

An optimization-based method for unit commitment

X Guan, P B Luh and H Yan

Department of Electrical and Systems Engineering,
University of Connecticut, Storrs, CT 06269-3157,
USA

J A Amalfi

Northeast Utilities Service Company, Berlin, CT
06037-1616, USA

An optimization-based method for unit commitment using the Lagrangian relaxation technique is presented. The salient features of this method includes nondiscretization of generation levels, a systematic method to handle ramp rate constraints, and a good initialization procedure. By using Lagrange multipliers to relax system-wide demand and reserve requirements and ramp rate constraints, the problem is decomposed into the scheduling of individual units. The optimal generation level of a unit at each hour can be easily calculated since there are no system dynamics, and the cost function is stage-wise additive and piecewise linear with only a few corner points. A relaxed subproblem can therefore be efficiently solved by using the dynamic programming technique without discretizing generation levels. A subgradient algorithm with adaptive step sizing is used to update Lagrange multipliers. An effective method based on priority-list commitment and dispatch is adopted to initialize these multipliers, and a heuristic approach is developed to generate a good feasible schedule based on the dual solution. Numerical results based on data sets from Northeast Utilities show that this algorithm is efficient, and near-optimal solutions are obtained.

Keywords: unit commitment, power system scheduling, mathematical programming, Lagrangian relaxation

I. Introduction

Unit commitment of a thermal power system is used to determine when to start up and/or shut down thermal units, and how to dispatch the committed units to meet system-wide demand and reserve requirements over a period of up to one week. Each unit may have minimum up and down times, ramp rate and other constraints. This class of mixed integer programming problems has been an active research subject for several decades

because of potential cost savings. According to a recent analysis, a 1% reduction in operating costs can result in savings of 10 to 30 million US\$ per year for an electrical utility with 10 000 MW of installed capacity¹. However, consistently generating optimal schedules has proved to be very difficult, because the problem belongs to the class of NP-hard combinatorial problems, and is considered to be extremely difficult to solve for systems of practical size (e.g. 100 units).

Recently, impressive results have been obtained by using the Lagrangian relaxation approach for obtaining near optimal solutions²⁻⁹. The basic idea is to relax system-wide demand and reserve requirements by using Lagrange multipliers. The problem can then be decomposed into individual unit commitment subproblems, which are much easier to solve. The high-level problem is to optimize Lagrange multipliers, and can be solved efficiently by using continuous variable optimization techniques. The disadvantage of this method is that the dual solution is generally infeasible, i.e. the once relaxed system-wide constraints are not satisfied. Some techniques, usually heuristics, are needed to modify the dual solution to obtain a good feasible schedule. Nevertheless, since the value of the dual function is a lower bound on the optimal cost, the quality of the feasible solution can be quantitatively measured.

A method for unit commitment based on the Lagrangian relaxation technique is presented here. The salient features of this method include nondiscretization of generation levels, a systematic method to handle ramp rate constraints, and a good initialization procedure. For a relaxed subproblem without ramp rate constraint, the optimal generation level of each hour can easily be determined since there are no system dynamics, and the cost function is stage-wise additive and piecewise linear with only a few corner points. This subproblem is solved by first constructing a state transition diagram where the optimal generation levels of all up states are computed without discretizing generation levels. Dynamic programming technique is then applied with only a few well structured states. This eliminates the difficult trade-off

Received 1 August 1991; revised 8 November 1991; accepted 6 February 1992

between computational requirements and accuracy as needed by most approaches that discretize generation levels.

The ramp rate constraint of a unit couples the generation levels of two consecutive hours. If the generation levels are discretized, this constraint can be handled by using the standard dynamic programming technique^{2,6}. The computational requirements, however, would increase significantly (possibly by an order of magnitude) as compared to the case without ramp rate constraints². If the generation levels are not discretized, the constraint is very difficult to deal with. A straightforward application of the dynamic programming technique may lead to suboptimal results (see Sections III.2 and III.7). This constraint is handled by using an *ad hoc* approach in Reference 8, and is not considered in Reference 9. In Reference 10, a subproblem is solved by dynamic programming with an extended state space for ramp-down logic, where limits of generation levels at each hour are established. The optimality and feasibility of the solution, however, are not guaranteed. Rather, ramp constraints are satisfied in a sophisticated economic dispatch process by using a 'look-ahead' heuristic method.

In our paper, ramp rate constraints are relaxed by introducing an additional set of multipliers for a unit with the constraints. The subproblem is then solved as if there were no ramp rate constraint. An intermediate level is introduced to update this set of multipliers, and a three-level framework is formed as shown in Figure 1. Note that intermediate subproblems are only needed for units with ramp rate constraints. An efficient subgradient algorithm¹¹ is modified to update Lagrange multipliers associated with system demand, reserve and ramp rate constraints. An effective method is developed to initialize these multipliers based on the priority-list commitment^{8,12} and dispatch. Since a dual solution is generally infeasible, a heuristic method is also developed to construct a feasible schedule based on the dual result.

The purpose of the research is to generate hourly schedules for Northeast Utilities Service Company (NU) which has about 70 thermal units. The maximum time horizon is 240 h (10 days). Numerical results based on NU data show that this method is efficient, and

near-optimal solutions are obtained. The algorithm developed has been embedded in the daily scheduling package of NU and used by NU engineers, mathematical problem formulation is given in Section II.

II. Problem formulation

Consider a thermal power system with I units. It is required to determine the startup, shutdown, and generation levels of all units over a specified time period T . The objective is to minimize the total cost subject to system demand and spinning reserve requirements, and other individual unit constraints. The time unit is one hour and the planning horizon may vary from one day to ten days. To formulate the problem mathematically, the following notation is introduced.

$C_i(p_i(t))$	fuel cost of unit i for generating power $p_i(t)$ at time t , a piecewise linear function of $p_i(t)$, in US\$
i	index of units, $i = 1, \dots, I$
I	number of units
$P_d(t)$	system demand at time t , in MW
$p_i(t)$	power generated by unit i at time t , in MW
$\bar{p}_i(t)$	maximum generation level of unit i at time t , in MW
$\underline{p}_i(t)$	minimum generation level of unit i at time t , in MW
$P_r(t)$	system spinning reserve requirement at time t , in MW
\bar{r}_i	maximum spinning reserve contribution of unit i , in MW
$r_i(x_i(t), p_i(t))$	spinning reserve contribution of unit i at time t , $r_i(\cdot) = 0$ if unit is down ($x_i(t) < 0$) and $r_i(\cdot) = \min\{\bar{p}_i(t) - p_i(t), \bar{r}_i\}$ if unit is up ($x_i(t) > 0$), in MW
R_i	ramp rate of unit i , in MW/h
$S_i(x_i(t), u_i(t))$	startup cost of unit i , in US\$
S_i^c	cold start-up cost of unit i , in US\$
S_i^h	hot start-up cost of unit i , in US\$
t	time index, $t = 1, \dots, T$
T	time horizon of commitment, in hours
$u_i(t)$	discrete decision variable of unit i at time t for up (1) or down (-1) of the unit at time $t + 1$
$x_i(t)$	state of unit i at time t , denoting number of hours that unit i has been on (positive values) or off (negative values)
Δ_i	maximum allowable change in generation between two consecutive hours, $\Delta_i = 1 \cdot R_i$, in MW
$\bar{\tau}_i$	minimum up-time of unit i , in hours
$\underline{\tau}_i$	minimum down-time of unit i , in hours
τ_i^c	cold start-up time of unit i , in hours

The problem is then formulated as the following mixed integer programming problem

(P)

$$\min_{\substack{u_i(t) \\ p(t)}} J, \quad \text{with } J \equiv \sum_{t=1}^T \sum_{i=1}^I [C_i(p_i(t)) + S_i(x_i(t), u_i(t))]$$

(1)

subject to system-wide constraints which include the

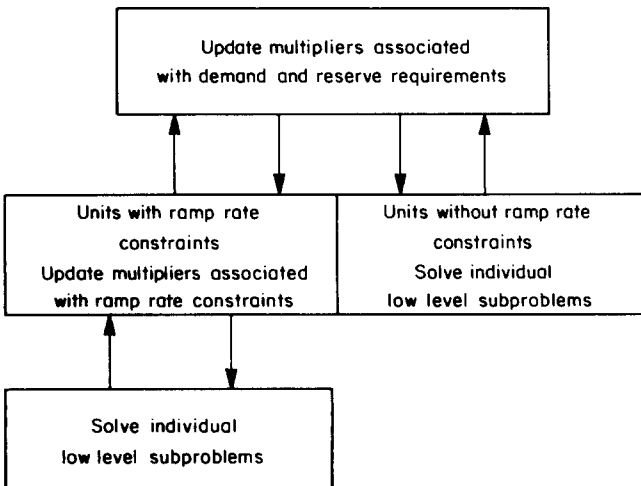


Figure 1. Three-level framework of Lagrangian relaxation

following

System demand

$$\sum_{i=1}^I p_i(t) = P_d(t) \quad (2)$$

Spinning reserve

$$\sum_{i=1}^I r_i(x_i(t), p_i(t)) \geq P_r(t) \quad (3)$$

Individual unit constraints include

State transition

$$x_i(t+1) = x_i(t) + u_i(t) \quad \text{if } x_i(t) \cdot u_i(t) > 0 \quad (4)$$

$$x_i(t+1) = u_i(t) \quad \text{if } x_i(t) \cdot u_i(t) < 0 \quad (5)$$

i.e. the number of hours being up or down accumulates if no start-up or shut-down occurs, otherwise the number of hours being up or down equals 1

Capacity

$$\underline{p}_i(t) \leq p_i(t) \leq \bar{p}_i(t) \quad \text{if } x_i(t) > 0 \quad (6)$$

$$p_i(t) = 0 \quad \text{if } x_i(t) < 0 \quad (7)$$

Some units may also have one or more of the following constraints.

Ramp rate

$$[p_i(t+1) - \Delta_i] \leq p_i(t) \leq [p_i(t+1) + \Delta_i] \quad \text{if } x_i(t) \geq 1 \text{ and } x_i(t+1) \geq 1 \quad (8)$$

Minimum up/down time

$$u_i(t) = 1 \quad \text{if } 1 \leq x_i(t) < \bar{\tau}_i \quad (9)$$

$$u_i(t) = -1 \quad \text{if } -\underline{\tau}_i < x_i(t) \leq -1 \quad (10)$$

implying that unit i must be kept on if it is up for less than the minimum-up time or be kept off if down for less than the minimum down time.

Minimum generation for the first and last hour (this is required by the New England Power Pool for steam units)

$$p_i(t) = \underline{p}_i(t), \quad r_i(x_i(t), p_i(t)) = 0 \quad \text{if } x_i(t-1) < 0 \text{ and } x_i(t) > 0 \\ \text{or if } x_i(t) > 0 \text{ and } x_i(t+1) < 0 \quad (11)$$

Must-run or must-not-run

$$x_i(t) > 0 \quad \text{for } t_{i1} \leq t \leq t_{i2} \quad (12)$$

if unit i is must-run for $t \in [t_{i1}, t_{i2}] \subset [1, T]$, and

$$x_i(t) < 0 \quad \text{for } t_{i3} \leq t \leq t_{i4} \quad (13)$$

if unit i is must-not-run for $t \in [t_{i3}, t_{i4}] \subset [1, T]$. Multiple must-run or must-not-run periods are also possible.

III. Solution methodology

III.1 The Lagrangian relaxation framework

As mentioned before, the basic idea of the Lagrangian relaxation technique is to relax system-wide constraints on demand and spinning reserve (equations (2) and (3)) by using Lagrange multipliers and to formulate a two-level structure. According to the cost function (1) and constraints (2) and (3), the Lagrangian is formulated

as follows

$$L = \sum_{t=1}^T \left\{ \sum_{i=1}^I [C_i(p_i(t)) + S_i(x_i(t), u_i(t))] \right. \\ \left. + \lambda(t)[P_d(t) - \sum_{i=1}^I p_i(t)] \right. \\ \left. + \mu(t)[P_r(t) - \sum_{i=1}^I r_i(x_i(t), p_i(t))] \right\} \quad (14)$$

where $\lambda(t)$ and $\mu(t)$ are Lagrange multipliers associated with demand and spinning reserve requirements at time t , respectively. For notational convenience, define

$$\lambda \equiv [\lambda(1), \lambda(2), \dots, \lambda(T)]^T \quad (15)$$

$$\mu \equiv [\mu(1), \mu(2), \dots, \mu(T)]^T \quad (16)$$

By using the duality theorem^{4,13} and exploiting the decomposable structure of equation (14), a two-level maximum-minimum optimization problem can be formed. Given multipliers λ and μ , the low level consists of individual unit subproblems

$$(\mathbf{P-i}), i = 1, 2, \dots, I \\ \min_{\substack{u_i(t) \\ p_i(t)}} L, \quad \text{with } L \equiv \sum_{t=1}^T \{ [C_i(p_i(t)) + S_i(x_i(t), u_i(t))] \\ - \lambda(t)p_i(t) - \mu(t)r_i(x_i(t), p_i(t))] \} \quad (17)$$

subject to equations (4)–(13).

Let $L_i^*(\lambda, \mu)$ denote the optimal Lagrangian for (P-i) with the given λ and μ . Then the high level dual problem is

$$(\mathbf{P-D}) \\ \max_{\lambda, \mu} \Phi(\lambda, \mu), \quad \text{with } \Phi(\lambda, \mu) \equiv \sum_{i=1}^I L_i^*(\lambda, \mu) \\ + \sum_{t=1}^T [\lambda(t)P_d(t) + \mu(t)P_r(t)] \quad (18)$$

subject to

$$\mu(t) \geq 0, \quad t = 1, 2, \dots, T \quad (19)$$

The above derivation presents a decomposition framework for solving the unit commitment problem. There are several steps to obtaining a near optimal solution: solving subproblems, solving the dual problem, and constructing a feasible solution, which are considered below.

III.2 Solving individual unit subproblems

The solution methodology for a subproblem without ramp rate constraints is presented first. For the cost function in equation (17) with λ and μ given, define the non-start-up cost as

$$f_i(p_i(t), x_i(t)) \equiv [C_i(p_i(t) - \lambda(t)p_i(t) \\ - \mu(t)r_i(x_i(t), p_i(t)))] \quad (20)$$

Clearly, if the unit is down (i.e. $x_i(t) < 0$), then

$$f_i(p_i(t), x_i(t)) = 0 \quad (21)$$

If the unit is up, this non-start-up cost function does not depend on the state $x_i(t)$ since the generation cost $C_i(p_i(t))$ and spinning reserve $r_i(x_i(t), p_i(t))$ generally depend only on the generation level if unit i is up. This

is true except for the first hour and last hour generation when equation (11) is active. That is, for any two arbitrary $x_i(t)$ and $x'_i(t)$ ($1 < x_i(t) \leq \bar{\tau}_i$ and $1 < x'_i(t) \leq \bar{\tau}_i$)

$$f_i(p_i(t), x_i(t)) = f_i(p_i(t), x'_i(t)) \quad (22)$$

Now the cost function of P-i in equation (17) can be rewritten as

$$L_i = \sum_{t=1}^T [f_i(p_i(t), x_i(t)) + S_i(x_i(t), u_i(t))]$$

Note that L_i is stage-wise additive, there are no dynamics on the generation levels, and the start-up cost $S_i(x_i(t), u_i(t))$ is independent of generation $p_i(t)$. The optimal generation level at time t for an up state ($x_i(t) > 0$) can therefore be obtained by minimizing $f_i(p_i(t), x_i(t))$ subject to the first hour and last hour generation constraint, equation (11). That is,

$$p_i^*(t) = \arg \min_{p_i(t)} f_i(p_i(t), x_i(t)) \quad (23)$$

if equation (11) is not active. Otherwise, $p_i^*(t) = p_i(t)$. To solve equation (23), note that the generation cost $C_i(p_i(t))$ and spinning reserve

$$r_i(x_i(t), p_i(t)) = \min \{ \bar{p}_i(t) - p_i(t), \bar{r}_i \}$$

are piecewise linear functions of $p_i(t)$, therefore, $f_i(p_i(t), x_i(t))$ defined in equation (20) is also piecewise linear with only a few corner points. The solution to equation (23) can thus be easily obtained by checking the corner points of f_i as shown in Figure 2. The corresponding optimal fuel cost can also be computed.

According to the billing rules of the New England Power Tool, the time varying start-up cost is a linear function of time since last shut down as shown in Figure 3. It is given by

$$S_i(x_i(t), u_i(t)) = S_i^h + (|x_i(t)| - \tau_i)(S_i^c - S_i^h)/(\tau_i^c - \tau_i) \quad \text{for } \tau_i \leq |x_i(t)| \leq \tau_i^c \quad (24)$$

The start-up cost remains constant after the cold start-up time. (This start-up cost function is different from the exponential start-up cost function used by some utilities. The effect of this difference is believed to be small, especially when $S_i^c - S_i^h$ is small. The exact difference has not been determined as no data is available. The algorithm, however, is not limited by the specific form of the start-up cost function.) The number of down states

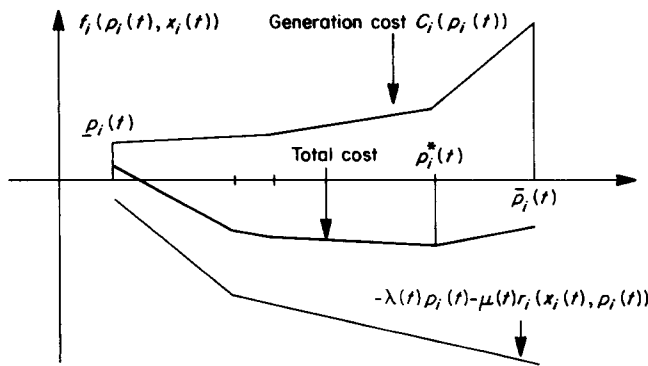


Figure 2. Stage size cost function of subproblem i

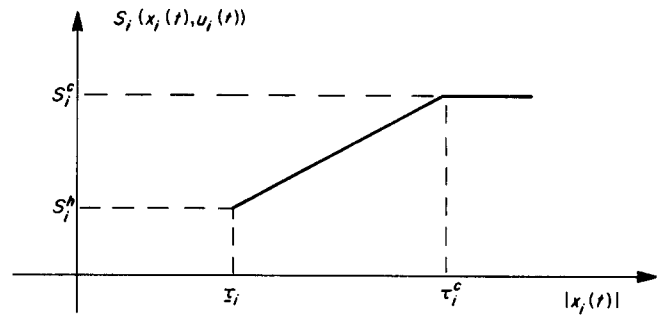


Figure 3. Linear start-up cost $S_i(x_i(t), u_i(t))$

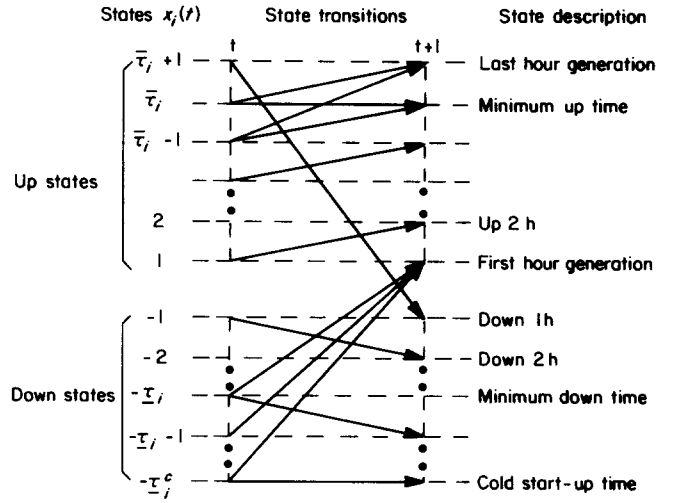


Figure 4. State transition diagram

needed to describe the different start-up costs at a particular hour is therefore equal to the cold start-up time. Since a unit can be kept on or shut down after it is up for minimum up time, the required number of up states is the minimum up time plus one, where the extra one is needed to consider last hour generation. By combining the above analysis for up and down states, the state transition diagram can then be constructed as in Figure 4. In the figure, each node represents a state and each edge with an arrow indicates a possible state transition. The non-start-up and start-up costs are associated with nodes (states) and edges (state transition), respectively. It should be noted that all the nodes corresponding to up states at hour t have the same generation level and therefore the same non-start-up cost with the possible exceptions of the first and last hour generations. The first and last hour generations for units with constraint (11) must comply with the constraint. The generation level and cost are zero for all the down state nodes. Based on this state transition diagram, the optimal commitment and generation of unit i can be obtained by using dynamic programming with a few states and well structured state transitions at each hour.

For a unit with ramp rate constraint (equation (8)), the generation levels of two consecutive hours are coupled. The optimal generation at hour t cannot be obtained by just considering the stage-wise cost of that hour as shown in Figure 2. It may not be obtained by a straightforward application of dynamic programming with $p_i(t)$ constrained to lie within

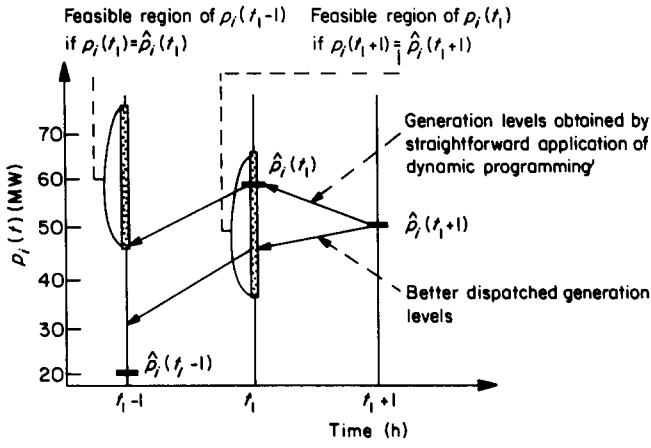


Figure 5. Effects of ramp rate constraint

$[p_i(t+1) - \Delta_i, p_i(t+1) + \Delta_i]$ either as shown in Figure 5. A numerical example associated with Figure 5 is provided in Section III.7. Obtaining the optimal solution for the subproblem is very difficult. The approach taken here is to use additional sets of multipliers v_{i1} and v_{i2} to relax the ramp up and ramp down constraints, respectively. The cost function in equation (17) then becomes

$$L'_i(\lambda, \mu, v_{i1}, v_{i2}) \equiv L_i(\lambda, \mu) + \sum_{t=1}^{T-1} \{v_{i1}(t) \times [p_i(t+1) - \Delta_i - p_i(t)] + v_{i2}(t)[p_i(t) - p_i(t+1) - \Delta_i]\} \quad (25)$$

By regrouping terms in equation (25) according to hours, subproblem P-i can be rewritten as:

$$\begin{aligned} \min_{\substack{u_i(t) \\ p_i(t)}} L'_i(\lambda, \mu, v_{i1}, v_{i2}) \\ \text{with } L'_i(\lambda, \mu, v_{i1}, v_{i2}) \equiv \sum_{t=1}^T [h_i(p_i(t), x_i(t), v_{i1}, v_{i2}) + S_i(x_i(t), u_i(t)) - \Delta_i(v_{i1}(t) + v_{i2}(t))] \end{aligned} \quad (26)$$

subject to equations (4)–(7) and (9)–(13). In equation (26)

$$\begin{aligned} h_i(p_i(t), x_i(t), v_{i1}, v_{i2}) \\ \equiv f_i(p_i(t), x_i(t)) + [v_{i2}(t) - v_{i1}(t)]p_i(t) + v_{i1}(t-1) - v_{i2}(t-1)]p_i(t), \\ t = 2, 3, \dots, T-1 \end{aligned} \quad (27)$$

$$\begin{aligned} h_i(p_i(1), x_i(1), v_{i1}, v_{i2}) \\ \equiv f_i(p_i(1), x_i(1)) + [v_{i2}(1) - v_{i1}(1)]p_i(1) \end{aligned} \quad (28)$$

$$\begin{aligned} h_i(p_i(T), x_i(T), v_{i1}, v_{i2}) \\ \equiv f_i(p_i(T), x_i(T)) + [v_{i1}(T-1) - v_{i2}(T-1)]p_i(T) \end{aligned} \quad (29)$$

where v_{i1} and v_{i2} are stack vectors like λ and μ . The optimal generation for each hour can be obtained

following equation (23)

$$p_i^*(t) = \arg \min_{p_i(t)} h_i(p_i(t), x_i(t)) \quad (30)$$

if equation (11) is not active. Otherwise, $p_i^*(t) = p_i(t)$. Dynamic programming can then be applied to solve P-i based on the diagram depicted in Figure 4.

Let $L_i^*(\lambda, \mu, v_{i1}, v_{i2})$ denote the optimal Lagrangian for equation (26). The multipliers $v_{i1}(t)$ and $v_{i2}(t)$ are updated at the intermediate level as shown in Figure 1 by a subgradient algorithm to maximize the Lagrangian, i.e.

$$\max_{\substack{v_{i1} \\ v_{i2}}} L_i^*(\lambda, \mu, v_{i1}, v_{i2}) \quad (31)$$

The subgradient algorithm to update the multipliers λ , μ , v_{i1} and v_{i2} is presented below.

III.4 Solving the dual problem

The high level dual problem is to update the multipliers λ and μ associated with demand and reserve requirements so as to maximize the dual function (equation (18)). Since discrete decision variables are involved at the low level, the objective function $\Phi(\lambda, \mu)$ in equation (18) may not be differentiable at certain points. Therefore, a subgradient algorithm is used to update λ and μ as follows^{11,14–16}

$$\lambda^{l+1}(t) = \max[0, \lambda^l(t) + \alpha^l g_\lambda(t)] \quad (32)$$

$$\mu^{l+1}(t) = \max[0, \mu^l(t) + \alpha^l g_\mu(t)] \quad (33)$$

where l is the high level iteration index, α is the step size,

$$g_\lambda(t) \equiv P_d(t) - \sum_{i=1}^I p_i(t) \quad (34)$$

is the subgradient of $\Phi(\lambda, \mu)$ with respect to $\lambda(t)$, and

$$g_\mu(t) \equiv P_d(t) - \sum_{i=1}^I r_i(x_i(t), p_i(t)) \quad (35)$$

the subgradient of $\Phi(\lambda, \mu)$ with respect to $\mu(t)$. From equation (14), the multiplier $\lambda(t)$ is the marginal cost for demand at time t , i.e. the cost to generate one more MW power at time t . The multiplier $\lambda(t)$ is therefore positive although it is associated with an equality constraint.

The adaptive step sizing method of Reference 11 is modified to obtain the step size α at iteration l . It is given by

$$\alpha = \gamma \frac{\bar{L} - L}{[g_\lambda^T, g_\mu^T][g_\lambda^T, g_\mu^T]^T}$$

where \bar{L} is an estimate of the optimal value of L with $\bar{L} \geq L$, g_λ and g_μ are stack vectors of $g_\lambda(t)$ and $g_\mu(t)$ as in equation (15), and γ is a positive scaling constant. Proof of convergence of the above formula is given in Reference 14. The adaptive step sizing is for the selection of \bar{L} and γ . The estimate \bar{L} generally decreases as the number of iterations increases. Since the dual cost L is generally increasing with the number of iterations, step size α decreases. When $\bar{L} - L$ comes within a certain threshold δ_d , \bar{L} is increased by a certain percentage causing α to jump. This jump is often too large resulting in a drop in L as shown in Figure 6. The estimate \bar{L} then decreases with the number of iterations and the process

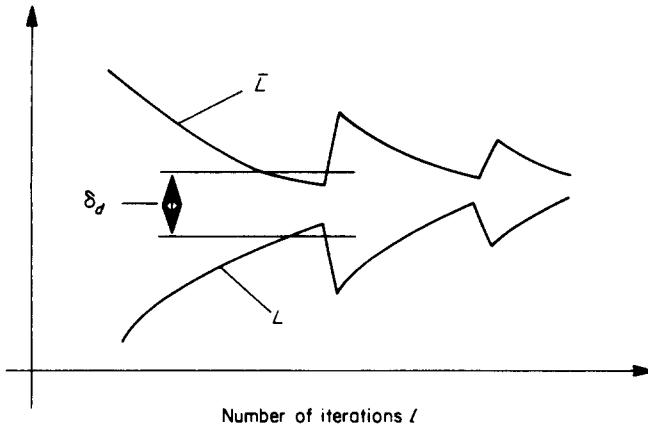


Figure 6. Adaptive step sizing method

repeats. The scaling constant γ is reduced by half if the dual cost L does not increase as \bar{L} decreases, or if the values of L just before the jumps (the peak values of L as in Figure 6) do not increase. This decrease in γ prevents oscillation caused by too large step sizes. The high level iteration terminates when the dual cost L cannot be further improved, or a pre-set number of iterations has been reached.

By using this step sizing method, the total number of high level iterations required is not sensitive to the initial values of γ and \bar{L} , nor to the increment or decrement of \bar{L} . Therefore, trial-and-error parameter tuning efforts can be greatly reduced, and the algorithm is robust for various data sets tested.

The multipliers v_{i1} and v_{i2} for relaxing the ramp rate constraint are updated in the same way as in equations (32)–(35) with λ and μ replaced by v_{i1} and v_{i2} . The iteration index, however, is different. For each high level iteration with λ and μ given, v_{i1} and v_{i2} are updated at the intermediate level (Figure 1) until the dual cost function in equation (31) cannot be improved. The subgradients of L_i^* with respect to v_{i1} and v_{i2} are

$$g_{v1}(t) = p_i(t+1) - \Delta_i - p_i(t) \quad (36)$$

and

$$g_{v2}(t) = p_i(t) - [p_i(t+1) + \Delta_i] \quad (37)$$

III.5 Initializing the multipliers

Good initialization of multipliers can significantly reduce the number of high level iterations needed. In our algorithm, the multipliers associated with system demand are initialized to be the system marginal costs based on the priority-list commitment and dispatch. For each hour, a unit priority table based on unit full load average rates is first constructed. The units are committed in the ascending order of their full load average rates until the committed capacity meets the sum of the demand and reserve requirements. A block priority table of those committed units is then constructed where a block of a unit represents a segment of generation with equal marginal cost. Blocks are dispatched in the ascending order of block rate until the system demand is satisfied. The multiplier $\lambda(t)$ is then initialized to be the system marginal cost, i.e. the rate of the last block dispatched at time t . All other multipliers are initialized to zero since the reserve and ramp rate constraints are too difficult to be considered in initialization.

III.6 Obtaining feasible solutions

The dual solution is generally infeasible, i.e. the demand and reserve constraints generally are not satisfied⁶. A heuristic method is developed to generate a good feasible solution based on the dual results. The dual solution is first checked to determine whether it can be made feasible by adjusting generation levels of committed units only. Since generation levels of some units cannot be adjusted or are difficult to adjust, e.g. units at first or last hour generation and units with ramp rate constraints, the committed units at hour t are divided into three nonoverlapping sets. Let E_{2t} denote the set of units at first or last hour generation at hour t with equation (11) active, E_{3t} the set of units with ramp rate constraints but not in E_{2t} , and E_{1t} the rest of committed units at hour t . Then the following two inequalities form a set of sufficient conditions for the feasible dispatch of the committed units

$$\sum_{i \in E_{1t}} \bar{p}_i(t) \geq P_d(t) + P_r(t) - \sum_{j \in E_{2t}} p_j(t) - \sum_{k \in E_{3t}} p_k(t) \quad (38)$$

$$\sum_{i \in E_{1t}} \bar{r}_i \geq P_r(t) - \sum_{k \in E_{3t}} r_k(x_k(t), p_k(t)) \quad (39)$$

Equation (38) requires the committed capacity to meet the total demand and reserve requirements. Equation (39) then guarantees that the committed units can provide enough reserve. If equations (38) and (39) are satisfied, a feasible economic dispatch solution can be constructed by manipulating generation levels of the units in E_{1t} while fixing the generation levels and reserve contributions of the units in E_{2t} and E_{3t} . The first step is to preserve enough reserve from the units in E_{1t} to meet the reserve requirement left by the units in E_{3t} (the units in E_{2t} do not provide reserve as can be seen from equation (11)). This is done by preserving blocks in the descending order of block rate until the reserve requirement is satisfied, subject to maximum reserve contributions of individual units. The unpreserved blocks of the units in E_{1t} are then dispatched in the ascending order of block rates until the system demand is satisfied.

If equations (38) and (39) are not satisfied, more units are to be committed at hour t . Two feasible solutions are generated, and the better one is selected. The first solution is obtained by committing turbine units without minimum up and down time constraints. This is good for isolated infeasible hours. The second solution is generated by adjusting steam units based on unit full load average rates subject to minimum up and down time constraints. A feasible solution is generally obtained by starting up some units earlier or shutting down some units later than scheduled in the dual solution. A steam unit may also be started up just to cover a number of consecutive infeasible hours. The infeasibility caused by too many committed units has not yet occurred in our testing. If so, some units can be shut down in a similar way based on full load average rates.

The ramp rate constraint in the dual solution is generally satisfied or with only a few minor violations at the convergence of an intermediate subproblem as observed in our testing. A simple heuristic has been developed to adjust the generation levels to meet the ramp rate constraint before checking equations (38) and (39). This is done by adjusting generation levels to be within the ramp rate starting from the first hour generation forward in time and also from the last hour

generation backward in time until the ramp rate is satisfied for all hours.

A few additional high level iterations, each going through the heuristics to obtain a feasible schedule, are then carried out (the so-called 'heuristic iterations') to select a good feasible solution. The final cost and the maximum dual function value are used to calculate the duality gap, a measure of the quality of the feasible schedule.

III.7 Summary of the algorithm

The algorithm is summarized as follows

- (1) [Initialize] Initialize system demand multiplier λ according to priority-list commitment and dispatch. Initialize all other multipliers μ , v_{i1} and v_{i2} to zero.
- (2) [Solve subproblems] For a unit without ramp constraint, go to Step 2a. Otherwise go to Step 2b. If all the subproblems have been solved, go to Step 3.
 - (a) Solve the subproblem without the ramp rate constraint for the given λ and μ according to Section III.2. Go to Step 2.
 - (b) Solve the subproblem with the ramp rate constraints for the given λ , μ , v_{i1} and v_{i2} according to Section III.2.
 - (c) Update v_{i1} and v_{i2} by using the subgradient method.
 - (d) If the stopping criteria for v_{i1} and v_{i2} are satisfied, go to Step 2. Otherwise go to Step 2b.
- (3) [Update multipliers] Update λ and μ according to equations (32) and (33).
- (4) [Check convergence] If the stopping criteria for the high level problem (equation (18)) have not been satisfied, go to Step 2.
- (5) [Generate feasible solutions] If a feasible solution can be obtained without changing commitment, go to Step 5a. Otherwise go to Step 5b.
 - (a) Generate a feasible solution by economic dispatch. Go to Step 6.
 - (b) Obtain two feasible solutions by using the heuristic methods discussed in Section III.5.
- (6) [Select the best feasible solution] Select the best feasible solution obtained. If the desired number of heuristic iterations is reached, stop.
- (7) [Update multipliers] Follow Step 3 to update λ and μ , and follow Step 2 to solve low level subproblems. Go to Step 5.

III.8 Discussion on methods for solving subproblems with ramp rate constraints

The optimal solution of a subproblem with ramp rate constraint may not be obtained by a straightforward application of dynamic programming as used in Reference 8. This can be illustrated by the following simple example. For convenience of presentation, suppose that the generation cost $C_i(p_i(t))$ is a quadratic function of $p_i(t)$ (approximation to the piecewise linear function shown in Figure 2). The stage-wise dual cost can be written as

$$f_i(p_i(t), x_i(t)) = a_i p_i^2(t) + b_i p_i(t) + c_i - \lambda(t) p_i(t) - \mu(t) r_i(p_i(t), x_i(t)), \text{ for } x_i(t) \geq 1$$

where a_i , b_i and c_i are appropriate coefficients. If the ramp rate constraint is ignored and $\mu(t) = 0$, the optimal generation level at hour t can be obtained by differentiating the above equation and considering the

minimum and maximum generation levels, i.e.

$$\hat{p}_i(t) = \min\{\max\{-(b_i - \lambda(t))/(2a_i), \underline{p}_i(t)\}, \bar{p}_i(t)\}$$

With parameters $a_i = 0.05$, $b_i = 24$, $c_i = 50$, $p_i(t) = 15$ MW, $\bar{p}_i(t) = 75$ MW, and the multipliers $\lambda(t_1 - 1) = 26$, $\lambda(t_1) = 30$, $\lambda(t_1 + 1) = 29$, $\mu(t_1 - 1) = \mu(t_1) = \mu(t_1 + 1) = 0$ for a particular t_1 , the optimal generation levels at $t_1 - 1$, t_1 and $t_1 + 1$ without the ramp rate constraint are found to be

$$\hat{p}_i(t_1 - 1) = 20 \text{ MW}$$

$$\hat{p}_i(t_1) = 60 \text{ MW}$$

$$\hat{p}_i(t_1 + 1) = 50 \text{ MW}$$

Now consider the ramp rate constraint with $\Delta_i = 15$ MW. By straightforward application of dynamic programming as shown in Figure 5 with $p_i(t_1 + 1) = \hat{p}_i(t_1 + 1) = 50$ MW, the generation levels at t_1 and $t_1 - 1$ are obtained as

$$p_i(t_1) = \hat{p}_i(t_1) = 60 \text{ MW}, \quad p_i(t_1 - 1) = 45 \text{ MW}$$

with total dual cost -143.75 . Note that $p_i(t_1 - 1)$ is on the boundary of the feasible region delineated by $p_i(t_1) = 60$ MW. However, taking

$$p_i(t_1 + 1) = 50 \text{ MW}$$

$$p_i(t_1) = 45 \text{ MW}$$

$$p_i(t_1 - 1) = 30 \text{ MW}$$

one obtains a better result with total dual cost -158.75 .

The methods developed in Reference 10 for solving subproblems with ramp rate constraints are efficient. The schedules generated by using these methods are more likely to meet the constraints than if these methods were not used. However, these methods do not guarantee that a schedule satisfying the ramp rate constraints will be produced in solving subproblems (conclusion in Reference 10), and optimality is not systematically considered.

The computational requirements of our ramp rate relaxation method are generally higher than the requirements of methods in References 8 and 10 since a number of intermediate iterations may be needed. The computational time for economic dispatch, however, would be lower in comparison with other methods because our method for constructing a feasible solution, as can be seen from Section III.5, is simpler.

As compared to the methods that discretize generation levels, an advantage of our ramp rate relaxation method is that the computational complexity is related to the number of active ramp rate constraints. If no ramp constraint is active, the complexity is about the same as for the case without the constraint since v_{i1} and v_{i2} are initialized to zero. The complexity increases as more ramp rate constraints become active. This is in contrast to the discretized dynamic programming approach, where computational complexity is fixed whether the constraint is active or not.

IV. Numerical results

Numerical testing of our algorithm was performed by using five NU data sets: week 2 in August 1989; week 2 in December 1989; week 4 in February 1990; week 3 in April 1990 and week 4 in March 1991. These data sets cover weeks in various seasons and were randomly

selected from NU billing data files. Hydro and pumped storage contributions, together with power provided by cogenerators and nondispatchable contracts, were deducted from system demand and reserve requirements. The number of thermal units and dispatchable contracts is about 70, and the commitment horizon is one week, i.e. 168 h. The reserve requirement considered is the 10 min spinning reserve. Each dispatchable contract is modelled as a one-block unit without minimum up and down time constraints, and without contribution to reserve. A summary of the major system parameters is given in Table 1.

All the billing rules of New England Power Pool are complied with and many practical considerations are included. For example, unit characteristics and fuel prices are allowed to change within the commitment horizon in view of maintenance, unit entitlement change, etc., and generation levels of some units may be fixed for several time intervals for testing or maintenance purposes. Some units have extremely long minimum down times (e.g. 55 h), and some units have very long cold start-up times (e.g. 60 h and some even up to 110 h). These result in large numbers of down states. All the above add to the complexity of the testing.

The algorithm was implemented in FORTRAN on SUN Sparc Station 2. Numerical results for the five data sets are summarized in Table 2. For each data set, the case with the ramp rate dropped and the case with all constraints considered were tested to see the effects of

ramp rate on algorithm performance. There are only four units with ramp rate constraints (excluding nuclear units whose ramp rate constraints are seldom active). For the data sets tested, ramp rate violations do occur if this constraint is dropped.

It can be seen from Table 2 that the CPU times for all five data sets (and for many others not shown here) are only a few minutes on a Sparc Station 2. As expected, it takes more time to solve problems with ramp rate constraints than problems without them since a number of intermediate iterations are needed to update the ramp rate multipliers. To speed up computation, the constraints are not imposed until a certain number of high level iterations have been completed. The total numbers of high level iterations for cases with or without the constraints, however, are very close, and the increase in CPU time is not as bad as reported². As the number of ramp rate constrained units increases, the increase in computational time would mostly be attributed to the increased effort spent on intermediate level iterations. It has also been observed from testing that the number of high level iterations is insensitive to system size and time horizon, and the computational time increases about linearly as the number of units or commitment horizon increases.

The quality of a solution is measured by the duality gap, defined as the relative difference between the final cost and the maximum value of the dual function. From Table 2, it can be seen that ramp rate constraints cause

Table 1. Summary of system characteristics

System characteristics	Total number					Total capacity or requirement (MW)				
	Aug w2 89	Dec w2 89	Feb w4 90	Apr w3 90	Mar w4 91	Aug w2 89	Dec w2 89	Feb w4 90	Apr w3 90	Mar w4 91
Steam units	24	25	32	31	25	1684	1583	1804	1776	1547
Turbine units	27	26	23	24	26	257	288	293	238	204
Nuclear units	7	6	7	7	7	2471	2406	1245	2102	1792
Dispatchable contracts	8	19	8	6	6	350	720	689	292	245
All units	67	76	70	68	64	4762	4997	4031	4408	3788
Peak load						3380	3680	3132	3771	3374
Minimum load						2200	2800	2036	2559	2524
Maximum reserve						140	125	132	149	166

Table 2. Summary of testing results

	Data sets	Number of high level itera- tions	CPU time (sec)	Max. dual cost (\$)	Best feasible cost (\$)	Duality gap (%)
Without ramp rate	Aug w2, 89	39	231.12	4,551,996.65	4,565,930.38	0.30
	Dec w2, 89	55	235.34	10,494,111.07	10,526,993.52	0.39
	Feb w4, 90	57	240.24	5,693,638.36	5,698,307.96	0.08
	Apr w3, 90	29	183.11	5,856,300.91	5,863,637.03	0.13
	Mar w4, 91	43	242.87	4,575,202.30	4,579,412.09	0.09
With ramp rate	Aug w2, 89	37	282.24	4,550,967.85	4,565,506.38	0.32
	Dec w2, 89	49	365.31	10,498,197.37	10,536,055.69	0.49
	Feb w4, 90	66	423.94	5,700,291.95	5,711,858.93	0.20
	Apr w3, 90	34	343.51	5,857,696.82	5,868,396.14	0.18
	Mar w4, 91	47	324.09	4,575,897.71	4,581,127.17	0.11

the duality gaps to increase, however, all the gaps are below 0.5% with most of them below 0.3%. The duality gaps of other data sets tested but not reported here are in the same range. The results thus demonstrate consistent convergence of the algorithm, and support our claim that near optimal solutions are obtained by using this algorithm. Note that duality gap is related to system characteristics. It has been reported that the gap increases as the minimum up and down times and start up costs increase⁶. The extremely long minimum down times of some NU units, the fixing of generation levels in some intervals, and the presence of ramp rate constraints are factors contributing to the duality gap.

The advantage of having a good initialization of multipliers is clear. If one starts with $\lambda = \mu = 0$, more than ten high level iterations are often needed to let λ approach the value obtained from our initialization scheme.

It is difficult to compare CPU times of different algorithms reported in the literature since the results were obtained based on systems with different sizes, characteristics, constraints, and obtained by using different types of computers with different configurations. Precise comparison should be based on the same testing environments. Through rough conversion, however, it can be seen that the CPU times of our algorithm are in the similar range of the CPU times of References 2, 6, 8 and 9. For example, the CPU time of the thermal unit commitment problem without the ramp rate constraint presented in Reference 9 is 58 min on a VAX-11/780, while our algorithm takes about 4 min on a Sparc 2 for our data sets without the ramp rate constraint. Considering that Sparc 2 is roughly 10 times faster than VAX-11/780 and the system size in Reference 9 is larger than ours, our CPU times are in the similar range of the CPU times of Reference 9. It is difficult to compare methods for handling the ramp rate constraint since little information is available on the number of units with ramp rate constraints and the number of ramp rate constraints that are actually active.

V. Conclusions

A new algorithm has been presented to solve thermal unit commitment problems by using the Lagrangian relaxation approach. For individual subproblems, dynamic programming without discretizing generation levels proved to be an efficient approach. The ramp rate constraint is handled through relaxation. This method provides the advantages of nondiscretization of generation levels and is proved to be efficient for systems with a few ramp rate constrained units. Good initialization of multipliers associated with system demand by using the priority-list commitment and dispatch can significantly cut down the computational time. The heuristic method developed to obtain feasible solutions is effective, and

near optimal solutions are obtained. This algorithm has been embedded in the daily scheduling package of NU, and used by NU engineers.

VI. Acknowledgment

This work was supported in part by the National Science Foundation under Grant ECS-8717167 and a contract from the Northeast Utilities Service Company. The authors would like to thank Dr R W Goodrich, Mr M B Gabris, Mr C Larson and Mr P Rogan of Northeast Utilities Service Company for their valuable suggestions and support.

VII. References

- 1 Grimes, R C and Jabbour, S J 'The DYNAMICS model for measuring dynamic operating benefits' Technical Report, Decision Focus Incorporated, Los Altos, CA, June 1989
- 2 Bard, J F 'Short-term scheduling of thermal-electric generators using Lagrangian relaxation' *Oper Res* Vol 36 No 5 (1988) pp 756-766
- 3 Batut, J, Renaud, A and Sandrin, P 'New software for the generation rescheduling in the future EDF national control center' *Proc Tenth Power Systems Computation Conf* Graz, Austria, 19-24 August 1990, pp 1163-1170
- 4 Bersekas, D P, Lauer, G S, Sandell, N R and Posbergh, T A 'Optimal short-term scheduling of large-scale power systems' *IEEE Trans Automatic Control* Vol AC-28 No 1 (1983) pp 1-11
- 5 Cohen, A I 'Optimization-based methods for operations scheduling' *Proc IEEE* Vol 75 No 12 (1987) pp 1574-1591
- 6 Ferreira, L A F M, Andersson, T, Imparato, C F, Miller, T E, Pang, C K, Svoboda, A and Vojdani, A F 'Short-term resource scheduling in multi-area hydrothermal power systems' *Electric Power & Energy Syst* Vol 11 No 3 (1989) pp 200-212
- 7 Merlin, A and Sandrin, P 'A new method for unit commitment at Electricite de France' *IEEE Trans Power Appar & Syst* Vol PAS-102 (1983) pp 1218-1225
- 8 Ruzic, S and Rajakovic, N 'A new approach for solving extended unit commitment problem' *IEEE Trans Power Syst* Vol 6 No 2 (1991) pp 269-277
- 9 Shaw, J J and Bersekas, D P 'Optimal scheduling of large hydrothermal power systems' *IEEE Trans Power Appar & Syst* Vol PAS-104 (1985) pp 286-293
- 10 Cohen, A I 'Modeling unit ramp limitation in unit commitment' *Proc Tenth Power Systems Computation Conf*, Graz, Austria, 19-24 August 1990, pp 1107-1114
- 11 Luh, P B, Hootomt, D J, Max, E and Pattipati, K R 'Schedule generation and reconfiguration for parallel machines' *IEEE Trans Robotics & Automation* Vol 6 No 6 (1990) pp 687-696
- 12 Wood, A J and Wollenberg, B F *Power generation and control* Wiley, Chichester, UK (1984)
- 13 Nemhauser, G L and Wolsey, L A *Integer and combinatorial optimization* Wiley, Chichester, UK (1988)
- 14 Held, M, Wolfe, P and Crowder, H P 'Validation of subgradient optimization' *Mathematical Programming* Vol 6 (1974) pp 62-88
- 15 Polyak, B T 'Minimization of unsmooth functionals' *USSR Computational Math and Math Physics* Vol 9 (1969) pp 14-29
- 16 Shor, N Z 'On the structure of algorithms for the numerical solution of optimal planning and design problems' *Dissertation, Cybernetics Institute, Academy of Sciences, USSR* (1964)