))/v_L(T_0+1)$, which gives an upper bound on the gap between the Algorithm 1's solution $\mathbf{z}$ and the optimal $\Pi_{\text{PA}}(T_0+1, 1)$ solution. Also define ImprvGap $:= (v_H(T_0+1) - v(\mathbf{z}))/v_L(T_0+1)$, which measures how much the $\mathbf{z}$ solution improves on the best solution found by directly solving $\Pi_{\text{PA}}(T_0+1, 1)$ (negative value means $\mathbf{z}$ solution is worse).

Table 3: Computation Study on Algorithm 1

T	T_0	$v(\mathbf{z})$ (million \$)	OptGap (%)	ImprvGap (%)	PA time (sec.)	Alg 1 Time (sec.)
3	2	2138.57	0.28	0.00	0.08	0.35
4	2	3451.26	0.44	0.06	5.45	2.88
	3	3200.11	0.50	0.00	4.11	3.63
5	2	5073.68	0.87	−0.02	18004.46	27.27
	4	4353.10	0.68	0.07	18002.52	32.54
6	3	6255.76	0.83	0.18	18003.85	106.19
	5	5621.10	1.18	0.17	18003.95	119.86
7	3	8081.19	0.59	0.21	18010.35	360.23
	6	6987.25	1.05	0.37	18002.95	519.12
8	4	9463.10	0.66	0.50	18000.91	1324.39
	7	8440.01	1.10	0.51	18000.30	1833.46
9	3	11923.71	0.62	0.15	18001.53	1749.95
	6	10465.37	0.94	0.42	18017.75	6921.59
	8	9951.13	1.53	1.71	18007.46	7141.51
10	4	13302.09	0.58	0.63	18020.00	6838.51
	7	11934.25	1.19	1.64	18004.19	14771.03
	9	11462.33	1.48	3.66	18083.68	15446.97

Table 3 clearly shows that Algorithm 1 significantly reduces the computation time compared to directly solving the PA model (see Columns 6 and 7). Also as shown in Column 4 "OptGap", Algorithm 1 is able to produce a $\mathbf{z}$ solution that is within at most 1.5% from the optimal $\Pi_{\text{PA}}(T_0+1, 1)$ solution. Furthermore, as shown in Column 5 "ImprvGap", the $\mathbf{z}$ solution constructed by Algorithm 1 actually improves over the best feasible solution found by directly solving the $\Pi_{\text{PA}}(T_0+1, 1)$ for almost all instances, except for $T = 5, T_0 = 2$ where the $\mathbf{z}$ solution is slightly worse by 0.02%. In fact, the improvement steadily increases as T and T_0 increase. For the 9-period problem, Algorithm 1 stopping at $T_0 = 8$ improves over the direct method by 1.71% in expected cost, while the computation time is only 40% of the latter. For the 10-period problem, Algorithm 1 stopping at $T_0 = 9$ improves over the direct method by 3.66% in expected cost with computation time reduced by 15%. Also note that for these two instances, the $\Pi_{\text{PA}}(T_0+1, 1)$ model is the full multi-stage model. Therefore, the last instance shows that, by recursively traversing the tree with solving small PA models, the resulting solution is within 1.48% from the true optimal full multi-stage solution for the 10-year planning problem. The computation is done within 4.3 hours.

5.4 Effect of different node orderings

As alluded to in Section 4.1, Algorithm 1 works for any topological ordering chosen for the nodes of the scenario tree. We distinguish four typical orderings, corresponding to breath-first-search (BFS) and depth-

first-search (DFS) on the tree. For the simplicity of exposition, suppose for every node $n \in \mathcal{T}$, all its children nodes are positioned from top to bottom in the order of increasing demand. This is possible because the scenario tree is generated in a way that no two nodes sharing the same parent node have the same demand. Figure 3 illustrates four types of typological orderings on a simple example.

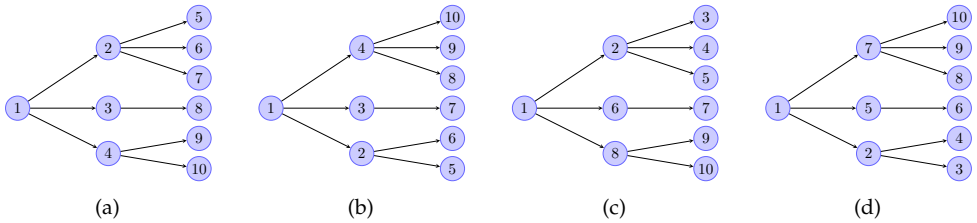

Figure 3: Four topological orderings on a 3-period tree
Corresponding search strategies: (a) `BFS_LOW`; (b) `BFS_HIGH`; (c) `DFS_LOW`; (d) `DFS_HIGH`.

In particular, we have the following four orderings and their corresponding search strategies:

1. `BFS_LOW`: breath-first-search on the tree, and within each level, nodes are visited from top to bottom;
2. `BFS_HIGH`: breath-first-search, and within each level, nodes are visited from bottom to top;
3. `DFS_LOW`: depth-first-search, and selects the child node with *lowest* demand;
4. `DFS_HIGH`: depth-first-search, and selects the child node with *highest* demand.

All the experiments in Section 5.3 are conducted with the `BFS_LOW` ordering. It is interesting to see the performance of Algorithm 1 under different node orderings, because traversing the scenario tree in different orders may reveal how fast the solution improves as we explore the node sequence, which may also shed some light on which nodes are important to explore, thus helps the decision maker decide whether to refine the current solution by branching through a specific node.

Figure 4 contains the cost improvement curves in an 8-year GEP instance under four different nodes orderings: `BFS_LOW`, `BFS_HIGH`, `DFS_LOW`, and `DFS_HIGH`. The upper-left figure shows how cost of the $\mathbf{z}$ solution is improved as allowed computation time limit increases. The other four figures correspond to the cost improvements as the number of visited nodes increases in these four orderings. Indeed, there is one subproblem associated with each node of the tree. The more subproblems Algorithm 1 solves, the lower cost the $\mathbf{z}$ solution incurs.

Figure 4 shows that, if there is no limit on computation time, i.e. Algorithm 1 is allowed to run until traversing all nodes in the scenario tree, then, as expected, there is no difference in the final policies corresponding to the above four different node orderings. The total computation times required are also almost the same. This is not surprising since all nodes are explored in a topological ordering, and the sets of subproblems solved are identical under each of these four orderings. The only difference is the sequence of solving these subproblems. However, if computation time is limited, we notice from Figure 4 that `DFS_HIGH` produces a better solution than the other three. This suggests that refining the solution on the subtree corresponding to higher demands (as well as higher natural gas prices) improves the solution value faster. Numerical results suggest that as the solution gets refined along this subtree, capacity expansion decisions are postponed. While generation cost is also higher in later periods, the savings of postponing capacity expansion dominates the increment in generation cost.

In the other four figures, we observe that subproblems in earlier periods take more time since they contains more variables and constraints. In addition, subproblems on some nodes take more time or reduce

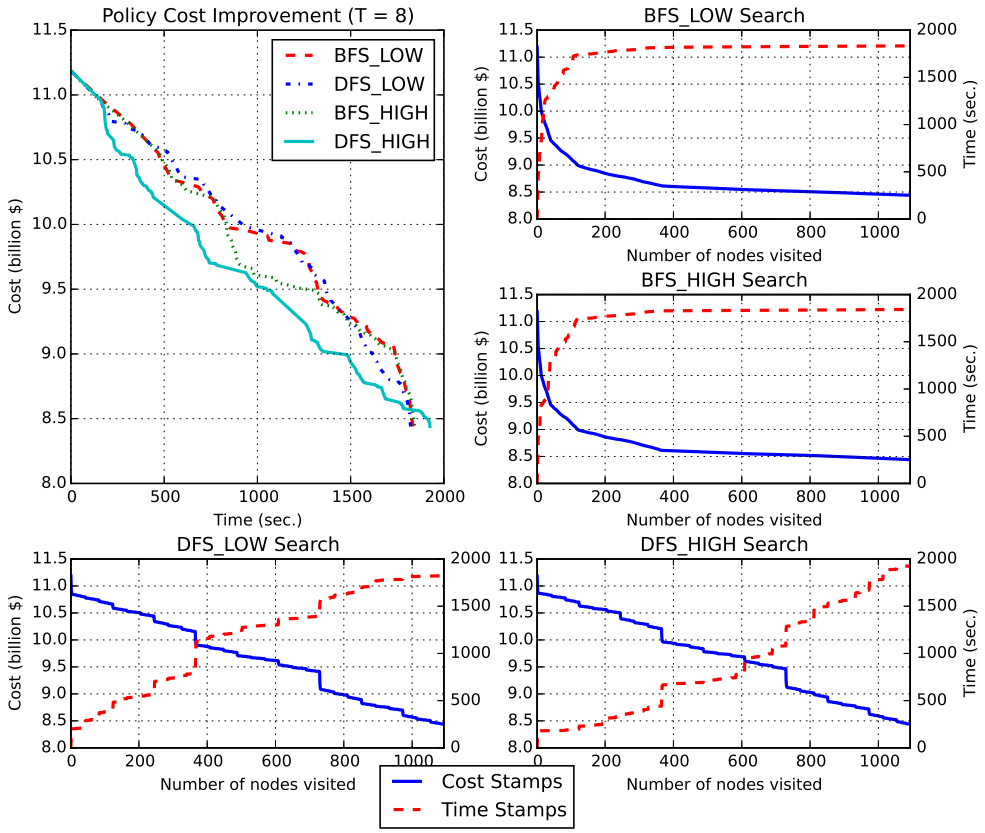

Figure 4: Solution cost improvement under four different orderings.

the total expected cost in a larger scale than others, and these two effects usually show up as a pair. These nodes correspond to the ones in early periods of the scenario tree. However, there do exist some nodes whose subproblems take a long time to solve but do not substantially improve the total cost. In other words, the subproblems on such nodes do not contribute to a significant improvement to the overall solution. It could be interesting if the decision maker can identify such nodes beforehand and skip over them when implementing Algorithm 1. It could also be interesting to see how Algorithm 1 performs on a scenario tree with more branches at each node. We left these issues for future research.

6 Concluding remarks

We consider a long-term power generation expansion planning problem and propose a new framework of partially adaptive stochastic mixed integer optimization models for the GEP problem and generic capacity expansion planning problem. Our model unifies the two-stage and multi-stage approaches, and provides the decision maker with the flexibility to adjust the adaptivity of the capacity expansion decisions with respect to uncertainty realizations. Since an optimal solution to the PA model is always feasible to the multi-stage model, we present nontrivial bounds for the gap between these two models. Furthermore, we propose an approximation algorithm which recursively solves a sequence of PA models and returns a

feasible multi-stage solution. We identify a set of sufficient conditions under which the algorithm produces an optimal multi-stage solution.

We conduct extensive computational experiments on the PA models and the proposed recursive algorithm on a realistic scale generation expansion planning problem. Numerical results show that PA model provides significant value to the long-term generation expansion problem. It considerably reduces the expected total cost of the GEP problem comparing to the traditional two-stage model. Computation also shows that, with a small amount of flexibility in the expansion decision (i.e. small μ in the $\Pi_{\mathrm{PA}}(\mu)$ model), the PA model can approximate the multi-stage model fairly well. Computational experiments further shows that it is not necessary to directly solve the PA model, but rather recursively traversing the scenario tree and solving a sequence of small sized PA models can produce near-optimal solution with a much reduced computation time. We also explore the impact of different search orderings on the performance of the algorithm and its implications.

References

Shabbir Ahmed and Nikolaos V. Sahinidis. An approximation scheme for stochastic integer programs arising in capacity expansion. *Operations Research*, 51(3):461–471, 2003.

Shabbir Ahmed, Alan J. King, and Gyana Parija. A multi-stage stochastic integer programming approach for capacity expansion under uncertainty. *Journal of Global Optimization*, 26(1):3–24, 2003.

James C. Bean, Julia L. Higle, and Robert L. Smith. Capacity expansion under stochastic demands. *Operations Research*, 40(3-supplement-2):S210–S216, 1992.

Oded Berman, Zvi Ganz, and Janet M. Wagner. A stochastic optimization model for planning capacity expansion in a service industry under uncertain demand. *Naval Research Logistics (NRL)*, 41(4):545–564, 1994.

Daniel Bienstock and Jeremy F. Shapiro. Optimizing resource acquisition decisions by stochastic programming. *Management Science*, 34(2):215–229, 1988.

Jeremy A Bloom. Solving an electricity generating capacity expansion planning problem by generalized benders' decomposition. *Operations Research*, 31(1):84–100, 1983.

Jeremy A Bloom, Michael Caramanis, and Leonid Charny. Long-range generation planning using generalized benders' decomposition: Implementation and experience. *Operations Research*, 32(2):290–313, 1984.

Zhi-Long Chen, Shanling Li, and Devanath Tirupati. A scenario-based stochastic programming approach for technology and capacity planning. *Computers & Operations Research*, 29(7):781–806, 2002.

M. H. A. Davis, M. A. H. Dempster, S. P. Sethi, and D. Vermes. Optimal capacity expansion under uncertainty. *Advances in Applied Probability*, pages 156–176, 1987.

MAH Dempster, N Hicks Pedron, EA Medova, JE Scott, and A Sembos. Planning logistics operations in the oil industry. *Journal of the Operational Research Society*, pages 1271–1288, 2000.

Jitka Dupačová, Marida Bertocchi, and Vittorio Moriggia. Testing the structure of multistage stochastic programs. *Computational Management Science*, 6(2):161–185, 2009.

Donald Erlenkotter. Optimal plant size with time-phased imports. *Investments for Capacity Expansion: Size, Location, and Time-Phasing*, 5:157, 1967.

John Freidenfelds. Capacity expansion when demand is a birth-death random process. *Operations Research*, 28(3-part-ii):712–721, 1980.

Richard J. Giglio. Stochastic capacity models. *Management Science*, 17(3):174–184, 1970.

Richard C Grinold. Infinite horizon stochastic programs. *SIAM journal on control and optimization*, 24(6): 1246–1260, 1986.

Kjetil Høyland and Stein W. Wallace. Generating scenario trees for multistage decision problems. *Management Science*, 47(2):295–307, 2001.

Kai Huang and Shabbir Ahmed. The value of multistage stochastic programming in capacity planning under uncertainty. *Operations Research*, 57(4):893–904, 2009.

Shan Jin, Sarah M. Ryan, Jean-Paul Watson, and David L. Woodruff. Modeling and solving a large-scale generation expansion planning problem under uncertainty. *Energy Systems*, 2(3-4):209–242, 2011.

Alan S. Manne. Capacity expansion and probabilistic growth. *Econometrica: Journal of the Econometric Society*, pages 632–649, 1961.

Soren S Nielsen and Stavros A Zenios. A stochastic programming model for funding single premium deferred annuities. *Mathematical Programming*, 75(2):177–200, 1996.

Andrzej Ruszczynski and Alexander Shapiro. Stochastic programming. *Handbooks in Operations Research and Management Science*, 10:1–682, 2003.

Sarah M. Ryan, James McCalley, and David L. Woodruff. Long term resource planning for electric power systems under uncertainty. *Computational Needs for the Next Generation Electric Grid*, pages 6–1–6–41, 2011.

Kavinesh J. Singh, Andy B. Philpott, and R. Kevin Wood. Dantzig-wolfe decomposition for solving multistage stochastic capacity-planning problems. *Operations Research*, 57(5):1271–1286, 2009.

Stein W. Wallace and Stein-Erik Fleten. Stochastic programming models in energy. *Handbooks in Operations Research and Management Science*, 10:637–677, 2003.

Online Supplement[1]

Proof of Proposition 1

PROPOSITION 1. *In problem* (10)*, the left-hand-side coefficient matrix is totally unimodular.*

Proof of Proposition 1. Let us denote the left-hand-side coefficient matrix corresponding to constraints (10b) by $\mathbf{D}$. The left-hand-side coefficient matrix corresponding to constraints (10c) are the same as some rows in $-\mathbf{D}$. In fact, they correspond to each scenario (path) in the tree $\hat{\mathcal{T}}$. Therefore, it is sufficient to show that $\mathbf{D}$ is totally unimodular. We know that every entry of $\mathbf{D}$ is either 0 or 1. For each column j, there are exactly $|\hat{\mathcal{T}}(j)|$ 1's, in particular, $\mathbf{D}_{ij} = 1$ if $i \in \hat{\mathcal{T}}(j)$. We can traverse the tree by depth-first-search, and rearrange the rows according to the sequence. After rearrangement, $\mathbf{D}$ is an interval matrix hence is totally unimodular (cf. Schrijver 1998). $\square$

Proof of Proposition 2

PROPOSITION 2. *If $\hat{p}_n \hat{a}_{in} > 0$ for all $n \in \hat{\mathcal{T}}$ and $i \in \mathcal{I}$, then in the LP relaxation of problem* (10)*, constraints* (10c) *are redundant.*

Proof of Proposition 2. It is sufficient to show that any optimal solution to the problem

$$\min \left\{ \sum_{n \in \hat{\mathcal{T}}} \hat{p}_n \hat{a}_{in} x_{in} \quad \text{s.t.} \sum_{m \in \mathcal{P}(n)} x_{im} \geq \lceil \hat{\delta}_{in} \rceil, x_{in} \geq 0, \forall n \in \hat{\mathcal{T}} \right\} \tag{$*$}$$

satisfies constraints (10c). Let $\tilde{\mathbf{x}}_i$ be an optimal solution to $(*)$. Suppose there exists $n_0 \in \hat{\mathcal{T}}$ such that $\sum_{m \in \mathcal{P}(n_0)} \tilde{x}_{im} = w > u_i$. Recall that $u_i \geq \lceil \hat{\delta}_{in} \rceil$ for all $n \in \hat{\mathcal{T}}$, it follows that $\sum_{m \in \mathcal{P}(n_0)} \tilde{x}_{im} > \lceil \hat{\delta}_{in} \rceil$ for all $n \in \hat{\mathcal{T}}(n_0)$. If $\tilde{x}_{in_0} > 0$, since $\hat{p}_{n_0} \hat{a}_{in_0} > 0$, by optimality of $\tilde{\mathbf{x}}_i$, we know there must be some $n_0' \in \hat{\mathcal{T}}(n_0)$, such that $\lceil \hat{\delta}_{in_0'} \rceil = w > u_i$,

[1] All the reference numbers match those in the main paper.

which contradicts to the fact that $u_i \geq \lceil \hat{\delta}_{in} \rceil$ for all $n \in \hat{\mathcal{T}}$. If $\tilde{x}_{in_0} = 0$, we traverse back along $\mathcal{P}(n_0)$ to find the first node n_0'' with $\tilde{x}_{in_0''} > 0$. Notice such a node must exist since $\sum_{m \in \mathcal{P}(n_0)} \tilde{x}_{im} > 0$. It follows that $\sum_{m \in \mathcal{P}(n_0'')} \tilde{x}_{im} = w > u_i$. We go back to the first case. Therefore, the result holds. $\square$

Proof of Proposition 3

PROPOSITION 3. *$o_i^P(\mu) \leq v_i^P(\mu) + a_{i1} \cdot \lambda_i$, where $\lambda_i = \max_{n \in \mathcal{T}}\{\lceil \delta_{in} \rceil - \delta_{in}\}$. If $\{\delta_{in}\}_{n \in \mathcal{T}}$ are all integers, the inequality is tight.*

Proof of Proposition 3. In fact, with linear programming duality, Proposition 1 and 2, we have

$$
\begin{aligned}
o_i^{\mathrm{P}}(\mu) &= \min \left\{ \sum_{n \in \hat{\mathcal{T}}} \hat{p}_n \hat{a}_{in} x_{in} \quad \text{s.t.} \sum_{m \in \mathcal{P}(n)} x_{im} \geq \hat{\delta}_{in}, \sum_{m \in \mathcal{P}(n)} x_{im} \leq u_i,\ x_{in} \in \mathbb{Z}_+,\ \forall n \in \hat{\mathcal{T}} \right\} \\
&\overset{\text{(i)}}{=} \min \left\{ \sum_{n \in \hat{\mathcal{T}}} \hat{p}_n \hat{a}_{in} x_{in} \quad \text{s.t.} \sum_{m \in \mathcal{P}(n)} x_{im} \geq \lceil \hat{\delta}_{in} \rceil, \sum_{m \in \mathcal{P}(n)} x_{im} \leq u_i,\ x_{in} \in \mathbb{R}_+,\ \forall n \in \hat{\mathcal{T}} \right\} \\
&\overset{\text{(ii)}}{=} \min \left\{ \sum_{n \in \hat{\mathcal{T}}} \hat{p}_n \hat{a}_{in} x_{in} \quad \text{s.t.} \sum_{m \in \mathcal{P}(n)} x_{im} \geq \lceil \hat{\delta}_{in} \rceil,\ x_{in} \in \mathbb{R}_+,\ \forall n \in \hat{\mathcal{T}} \right\} \\
&\overset{\text{(iii)}}{=} \max \left\{ \sum_{n \in \hat{\mathcal{T}}} \lceil \hat{\delta}_{in} \rceil \pi_{in} \quad \text{s.t.} \sum_{m \in \hat{\mathcal{T}}(n)} \pi_{im} \leq \hat{p}_n \hat{a}_n,\ \pi_{in} \in \mathbb{R}_+,\ \forall n \in \hat{\mathcal{T}} \right\} \\
&= \max \left\{ \sum_{n \in \hat{\mathcal{T}}} (\hat{\delta}_{in} + \lceil \hat{\delta}_{in} \rceil - \hat{\delta}_{in}) \pi_{in} \quad \text{s.t.} \sum_{m \in \hat{\mathcal{T}}(n)} \pi_{im} \leq \hat{p}_n \hat{a}_n,\ \pi_{in} \in \mathbb{R}_+,\ \forall n \in \hat{\mathcal{T}} \right\} \\
&\leq \max \left\{ \sum_{n \in \hat{\mathcal{T}}} \hat{\delta}_{in} \pi_{in} \quad \text{s.t.} \sum_{m \in \hat{\mathcal{T}}(n)} \pi_{im} \leq \hat{p}_n \hat{a}_n,\ \pi_{in} \in \mathbb{R}_+,\ \forall n \in \hat{\mathcal{T}} \right\} \\
&\quad + \max \left\{ \sum_{n \in \hat{\mathcal{T}}} \pi_{in} \quad \text{s.t.} \sum_{m \in \hat{\mathcal{T}}(n)} \pi_{im} \leq \hat{p}_n \hat{a}_n,\ \pi_{in} \in \mathbb{R}_+,\ \forall n \in \hat{\mathcal{T}} \right\} \cdot \max_{n \in \mathcal{T}} \{ \lceil \hat{\delta}_{in} \rceil - \hat{\delta}_{in} \} \\
&\overset{\text{(iv)}}{=} \min \left\{ \sum_{n \in \hat{\mathcal{T}}} \hat{p}_n \hat{a}_{in} x_{in} \quad \text{s.t.} \sum_{m \in \mathcal{P}(n)} x_{im} \geq \hat{\delta}_{in},\ x_{in} \in \mathbb{R}_+,\ \forall n \in \hat{\mathcal{T}} \right\}
\end{aligned}
$$

$$+\min\left\{\sum_{n\in\hat{\mathcal{T}}}\hat{p}_n\hat{a}_{in}x_{in}\quad \text{s.t.}\sum_{m\in\mathcal{P}(n)}x_{im}\geq 1,\ x_{in}\in\mathbb{R}_+,\ \forall n\in\hat{\mathcal{T}}\right\}\cdot\max_{n\in\mathcal{T}}\{\lceil\hat{\delta}_{in}\rceil-\hat{\delta}_{in}\}$$
$$\overset{\text{(v)}}{\leq} v_i^{\mathrm{P}}(\mu)+a_{i1}\cdot\lambda_i.$$

Specifically, (i) follows from Proposition 1; (ii) follows from Proposition 2; (iii) and (iv) follow from linear programming duality; (v) follows from Proposition 2, the definition of $\{\hat{\delta}_{in}\}_{n\in\hat{T}}$, the fact that $p_1=1$, and the optimal solution to a single-technology GEP problem with demand 1 at each node is nothing but building one generator at the beginning of the planning horizon. The tightness of the inequality follows from Proposition 1. □

Proof of Proposition 4

Proposition 4. *For any $\mu\in\{1,\ldots,T\}$, the following relation holds, $\delta^{(\mu-)}\leq\delta^{(T)}\leq\delta^{(\mu)}\leq\delta^{(1)}$.*

Proof of Proposition 4. Recall that in a scenario tree, the probability associated with a node equals the sums of probabilities of its children nodes. By definition, we have

$$\begin{aligned}
\delta^{(\mu-)} &= \sum_{n\in\mathcal{S}_{\mu-1}} p_n \max_{m\in\mathcal{P}(n)}\{\delta_m\}\\
&= \sum_{n\in\mathcal{S}_{\mu-1}}\left(\sum_{k\in\mathcal{S}_T\cap\mathcal{T}(n)} p_k \max_{m\in\mathcal{P}(n)}\{\delta_m\}\right)\leq\sum_{n\in\mathcal{S}_{\mu-1}}\left(\sum_{k\in\mathcal{S}_T\cap\mathcal{T}(n)} p_k \max_{m\in\mathcal{P}(k)}\{\delta_m\}\right)\\
(\delta^{(T)}) &= \sum_{k\in\mathcal{S}_T} p_k \max_{m\in\mathcal{P}(k)}\{\delta_m\}\\
&= \sum_{n\in\mathcal{S}_\mu}\sum_{k\in\mathcal{S}_T\cap\mathcal{T}(n)} p_k \max_{m\in\mathcal{P}(k)}\{\delta_m\}\leq\sum_{n\in\mathcal{S}_\mu}\left(\sum_{k\in\mathcal{S}_T\cap\mathcal{T}(n)} p_k\right)\max_{m\in\mathcal{P}(n)\cup\mathcal{T}(n)}\{\delta_m\}\\
(\delta^{(\mu)}) &= \sum_{n\in\mathcal{S}_\mu} p_n \max_{m\in\mathcal{P}(n)\cup\mathcal{T}(n)}\{\delta_m\}\leq\sum_{n\in S_\mu} p_n\delta^{(1)}=\delta^{(1)}.\quad\square
\end{aligned}$$

Proof of Theorem 1

THEOREM 1. $(\underline{a}_{\mu-} - a_*)\delta^{(\mu-)} + a_*\delta^{(\mu)} \le v^P(\mu) \le (\bar{a}_{\mu-} - \bar{a}_{\mu+})\delta^{(\mu-)} + \bar{a}_{\mu+}\delta^{(\mu)}$.

Proof of Theorem 1. We change the notation of decision variables for capacity expansion decisions starting from period μ into a different representation. In particular, for any $n \in \mathcal{S}_\mu$ and $t > \mu$, $\{x_m : m \in \mathcal{T}(n) \cap \mathcal{S}_t\}$ share the same value, let $x_{n,t}$ denote the new variable that represents the common value of these variables. Given a feasible solution $\mathbf{x}$ to the LP relaxation of problem (7), for any $n \in S_{\mu-1}$, by feasibility we have

$$\begin{aligned}
\sum_{m\in\mathcal{P}(n)} x_m \ge \max_{m\in\mathcal{P}(n)}\{\delta_m\} \Rightarrow \sum_{n\in\mathcal{S}_{\mu-1}} p_n \sum_{m\in\mathcal{P}(n)} x_m &\ge \sum_{n\in\mathcal{S}_{\mu-1}} p_n \max_{m\in\mathcal{P}(n)}\{\delta_m\} \\
&\Leftrightarrow \sum_{t=1}^{\mu-1}\sum_{k\in\mathcal{S}_t}\left(\sum_{m\in S_{\mu-1}\cap\mathcal{T}(k)} p_m\right) x_k \ge \delta_{\mu-} \\
&\Leftrightarrow \sum_{t=1}^{\mu-1}\sum_{k\in\mathcal{S}_t} p_k x_k \ge \delta_{\mu-},
\end{aligned}$$

where the first equivalence follows from changing the summation sequence; and the second equivalence follows from the fact that $\sum_{m\in S_\mu\cap\mathcal{T}(k)} p_m = p_k$ for all $k \in \mathcal{T}$ such that $t_k < \mu$. In addition, for any $n \in \mathcal{S}_\mu$, by feasibility we have

$$\sum_{m\in\mathcal{P}(a(n))} x_m + \sum_{t=\mu}^{T} x_{n,t} \ge \max_{\mathcal{P}(n)\cup\mathcal{T}(n)}\{\delta_m\} \quad\Leftrightarrow\quad \sum_{t=\mu}^{T} x_{n,t} \ge \max_{\mathcal{P}(n)\cup\mathcal{T}(n)}\{\delta_m\} - \sum_{m\in\mathcal{P}(a(n))} x_m.$$

Then if $\mathbf{x}^*$ is an optimal solution to the LP relaxation of problem (7), we have

$$\begin{aligned}
v^{\mathrm{P}}(\mu) = \sum_{n\in\mathcal{T}} p_n a_n x_n^* &= \sum_{t=1}^{\mu-1}\sum_{n\in\mathcal{S}_t} p_n a_n x_n^* + \sum_{t=\mu}^{T}\sum_{n\in\mathcal{S}_t} p_n a_n x_n^* \\
&\ge \underline{a}_{\mu-}\sum_{t=1}^{\mu-1}\sum_{n\in\mathcal{S}_t} p_n x_n^* + a_*\sum_{t=\mu}^{T}\sum_{n\in\mathcal{S}_t} p_n x_n^* \\
&= \underline{a}_{\mu-}\sum_{t=1}^{\mu-1}\sum_{n\in\mathcal{S}_t} p_n x_n^* + a_*\sum_{n\in\mathcal{S}_\mu} p_n \sum_{t=\mu}^{T} x_{n,t}^*
\end{aligned}$$

$$
\begin{aligned}
&\geq \underline{a}_{\mu-} \sum_{t=1}^{\mu-1} \sum_{n\in\mathcal{S}_t} p_n x_n^* + a_* \sum_{n\in\mathcal{S}_\mu} p_n \left(\max_{\mathcal{P}(n)\cup\mathcal{T}(n)} \{\delta_m\} - \sum_{m\in\mathcal{P}(a(n))} x_m^* \right) \\
&= (\underline{a}_{\mu-} - a_*) \sum_{t=1}^{\mu-1} \sum_{n\in\mathcal{S}_t} p_n x_n^* + a_* \sum_{n\in S_\mu} p_n \max_{\mathcal{P}(n)\cup\mathcal{T}(n)} \{\delta_m\} \\
&\geq (\underline{a}_{\mu-} - a_*)\delta^{(\mu-)} + a_*\delta^{(\mu)},
\end{aligned}
$$

where the last inequality follows from $\underline{a}_{\mu-} \geq a_*$ and the definitions of $\delta^{(\mu)}$ and $\delta^{(\mu-)}$.

Next, we consider a feasible solution $\hat{\mathbf{x}}$ to the LP relaxation of problem (7). For any $n \in \mathcal{T}$ such that $t_n \leq \mu - 1$, let $\hat{x}_n = \max\{\delta_m : m \in \mathcal{P}(n)\} - \max\{\delta_m : m \in \mathcal{P}(a(n))\}$, and $\max\{\delta_m : m \in \mathcal{P}(a(1))\} = 0$; for any $n \in S_\mu$, $t \geq \mu$, let $\hat{x}_{n,t} = \max\{\delta_m : m \in \mathcal{P}(a(n)) \cup \mathcal{T}(n,t)\} - \max\{\delta_m : m \in \mathcal{P}(a(n)) \cup \mathcal{T}(n,t-1)\}$. Then we have

$$
\begin{aligned}
v^{\mathrm{P}}(\mu) \leq \sum_{n\in\mathcal{T}} p_n a_n \hat{x}_n &= \sum_{t=1}^{\mu-1} \sum_{n\in\mathcal{S}_t} p_n a_n \hat{x}_n + \sum_{t=\mu}^{T} \sum_{n\in\mathcal{S}_t} p_n a_n \hat{x}_n \\
&\leq \bar{a}_{\mu-} \sum_{t=1}^{\mu-1} \sum_{n\in\mathcal{S}_t} p_n \hat{x}_n + \bar{a}_{\mu+} \sum_{t=\mu}^{T} \sum_{n\in\mathcal{S}_t} p_n \hat{x}_n \\
&= \bar{a}_{\mu-} \sum_{t=1}^{\mu-1} \sum_{n\in\mathcal{S}_t} p_n \hat{x}_n + \bar{a}_{\mu+} \sum_{n\in\mathcal{S}_\mu} p_n \sum_{t=\mu}^{T} \hat{x}_{n,t} \\
&= \bar{a}_{\mu-} \sum_{t=1}^{\mu-1} \sum_{n\in\mathcal{S}_t} p_n \left(\max_{m\in\mathcal{P}(n)} \{\delta_m\} - \max_{m\in\mathcal{P}(a(n))} \{\delta_m\} \right) \\
&\quad + \bar{a}_{\mu+} \sum_{n\in\mathcal{S}_\mu} p_n \sum_{t=\mu}^{T} \left(\max_{m\in\mathcal{P}(a(n))\cup\mathcal{T}(n,t)} \{\delta_m\} - \max_{m\in\mathcal{P}(a(n))\cup\mathcal{T}(n,t-1)} \{\delta_m\} \right) \\
&= \bar{a}_{\mu-} \sum_{n\in\mathcal{S}_{\mu-1}} p_n \max_{m\in\mathcal{P}(n)} \{\delta_m\} + \bar{a}_{\mu+} \sum_{n\in\mathcal{S}_\mu} p_n \left(\max_{m\in\mathcal{P}(a(n))\cup\mathcal{T}(n)} \{\delta_m\} - \max_{m\in\mathcal{P}(a(n))} \{\delta_m\} \right) \\
&= (\bar{a}_{\mu-} - \bar{a}_{\mu+}) \sum_{n\in\mathcal{S}_{\mu-1}} p_n \max_{m\in\mathcal{P}(n)} \{\delta_m\} + \bar{a}_{\mu+} \sum_{n\in S_\mu} p_n \max_{m\in\mathcal{P}(n)\cup\mathcal{T}(n)} \{\delta_m\} \\
&= (\bar{a}_{\mu-} - \bar{a}_{\mu+})\, \delta^{(\mu-)} + \bar{a}_{\mu+}\delta^{(\mu)},
\end{aligned}
$$

where the third to last equality follows from the fact that the probability of node n equals to the sum of probabilities of its children nodes. □

Proof of Theorem 5

Theorem 5 *For a multi-stage generation expansion planning problem* (4)*, if the instance satisfies the following conditions:*

i) the unit investment costs of each type of generators are stationary, i.e., in (1a)*,* $c_{it_n} = c_i$ *for all* $i \in \mathcal{I}$ *and* $n \in \mathcal{T}$*;*

ii) all generators share the same unit generation cost, i.e., in (1a)*,* $b_{ink} = b_{nk}$ *for all* $k \in \mathcal{K}_{t_n}$ *and* $n \in \mathcal{T}$*;*

iii) demand must be satisfied by generation, i.e., no penalty is allowed;

then the solution returned by Algorithm 1 is optimal to the multi-stage problem.

To prove Theorem 5, we need the following lemma.

Lemma 1. *Suppose condition (i) in Theorem 5 is satisfied. Let* $\{\mathbf{x}^*, \mathbf{y}^*\}$ *be an optimal solution to problem* (2)*, then* $x^*_{i1} = \max\{\lceil [\mathbf{A}_{1k}\mathbf{y}^*_{1k}]_i \rceil : k \in \mathcal{K}_1\}$ *for all* $i \in \mathcal{I}$*.*

Proof of Lemma 1. Suppose there exists $i_0 \in \mathcal{I}$ such that $x^*_{i_0 1} > \max\{\lceil [\mathbf{A}_{1k}\mathbf{y}^*_{1k}]_{i_0} \rceil : k \in \mathcal{K}_1\}$. Since $x^*_{i_0 1} \in \mathbb{Z}_+$, $x^*_{i_0 1} \geq \max\{\lceil [\mathbf{A}_{1k}\mathbf{y}^*_{1k}]_{i_0} \rceil : k \in \mathcal{K}_1\} + 1$. Let $\tilde{x}_{i_0 1} = x^*_{i_0 1} - 1$, $\tilde{x}_{i_0 n} = x^*_{i_0 n} + 1$ for all $n \in C(1)$, $\tilde{x}_{i1} = x^*_{i1}$ for all $i \neq i_0$, and $\tilde{\mathbf{x}}_n = \mathbf{x}^*_n$ for n such that $t_n \geq 3$, $\tilde{\mathbf{y}}_n = \mathbf{y}^*_n$ for $n \in \mathcal{T}$. It is clear that $\{\tilde{\mathbf{x}}, \tilde{\mathbf{y}}\}$ is still feasible, but it changes the total cost by $-c_{i_0} + \sum_{n \in C(1)} \frac{p_n c_{i_0}}{(1+r)} = -\frac{r}{1+r} c_{i_0} < 0$, where we use the fact that $\sum_{n \in C(1)} p_n = p_1 = 1$. This contradicts the optimality of $\{\mathbf{x}^*, \mathbf{y}^*\}$. □

Lemma 1 indicates that in every optimal solution to model (2), one would never build a generator of any type that is not used for generation at the first planning period.

Proof of Theorem 5. Suppose conditions (i)-(iii) hold, in the objective function of both multi-stage and PA models, the generation cost becomes

$$\sum_{n\in\mathcal{T}}\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}_{t_n}}\frac{p_n h_{nk} b_{ink}}{(1+r)^{t_n-1}}v_{ink}=\sum_{n\in\mathcal{T}}\sum_{k\in\mathcal{K}_{t_n}}\frac{p_n b_{nk}}{(1+r)^{t_n-1}}\sum_{i\in\mathcal{I}}v_{ink}=\sum_{n\in\mathcal{T}}\sum_{k\in\mathcal{K}_{t_n}}\frac{p_n b_{nk} d_{nk}}{(1+r)^{t_n-1}},$$

which is a constant. This implies that in both multi-stage and PA models, the choice of generators to satisfy demand will only depend on the investment costs. Moreover, for any subproblem solved during the course of Algorithm 1, it follows from Lemma 1 that both multi-stage and PA model will expand the capacity in the most economic way to meet the current demand, but will not build any generator that is not used in the current period. In other words, multi-stage and PA models will make the same capacity expansion decisions at the root node of the subtree corresponding to that subproblem. Therefore, the solution output by Algorithm 1 is optimal to multi-stage problem (4). □

References

Schrijver, Alexander. 1998. *Theory of linear and integer programming*. John Wiley & Sons.