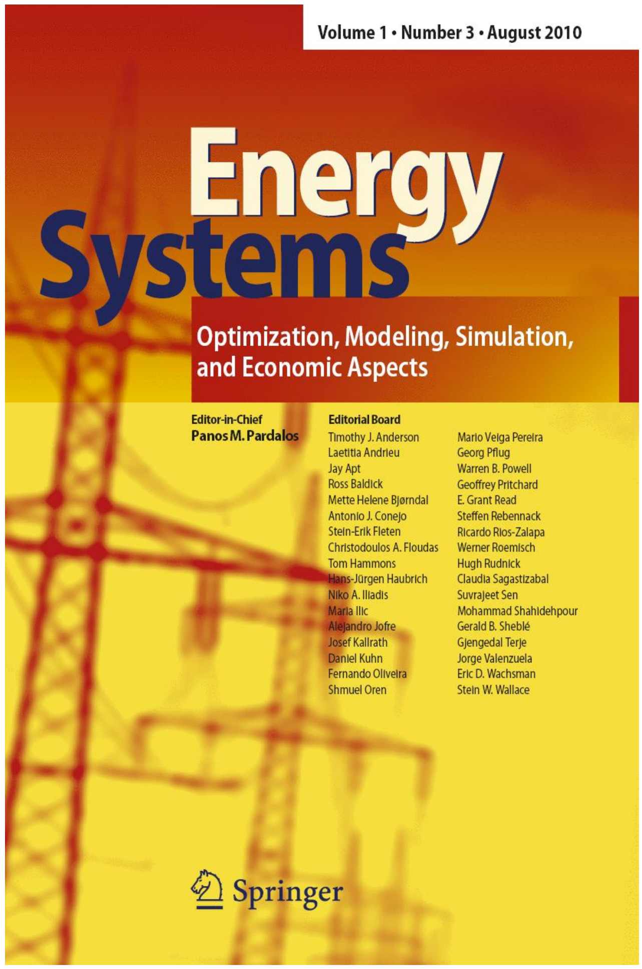
ISSN 1868-3967, Volume 1, Number 3



Energy Syst (2010) 1: 339–376
DOI 10.1007/s12667-010-0015-4

ORIGINAL PAPER

Accelerating the Benders decomposition for network-constrained unit commitment problems

Lei Wu · Mohammad Shahidehpour

Received: 21 February 2010 / Accepted: 31 May 2010 / Published online: 16 June 2010

Abstract This paper presents an optimization method by generating multiple strong Benders cuts for accelerating the convergence of Benders Decomposition (BD) when solving the network-constrained generation unit commitment (NCUC) problem. In NCUC, dc transmission network evaluation subproblems are highly degenerate, which would lead to many dual optimal solutions. Furthermore, the classical BD cuts are often low-density which involve only a limited number of decision variables in the master problem. Therefore, the dual optimal solutions and the corresponding Benders cuts are of crucial importance for improving the efficiency of the BD algorithm. The proposed method would generate multiple strong Benders cuts, which are pareto optimal, among candidates from multiple dual optimal solutions. Such cuts would be high-density in comparison with low-density cuts produced by the classical BD. The proposed multiple strong Benders cuts are efficient in terms of reducing the total iteration number and the overall computing time. The high-density cuts may restrict the feasible region of the master unit commitment (UC) problem in each iteration as they incorporate more decision variables in each Benders cut. The multiple strong Benders cuts would accordingly reduce the iteration number and overall computing time. Numerical tests demonstrate the efficiency of the proposed multiple strong Benders cuts method in comparison with the classical BD algorithm and the linear sensitivity factors (LSF) method. The proposed method can be extended to other applications of BD for solving the large-scale optimization problems in power systems operation, maintenance, and planning.

This work is supported in part by the National Science Foundation grants ECCS-0725666 and ECCS-0801853.

L. Wu · M. Shahidehpour (✉)
Electrical and Computer Engineering Department, Illinois Institute of Technology, Chicago, IL 60616, USA
e-mail: ms@iit.edu

L. Wu
e-mail: lwu10@iit.edu

Keywords Electric power systems · Network-constrained unit commitment · Benders decomposition · Multiple strong Benders cuts · Degenerate · Multiple optimal dual solutions · Low-density · Pareto optimal · LSF

Nomenclature

Indices

a, b, m, n	Denote a bus
d, d'	Denote a BD iteration
i, j	Denote a unit
k	Denote a segment of a cost curve function
l, l'	Denote a transmission line
o, o'	Denote a subset
p	Denote a phase shifter
t, τ	Hourly index

Sets and vectors

$\mathbf{B}, \mathbf{B}_o, \mathbf{B}_{o'}$	Set/subsets of buses
$\mathbf{f}(b), \mathbf{t}(b)$	Set of transmission lines starting from bus b/ending at bus b
$\mathbf{PL}_t$	Vector of power flow variables at hour t
$\mathbf{PL}^{\max}$	Vector of power flow upper limit
$\mathbf{U}(b)$	Set of generators located at bus b
$\mathbf{Z}, \mathbf{Z}^c, \mathbf{SZ}_o, \mathbf{SZ}_{o'}$	Sets of buses
$\gamma(m)$	Denote a set of buses, for each bus n in the set, there is a phase shifter at line m to n, and m is the tap side and n is the non-tap side
γ_t	Vector for phase shifter at hour t
$\gamma^{\min}, \gamma^{\max}$	Vector of phase shifter lower/upper limits

Variables

DT_i, UT_i	Number of hours unit i must be initially offline/online due to its minimum off/on time limits
G	Objective value of the global optimal solution to the original problem
I_{it}, P_{it}	Unit commitment decision/generation dispatch of unit i at hour t
LB_d, UB_d	Lower/upper bound of the original problem objective value obtained at BD iteration d
LB_{best}, UB_{best}	Best obtained lower/upper bounds for the objective value to the original problem
offset	Offset term in the linear expression of the power loss
OR_{it}, SR_{it}	Non-spinning/spinning reserve provided by unit i at hour t
P_{ikt}	Generation dispatch of unit i at hour t at segment k
P_{Losst}	System loss at hour t
PL_{lt}	Power flow of transmission line l at hour t
q, q_1, q_2, Q	Dual variables for the dual problem
s	Slack variable
SD_{it}, SU_{it}	Shutdown/startup cost of unit i at hour t
θ_{at}, θ_{bt}	Phase angle of bus a and bus b at hour t
θ_{ref}	Phase angle of the reference bus

γ_{abt}	Phase shifter value on the line from bus a to bus b at hour t
λ, κ, π	Dual variables for the primal problem
$\|\mathbf{Z}\|$	Number of buses that belong to set $\mathbf{Z}$
$\hat{\cdot}, \tilde{\cdot}$	Value for variables calculated at previous BD iterations

Constants

c	A constant integer number determined during the algorithm
C_b	Loss distribution factor corresponding to bus b
c_{ik}, N_i	Incremental cost for segment k and no-load cost of unit i
D_{bt}	System load at bus b hour t
DP_i, UP_i	Shutdown/startup ramp limits of unit i
DR_i, UR_i	Ramping down/ramping up limits of unit i
LF_b	Loss factor related to bus b
LSF_{ab}^m	The sensitivity of the power flow on line l (from bus a to bus b) to power injection at bus m
M, ε	Very large/small positive number
MSR_i	Spinning reserve that can be provided by unit i in one minute
NB	Number of buses
NT	Number of hours under study
P_{Dt}, P_{Lt}	System load/loss at hour t
$P_i^{\min}, P_i^{\max}$	Minimum/maximum capacity of unit i
$P_{ik}^{\max}$	Maximum capacity of unit i of segment k
QSC_i	Quick start capacity of unit i
R_{Ot}, R_{St}	System non-spinning/spinning reserve requirements at hour t
T_i^{on}, T_i^{off}	Minimum on/off time limits of unit i
x_{ab}	Line reactance between bus a and bus b
$X_{i0}^{on}, X_{i0}^{off}$	On/off time counter of unit i at the initial status

1 Introduction

In restructured power systems, self-interested entities, including generation companies (GENCOs), transmission companies (TRANSCOs), and distribution companies (DISCOs), are to maximize their own profits and minimize potential risks when participating in power markets. The independent system operator (ISO) coordinates market participants and operates the competitive market efficiently for ensuring a secure and economic operation. Such decisions corresponding to the power system operation, maintenance, and planning are to be made via efficient optimization models and methodologies [1]. Efficient decomposition methods are utilized in large power system optimization problems since most of the optimization problems belong to non-deterministic polynomial-time hard (NP-hard) problems and the solution to the original problem in such cases would be an intractable task without decomposition.

The BD algorithm was first introduced by J.F. Benders [2] for solving large-scale, mixed-integer programming (MIP) problems. BD is a common technique for separating large scale power system problems into several easy-to-solve subproblems. The algorithm is widely adopted in various power system optimization problems including security-constrained unit commitment (SCUC) and security-constrained optimal power flow (SCOPF) [3–9]. Reference [3] applied BD to SCUC problems for separating the unit commitment (UC) in the master problem from the hourly network security check in subproblems. The subproblems checked the ac network security constraints based on the UC solution to determine whether a converged and secure ac power flow solution would be feasible. If the feasibility conditions could not be satisfied, the feasibility Benders cuts were fed back to the master UC problem for seeking a feasible UC solution based on the added cuts. Reference [4] considered BD applications to generation and transmission planning problems by decoupling the master planning problems from the network evaluation and network security checking subproblems. Reference [5] used the feasibility theorem of BD to derive an analytical condition for determining whether feasible solutions of the original SCUC problems could be obtained from the current dual solution of Lagrangian Relaxation. Reference [6] proposed a BD-based algorithm for calculating a preventive dispatch solution based on a full ac power flow. It solved the optimal power flow in the master problem, and then minimized the real and reactive power by fictitious sources in subproblems. Reference [7] used the generalized BD to solve the large-scale multi-period optimal power flow problem. The problem considered the start-up and shut-down characteristics of thermal units, the transmission network model, and hydraulic equations for integrated hydroelectric plants in a river system. Optimality cuts generated at each iteration provided a lower estimate of total operation cost in the Benders subproblem. Reference [8] applied generalized feasibility Benders cuts for optimizing the post-contingency corrective action in SCUC, which accurately formulated the hourly UC of quick-start units in post-contingency corrective actions via a MIP subproblem instead of linear programming (LP). Reference [9] discussed a general structure of BD for power system decision making applications. Three categories of decision-making problems in power systems for economic, reliability, and risk evaluations were mapped onto Benders subproblems. The applications demonstrated the potential use of BD for solving special structured MIP problems in power systems.

The classical BD algorithm often converges slowly, i.e., the slow convergence would introduce major computational bottlenecks for a MIP problem which has to be solved repeatedly in practical cases. Several studies discussed possible improvements for accelerating the BD convergence. Reference [10] discussed several ways of improving the performance of BD including a judicious selection of optima from the Benders subproblems to generator strong cuts and properly formulate the subsequent BD problems. Reference [11] proposed to generate cuts from the solution of a LP-approximated master problem by relaxing the integrality request for the first several iterations. Similarly, [12] discussed that only a feasible, rather than an optimal, solution of master problem was necessary for subproblems to generate cuts. Therefore, heuristic methods were sought, in place of the optimal solution of master problem, to quickly locate feasible solutions for generating cuts. In this situation, the convergence was not guaranteed since the cuts were only associated with a set of feasible master solutions. On the other hand, [13] proposed to generate inexact cuts which were to derive a good enough dual solution, instead of the optimal dual solution, for subproblems. In such cases, subproblems were very large-scale LP problems which required a substantial computation time to get the solution. Reference [14] applied the subsystems of an infeasible LP to generate feasibility cuts, and computational results showed that the method converged faster than that with feasibility cuts generated via extreme rays located by the simplex method. Recently, [15] proposed an algorithm to accelerate the BD by exploring the local branching in the neighborhood of the master problem solution. The lower and the upper bounds were simultaneously improved via local branching in which each located feasible solution was used to generator optimality/feasibility cuts.

This paper proposes multiple strong Benders cuts at each iteration to accelerate the convergence of the NCUC problem. The experiments have shown that the drawback of classical BD algorithm resides in the slow convergence for large-scale power system applications [4]. As will be proved in this paper, each strong Benders cut generated through the proposed algorithm is a pareto optimal cut, and the proposed method reduces the number of Benders iterations as well as the overall computing time. The proposed BD strategy can be easily extended to other large-scale optimization problems in power systems.

The rest of the paper is organized as follows. Section 2 presents the NCUC model and the classical BD algorithm. Section 3 proposes two methods for generating multiple strong cuts at each iteration for accelerating convergence, and in the appendix we prove that each generated cut is pareto optimal. Section 4 presents three cases to discuss the effectiveness of the proposed method. The conclusions are drawn in Sect. 5.

2 NCUC problem formulation and solution methodology

The NCUC problem is formulated as a MIP problem shown in (1)–(8). The objective function is to minimize the total operation cost, including incremental cost, no-load cost, and startup/shutdown cost, as shown in (1), subject to a set of system and generating unit constraints. System constraints include the system power balance in (2), and the system spinning and non-spinning reserve requirements in (3). Unit constraints include ramping up/down limits as well as startup/shutdown ramping for individual units in (4), minimum on and off time constraints in (5), real power generation limits in (6), and constraints on spinning and non-spinning reserves provided by individual generators in (7). Only thermal units are modeled in (4)–(7). The formulation of other types of generating units including combined-cycle, cascaded hydro, and pumped storage, is discussed in the authors' previous work [19]. Here, (8) represents the dc network evaluation constraints, which include power balance for each bus b, power flow equation for each transmission line l, and phase shifter and transmission power flow limits. For simplicity, P_{Losst} is given initially as an input which is estimated as a percentage of the system load P_{Dt}. A more accurate loss formulation is given in (9) if the loss factor LF_b is known, which represents the sensitivity of power loss to the bus power injection

$$G = \text{Min} \sum_t \sum_i \left[\sum_k c_{ik} \cdot P_{ikt} + N_i \cdot I_{it} + SU_{it} + SD_{it} \right] \tag{1}$$

$$\text{S.t.} \quad \sum_i P_{it} = P_{Dt} + P_{Losst} \tag{2}$$

$$\sum_i SR_{it} \geq R_{St}$$
$$\sum_i OR_{it} \geq R_{Ot} \tag{3}$$

$$P_{it} - P_{i(t-1)} \leq UR_i \cdot I_{i(t-1)} + UP_i \cdot (I_{it} - I_{i(t-1)}) + P_i^{\max} \cdot (1 - I_{it})$$
$$P_{i(t-1)} - P_{it} \leq DR_i \cdot I_{it} + DP_i \cdot (I_{i(t-1)} - I_{it}) + P_i^{\max} \cdot (1 - I_{i(t-1)}) \tag{4}$$

$$\sum_{t=1}^{UT_i} (1 - I_{it}) = 0, \quad \text{where } UT_i = \max\{0, \min[NT, (T_i^{on} - X_{i0}^{on}) \cdot I_{i0}]\}$$
$$\sum_{\tau=t}^{t+T_i^{on}-1} I_{i\tau} \geq T_i^{on} \cdot (I_{it} - I_{i(t-1)}), \quad \forall t = UT_i + 1, \ldots, NT - T_i^{on} + 1$$
$$\sum_{\tau=t}^{NT} [I_{i\tau} - (I_{it} - I_{i(t-1)})] \geq 0, \quad \forall t = NT - T_i^{on} + 2, \ldots, NT$$
$$\sum_{t=1}^{DT_i} I_{it} = 0, \quad \text{where } DT_i = \max\{0, \min[NT, (T_i^{off} - X_{i0}^{off}) \cdot (1 - I_{i0})]\} \tag{5}$$
$$\sum_{\tau=t}^{t+T_i^{off}-1} (1 - I_{i\tau}) \geq T_i^{off} \cdot (I_{i(t-1)} - I_{it}),$$
$$\forall t = DT_i + 1, \ldots, NT - T_i^{off} + 1$$
$$\sum_{\tau=t}^{NT} [1 - I_{i\tau} - (I_{i(t-1)} - I_{it})] \geq 0, \quad \forall t = NT - T_i^{off} + 2, \ldots, NT$$

$$P_{it} = P_i^{\min} \cdot I_{it} + \sum_k P_{ikt}$$
$$0 \leq P_{ikt} \leq P_{ik}^{\max} \cdot I_{it} \tag{6}$$
$$P_i^{\min} \cdot I_{it} \leq P_{it} \leq P_i^{\max} \cdot I_{it}$$

$$P_{it} + SR_{it} \leq P_i^{\max} \cdot I_{it}$$
$$SR_{it} \leq 10 \cdot MSR_i \cdot I_{it} \tag{7}$$
$$OR_{it} = SR_{it} + (1 - I_{it}) \cdot QSC_i$$

$$\sum_{l \in \mathbf{f}(b)} PL_{lt} - \sum_{l \in \mathbf{t}(b)} PL_{lt} = \sum_{i \in \mathbf{U}(b)} P_{it} - D_{bt} - C_b \cdot P_{Losst}$$
$$PL_{lt} - (\theta_{at} - \theta_{bt} - \gamma_{abt})/x_{ab} = 0 \quad \text{line } l \text{ is from bus } a \text{ to bus } b \tag{8}$$
$$\gamma^{\min} \leq \gamma_t \leq \gamma^{\max}$$

$$-\mathbf{PL}^{\max} \le \mathbf{PL}_t \le \mathbf{PL}^{\max}$$
$$\theta_{ref} = 0$$
$$P_{Lt} - \sum_b LF_b \cdot \left(\sum_{i \in \mathbf{U}(b)} P_{it} - D_{bt} \right) - offset = 0 \tag{9}$$

BD is employed to solve the large-scale NCUC problem with a large number of constraints including those in the hourly network evaluation subproblems (8). In applying BD, the original problem described as (1)–(8) is decomposed into a master UC problem (10) and several hourly network evaluation subproblems (11). The master is a MIP problem and subproblems are LP problems. The subproblems examine the master problem solution for satisfying the network constraints. If the subproblem at hour t is infeasible, the corresponding Benders cut in (12) is generated and added to the next iteration of the master UC problem. The iterative process will continue by successively adding Benders cuts until network violations are mitigated. In (11) a unique non-negative slack variable s_t is introduced for all buses. The physical interpretation of a single non-negative slack variable is to diminish the most violated bus power balance iteratively. Reference [16] reported that a single non-negative slack variable strategy would outperform multiple non-negative slack variables at different buses by reducing the overall number of iterations

$$\begin{aligned} & LB = \text{Min} \sum_t \sum_i \left[\sum_k c_{ik} \cdot P_{ikt} + N_i \cdot I_{it} + SU_{it} + SD_{it} \right] \\ & \text{S.t.} \quad \text{Equations (2)–(7)} \\ & \qquad \text{Benders cuts from all previous iterations} \end{aligned} \tag{10}$$

$$\begin{aligned} & \text{Min} \quad s_t \\ & \text{S.t.} \quad \sum_{l \in \mathbf{f}(b)} PL_{lt} - \sum_{l \in \mathbf{t}(b)} PL_{lt} - s_t \le \sum_{i \in \mathbf{U}(b)} P_{it} - D_{bt} - C_b \cdot P_{Losst} && \lambda_{1,bt} \\ & \qquad - \sum_{l \in \mathbf{f}(b)} PL_{lt} + \sum_{l \in \mathbf{t}(b)} PL_{lt} - s_t \le - \sum_{i \in \mathbf{U}(b)} P_{it} + D_{bt} + C_b \cdot P_{Losst} && \lambda_{2,bt} \\ & \qquad PL_{lt} - (\theta_{at} - \theta_{bt} - \gamma_{abt})/x_{ab} = 0 && \kappa_{lt} \\ & \qquad \gamma_{pt} \le \gamma_p^{\max} && \pi_{1,pt} \\ & \qquad -\gamma_{pt} \le -\gamma_p^{\min} && \pi_{2,pt} \\ & \qquad PL_{lt} \le PL_l^{\max} && \pi_{3,lt} \\ & \qquad -PL_{lt} \le PL_l^{\max} && \pi_{4,lt} \\ & \qquad \theta_{ref,t} = 0 && \pi_t \\ & \qquad 0 \le s_t \end{aligned} \tag{11}$$

If the optimal objective value $\hat{s}_t$ in (11) is larger than the predefined threshold ε, the feasibility Benders cut represented by (12) will be utilized by grouping generators that are located at the same bus

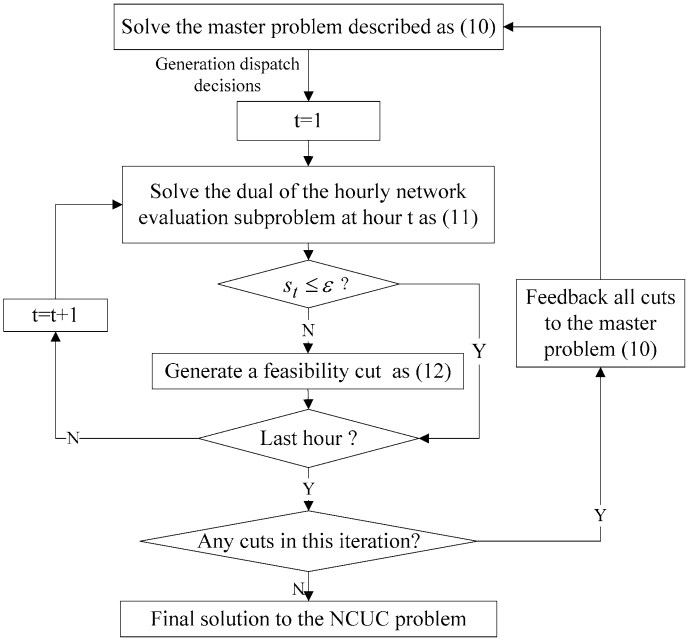


Fig. 1 NCUC procedure with the classical BD method

$$\sum_{b=1}^{NB} \sum_{i\in \mathbf{U}(b)} (\hat{\lambda}_{1b,t} - \hat{\lambda}_{2b,t}) \cdot (P_{it} - \hat{P}_{it}) + \hat{s}_t \le 0 \tag{12}$$

$\mathbf{U}(b)$ represents the set of generators located at the same bus b that shares the same dual variables $\hat{\lambda}_{1b,t}$ and $\hat{\lambda}_{2b,t}$, and $\hat{s}_t$ is the current optimal objective value of (11). Figure 1 shows the flowchart of the NCUC problem with the classical BD method. The master UC problem (10) is solved and the dispatch solution $\hat{P}_{it}$ is passed on to the hourly network evaluation subproblems (11). If there is any network violations (i.e. $s_t > \varepsilon$), a feasibility cut in (12) corresponding to the violated hour is generated and fed back to the master UC problem for further iterations. The procedure stops when $s_t \le \varepsilon$ is satisfied at all hours, no more feasibility Benders cuts are generated, and the final solution to the original NCUC problem is obtained. At each Benders iteration d, the objective value of (10) provides a lower bound LB_d to the original problem, and $UB_d = LB_d + M \cdot \sum_t s_{dt}$ provides an upper bound to the original problem where s_{dt} is the solution of (11) representing the violation at hour t at Benders iteration d. It is noted that by successively appending feasibility cuts to (10) the lower bound obtained at each iteration will be monotonically increasing, i.e. $LB_{d-1} \le LB_d \le G$ and $LB_{best} = LB_d$, but that will not be the case with the upper bound. The upper bound for the current iteration may be lower or higher than that in previous iterations, and the best upper bound is the minimum of all obtained values, i.e. $G \le UB_{best} = \min_d\{UB_d\}$. Assuming the current Benders iteration is d, and the best upper bound is obtained at iteration d', the stopping criterion $s_t \le \varepsilon$ is $\frac{UB_{best}-LB_{best}}{LB_{best}} = \frac{(LB_{d'}+M\cdot\sum_t s_{d't})-LB_d}{LB_d} \le \frac{M}{LB_d} \cdot NT \cdot \varepsilon$ where $s_{d't}$ is the violation at hour

t obtained at the Benders iteration d'. That is, with a small enough ε, the stopping criterion will force LB_{best} and UB_{best} to be close, which will restrict the final solution to be close enough to the global optimal solution of the original problem. Here, s_t in (11) represents the maximum power imbalance among all buses at hour t; thus a non-zero s_t means the final solution is not strictly feasible to the original NCUC problem in the sense that there will be at most ε MW violation at each bus. The ε values of 0.01–0.1 are used in our case studies. For power systems with peak loads of thousands to tens of thousands of MW, the violation of less than 0.01–0.1 MW is tolerable and practical. The NCUC is used to determine day-head UC schedules (i.e. 24 hours), and small violations less than 0.01–0.1 MW can be handled by the real-time power system operation with automatic generation control schemes [1].

The linear sensitivity factor (LSF) method could also be used to solve the NCUC problem [1], which replaces (8) with (13). LSF represents the sensitivity of line flows to the bus power injection. The power flow in each line can be calculated by (13), where LSF_{ab}^{m} represents the sensitivity of power flow of line l (from bus a to bus b) to power injection at bus m. The detailed calculation procedure of LSF_{ab}^{m} is discussed in [1]. Here, $(\sum_{i:i\in\mathbf{U}(m)} P_{it} - D_{mt} - C_m \cdot P_{Losst} + \sum_{n:n\in\gamma(m)} \gamma_{mn}/x_{mn} - \sum_{n:m\in\gamma(n)} \gamma_{mn}/x_{mn})$ represents the total power injection to bus m, where items represent total generation, load demand, losses distributed at bus m, and equivalent power injection from phase shifters to bus m, respectively. Using the LSF method for the solution of NCUC problem, the UC problem described as (1)–(7) is solved first, then we check the power flow on each transmission line at each hour using (13), and append the violated (13) back to the master UC problem for mitigating violations. The procedure will stop when (13) is satisfied for all lines at all hours, and no transmission power flow violation exists. In the case study section, we will see that the LSF method may be efficient for small systems. However, since there is one constraint for each violated transmission line at each hour, LSF may introduce many more constraints than the BD method for large-scale power systems, which makes the UC problem difficult to solve in real time. Thus, decomposition is the only viable option for the solution of the large-scale NCUC problem in real time

$$\begin{aligned} -PL_l^{\max} \le PL_{lt} = \sum_m LSF_{ab}^m \cdot \Bigg(\sum_{i\in\mathbf{U}(m)} P_{it} - D_{mt} - C_m \cdot P_{Losst} \\ + \sum_{n\in\gamma(m)} \gamma_{mn}/x_{mn} - \sum_{m\in\gamma(n)} \gamma_{mn}/x_{mn} \Bigg) \le PL_l^{\max} \end{aligned} \tag{13}$$

3 Accelerating BD via multiple strong Benders cuts

The drawback of the classical BD algorithm presented in Sect. 2 resides in its slow convergence. In this section, we discuss our methodology for accelerating the BD procedure via multiple strong Benders cuts at each iteration. The proposed methodology for generating multiple strong Benders cuts reduces the number of Benders iterations as well as the overall computing time.

3.1 Generate multiple strong Benders cuts from multiple optimal dual solutions

The dual problem of (11) is formulated as (14) which is highly degenerate. In this case, there are many zeros on the right-hand-side with multiple optimal solutions. If one or more of the variables in the basis of a LP problem is zero, the basis is called degenerate. And the LP is highly degenerate if there are many vertices of the feasible region for which the associated basis is degenerate. Since the coefficients of Benders cut (12) are generated using the optimal dual solution of (14), the proper choice of the optimal dual solution and the corresponding Benders cuts would be of crucial importance for the efficiency of the BD algorithm

$$
\begin{aligned}
\text{Max} \quad & \sum_b \left(\sum_{i \in \mathbf{U}(b)} P_{it} - D_{bt} - C_b \cdot P_{Losst} \right) \cdot (\lambda_{1,bt} - \lambda_{2,bt}) \\
& + \sum_p (\gamma_p^{\max} \cdot \pi_{1,pt} - \gamma_p^{\min} \cdot \pi_{2,pt}) + \sum_l PL_l^{\max} \cdot (\pi_{3,lt} + \pi_{4,lt}) \\
\text{S.t.} \quad & (\lambda_{1,at} - \lambda_{2,at}) + (-\lambda_{1,bt} + \lambda_{2,bt}) + \kappa_{lt} + \pi_{3,lt} - \pi_{4,lt} = 0 \\
& \quad \text{line } l \text{ is from bus } a \text{ to } b \\
& \sum_a \kappa_{lt}/x_{am} - \sum_b \kappa_{l't}/x_{mb} = 0 \\
& \quad \text{line } l \text{ is from bus } a \text{ to } m, \text{ line } l' \text{ is from bus } m \text{ to } b, \\
& \quad m \text{ is not reference bus} \\
& \sum_a \kappa_{lt}/x_{am} - \sum_b \kappa_{l't}/x_{mb} + \pi_t = 0 \\
& \quad \text{line } l \text{ is from bus } a \text{ to } m, \text{ line } l' \text{ is from} \\
& \quad \text{bus } m \text{ to } b, m \text{ is reference bus} \\
& \kappa_{lt}/x_{ab} + \pi_{1,pt} - \pi_{2,pt} = 0 \\
& \quad \text{phase shifter } p \text{ is located at line } l, \text{ which is from bus } a \text{ to } b \\
& -\sum_{b=1}^{NB} (\lambda_{1b,t} + \lambda_{2b,t}) \le 1 \\
& \lambda_{1,bt}, \lambda_{2,bt}, \pi_{1,pt}, \pi_{2,pt}, \pi_{3,lt}, \pi_{4,lt} \le 0, \quad \kappa_{lt}, \pi_t \text{ free}
\end{aligned}
\tag{14}
$$

All buses are divided into exclusive subsets $\mathbf{B}_o$ that satisfy $\bigcup_o \mathbf{B}_o = \mathbf{B}$ and $\mathbf{B}_o \cap \mathbf{B}_{o'} = \mathbf{\Phi}$, where $\mathbf{B}$ is the set of all buses in the system. A set of new subproblems (15) is generated corresponding to each subset $\mathbf{B}_o$. The idea behind (15) is to find an optimal dual solution which can generate a cut for restricting the feasible region of P_i $i \in \mathbf{U}(b)$ and $b \in \mathbf{B}_o$ to the highest possible extent. An optimal solution of (15) is also an optimal solution to the dual problem (14) since all constraints in (14) are included in (15) and the fifth constraint in (15) would force the objective value to be equal to its optimal value $\hat{s}$, calculated by (11)

$$
\begin{aligned}
\text{Max} \quad & \sum_{b\in\mathbf{B}_o}\sum_{i\in\mathbf{U}(b)}(P_i^{\max}-\varepsilon)\cdot(\lambda_{1b,t}-\lambda_{2b,t}) \\
& \quad + \sum_{b\notin\mathbf{B}_o}\sum_{i\in\mathbf{U}(b)}(P_i^{\min}+\varepsilon)\cdot(\lambda_{1b,t}-\lambda_{2b,t}) \\
\text{S.t.} \quad & (\lambda_{1,at}-\lambda_{2,at})+(-\lambda_{1,bt}+\lambda_{2,bt})+\kappa_{lt}+\pi_{3,lt}-\pi_{4,lt}=0 \\
& \quad \text{line } l \text{ is from bus } a \text{ to } b \\
& \sum_a \kappa_{lt}/x_{am}-\sum_b \kappa_{l't}/x_{mb}=0 \\
& \quad \text{line } l \text{ is from bus } a \text{ to } m, \text{ line } l' \text{ is} \\
& \quad \text{from bus } m \text{ to } b, m \text{ is not reference bus} \\
& \sum_a \kappa_{lt}/x_{am}-\sum_b \kappa_{l't}/x_{mb}+\pi_t=0 \\
& \quad \text{line } l \text{ is from bus } a \text{ to } m, \text{ line } l' \text{ is from bus } m \text{ to } b, m \text{ is reference bus} \\
& \kappa_{lt}/x_{ab}+\pi_{1,pt}-\pi_{2,pt}=0 \\
& \quad \text{phase shifter } p \text{ is located at line } l, \text{ which is from bus } a \text{ to } b \\
& \sum_b\Bigg(\sum_{i\in\mathbf{U}(b)}P_{it}-D_{bt}-C_b\cdot P_{Losst}\Bigg)\cdot(\lambda_{1,bt}-\lambda_{2,bt}) \\
& \quad +\sum_p(\gamma_p^{\max}\cdot\pi_{1,pt}-\gamma_p^{\min}\cdot\pi_{2,pt})+\sum_l PL_l^{\max}\cdot(\pi_{3,lt}+\pi_{4,lt})=\hat{s}_t \quad Q \\
& -\sum_{b=1}^{NB}(\lambda_{1b,t}+\lambda_{2b,t})\le 1 \\
& \lambda_{1,bt},\lambda_{2,bt},\pi_{1,pt},\pi_{2,pt},\pi_{3,lt},\pi_{4,lt}\le 0, \quad \kappa_{lt},\pi_t \text{ free}
\end{aligned}
\tag{15}
$$

The new cut (16) is generated from the optimal solution of (15).

$$
\sum_{b=1}^{NB}\sum_{i\in\mathbf{U}(b)}(\tilde{\lambda}_{1,bt}-\tilde{\lambda}_{2,bt})\cdot P_{it}+\hat{s}_t\le 0 \tag{16}
$$

where $\tilde{\lambda}$ is the new optimal solution to (15).

Definition 1 (See also Magnanti and Wong 1981 [10]) A cut for the minimization of (1)–(8) as $\sum_{b=1}^{NB}\sum_{i\in\mathbf{U}(b)}(\lambda_{1b,t}-\lambda_{2b,t})\cdot(P_{it}-\hat{P}_{it})+\hat{s}_t\le 0$ would dominate or is stronger than $\sum_{b=1}^{NB}\sum_{i\in\mathbf{U}(b)}(\lambda'_{1b,t}-\lambda'_{2b,t})\cdot(P_{it}-\hat{P}_{it})+\hat{s}_t\le 0$ as another cut, if for all $\mathbf{P}_t$ in their domain $\{\mathbf{0}, [\mathbf{P}^{\min}, \mathbf{P}^{\max}]\}$, (17) is satisfied with a strict inequality for any point in the feasible region.

$$\left[\sum_{b=1}^{NB}\sum_{i\in\mathbf{U}(b)}(\lambda'_{1b,t}-\lambda'_{2b,t})\cdot(P_{it}-\hat{P}_{it})+\hat{s}_t\right]$$
$$-\left[\sum_{b=1}^{NB}\sum_{i\in\mathbf{U}(b)}(\lambda_{1b,t}-\lambda_{2b,t})\cdot(P_{it}-\hat{P}_{it})+\hat{s}_t\right]\leq 0 \tag{17}$$

Definition 2 (See also Magnanti and Wong 1981 [10]) A cut is pareto optimal if no other cut could dominate it.

Theorem 1 *The cut* (16) *generated via multiple optimal dual solutions is pareto optimal.*

The proof of Theorem 1 is provided in Appendix A. The problem considered in Magnanti and Wong 1981 [10] is a convex hull, and the points $\mathbf{P}_t$ we discuss here belong to a non-convex set $\{\mathbf{0}, [\mathbf{P}^{\min}, \mathbf{P}^{\max}]\}$. The dual to (15), which corresponds to the primal form of the original problem (11), is given in (18). Since dual problems (14)–(15) would introduce many more variables than the primal problem (11), we have experienced that it is more efficient to solve the NCUC problem in its primal form (11) and (18) than its dual form (14)–(15). Furthermore, subproblems (11) and (18) are large and sparse and could experience a higher convergence speed by using a primal-dual logarithmic barrier algorithm provided by CPLEX [17]. This algorithm is suitable for large LP problems with dense columns; that is, a relatively high number of nonzero entries in each column

$$\begin{aligned}
\text{Min}\quad & \hat{s}_t\cdot Q+s_t\\
\text{S.t.}\quad & \sum_{l\in\mathbf{f}(b)}PL_{lt}-\sum_{l\in\mathbf{t}(b)}PL_{lt}-s_t+\left(\sum_{i\in\mathbf{U}(b)}P_{it}-D_{bt}-C_b\cdot P_{Losst}\right)\cdot Q\\
&\quad\leq RQ1_{bt}\\
& -\sum_{l\in\mathbf{f}(b)}PL_{lt}+\sum_{l\in\mathbf{t}(b)}PL_{lt}-s_t-\left(\sum_{i\in\mathbf{U}(b)}P_{it}-D_{bt}-C_b\cdot P_{Losst}\right)\cdot Q\\
&\quad\leq RQ2_{bt}\\
& PL_{lt}-(\theta_{at}-\theta_{bt}-\gamma_{abt})/x_{ab}=0\\
& \gamma_{pt}+\gamma_p^{\max}\cdot Q\leq\gamma_p^{\max}\\
& -\gamma_{pt}-\gamma_p^{\min}\cdot Q\leq-\gamma_p^{\min}\\
& PL_{lt}+PL_l^{\max}\cdot Q\leq PL_l^{\max}\\
& -PL_{lt}+PL_l^{\max}\cdot Q\leq PL_l^{\max}\\
& \theta_{ref,t}=0\\
& 0\leq s_t,\quad Q\text{ free}
\end{aligned}\tag{18}$$

where

$$RQ1_{bt} = \begin{cases} \sum_{i\in \mathbf{U}(b)} (P_i^{\max} - \varepsilon), & b \in \mathbf{B}_o \\ \sum_{i\in \mathbf{U}(b)} (P_i^{\min} + \varepsilon), & b \notin \mathbf{B}_o \end{cases} \quad \text{and}$$

$$RQ2_{bt} = \begin{cases} -\sum_{i\in \mathbf{U}(b)} (P_i^{\max} - \varepsilon), & b \in \mathbf{B}_o \\ -\sum_{i\in \mathbf{U}(b)} (P_i^{\min} + \varepsilon), & b \notin \mathbf{B}_o \end{cases}$$

In the implementation of (18), a strong Benders cut (16) is generated for each subset $\mathbf{B}_o$. We should point out that different forms of subsets $\mathbf{B}_o$ might conceivably generate different pareto optimal cuts. The number of subsets may also impact the convergence. More pareto optimal cuts generated in each iteration could result in fewer iterations between the master UC and network evaluation subproblems. However, the solution of the master problem may take longer because of the large number of cuts. Here, the optimal number of subsets would be a tradeoff between the computational burden of solving the master problem and the number of iterations. We use 2 subsets for off peak hours and 6 for peak hours in our case studies. Also, buses that are geographically close are placed into the same subset $\mathbf{B}_o$. In practice, the parameter tuning will be necessary in order to reduce the number of iterations and the overall computing time.

3.2 Generate multiple strong Benders cuts by enhancing the density of cuts

For an optimal dual solution of (14), $\hat{\lambda}_{1,bt}, \hat{\lambda}_{2,bt}, \hat{\pi}_{1,pt}, \hat{\pi}_{2,pt}, \hat{\pi}_{3,lt}, \hat{\pi}_{4,lt}, \hat{\kappa}_{lt}, \hat{\pi}_t$, all buses are divided into two subsets $\mathbf{Z}$ and $\mathbf{Z}^c$, with $\mathbf{Z}$ representing the set of buses b that satisfy $(\hat{\lambda}_{1,bt} - \hat{\lambda}_{2,bt}) = 0$, and $\mathbf{Z}^c$ representing the set of all other buses. Usually the Benders cut (12) is low density which means that a small number of decision variables $\mathbf{P}_t$ is involved. That is, very few coefficients of $\mathbf{P}_t$ are nonzero, which would limit the contribution of Benders cut (12) to strengthening the feasible region of the master UC problem. The idea here is to generate strong Benders cuts for strengthening the master UC problem at each iteration and incorporating more decision variables $\mathbf{P}_t$ at each Benders cut for reducing the number of iterations and the overall computing time.

Based on this idea, we divide the set $\mathbf{Z}$ into exclusive subsets $\mathbf{SZ}_o$, which satisfy $\bigcup_o \mathbf{SZ}_o = \mathbf{Z}$ and $\mathbf{SZ}_o \bigcap \mathbf{SZ}_{o'} = \boldsymbol{\Phi}$. Here buses in the set $\mathbf{Z}$ that are geographically close are included in the same subset $\mathbf{SZ}_o$. Different partitioning subsets may result in deriving different high density cuts and different convergence performances, which would require more research. A new LP optimization problem (19) is formulated corresponding to each subset $\mathbf{SZ}_o$ for generating high-density Benders cuts. The last inequality constraint in (19) is to guarantee that the problem (19) is bounded. We experience that the value of $2 \cdot |\mathbf{Z}|$ used in problem (19) does not impact the convergence performance of the proposed method significantly as illustrated in Case study 4.2 in which we replace $2 \cdot |\mathbf{Z}|$ with 1. The optimal solution of (19), $\hat{\lambda}_{1,bt}\hat{\lambda}_{2,bt}$ for $b \in \mathbf{Z}^c$ and $\bar{\lambda}_{1,bt}\bar{\lambda}_{2,bt}$ for $b \in \mathbf{Z}$, is not feasible for the network evaluation problem (11) or its dual (14) as it violates $-\sum_{b=1}^{NB}(\lambda_{1,bt} + \lambda_{2,bt}) \le 1$ in (14). Thus (20) is solved to get the final high-density Benders cut given in (21), where $c = \sum_{b\in \mathbf{Z}^c}(-\hat{\lambda}_{1,bt} - \hat{\lambda}_{2,bt}) + 2 \cdot |\mathbf{Z}|$

is constant with the given $\hat{\lambda}_{1,bt}$ and $\hat{\lambda}_{2,bt}$ for $b \in \mathbf{Z}^c$

$$
\begin{aligned}
\text{Max} \quad & \sum_{b\in \mathbf{SZ}_o}\sum_{i\in \mathbf{U}(b)} (P_i^{\max} - \varepsilon)\cdot(\lambda_{1b,t} - \lambda_{2b,t}) \\
& \quad + \sum_{b\in\{\mathbf{Z}-\mathbf{SZ}_o\}}\sum_{i\in \mathbf{U}(b)} (P_i^{\min} + \varepsilon)\cdot(\lambda_{1b,t} - \lambda_{2b,t}) \\
\text{S.t.} \quad & (\lambda_{1,at} - \lambda_{2,at}) + (-\lambda_{1,bt} + \lambda_{2,bt}) + \kappa_{lt} + \pi_{3,lt} - \pi_{4,lt} = 0 \\
& \quad \text{line } l \text{ is from bus } a \text{ to } b \\
& \sum_a \kappa_{lt}/x_{am} - \sum_b \kappa_{l't}/x_{mb} = 0 \\
& \quad \text{line } l \text{ is from bus } a \text{ to } m\text{, line } l' \text{ is from} \\
& \quad \text{bus } m \text{ to } b\text{, } m \text{ is not reference bus} \\
& \sum_a \kappa_{lt}/x_{am} - \sum_b \kappa_{l't}/x_{mb} + \pi_t = 0 \\
& \quad \text{line } l \text{ is from bus } a \text{ to } m\text{, line } l' \text{ is} \\
& \quad \text{from bus } m \text{ to } b\text{, } m \text{ is reference bus} \\
& \kappa_{lt}/x_{am} + \pi_{1,pt} - \pi_{2,pt} = 0 \\
& \quad \text{phase shifter } p \text{ is located at line } l\text{, which is from bus } a \text{ to } b \\
& \sum_b \left(\sum_{i\in \mathbf{U}(b)} P_{it} - D_{bt} - C_b\cdot P_{Losst} \right)\cdot(\lambda_{1,bt} - \lambda_{2,bt}) \\
& \quad + \sum_p (\gamma_p^{\max}\cdot \pi_{1,pt} - \gamma_p^{\min}\cdot \pi_{2,pt}) + \sum_l PL_l^{\max}\cdot(\pi_{3,lt} + \pi_{4,lt}) = \hat{s}_t \quad Q \\
& \lambda_{1,bt} = \hat{\lambda}_{1,bt}, \quad b \in \mathbf{Z}^c, \quad q_{1,bt} \\
& \lambda_{2,bt} = \hat{\lambda}_{2,bt}, \quad b \in \mathbf{Z}^c, \quad q_{2,bt} \\
& \sum_{b\in \mathbf{Z}} (-\lambda_{1,bt} - \lambda_{2,bt}) \le 2\cdot|\mathbf{Z}| \quad q \\
& \lambda_{1,bt}, \lambda_{2,bt}, \pi_{1,pt}, \pi_{2,pt}, \pi_{3,lt}, \pi_{4,lt} \le 0, \quad \kappa_{lt}, \pi_t \text{ free}
\end{aligned}
\tag{19}
$$

$$
\begin{aligned}
\text{Max} \quad & \breve{s}_t = \sum_b \left(\sum_{i\in \mathbf{U}(b)} P_{it} - D_{bt} - C_b\cdot P_{Losst} \right)\cdot(\lambda_{1,bt} - \lambda_{2,bt}) \\
& \quad + \sum_p (\gamma_p^{\max}\cdot \pi_{1,pt} - \gamma_p^{\min}\cdot \pi_{2,pt}) + \sum_l PL_l^{\max}\cdot(\pi_{3,lt} + \pi_{4,lt})
\end{aligned}
$$

$$
\begin{aligned}
\text{S.t.} \quad & (\lambda_{1,at} - \lambda_{2,at}) + (-\lambda_{1,bt} + \lambda_{2,bt}) + \kappa_{lt} + \pi_{3,lt} - \pi_{4,lt} = 0 \\
& \quad \text{line } l \text{ is from bus } a \text{ to } b \\
& \sum_a \kappa_{lt}/x_{am} - \sum_b \kappa_{l't}/x_{mb} = 0 \\
& \quad \text{line } l \text{ is from bus } a \text{ to } m, \text{ line } l' \text{ is from} \\
& \quad \text{bus } m \text{ to } b, m \text{ is not reference bus} \\
& \sum_a \kappa_{lt}/x_{am} - \sum_b \kappa_{l't}/x_{mb} + \pi_t = 0 \\
& \quad \text{line } l \text{ is from bus } a \text{ to } m, \text{ line } l' \text{ is} \\
& \quad \text{from bus } m \text{ to } b, m \text{ is reference bus} \qquad (20) \\
& \kappa_{lt}/x_{am} + \pi_{1,pt} - \pi_{2,pt} = 0 \\
& \quad \text{phase shifter } p \text{ is located at line } l, \text{ which is from bus } a \text{ to } b \\
& -\sum_{b=1}^{NB} (\lambda_{1,bt} + \lambda_{2,bt}) \le 1 \\
& \lambda_{1,bt} = \hat{\lambda}_{1,bt}/c, \quad \lambda_{2,bt} = \hat{\lambda}_{2,bt}/c, \quad b \in \mathbf{Z}^c, \quad q_{1,bt}, q_{2,bt} \\
& \lambda_{1,bt} = \bar{\lambda}_{1,bt}/c, \quad \lambda_{2,bt} = \bar{\lambda}_{2,bt}/c, \quad b \in \mathbf{Z}, \quad q_{1,bt}, q_{2,bt} \\
& \lambda_{1,bt}, \lambda_{2,bt}, \pi_{1,pt}, \pi_{2,pt}, \pi_{3,lt}, \pi_{4,lt} \le 0, \quad \kappa_{lt}, \pi_t \text{ free}
\end{aligned}
$$

Theorem 2 *The cut* (21) *is pareto optimal, given the fixed coefficients* $\hat{\lambda}_{1,bt}$ *and* $\hat{\lambda}_{2,bt}$ *corresponding to all buses in the subset* $b \in \mathbf{Z}^c$, *which is the optimal solution of* (14).

$$
\sum_{b \in \mathbf{Z}^c} \sum_{i \in \mathbf{U}(b)} (\hat{\lambda}_{1,bt}/c - \hat{\lambda}_{2,bt}/c) \cdot (P_{it} - \hat{P}_{it}) + \sum_{b \in \mathbf{Z}} \sum_{i \in \mathbf{U}(b)} (\bar{\lambda}_{1,bt}/c - \bar{\lambda}_{2,bt}/c) \cdot P_{it} + \breve{s}_t \le 0 \tag{21}
$$

The proof of Theorem 2 is provided in Appendix B. The duals to (19)–(20), which correspond to the primal form of the original problem (11), are given in (22)–(23). Similar to (14)–(15), the dual forms (19)–(20) introduce many more variables than their corresponding primal forms (22)–(23). Thus a higher efficiency could be achieved by solving primal problems (22)–(23)

$$
\begin{aligned}
\text{Min} \quad & \hat{s}_t \cdot Q + \sum_{b \in \mathbf{Z}^c} (\hat{\lambda}_{1,bt} \cdot q_{1,bt} + \hat{\lambda}_{2,bt} \cdot q_{2,bt}) + 2 \cdot |\mathbf{Z}| \cdot q \\
\text{S.t.} \quad & \sum_{l \in \mathbf{f}(b)} PL_{lt} - \sum_{l \in \mathbf{t}(b)} PL_{lt} \\
& \quad + \left(\sum_{i \in \mathbf{U}(b)} P_{it} - D_{bt} - C_b \cdot P_{Losst} \right) \cdot Q + Q1_{bt} \le RQ1_{bt}
\end{aligned}
$$

$$
\begin{aligned}
&-\sum_{l\in\mathbf{f}(b)} PL_{lt} + \sum_{l\in\mathbf{t}(b)} PL_{lt} \\
&\quad -\Bigg(\sum_{i\in\mathbf{U}(b)} P_{it} - D_{bt} - C_b\cdot P_{Losst}\Bigg)\cdot Q + Q2_{bt} \le RQ2_{bt} \\
&PL_{lt} - (\theta_{at} - \theta_{bt} - \gamma_{abt})/x_{ab} = 0 \\
&\gamma_{pt} + \gamma_p^{\max}\cdot Q \le 0 \\
&-\gamma_{pt} - \gamma_p^{\min}\cdot Q \le 0 \\
&PL_{lt} + PL_l^{\max}\cdot Q \le 0 \\
&-PL_{lt} + PL_l^{\max}\cdot Q \le 0 \\
&\theta_{ref,t} = 0 \\
&0 \le q, \quad Q, q_{1,bt}, q_{2,bt} \text{ free}
\end{aligned}
\tag{22}
$$

where

$$
RQ1_{bt} = \begin{cases} \sum_{i\in\mathbf{U}(b)}(P_i^{\max} - \varepsilon), & b\in\mathbf{SZ}_o, \\ \sum_{i\in\mathbf{U}(b)}(P_i^{\min} + \varepsilon), & b\in\mathbf{Z} - \mathbf{SZ}_o, \\ 0, & b\in\mathbf{Z}^c, \end{cases} \qquad RQ2_{bt} = -RQ1_{bt}
$$

$$
Q1_{bt} = \begin{cases} -q, & b\in\mathbf{Z}, \\ q_{1,bt}, & b\in\mathbf{Z}^c, \end{cases} \qquad Q2_{bt} = \begin{cases} -q, & b\in\mathbf{Z} \\ q_{2,bt}, & b\in\mathbf{Z}^c \end{cases}
$$

$$
\begin{aligned}
\text{Min} \quad & s_t + \sum_{b\in\mathbf{Z}^c}(\hat{\lambda}_{1,bt}\cdot q_{1,bt} + \hat{\lambda}_{2,bt}\cdot q_{2,bt})/c + \sum_{b\in\mathbf{Z}}(\hat{\lambda}_{1,bt}\cdot q_{1,bt} + \hat{\lambda}_{2,bt}\cdot q_{2,bt})/c \\
\text{S.t.} \quad & \sum_{l\in\mathbf{f}(b)} PL_{lt} - \sum_{l\in\mathbf{t}(b)} PL_{lt} - s_t + q_{1,bt} \le \sum_{i\in\mathbf{U}(b)} P_{it} - D_{bt} - C_b\cdot P_{Losst} \\
& -\sum_{l\in\mathbf{f}(b)} PL_{lt} + \sum_{l\in\mathbf{t}(b)} PL_{lt} - s_t + q_{2,bt} \le -\sum_{i\in\mathbf{U}(b)} P_{it} + D_{bt} + C_b\cdot P_{Losst} \\
& PL_{lt} - (\theta_{at} - \theta_{bt} - \gamma_{abt})/x_{ab} = 0 \\
& \gamma_{pt} \le \gamma_p^{\max} \\
& -\gamma_{pt} \le -\gamma_p^{\min} \\
& PL_{lt} \le PL_l^{\max} \\
& -PL_{lt} \le PL_l^{\max} \\
& \theta_{ref,t} = 0 \\
& 0 \le s_t, \quad q_{1,bt}, q_{2,bt} \text{ free}
\end{aligned}
\tag{23}
$$

In the implementation of (22)–(23), there is one high-density Benders cut (14) corresponding to each subsets $\mathbf{SZ}_o$, and different divisions of subsets $\mathbf{SZ}_o$ may conceivably generate different cuts. The number of subsets may also impact the convergence. The more cuts added to each iteration would result in fewer iterations between

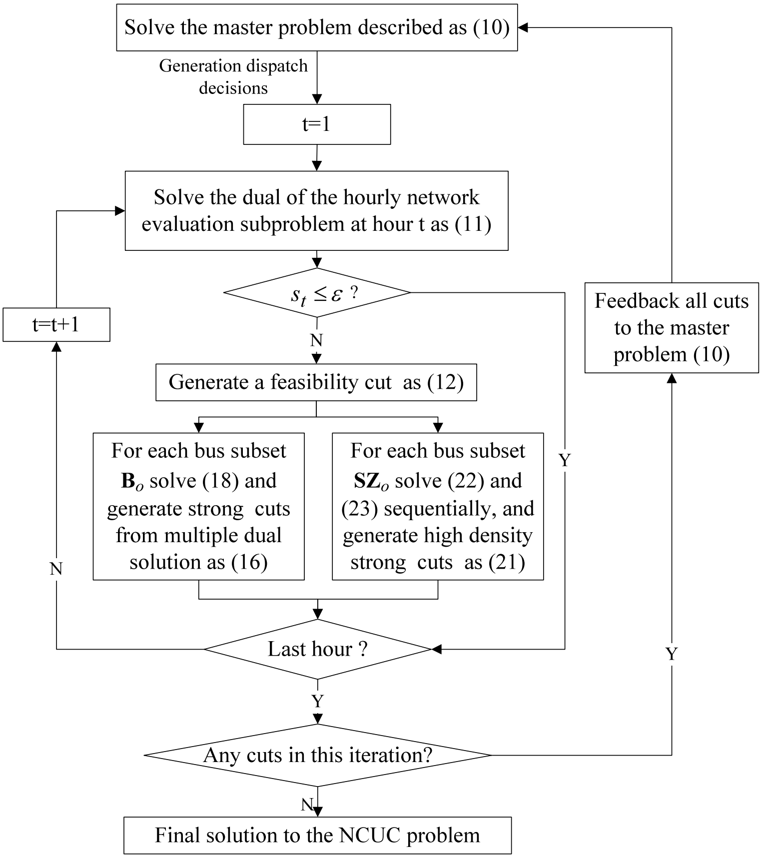


Fig. 2 NCUC procedure with the proposed accelerating BD algorithm

master UC and network evaluation subproblems, but the solution of master problem may take longer because of the large number of added cuts. Here, 5 subsets are used in our case studies.

The overall computing time is the time for solving the master UC problem and a series of network evaluation subproblems iteratively. Fewer iterations would result in a significant reduction in computing time. However, the calculation of additional cuts could be time consuming. Here subproblems (18) and (22)–(23) with respect to disjunctive subsets would be solved in parallel since they are independent, which will reduce the overall computing time.

Figure 2 shows the flowchart for solving the NCUC problem with the proposed accelerating BD method. In comparison to Fig. 1, if there is a network violation (i.e. $s_t > \varepsilon$), besides the feasibility cuts obtained in (12), we also solve a set of subproblems and generate strong Benders cuts for accelerating the BD algorithm. We solve (18) for each subset $\mathbf{B}_o$ and generate strong cuts from multiple dual solutions as in (16). Concurrently, we also solve (22)–(23) for each subset $\mathbf{SZ}_o$ and generate high density strong cuts as (21). Here, (18) corresponding to different subsets $\mathbf{B}_o$ and

(22)–(23) corresponding to different subset $\mathbf{SZ}_o$ can be solved in parallel. All cuts from (12), (16) and (21) are fed back to the master UC problem for further iterations. The same stopping criterion as that of the classical BD discussed in Sect. 2 is used. This paper does not intend to prove that high density cuts (21) are tighter than cuts (12) generated by the classical BD method. In fact, cuts (16) and (21) generated by the proposed method and the cut (12) generated from the classical BD are all pareto; that is, non of them dominates the others. We observed that the classical BD usually generates low density cuts which involve a small number of decision variables $\mathbf{P}_t$. That is, very few coefficients of $\mathbf{P}_t$ are nonzero. These cuts may not restrict significantly the feasible region of the master UC problem, thus more iterations may be needed. The idea here is to generate additional strong Benders cuts for incorporating more decision variables $\mathbf{P}_t$ at each generated Benders cut and strengthening the master UC problem at each iteration. As shown in numerical examples, the incorporation of multiple strong Benders cuts proposed here can strengthen the master UC problem at each iteration as compared with the single cut generated by the classical BD, and reduce the iteration number and overall computing time.

4 Case studies

Three cases are studied in this section to demonstrate the effectiveness of the proposed strong Benders cuts approach in terms of offering fewer Benders iterations and a shorter CPU time. The first case is a mathematical example which shows how Benders cuts are generated for multiple optimal dual solutions, which is common in large-scale power systems. The second one is a 3-bus power system which shows the effectiveness of high-density cuts. The third example is a large-scale 5663-bus system which shows both aspects of the previous examples. All cases are solved using CPLEX 11.0.0 on a 2.4 GHz personal computer.

4.1 Mathematical example

The problem is described as:

$$
\begin{aligned}
\text{Min}\quad & 6 \cdot P_1 + 8 \cdot P_2 + 10 \cdot P_3 \\
\text{S.t.}\quad & 1 \cdot I_1 \le P_1 \le 10 \cdot I_1, \qquad 1 \cdot I_2 \le P_2 \le 10 \cdot I_2, \qquad 1 \cdot I_3 \le P_3 \le 10 \cdot I_3 \\
& P_1 + P_2 + P_3 = 10, \qquad 2 \cdot x_1 + x_2 - 4 \cdot x_4 + P_1 \le 13 \\
& x_1 + x_3 - 0.5 \cdot P_2 \le 2, \qquad 3 \cdot x_2 + 0.5 \cdot x_3 + x_4 + P_3 \le 12.5 \\
& -x_1 - x_2 - x_3 \le -7, \qquad 0 \le x_1 \le 5, \qquad 0 \le x_2 \le 4, \qquad 0 \le x_3 \le 3 \\
& 0 \le x_4 \le 2, \qquad I_1, I_2, I_3 \in \{0, 1\}
\end{aligned}
$$

Table 1 shows the solution procedure with the classical BD algorithm. The optimal solution is reached after 3 iterations. The primal and dual subproblems for the first iteration are presented in Table 2. By solving the primal subproblem, we get an optimal dual solution of $\lambda_1 = 0$, $\lambda_2 = -1$, $\lambda_3 = 0$ with a feasibility Benders cut of $-P_2 \le -2$.

Table 1 Solution procedure with the classical BD algorithm

Iteration	Solution						Benders cuts
	I_1	I_2	I_3	P_1	P_2	P_3	
1	1	0	0	10	0	0	$-P_2 \le -2$
2	1	1	0	8	2	0	$0.0625 \cdot P_1 - 0.34375 \cdot P_2 + 0.25 \cdot P_3 \le -0.375$
3	1	1	0	7.538	2.462	0	–

Table 2 Primal and dual subproblems in the first iteration

Subproblem in the primal form			Subproblem in the dual form	
Min	s		Max	$3 \cdot \lambda_1 + 2 \cdot \lambda_2 + 12.5 \cdot \lambda_3 - 7 \cdot \pi_1 + 5 \cdot \pi_2$
S.t.	$2 \cdot x_1 + x_2 - 4 \cdot x_4 - s \le 13 - \hat{P}_1$	λ_1		$+ 4 \cdot \pi_3 + 3 \cdot \pi_4 + 2 \cdot \pi_5$
	$x_1 + x_3 - s \le 2 + 0.5 \cdot \hat{P}_2$	λ_2	S.t.	$2 \cdot \lambda_1 + \lambda_2 - \pi_1 + \pi_2 \le 0$
	$3 \cdot x_2 + 0.5 \cdot x_3 + x_4 - s \le 12.5 - \hat{P}_3$	λ_3		$\lambda_1 + 3 \cdot \lambda_3 - \pi_1 + \pi_3 \le 0$
	$-x_1 - x_2 - x_3 \le -7$	π_1		$\lambda_2 + 0.5 \cdot \lambda_3 - \pi_1 + \pi_4 \le 0$
	$0 \le x_1 \le 5$	π_2		$-4 \cdot \lambda_1 + \lambda_3 + \pi_5 \le 0$
	$0 \le x_2 \le 4$	π_3		$-\lambda_1 - \lambda_2 - \lambda_3 \le 1$
	$0 \le x_3 \le 3$	π_4		$\lambda_1, \lambda_2, \lambda_3, \pi_1, \pi_2, \pi_3, \pi_4, \pi_5 \le 0$
	$0 \le x_4 \le 2$	π_5		
	$0 \le s$			

The dual subproblem in Table 2 has multiple optimal solutions and the above solution is the one chosen by the LP solver. Based on the LP theory, there are infinite optimal solutions for this dual subproblem because the linear combination of any two optimal solutions is also an optimal solution. According to (18), we optimize two extra problems (i.e., subset $\mathbf{B}_1$ including all three buses, and subset $\mathbf{B}_2$ including non) as presented in Table 3. When shifted to the right-hand side, coefficients of P_1 and P_3 are negative and coefficients of P_2 is positive, by considering the $\mathbf{B}_1$ subset the most strict feasibility region of P_2 is found that can be fed back to the master problem. Similarly, by considering non in $\mathbf{B}_2$ subset, we find the most restrict feasibility region of P_1 and P_3. Accordingly, a new dual optimal solution is obtained for both problems as $\lambda_1 = -0.0625, \lambda_2 = -0.6875, \lambda_3 = -0.25$ with the corresponding cut of $0.0625 \cdot P_1 - 0.34375 \cdot P_2 + 0.25 \cdot P_3 \le -0.375$. The problem is solved in two iterations with multiple strong Benders cuts generated via multiple dual optimal solutions and shown in Table 4. Here, a reduction of 33.3% is achieved in the number of iterations while the same global optimal solution is obtained. For comparison, this problem is solved directly by CPLEX without decomposition. The computation time with CPLEX is 0.11 second, with the global optimal solution of $P_1 = 7.538, P_2 = 2.462, x_1 = 3.231, x_2 = 3.769, x_4 = 1.192, I_1 = I_2 = 1$ which is the same as that obtained by the proposed accelerating BD algorithm.

Table 3 Subproblems for obtaining multiple dual optimal solutions as shown in (18)

Subproblem corresponding to subset $\mathbf{B}_1$		Subproblem corresponding to subset $\mathbf{B}_2$	
Min	$s+Q$	Min	$s+Q$
S.t.	$2\cdot x_1+x_2-4\cdot x_4-s+3\cdot Q\le -9.99$	S.t.	$2\cdot x_1+x_2-4\cdot x_4-s+3\cdot Q\le -1.01$
	$x_1+x_3-s+2\cdot Q\le 4.995$		$x_1+x_3-s+2\cdot Q\le 0.505$
	$3\cdot x_2+0.5\cdot x_3+x_4-s+12.5\cdot Q\le -9.99$		$3\cdot x_2+0.5\cdot x_3+x_4-s+12.5\cdot Q\le -1.01$
	$-x_1-x_2-x_3-7\cdot Q\le 0$		$-x_1-x_2-x_3-7\cdot Q\le 0$
	$x_1+5\cdot Q\le 5$		$x_1+5\cdot Q\le 5$
	$x_2+4\cdot Q\le 4$		$x_2+4\cdot Q\le 4$
	$x_3+3\cdot Q\le 3$		$x_3+3\cdot Q\le 3$
	$x_4+2\cdot Q\le 2$		$x_4+2\cdot Q\le 2$
	$0\le x_1,x_2,x_3,x_4,s \quad Q$ free		$0\le x_1,x_2,x_3,x_4,s \quad Q$ free

Table 4 Solution procedure using the proposed multiple cuts from multiple optimal dual solutions

Iteration	Solution						Benders cuts
	I_1	I_2	I_3	P_1	P_2	P_3	
1	1	0	0	10	0	0	$-P_2\le -2$
							$0.0625\cdot P_1-0.34375\cdot P_2+0.25\cdot P_3\le -0.375$
2	1	1	0	7.538	2.462	0	–

4.2 3-bus system

A 3-bus system shown in Fig. 3 with three generators, three transmission lines, and one load is studied. Three generators are located at three buses respectively, with parameters listed in Table 5. Transmission line capacities are also given in Fig. 3. Reactance of all lines is 0.1 p.u. and bus 1 is the reference bus. A one-hour NCUC problem described by (1)–(2), (6) and (8) is considered to find the optimal UC decisions and generation dispatch with minimum operation cost, while satisfying the system load of 450 MW located at bus 3. For this one-hour case study, other constraints, such as reserve requirement, system losses, generator minimum on/off time constraints, and ramping up/down limits, are relaxed for the sake of discussion. The problem is formulated as follows with constraints in each row representing the system load balance, generation capacity limits for each unit, power balance for each bus, dc power flow equation for each transmission line, power flow capacity limits for each line, and the reference bus identification

$$\begin{aligned}
\text{Min}\quad & 10\cdot P_{11}+11\cdot P_{21}+12\cdot P_{31}\\
\text{S.t.}\quad & P_{11}+P_{21}+P_{31}=450\\
& 100\cdot I_{11}\le P_{11}\le 500\cdot I_{11},\qquad 90\cdot I_{21}\le P_{21}\le 400\cdot I_{21}\\
& 30\cdot I_{31}\le P_{31}\le 200\cdot I_{31}\\
& PL_{11}+PL_{31}=P_{11},\qquad PL_{21}-PL_{11}=P_{21}
\end{aligned}$$

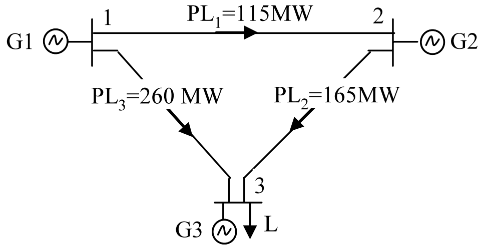


Fig. 3 One-line diagram of the 3-bus system

Table 5 Generator parameters

Generator	$P^{\min}$ (MW)	$P^{\max}$ (MW)	Cost function ($/MWh)
1	100	500	10
2	90	400	11
3	30	200	12

Table 6 Solution procedure using classical BD algorithm

Iteration	UC solution (MW)			Benders cuts
	P_{11}	P_{21}	P_{31}	
1	450	0	0	$P_{11} \le 375$
2	360	90	0	$-P_{31} \le -25$
3	375	0	75	$P_{11} - P_{21} \le 345$
4	330	90	30	$P_{21} - P_{31} \le 45$
5	315	90	45	–

$$
\begin{aligned}
&-PL_{21} - PL_{31} = P_{31} - 450\\
&PL_{11} - (\theta_{11} - \theta_{21})/0.1 = 0, \qquad PL_{21} - (\theta_{21} - \theta_{31})/0.1 = 0\\
&PL_{31} - (\theta_{11} - \theta_{31})/0.1 = 0\\
&-115 \le PL_{11} \le 115, \qquad -165 \le PL_{21} \le 165, \qquad -260 \le PL_{31} \le 260\\
&\theta_{11} = 0, \qquad I_{11}, I_{21}, I_{31} \in \{0, 1\}
\end{aligned}
$$

Applying the classical BD algorithm as described by (10)–(12), we would need 5 iterations between the master UC and the network evaluation subproblems with only 3 transmission lines to eliminate the network violations. Four feasibility Benders cuts are generated as shown in the last row of Table 6. The final optimal solution is to commit all three generators with a dispatch of 315 MW, 90 MW and 45 MW. This example shows the drawback of the classical BD algorithm, which is partly due to the low-density cuts generated in the first two iterations. That is, only one variable is incorporated in each cut, which would limit the algorithm's capability for reducing the feasible region of the master problem.

Next, we show that the proposed method would generate high-density strong Benders cuts for reducing the number of Benders iterations to 3 and the overall computing time. First the network evaluation subproblem (11) is solved, which is shown in the first row of Table 7. We obtain a feasibility Benders cut $P_{11} \le 375$ with the opti-

Table 7 Subproblems (11), (22)–(23) corresponding to $\mathbf{SZ}_1 = \{2\}$

Subproblem (11)	Min	s_1
	S.t.	$PL_{11} + PL_{31} - s_1 \le 450 \quad \lambda_{1,11}$
		$-PL_{11} - PL_{31} - s_1 \le -450 \quad \lambda_{2,11}$
		$PL_{21} - PL_{11} - s_1 \le 0 \quad \lambda_{1,21}$
		$-PL_{21} + PL_{11} - s_1 \le 0 \quad \lambda_{2,21}$
		$-PL_{21} - PL_{31} - s_1 \le -450 \quad \lambda_{1,31}$
		$PL_{21} + PL_{31} - s_1 \le 450 \quad \lambda_{2,31}$
		$PL_{11} - (\theta_{11} - \theta_{21})/0.1 = 0 \quad \kappa_{11}, \qquad PL_{21} - (\theta_{21} - \theta_{31})/0.1 = 0 \quad \kappa_{21}$
		$PL_{31} - (\theta_{11} - \theta_{31})/0.1 = 0 \quad \kappa_{31}$
		$PL_{11} \le 115 \quad \pi_{3,11}, \qquad -PL_{11} \le 115 \quad \pi_{4,11}$
		$PL_{21} \le 165 \quad \pi_{3,21}, \qquad -PL_{21} \le 165 \quad \pi_{4,21}$
		$PL_{31} \le 260 \quad \pi_{3,31}, \qquad -PL_{31} \le 260 \quad \pi_{4,31}$
		$\theta_{11} = 0 \quad \pi_1, \qquad 0 \le s_1$
Subproblem (22)	Min	$75 \cdot Q + 0 \cdot q_{1,11} + (-1) \cdot q_{2,11} + 4 \cdot q$
	S.t.	$PL_{11} + PL_{31} + 450 \cdot Q + q_{1,11} \le 0 \quad \lambda_{1,11}$
		$-PL_{11} - PL_{31} - 450 \cdot Q + q_{2,11} \le 0 \quad \lambda_{2,11}$
		$PL_{21} - PL_{11} + 0 \cdot Q - q \le 399.99 \quad \lambda_{1,21}$
		$-PL_{21} + PL_{11} - 0 \cdot Q - q \le -399.99 \quad \lambda_{2,21}$
		$-PL_{21} - PL_{31} + (-450) \cdot Q - q \le 30.01 \quad \lambda_{1,31}$
		$PL_{21} + PL_{31} - (-450) \cdot Q - q \le -30.01 \quad \lambda_{2,31}$
		$PL_{11} - (\theta_{11} - \theta_{21})/0.1 = 0, \qquad PL_{21} - (\theta_{21} - \theta_{31})/0.1 = 0$
		$PL_{31} - (\theta_{11} - \theta_{31})/0.1 = 0$
		$PL_{11} + 115 \cdot Q \le 0, \qquad -PL_{11} + 115 \cdot Q \le 0$
		$PL_{21} + 165 \cdot Q \le 0, \qquad -PL_{21} + 165 \cdot Q \le 0$
		$PL_{31} + 260 \cdot Q \le 0, \qquad -PL_{31} + 260 \cdot Q \le 0$
		$\theta_{ref,t} = 0$
		$0 \le q, \quad Q, q_{1,11}, q_{2,11}$ free
Subproblem (23)	Min	$s_1 + (-1/5) \cdot q_{2,11} + (-1.263/5) \cdot q_{2,21} + (-2.737/5) \cdot q_{2,31}$
	S.t.	$PL_{11} + PL_{31} - s_1 + q_{1,11} \le 450 \quad \lambda_{1,11}$
		$-PL_{11} - PL_{31} - s_1 + q_{2,11} \le -450 \quad \lambda_{2,11}$
		$PL_{21} - PL_{11} - s_1 + q_{1,21} \le 0 \quad \lambda_{1,21}$
		$-PL_{21} + PL_{11} - s_1 + q_{2,21} \le 0 \quad \lambda_{2,21}$
		$-PL_{21} - PL_{31} - s_1 + q_{1,31} \le -450 \quad \lambda_{1,31}$
		$PL_{21} + PL_{31} - s_1 + q_{2,31} \le 450 \quad \lambda_{2,31}$
		$PL_{11} - (\theta_{11} - \theta_{21})/0.1 = 0 \quad \kappa_{11}, \qquad PL_{21} - (\theta_{21} - \theta_{31})/0.1 = 0 \quad \kappa_{21}$
		$PL_{31} - (\theta_{11} - \theta_{31})/0.1 = 0 \quad \kappa_{31}$
		$PL_{11} \le 115 \quad \pi_{3,11}, \qquad -PL_{11} \le 115 \quad \pi_{4,11}$
		$PL_{21} \le 165 \quad \pi_{3,21}, \qquad -PL_{21} \le 165 \quad \pi_{4,21}$
		$PL_{31} \le 260 \quad \pi_{3,31}, \qquad -PL_{31} \le 260 \quad \pi_{4,31}$
		$\theta_{ref,t} = 0$
		$0 \le s_1, \quad q_{1,11}, q_{2,11}, q_{1,21}, q_{2,21}, q_{1,31}, q_{2,31}$ free

Table 8 Solution procedure using the proposed high-density cuts

Iteration	UC solution (MW)			Benders cuts
	P_{11}	P_{21}	P_{31}	
1	450	0	0	$P_{11} \leq 375$
				$0.2 \cdot P_{11} + 0.253 \cdot P_{21} - 0.547 \cdot P_{31} \leq 75$
				$0.2 \cdot P_{11} - 0.0875 \cdot P_{21} + 0.0571 \cdot P_{31} \leq 75$
2	330	90	30	$P_{21} - P_{31} \leq 45$
3	315	90	45	–

mal dual solution of $\lambda_{2,11} = \pi_{3,11} = \pi_{3,31} = -1$, all others are zero, and $\hat{s} = 75$, which is a low density cut that only contains variable P_{11}. Thus at the first iteration, the set $\mathbf{Z}$ contains two buses 2 and 3, and set $\mathbf{Z}^c$ contains bus 1, since the corresponding dual variables for generators 2 and 3 (i.e., $\lambda_{1,21}, \lambda_{2,21}, \lambda_{1,31}, \lambda_{2,31}$) are all zero. ε is assumed to be 0.01 in (22). Two optimization subproblems are constructed according to the primal form (22) with $\mathbf{SZ}_1 = \{2\}$ and $\mathbf{SZ}_2 = \{3\}$ respectively for the fixed $\lambda_{1,11} = 0, \lambda_{2,11} = -1, \hat{s}_t = 75$. The two solutions are $\lambda_{1,21} = 0, \lambda_{2,21} = -1.263, \lambda_{1,31} = -2.737, \lambda_{2,31} = 0$ and $\lambda_{1,21} = -0.429, \lambda_{2,21} = 0, \lambda_{1,31} = 0, \lambda_{2,31} = 0.286$. We apply these to the primal problem (23) which would generate the following two high-density cuts: $0.2 \cdot P_{11} + 0.253 \cdot P_{21} - 0.547 \cdot P_{31} \leq 75$ and $0.2 \cdot P_{11} - 0.0875 \cdot P_{21} + 0.0571 \cdot P_{31} \leq 75$. Here c is equal to 5. In Table 7, we give the detailed mathematical formulations for subproblems (22)–(23) corresponding to $\mathbf{SZ}_1 = \{2\}$.

These two cuts are high density because decision variables in the master problem, P_{11}, P_{21} and P_{31}, are all incorporated in these two cuts; thus they can better restrict the feasible region of the master problem. Appending these three feasibility cuts to the master problem, a solution of 330 MW, 90 MW and 30 MW is obtained. Repeating the above procedures by solving the network evaluation subproblem (11), we obtain the feasibility Benders cut $P_{21} - P_{31} \leq 45$ for the second iteration, with $\lambda_{2,21} = \lambda_{1,31} = \kappa_{31} = -0.5, \kappa_{11} = \kappa_{21} = 0.5, \pi_{3,21} = -1.5$, all others are zero, and $\hat{s} = 7.5$. Thus, at this iteration, the set $\mathbf{Z}$ contains only bus 1 and set $\mathbf{Z}^c$ contains buses 2 and 3. Here, one optimization problem is constructed according to the primal form (22), with $\mathbf{SZ}_1 = \{1\}$, for the fixed $\lambda_{1,21} = 0, \lambda_{2,21} = -0.5, \lambda_{1,31} = -0.5, \lambda_{2,31} = 0, \hat{s}_t = 7.5$. The solution is $\lambda_{1,11} = 0, \lambda_{2,11} = 0$, which means no more high-density cuts can be generated at this iteration, and there is no need to further optimize the primal form (23). If we append the feasibility cut $P_{21} - P_{31} \leq 45$ obtained in this iteration to the master problem, a solution of 315 MW, 90 MW and 45 MW is obtained.

If we repeat the above procedure, the optimal objective value of (11) will be equal to zero which means that there will be no more violations in the network evaluation subproblem. Thus the current solution is optimal and the BD procedure stops. The procedure is shown in Table 8.

Benders cuts presented in the last column of Tables 6 and 8 are mapped onto the $P_{21} - P_{31}$ plane for comparison, which are shown in Figs. 4 and 5, respectively. The mapping is processed by substituting P_{11} with $450 - P_{21} - P_{31}$, which is derived from the system power balance constraint. The domains of P_{21} and P_{31}, $\{P_{21} =$

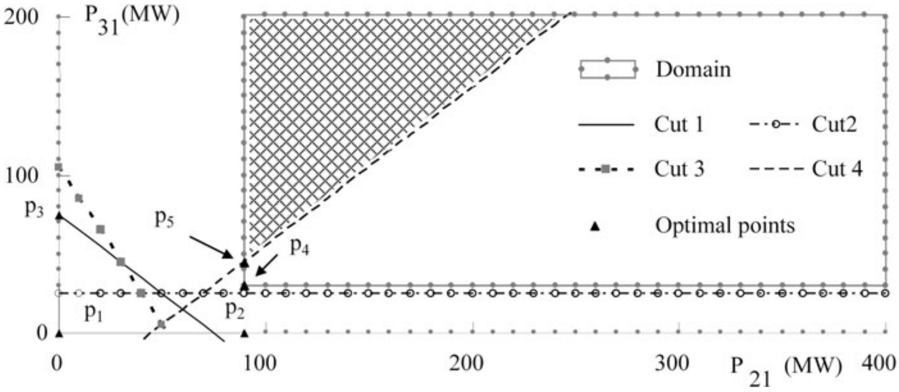


Fig. 4 Benders cuts from the classical BD algorithm mapped onto $P_{21} - P_{31}$ plane

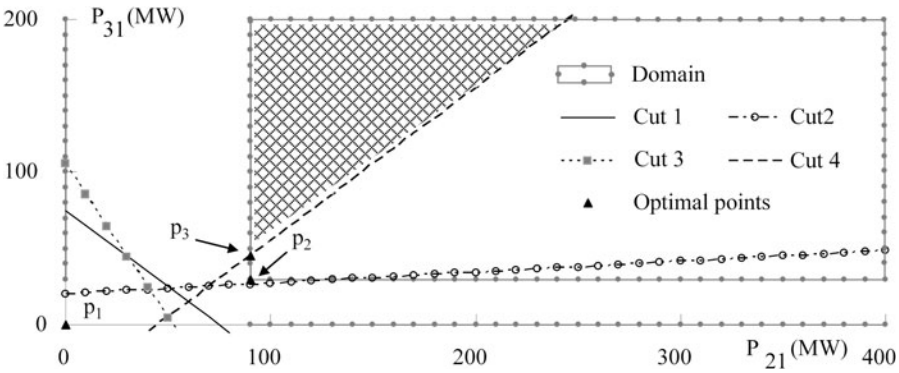


Fig. 5 Benders cuts from the proposed method mapped onto $P_{21} - P_{31}$ plane

$0, P_{31} = 0\} \cup \{90 \le P_{21} \le 400, P_{31} = 0\} \cup \{P_{21} = 0, 30 \le P_{31} \le 200\} \cup \{90 \le P_{21} \le 400, 30 \le P_{31} \le 200\}$, are also shown in Figs. 4 and 5. The Benders cuts 1–4 correspond to the four cuts in Tables 6 and 8 respectively. The optimal points $p_1 - p_5$ obtained at each iteration in Table 6 and optimal points $p_1 - p_3$ at each iteration in Table 8 are also shown. Here, Cuts 1 and 4 are the same in both methods. The Cut 3 coincides in the two methods, i.e., by substituting P_{11} with $450 - P_{21} - P_{31}$, the Cut 3 in Table 6 becomes $2P_{21} + P_{31} \ge 105$ and the Cut 3 in Table 8 becomes $0.2875P_{21} + 0.1429P_{31} \ge 15$, which is $2.0119P_{21} + P_{31} \ge 104.9685$. The Cut 2 from Table 8 provides a tighter formulation than Cut 2 from Table 6, i.e., it passes through the domain boundary $\{90 \le P_{21} \le 400, P_{31} = 0\}$ and restricts a tighter feasibility region. The meshed areas in Figs. 4 and 5 show the feasibility regions at the final iteration bounded by Benders cuts and the domain. In Fig. 5, the high density Cuts 2 and 3 generated at the first iteration would tighten the feasibility region of the master problem. These two high density cuts would eliminate the two feasible solutions obtained in the classical BD algorithm and shown as p_2 and p_3 in Fig. 4; thus they would reduce the number of iterations by two as compared with that of the classical BD algorithm. By comparison, the proposed method outperforms the classical BD

Table 9 Solution procedure using the proposed method with tuning parameter

Iteration	UC Solution (MW)			Benders cuts
	P_{11}	P_{21}	P_{31}	
1	450	0	0	$P_{11} \le 375$
				$0.2 \cdot P_{11} + 0.0947 \cdot P_{21} - 0.1053 \cdot P_{31} \le 75$
				$0.2 \cdot P_{11} - 0.0875 \cdot P_{21} + 0.0571 \cdot P_{31} \le 75$
2	330	90	30	$P_{21} - P_{31} \le 45$
3	315	90	45	–

algorithm in the sense that it would offer the same reduced feasible region in fewer iterations and smaller overall computing time.

In the above study, $2 \cdot |\mathbf{Z}|$ is used as the right-hand-side for the last inequality constraint in (22) to guarantee that the problem (22) is bounded. In comparison, we use 1 instead of $2 \cdot |\mathbf{Z}|$ to show the impact of this parameter on convergence. Table 9 shows that the problem is solved in three iterations and the same global optimal solution is derived. As compared with Table 8, when using 1 instead of $2 \cdot |\mathbf{Z}|$, the strong cuts are not exactly the same but the UC solution at iteration 2 is the same as that in Table 8. The total number of iterations and generated cuts are the same in the two cases with the same final solution, which shows that the parameter $2 \cdot |\mathbf{Z}|$ does not significantly impact the convergence in this case. This small problem is also directly solved by CPLEX with the computation time of 0.04 second and the global optimal solution of $P_{11} = 315$, $P_{21} = 90$, $P_{31} = 45$, $I_{11} = I_{21} = I_{31} = 1$, $PL_{11} = 75$, $PL_{21} = 165$, $PL_{31} = 240$, $\theta_{11} = 0$, $\theta_{21} = -7.5$, and $\theta_{31} = -24$, which is the same as that obtained by the proposed accelerating BD algorithm. For Cases 4.1 and 4.2, the purpose of reporting the solutions with CPLEX is to show the effectiveness of the proposed accelerating BD algorithm. In the following Case 4.3, we will see that the NCUC problem for large-scale power systems may take much longer to be solved directly by CPLEX, and the proposed accelerating BD algorithm can efficiently solve the problem by delivering good enough solutions within acceptable time frames.

4.3 5663-bus system

A large-scale system with 599 generators, 5663 buses, 7036 transmission lines, and 251 phase shifters is used to illustrate the effectiveness of the proposed methodology. A 24-hour NCUC problem is considered to find the optimal UC decisions and generation dispatch with the minimum operation cost, while satisfying the system loads, reserve requirements, and losses for a 24-hour horizon shown in Fig. 6. All UC and dc transmission network constraints described by (2)–(8) are considered in this case. Table 10 shows the complexity of the large-scale power system optimization problem. It is noted that the number of dc network constraints in (8) (excluding upper and lower limits for phase shifters and transmission flow variables) is even larger than the number of UC constraints in (2)–(7), which indicates that the decomposition is the only option for handling large-scale power system optimization problems with a substantial network structure.

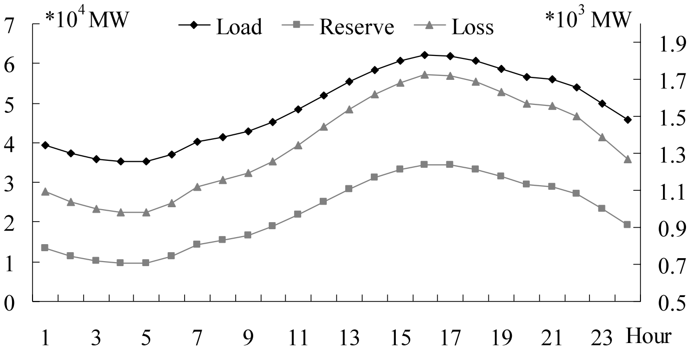


Fig. 6 System loads, reserves, and losses for a 24-hour horizon

Table 10 Complexity of the 5663-bus system

	# of binary variables	# of continuous variables	# of UC constraints in (2)–(7)	# of dc network constraints in (8)
Peak hour problem	1,797	19,318	10,323	12,700
24-hour problem	43,128	457,608	274,268	304,800

- Case 1: Comparison of the proposed method with the classical BD algorithm at the peak hour 16 with a peak load of 62,000 MW.
- Case 2: Comparison of the proposed method with the classical BD algorithm for the 24-hour study.
- Case 3: Appending the initial cuts and studying its impact on convergence.

Case 1: The peak hour is studied using the classical BD algorithm (12) and the proposed strong Benders cuts (16) and (21). Figure 7 shows the NCUC convergence. The proposed method would give a steeper ascent to the violation and speed up the convergence. Table 11 compares the number of iterations and cuts, total CPU time, and the operation cost between the classical BD algorithm and the proposed strong Benders cuts method. We also provide the solution by using the LSF for comparison. The CPU time is based on multiple CPUs, which enables the parallel calculation of hourly network checking subproblems, and the parallel calculation of (18) for all subset $\mathbf{B}_o$ and (22)–(23) for all subset $\mathbf{SZ}_o$. The classical BD algorithm would need 18 iterations with a total of 17 cuts and a total CPU time of 118 s, most of which is consumed by iterations between UC and hourly network evaluation subproblems. In comparison, with the proposed strong Benders cuts, only 8 iterations are needed and the total CPU time is 25 s, which is almost one-fifth of the time used by the classical BD algorithm. A total of 23 cuts are generated in which 13 are from multiple optimal dual solutions and another 3 for enhancing the cut density. For this case, the difference between the proposed method and the classical BD algorithm is even more evident. The proposed method converges more rapidly in terms of the total number of iterations by restricting the feasible region of the master problem at each iteration via strong Benders cuts.

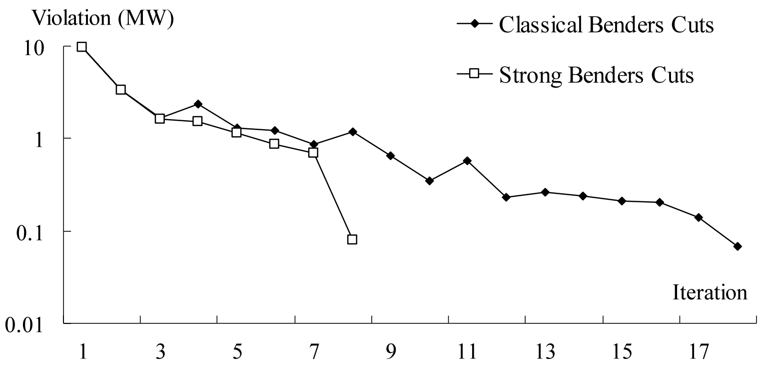


Fig. 7 Classical and multiple strong Benders cuts for the peak hour

Table 11 Comparison of the results in Case 1

	# of iterations	# of cuts	CPU time (s)	Operation cost ($)
Classical BD algorithm	18	17	118	1,745,451.8
LSF	4	129	11	1,747,353.5
Multiple strong Benders cuts	8	23	25	1,777,264.2
CPLEX Directly	–	–	67	1,747,360.4

The violation threshold of 0.1 MW is used here. The LSF uses fewer iterations and less CPU time than the strong Benders cuts method for this one-hour NCUC problem. However, LSF introduces many more network evaluation constraints than the strong Benders cuts method, which will cause a longer study period for the master UC problem since the master UC is a NP-hard problem. As shown in Cases 2–3, the CPU time for LSF and strong Benders cuts are comparable. The last row of Table 11 shows the operation cost for the three methods. The three solutions are suboptimal since the problem is NP-hard. In this case, the operation cost based on the proposed multiple strong Benders cuts is about 1.82% (i.e. $(1,777,264.2 - 1,745,451.8)/1,745,451.8$) higher than that of the classical BD algorithm. However, the proposed multiple strong Benders cut method reduces of the number of iterations by 55.56% and the CPU time by 78.81%, and provides a similar operation cost as compared to the classical BD algorithm. Hence the proposed method outperforms the classical BD algorithm by providing a good enough solution with a much smaller computation time. This peak hour NCUC problem can be directly solved by CPLEX in 67 seconds. The objective is $1,747,360.4 with the MIP gap tolerance of 0.05%. In comparison to the solutions given in Table 11, the proposed multiple strong Benders cuts obtained an operation cost that is 1.71% (i.e. $(1,777,264.2 - 1,747,360.4)/1,747,360.4$) higher than that of CPLEX, while saving a computation time of more than 62.69% (i.e. $(67 - 25)/67$).

Case 2: In this case, the system is studied for a 24-hour period. Figure 8 compares the NCUC convergence of the classical BD with that of the proposed strong Benders cuts method. The classical BD algorithm takes 27 iterations vs. 16 for the proposed

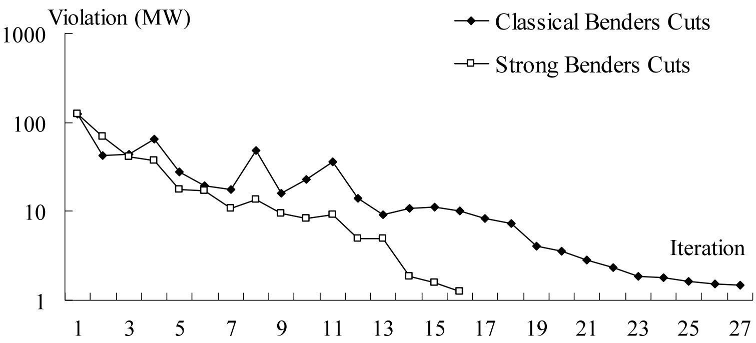


Fig. 8 Classical and strong Benders cuts for the 24-hour case study

Fig. 9 CPU time at each iteration in Case 2

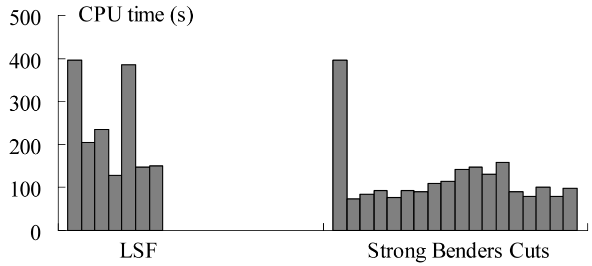


Table 12 Comparison of the results for the 24-hour study

	# of iterations	# of cuts	CPU time (s)	Operation cost ($)
Classical BD algorithm	27	409	3008	28,652,220.7
LSF	7	1693	1713	28,607,073.8
Multiple strong Benders cuts	16	881	2159	28,660,027.6

strong Benders cuts method. A total of 881 cuts are generated for limiting the feasible region of the master problem during the iterative process in which 491 strong cuts are derived from the multiple optimal dual solutions and another 110 for enhancing the cut density. Table 12 compares the number of iterations and cuts, total CPU time, and the operation cost of the classical BD algorithm, the strong Benders cuts, and the LSF method. The proposed strong Benders cuts method reduces the number of iterations by 40.74% (i.e. $(27-16)/27$) and saves the CPU time by 28.22% (i.e. $(3008-2159)/3008$) as compared to the classical BD algorithm. In this case, the operation cost obtained by LSF is the lowest, and the cost by multiple strong Benders cuts is about 0.027% (i.e. $(28,660,027.6-28,652,220.7)/28,652,220.7$) higher than that of the classical BD algorithm. Figure 9 shows that it takes more time to solve the master UC at each iteration with the LSF method since the LSF introduces many more constraints than the strong Benders cuts method. The LSF introduces 1,693 network evaluation constraints vs. 812 in the strong Benders cuts method. Although LSF takes 7 iterations, which is less than half of the 16 for the strong Benders cuts

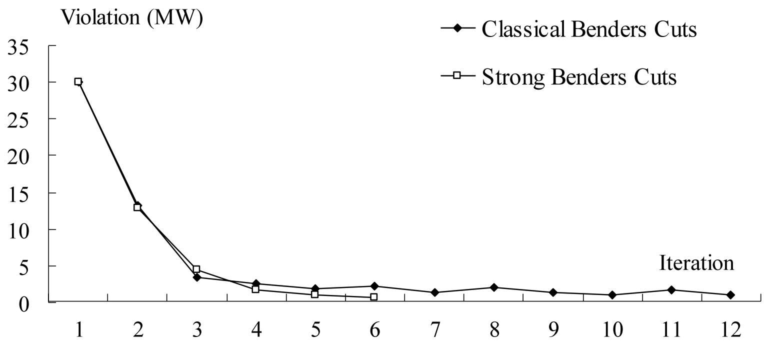


Fig. 10 Classical and strong Benders cuts for the 24-hour study with initial cuts

method, the required CPU time of LSF is only 20.66% (i.e. (2159 – 1713)/2159) smaller than that of the strong Benders cuts method. When solving the master UC problem successively, the solution at the previous iteration can be used as a starting point for the existing UC. Figure 9 shows that a smaller CPU time is used progressively at each iteration of the BD algorithm since only a limited number of Benders cuts are introduced into the master problem at each iteration and the solution at the previous iteration always provides a good starting point. In contrast, the LSF method introduces more constraints at each iteration, all generation dispatch decision variables are considered which would make the problem more difficult to be solved, and the solution of previous iteration may not be used as the starting point. Hence, the CPU time for each LSF iteration is not dramatically reduced as compared with that of the BD algorithm. This 24-hour NCUC problem can not be solved without decomposition by the current version of CPLEX. It takes 2,389 seconds to just solve the root node LP problem, which is generated by relaxing integrality requests of the NCUC problem. Furthermore, after running for more than 3 hours, the CPLEX could not locate a feasible solution to the NCUC problem at any MIP gap level.

Case 3: A good set of effective initial Benders cuts could reduce the number of iterations between the master problem and subproblems [10]. In this case, the hourly NCUCs are first solved in parallel while ignoring startup and shutdown costs and coupling constraints among UC hours (e.g. min on/off time and ramping up/down rate limits) to generate an initial set of Benders cuts [18]. The Benders cuts resulting from the hourly NCUC solution are considered as initial Benders cuts for the 24-hour NCUC solution. Figure 10 shows the NCUC convergence for the classical BD and the strong Benders cuts methods with 603 initial Benders cuts generated by solving the hourly NCUC. The classical BD algorithm takes 12 iterations with a total of 90 cuts. The strong Benders cuts method takes 6 iterations with a total of 294 cuts in which 185 strong cuts are derived from the multiple optimal dual solutions and another 26 are high-density cuts. For a fair comparison, the same set of initial Benders cuts are adopted in the LSF method. The difference between the three operation costs is less than 1% (i.e., $28,566,540.3 - 28,548,341.2)/28,548,341.2 = 0.064\%$). Compared with the solutions given in Tables 12, 13 shows that the inclusion of initial Benders cuts would dramatically reduce the number of iterations and cuts for both the classical BD algorithm and the strong Benders cuts methods.

Table 13 Comparison of the results for the 24-hour study with initial cuts

	# of iterations	# of cuts	CPU time (s)	Operation cost ($)
Classical BD algorithm	12	90	2797	28,548,341.2
LSF	5	711	1665	28,566,540.3
Strong Benders cuts	6	294	1584	28,554,679.8

The number of iterations is reduced by 55.56% and 62.5% respectively (i.e., $(27-12)/27$ and $(16-6)/16$), and the number of cuts is decreased by about 78.00% (i.e., $(409-90)/409$) for the classical BD algorithm and 66.63% (i.e., $(881-294)/881$) for the strong Benders cuts. However, the total CPU time for the classical BD algorithm is decreased only by about 7% (i.e., $(3008-2797)/3008$) and 21.63% (i.e., $(2159-1584)/2159$) for the strong Benders cuts. Here, the total CPU time reduction is not significant as may be expected because by appending 603 initial cuts, the required CPU time for each UC master problem would be longer which would diminish the saving in time when initial cuts are used. With initial cuts, the LSF takes 5 iterations as compared with 6 for the strong Benders cuts method, and the CPU time for these two methods are comparable in this case. The strong Benders cuts method would take 4.86% (i.e., $(1665-1584)/1665$) less CPU time than the LSF method.

5 Conclusions

This paper proposed a method for generating multiple strong Benders cuts to accelerate the convergence of NCUC with BD. The generated strong Benders cuts are proved to be pareto optimal. Numerical tests show that as compared to the classical BD algorithm, multiple strong Benders cuts generated by the proposed method result in a significant reduction in terms of total number of iterations and the CPU time. For large systems, the proposed multiple strong Benders cuts method and the LSF method are comparable in terms of good enough solutions and CPU time. The proposed multiple strong Benders cuts method also requires fewer cuts than that of the LSF methods. The proposed strong Benders cuts method can be easily extended to other applications of BD on large-scale optimization problems in power systems operation, maintenance and planning.

Appendix A

Proof for Theorem 1 Suppose to the contrary that (16) is not pareto optimal. That is, there exists a cut (A.1) that dominates (16)

$$\sum_{b=1}^{NB} \sum_{i\in \mathbf{U}(b)} (\lambda'_{1,bt} - \lambda'_{2,bt}) \cdot P_{it} + \hat{s}_t \le 0 \tag{A.1}$$

Then from Definition 2, it is true that for all $\mathbf{P}_t$ in their domain of $\{\mathbf{0}, [\mathbf{P}^{\min}, \mathbf{P}^{\max}]\}$

$$\left\{\left[\sum_{b=1}^{NB}\sum_{i\in\mathbf{U}(b)}(\tilde{\lambda}_{1,bt}-\tilde{\lambda}_{2,bt})\cdot P_{it}+\hat{s}_t\right]-\left[\sum_{b=1}^{NB}\sum_{i\in\mathbf{U}(b)}(\lambda'_{1,bt}-\lambda'_{2,bt})\cdot P_{it}+\hat{s}_t\right]\right\}\leq 0 \tag{A.2}$$

Thus (A.2) is true for $P_{it}=P_i^{\max}-\varepsilon, \forall i\in\mathbf{U}(b), b\in\mathbf{B}_o$, and $P_{it}=P_i^{\min}+\varepsilon, \forall i\in\mathbf{U}(b), b\notin\mathbf{B}_o$, which is

$$\left\{\begin{array}{l}\left[\sum_{b\in\mathbf{B}_o}\sum_{i\in\mathbf{U}(b)}(\tilde{\lambda}_{1,bt}-\tilde{\lambda}_{2,bt})\cdot(P_i^{\max}-\varepsilon)\right.\\ \quad\left.+\sum_{b\notin\mathbf{B}_o}\sum_{i\in\mathbf{U}(b)}(\tilde{\lambda}_{1,bt}-\tilde{\lambda}_{2,bt})\cdot(P_i^{\min}+\varepsilon)\right]\\ \quad-\left[\sum_{b\in\mathbf{B}_o}\sum_{i\in\mathbf{U}(b)}(\lambda'_{1,bt}-\lambda'_{2,bt})\cdot(P_i^{\max}-\varepsilon)\right.\\ \quad\left.+\sum_{b\notin\mathbf{B}_o}\sum_{i\in\mathbf{U}(b)}(\lambda'_{1,bt}-\lambda'_{2,bt})\cdot(P_i^{\min}+\varepsilon)\right]\end{array}\right\}\leq 0 \tag{A.3}$$

However, solving the maximization problem (15), we obtain the optimal objective value as $\sum_{b\in\mathbf{B}_o}\sum_{i\in\mathbf{U}(b)}(\tilde{\lambda}_{1,bt}-\tilde{\lambda}_{2,bt})\cdot(P_i^{\max}-\varepsilon)+\sum_{b\notin\mathbf{B}_o}\sum_{i\in\mathbf{U}(b)}(\tilde{\lambda}_{1,bt}-\tilde{\lambda}_{2,bt})\cdot(P_i^{\min}+\varepsilon)$, which means

$$\left\{\begin{array}{l}\left[\sum_{b\in\mathbf{B}_o}\sum_{i\in\mathbf{U}(b)}(\tilde{\lambda}_{1,bt}-\tilde{\lambda}_{2,bt})\cdot(P_i^{\max}-\varepsilon)\right.\\ \quad\left.+\sum_{b\notin\mathbf{B}_o}\sum_{i\in\mathbf{U}(b)}(\tilde{\lambda}_{1,bt}-\tilde{\lambda}_{2,bt})\cdot(P_i^{\min}+\varepsilon)\right]\\ \quad-\left[\sum_{b\in\mathbf{B}_o}\sum_{i\in\mathbf{U}(b)}(\lambda'_{1,bt}-\lambda'_{2,bt})\cdot(P_i^{\max}-\varepsilon)\right.\\ \quad\left.+\sum_{b\notin\mathbf{B}_o}\sum_{i\in\mathbf{U}(b)}(\lambda'_{1,bt}-\lambda'_{2,bt})\cdot(P_i^{\min}+\varepsilon)\right]\end{array}\right\}\geq 0 \tag{A.4}$$

Using (A.3) and (A.4), we have

$$\left\{\begin{array}{l}\left[\sum_{b\in\mathbf{B}_o}\sum_{i\in\mathbf{U}(b)}(\tilde{\lambda}_{1,bt}-\tilde{\lambda}_{2,bt})\cdot(P_i^{\max}-\varepsilon)\right.\\ \quad\left.+\sum_{b\notin\mathbf{B}_o}\sum_{i\in\mathbf{U}(b)}(\tilde{\lambda}_{1,bt}-\tilde{\lambda}_{2,bt})\cdot(P_i^{\min}+\varepsilon)\right]\\ \quad-\left[\sum_{b\in\mathbf{B}_o}\sum_{i\in\mathbf{U}(b)}(\lambda'_{1,bt}-\lambda'_{2,bt})\cdot(P_i^{\max}-\varepsilon)\right.\\ \quad\left.+\sum_{b\notin\mathbf{B}_o}\sum_{i\in\mathbf{U}(b)}(\lambda'_{1,bt}-\lambda'_{2,bt})\cdot(P_i^{\min}+\varepsilon)\right]\end{array}\right\}=0 \tag{A.5}$$

Since (A.1) dominates (16), according to Definition 1 there must exits a $\tilde{\mathbf{P}}$ in the domain of $\{\mathbf{0}, [\mathbf{P}^{\min}, \mathbf{P}^{\max}]\}$ by which the strict inequality of (A.2) is satisfied as shown in (A.6)

$$\left\{\left[\sum_{b=1}^{NB}\sum_{i\in\mathbf{U}(b)}(\tilde{\lambda}_{1,bt}-\tilde{\lambda}_{2,bt})\cdot\tilde{P}_i\right]-\left[\sum_{b=1}^{NB}\sum_{i\in\mathbf{U}(b)}(\lambda'_{1,bt}-\lambda'_{2,bt})\cdot\tilde{P}_i\right]\right\}<0 \quad \text{(A.6)}$$

Next we show that there exits a scalar $\theta > 1$ such that $\theta\cdot(P_i^{\max}-\varepsilon)+(1-\theta)\cdot\tilde{P}_i$ and $\theta\cdot(P_i^{\min}+\varepsilon)+(1-\theta)\cdot\tilde{P}_i$ are both in the range of $\{\mathbf{0}, [\mathbf{P}^{\min}, \mathbf{P}^{\max}]\}$.

1. If $\tilde{P}_i=0, \theta\cdot(P_i^{\max}-\varepsilon)+(1-\theta)\cdot\tilde{P}_i=\theta\cdot(P_i^{\max}-\varepsilon)$ and $\theta\cdot(P_i^{\min}+\varepsilon)+(1-\theta)\cdot\tilde{P}_i=\theta\cdot(P_i^{\min}+\varepsilon)$. Thus for small $0\le\varepsilon$ and $1<\theta\le\min\{\frac{P_i^{\max}}{P_i^{\max}-\varepsilon}, \frac{P_i^{\max}}{P_i^{\min}+\varepsilon}\}$, $\theta\cdot(P_i^{\max}-\varepsilon)+(1-\theta)\cdot\tilde{P}_i$ and $\theta\cdot(P_i^{\min}+\varepsilon)+(1-\theta)\cdot\tilde{P}_i$ are both in the range of $[\mathbf{P}^{\min}, \mathbf{P}^{\max}]$.
2. If $\tilde{P}_i\in[P_i^{\min}, P_i^{\max}]$, for θ that satisfies $1<\theta\le\frac{P_i^{\max}-P_i^{\min}}{P_i^{\max}-P_i^{\min}-\varepsilon}$, we have $\frac{\theta}{\theta-1}\cdot\varepsilon\ge P_i^{\max}-P_i^{\min}$. Thus $\frac{\theta}{\theta-1}\cdot\varepsilon\ge P_i^{\max}-P_i^{\min}\ge P_i^{\max}-\tilde{P}_i$, and we have $\theta\cdot(P_i^{\max}-\varepsilon)+(1-\theta)\cdot\tilde{P}_i-P_i^{\max}=(\theta-1)\cdot(P_i^{\max}-\tilde{P}-\frac{\theta}{\theta-1}\cdot\varepsilon)\le 0$, which derives $\theta\cdot(P_i^{\max}-\varepsilon)+(1-\theta)\cdot\tilde{P}_i\le P_i^{\max}$. For small $0\le\varepsilon$, it usually holds that $P_i^{\max}-P_i^{\min}-\varepsilon\ge\varepsilon$ since $P_i^{\max}$ and $P_i^{\min}$ are usually in tens and hundreds of MW range, and an ε of 0.01–0.1 MW is used in this paper. Thus $1<\theta\le\frac{P_i^{\max}-P_i^{\min}}{P_i^{\max}-P_i^{\min}-\varepsilon}\le\frac{P_i^{\max}-P_i^{\min}}{\varepsilon}$, and we have $\theta\cdot(P_i^{\max}-\varepsilon)+(1-\theta)\cdot\tilde{P}_i-P_i^{\min}\ge\theta\cdot(P_i^{\max}-\varepsilon)+(1-\theta)\cdot P_i^{\max}-P_i^{\min}=P_i^{\max}-P_i^{\min}-\theta\cdot\varepsilon\ge 0$, which derives $P_i^{\min}\le\theta\cdot(P_i^{\max}-\varepsilon)+(1-\theta)\cdot\tilde{P}_i$. Similarly, we derive that $\theta\cdot(P_i^{\min}+\varepsilon)+(1-\theta)\cdot\tilde{P}_i$ is in the range of $[\mathbf{P}^{\min}, \mathbf{P}^{\max}]$.

Based on the above discussion, we multiply (A.5) by θ and (A.6) by $(1-\theta)$ which is negative and preserves the inequality, and add them up to derive

$$\begin{aligned}
&\sum_{b\in\mathbf{B}_o}\sum_{i\in\mathbf{U}(b)}(\tilde{\lambda}_{1,bt}-\tilde{\lambda}_{2,bt})\cdot[\theta\cdot(P_i^{\max}-\varepsilon)+(1-\theta)\cdot\tilde{P}_i]\\
&\quad+\sum_{b\notin\mathbf{B}_o}\sum_{i\in\mathbf{U}(b)}(\tilde{\lambda}_{1,bt}-\tilde{\lambda}_{2,bt})\cdot[\theta\cdot(P_i^{\min}+\varepsilon)+(1-\theta)\cdot\tilde{P}_i]\\
&\quad>\sum_{b\in\mathbf{B}_o}\sum_{i\in\mathbf{U}(b)}(\lambda'_{1,bt}-\lambda'_{2,bt})\cdot[\theta\cdot(P_i^{\max}-\varepsilon)+(1-\theta)\cdot\tilde{P}_i]\\
&\quad+\sum_{b\notin\mathbf{B}_o}\sum_{i\in\mathbf{U}(b)}(\lambda'_{1,bt}-\lambda'_{2,bt})\cdot[\theta\cdot(P_i^{\min}+\varepsilon)+(1-\theta)\cdot\tilde{P}_i] \qquad \text{(A.7)}
\end{aligned}$$

According to (A.7)

$$P_{it}=\begin{cases}\theta\cdot(P_i^{\max}-\varepsilon)+(1-\theta)\cdot\tilde{P}_i, & \forall i\in\mathbf{U}(b), b\in\mathbf{B}_o\\ \theta\cdot(P_i^{\min}+\varepsilon)+(1-\theta)\cdot\tilde{P}_i, & \forall i\in\mathbf{U}(b), b\notin\mathbf{B}_o\end{cases}$$

$$\left\{\begin{array}{l}\left[\sum_{b\in\mathbf{B}_o}\sum_{i\in\mathbf{U}(b)}(\hat{\lambda}_{1,bt}-\hat{\lambda}_{2,bt})\cdot(P_{it}-\hat{P}_{it})\right.\\ \quad\left.+\sum_{b\notin\mathbf{B}_o}\sum_{i\in\mathbf{U}(b)}(\tilde{\lambda}_{1,bt}-\tilde{\lambda}_{2,bt})\cdot P_{it}+\hat{s}_t\right]\\ \quad-\left[\sum_{b\in\mathbf{B}_o}\sum_{i\in\mathbf{U}(b)}(\hat{\lambda}_{1,bt}-\hat{\lambda}_{2,bt})\cdot(P_{it}-\hat{P}_{it})\right.\\ \quad\left.+\sum_{b\notin\mathbf{B}_o}\sum_{i\in\mathbf{U}(b)}(\lambda'_{1,bt}-\lambda'_{2,bt})\cdot P_{it}+\hat{s}_t\right]\end{array}\right\}>0$$

which is inconsistent with (A.2). Hence our assumption of (16) not being pareto optimal is untenable. Thus (16) is pareto optimal. □

Appendix B

The following Theorem B.1 is a prerequisite for the proof of Theorem 2.

Theorem B.1 *Let* $\{\hat{\lambda}_{1,bt}/c, \hat{\lambda}_{2,bt}/c, b\in\mathbf{Z}^c, \bar{\lambda}_{1,bt}/c, \bar{\lambda}_{2,bt}/c, b\in\mathbf{Z}\}$ *together with* $\{\bar{\pi}_{1,pt}, \bar{\pi}_{2,pt}, \bar{\pi}_{3,lt}, \bar{\pi}_{4,lt}, \bar{\pi}_t, \bar{\kappa}_{lt}\}$ *be an optimal solution of* (20) *with the objective value of* $\breve{s}_t$. *Thus* $\hat{s}\le c\cdot\breve{s}_t$ *is satisfied, where* $\hat{s}_t$ *is the optimal objective value of* (11) *used in* (19).

Proof for Theorem B.1 First let us show that the following problem (B.1) has the optimal objective value of $c\cdot\breve{s}_t$. Assume the optimal objective value of (B.1) is O

$$\begin{aligned}
\text{Max}\quad & \sum_b\left(\sum_{i\in\mathbf{U}(b)}P_{it}-D_{bt}-C_b\cdot P_{Losst}\right)\cdot(\lambda_{1,bt}-\lambda_{2,bt})\\
&\quad+\sum_p(\gamma_p^{\max}\cdot\pi_{1,pt}-\gamma_p^{\min}\cdot\pi_{2,pt})+\sum_l PL_l^{\max}\cdot(\pi_{3,lt}+\pi_{4,lt})\\
\text{S.t.}\quad & (\lambda_{1,at}-\lambda_{2,at})+(-\lambda_{1,bt}+\lambda_{2,bt})+\kappa_{lt}+\pi_{3,lt}-\pi_{4,lt}=0\\
&\quad\text{line } l \text{ is from bus } a \text{ to bus } b\\
&\sum_a\kappa_{lt}/x_{am}-\sum_b\kappa_{l't}/x_{mb}=0\\
&\quad\text{line } l \text{ is from bus } a \text{ to bus } m\text{, line } l' \text{ is from bus } m \text{ to bus } b,\\
&\quad m \text{ is not reference bus}\\
&\sum_a\kappa_{lt}/x_{am}-\sum_b\kappa_{l't}/x_{mb}+\pi_t=0 \qquad\qquad (\text{B.1})\\
&\quad\text{line } l \text{ is from bus } a \text{ to bus } m\text{, line } l' \text{ is from bus } m \text{ to bus } b,\\
&\quad m \text{ is reference bus}\\
&\kappa_{lt}/x_{ab}+\pi_{1,pt}-\pi_{2,pt}=0\\
&\quad\text{phase shifter } p \text{ is located at line } l\text{, which is from bus } a \text{ to bus } b\\
&-\sum_{b\in Z}(\lambda_{1,bt}+\lambda_{2,bt})\le 2\cdot|\mathbf{Z}|
\end{aligned}$$

$$\lambda_{1,bt} = \hat{\lambda}_{1,bt}, \quad \lambda_{2,bt} = \hat{\lambda}_{2,bt}, \quad b \in \mathbf{Z}^c$$
$$\lambda_{1,bt} = \bar{\lambda}_{1,bt}, \quad \lambda_{2,bt} = \bar{\lambda}_{2,bt}, \quad b \in \mathbf{Z}$$
$$\lambda_{1,bt}, \lambda_{2,bt}, \pi_{1,pt}, \pi_{2,pt}, \pi_{3,lt}, \pi_{4,lt} \le 0, \quad \kappa_{lt}, \pi_t \text{ free}$$

It is obvious that $\{\hat{\lambda}_{1,bt}, \hat{\lambda}_{2,bt}, b \in \mathbf{Z}^c, \bar{\lambda}_{1,bt}, \bar{\lambda}_{2,bt}, b \in \mathbf{Z}\}$ together with $\{c \cdot \bar{\pi}_{1,pt}, c \cdot \bar{\pi}_{2,pt}, c \cdot \bar{\pi}_{3,lt}, c \cdot \bar{\pi}_{4,lt}, c \cdot \bar{\pi}_t, c \cdot \bar{\kappa}_{lt}\}$ is a feasible solution of (B.1), with the objective value of $c \cdot \breve{s}_t$. Thus the optimal objective value of (B.1) is not less than $c \cdot \breve{s}_t$, i.e.,

$$O \ge c \cdot \breve{s}_t \tag{B.2}$$

Assume $\{\hat{\lambda}_{1,bt}, \hat{\lambda}_{2,bt}, b \in \mathbf{Z}^c, \bar{\lambda}_{1,bt}, \bar{\lambda}_{2,bt}, b \in \mathbf{Z}\}$ together with $\{\pi'_{1,pt}, \pi'_{2,pt}, \pi'_{3,lt}, \pi'_{4,lt}, \pi'_t, \kappa'_{lt}\}$ is the optimal solution of (B.1) with the optimal objective value of O. Thus it is obvious that $\{\hat{\lambda}_{1,bt}/c, \hat{\lambda}_{2,bt}/c, b \in \mathbf{Z}^c, \bar{\lambda}_{1,bt}/c, \bar{\lambda}_{2,bt}/c, b \in \mathbf{Z}\}$ together with $\{\pi'_{1,pt}/c, \pi'_{2,pt}/c, \pi'_{3,lt}/c, \pi'_{4,lt}/c, \pi'_t/c, \kappa'_{lt}/c\}$ is a feasible solution of (20) with the objective value of O/c. Since the optimal objective value to (20) is $\breve{s}_t$, we have

$$O/c \le \breve{s}_t \tag{B.3}$$

Using (B.2) and (B.3), we conclude that $O = c \cdot \breve{s}_t$. That is the optimal objective value of (B.1) O is equal to $c \cdot \breve{s}_t$. Comparing (B.1) with (19), the optimal solution of (19), which is $\hat{\lambda}_{1,bt}\hat{\lambda}_{2,bt}$ for $b \in \mathbf{Z}^c$ and $\bar{\lambda}_{1,bt}\bar{\lambda}_{2,bt}$ for $b \in \mathbf{Z}$ together with the corresponding $\{\pi'_{1,pt}, \pi'_{2,pt}, \pi'_{3,lt}, \pi'_{4,lt}, \pi'_t, \kappa'_{lt}\}$ is a feasible solution to (B.1) as this solution set satisfies all constraints in (B.1) with the objective value of $\hat{s}_t$. Thus, $\hat{s}_t$ is no larger than the optimal objective value $c \cdot \breve{s}_t$, that is $\hat{s} \le c \cdot \breve{s}_t$. □

Proof for Theorem 2 Suppose that contrary to our assumption, (21) is not pareto optimal. That is, there exists a cut (B.4) that dominates (21)

$$\sum_{b \in \mathbf{Z}^c} \sum_{i \in \mathbf{U}(b)} (\hat{\lambda}_{1,bt}/c - \hat{\lambda}_{2,bt}/c) \cdot (P_{it} - \hat{P}_{it}) + \sum_{b \in \mathbf{Z}} \sum_{i \in \mathbf{U}(b)} (\lambda'_{1,bt} - \lambda'_{2,bt}) \cdot P_{it} + \breve{s}_t \le 0 \tag{B.4}$$

Then from Definition 2, it is true that for all $\mathbf{P}_t$ in the domain of $\{\mathbf{0}, [\mathbf{P}^{\min}, \mathbf{P}^{\max}]\}$, we have

$$\left\{ \begin{array}{l} \left[\sum_{b \in \mathbf{Z}^c} \sum_{i \in \mathbf{U}(b)} (\hat{\lambda}_{1,bt}/c - \hat{\lambda}_{2,bt}/c) \cdot (P_{it} - \hat{P}_{it}) \right. \\ \left. \quad + \sum_{b \in \mathbf{Z}} \sum_{i \in \mathbf{U}(b)} (\bar{\lambda}_{1,bt}/c - \bar{\lambda}_{2,bt}/c) \cdot P_{it} + \breve{s}_t \right] \\ \quad - \left[\sum_{b \in \mathbf{Z}^c} \sum_{i \in \mathbf{U}(b)} (\hat{\lambda}_{1,bt}/c - \hat{\lambda}_{2,bt}/c) \cdot (P_{it} - \hat{P}_{it}) \right. \\ \left. \quad + \sum_{b \in \mathbf{Z}} \sum_{i \in \mathbf{U}(b)} (\lambda'_{1,bt} - \lambda'_{2,bt}) \cdot P_{it} + \breve{s}_t \right] \end{array} \right\} \le 0 \tag{B.5}$$

Thus (B.5) is true for $P_{it} = \hat{P}_{it}, \forall i \in \mathbf{U}(b), b \in \mathbf{Z}^c$, $P_{it} = P_i^{\max} - \varepsilon, \forall i \in \mathbf{U}(b), b \in \mathbf{SZ}_o$, and $P_{it} = (P_i^{\min} + \varepsilon), \forall i \in \mathbf{U}(b), b \in \{\mathbf{Z} - \mathbf{SZ}_o\}$, which indicates that

$$\left\{\begin{array}{l}\left[\sum_{b\in\mathbf{SZ}_o}\sum_{i\in\mathbf{U}(b)}(\bar{\lambda}_{1,bt}/c-\bar{\lambda}_{2,bt}/c)\cdot(P_i^{\max}-\varepsilon)\right.\\ \quad\left.+\sum_{b\in\{\mathbf{Z}-\mathbf{SZ}_o\}}\sum_{i\in\mathbf{U}(b)}(\bar{\lambda}_{1,bt}/c-\bar{\lambda}_{2,bt}/c)\cdot(P_i^{\min}+\varepsilon)\right]\\ \quad-\left[\sum_{b\in\mathbf{SZ}_o}\sum_{i\in\mathbf{U}(b)}(\lambda'_{1,bt}-\lambda'_{2,bt})\cdot(P_i^{\max}-\varepsilon)\right.\\ \quad\left.+\sum_{b\in\{\mathbf{Z}-\mathbf{SZ}_o\}}\sum_{i\in\mathbf{U}(b)}(\lambda'_{1,bt}-\lambda'_{2,bt})\cdot(P_i^{\min}+\varepsilon)\right]\end{array}\right\}\le 0 \qquad \text{(B.6)}$$

With (B.4), $\{\hat{\lambda}_{1,bt}/c, \hat{\lambda}_{2,bt}/c, b\in\mathbf{Z}^c, \lambda'_{1,bt}, \lambda'_{2,bt}, b\in\mathbf{Z}\}$ together with $\{\pi'_{1,pt}, \pi'_{2,pt}, \pi'_{3,lt}, \pi'_{4,lt}, \pi'_t, \kappa'_{lt}\}$ is an optimal solution of (20) since its corresponding objective value is also $\breve{s}_t$. According to Theorem B.1, $\hat{s}\le c\cdot\breve{s}_t$, thus there must exit a solution of $\{\hat{\lambda}_{1,bt}/c, \hat{\lambda}_{2,bt}/c, b\in\mathbf{Z}^c, \lambda'_{1,bt}, \lambda'_{2,bt}, b\in\mathbf{Z}\}$ together with $\{\pi''_{1,pt}, \pi''_{2,pt}, \pi''_{3,lt}, \pi''_{4,lt}, \pi''_t, \kappa''_{lt}\}$ for (20), which has the objective value of $\hat{s}/c$. Thus $\{\hat{\lambda}_{1,bt}, \hat{\lambda}_{2,bt}, b\in\mathbf{Z}^c, c\cdot\lambda'_{1,bt}, c\cdot\lambda'_{2,bt}, b\in\mathbf{Z}\}$ together with $\{c\cdot\pi''_{1,pt}, c\cdot\pi''_{2,pt}, c\cdot\pi''_{3,lt}, c\cdot\pi''_{4,lt}, c\cdot\pi''_t, c\cdot\kappa''_{lt}\}$ is a feasible solution of (19).

Since $\{\hat{\lambda}_{1,bt}, \hat{\lambda}_{2,bt}, b\in\mathbf{Z}^c, \bar{\lambda}_{1,bt}, \bar{\lambda}_{2,bt}, b\in\mathbf{Z}\}$ is the optimal solution of (19), and $\{\hat{\lambda}_{1,bt}, \hat{\lambda}_{2,bt}, b\in\mathbf{Z}^c, c\cdot\lambda'_{1,bt}, c\cdot\lambda'_{2,bt}, b\in\mathbf{Z}\}$ is one of its feasible solutions, we learn that

$$\left\{\begin{array}{l}\left[\sum_{b\in\mathbf{SZ}_o}\sum_{i\in\mathbf{U}(b)}(\bar{\lambda}_{1,bt}-\bar{\lambda}_{2,bt})\cdot(P_i^{\max}-\varepsilon)\right.\\ \quad\left.+\sum_{b\in\{\mathbf{Z}-\mathbf{SZ}_o\}}\sum_{i\in\mathbf{U}(b)}(\bar{\lambda}_{1,bt}-\bar{\lambda}_{2,bt})\cdot(P_i^{\min}+\varepsilon)\right]\\ \quad-\left[\sum_{b\in\mathbf{SZ}_o}\sum_{i\in\mathbf{U}(b)}(c\cdot\lambda'_{1,bt}-c\cdot\lambda'_{2,bt})\cdot(P_i^{\max}-\varepsilon)\right.\\ \quad\left.+\sum_{b\in\{\mathbf{Z}-\mathbf{SZ}_o\}}\sum_{i\in\mathbf{U}(b)}(c\cdot\lambda'_{1,bt}-c\cdot\lambda'_{2,bt})\cdot(P_i^{\min}+\varepsilon)\right]\end{array}\right\}\ge 0 \qquad \text{(B.7)}$$

Multiplying both sides of (B.7) by c, we get (B.8) where $c=\sum_{b\in\mathrm{NZ}}(-\hat{\lambda}_{1,bt}-\hat{\lambda}_{2,bt})+2\cdot|\mathbf{Z}|$ is a positive constant

$$\left\{\begin{array}{l}\left[\sum_{b\in\mathbf{SZ}_o}\sum_{i\in\mathbf{U}(b)}(\bar{\lambda}_{1,bt}/c-\bar{\lambda}_{2,bt}/c)\cdot(P_i^{\max}-\varepsilon)\right.\\ \quad\left.+\sum_{b\in\{\mathbf{Z}-\mathbf{SZ}_o\}}\sum_{i\in\mathbf{U}(b)}(\bar{\lambda}_{1,bt}/c-\bar{\lambda}_{2,bt}/c)\cdot(P_i^{\min}+\varepsilon)\right]\\ \quad-\left[\sum_{b\in\mathbf{SZ}_o}\sum_{i\in\mathbf{U}(b)}(\lambda'_{1,bt}-\lambda'_{2,bt})\cdot(P_i^{\max}-\varepsilon)\right.\\ \quad\left.+\sum_{b\in\{\mathbf{Z}-\mathbf{SZ}_o\}}\sum_{i\in\mathbf{U}(b)}(\lambda'_{1,bt}-\lambda'_{2,bt})\cdot(P_i^{\min}+\varepsilon)\right]\end{array}\right\}\ge 0 \qquad \text{(B.8)}$$

Using (B.6) and (B.8), we have

$$\left\{\begin{array}{l}\left[\sum_{b\in\mathbf{SZ}_o}\sum_{i\in\mathbf{U}(b)}(\bar\lambda_{1,bt}/c-\bar\lambda_{2,bt}/c)\cdot(P_i^{\max}-\varepsilon)\right.\\ \quad\left.+\sum_{b\in\{\mathbf{Z}-\mathbf{SZ}_o\}}\sum_{i\in\mathbf{U}(b)}(\bar\lambda_{1,bt}/c-\bar\lambda_{2,bt}/c)\cdot(P_i^{\min}+\varepsilon)\right]\\ \quad-\left[\sum_{b\in\mathbf{SZ}_o}\sum_{i\in\mathbf{U}(b)}(\lambda'_{1,bt}-\lambda'_{2,bt})\cdot(P_i^{\max}-\varepsilon)\right.\\ \quad\left.+\sum_{b\in\{\mathbf{Z}-\mathbf{SZ}_o\}}\sum_{i\in\mathbf{U}(b)}(\lambda'_{1,bt}-\lambda'_{2,bt})\cdot(P_i^{\min}+\varepsilon)\right]\end{array}\right\}=0 \tag{B.9}$$

Since (B.4) dominates (21), according to the Definition 1 there must exits a $\tilde{\mathbf{P}}$ in the domain of $\{\mathbf{0}, [\mathbf{P}^{\min}, \mathbf{P}^{\max}]\}$ which satisfies the strict inequality of (B.5) as shown in (B.10)

$$\left\{\begin{array}{l}\left[\sum_{b\in\mathbf{Z}^c}\sum_{i\in\mathbf{U}(b)}(\hat\lambda_{1,bt}/c-\hat\lambda_{2,bt}/c)\cdot(\tilde P_i-\hat P_{it})\right.\\ \quad\left.+\sum_{b\in\mathbf{Z}}\sum_{i\in\mathbf{U}(b)}(\bar\lambda_{1,bt}/c-\bar\lambda_{2,bt}/c)\cdot\tilde P_i+\breve{s}_t\right]\\ \quad-\left[\sum_{b\in\mathbf{Z}^c}\sum_{i\in\mathbf{U}(b)}(\hat\lambda_{1,bt}/c-\hat\lambda_{2,bt}/c)\cdot(\tilde P_i-\hat P_{it})\right.\\ \quad\left.+\sum_{b\in\mathbf{Z}}\sum_{i\in\mathbf{U}(b)}(\lambda'_{1,bt}-\lambda'_{2,bt})\cdot\tilde P_i+\breve{s}_t\right]\end{array}\right\}<0 \tag{B.10}$$

Applying $\tilde{\mathbf{P}}$ to (B.9), we get

$$\left\{\begin{array}{l}\left[\sum_{b\in\mathbf{Z}^c}\sum_{i\in\mathbf{U}(b)}(\hat\lambda_{1,bt}/c-\hat\lambda_{2,bt}/c)\cdot(\tilde P_i-\hat P_{it})\right.\\ \quad+\sum_{b\in\mathbf{SZ}_o}\sum_{i\in\mathbf{U}(b)}(\bar\lambda_{1,bt}/c-\bar\lambda_{2,bt}/c)\cdot(P_i^{\max}-\varepsilon)\\ \quad\left.+\sum_{b\in\{\mathbf{Z}-\mathbf{SZ}_o\}}\sum_{i\in\mathbf{U}(b)}(\bar\lambda_{1,bt}/c-\bar\lambda_{2,bt}/c)\cdot(P_i^{\min}+\varepsilon)\right]\\ \quad-\left[\sum_{b\in\mathbf{Z}^c}\sum_{i\in\mathbf{U}(b)}(\hat\lambda_{1,bt}/c-\hat\lambda_{2,bt}/c)\cdot(\tilde P_i-\hat P_{it})\right.\\ \quad+\sum_{b\in\mathbf{SZ}_o}\sum_{i\in\mathbf{U}(b)}(\lambda'_{1,bt}-\lambda'_{2,bt})\cdot(P_i^{\max}-\varepsilon)\\ \quad\left.+\sum_{b\in\{\mathbf{Z}-\mathbf{SZ}_o\}}\sum_{i\in\mathbf{U}(b)}(\lambda'_{1,bt}-\lambda'_{2,bt})\cdot(P_i^{\min}+\varepsilon)\right]\end{array}\right\}=0 \tag{B.11}$$

Theorem 1 points out that there exits a scalar $\theta > 1$ such that $\theta\cdot(P_i^{\max}-\varepsilon)+(1-\theta)\cdot\tilde P_i$ and $\theta\cdot(P_i^{\min}+\varepsilon)+(1-\theta)\cdot\tilde P_i$ are in the range of $\{\mathbf{0}, [\mathbf{P}^{\min}, \mathbf{P}^{\max}]\}$. Thus multiplying (B.11) by θ and (B.10) by $(1-\theta)$ which is negative and preserves the inequality, and adding the two quantities up, we get

$$
\begin{aligned}
&\sum_{b\in\mathbf{Z}^c}\sum_{i\in\mathbf{U}(b)}(\hat{\lambda}_{1,bt}/c-\hat{\lambda}_{2,bt}/c)\cdot(\tilde{P}_i-\hat{P}_{it})\\
&\quad+\sum_{b\in\mathbf{SZ}_o}\sum_{i\in\mathbf{U}(b)}(\bar{\lambda}_{1,bt}/c-\bar{\lambda}_{2,bt}/c)\cdot[\theta\cdot(P_i^{\max}-\varepsilon)+(1-\theta)\cdot\tilde{P}_i]\\
&\quad+\sum_{b\in\{\mathbf{Z}-\mathbf{SZ}_o\}}\sum_{i\in\mathbf{U}(b)}(\bar{\lambda}_{1,bt}/c-\bar{\lambda}_{2,bt}/c)\cdot[\theta\cdot(P_i^{\min}+\varepsilon)+(1-\theta)\cdot\tilde{P}_i]\\
&\;>\sum_{b\in\mathbf{Z}^c}\sum_{i\in\mathbf{U}(b)}(\hat{\lambda}_{1,bt}/c-\hat{\lambda}_{2,bt}/c)\cdot(\tilde{P}_i-\hat{P}_{it})\\
&\quad+\sum_{b\in\mathbf{SZ}_o}\sum_{i\in\mathbf{U}(b)}(\lambda'_{1,bt}-\lambda'_{2,bt})\cdot[\theta\cdot(P_i^{\max}-\varepsilon)+(1-\theta)\cdot\tilde{P}_i]\\
&\quad+\sum_{b\in\{\mathbf{Z}-\mathbf{SZ}_o\}}\sum_{i\in\mathbf{U}(b)}(\lambda'_{1,bt}-\lambda'_{2,bt})\cdot[\theta\cdot(P_i^{\min}+\varepsilon)+(1-\theta)\cdot\tilde{P}_i]
\end{aligned}
\tag{B.12}
$$

Equation (B.12) indicates that

$$
P_{it}=\begin{cases}\tilde{P}_i & \forall i\in\mathbf{U}(b), b\in\mathbf{Z}^c\\ \theta\cdot(P_i^{\max}-\varepsilon)+(1-\theta)\cdot\tilde{P}_i, & \forall i\in\mathbf{U}(b), b\in\mathbf{SZ}_o\\ \theta\cdot(P_i^{\min}+\varepsilon)+(1-\theta)\cdot\tilde{P}_i, & \forall i\in\mathbf{U}(b), b\in\{\mathbf{Z}-\mathbf{SZ}_o\}\end{cases}
$$

then

$$
\left\{
\begin{aligned}
&\Bigg[\sum_{b\in\mathbf{Z}^c}\sum_{i\in\mathbf{U}(b)}(\hat{\lambda}_{1,bt}/c-\hat{\lambda}_{2,bt}/c)\cdot(P_{it}-\hat{P}_{it})\\
&\quad+\sum_{b\in\mathbf{Z}}\sum_{i\in\mathbf{U}(b)}(\bar{\lambda}_{1,bt}/c-\bar{\lambda}_{2,bt}/c)\cdot P_{it}+\breve{s}_t\Bigg]\\
&\quad-\Bigg[\sum_{b\in\mathbf{Z}^c}\sum_{i\in\mathbf{U}(b)}(\hat{\lambda}_{1,bt}/c-\hat{\lambda}_{2,bt}/c)\cdot(P_{it}-\hat{P}_{it})\\
&\quad+\sum_{b\in\mathbf{Z}}\sum_{i\in\mathbf{U}(b)}(\lambda'_{1,bt}-\lambda'_{2,bt})\cdot P_{it}+\breve{s}_t\Bigg]
\end{aligned}
\right\}>0
$$

which is inconsistent with (B.5) and shows that our assumption that (21) is not pareto optimal is untenable. Thus (21) is pareto optimal. □

References

1. Shahidehpour, M., Yamin, H., Li, Z.: Market Operations in Electric Power Systems. Wiley, New York (2002)
2. Benders, J.F.: Partitioning procedures for solving mixed-variables programming problems. Numer. Math. **4**, 238–252 (1962)
3. Fu, Y., Shahidehpour, M., Li, Z.: Security-constrained unit commitment with ac constraints. IEEE Trans. Power Syst. **20**, 1538–1550 (2005)

4. Shahidehpour, M., Fu, Y.: Benders decomposition—applying Benders decomposition to power systems. IEEE Power Energy Mag. **3**, 20–21 (2005)
5. Guan, X., Guo, S., Zhai, Q.: The conditions for obtaining feasible solutions to security-constrained unit commitment problems. IEEE Trans. Power Syst. **20**, 1746–1756 (2005)
6. Martínez-Crespo, J., Usaola, J., Fernández, J.L.: Security-constrained optimal generation scheduling in large-scale power systems. IEEE Trans. Power Syst. **21**, 321–332 (2006)
7. Alguacil, N., Conejo, A.J.: Multiperiod optimal power flow using Benders decomposition. IEEE Trans. Power Syst. **15**, 196–201 (2000)
8. Wu, L., Shahidehpour, M., Liu, C.: MIP-based post-contingency corrective action with quick-start units. IEEE Trans. Power Syst. **24**, 1898–1899 (2009)
9. Li, Y., McCalley, J.D.: A general Benders decomposition structure for power system decision problems. In: EIT 2008 IEEE International Conference, pp. 72–77 (2008)
10. Magnanti, T.L., Wong, R.T.: Accelerating Benders decomposition: algorithmic enhancement and model selection criteria. Oper. Res. **29**, 464–484 (1981)
11. McDaniel, D., Devine, M.: A modified Benders' partitioning algorithm for mixed integer programming. Manag. Sci. **24**, 312–319 (1977)
12. Côté, G., Laughton, M.A.: Large-scale mixed integer programming: Benders-type heuristics. Eur. J. Oper. Res. **16**, 327–333 (1984)
13. Zakeri, G., Philpott, A.B., Ryan, D.M.: Inexact cuts in Benders decomposition. SIAM J. Optim. **10**, 643–657 (2000)
14. Fischetti, M., Salvagnin, D., Zanette, A.: Minimal infeasible subsystems and Benders cuts. http://www.dei.unipd.it/~fisch/papers/Benders.pdf (2010). Accessed 12 April 2010
15. Rei, W., Gendreau, M., Cordeau, J.F., Soriano, P.: Accelerating Benders decomposition by local branching. Informs J. Comput. **21**, 333–345 (2009)
16. Fu, Y., Li, Z., Shahidehpour, M.: Accelerated Benders decomposition in SCUC. IEEE Tran. Power Syst. (2010)
17. CPLEX 11.0. User's manual. http://www-01.ibm.com/software/integration/optimization/cplex/ (2010). Accessed 12 April 2010
18. Fu, Y., Shahidehpour, M.: Fast SCUC for large-scale power systems. IEEE Trans. Power Syst. **22**, 2144–2151 (2007)
19. Li, T., Shahidehpour, M.: Price-based unit commitment: a case of Lagrangian relaxation versus mixed integer programming. IEEE Trans. Power Syst. **20**, 2015–2025 (2005)