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. □

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}$$

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

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 $\mu\in\{1,\dots,T\}$, 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

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

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. $\square$

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 $x^*_{i_01} > \max\{\lceil [\mathbf{A}_{1k}\mathbf{y}^*_{1k}]_{i_0} \rceil : k \in \mathcal{K}_1\}$. Since $x^*_{i_01} \in \mathbb{Z}_+$, $x^*_{i_01} \geq \max\{\lceil [\mathbf{A}_{1k}\mathbf{y}^*_{1k}]_{i_0} \rceil : k \in \mathcal{K}_1\} + 1$. Let $\tilde{x}_{i_01} = x^*_{i_01} - 1$, $\tilde{x}_{i_0n} = x^*_{i_0n} + 1$ 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 $\{\mathbf{x}^*, \mathbf{y}^*\}$. $\square$

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.