Optimal Short-Term Scheduling of Large-Scale Power Systems

DIMITRI P. BERTSEKAS, SENIOR MEMBER, IEEE, GREGORY S. LAUER, MEMBER, IEEE,
NILS R. SANDELL, JR., MEMBER, IEEE, AND THOMAS A. POSBERGH, MEMBER, IEEE

Abstract—This paper is concerned with the longstanding problem of optimal unit commitment in an electric power system. We follow the traditional formulation of this problem which gives rise to a large-scale, dynamic, mixed-integer programming problem. We describe a solution methodology based on duality, Lagrangian relaxation, and nondifferentiable optimization that has two unique features. First, computational requirements typically grow only linearly with the number of generating units. Second, the duality gap decreases in relative terms as the number of units increases, and as a result our algorithm tends to actually perform better for problems of large size. This allows for the first time consistently reliable solution of large practical problems involving several hundreds of units within realistic time constraints. Aside from the unit commitment problem, this methodology is applicable to a broad class of large-scale dynamic scheduling and resource allocation problems involving integer variables.

I. INTRODUCTION

A PROBLEM that must be solved on a daily basis by a power utility is to determine a schedule of what units will be used to meet the demand anticipated over a future 24-hour period. This is commonly referred to as the *unit commitment problem*, and its solution has been the subject of intensive efforts over the last 20 years. References [1] and [2] give an extensive account of these efforts and an appraisal of the state of the art. Generally, the available approaches can be categorized into two groups. The first group consists of rigorous optimization approaches that are impractical for application to problems of realistic size. The second group consists of heuristic approaches that are actually used in practice, but give no assurance that the schedules produced are optimal or even close to optimal.

The purpose of this paper is to describe an optimization methodology which is, for the first time, capable of solving realistic unit commitment problems. The approach is based on a duality transformation of the original problem and optimal solution of the associated (nondifferentiable) dual

problem. This approach coupled with the branch-and-bound technique is common in integer programming and is referred to as Lagrangian relaxation (see [1], [3]–[5]). It has been used earlier for solution of the unit commitment problem by Muckstadt and Koenig [1]. It is theoretically possible with this approach to obtain an optimal solution by examining a sufficient number of nodes of the branch-and-bound tree, but in practice, due to computation time limitations, one can only examine a small number of nodes and hope to find a good suboptimal feasible solution. Verification of the quality of this feasible solution is based on its cost (which is an upper bound to the optimal cost) and the optimal value of the corresponding dual problem (which is a lower bound to the optimal cost). Typically, the branch-and-bound tree search is terminated when the best feasible solution obtained is demonstrably (according to the preceding criterion) within a certain percentage of the optimum (1 percent is the tolerance used in [1]).

An approach of this type can be successful in solving unit commitment problems of realistic size only if it meets the following two prerequisites.

1) The difference between the optimal values of the primal and dual problems (the duality gap) is small.

2) The nondifferentiable optimization method used for solving the dual problem provides sufficient information for generating a nearly optimal feasible solution of the primal problem.

It turns out that the duality gap for the unit commitment problem is not only typically small, but *also becomes smaller (in relative terms) as the number of units increases*. This fact, which has not been observed earlier, is established rigorously in this paper and suggests that Lagrangian relaxation is particularly well suited for unit commitment problems of large size.

The approach of Muckstadt and Koenig [1] is not practically viable primarily because it does not satisfy prerequisite 2). The method used in [1] is a version of Shor's original subgradient method (see [5]–[7]) which solves satisfactorily the dual problem (as verified also by us), but does not provide sufficient information for finding consistently a good feasible solution of the primal problem. This is reflected in computational results [1] which show that the number of nodes of the branch-and-bound tree that need to be examined before a satisfactory solution

(within 1 percent of the optimum) is found varies widely from a few nodes to nearly a hundred. It is therefore not surprising that the largest problem solved in [1] involves only 15 generating units over 12 time periods.

Our own success with large-size problems is due to the different method we use for solving the nondifferentiable dual problem. This method (due to Bertsekas [8] and [9]) is related to a method of multipliers with exponential penalty function (Kort and Bertsekas [10]), and is based on approximating the dual problem with a twice-differentiable problem which is subsequently solved by a constrained version of Newton's method [11]. The crucial fact is that this method not only solves the dual problem, but also provides additional information in the form of certain multipliers that solve a problem which is a *relaxed* version of the original unit commitment problem (in the usual relaxed control sense). These multipliers form the basis for generating a feasible unit commitment schedule which, based on our computational experience, is consistently within 1 percent (and usually within 0.5 percent) of the optimum. This in turn allows us to *entirely abandon the branch-and-bound philosophy* or equivalently satisfy ourselves with examining *only one node of the branch-and-bound tree*. Since the number of iterations of Newton's method needed to solve the dual problem is quite insensitive to the number of generating units, it follows that *the computational requirements for solving the unit commitment problem in this manner typically grow only linearly with problem size*. We are currently able to solve consistently with our experimental code problems involving 200 units over 24 time periods in 10–12 min of VAX-11/780 CPU time. This indicates that our optimization methodology is applicable to even the largest practical unit commitment problems to be encountered presently.

This paper is organized as follows. In the next section, we formulate the unit commitment problem as a deterministic, dynamic optimization problem involving integer and continuous variables. The formulation is essentially the same as earlier formulations (see, e.g., [1]). A discussion of the underlying physical assumptions and the relevance of the mathematical model is given in a separate paper [12]. For simplicity, we have restricted ourselves to thermal generating units exclusively. Thus we have assumed in effect that all hydro power generation has already been scheduled. The combined thermal and hydro unit scheduling problem is the subject of a separate publication [23] in which we show how it is possible (again using duality and nondifferentiable optimization) to schedule optimally the hydro units first over a period of one week and then the thermal units using the method of this paper. We have also assumed for simplicity the absence of interchange contracts, but the basic model can be easily modified to account for this possibility.

In Section III we formulate a problem which is dual to the unit commitment problem and which involves a nondifferentiable objective function. In Section IV we describe the approximation method for solving the dual problem. In Section V we introduce a relaxed version of the unit commitment problem and show how the approximation method of Section IV solves this problem simultaneously with the dual problem. We also show how various derivatives needed in the approximation method can be computed efficiently by making use of the structure of the relaxed problem.

In Section VI we show that the duality gap, as a percentage of the minimum cost, is inversely proportional to the number of units. This analysis is carried out by reformulating the relaxed version of the problem as a linear program and applying known results of linear programming theory (compare with [4, p. 172]). In Section VII we show how a good feasible solution of the original unit commitment problem can be generated by making use of the solution of the relaxed problem. In Section VIII we report on our computational experience, and in Section IX we summarize our results.

II. PROBLEM FORMULATION

Given a power system consisting of I thermal units, the problem is to schedule startup, shutdown, and power generation of these units over N time periods so as to minimize fuel costs while meeting given demand and reserve requirements.

For each unit i, we denote the following.

g_t^i The average output power in period t.

$\underline{g}_i$ The minimum output power of unit i.

$\bar{g}_i$ The maximum output power of unit i.

$C_i(g)$ The fuel cost for operating unit i at power level g over one period.

u_i^t The startup/shutdown decision variable for unit i at period t (one for startup, zero for shutdown).

We will assume that each unit can be in one of two states: up (denoted by 1) and down (denoted by 0). Thus the state transition diagram of each unit is as given in Fig. 1, and if we denote by x_i^t the state of unit i at time t, the equation governing its evolution is given by

$$x_i^{t+1} = u_i^t \qquad (i = 1, \cdots, I; t = 0, 1, \cdots, N-1). \quad (2.1)$$

This assumption neglects typical practical constraints that require that once a unit is started up (shut down), it cannot be shut down (started up) for a given number of periods. However, it is possible to take into account these constraints by introducing additional states into the state transition graph. The methodology described in this paper can be trivially modified to account for this possibility. The necessary changes are described in [12] and [13], and in fact our implemented code (which was used to produce the computational results of Section VIII) fully takes into account these uptime and downtime constraints. We have decided to adopt a simpler model in this paper to make the presentation more clear and to cope better with an already overburdened notation.

The startup/shutdown cost for unit i will be denoted by $S_i(x_i, u_i)$ and depends on the state x_i and the decision variable u_i. It is possible to allow for startup costs that
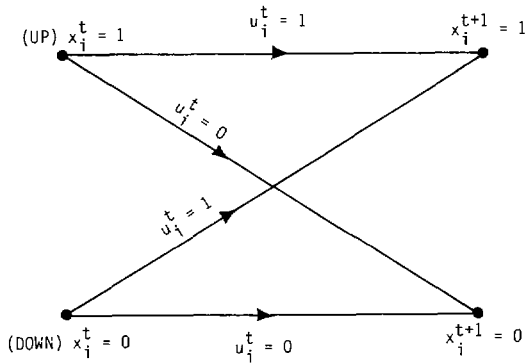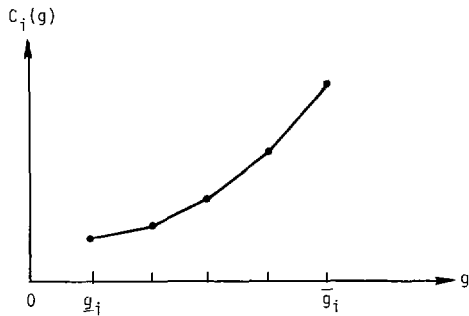
Fig. 1. State transition diagram for unit i.

Fig. 2. Cost-power generation curve.

depend on the number of periods that the unit has been down by introducing additional down states in the state transition graph. We make the natural assumption that

$$S_i(1,1) = S_i(0,0) = 0. \tag{2.2}$$

The constraints on the output power generated for each unit i and period t are

$$\underline{g}_i \leqslant g_i^t \leqslant \bar{g}_i \qquad \text{if } u_i^t = 1. \tag{2.3a}$$

$$g_i^t = 0 \qquad \text{if } u_i^t = 0. \tag{2.3b}$$

The cost-power generation function $C_i(g)$ is assumed to be convex and piecewise linear in the interval $[\underline{g}_i, \bar{g}_i]$ as shown in Fig. 2. We also assume that $C_i(0) = 0$.

The demand and reserve constraints are expressed as

$$\sum_{i=1}^{I} g_i^t \geqslant D^t \qquad (\forall t = 0, 1, \cdots, N-1) \tag{2.4}$$

$$\sum_{i=1}^{I} r_i(u_i^t) \geqslant R^t \qquad (\forall t = 0, 1, \cdots, N-1) \tag{2.5}$$

where D^t is the expected average demand (not met by hydroelectric generation) in period t, and R^t is a threshold chosen to ensure that, with high probability, the demand will be met even if units fail or the actual demand varies from the expected demand. The quantity $r_i(u_i^t)$ is the maximum power that a unit can provide within a specified, short interval of time (e.g., 10 min).

We assume for simplicity that we have

$$r_i(u_i^t) = g_i^e \qquad \text{if } u_i^t = 1 \tag{2.6a}$$

$$r_i(u_i^t) = 0 \qquad \text{if } u_i^t = 0 \tag{2.6b}$$

where g_i^e is the maximum emergency power available for unit i. A possible generalization which can be readily incorporated in our solution methodology is to let r_i be a piecewise linear concave function of g_i^t when $u_i^t = 1$—for example, $r_i^t(1, g_i^t) = \min\{g_i^t + g_i^\Delta, g_i^e\}$, where g_i^Δ is the maximum possible increase in power for unit i during the time interval associated with the reserve constraint (e.g., 10 min).

We can now state our problem as one of finding optimal control variables $(u, g) = \{(u_i^t, g_i^t) | i = 1, \cdots, I, \; t = 0, \cdots, N-1\}$ that minimize the total cost:

$$J(u, g) = \sum_{t=0}^{N-1} \sum_{i=1}^{I} \{C_i(g_i^t) + S_i(x_i^t, u_i^t)\} \tag{2.7}$$

subject to the system (2.1) and the constraints (2.3a)–(2.5). This problem involves $N \cdot I$ continuous variables (g_i^t) and $N \cdot I$ integer $0-1$ variables (u_i^t). For the typical value $N = 24$ and I being in the hundreds, we are faced with a mixed-integer programming problem with several thousands of integer variables and as many continuous variables. Because of such large dimensionality, it appears that a direct attack on this problem with standard methods (e.g., branch-and-bound) is hopeless. However, we can exploit the separable structure of this problem and pass into a dual problem that involves only $2N$ variables.

III. THE DUAL PROBLEM

By assigning nonnegative Lagrange multipliers λ^t and μ^t to the constraints (2.4) and (2.5), respectively, we can consider the corresponding dual functional

$$q(\lambda, \mu) = \min_{u, g} \left\{ J(u, g) + \sum_{t=0}^{N-1} \left[\lambda^t \left(D^t - \sum_{i=1}^{I} g_i^t \right) \right. \right.$$
$$\left. \left. + \mu^t \left(R^t - \sum_{i=1}^{I} r_i(u_i^t) \right) \right] \right\} \tag{3.1}$$

where the minimization is subject to the constraints (2.1), (2.3a), and (2.3b). In view of the separable structure of J [cf. (2.7)], we can also write q as

$$q(\lambda, \mu) = \sum_{i=1}^{I} q_i(\lambda, \mu) + \sum_{t=0}^{N-1} (\lambda^t D^t + \mu^t R^t) \tag{3.2}$$

where

$$q_i(\lambda, \mu) = \min_{u_i, g_i} \sum_{t=0}^{N-1} \left[C_i(g_i^t) + S_i(x_i^t, u_i^t) - \lambda^t g_i^t - \mu^t r(u_i^t) \right] \tag{3.3}$$

and the minimization in (3.3) is subject to the constraints

(2.1), (2.3a), and (2.3b). This minimization represents an optimal control problem involving a single unit and a state space with only two states which can be readily solved by dynamic programming. Thus the value $q(\lambda, \mu)$ for any given (λ, μ) can be obtained relatively easily. The dual problem is

$$\max q(\lambda, \mu)$$
$$\text{subject to } \lambda \geqslant 0, \qquad \mu \geqslant 0. \qquad (3.4)$$

It is well known that, if J^* is the optimal value of $J(u, g)$, we have $q(\lambda, \mu) \leqslant J^*$ for all $\lambda \geqslant 0$, $\mu \geqslant 0$. Thus, if we denote by q^* the dual optimal value, we have

$$q^* = \max_{\substack{\lambda \geqslant 0 \\ \mu \geqslant 0}} q(\lambda, \mu) \leqslant J^*. \qquad (3.5)$$

The difference $(J^* - q^*)$ is termed the *duality gap* and can be expected to be strictly positive in view of the fact that the constraint set for (u, g) is nonconvex. Our solution method consists of first solving the dual problem (3.4) and then using its solution to obtain a feasible solution $(\bar{u}, \bar{g})$ of the primal problem. The criterion for accepting $(\bar{u}, \bar{g})$ as the final solution will be the magnitude of the ratio

$$\frac{J(\bar{u}, \bar{g}) - q^*}{q^*}. \qquad (3.6)$$

For this approach to have a chance of success, it is necessary that the "relative duality gap," i.e., the ratio

$$\frac{J^* - q^*}{q^*} \qquad (3.7)$$

be small. We will show in Section VI via analysis and in Section VIII via computational experiment that indeed this ratio is relatively small and tends to become smaller as the number of units I increases.

IV. Solution of the Dual Problem

As mentioned earlier, it is possible to obtain the value of each component $q_i(\lambda, \mu)$ [cf. (3.3)] of the dual functional using dynamic programming. The corresponding equations for unit i and fixed $\lambda \geqslant 0$ and $\mu \geqslant 0$ are

$$J_i^N(x_i^N; \lambda, \mu) = 0 \qquad (4.1a)$$

$$J_i^t(x_i^t; \lambda, \mu) = \min_{(u_i^t, g_i^t) \in W_i} \{C_i(g_i^t) + S_i(x_i^t, u_i^t)$$
$$- \lambda^t g_i^t - \mu^t r_i(u_i^t) + J_i^{t+1}(u_i^t; \lambda, \mu)\} \qquad (4.1b)$$

where $t = 0, 1, \cdots, N-1$, $x_i^t \in \{0, 1\}$, and the constraint W_i is given by

$$W_i = (0,0) \cup \{(1, g_i^t) | \underline{g}_i \leqslant g_i^t \leqslant \bar{g}_i\}. \qquad (4.2)$$

The value $q_i(\lambda, \mu)$ is obtained from the last step of the algorithm as

$$q_i(\lambda, \mu) = J_i^0(x_i^0; \lambda, \mu). \qquad (4.3)$$

For $u_i^t = 1$ and $g_i^t \in [\underline{g}_i, \bar{g}_i]$, the function within braces in (4.1b) is a piecewise linear convex function of g_i^t with breakpoints denoted $g_i(1), \cdots, g_i(n_i)$ where $g_i(1) = \underline{g}_i$ and $g_i(n_i) = \bar{g}_i$. As a result, this function attains its minimum at one of these breakpoints. This leads to the following alternative form of (4.1b):

$$J_i^t(x_i^t; \lambda, \mu) = \min\{a_i^t(x_i^t; \lambda, \mu), b_i^t(x_i^t; \lambda, \mu)\} \qquad (4.4)$$

where

$$a_i^t(x_i^t; \lambda, \mu) = \min_{j=1, \cdots, n_i} \{C_i[g_i(j)] - \lambda^t g_i(j)\}$$
$$+ S_i(x_i^t, 1) - \mu^t g_i^e + J_i^{t+1}(1; \lambda, \mu) \qquad (4.5)$$

$$b_i^t(x_i^t; \lambda, \mu) = S_i(x_i^t, 0) + J_i^{t+1}(0; \lambda, \mu) \qquad (4.6)$$

and a_i^t, b_i^t are the "costs to go" corresponding to an up decision and down decision, respectively, for unit i at time t. It can be seen that, for each t and x_i^t, the function $J_i^t(x_i^t; \lambda, \mu)$ is piecewise linear and concave, and the same is true for the dual functional q which is given by [cf. (3.2), (4.3)]

$$q(\lambda, \mu) = \sum_{i=1}^{I} J_i^0(x_i^0; \lambda, \mu) + \sum_{t=0}^{N-1} (\lambda^t D^t + \mu^t R^t). \qquad (4.7)$$

Thus the dual problem of maximizing q subject to $\lambda \geqslant 0$, $\mu \geqslant 0$ is a linear programming problem. Its solution, however, by simplex-like methods seems hopelessly time consuming. For the purposes of obtaining an approximately optimal solution (which is sufficient for our purposes), it appears that nondifferentiable optimization methods are far preferable. One possibility is to use Shor's original subgradient method as in [1]. The advantage of this method over other competing methods [16]–[21] based on subgradients is that it does not require a line search or other operations which for the problem of this paper are unacceptably time consuming. Another possibility is to use an approximation method [8], [9] which has the advantage that, in addition to solving the dual problem, it also yields at no extra cost the solution to another problem which may be viewed both as "dual-to-the-dual" as well as a relaxed version of the original unit commitment problem.

The main idea of the approximation method we use is to recursively replace every function of the form

$$\min\{h_1(\lambda, \mu), \cdots, h_m(\lambda, \mu)\} \qquad (4.8)$$

in the dynamic programming algorithm, where $h_1, \cdots, h_m$ are some generic functions, by a function of the form

$$-\frac{1}{d} \log \left\{ \sum_{j=1}^{m} y_j e^{-dh_j(\lambda, \mu)} \right\} \qquad (4.9)$$

where d is a positive parameter and y_j are positive multipliers with $\sum_{j=1}^{m} y_j = 1$. Thus the dynamic programming algorithm [(4.4)–(4.6)] is replaced by the "approximate

dynamic programming algorithm"

$$\tilde{J}_i^N(x_i^N; \lambda, \mu; y, z, d) = 0 \qquad (4.10)$$

$$\tilde{J}_i^t(x_i^t; \lambda, \mu; y, z, d) = -\frac{1}{d_i^t}\log\Big\{ z_i^t(x_i^t) e^{-d_i^t \tilde{a}_i(x_i^t; \lambda, \mu; y, z, d)}$$
$$+ \big[1 - z_i^t(x_i^t)\big] e^{-d_i^t \tilde{b}_i(x_i^t; \lambda, \mu; y, z, d)} \Big\} \qquad (4.11)$$

where

$$\tilde{a}_i^t(x_i^t; \lambda, \mu; y, z, d)$$
$$= -\frac{1}{d_i^t}\log\Big\{ \sum_{j=1}^{n_i} y_{ij}^t e^{-d_i^t[C_i[g_i(j)] - \lambda' g_i(j)]} \Big\}$$
$$+ S_i(x_i^t, 1) - \mu^t g_i^e + \tilde{J}_i^{t+1}(1; \lambda, \mu; y, z, d) \qquad (4.12)$$

and

$$\tilde{b}_i^t(x_i^t; \lambda, \mu; y, z, d) = S_i(x_i^t, 0) + \tilde{J}_i^{t+1}(0; \lambda, \mu; y, z, d). \qquad (4.13)$$

Here y, z, and d are the vectors with components y_{ij}^t, $z_i^t(x_i^t)$, and d_i^t, $t = 0, \cdots, N-1$, $i = 1, \cdots, I$, $j = 1, \cdots, n_i$, $x_i^t \in \langle 0, 1 \rangle$, and for all t, i, j, and x_i^t, we require

$$0 < y_{ij}^t, \qquad 0 < z_i^t(x_i^t) < 1,$$
$$0 < d_i^t, \qquad \sum_{j=1}^{n_i} y_{ij}^t = 1. \qquad (4.14)$$

Finally, the "approximate dual problem" is given by

$$\max \tilde{q}(\lambda, \mu; y, z, d)$$
$$= \sum_{i=1}^{I} \tilde{q}_i(\lambda, \mu; y, z, d) + \sum_{t=0}^{N-1} (\lambda' D^t + \mu' R^t)$$
$$\text{subject to } \lambda \geqslant 0, \qquad \mu \geq 0 \qquad (4.15)$$

where

$$\tilde{q}_i(\lambda, \mu; y, z, d) = \tilde{J}_i^0(x_i^0; \lambda, \mu; y, z, d). \qquad (4.16)$$

Note that for each set of parameters (y, z, d), problem (4.15) is a twice-differentiable constrained optimization problem which can be solved by Newton-like methods.

Our method for maximizing the (exact) dual functional q consists of solving the approximate dual problem (4.15) for some initial set of parameters (y, z, d) and then updating these parameters according to the formulas

$$y_{ij}^t \leftarrow \bar{y}_{ij}^t, \quad z_i^t(x_i^t) \leftarrow \bar{z}_i^t(x_i^t), \quad d_i^t \leftarrow w d_i^t \qquad (4.17)$$

where w is a scalar greater than unity and $\bar{y}_{ij}^t$ and $\bar{z}_i^t(x_i^t)$ are given by

$$\bar{y}_{ij}^t = \frac{y_{ij}^t e^{-d_i^t[C_i[g_i(j)] - \lambda' g_i(j)]}}{\sum_{r=1}^{n_i} y_{ir}^t e^{-d_i^t[C_i[g_i(r)] - \lambda' g_i(r)]}} \qquad (4.18)$$

$$\bar{z}_i^t(x_i^t) = \Big(z_i^t(x_i^t) e^{-d_i^t \tilde{a}_i(x_i^t; \lambda, \mu; y, z, d)} \Big)$$
$$\cdot \Big(z_i^t(x_i^t) e^{-d_i^t \tilde{a}_i(x_i^t; \lambda, \mu; y, z, d)}$$
$$+ \big[1 - z_i^t(x_i^t)\big] e^{-d_i^t \tilde{b}_i(x_i^t; \lambda, \mu; y, z, d)} \Big)^{-1} \qquad (4.19)$$

and these expressions are evaluated at the solution (λ, μ) of the corresponding approximate dual problem. This procedure is repeated as many times as is necessary for the difference between the optimal value of the approximate dual functional and the corresponding value of the exact dual functional to become sufficiently small.

Convergence of this method to a solution of the dual problem can be readily shown based on the fact $d_i^t \to \infty$, provided a mechanism is incorporated that keeps all parameters $z_i^t(x_i^t)$, y_{ij}^t bounded away from zero. However, the method is capable of convergence even when d_i^t is kept constant on the power of the multiplier updating formulas (4.18), (4.19) alone. This is a generic property of the exponential method of multipliers on which the method is based (see [10], [14], and [15]). Typically, the number of approximate dual problems that need to be solved to obtain acceptable convergence is around four, and each one of these needs only to be solved approximately with the convergence tolerance becoming progressively more stringent with each new approximate dual problem.

The multipliers $\bar{y}_{ij}^t$, $\bar{z}_i^t(x_i^t)$ obtained via (4.18), (4.19) at the end of each maximization of the approximate dual functional form sequences which typically converge[1] to values $\hat{y}_{ij}^t$, $\hat{z}_i^t(x_i^t)$ satisfying

$$\hat{z}_i^t(x_i^t) = 1 \qquad \text{if } a_i^t(x_i^t; \hat{\lambda}, \hat{\mu}) > b_i^t(x_i^t; \hat{\lambda}, \hat{\mu}) \qquad (4.20)$$
$$\hat{z}_i^t(x_i^t) = 0 \qquad \text{if } a_i^t(x_i^t; \hat{\lambda}, \hat{\mu}) < b_i^t(x_i^t; \hat{\lambda}, \hat{\mu}) \qquad (4.21)$$
$$\hat{y}_{ij}^t > 0 \Rightarrow j = \arg \min_{j = 1, \cdots, n_i} \big\{ C_i[g_i(j)] - \hat{\lambda}' g_i(j) \big\} \qquad (4.22)$$

where $(\hat{\lambda}, \hat{\mu})$ is the maximizing vector of the dual functional obtained in the limit via the algorithm, and a_i^t, b_i^t are given by (4.5), (4.6).

Regarding the numerical solution of the approximate dual problem, we have been using a recently developed, quadratically convergent, constrained version of Newton's method [11] which makes use of the gradient and Hessian matrix of the approximate dual functional. As will be discussed in the next section, it is possible to compute recursively (and quite efficiently) these quantities, and this is another key factor for the success of the overall method. Typically, the number of Newton iterations needed to solve with sufficient accuracy the dual problem is around ten, and this number appears to be insensitive to the number of units I. This fact is the basis of our claim that the computa-

---

[1]It is only possible to show rigorously that all limit points of these sequences satisfy (4.20)–(4.22), but it appears that the possibility of the sequences $\langle \bar{y}_{ij}^t \rangle$ and $\langle \bar{z}_i^t(x_i^t) \rangle$ having more than one limit point is remote, and in any case, it has never been observed by us.

tional requirements for solving both the dual problem and (within a practically acceptable tolerance) the unit commitment problem grow linearly with the number of units.

V. GRADIENT COMPUTATION AND THE RELAXED PROBLEM

We first derive an expression for the gradient of the approximate dual functional with respect to (λ, μ). It is straightforward to verify from (4.10)–(4.13) that for each state x_i^t, we have

$$\frac{\partial \tilde{J}_i^t(x_i^t)}{\partial \lambda^\tau} = \frac{\partial \tilde{J}_i^t(x_i^t)}{\partial \mu^\tau} = 0 \qquad \text{if } \tau < t \tag{5.1}$$

$$\frac{\partial \tilde{J}_i^t(x_i^t)}{\partial \lambda^t} = -\bar{z}_i^t(x_i^t) \sum_{j=1}^{n_i} \bar{y}_{ij}^t g_i(j) \tag{5.2}$$

$$\frac{\partial \tilde{J}_i^t(x_i^t)}{\partial \mu^t} = -\bar{z}_i^t(x_i^t) g_i^e \tag{5.3}$$

$$\frac{\partial \tilde{J}_i^t(x_i^t)}{\partial \lambda^\tau} = \bar{z}_i^t(x_i^t) \frac{\partial \tilde{J}_i^{t+1}(1)}{\partial \lambda^\tau}$$
$$+ \left[1 - \bar{z}_i^t(x_i^t) \right] \frac{\partial \tilde{J}_i^{t+1}(0)}{\partial \lambda^\tau} \qquad \text{if } \tau > t \tag{5.4}$$

$$\frac{\partial \tilde{J}_i^t(x_i^t)}{\partial \mu^\tau} = \bar{z}_i^t(x_i^t) \frac{\partial \tilde{J}_i^{t-1}(1)}{\partial \mu^\tau}$$
$$+ \left[1 - \bar{z}_i^t(x_i^t) \right] \frac{\partial \tilde{J}_i^{t+1}(0)}{\partial \mu^\tau} \qquad \text{if } \tau > t \tag{5.5}$$

where $\bar{z}_i^t$ and $\bar{y}_{ij}^t$ are as in (4.18), (4.19) and for notational convenience, we have dropped all the arguments of $\tilde{J}_i^t$ except for x_i^t. From these equations, it is possible to calculate recursively all first partial derivatives of $\tilde{J}_i^0$ and hence also of the approximate dual functional $\tilde{q}$ [cf. (4.15), (4.16)].

An interesting and relevant observation is that the parameter $\bar{z}_i^t(x_i^t)$ may be viewed as a *transition probability* of going from state x_i^t to the up state 1 when the states of the unit are viewed as states of a (nonstationary) Markov chain, as shown in Fig. 3. With this interpretation, it is easily seen from (5.1)–(5.5) that for all i, t, and $\tau \geq t$, we have

$$\frac{\partial \tilde{J}_i^t(x_i^t)}{\partial \lambda^\tau} = - p_i(x_i^t; t, \tau) \sum_{j=1}^{n_i} \bar{y}_{ij}^\tau g_i(j) \tag{5.6}$$

$$\frac{\partial \tilde{J}_i^t(x_i^t)}{\partial \mu^\tau} = - p_i(x_i^t; t, \tau) g_i^e \tag{5.7}$$

where $\bar{y}_{ij}^\tau$ is given by (4.18), and $p_i(x_i^t; t, \tau)$ is the probability that an up decision for unit i will be made at time τ ($u_i^\tau = 1$) when the state at time t is x_i^t. Thus the gradient of q_i can be calculated by propagating forward the state probability vector as in
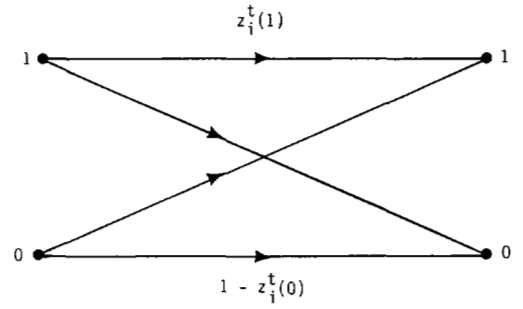
$$p_i^{t+1} = p_i^t P_i^t \tag{5.8}$$

Fig. 3. Transition probability graph for the relaxed problem.

where p_i^t is the row vector of probability distribution of x_i^t given the initial distribution p_i^0 where

$$p_i^0 = (1, 0) \qquad \text{if } x_i^0 = 1 \text{ or} \tag{5.9a}$$

$$p_i^0 = (0, 1) \qquad \text{if } x_i^0 = 0 \tag{5.9b}$$

and P_i^t is the transition probability matrix

$$P_i^t = \begin{bmatrix} \bar{z}_i^t(1) & 1 - \bar{z}_i^t(1) \\ \bar{z}_i^t(0) & 1 - \bar{z}_i^t(0) \end{bmatrix}. \tag{5.10}$$

A further differentiation of the expressions $\partial[\tilde{J}_i^0(x_i^0)]/\partial \lambda^\tau$, $\partial[\tilde{J}_i^0(x_i^0)]/\partial \mu^\tau$, $\tau > 0$ as given by (5.6), (5.7) yields an expression for the Hessian matrix of $\tilde{q}_i$. The details for this are straightforward but lengthy and will not be given (see [13]). We mention that it is possible to develop a recursive scheme for computing this Hessian matrix by taking advantage of the transition probability graph. It is also possible to show that each Hessian matrix $\nabla^2 \tilde{q}_i$ is negative definite (see [13]). It follows from (4.15) that $\nabla^2 \tilde{q}$ is also negative definite which is an essential requirement for good performance of the Newton-like method that we have been using for maximization of $\tilde{q}$.

The connection with the probabilistic state transition graph of Fig. 3 can be carried further by considering the following *relaxed version of the unit commitment problem.*

Consider the unit commitment problem of Section II, where at each time period t and state $\{x_i^t | i = 1, \cdots, I\}$, we choose for each unit i (instead of $0 - 1$ decision variables u_i^t and generation variables g_i^t) a probability $z_i^t(x_i^t)$ according to which $u_i^t = 1$, and probabilities y_{ij}^t according to which $g_i^t = g_i(j)$ when $u_i^t = 1$. In other words, the decision variables in the relaxed problem are

$$z_i^t(x_i^t) = P\{u_i^t = 1 | x_i^t\} \tag{5.11a}$$

$$y_{ij}^t = P\{g_i^t = g_i(j) | u_i^t = 1\}. \tag{5.11b}$$

The problem is to find $z_i^t(x_i^t)$ and y_{ij}^t that solve the problem

$$\min \hat{J}(y, z) = E\left\{ \sum_{t=0}^{N-1} \sum_{i=1}^{I} \left[C_i(g_i^t) + S_i(x_i^t, u_i^t) \right] \right\} \tag{5.12}$$

subject to $E\left\{ \sum_{i=1}^{I} g_i^t \right\} \geq D^t, \qquad t = 0, 1, \cdots, N-1$

$$\tag{5.13}$$

$$E\left\{\sum_{i=1}^{I} r_i(u_i^t)\right\} \geqslant R^t, \qquad t = 0, 1, \cdots, N-1$$

(5.14)

where $E\{\cdot\}$ denotes expected value, and the probabilistic evolution of x_i^t is specified by the transition probability graph of Fig. 3.

This problem is the same as the unit commitment problem except that we have expanded the set of feasible decisions to allow for *randomized decisions*.

Suppose that we have obtained an optimal solution $(\hat{\lambda}, \hat{\mu})$ of the dual problem (3.4) via the approximation method of the previous section, and a corresponding set of multipliers $(\hat{y}, \hat{z})$ satisfying (4.20)–(4.22). We will show that $(\hat{y}, \hat{z})$ *is a solution of the relaxed unit commitment problem*.

Indeed, let $\langle(\lambda_k, \mu_k)\rangle$ and $\langle(y_k, z_k, d_k)\rangle$ be the sequences generated by the approximation method. Then (λ_k, μ_k) solves the approximate dual problem corresponding to (y_k, z_k, d_k), so from a necessary condition for optimality, we obtain

$$\frac{\partial \tilde{q}}{\partial \lambda^t} \geqslant 0, \quad \frac{\partial \tilde{q}}{\partial \mu^t} \geqslant 0, \quad t = 0, \cdots, N-1$$

(5.15)

$$\lambda_k^t > 0 \Rightarrow \frac{\partial \tilde{q}}{\partial \lambda^t} = 0$$

(5.16)

$$\mu_k^t > 0 \Rightarrow \frac{\partial \tilde{q}}{\partial \mu^t} = 0$$

(5.17)

where all partial derivatives above are evaluated at $(\lambda_k, \mu_k, y_k, z_k, d_k)$. Using the expression for $\partial \tilde{q}/\partial \lambda^t$, $\partial \tilde{q}/\partial \mu^t$ obtained via (4.15), (4.16), (5.6), and (5.7), and the fact $(\lambda_k, \mu_k) \to (\hat{\lambda}, \hat{\mu})$, $(y_k, z_k) \to (\hat{y}, \hat{z})$, we obtain from (5.15)

$$\sum_{i=1}^{I} p_i(x_i^0; 0, t) \sum_{j=1}^{n_i} \hat{y}_{ij}^t g_i(j) - D^t \geqslant 0$$

(5.18)

$$\sum_{i=1}^{I} p_i(x_i^0; 0, t) g_i^e - R^t \geqslant 0$$

(5.19)

where $p_i(x_i^0; 0, t)$ is the probability that the decision made for unit i at time t corresponding to the randomized decision variables $(\hat{y}, \hat{z})$ will be to have the unit up ($u_i^t = 1$). Thus (5.18), (5.19) simply say that $(\hat{y}, \hat{z})$ satisfies the constraints (5.13), (5.14) of the relaxed problem. Similarly, (5.16), (5.17) can be written as

$$\hat{\lambda}^t > 0 \Rightarrow E\left\{\sum_{i=1}^{I} g_i^t\right\} = D^t$$

(5.20)

$$\hat{\mu}^t > 0 \Rightarrow E\left\{\sum_{i=1}^{I} r_i(u_i^t)\right\} = R^t.$$

(5.21)

Finally, if we form the Lagrangian function of the relaxed problem

$$\hat{J}(y, z) + \sum_{t=0}^{N-1}\left[\hat{\lambda}^t\left(D^t - E\left\{\sum_{i=1}^{I} g_i^t\right\}\right) \right.$$

$$\left. + \hat{\mu}^t\left(R^t - E\left\{\sum_{i=1}^{I} r_i(u_i^t)\right\}\right)\right], \quad (5.22)$$

we can verify easily using (4.20)–(4.22) that it is minimized at $(\hat{y}, \hat{z})$. It follows from a well-known saddle-point theorem (see [5, pp. 144–145]) that $(\hat{y}, \hat{z})$ is an optimal solution to the relaxed problem, and furthermore the optimal value of this problem is equal to the dual optimal value q^*.

It is important to note that to find a solution of the relaxed problem, it is not sufficient to find a solution $(\hat{\lambda}, \hat{\mu})$ of the dual problem and then minimize the Lagrangian function (5.22) with respect to (y, z). This is due to the fact that minimizing pairs (y, z) of (5.22) need not satisfy the conditions (5.18)–(5.21). Thus solving the dual problem is by itself of little help in solving the relaxed problem. The point of view advanced in this paper is that the solution of the relaxed problem is far more important than the solution of the dual problem (see Section VII). Therefore, a dual problem solution method such as the subgradient method used in [1], which does not simultaneously solve the relaxed problem, is inadequate for our purposes, while the approximation method that we have been using is fully satisfactory.

VI. AN ESTIMATE OF THE DUALITY GAP

In this section, we reformulate the relaxed version of the unit commitment problem as a linear programming problem. We then utilize basic linear programming results to show that there exists an optimal solution in which the number of units which have "relaxed" controls (i.e., unit commitment decisions that are noninteger) is at most equal to the number of demand and reserve constraints. This result is used to derive a bound on the duality gap associated with dualizing the unit commitment problem. This is done by bounding the increase in cost associated with modifying the relaxed solution into a feasible solution.[2] Finally, since this increase in cost is independent of the number of units, we show that the duality gap goes to zero (in relative terms) as the number of units increases. An alternative and more general method for estimating the duality gap is given in [24].

To reformulate the relaxed problem as a linear programming problem, we show that selecting the $z_i^t(x_i^t)$ and the y_{ij}^t is equivalent to assigning a probability to every possible sequence of controls (u_i^t, g_i^t) and that the cost and constraints can be formulated in terms of these sequences.

Let us enumerate all possible sequences of controls for each unit (this can be done since the values of u_i^t are zero

---

[2] The term "feasible" when applied to a unit means that the $u_i^t(x_i^t)$ must be zero or one; when applied to the unit commitment schedule it means that each unit's schedule is feasible and that all constraints are satisfied.

and one, and the only values we allow g_i^t to assume are those of the breakpoints $g_i(j)$) and denote the kth such sequence by w_i^k. Define

$$C_{ik} = \sum_{t=0}^{N-1} \left\{ C_i(g_i^t) + S_i(x_i^t, u_i^t) \right\} \Big|_{(g_i^t, u_i^t) = w_i^k}, \quad (6.1)$$

$$g_{ik}^t = g_i^t|_{(g_i^t, u_i^t) = w_i^k}, \quad \text{and} \quad (6.2)$$

$$r_{ik}^t = r_i^t(u_i^t)|_{(g_i^t, u_i^t) = w_i^k}. \quad (6.3)$$

Assign a "probability" p_{ik} to each sequence w_i^k so that

$$\sum_k p_{ik} = 1. \quad (6.4)$$

If the $z_i^t(x_i^t)$ and the y_{ij}^t have been selected, then the p_{ik} assigned to each sequence can be computed from

$$p_{ik} = \prod_{t=0}^{N-1} z_i^t(x_i^t) y_{ij}^t \Big|_{(g_i^t, u_i^t) = w_i^k} \quad (6.5)$$

where the x_i^t correspond to the control sequence w_i^k and y_{ij}^t is defined to be one if u_i^t is zero.

Equivalently, if the p_{ik} are assigned to each sequence, we can compute a consistent set of $z_i^t(x_i^t)$ and y_{ij}^t. This can be done recursively by computing the $z_i^0(x_i^0)$ first, then the y_{ij}^0, then $z_i^1(x_i^1)$, etc. To compute $z_i^0(x_i^0)$, simply add all the p_{ik} which correspond to trajectories leaving x_i^0 via a decision of $u_i^0 = 1$. If $z_i^0(x_i^0) = 0$, then the y_{ij}^0 can be assigned arbitrarily; otherwise they can be computed by summing over all trajectories leaving x_i^0, having $u_i^0 = 1$, and having $g_i^t = g_i(j)$, and then normalizing by the probability $z_i^0(x_i^0)$.

Similarly, $z_i^t(x_i^t)$ can be computed for any state which can be reached via nonzero probability decisions: sum over the trajectories leading through that state and decision, and normalize by the probability of being in that state. This procedure leads to a set of $z_i^t(x_i^t)$ and y_{ij}^t which are consistent with any assignment of p_{ik}. The conclusion is that selecting the p_{ik} is equivalent to selecting the $\{z_i^t(x_i^t), y_{ij}^t\}$.

Thus the relaxed unit commitment problem can be written as

$$\min_{p_{ik}} \sum_{i=1}^{I} \sum_k p_{ik} C_{ik} \quad (6.6)$$

$$\text{subject to } \sum_k p_{ik} = 1, \quad i = 1, \cdots, I \quad (6.7)$$

$$p_{ik} \geq 0 \quad (6.8)$$

$$\sum_{i=1}^{I} \sum_k p_{ik} g_{ik}^t \geq D^t, \quad t = 0, \cdots, N-1 \quad (6.9)$$

$$\sum_{i=1}^{I} \sum_k p_{ik} r_{ik}^t \geq R^t, \quad t = 0, \cdots, N-1. \quad (6.10)$$

This reformulation is clearly a linear programming problem with $I + 2N$ constraints.

If $I > 2N$, then (compare with [5, p. 172]) *any basic feasible solution of the above linear program has the property*

that at least $I - 2N$ of the indexes i have precisely one p_{ik} positive (and hence unity). To see this, note that any basic solution has at most $I + 2N$ variables positive. Each of the I constraints (6.7) requires at least one positive variable to satisfy it. The remaining $2N$ basic variables are then associated with at most $2N$ different units, leaving at most $(I - 2N)$ units i with exactly one positive variable p_{ik}. If we consider the set of multipliers $\{z_i^t(x_i^t), y_{ij}^t\}$ corresponding to a basic feasible solution $\{p_{ik}\}$ as discussed earlier, we also find, based on the argument above, that at most $2N$ of the multipliers $z_i^t(x_i^t)$, which correspond to states x_i^t along which some variable p_{ik} is positive, are neither zero or one.

If there exists any feasible solution to the linear program (6.6)–(6.10) (i.e., if operating all units at maximum output will satisfy demand and reserve constraints), then by noting that every feasible solution has finite cost, we can apply the fundamental theorem of linear programming to conclude that there exists an optimal basic solution to the linear program above.

Therefore it follows that there exists an optimal solution to the relaxed problem for which the schedules associated with $I - 2N$ units are feasible with respect to the original unit commitment problem (i.e., for which the values of u_i^t are zero or one). The important point to note is that *at most $2N$ units have schedules which do not meet the original integrality constraints and this bound is independent of the number of units.* In fact, at most $2N$ multipliers $z_i^t(x_i^t)$ in an optimal solution to the relaxed unit commitment problem are noninteger, and thus a feasible solution to the unit commitment problem can be obtained by modifying the schedules of at most $2N$ units.

Since the cost of this feasible solution overbounds the optimal unit commitment cost, and since the difference in the cost of the feasible and optimal relaxed solutions is constant, we can compute a bound on the duality gap which is independent of the number of units. The cost of the optimal unit commitment schedule grows as the number of units increases. Thus the duality gap, in relative terms, goes to zero as the number of units increases.

To determine the duality gap more precisely, we need to overbound the cost of modifying $2N$ relaxed schedules into $2N$ feasible schedules. Clearly this cost is bounded by $2N$ times the most expensive cost of modifying any one schedule. In Fig. 4, we plot the costs associated with operating a real unit (solid line) and the relaxed cost of operation (dashed line). Note that a demand and reserve feasible schedule which satisfies unit constraints can be obtained by turning the unit on as soon as possible and operating it at either the minimum output (if $g_i^t \in [0, g_i]$) or at g_i^t (if $g_i^t \in [g_i, \bar{g}_i]$). The maximum cost associated with this modification, since the generation cost is convex and increasing on $[g_i, \bar{g}_i]$, is N times that associated with producing the minimum output instead of $\epsilon > 0$, where ϵ is very small, plus the cost of starting up the unit.

If we define

$$S^* = \max_i S_i(0, 1) \quad (6.11)$$

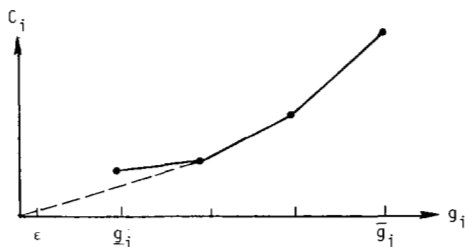$$C^* = \max_i \{ C_i(g_i) \}, \quad (6.12)$$

Fig. 4. Real and relaxed unit operation costs.

then the cost of modifying any unit schedule is less than $S^* + NC^*$ and the cost of the feasible schedule is less than

$$\hat{q} = q^* + 2N(S^* + NC^*) \qquad (6.13)$$

where q^* is the cost of the relaxed (or dual) solution. The relative duality gap is thus bounded by

$$\frac{\hat{q} - q^*}{q^*} = \frac{2N(S^* + NC^*)}{q^*} \qquad (6.14)$$

and since

$$\lim_{I \to \infty} q^* = + \infty, \qquad (6.15)$$

we have that the relative duality gap $(J^* - q^*)/q^*$ [cf. (3.7)] goes to zero as the number of units I increases.

VII. Obtaining a Feasible Unit Commitment Schedule

The preceding analysis has established that by solving the dual problem using the approximation method of Section IV, we obtain an optimal solution $(\hat{y}, \hat{z})$ of the relaxed unit commitment problem and that the corresponding optimal value is equal to the dual optimal value q^*. We know from the analysis of Section VI that the duality gap is expected to be small and in fact becomes smaller in relative terms as the number of units increases. Therefore it follows that the difference in the optimal values of the unit commitment problem and its relaxed version is small. Thus if we were able to perturb slightly the relaxed optimal solution $(\hat{y}, \hat{z})$ and obtain a feasible unit commitment schedule, then we could expect that this schedule would be close to being optimal. We next observe that once the decision variables u_i^t are selected, then the up or down status of all units at each time is determined and the generation variables g_i^t can be chosen optimally by means of a simple standard economic dispatch calculation. Thus the variables $\hat{y}_{ij}^t$ are not needed to provide near optimal values for the generation variables g_i^t.

There are two issues that arise in selecting the u_i^t on the basis of the solution to the relaxed unit commitment problem [i.e., on the basis of the $\hat{z}_i^t(x_i^t)$]. The first is that, since we do not solve the dual problem exactly, many of the $\hat{z}_i^t(x_i^t)$ may not have converged to zero or one, so there arises a question of deciding whether a particular unit i is "relaxed" or not at any particular time t on the basis of approximately optimal values of $z_i^t(x_i^t)$. The second is that of how to modify a "relaxed" unit schedule. These issues

are resolved by treating a value of $\hat{z}_i^t(x_i^t)$ which is within some small ϵ of zero (one) as indicating that if unit i is in state x_i^t at time t, then it should be up (down) in the next time period. Typically, less than $2N$ units are "relaxed" and for the remaining units, the $\hat{z}_i^t(x_i^t)$ have converged to within 0.001 of zero or one by the time we have a satisfactory solution to the dual problem. An exception is when there is a class of units which are identical, in which case if one of these units is "relaxed," all are "relaxed" and a large number of $\hat{z}_i^t(x_i^t)$ may be far from zero or one. The solution to this type of problem falls under the second issue.

We modify a "relaxed" unit schedule into a feasible schedule by setting $u_i^t(x_i^t)$ to one (zero) if $z_i^t(x_i^t)$ is over (under) a selected threshold. This threshold is selected so that the power generated by the units turned on is roughly equal to the "expected" power generated by these units in the relaxed solution. Note that this strategy handles classes of identical "relaxed" units in an intuitive manner: it turns on that fraction of units indicated by the $\hat{z}_i^t(x_i^t)$.

VIII. Computational Results

A more sophisticated version of the preceding algorithm has been implemented in a Fortran code[3] and was used to generate the results of this section. The implemented algorithm differs from the one discussed in this paper in two ways. First, we allow the imposition of minimum up and down time constraints on the operation of generators. The implemented algorithm produces a unit commitment schedule which satisfies all these constraints. Second, the implemented algorithm includes a variety of safeguards (e.g., against the Hessian becoming indefinite due to finite precision arithmetic) and rules (e.g., step-size selection, convergence criteria for the approximate problems, etc.) which speed up convergence to the solution.

These safeguards and rules introduce parameters which allow one to tune the algorithm to the problem at hand. However, the overall method is quite insensitive to variations in these parameters within a broad range. It should also be noted that the same parameters have been used for each combination of demand curve and generator set in Table I and that the same initial set of dual multipliers have been used for each example.

The data describing the generators and demand curves were taken from an EPRI report [22]. The units range in size from 50 MW to 1200 MW and include combustion turbine, oil- and coal-fired, and nuclear units. The demand curves for summer and winter (Figs. 5 and 6, respectively) have been scaled so that the peak demand is 0.7, 0.8, or 0.9 of the total generator capacity.

This information, denoted D/G, and the number of units in each example is given in the two leftmost columns of Table I. The top two rows of Table I indicate the number of time periods in the example (12 two-hour periods or 24 one-hour periods) as well as the type of

---

[3] On a VAX-11/780 without a floating-point accelerator.

TABLE I
SUMMARY OF COMPUTATIONAL RESULTS

| No. of Units	D/G	12 Time Periods		24 Time Periods	
		Summer	Winter	Summer	Winter
30	0.7	12, 0.32%, :36	9, 0.35%, :29	12, 0.38%, 1:48	11, 0.25%, 1:35
	0.8	12, 0.60%, :34	9, 0.35%, :27	10, 0.31%, 1:32	13, 0.15%, 1:57
	0.9	7, 0.31%, :22	9, 0.12%, :26	16, 0.22%, 2:13	10, 0.19%, 1:27
60	0.7	10, 0.29%, 1:03	12, 0.19%, 1:14	10, 0.31%, 2:55	12, 0.29%, 3:18
	0.8	10, 0.17%, 1:02	11, 0.16%, 1:06	8, 0.41%, 2:25	17, 0.16%, 4:32
	0.9	9, 0.31%, :55	7, 0.13%, :48	16, 0.13%, 4:16	18, 0.21%, 4:36
120	0.7	10, 0.08%, 2:07	8, 0.17%, 1:41	11, 0.24%, 6:28	9, 0.24%, 5:21
	0.8	9, 0.26%, 1:50	8, 0.34%, 1:42	9, 0.17%, 4:57	14, 0.11%, 7:26
	0.9	7, 0.31%, 1:34	8, 0.26%, 1:40	15, 0.09%, 7:50	14, 0.10%, 7:21
200	0.7	8, 0.31%, 2:48	7, 0.09%, 2:38	12, 0.10%, 10:57	13, 0.27%, 11:49
	0.8	11, 0.30%, 3:43	10, 0.21%, 3:30	12, 0.26%, 11:02	17, 0.09%, 12:12
	0.9	9, 0.06%, 3:08	8, 0.06%, 2:50	12, 0.47%, 10:43	11, 0.15%, 9:48
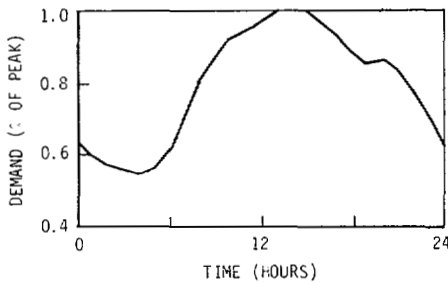
Fig. 5. Summer demand curve.
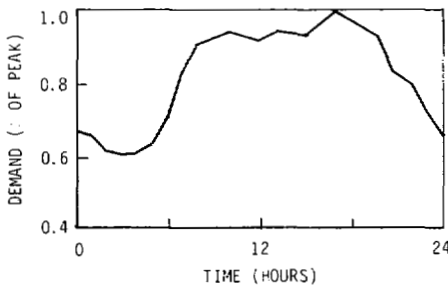
Fig. 6. Winter demand curve.

demand curve (i.e., summer or winter). Each entry in the table is a triple of numbers: the number of Newton steps required to solve the problem, an upper bound on suboptimality of the unit commitment schedule as given by $(\hat{J} - \hat{q})/\hat{q}$ where $\hat{J}$ is the cost of the feasible schedule computed and $\hat{q}$ is the highest value of the dual functional obtained, and the CPU time required to solve the problem (min : s).

From this table we conclude that for large problems (over 100 units), we typically require 8 Newton steps or less and that the CPU time required to solve large problems (with 24 time steps) with this experimental (and as yet not optimized) code is roughly 4–6 min/100 units. The number of Newton steps required is an important quantity, since the computation of the Hessian matrix of the approximate dual functional dominates the computation time. This number ranges from 3–14 steps and varies mainly

with the number of approximate dual problems that need to be solved (3 steps for 3 problems, 5–8 steps for 4 problems, and 9–14 steps for 5 approximate problems). The variation of computation time with problem data is quite small, however, and is not a cause for concern.

Table I also indicates the dependence of the computation time on problem size: roughly linear in the number of units and between linear and quadratic in the number of time steps. The dependence of the duality gap on the problem size is somewhat less clear, but does roughly decrease with increasing numbers of units or time steps. Actually, the percentages shown in Table I are rather loose upper bounds on the actual value of the ratio $(\hat{J} - J^*)/J^*$ which characterizes the degree of suboptimality of the feasible suboptimal schedule obtained. This is true not only because of the presence of a duality gap, but also because the dual functional q is maximized only approximately, and therefore the best dual value of $\hat{q}$ obtained is lower than the maximal value q^* (and *a fortiori* J^*). We found by experimentation that the difference $(q^* - \hat{q})$ can account for a substantial portion of the percentages shown in Table I.

The success achieved with our methodology in solving the optimal unit commitment problem is apparent from Table I. Even the smallest problem in the table (30 units, 12 time periods) is twice as large as the largest optimal unit commitment problem that to our knowledge has ever been solved using other methods [1].

IX. SUMMARY AND EXTENSIONS

In this paper, we have introduced an algorithm for the solution of large (100+) unit commitment problems and shown that it can compute very accurate solutions (typically within 0.5 percent of optimal) very quickly (typically 4–6 min of CPU time per 100 units for 24-time-period problems). While the algorithm presented in Sections II–VII is valid only for state transition diagrams involving

two states, the algorithm implemented allows for the inclusion of minimum up and down time constraints.

The algorithm is also applicable to scheduling power systems involving both thermal and hydro units. For this purpose, a preliminary optimization phase is necessary to determine an optimal hydro generation schedule which is then used to determine the effective demand profile for the thermal units. A separate paper deals with the details of this procedure [23].

REFERENCES

[1] J. A. Muckstadt and S. A. Koenig, "An application of Lagrangian relaxation to scheduling in power-generation systems," *Operations Res.*, vol. 25, pp. 387–403, 1977.

[2] J. Gruhl, F. Schweppe, and M. Ruane, "Unit commitment scheduling of electric power systems," in *Systems Engineering for Power: Status and Prospects, Eng. Foundation Conf.*, L. H. Fink and K. Carlsen, Eds., Henniker, NH, 1975.

[3] A. Geoffrion, "Lagrangian relaxation for integer programming," *Math. Programming Study*, vol. 2, pp. 82–114, 1974.

[4] L. S. Lasdon, *Optimization Theory for Large Systems*. New York: Macmillan, 1970.

[5] J. E. Shapiro, *Mathematical Programming: Structures and Algorithms*. New York: Wiley, 1979.

[6] B. T. Poljak, "Minimization of unsmooth functionals," *USSR Comput. Math. Math. Phys.*, vol. 9, pp. 14–29, 1969.

[7] M. Held, P. Wolfe, and H. Crowder, "Validation of subgradient optimization," *Math. Program.*, vol. 6, pp. 62–88, 1974.

[8] D. P. Bertsekas, "Nondifferentiable optimization via approximation," *Math. Programming Study*, vol. 3, pp. 1–25, 1975.

[9] ——, "Approximation procedures based on the method of multipliers," *J. Optimiz. Theory Appl.*, vol. 23, pp. 487–510, 1977.

[10] B. W. Kort and D. P. Bertsekas, "A new penalty function method for constrained minimization," in *Proc. 1972 IEEE Conf. Decision Contr.*, New Orleans, LA, 1972, pp. 162–166.

[11] D. P. Bertsekas, "Projected Newton methods for optimization problems with simple constraints," Lab. Inform. Decision Syst., M.I.T., Cambridge, MA, Rep. R-1025, Aug. 1980; see also *SIAM J. Contr. Optimiz.*, vol. 20, no. 2, pp. 221–246, 1982.

[12] G. S. Lauer, D. P. Bertsekas, N. R. Sandell, Jr., and T. A. Posbergh, "Solution of large-scale optimal unit commitment problems," *IEEE Trans. Power App. Syst.*, vol. PAS-101, pp. 74–86, 1982.

[13] G. S. Lauer, D. P. Bertsekas, and N. R. Sandell, Jr., "Optimal stochastic operations scheduling first annual report," TR-108, Alphatech, Inc., Burlington, MA, Rep. TR-108, Sept. 1980.

[14] V. H. Nguyen and J. J. Strodiot, "On the convergence rate of a penalty function method of exponential type," *J. Optimiz. Theory Appl.*, vol. 27, pp. 495–508, 1979.

[15] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*. New York: Academic, 1982.

[16] D. P. Bertsekas and S. K. Mitter, "A descent numerical method for optimization problems with nondifferentiable cost functionals," *SIAM J. Contr. Optimiz.*, vol. 11, pp. 637–652, 1973.

[17] C. Lemarechal, "An algorithm for minimizing convex functions," in *Proc. IFIP Congr. 1974*, J. L. Rosenfeld, Ed. Amsterdam, The Netherlands: North-Holland, 1974, pp. 552–557.

[18] N. Z. Shor, "Convergence of a gradient method with space dilation in the direction of the difference between two successive gradients," *Cybernetics*, vol. 11, pp. 564–570, 1975.

[19] R. Mifflin, "An algorithm for constrained optimization with semismooth functions," *Math. Operations Res.*, vol. 2, pp. 191–207, 1977.

[20] P. Wolfe, "A method of conjugate subgradients for minimizing nondifferentiable functions," *Math. Programming Study*, vol. 3, pp. 145–173, 1975.

[21] A. A. Goldstein, "Optimization of Lipschitz continuous functions," *Math. Programming*, vol. 13, pp. 14–22, 1977.

[22] H. W. Zaininger, A. J. Wood, H. K. Clark, T. F. Laskowski, and J. D. Burns, "Synthetic electric utilities systems for evaluating advanced technologies," Final Rep. EPRI-EM-285, 1977.

[23] N. R. Sandell, Jr., D. P. Bertsekas, J. J. Shaw, S. W. Gully, and R. Gendron, "Optimal scheduling of large scale hydrothermal power systems," in *Proc. IEEE Large Scale Syst. Conf.*, Virginia Beach, VA, Oct. 1982.

[24] D. P. Bertsekas and N. R. Sandell, Jr., "Estimates of the duality gap for large scale separable integer programming problems," in *Proc. 1982 IEEE Conf. Decision and Contr.*, Miami Beach, FL, Dec. 1982.

Dimitri P. Bertsekas (S'70–SM'77), for a photograph and biography, see p. 616 of the June 1982 issue of this TRANSACTIONS.

Gregory S. Lauer (S'77–M'79) was born in Boston, MA, in 1953. He received the B.S.E.E. and M.S.E.E. degrees from the University of Notre Dame, IN, in 1975 and 1976, and the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, in 1979.

He has worked at Alphatech, Inc., Burlington, MA, since 1979. He has done research and is interested in the areas of optimization, estimation, detection, and communication.

Dr. Lauer is a member of Eta Kappa Nu and Sigma Xi.

Nils R. Sandell, Jr. (S'70–M'74), for a photograph and biography, see p. 838 of the August 1982 issue of this TRANSACTIONS.
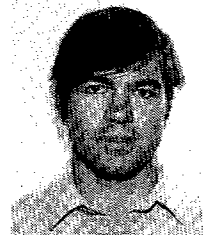
Thomas A. Posbergh (S'74–M'78–M'79) was born in Elizabeth, NJ, on April 24, 1954. He received the B.S. degree from New Jersey Institute of Technology, Newark, in 1976 and the S.M. degree from the Massachusetts Institute of Technology, Cambridge, in 1979, both in electrical engineering.

From 1976 to 1979 he was associated with the Charles Stark Draper Laboratory, Cambridge, MA. From 1979 to 1981 he was a Staff Engineer with Alphatech, Inc., Burlington, MA. At present he is employed by the U.S. Army Armament Research and Development Command, Dover, NJ. His current interests include decision making under uncertainty, stochastic systems, numerical methods, and microprocessor implementation of control algorithms.