

Exploiting Soft Constraints within Decomposition and Coordination Methods for Sub-hourly Unit Commitment

Niranjan Raghunathan, Mikhail A. Bragin, *Member, IEEE*, Bing Yan, *Member, IEEE*, Peter B. Luh, *Life Fellow, IEEE*, Khosrow Moslehi, *Member, IEEE*, Xiaoming Feng, *Member, IEEE*, Yaowen Yu, *Member, IEEE*, Chien-Ning Yu, *Member, IEEE*, and Chia-Chun Tsai

Abstract--Unit commitment (UC) is an important problem solved on a daily basis with a strict time limit. While hourly UC problems are currently used, they may not be flexible enough with the fast-changing demand and the increasing penetration of intermittent renewables. Sub-hourly UC is therefore recommended. This, however, will significantly increase problem complexity even under the deterministic setting, and current methods may not be able to obtain good solutions within the time limit. In this paper, deterministic sub-hourly UC is considered with an innovative exploitation of “soft constraints” - constraints that do not need to be strictly satisfied but their violations are penalized by predetermined coefficients. Based on the “surrogate optimization” concept, there is no need to optimally solve the “relaxed problem,” and the convergence of multipliers is guaranteed as long as the “surrogate optimality condition” is satisfied. Consequently, subproblems can be formed and efficiently optimized when soft constraints are not relaxed. The overall scheme leads to a drastic reduction of the number of multipliers, decreased computational requirements, and improved performance. To further enhance the method, a parallel version is developed. Testing results based on the Polish system demonstrate the effectiveness and robustness of both the sequential and parallel versions at finding high-quality solutions within the time limit.

Index Terms-- Sub-hourly Unit Commitment, Soft Constraints, Surrogate Lagrangian Relaxation (SLR), Surrogate Absolute-Value Lagrangian Relaxation (SAVLR), Parallel Algorithms

I. INTRODUCTION

Unit commitment (UC) problems are solved by Independent System Operators (ISOs) on a daily basis to determine generators’ commitment statuses and their generation levels to minimize the total cost subject to system demand, reserve, transmission capacity, and unit-level constraints. The problem is generally formulated in a mixed-integer linear programming (MILP) form, with a piece-wise linear cost function, linear constraints, and both binary and continuous decision variables. There are variations in how constraints are formulated. When constraints are required to be strictly satisfied (e.g., system demand), they are “hard” constraints.

Otherwise, they are “soft” constraints (e.g., reserve and transmission capacity constraints), where violations incur penalties with predetermined coefficients [1]. Such problems are solved with a strict time limit (e.g., 15 or 20 minutes).

In the current practice, hourly intervals are used with a typical planning horizon of 24 hours. With the fast-changing demand and the increasing penetration of intermittent renewables, high sub-hourly variations cannot be adequately accounted for. Hourly UC may thus lead to expensive solutions and result in renewable curtailment or load shedding [2]. Sub-hourly UC, e.g., with 15 minutes as the time interval, is therefore recommended for higher flexibility and efficiency.

Although sub-hourly UC has been studied, e.g., in [3]–[5], its major difficulty is the increased number of binary variables as compared to that of hourly UC, roughly quadrupling for the 15-minute case with significantly increased complexity. Also, a unit’s ramp rate per interval is much smaller than that of the hourly UC, rendering more ramping constraints active for intervals with fast changing demand and renewable generation, and resulting in increased computational requirements. Consequently, sub-hourly UC is difficult to solve even under the deterministic setting, and current methods may not be able to obtain near-optimal solutions within the required time limit.

In this paper, deterministic sub-hourly UC with 15-minute time intervals is considered, and the goal is to obtain near-optimal solutions within 20 minutes. After reviewing the literature in Section II, the problem formulation is presented in Section III, with system demand modeled as “hard” constraints and reserve requirements and transmission capacities modeled as “soft” constraints.

In Section IV, soft constraints are innovatively exploited within our recent decomposition and coordination framework of Surrogate Absolute-Value Lagrangian Relaxation (SAVLR) [6]. This method as well as its predecessor - Surrogate Lagrangian Relaxation (SLR) [7], overcame all major difficulties of traditional Lagrangian relaxation (LR) of long subproblem solving times, zigzagging of multipliers, and the

This work was supported in part by the National Science Foundation under Grants ECCS-1810108 and ECCS-1831811, and by a project funded by ABB. Any opinions, findings, conclusions or recommendations expressed in this paper are those of the authors and do not reflect the views of NSF or ABB.

Niranjan Raghunathan, Mikhail A. Bragin, and Peter B. Luh are with the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT 06269-4157, USA (e-mail: niranjan.raghunathan@uconn.edu, mikhail.bragin@uconn.edu and peter.luh@uconn.edu).

Bing Yan is with the Department of Electrical and Microelectronic Engineering at Rochester Institute of Technology, Rochester, NY 14623, USA (e-mail: bxyee@rit.edu).

Khosrow Moslehi, Xiaoming Feng, Chien-Ning Yu, and Chia-Chun Tsai are with ABB (e-mail: khosrow.moslehi@hitachi-powergrids.com, xiaoming.feng@hitachi-powergrids.com, chien-ning.yu@hitachi-powergrids.com, chia-chun.tsai@hitachi-powergrids.com).

Yaowen Yu was with ABB, and is now with the Key Laboratory of Image Processing and Intelligent Control, School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: yaowen_yu@hust.edu.cn).

need to guesstimate the optimal dual values. When solving the sub-hourly UC, however, if all system-wide constraints are relaxed, a very large number of multipliers will be generated, resulting in slow convergence and poor solution quality. To overcome these difficulties, only hard system demand constraints are relaxed, but not soft reserve and transmission capacity constraints. This is possible since, under the “surrogate optimization” concept, the multipliers can be updated and will converge to their optimal values as long as the “surrogate optimality condition” is satisfied without optimally solving the “relaxed problem.” Consequently, subproblems can still be formed and optimized by using MILP methods, e.g., branch-and-cut (B&C), when soft constraints are not relaxed. A theorem with the convergence proof of the new method is presented. This is a unique feature of SLR/SAVLR and is not possible under the traditional LR framework.

The innovative exploitation of “soft” constraints is then extended to problems formulated with hard reserve and transmission capacity constraints. This is done by converting these hard constraints to soft in the iterative solution process, while enforcing the original hard constraints at the stage of finding feasible solutions toward the end. This method thus opens up a new way to solve difficult problems with hard constraints. To further improve performance, a parallel version is developed, which utilizes multiple cores of processors to construct and solve subproblems in parallel.

In Section V, three numerical examples are presented. The first is an hourly UC problem for a 6-bus system, and is used to demonstrate the effects of our exploitation of soft constraints within the SAVLR framework. The second example is a 15-minute UC problem based on the 2383-bus Polish system, and is used to demonstrate the ability of our methods to solve large practical problems. In the third example, Monte Carlo testing based on six load profiles is performed for the same Polish system. Testing results demonstrate the effectiveness and robustness of both the sequential and parallel versions at finding high-quality solutions within the time limit. Our methods are generic and can be used to solve other complex MILP problems in power systems and beyond.

II. LITERATURE REVIEW

In subsection II-A, the use of soft constraints within the UC context is examined. In subsection II-B, the literature on sub-hourly UC is reviewed. In subsection II-C, Surrogate Lagrangian Relaxation (SLR) and its extension with accelerated convergence, Surrogate Absolute-Value Lagrangian Relaxation (SAVLR), are presented. Branch-and-cut (B&C), which is used to solve subproblems, is reviewed in subsection II-D.

A. Use of soft constraints in unit commitment

In [1], the use of soft constraints by ISOs in the United States is reviewed. For UC problems, constraints are modeled as soft when they do not need to be strictly satisfied. For example, ISO-New England models reserve requirements as soft, with different predetermined penalty coefficients for different types of reserves. These coefficients are typically high so that requirements are significantly violated only when satisfying them as hard is difficult or impossible. Constraint violations may also be penalized by step-wise “demand curves” [1]. Procedures for selecting the values of penalty coefficients for

reserve and transmission capacity constraints are different for individual ISOs, and detailed information may not be available.

Certain ISOs also model the demand constraints as soft. The predetermined penalty coefficient for soft demand constraints serves as a cap on the energy market price. Reference [1] also provides an analysis of the market implications of using soft constraints. While soft constraints are often used in power systems and other fields, their use within the LR framework for MILP problems has not been explored.

B. Sub-hourly unit commitment

With the fast-changing demand and the increasing penetration of intermittent renewables, high sub-hourly variations cannot be adequately accounted for in hourly UC. Sub-hourly UC has, therefore, been recommended to address the difficulties in [2 - 5].

In [2], the deterministic and stochastic UC based on the IEEE 118-bus system with sub-hourly intervals (15 or 20 minutes) are compared with the hourly UC. Results obtained by using a modified Benders decomposition method show that costs can be significantly reduced if sub-hourly UC is considered for both deterministic and stochastic problems.

In [3], deterministic UC based on the Irish system (72 units and 1 pumped-storage plant) with sub-hourly time intervals (5, 15, 30 min) is compared with the hourly UC. System demand and reserve constraints are considered, but not transmission capacity constraints. By using FICO Xpress-MP, the study shows that the increased temporal resolution captures more variability in system load and renewable generation, leading to more realistic estimations of total generation costs. It also shows increased cycling and ramping of units. Thus, if the intermittent nature of wind and solar is explicitly modeled, the sub-hourly UC should obtain more economic solutions than the hourly UC.

In addition to the above, the impact of stochastic sub-hourly UC on power system dynamics was discussed in [4] based on the IEEE 39-bus system with the UC problem solved by using GUROBI. The study shows that with smaller time intervals, long-term system frequency deviation caused by the volatility of wind power is reduced, leading to improved system reliability. In [5], the impact of sub-hourly UC on spinning reserve in the presence of intermittent renewables was investigated for an isolated island (with ten diesel generators, three wind turbines, and one photovoltaic power plant) and the IEEE 118-bus system using dynamic programming plus priority listing. The study shows that reserve costs are lowered by considering sub-hourly intervals as compared with those of hourly intervals.

C. Surrogate Lagrangian relaxation and Surrogate Absolute-Value Lagrangian Relaxation

For a system consisting of multiple interconnected areas, an optimization problem is referred to as “separable” if its objective function and the constraints coupling the areas are additive in terms of area variables. LR has been a price-based decomposition and coordination approach to exploit such separable structures, especially for problems with discrete variables. Within the approach, constraints that couple areas are relaxed by using Lagrangian multipliers to form the “relaxed problem,” with its objective function called the “Lagrangian.” The relaxed problem is decomposed into subproblems, one for

each area. Each subproblem is of much-reduced complexity as compared to that of the original problem because of its reduced size. For a given set of multipliers, all subproblems are optimized. The resulting levels of constraint violation form a subgradient, which is used to update the multipliers, and the process iterates. At the convergence of the multiplier updating process, heuristics are used to obtain a feasible solution satisfying all constraints.

Traditional LR methods, however, suffer from major difficulties of long subproblem solving times, zigzagging of multipliers, and the need to guesstimate the optimal dual values. One of the root causes for these difficulties is that the relaxed problem is required to be fully optimized (or all subproblems are required to be optimized) to obtain a subgradient to update multipliers. This renders long subproblem solving times. Also, the resulting subgradient directions may change drastically from one iteration to the next because of the geometry of the dual functions of MILP problems, causing multipliers to zigzag. A key question is therefore whether the relaxed problem (or all subproblems) should be fully optimized in order to update multipliers. In our recent “Surrogate Lagrangian Relaxation” (SLR) method [7], it was proved that the relaxed problem does not need to be fully optimized. Rather, only partial optimization subject to a simple inequality constraint – the “surrogate optimality condition” (see equation (10) in subsection IV-A) – is sufficient. The resulting “surrogate subgradients” (as opposed to “subgradients” since the relaxed problem is not fully optimized) can be used to update multipliers. In terms of subproblems, only a subset of them needs to be optimized subject to the same condition to update multipliers. Such an iterative process requires much-reduced computational effort, yields much less changing of surrogate subgradient directions from one iteration to the next, and results in much-reduced multiplier zigzagging as compared to that of the traditional LR. A stepsizing rule based on contraction mapping not requiring to guesstimate the optimal dual value is also provided in [7]. Consequently, all major difficulties of the traditional LR have been overcome by SLR.

If the original problem is not separable, subproblems can still be formed within the SLR framework. The objective function of a subproblem is obtained by collecting from the Lagrangian all the terms related to the area under consideration, and is minimized with respect to the decision variables belonging to that area while keeping other decision variables at their latest available values, subject to all non-relaxed constraints of relevance to that area. In this way, the LR approach is extended to non-separable problems. Also, not all subproblems need to be solved to obtain a surrogate subgradient to update multipliers. Rather, only one subproblem needs to be solved, and even that subproblem does not need to be optimally solved – only subject to the surrogate optimality condition. The resulting surrogate subgradients are used to update multipliers without requiring the optimal dual value, leading to much reduced computational effort and zigzagging of multipliers.

Subproblem solutions of LR or SLR, when put together, are generally not feasible with respect to the original problem since coupling constraints have been relaxed during the iterative solution process. Heuristics, e.g., those in [8] and [9], are needed to satisfy feasibility toward the end.

Recently, the convergence of SLR has been accelerated by adding absolute-value penalties on constraint violations in Surrogate Absolute-Value Lagrangian Relaxation (SAVLR) in [6]. Absolute-value functions are used instead of quadratic functions as in Augmented Lagrangian Relaxation [10] since they can be exactly linearized to be solved by using MILP solvers. Furthermore, since constraint violations are penalized, subproblem solutions naturally approach feasibility, and advanced heuristics may not be needed to satisfy feasibility.

For the sub-hourly UC problem under consideration, SAVLR may still have troubles in obtaining near-optimal solutions within the time limit for practical cases because of the inherent difficulties introduced in Section I. In particular, when all coupling constraints are relaxed, a large number of multipliers will be introduced, leading to slow convergence and poor solution quality.

D. Branch-and-cut

Branch-and-cut is a widely used method for solving MILP problems [11]. In the process, integrality constraints are relaxed and the resulting problem is solved by using linear programming (LP) methods. Then, valid cuts are added trying to obtain the convex hull. If the convex hull is obtained, then the optimal LP solution is optimal to the original problem. Types of cuts include flow cover cuts, Gomory fractional cuts, implied bound cuts, and others. If the convex hull is difficult to obtain, then the method relies on time-consuming branch-and-bound operations to obtain an optimal or near-optimal solution. LP solutions also serve as lower bounds to the objective function. For practical sub-hourly UC problems, the method may suffer from difficulties. Nevertheless, B&C is effective for solving subproblems within SAVLR because of the much-reduced problem sizes.

III. PROBLEM FORMULATION

Consider a power system consisting of J units which are distributed among A areas. The 15-minute UC is formulated as an MILP problem, the same as that of the hourly UC (e.g., [12]), with ramp rates and other parameters appropriately scaled. The objective is to minimize the sum of generation, startup, no-load, and reserve costs as well as soft-constraint penalties on reserve and transmission capacity over T time intervals, subject to unit-level, area-level, and system-level constraints. To penalize the violations of soft constraints, for simplicity, a single penalty coefficient is used. In practice, different predetermined penalty coefficients or functions can be used. The objective function is formulated as:

$$\begin{aligned} \sum_t \left(\sum_j \left(\sum_{b \in B_j} C_{b,j} p_{b,j}(t) + C_j^{\text{SU}} u_j(t) \right. \right. \\ \left. \left. + C_j^{\text{NL}} x_j(t) + \sum_m C_{j,m}^{\text{R}} r_{j,m}(t) \right) + \sum_a \sum_n C_{a,n}^{\text{P,R}} v_{a,n}^{\text{R}}(t) \right. \\ \left. + \sum_i C_i^{\text{P,TC}} (v_i^{\text{TC},+}(t) + v_i^{\text{TC},-}(t)) \right). \end{aligned} \quad (1)$$

In the above, unit $j, j \in \{1, 2, \dots, J\}$, has the following decision variables: $\{p_{b,j}(t)\}$ are continuous generation-levels of block $b \in B_j$ at time t ; $\{x_j(t)\}$ are binary unit commitment variables; $\{u_j(t)\}$ are binary startup variables; $\{r_{j,m}(t)\}$ are continuous

variables for reserve level of type m from a unit's perspective, including the following of decreasing quality: regulation, 10-minute spinning, ten-minute non-spinning, thirty-minute spinning, and thirty-minute non-spinning. Area a , $a \in \{1, 2, \dots, A\}$, has the following decision variables: $\{v_{a,n}^R(t)\}$ are continuous variables for soft reserve violation of type n from area a 's perspective, including regulation, 10-minute spinning, 10-minute total, and 30-minute total, each can be satisfied by unit contributions of equal or higher quality. Line l , $l \in \{1, 2, \dots, L\}$, where L is the total number of transmission lines, has the following decision variables: nonnegative continuous variables $\{v_l^{TC,+}(t)\}$ and $\{v_l^{TC,-}(t)\}$ for soft transmission capacity violations on the positive and negative directions, respectively. Costs and penalties within (1) include generation costs $\{C_{b,j} p_{b,j}(t)\}$, startup costs $\{C_j^{SU} u_j(t)\}$, no-load costs $\{C_j^{NL} x_j(t)\}$, reserve costs $\{C_{j,m}^R r_{j,m}(t)\}$, soft reserve penalties $\{C_{a,n}^{P,R} v_{a,n}^R(t)\}$, and soft transmission capacity penalties $\{C_l^{P,TC} (v_l^{TC,+}(t) + v_l^{TC,-}(t))\}$.

The problem is subject to the following unit-level constraints: generation capacity, ramp rates, startup, min up/down-time, and reserve capacity given in (1), (4-5), (11-13), (14), and (15-18) of [12], respectively.

Area-wise soft regulation constraints (type $n = 1$) for individual areas are formulated as:

$$\sum_j (r_{1,j}(t) \times pa_{j,a}) + v_{a,1}(t) \geq RR_{a,1}(t), \quad \forall a, t, \quad (2)$$

where $pa_{j,a}$ is the binary area participation indicator of unit j in area a , which is 1 if the unit is in area a and 0 otherwise, and $RR_{a,1}$ is the regulation requirement for area a . Reserve constraints for other types are formulated similarly.

The system-level demand constraints are given as

$$\sum_j p_j(t) = \sum_i P_i^D(t) \quad \forall t, \quad (3)$$

where $p_j(t)$ is the total generation level of unit j at time t , and is the sum of outputs of all the blocks of unit j at time t , and $P_i^D(t)$ is the demand at node i and time t . Soft transmission capacity constraints are formulated as:

$$\begin{aligned} -f_l^{\max} - v_l^{TC,-}(t) &\leq \sum_i \alpha_{i,l} \left(\sum_{j \in J_i} p_j(t) - P_i^D(t) \right) \\ &\leq f_l^{\max} + v_l^{TC,+}(t) \quad \forall t, l, \end{aligned} \quad (4)$$

where f_l^{\max} is the capacity of transmission line l ; J_i is the set of generators at node i ; and $\{\alpha_{i,l}\}$ are generation shift factors, specifying the relationship between power injection at node i and power flow through line l .

IV. SOLUTION METHODOLOGY

In subsection IV-A, soft constraints are exploited within the SAVLR framework to solve the problem with step-by-step derivations. In IV-B, initialization, parameter tuning, and extending the method to solve problems with hard reserve and transmission capacity constraints are presented. To speed up computation, a parallel version of IV-A is developed in IV-C.

A. Exploiting soft constraints within the SAVLR framework

When solving a large sub-hourly UC problem, relaxing all coupling constraints will lead to an excessive number of multipliers, slow convergence, and poor solution quality – the more constraints relaxed, generally the poorer solution quality. To significantly reduce the number of multipliers, only system demand constraints are relaxed but not soft reserve and transmission capacity constraints. With soft constraints not relaxed, subproblems can still be formed and optimized, subject to the surrogate optimality condition. Furthermore, the ability of subproblem solutions to satisfy the surrogate optimality condition is high in view of the flexibility of subproblem solutions with soft constraints. Consequently, convergence can be established with a much-reduced number of multipliers, a feature not possible under the traditional LR framework.

Relaxing system demand constraints

With system demand constraints (3) relaxed by using multipliers $\{\lambda(t)\}$ and constraint violations penalized by using the penalty coefficient c , the relaxed problem is:

$$\begin{aligned} \min_{\{p,u,x,r,v\}} L_c, \text{ with} \\ L_c \equiv \sum_t \left(\sum_j \left(\sum_{b \in B_j} C_{b,j} p_{b,j}(t) + C_j^{SU} u_j(t) + C_j^{NL} x_j(t) \right. \right. \\ \left. \left. + \sum_m C_{j,m}^R r_{j,m}(t) \right) + \sum_a \sum_n C_{a,n}^{P,R} v_{a,n}^R(t) \right. \\ \left. + \sum_{l \in L} C_l^{P,TC} (v_l^{TC,+}(t) + v_l^{TC,-}(t)) \right. \\ \left. + \lambda(t) \left(\sum_j p_j(t) - \sum_i P_i^D(t) \right) \right. \\ \left. + c \left(\left| \sum_j p_j(t) - \sum_i P_i^D(t) \right| \right) \right), \end{aligned} \quad (5)$$

subject to individual unit-level constraints, and soft reserve and transmission capacity constraints which are not relaxed. In the above, L_c is the Lagrangian.

Linearizing absolute value penalties

The Lagrangian (5) is nonlinear due to the absolute-value penalty terms, which are linearized in a standard way [13] (Pg. 63). The linearized relaxed problem is:

$$\begin{aligned} \min_{\{p,u,x,r,v,z\}} L'_c, \text{ with} \\ L'_c \equiv \sum_t \left(\sum_j \left(\sum_{b \in B_j} C_{b,j} p_{b,j}(t) + C_j^{SU} u_j(t) + C_j^{NL} x_j(t) \right. \right. \\ \left. \left. + \sum_m C_{j,m}^R r_{j,m}(t) \right) + \sum_a \sum_n C_{a,n}^{P,R} v_{a,n}^R(t) \right. \\ \left. + \sum_{l \in L} C_l^{P,TC} (v_l^{TC,+}(t) + v_l^{TC,-}(t)) \right. \\ \left. + \lambda(t) (z^+(t) - z^-(t)) + c(z^+(t) + z^-(t)) \right), \end{aligned} \quad (6)$$

subject to individual unit-level constraints and soft reserve and transmission capacity constraints. In the above, $z^+(t)$ and

$z^-(t)$ are nonnegative continuous linearization variables satisfying the following newly introduced constraints:

$$z^+(t) - z^-(t) = \sum_j p_j(t) - \sum_i P_i^D(t), \forall t. \quad (7)$$

Formulating and solving area subproblems

The objective function of subproblem a is formed by collecting all the terms from Lagrangian L_c in (6) related to area a . It is minimized with respect to the decision variables belonging to that area while keeping all other decision variables at their latest available values following the procedures of SAVLR as reviewed in subsection II-C. The subproblem a solved at iteration k is:

$$\begin{aligned} \min_{\{p, u, x, r, v, z\}} L_s^k, \text{ where} \\ L_a^k \equiv \sum_i \left(\sum_{j: j \in G_a} \left(\sum_{b \in B_j} C_{b,j} p_{b,j}(t) + C_j^{SU} u_j(t) + C_j^{NL} x_j(t) \right. \right. \\ \left. \left. + \sum_m C_{j,m}^R r_{j,m}(t) \right) + \sum_a \sum_n C_{a,n}^{P,R} v_{a,n}^R(t) \right. \\ \left. + \sum_{l \in L} C_l^{P,TC} (v_l^{TC,+}(t) + v_l^{TC,-}(t)) \right. \\ \left. + \lambda^k(t) (z^{k,+}(t) - z^{k,-}(t)) + c^k (z^{k,+}(t) + z^{k,-}(t)) \right), \end{aligned} \quad (8)$$

where G_a is the set of units belonging to area a . For brevity, constant terms are omitted in (8). This subproblem is subject to unit-level constraints for units belonging to the area, soft reserve and transmission capacity constraints (not relaxed), and the following linearization constraints (not relaxed):

$$\begin{aligned} z^{k,+}(t) - z^{k,-}(t) = & \left(\sum_{j: j \notin G_a} p_j^{k-1}(t) + \sum_{j: j \in G_a} p_j^k(t) \right) \\ & - \sum_i P_i^D(t), \end{aligned} \quad (9)$$

where $p_j^{k-1}(t)$ denote the most recent decision variable obtained at iteration $k - 1$.

Subproblems are solved by using B&C subject to the satisfaction of the surrogate optimality condition [6] (Eq. (14))

$$\tilde{L}_{c^k, \lambda^k}^k < \tilde{L}_{c^k, \lambda^k}^{k-1}, \quad (10)$$

where $\tilde{L}_{c^k, \lambda^k}^k$ is the “surrogate dual value,” which is the Lagrangian (6) evaluated using decision variables of area a at iteration k , decision variables of other areas at iteration $k - 1$, multipliers λ^k , and the penalty coefficient c^k ; and $\tilde{L}_{c^k, \lambda^k}^{k-1}$ is (6) evaluated using all decision variables at iteration $k - 1$, multipliers λ^k , and the penalty coefficient c^k .

Updating multipliers and the penalty coefficient

If the surrogate optimality condition (10) is satisfied, then the “surrogate subgradients” are obtained as:

$$\tilde{g}^k(t) = \left(\sum_{j \in G_a} p_j^k(t) + \sum_{j \notin G_a} p_j^{k-1}(t) \right) - \sum_i P_i^D(t), \quad \forall t, \quad (11)$$

and the multipliers $\lambda(t)$ are updated following (14) from [6] as:

$$\lambda^{k+1}(t) = \lambda^k(t) + s^k \cdot \tilde{g}^k(t), \quad \forall t. \quad (12)$$

In (12), the step size s^k is updated as:

$$s^k = \alpha_k \frac{s^{k-1} \|\tilde{g}^{k-1}\|_2}{\|\tilde{g}^k\|_2}, \quad 0 < \alpha_k < 1, \quad (13)$$

where

$$\alpha_k = 1 - \frac{1}{Mk^\rho}, \quad \rho = 1 - \frac{1}{k^r}, \quad M \geq 1, \quad 0 \leq r \leq 1. \quad (14)$$

The penalty coefficient c is updated as:

$$c^{k+1} = \min(c^{ub}, \beta \cdot c^k), \quad \beta > 1, \quad (15)$$

where c^{ub} is an upper bound on c , preventing c from getting too large to impede coordination.

If the surrogate optimality condition is not satisfied, the multipliers $\lambda(t)$ and the penalty coefficient c are not updated, and the next subproblem is solved. If the surrogate optimality condition cannot be satisfied after A consecutive subproblems, c is deemed too large, and is reduced follows [6] (Eq. (21))

$$c^{k+1} = \frac{c^k}{\beta}, \quad \beta > 1. \quad (16)$$

Eq. (16) is also used if subproblem solutions, when put together, are feasible with respect to the original problem to reduce the emphasis of the feasibility. After updating the multipliers, stepsize, and penalty coefficient, the next subproblem is solved, and the process is repeated until a time limit is reached or the multipliers converge based on appropriate criteria (e.g., if the distance between multipliers after solving A subproblems is within a predetermined tolerance). The following theorem establishes the convergence of SAVLR when not all coupling constraints are relaxed.

Theorem 1. Convergence of SAVLR in which not all coupling constraints are relaxed. Consider the relaxed problem (5) in which not all coupling constraints are relaxed. If the surrogate optimality condition (10) is satisfied for at least one subproblem after solving A subproblems, and multipliers are updated as in (12) using stepsize update rule (13), then multipliers converge to their optimal values λ^* .

Proof: The proof follows directly Theorem 1 of [6]. \square

Finding feasible solutions and evaluating solution quality

Since system demand constraints are relaxed, subproblem solutions, when put together, might not constitute a solution feasible to the original problem. To construct a feasible solution when the time limit is reached or appropriate stopping criteria are satisfied, all binary decision variables are fixed at their subproblem solution values, and the resulting LP problem is solved. If a feasible solution cannot be obtained in this way, then more complicated heuristics will be needed, e.g., by fixing a subset of binary decisions and solving the rest of the decision variables. This, however, was not encountered in our testing to be reported in Section V.

To measure the quality of a feasible solution, the following optimality gap is used:

$$\text{Gap (\%)} = 100 \times \frac{(f^{\text{best}} - lb^{\text{best}})}{f^{\text{best}}}, \quad (17)$$

where f^{best} is the best feasible cost and lb^{best} is the best known lower bound, which is obtained by using pure B&C in our numerical testing to be presented in Section V.

B. Initialization, Parameter Tuning and Solving Problems with Hard Constraints

In this subsection, initialization of subproblem solutions, tuning stepsize parameters, and extending our method to solve problems with hard reserve and transmission capacity constraints are presented.

Initializing subproblem solutions

Initial subproblem solutions are needed to start the iterative process (see the flowchart in Fig. 1). To do this, startup and generation level variables are chosen for each time interval such that generation capacity (1), ramp rate (4-5), startup (11-13), and minimum up/down-time (14) constraints are satisfied, while other constraints are ignored.

Tuning stepsize parameters

There are three parameters in our method: (M , r , and s^0), and they could be different for different classes of problems, e.g., a 6-bus system and the 2383-bus Polish system in the next section. To systematically select them for each class of problems, a tuning process is developed through testing a discrete subset of the parameter space. To avoid searching in three dimensions, the process is divided into two stages. In the first stage, r is fixed at a typical value determined from testing experience, and a search over values of M and s^0 is performed to obtain a good range for M and s^0 . Then in the second stage, s^0 is fixed at the value within the range identified in the first stage, and a search is performed over values r and a finer grid of M selected from the first stage. Once obtained, a parameter set is used for all instances of that class of problems.

Solving problems with hard constraints

Motivated by the effectiveness of our method, one idea to solve UC problems formulated with hard reserve and transmission capacity constraints is to convert these constraints to soft, and then follow the procedure developed above. Only in the last step, feasible solutions satisfying all the original constraints are obtained by using heuristics. Note that if hard constraints are not converted to soft, they should always be satisfied when solving subproblems, significantly increasing the difficulties of satisfying the surrogate optimality condition.

The key consideration here is the selection of soft constraint penalty coefficients. To facilitate the satisfaction of the surrogate optimality condition when solving subproblems, these penalty coefficients should not be too large to impede coordination. However, they should be larger than the values of multipliers. If these coefficients are of the same order of magnitude as multipliers, there might be an excessive violation of the original hard constraints during the iterative solution process, making it difficult to obtain feasible solutions at the end. Since values of multipliers are not known, these coefficients are set an order of magnitude higher than the estimated upper bounds on the multipliers, which are calculated in the pre-processing stage by, e.g., priority list commitment and dispatch ([8]). Our approach thus opens up a new way to solve difficult MILP problems with hard constraints. A flowchart for SAVLR is given in Fig. 1.

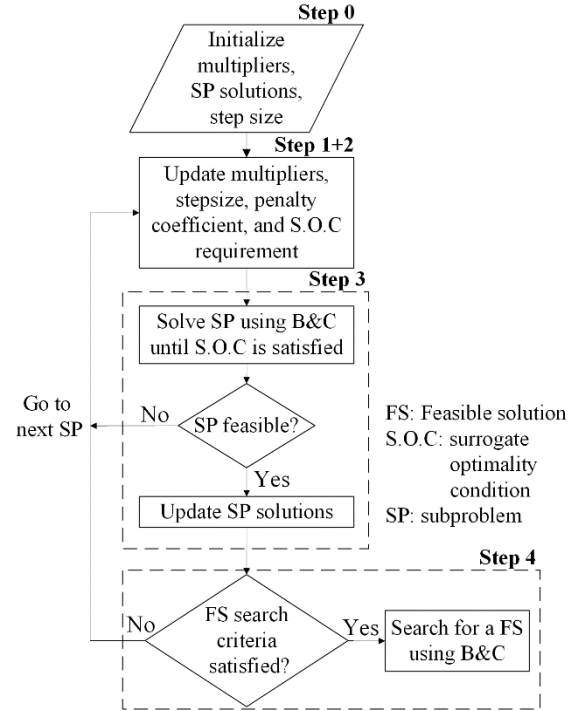


Figure 1: SAVLR flowchart

C. Synchronous parallel Surrogate Absolute-Value Lagrangian Relaxation

The method presented above can be further improved by taking advantages of multiple cores available in modern processors. The formulation of the relaxed problem and the decomposition process are identical to those for the sequential version. Subproblems are grouped in batches, and subproblems in a batch are assigned to individual cores. The subproblems in a batch are solved in parallel after all the subproblems in the previous batch are finished and multipliers and the penalty coefficient c updated, resulting in a synchronous parallel version. Subproblems in a batch are thus independent of each other, and there is no communication needed among cores when subproblems are solved.

After a batch of subproblems is solved, their solutions are combined. Although individual solutions may satisfy the surrogate optimality condition, the combined solution may not. If the combined solution satisfies the surrogate optimality condition, then the multipliers $\{\lambda(t)\}$ and the penalty coefficient c are updated as for sequential SAVLR, and the next batch of subproblems is solved. If the combined solution does not satisfy the surrogate optimality condition, then combinations of subsets of solutions are checked, with priority given to the largest combinations. Once a combination satisfying the surrogate optimality condition is obtained, multipliers and the penalty coefficient c are updated, and the next batch is solved.

In contrary to traditional wisdom, not all subproblems should be solved in parallel. This is because in this case, the method is similar to traditional LR, and suffers from drastic zigzagging of multipliers and slow convergence. Based on our testing results, 10% to 40% of subproblems may be solved in parallel subject to the number of available cores for effective convergence.

V. NUMERICAL TESTING AND ANALYSIS

Three UC problems are solved to demonstrate the performance of our method. The first is a simple 6-bus hourly UC over 24 hours to demonstrate the impacts of the exploitation of soft constraints within the SAVLR framework. The second is a 15-minute UC based on the 2383-bus Polish system, with hard demand, reserve, and transmission capacity constraints. After hard reserve and transmission capacity constraints are converted to soft, the ability of the sequential SAVLR to solve such large problems is presented. In the third example, to demonstrate the robustness and performance of both sequential and parallel methods, Monte Carlo testing for the same 15-minute Polish system but with soft reserve and transmission capacity constraints is performed based on six load profiles.

The implementation overhead is reported, and the parameter tuning procedure is demonstrated. For all the examples, our results are benchmarked against those obtained by using B&C (as available in CPLEX) with mixed-integer rounding, flow, implied bound, and other cuts applied. All testing is performed on an Intel Xeon CPU 3.1 GHz, 4 Cores, 32 GB laptop, with MATLAB R2018a and CPLEX 12.8.

Example 1: 6-bus system

An hourly UC over a 24-hour time horizon for a 6-bus system with 9 units and 11 transmission lines as shown in Fig. 2 is considered. The problem is subject to the standard unit-level constraints, hard system demand constraints, and soft transmission capacity constraints. For simplicity, reserves are not considered. A soft constraint penalty coefficient of \$1000/MW is set for not meeting the transmission capacity requirements.

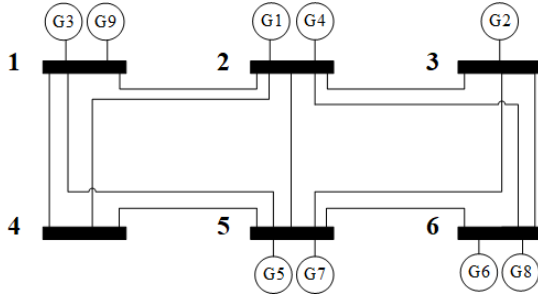


Figure 2: The 6-bus system considered in Example 1

The problem is solved by using B&C, and sequential and parallel SAVLR. For SAVLR, the problem is decomposed into 5 subproblems, which are solved by using B&C. For parallel SAVLR, these subproblems are grouped into 3 batches, with 2 batches containing 2 subproblems, and 1 batch consisting of a single subproblem (40% of subproblems solved in parallel). For comparison purposes, the problem is also solved by using sequential SAVLR with the soft constraints relaxed. The feasible costs obtained by each method, and performance metrics (algorithm run time, optimality gap, and the number of iterations performed) are provided in Table I. Because of the low complexity of this example, B&C finds a solution with a .01% gap in a fraction of a second. Both sequential and parallel SAVLR, when only hard demand constraints are relaxed, find the same solution in 25 seconds. They take longer because multipliers and subproblem solutions are set to update for 25 seconds before finding feasible solutions, which takes less than

TABLE I
RESULTS FOR EXAMPLE I

Method	Feasible Cost	Gap	Time	Iterations
B&C	\$623,956	0.01%	< .1s	-
Sequential SAVLR – not relaxing soft constraints	\$623,956	0.01%	25s	40
Parallel SAVLR – not relaxing soft constraints	\$623,956	0.01%	25s	66
Sequential SAVLR – relaxing soft constraints	\$666,280	6.05%	43s	68

a second. The sequential SAVLR solved 286 subproblems, of which 219 satisfied the surrogate optimality condition. The parallel SAVLR solved 219 batches, of which 11 did not satisfy the surrogate optimality condition after combining their subproblem solutions. As a result, 390 subproblems (36% more than the sequential version) are solved. When soft transmission capacity constraints are relaxed, convergence is slow because 528 additional multipliers are introduced, and the feasible cost is also higher, demonstrating the effectiveness of not relaxing these soft constraints.

Example 2: The 2386-bus Polish system with hard demand, reserve and transmission capacity constraints

In this example, the effectiveness of our sequential method for sub-hourly UC with hard constraints is demonstrated. A 15-minute UC problem over a 24-hour planning horizon is considered based on the 2383-bus Polish system power flow test case with 327 units and 2895 transmission lines and hard system demand, reserve, and transmission capacity constraints [14]. The hard reserve and transmission capacity constraints are first converted to soft. To select soft constraint penalty coefficients, the initial multipliers are first calculated based on priority list commitment, and the largest of which is \$25/MW. Since multipliers are expected to be in the same order of magnitude as their initial values, to estimate an upper bound, the largest initial multiplier is doubled (\$50/MW), and soft constraint penalty coefficients are set to be an order of magnitude higher at \$500/MW for all soft constraints. To increase problem complexity, all unit ramp rates are reduced to 60% of their nominal values. The problem is decomposed into 28 subproblems, each consisting of 9 to 12 units.

Both B&C and sequential SAVLR are run for 2 hours. For sequential SAVLR, a feasible solution is searched every 10 minutes to compare the solution quality over time, with each search taking approximately 30 seconds. The feasible costs and gaps obtained over time by each method are plotted in Fig. 3. SAVLR obtains a feasible solution with a cost of \$32.78 million and a gap of 1.52% in 20 minutes. B&C cannot find a feasible solution with a cost under \$36 million until after 115 minutes, and at that point, it finds a solution with a cost of \$32.3 million with a gap of .04%.

To demonstrate the effects of selecting the soft constraint penalty coefficient, the problem is also solved with its value set at \$200/MW and \$50,000/MW. Performance metrics

including $\|\tilde{g}\|_2$ (the L^2 norm of the relaxed demand constraints, see equation (11) in subsection IV-A) at the end, $\|v\|_1$, the L^1 norm of the soft reserve and transmission capacity constraints (see equation (1) in subsection III) at the end, feasible cost, and optimality gap for the 3 cases are presented in Table II.

TABLE II
PERFORMANCE WITH DIFFERENT PENALTY COEFFICIENTS FOR EXAMPLE 2

Sequential SAVLR with different soft constraint penalty coefficients				
Soft constraint penalty coefficient	$\ \tilde{g}\ _2$ (MW)	$\ v\ _1$ (MW)	Feasible cost	Gap
\$500/MW	13.04	1.73	3.278×10^7	1.52 %
\$200/MW	64.17	18.1	No feasible sol.	-
\$50,000/MW	2.81×10^3	~ 0	No feasible sol.	-

Over the course of the iterative multiplier updating process, multiplier values fall within the range of $-\$3.75$ and $\$74.43$. With the soft constraint penalty coefficient set at $\$200/\text{MW}$, the norm of reserve and transmission capacity constraint violations, $\|v\|_1$, is 18.1 MW. This value of $\$200/\text{MW}$ is too close to the range of multipliers. Consequently, subproblem solutions suffer from high levels of constraint violation, and a feasible solution to the original problem is not found. With the soft constraint penalty coefficient set at $\$50,000/\text{MW}$, the amount by which the original reserve and transmission capacity constraints are

violated by subproblem solutions is negligible. However, with such a high penalty coefficient, the soft constraints behave like hard constraints, impeding the subproblem coordination process. The norm of system demand constraint violations, $\|\tilde{g}\|_2$, remains large (2.81×10^3 MW) even after 20 minutes, and a feasible solution to the original problem is not found. With a penalty coefficient of $\$500/\text{MW}$, which is about an order of magnitude higher than the largest multiplier, $\|v\|_1$ is only 1.73 MW; $\|\tilde{g}\|_2$ is 13.4 MW; and a feasible solution with a gap of 1.52% is obtained.

Example 3: Monte Carlo simulations for the Polish system, with soft reserve and transmission capacity constraints

In this example, the robustness and performance of both the sequential and the parallel methods are examined by performing Monte Carlo testing of the 15-minute UC problem over a 24-hour planning horizon for the Polish system with soft reserve and transmission capacity constraints. The soft constraint penalty coefficient is $\$1000/\text{MW}$ for both constraints. For the parallel SAVLR, 8 batches with 3 subproblems each and 1 batch with 4 subproblems are formed (10%-14% of subproblems solved in parallel). Six cases are considered based on seasonal and weekday/weekend variations following Table 4 of [15]. For each case, three scenarios are created by varying the load at each node by up to 3%. All methods are run for 20 minutes, including 30 seconds for finding feasible solutions. The average results of the three scenarios for each case are shown in Table III. The average initial subproblem model construction time (building constraint

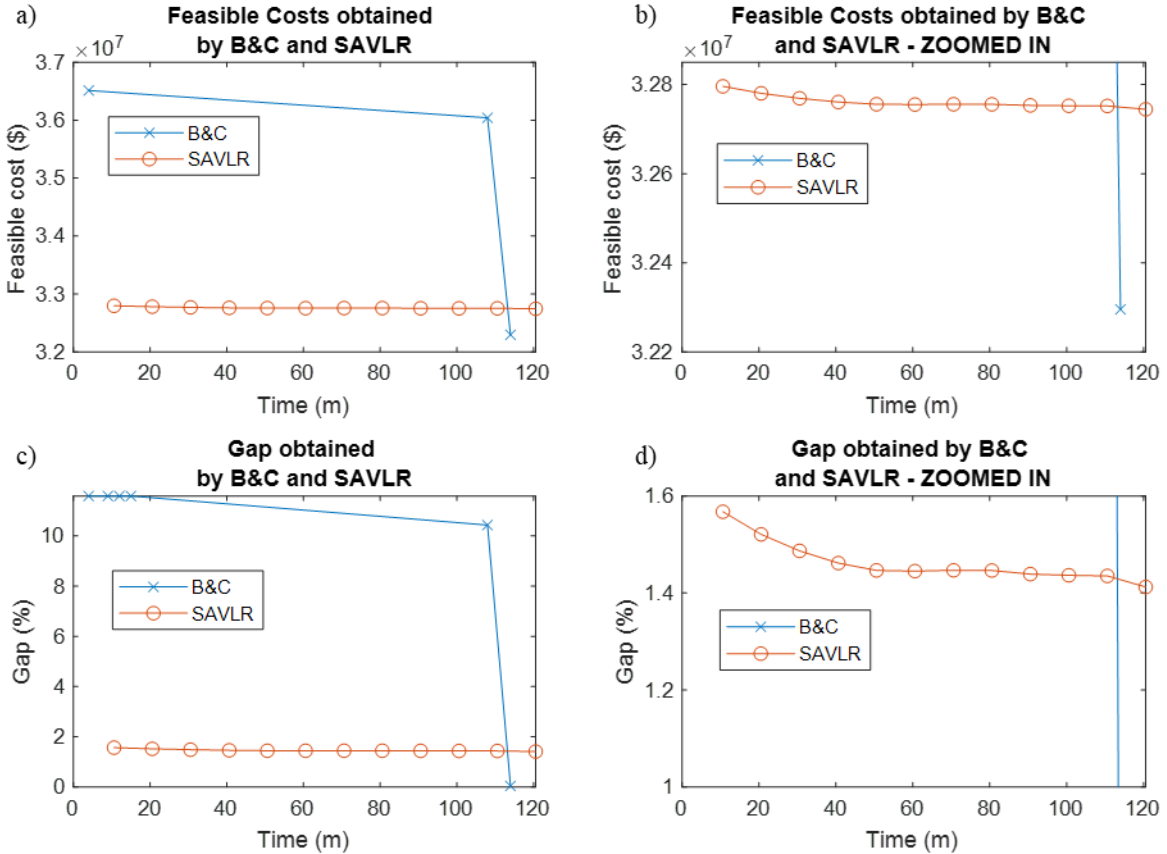


Figure 3: a) Feasible costs. b) Feasible costs obtained by SAVLR decrease over time. c) Optimality gaps. d) Optimality gaps zoomed in.

matrices) is 7.6 seconds, and the overhead at each iteration (defining costs and constraint requirements) is 0.3 seconds. The main cuts used are mixed-integer rounding, flow, and implied bound cuts.

TABLE III
RESULTS FOR EXAMPLE 3

Season – weekday/weekend	Average feasible cost (average gap)		
	Sequential SAVLR ($\times 10^7$)	Parallel SAVLR ($\times 10^7$)	B&C ($\times 10^7$)
Winter - weekday	3.2435 (2.68%)	3.2476 (2.80%)	3.5859 (11.97%)
Winter - weekend	3.2535 (3.06%)	3.2531 (3.05%)	3.5504 (11.17%)
Summer - weekday	3.2518 (2.95%)	3.2809 (3.77%)	3.1690 (.39%)*
Summer - weekend	3.2363 (2.64%)	3.2385 (2.71%)	3.2881 (3.88%)
Spring/Fall - weekday	3.2493 (2.61%)	3.2625 (3.01%)	3.5631 (11.19%)
Spring/Fall - weekend	3.2691 (2.6%)	3.2783 (2.88%)	3.5648 (10.68%)
*Average values are based on 2 scenarios. No solution found for the 3 rd scenario			

For 5 out of the 6 cases, sequential and parallel SAVLR find lower average costs than those of B&C. B&C finds the best average cost for the “Summer - weekday” case. However, only 2 out of 3 scenarios were solved, and no solution was found for the 3rd instance (the average cost and gap were calculated using the 2 scenarios for which solutions were found). Both versions of SAVLR give feasible solutions for all scenarios of all cases.

SAVLR consistently finds solutions with less than a 4% gap in 20 minutes, while B&C struggles for most scenarios due to the high problem complexity.

Sequential SAVLR finds solutions that cost less than those found by the parallel version for 5 out of 6 cases. This is because the parallel version still suffers from the convergence difficulties mentioned in subsection IV-C, although it solves more subproblems than those solved by the sequential version.

Tuning stepsizing parameters

The results from stages 1 and 2 of the parameter tuning procedure described in subsection IV-B are given in Tables IV and V, respectively. In both stages, a search for a feasible solution is performed after 20 minutes for each parameter set. In stage 1, based on testing experience, r is fixed at .1, and a search over values of $M \in \{1, 10, 20, 30\}$ and $s^0 \in \{5 \times 10^{-3}, 5 \times 10^{-2}, 5 \times 10^{-1}\}$ is performed. Runtime for stage 1 is 4 hours. Based on Table IV, $s^0 \in \{5 \times 10^{-3}, 5 \times 10^{-2}\}$ and $M \in \{1, 10\}$ give comparable feasible costs, and are suitable ranges for these parameters. Therefore in stage 2, s^0 is fixed at 5×10^{-3} and a search over $M \in \{5, 10, 15\}$ and $r \in \{.06, .08, .10, .12, .14\}$ is performed. Runtime for stage 2 is 5 hours.

Based on Table V, $r \in \{.06, .08, .10, .12, .14\}$ and $M \in \{5, 10\}$ give comparable feasible costs. Consequently, the parameter set $\{M = 10, r = .1, s^0 = 5 \times 10^{-3}\}$ is used in sequential SAVLR for all cases and scenarios. The same procedure is also used for tuning parallel SAVLR.

TABLE IV
TUNING FOR EXAMPLE 3: STAGE 1

Feasible costs values of M and s^0 with $r = .1$			
	$s^0 = 5 \times 10^{-3}$	$s^0 = 5 \times 10^{-2}$	$s^0 = 5 \times 10^{-1}$
$M = 1$	3.23×10^7	3.23×10^7	3.25×10^7
$M = 10$	3.23×10^7	3.24×10^7	No feasible sol.
$M = 20$	3.23×10^7	No feasible sol.	No feasible sol.
$M = 30$	No feasible sol.	No feasible sol.	No feasible sol.

TABLE V
TUNING FOR EXAMPLE 3: STAGE 2

Feasible costs values of M and r with $s^0 = 5 \times 10^{-3}$			
	$M = 5$	$M = 10$	$M = 15$
$r = .06$	3.24×10^7	3.22×10^7	3.23×10^7
$r = .08$	3.23×10^7	3.22×10^7	3.21×10^7
$r = .10$	3.22×10^7	3.23×10^7	3.21×10^7
$r = .12$	3.22×10^7	3.21×10^7	3.26×10^7
$r = .14$	3.22×10^7	3.22×10^7	No feasible sol.

VI. CONCLUSION

The soft reserve and transmission capacity constraints are innovatively exploited within the SAVLR framework for complicated sub-hourly UC problems. The approach leads to a drastic reduction of the number of multipliers, decreased computational requirements, and improved performance. Both the sequential and the parallel versions consistently provide near-optimal solutions in a computationally efficient manner for the 15-minute UC problems tested. These methods are generic and have major implications for solving other complex MILP problems in power systems and beyond.

VII. REFERENCES

- [1] Y. M. Al-Abdullah, A. Salloum, K. W. Hedman, and V. Vittal, “Analyzing the Impacts of Constraint Relaxation Practices in Electric Energy Markets,” IEEE Transactions on Power Systems, vol. 31, no. 4, pp. 2566–2577, 2016.
- [2] J. Wang, J. Wang, C. Liu, and J. P. Ruiz, “Stochastic unit commitment with sub-hourly dispatch constraints,” Applied Energy, vol. 105, pp. 418–422, 2013.
- [3] J. P. Deane, G. Drayton, B.P. Ó Gallachóir, “The impact of sub-hourly modelling in power systems with significant levels of renewable generation,” Applied Energy, Volume 113, Pages 152-158, 2014.
- [4] T. Kërçi, J. Giraldo, F. Milano, “Analysis of the impact of sub-hourly unit commitment on power system dynamics,” International Journal of Electrical Power & Energy Systems, Volume 119, 2020.
- [5] M. Kazemi, P. Siano, D. Sarno, and A. Goudarzi, “Evaluating the impact of sub-hourly unit commitment method on spinning reserve in presence of intermittent generators,” Energy, Volume 113, pp. 338-354, 2016.
- [6] M. A. Bragin, P. B. Luh, B. Yan, and X. Sun, “A Scalable Solution Methodology for Mixed-Integer Linear Programming Problems Arising

- in Automation,” IEEE Transactions on Automation Science and Engineering, vol. 16, no. 2, pp. 531–541, 2019.
- [7] M. A. Bragin, P. B. Luh, J. H. Yan, N. Yu, and G. A. Stern, “Convergence of the Surrogate Lagrangian Relaxation Method,” Journal of Optimization Theory and Applications, vol. 164, no. 1, pp. 173–201, 2015.
 - [8] X. Guan, P. B. Luh, H. Yan, J. A. Amalfi, “An optimization-based method for unit commitment,” International Journal of Electrical Power & Energy Systems, Volume 14, Issue 1, 1992, Pages 9-17.
 - [9] H. Wu, X. Guan, Q. Zhai and H. Ye, “A Systematic Method for Constructing Feasible Solution to SCUC Problem With Analytical Feasibility Conditions,” in IEEE Transactions on Power Systems, vol. 27, no. 1, pp. 526-534, Feb. 2012.
 - [10] D. P. Bertsekas, Nonlinear Programming, Third Edition, Belmont, MA, USA: Athena Scientific 2016.
 - [11] Cplex, I.I., 2009. V12. 1: User’s Manual for CPLEX. International Business Machines Corporation, 46(53), p.157.
 - [12] B. Yan, P. Luh, T. Zheng, D. Schiro, M. Bragin, F. Zhao, J. Zhao, and I. Lelic, “A Systematic Formulation Tightening Approach for Unit Commitment Problems,” IEEE Transactions on Power Systems, vol. 35, no. 1, pp. 782–794, 2020.
 - [13] J. Bisschop, AIMMS Optimization Modeling. Morrisville, NC, USA: Lulu Press, 2006
 - [14] <https://matpower.org/docs/ref/matpower5.0/case2383wp.html>
 - [15] Reliability Test System Task Force, “The IEEE Reliability Test System - 1996,” IEEE Transactions on Power Systems, vol. 14, no. 3, pp. 1010-1020, Aug. 1999.