Sensitivity of Distributed Optimization Convergence Performance to Reference Bus Location

A. Mohammadi, *Student Member, IEEE*, F. Safdarian, *Student Member, IEEE*, and A. Kargarian, *Member, IEEE*

***Abstract*—Distributed optimization are becoming popular to solve a large power system problem with the objective of reducing computational complexity. To this end, the convergence performance of distributed optimization plays an important role to solve an optimal power flow (OPF) problem. One of the critical factors that have a significant impact on the convergence performance is the reference bus location. Since selecting the reference bus location does not affect the result of centralized DC OPF, we can change the location of the reference bus to get more accurate results in distributed optimization. In this paper, our goal is to provide some insights into how to select reference bus location to have a better convergence performance. We modeled the power grid as a graph and based on some graph theory concepts, for each bus in the grid a score is assigned, and then we cluster buses to find out which buses are more suitable to be considered as the reference bus. We implement the analytical target cascading (ATC) on the IEEE 48-bus system to solve a DC OPF problem. The results show that by selecting a proper reference bus, we obtained more accurate results with an excellent convergence rate while improper selection may take much more iterations to converge.**

***Index Terms*—Analytical target cascading, reference bus, PageRank, optimal power flow.**

I. Introduction

DISTRIBUTED optimization methods have become popular to solve power system problems, such as optimal power flow (OPF). The main advantages of the distributed methods in comparison with centralized methods are: 1) respecting the information privacy of autonomous agents, 2) enhancing the system performance against a single point failure, and 3) taking advantage of parallel computing to reduce the computational burden of large-scale problems.

Several distributed and decentralized optimization algorithms are proposed in the literature to solve power system problems. In [1], several popular distributed and decentralized optimization algorithms (e.g., auxiliary problem principle [2, 3], alternating direction method of multipliers [2, 4], consensus+innovation [5], optimality condition decomposition [6], and analytical target cascading (ATC) [7]) are discussed, and their application on DC OPF is studied. To solve OPF in a distributed manner, the power system needs to be decomposed into several subproblems. The way that the system is decomposed and the number of subproblems have a significant impact on the accuracy of obtained results and the convergence performance of distributed algorithms [8, 9]. Various strategies have been presented in the literature to improve the convergence performance of the distributed algorithms. Most of these strategies try to make changes in the structure of distributed algorithms. For instance, in [10], authors enhance the convergence rate by updating Lagrange multipliers based on the concept of the Nesterov technique. Since proper initialization plays an important role in convergence rate, reference [7] proposes a hot start to achieve a better convergence speed. Finding the best way of network decomposition to take advantage of parallel computing is investigated in [11].

Since the reference bus in a power system is predetermined, the impact of the location of reference bus in distributed algorithms is neglected in the existing literature. For a DC OPF problem, if one varies the reference bus location and solves a new OPF problem, the same result will be obtained as the reference bus only set a reference point for bus voltage angles. Although the reference bus does not affect the accuracy of a centralized optimization, we cannot disregard its impact on distributed algorithms. By selecting different reference buses, the convergence performance of the algorithms varies significantly.

In this paper, we analyze the impact of reference bus location on the convergence performance of the ATC-based distributed DC OPF algorithm. We study how to find a proper reference bus based on graph theory techniques to speed up the convergence performance while the solution accuracy is excellent. The power network is modeled as a graph where buses and lines are modeled, respectively, as nodes and edges. The PageRank method is presented to assign a centrality score to each bus. The centrality score is used to find the most efficient reference bus location. The IEEE 48-bus test system is used. To make the results more tangible, we cluster the buses to several groups. Several insights are provided to choose a proper reference bus location.

II. Centralized OPF Formulation

To demonstrate the impact of reference bus location on distributed optimization, DC OPF is studied. The centralized OPF problem is formulated as follows:

$$\min \sum_{i=1}^{NG} \underbrace{a_i \cdot p_i^2 + b_i \cdot p_i + c_i}_{f_i(p)} \tag{1}$$

$s.t.$

The authors are with the department of Electrical and Computer Engineering, Louisiana State University, Baton Rouge, LA 70803 USA (email: amoha39@lsu.edu, fsafda1@lsu.edu, kargarian@lsu.edu)

978-1-7281-1981-6/19/$31.00 ©2019 IEEE

$$h(x)=0 \leftrightarrow \begin{cases} p_i - d_i = \sum_{j\in\Psi_i} \frac{\delta_i-\delta_j}{X_{ij}} \quad \forall i & (2) \\ \delta_{ref}=0 & (3) \end{cases}$$

$$g(x)\leq 0 \leftrightarrow \begin{cases} P_{L_{ij}}^{min} \leq P_{L_{ij}} = \frac{\delta_i-\delta_j}{X_{ij}} \leq P_{L_{ij}}^{Max} & (4) \\ P_i^{min} \leq p_i \leq P_i^{Max} \quad \forall i & (5) \end{cases}$$

where a_i, b_i and c_i are the cost coefficients of generating units, and NG is the number of generators in the system. $h(x)$ and $g(x)$ are the sets of equality and inequality constraints. p_i and d_i indicate the amount of generation and load at bus i. δ_i denotes voltage angle at bus i. Ψ_i is the set of indices of neighbor buses that are connected to bus i. $P_{L_{ij}}$ indicates the amount of power flow between bus i and j. P_L^{min} and P_L^{max} are the minimum and maximum line limits.

Remark: The role of a reference bus in DC OPF is different than the role of a slack bus in AC OPF. In DC OPF, the reference bus sets a reference point for bus voltage angles.

III. ATC STRUCTURE

Consider that the centralized OPF is divided into several subproblems each of which pertains to a geographical zone of the system. Local OPF problems need to become coordinated to determine the optimal solution for the whole system (i.e., the solution of the centralized OPF). ATC is implemented to solve the DC OPF problem in a distributed manner. ATC is a multi-level distributed optimization approach. Subproblems in upper levels act as parents for subproblems in the lower levels that are named children. Each child has one parent, while one parent may have several children. Voltage angles of cross-border buses (buses that connected subsystems) are shared variables between parent and children are voltage angles of boundary buses. ATC should be solved iteratively until both parent and children come to an agreement on the value of shared variables.

Consider that centralize OPF is formulated as the following compact form:

$$\begin{cases} \min_x F(x) & (6) \\ s.t. \quad g(x)\leq 0, \qquad h(X)=0 & (7) \end{cases}$$

The optimization formula for subproblems, while they are still considered within a centralized format, is as follows:

$$\min \sum_{\forall k}\sum_{\forall l} f_{kl}(x_{kl}, t_{(k+1)d_1}, \dots, t_{(k+1)d_{i_{kl}}}) \tag{8}$$

$$s.t. \quad g_{kl}\left(x_{kl}, t_{(k+1)d_1}, \dots, t_{(k+1)d_{i_{kl}}}\right) \leq 0 \tag{9}$$

$$h_{kl}\left(x_{kl}, t_{(k+1)d_1}, \dots, t_{(k+1)d_{i_{kl}}}\right) = 0 \tag{10}$$

$$\forall l \in E_k \quad k=\{1,\dots,K\}$$

where k and l denote the level number and subproblem number in level k. The sets of shared variables for subproblem kl is dedicated by t_{kl}. The shared variable (t) which is propagated from parent to children is called target. The number of subproblems in level k is denoted by $d_{i_{kl}}$, K denotes the number of levels, and E_k is the number of subproblems in level K.

To segregate the subproblems from each other, a response copier is defined for each target variable. The response variables are copies of targets and are sent from children to parents. Since each response is a duplication of a target, the response needs to be the same as the target value. To this end, a set of consistency constraints is defined for every two connected subproblems to obtain a feasible solution:

$$C_{kl} = t_{kl} - r_{kl} = 0 \tag{11}$$

Since C_{kl} is a hard constraint which is not desirable for distributed optimization, it can be relaxed by a penalty function using the concept of augmented Lagrangian.

$$\min \sum_{\forall k}\sum_{\forall l} f_{kl}(x_{kl}, t_{(k+1)d_1}, \dots, t_{(k+1)d_{i_{kl}}}) + \pi_{kl}(c_{(k+1)d_1}, \dots, c_{(k+1)d_{i_{kl}}}) \tag{12}$$

where π_{kl} is the penalty function for each set of consistency constraint in level k and subproblem l. The penalty function is not a unique function. For instance, it can be exponential or quadratic.

Augmented Lagrangian function with alternating direction method of multipliers (ALAD) is an ATC coordination strategy that penalizes the consistency constraint with a quadratic function an solves the subproblems hierarchically.

$$\pi_{ALAD}(t,r) = \lambda^T(t-r) + \|w \circ (t-r)\|_2^2 \tag{13}$$

where λ and w are multipliers, and "∘" denotes the Hadamard product.

For the sake of simplicity, we consider two OPF subproblems one acts as a parent in located level one, and the other one acts as a child in level two. The subproblems are coupled through shared variables (i.e., voltage angles of borer buses). The ATC-based distributed OPF is formulated as follows:

$$\min_{(x_{11},\delta_{22})} f_{11}(x_{11},\delta_{22}) + \lambda^T\left(\delta_{22} - \acute{\delta}_{22}^{\nu-1}\right) + \left\|w \circ \left(\delta_{22} - \acute{\delta}_{22}^{\nu-1}\right)\right\|_2^2 \tag{14}$$

$$\min_{(x_{11},\acute{\delta}_{22})} f_{22}(x_{11},\acute{\delta}_{22}) + \lambda^T\left(\delta_{22} - \acute{\delta}_{22}^{\nu-1}\right) + \left\|w \circ \left(\delta_{22} - \acute{\delta}_{22}^{\nu-1}\right)\right\|_2^2 \tag{15}$$

where δ and $\acute{\delta}$ are target and response variables, and ν denotes the number of iterations. The first subproblem is solved, while the second subproblem is idle. The share variable δ is propagate own to the second level subproblem. Then, the second level subproblem minimizes its objective function, calculates $\acute{\delta}$, an sends it toward the first level problem. The Lagrange multiplier is updated as

$$\lambda^\nu = \lambda^{\nu-1} + 2w^2 \circ (\delta^{\nu-1} - \acute{\delta}^{\nu-1}) \tag{16}$$

and the next iteration is carried out. This iterative procedure continues until the difference between δ and $\hat{\delta}$ becomes smaller than the desired threshold (ϵ_{thr}). The ATC-based ALAD algorithm is summarized in the following pseudo code.

Solution Algorithm of ALAD

1: **initialize** $X = \lambda, w, \delta$ and $\hat{\delta}$
2: **while** $\max(\|\boldsymbol{\delta}^k - \widetilde{\boldsymbol{\delta}}^k\|) \leq \epsilon_{thr}$, $k = k + 1$ **do**
3: Solve (18) and (19) in a hierarchical manner
5: Update multiplier λ^k
6: **end while**

IV. Reference Bus Location based on Graph Centrality

In this section, insights about the location of reference bus are provided, and an approach is presented to quantify the buses based on features of distributed optimization and graph centrality theory.

A. Graph modeling

The power network is considered as a graph in which buses are graph nodes, and lines are graph edges. The weight of each edge might be defined with respect to different factors, such as the impedance, reactance, and power flow. A simple approach is to set all weights equal to one. This merely means that when two buses are connected, the connected edge between two nodes is one, otherwise it is zero. This represents the topology of the network based on connectivity. Although connectivity of edges is important for distributed optimization studies, more detailed information is needed to have in-depth analysis. We propose to define the edge weights based on the line reactance ($1/X$). In comparison with setting all wights to one, setting weights to $1/X$ can be used as a good indicator to distinguishes whether the connection between each pair of nodes is strong or week.

B. Graph Centrality and Importance of Buses Based on Page Rank

The concept of centrality was initially presented for the social network suggestion to indicate which person (i.e., node) plays a more important role in the network. In other words, who is the central person. Different centrality measures are proposed recently and each of which is used for a specific purpose. Well-known methods for centrality measurement include degree centrality, closeness centrality, betweenness centrality, Eigen centrality, and page rank. For example, in degree centrality, nodes with higher degree (i.e., several edges are connected to such nodes) are the most important nodes in the graph.

We look for the node with the highest centrality score, and consider it as the reference bus for DC OPF. To do so, we have to find the reference buses that show the same convergence performance regardless of the value of tuning parameters and Lagrange multipliers. In other words, if the reference bus has a good convergence performance, by changing multipliers it should still show good convergence performance. In this regard, DC OPF is solved by distributed optimization for different tuning parameters. The results show ending buses of tie-lines connected between the two border buses has the most stable behavior. Regardless of the amounts of tuning parameters and Lagrange multipliers, when we select a border bus as the reference bus, the convergence performance is acceptable.

In the next step, the best reference bus location should be chosen among border buses. We implement the PageRank method to assign a centrality score to each border bus. The PageRank method was first proposed by Google to assign a score to websites. The website that has a higher score is more popular in the web. The PageRank method can assign a centrality score within nodes by counting the number and quality of links (neighbor node connection). The number of connections and the quality of links to the website are important, and based on them, the importance of the website is determined. Generally, if more links are received from another website, it means that this website is more important. Based on that, the algorithm assigns a weight to each node in a network. Several factors are important in the PageRank such as the number of links which are received from the other nodes, the tendency of other nodes, and the centrality of other connected nodes. We can conclude that in the PageRank, a node that is highly linked with other nodes is potentially important (i.e., the node has a high centrality score) if its neighboring nodes are important as well. To demonstrate how the PageRank algorithm works, an example is illustrated in Fig. 1 [12].

Step 1: First, each node score is initialized $1/n$ where n is the number of nodes. Since three nodes exist in this example, the score of each node is:

$$P(A) = P(B) = P(C) = 1/3 \tag{17}$$

Step 2: To calculate PageRank for each node, we first find connection links and then determine the contribution of each adjacent node on the corresponding node. Node C is connected to nodes A and B. Node A contributes ½ because it has one sending and one receiving link. Therefore, node A contributes $1/2 \times 1/3$ to the PageRank of C. Node B contributes $1 \times 1/3$. Hence, the PageRank of node C is calculated as follows:

$$P(C) = 1/2 * 1/3 + 1 * 1/3 = 1/2 \tag{18}$$

Likewise, for the rest of the nodes, we have:

$$P(B) = 1/2 * 1/3 + 0 * 1/3 = 1/6 \tag{19}$$
$$P(A) = 0 * 1/3 + 1 * 1/3 = 1/3 \tag{20}$$

Step 3: Since PageRank is an iterative algorithm, Step 2 need to be repeated several times until it converges and centrality of each node does not change. A node with a higher score is more central in the graph. This algorithm can be applicable for both directed and undirected graph.

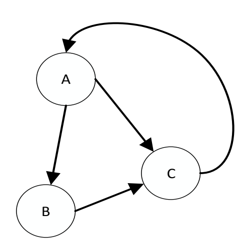

Fig. 1. PageRank Procedure

V. NUMERICAL RESULTS AND DISCUSSIONS

The objective is to find a proper reference bus location in the IEEE 48-bus system to enhance the convergence performance of ATC-based DC OPF. Based on our experience, by selecting the border buses as the reference bus, regardless of the values of multipliers, good convergence performance is achieved. We select the border buses as candidate options for the reference bus location. We, then, implement the PageRank to distinguish which border bus might have the best performance regarding to the convergence speed. The ALAD method is implemented to justify the results. For all cases, the algorithm starts from the same initial point to have a fair comparison. All simulations are carried out in MATLAB on a personal computer with 2.7 GHz CPU and 16GB of RAM.

A. Centrality Score using PageRank

To construct a graph, we use the reactance of each line as the edge weight. Using the GEPHI program, we find the PageRank value for each border bus. The main reason that border buses are selected as candidate reference buses is that by selecting a border bus as the reference bus, the error gap between the target and the response for that bus become zero. It means that the consistency constraint corresponding to one of the border buses is already satisfied. In Fig. 2, the border buses are shown by boldface font. Centrality scores for border buses {7, 27, 13, 23, 41, 39} are {0.0132, 0.0147, 0.0200, 0.0271, 0.272, 0.0309}, respectively. We expect that by selecting buses 41 and 39, we get the best convergence performance because these buses have the highest scores among the candidate reference buses.

B. Impact of Reference Bus Location on Distributed Algorithm

The ALAD method is implemented to find the DC OPF solution. We define three indices to evaluate the results. The first index is the relative distance of the overall cost which is calculated based on the centralized OPF and ALAD results.

$$rel_{cost} = \frac{|f_c - f_d|}{f_c} \tag{21}$$

where f_c and f_d are the optimal operation cost obtained, respectively, by centralized and distributed algorithms.

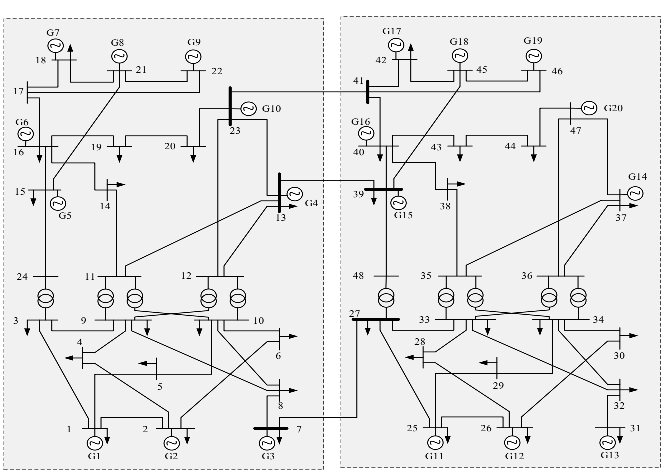

Fig. 2 The IEEE-48 bus system with two zones (subproblems).

The second index demonstrates the average of rel_{cost} over the course of iterations. Although rel_{cost} has a decreasing trend generally, it has some fluctuations and jumps during the convergence process. Therefore, we formulate $rel_{avg_{cost}}$ as follows:

$$rel_{avg_{cost}} = \frac{\sum_{\nu=1}^{\mathrm{N}} rel_{cost}^{\nu}}{\mathrm{N}} \tag{22}$$

where rel_{cost}^{ν} is the total cost function of the distributed algorithm in iteration ν. In addition, we calculate C_{avg}, i.e., the overall average of consistency constraints over the course of iterations:

$$C_{avg} = \frac{\sum_{\nu=1}^{\mathrm{N}} |\delta^{\nu} - \hat{\delta}^{\nu}|}{\mathrm{N} \times 2\rho} \tag{23}$$

where N is the maximum number of iterations. Parameter ρ is the number of connected tie-lines between subproblems (note that each tie-line has two shared variables).

To show impact of the reference bus location, we first demonstrate a wide range of rel_{cost} changes over the course of iterations. Then, we compare one of the best- and worst-case scenario errors over using $w = 250$, $\lambda = 250$ and $\nu \leq 50$ as initial values. By changing δ_{ref}, rel_{cost} changes from $7.5e-5$ to $1.2e-16$ after 50 iterations (see Fig.3). For the sake of explanation, we investigate two cases. Figure 4 demonstrates that by selecting bus 41 as the reference bus, the accuracy is much better as compared with the case in which bus 22 is selected as the reference bus. In this case, results for $\delta_{ref} = 22$ after 50 iterations are the same as $\delta_{ref} = 41$ after 10 iterations. This means 400% improvement regarding the convergence speed. This demonstrates the high impact of reference bus location on the convergence performance.

In another test, we justify that the results that come from the graph analysis are the same as what we expect. To this end, we select w and λ from {80,150,250} to analysis the impact of different ranges of multipliers on the results. For each w and λ, we run DC OPF by changing the reference bus location and obtain the $rel_{avg_{cost}}$ and C_{avg} indices. Using the Kmeans method, we cluster the results into two different classes. In Fig. 5, the blue color class is the one that has the best convergence performance while the red class refers to the rest of reference buses which may have normal or bad convergence performance. The results demonstrate that buses {41, 39, 13, 23} are always in the blue class regardless of multipliers value, but buses {7, 27} cannot achieve a good result. This conclusion follows what we expect from centrality scores obtained by PageRank

VI. CONCLUSION

In this paper, we investigated the impact of the reference bus location on a distributed DC OPF. We deploy the PageRank centrality technique to measure the centrality of border buses. Based on this score, we can anticipate which reference bus location can have a better convergence result. To this end, we introduce some indices to evaluate our proposed technique. We implemented ALAD method as a decentralized technique to solve a DC OPF problem for the IEEE-48 bus system. The results demonstrated that by changing the reference bus location the convergence speed and accuracy vary a lot. In this case, if we select the reference bus based on the graph analysis, we can reach to very good results.

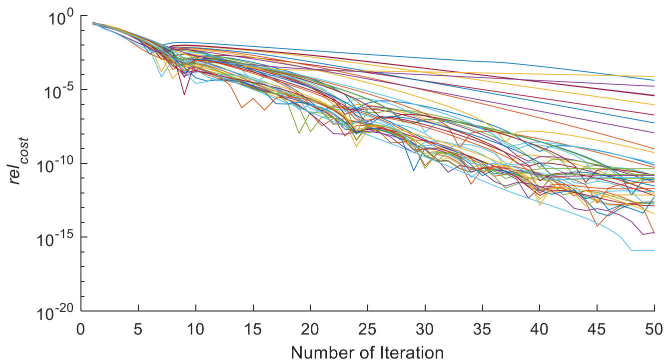

Fig. 3. The relative error over the course of iterations.

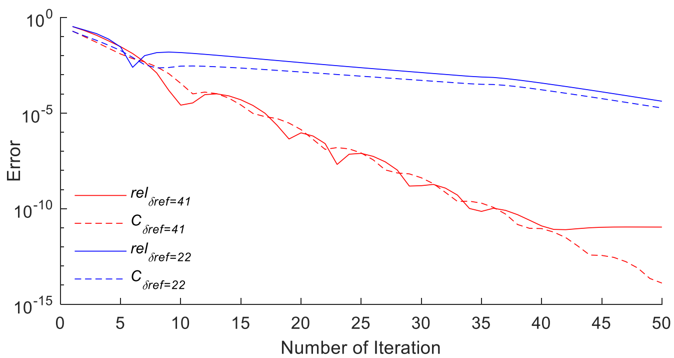

Fig. 4. Relative and consistency constraints errors over the course of iterations for two different buses.

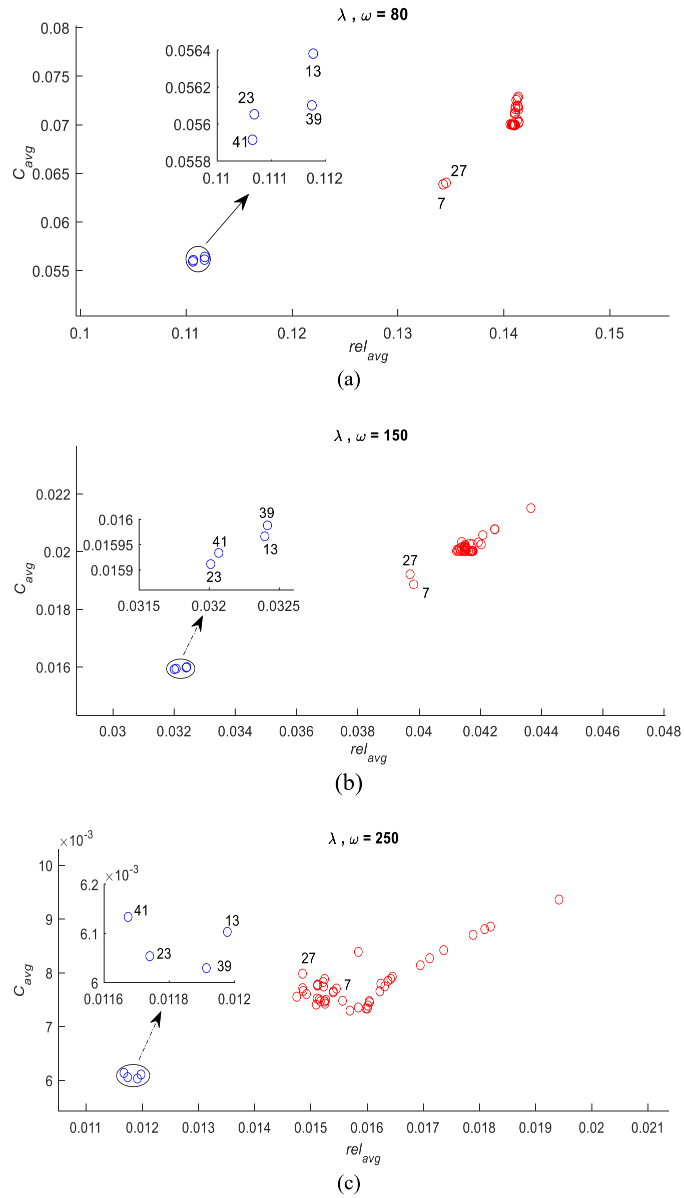

Fig. 5. Clustering the performance of reference bus location for different multipliers, a) $\lambda = w = 80$, b) $\lambda = w = 150$, and c) $\lambda = w = 250$.

REFERENCE

[1] A. Kargarian, J. Mohammadi, J. Guo, S. Chakrabarti, M. Barati, G. Hug, *et al.*, "Toward distributed/decentralized DC optimal power flow implementation in future electric power systems," *IEEE Transactions on Smart Grid,* vol. 9, pp. 2574-2594, 2018.

[2] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning,* vol. 3, pp. 1-122, 2011.

[3] J. Guo, O. Tonguz, and G. Hug, "Impact of power system partitioning on the efficiency of distributed multi-step optimization," in *PowerTech, 2017 IEEE Manchester*, 2017, pp. 1-6.

[4] S. Tosserams, L. Etman, P. Papalambros, and J. Rooda, "An augmented Lagrangian relaxation for analytical target cascading using the alternating direction method of multipliers," *Structural and multidisciplinary optimization,* vol. 31, pp. 176-189, 2006.

[5] S. Kar, G. Hug, J. Mohammadi, and J. M. Moura, "Distributed State Estimation and Energy Management in Smart Grids: A Consensus ${+} $ Innovations Approach," *IEEE Journal of selected topics in signal processing,* vol. 8, pp. 1022-1038, 2014.

[6] A. J. Conejo, F. J. Nogales, and F. J. Prieto, "A decomposition procedure based on approximate Newton directions," *Mathematical programming,* vol. 93, pp. 495-515, 2002.

[7] A. Mohammadi, M. Mehrtash, and A. Kargarian, "Diagonal quadratic approximation for decentralized collaborative TSO+ DSO optimal power flow," *IEEE Transactions on Smart Grid,* 2018.

[8] J. Guo, G. Hug, and O. Tonguz, "Impact of partitioning on the performance of decomposition methods for AC Optimal Power Flow," in *Innovative Smart Grid Technologies Conference (ISGT), 2015 IEEE Power & Energy Society*, 2015, pp. 1-5.

[9] A. Mohammadi, M. Mehrtash, A. Kargarian, M. Barati,, "Tie-Line Characteristics based Partitioning for Distributed Optimization of Power Systems," Power and Energy Society General Meeting (PESGM), Portland, OR, 2018.

[10] T. Goldstein, B. O'Donoghue, S. Setzer, and R. Baraniuk, "Fast alternating direction optimization methods," *SIAM Journal on Imaging Sciences,* vol. 7, pp. 1588-1623, 2014.

[11] J. M. Mulvey and A. Ruszczyn, "A diagonal quadratic approximation method for large scale linear programs," *Operations Research Letters,* vol. 12, pp. 205-215, 1992.

[12] Commonlounge. (2018,08,31). *Network Analysis, metrics for Centrality and The PageRank Algorithm*. Available: https://www.commonlounge.com/discussion/b4933ff1f59244ed9a83fcb01005696b