

An Iterative Approach to Improving Solution Quality for AC Optimal Power Flow Problems

Ling Zhang and Baosen Zhang

lzhang18@uw.edu, zhangbao@uw.edu

Department of Electrical and Computer Engineering, University of Washington
Seattle, WA, USA

ABSTRACT

The existence of multiple solutions to AC optimal power flow (ACOPF) problems has been noted for decades. Existing solvers are generally successful in finding local solutions, which satisfy first and second order optimality conditions, but may not be globally optimal. In this paper, we propose a simple iterative approach to improve the quality of solutions to ACOPF problems. First, we call an existing solver for the ACOPF problem. From the solution and the associated dual variables, we form a partial Lagrangian. Then we optimize this partial Lagrangian and use its solution as a warm start to call the solver again for the ACOPF problem. By repeating this process, we can iteratively improve the solution quality, moving from local solutions to global ones. We show the effectiveness of our algorithm on standard IEEE networks. The simulation results show that our algorithm can escape from local solutions to achieve global optimum within a few iterations.

CCS CONCEPTS

• Hardware → Power and energy.

ACM Reference Format:

Ling Zhang and Baosen Zhang. 2022. An Iterative Approach to Improving Solution Quality for AC Optimal Power Flow Problems. In *The Thirteenth ACM International Conference on Future Energy Systems (e-Energy '22)*, June 28–July 1, 2022, Virtual Event, USA. ACM, New York, NY, USA, 13 pages. <https://doi.org/10.1145/3538637.3538858>

1 INTRODUCTION

The optimal power flow (OPF) problem is a fundamental resource allocation problem in power system operations that minimizes the cost of power generation while satisfying demand. The ACOPF formulation of the problem uses nonlinear power flow equations, resulting in nonlinear and nonconvex optimization problems [6, 10, 19]. The consequence of the nonconvexity of ACOPF we study in this paper is the presence of multiple solutions.

Most ACOPF problems are solved via variations of nonlinear optimization algorithms, including Newton-Raphson, sequential programming, interior points and others (see [7, 19, 26] and the references within). These algorithms are in general only able to certify whether a solution is locally optimal, that is, they satisfy first order and/or second order optimality conditions. Because of

the existence of multiple local solutions, it is often difficult to find the globally optimal one and the system loses efficiency.

The existence of multiple local solutions of the OPF problem has been well-known for several decades [13, 17, 22]. Despite this, a common assumption is that OPF problems tend to have a single “practical” solution that is globally optimal, and therefore the fact that multiple solutions can exist do not impact day-to-day operations [21, 30]. However, an increasingly large body of work have pointed to that multiple solutions do occur under reasonable conditions and cannot be easily ruled out [5, 19, 31]. For example, during day-to-day operations, small changes in the load could lead to a solver jumping between solutions that are far apart from each other [8, 23]. In addition, [5] shows how modifications of the standard IEEE benchmarks can lead to each having more than one local solution. Statistical studies in [12, 15] show that there are more solutions than previously thought in many systems.

An open question in the field is to develop algorithms that can find globally optimal solutions, or at least improve upon local ones. In addition to lowering the operating cost, understanding and distinguishing between locally and globally optimal solutions can lead to important theoretical discoveries about the ACOPF problem. Consequently, several classes of algorithms have been developed. For example, holomorphic embedding has been used in [9, 14], but are slow and require very high numerical precision. Genetic algorithms can escape a local minimum, but are random in nature and require repeated trial and error [1, 3]. Robust optimal power flow can alleviate convergence issues, but may not improve on the quality of the solution [25]. Compared to meshed networks, radial networks are less complicated and there have been advances in the solving techniques for radial networks to avoid being kept at strictly local solutions. A common method is to convexify the nonlinear and nonconvex power flow equations [11, 16, 20, 32]. However, these relaxations can have difficulties when there are lower bounds on reactive or active power and the feasibility region of the problem becomes disconnected.

In this paper, we propose a simple algorithm that can effectively escape from strict local solutions to find better ones. By moving from one solution to another while reducing the cost, we can successively move towards the globally optimal solution. In contrast to algorithms that are launched repeatedly with random initializations, our proposed algorithm is deterministic. And it relies on duality theory to provide better warm starts to existing solvers.

Our process is outlined in Fig. 1. First, we solve the ACOPF problem using some solver (e.g., IPOPT [29] or Matpower [33]). From the solution and its associated dual variables, we form a partial Lagrangian by dualizing the power balance equations. We then optimize this partial Lagrangian, which leads to a different



This work is licensed under a Creative Commons Attribution International 4.0 License. *e-Energy '22*, June 28–July 1, 2022, Virtual Event, USA
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9397-3/22/06.
<https://doi.org/10.1145/3538637.3538858>

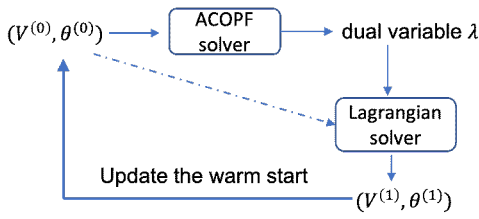


Figure 1: Outline of our algorithm. We form partial Lagrangians at local optima and solve it to provide better warm starting points. Note that the partial Lagrangian is solved using the same initialization as in the first call of the solver.

solution. Using this second solution as a warm start, we again call the solver for the ACOPF problem. Interestingly, this iterative process moves from higher cost to lower cost ones.

The key reason for the success of the algorithm is that the geometry of the partial Lagrangian is much more “friendly” than the geometry of the original ACOPF problem. We provide both theoretical analysis on tree networks and simulation results on standard 9-bus, 22-bus, 39-bus, 118-bus and 300-bus meshed networks and also on the IEEE 141-bus radial network. We show that our algorithm can quickly escape from local solutions and find lower cost solutions. This feature holds even for ACOPF problems with disconnected feasible spaces, for example, the 2-bus network shown in Section 4.1, which has been traditionally difficult to deal with [5, 19]. For networks with known global solutions (3, 9, 22, 118, 300-bus), we show that our algorithm can find the globally optimal solution in a single iteration, even starting from a strict local solution.

Our approach can be seen as a way to provide good warm starts to nonlinear optimization solvers. DCOPF is commonly used, although it can fail to find good starting points as shown by our simulations as well as existing results [2]. More sophisticated approaches either randomizes (stochastic gradient in [8] and load fluctuations in [23]) or uses a previous solution as the starting point [28]. The former tends to be time consuming, while the latter tends to lead to system being stuck in a strict local solution [5].

2 MODEL AND PROBLEM FORMULATION

Consider a power system network of n buses and m lines. For bus i , let V_i denote its voltage magnitude, θ_i its angle, P_i^G and Q_i^G the active and reactive output of the generator and P_i^D and Q_i^D the active and reactive load. The admittance between i and j is $g_{ij} - jb_{ij}$, the active power and reactive power flow from i to j is P_{ij}^f and Q_{ij}^f and θ_{ij} is used as a shorthand for $\theta_i - \theta_j$.

The ACOPF problem is to minimize the cost of active power generations while satisfying a set of constraints [5]. Out of the feasible solutions, we focus on two classes: local solutions and global solutions. Local solutions are all the solutions that satisfy local optimality conditions, for example, the KKT conditions or second order ones [4]. Out of this set, the solutions with the lowest cost are called the global ones. We sometimes refer to the local solutions that are not global as strict local solutions. We assume both strict local solutions and the global solutions are regular and the KKT conditions always hold at these solutions [4]. Therefore,

there always exist unique Lagrangian multipliers and we can use them to form the partial Lagrangian.

Specifically, the ACOPF problem is given by the following optimization problem:

$$\min_{\mathbf{V}, \boldsymbol{\theta}} \sum_i c_i (P_i^G) \quad (1a)$$

$$\text{s.t. } P_i^G = P_i^D + \sum_{j=1}^N P_{ij}^f \quad (1b)$$

$$Q_i^G = Q_i^D + \sum_{j=1}^N Q_{ij}^f \quad (1c)$$

$$P_{ij}^f = V_i^2 g_{ij} - V_i V_j (g_{ij} \cos(\theta_{ij}) - b_{ij} \sin(\theta_{ij})) \quad (1d)$$

$$Q_{ij}^f = V_i^2 \hat{b}_{ij} - V_i V_j (b_{ij} \cos(\theta_{ij}) + g_{ij} \sin(\theta_{ij})) \quad (1e)$$

$$\underline{V}_i \leq V_i \leq \bar{V}_i \quad (1f)$$

$$\underline{P}_i^G \leq P_i^G \leq \bar{P}_i^G \quad (1g)$$

$$\underline{Q}_i^G \leq Q_i^G \leq \bar{Q}_i^G \quad (1h)$$

$$(P_{ij}^f)^2 + (Q_{ij}^f)^2 \leq (S_{ij}^{\max})^2 \quad (1i)$$

where $\hat{b}_{ij} = b_{ij} + 0.5b_{ij}^C$ and b_{ij}^C is the line charging susceptance. The constraints (1b) and (1c) enforce power balance, (1d) and (1e) are the AC power flow equations, (1f) limits the bus voltage magnitudes, (1g) and (1h) represent the active and reactive limits and (1i) are the line flow limits. We assume the cost at each bus i , i.e., $c_i(\cdot)$, is increasing. Other than that, the cost can be linear, quadratic or other functions. We assume problem (1) is feasible in this paper.

Over the years, many nonlinear programming (NLP) solvers have been developed for the ACOPF problem, and their speed and efficiency have improved dramatically (e.g., see [6] and the references within). However, NLP solvers are typically only able to return local solutions. Since a local solution is not necessarily global, we propose an iterative approach to improve the solution quality by alternatively solving (1) and a partial Lagrangian. Any NLP solver can be used, and we use IPOPT [29] in this paper.

3 ALGORITHM

Our algorithm starts with a call to a NLP solver with some initial guess, denoted by $\boldsymbol{\theta}_{\text{init}}$, $\mathbf{V}_{\text{init}}$. For example, this can be the standard flat start with voltage magnitudes being 1 p.u. and angles set to 0. Then we assume the solver returns a feasible solution. Of course, we don't know whether this solution is globally optimal or not. At this solution, we record the dual variables associated with the power balance equations (1b) and (1c), denoted as $\bar{\boldsymbol{\mu}}^P$ and $\bar{\boldsymbol{\mu}}^Q$. Using these dual variables, we form the following partial Lagrangian by dualizing the power balance equations:

$$\begin{aligned} \mathcal{L}(\mathbf{V}, \boldsymbol{\theta}, \boldsymbol{\mu}^P, \boldsymbol{\mu}^Q) = & \sum_i c_i P_i^G + \sum_i \mu_i^P (P_i^D + \sum_{j=1}^N P_{ij}^f - P_i^G) \\ & + \sum_i \mu_i^Q (Q_i^D + \sum_{j=1}^N Q_{ij}^f - Q_i^G). \end{aligned}$$

We then minimize the partial Lagrangian by solving

$$\begin{aligned} \min_{\mathbf{V}, \boldsymbol{\theta}} \mathcal{L}(\mathbf{V}, \boldsymbol{\theta}, \boldsymbol{\mu}^P, \boldsymbol{\mu}^Q) \quad (3) \\ \text{s.t. } (1d) - (1i). \end{aligned}$$

The problem in (3) can be solved using any NLP solver. Since the partial Lagrangian has less constraints than the primal problem, its feasible solution space is larger. Therefore, the problem in (3) is feasible if the original ACOF problem is feasible.

We solve the problem in (3) *starting from the same initial point* ($\mathbf{V}_{\text{init}}, \theta_{\text{init}}$) that was used to solve the original primal problem in (1). Denote this solution to (3) by $(\bar{\mathbf{V}}, \bar{\theta})$. Note $(\bar{\mathbf{V}}, \bar{\theta})$ will not be the same as $(\mathbf{V}_{\text{init}}, \theta_{\text{init}})$ since they come from different problems. Then we start the NLP solver again to solve (1) but with the initial point $(\bar{\mathbf{V}}, \bar{\theta})$. This process can be repeated until the solutions stop changing or up to a predefined number of iterations.

The key observation in this paper is that the solution $(\bar{\mathbf{V}}, \bar{\theta})$ found by solving the partial Lagrangian is often a much better starting point than the original choice of $(\mathbf{V}_{\text{init}}, \theta_{\text{init}})$. Therefore, by repeating these steps, we can iteratively improve the solution quality (i.e., reducing the cost). The algorithm is summarized below as Algorithm 1. We illustrate the intuition behind this algorithm in the next section using 2-bus and 3-bus networks. Formal proofs are given in Section 5, and simulations results for larger IEEE benchmarks are presented in Section 6.

Algorithm 1: Solving ACOF iteratively

Inputs: $\theta_{\text{init}}^{(i)}, \mathbf{V}_{\text{init}}^{(i)}, i = 0$

- 1: At i -th iteration: Initialized at $\theta_{\text{init}}^{(i)}, \mathbf{V}_{\text{init}}^{(i)}$.
 - 2: Call NLP solver for (1), record $(\bar{\mu}_{(i)}^P, \bar{\mu}_{(i)}^Q)$.
 - 3: Given $(\bar{\mu}_{(i)}^P, \bar{\mu}_{(i)}^Q)$, call solver for the partial Lagrangian in (3), record the solutions as $(\bar{\theta}^{(i)}, \bar{\mathbf{V}}^{(i)})$.
 - 4: Call IPOPT for (1) initialized at $(\bar{\theta}^{(i)}, \bar{\mathbf{V}}^{(i)})$, record solutions $(\hat{\theta}^{(i)}, \hat{\mathbf{V}}^{(i)})$.
 - 5: If the solution from line 4 does not reduce the cost, terminate the algorithm.
 - 6: Otherwise, update initial points:
 $\theta_{\text{init}}^{(i+1)} = \hat{\theta}^{(i)}, \mathbf{V}_{\text{init}}^{(i+1)} = \hat{\mathbf{V}}^{(i)}$.
 - 7: Repeat 1-6 until the maximum number of iterations is reached.
-

In terms of computational overhead, each iteration of Algorithm 1 solves an ACOF problem twice and an OPF-like problem (minimizing the partial Lagrangian) once. In practice, we observe that the cost is reduced after every iteration and the global solution can be reached in a small number of iterations (for the cases where the global solution is known). Therefore, in contrast to algorithms that resolve the ACOF problem from a large number of random initialization points [8], Algorithm 1 is much more computationally efficient.

4 GEOMETRY AND INTUITION

In this section, we study the geometry of the ACOF problem to shed some light on why Algorithm 1 might be successful. We find that the main reason is that the optimization landscape of the partial Lagrangian is much “better” than the landscape of the original problem. To illustrate this geometric property, we use the 2-bus and 3-bus networks as examples. The formal proofs are provided in Section 5.

4.1 2-bus network

In this part, we consider a 2-bus network. For simplicity, we ignore the reactive power and set both voltage magnitudes to 1 p.u.. Suppose bus 1 is a generator and also the reference (slack) bus with an increasing cost function $c(\cdot)$, and bus 2 is the load bus with angle $-\theta$. The line admittance is $g - jb$. Given a load of l at bus 2 and ignoring all constraints except for the load balancing one, the ACOF in (1) becomes

$$\min_{\theta} c(g - g \cos(\theta) + b \sin(\theta)) \quad (4a)$$

$$\text{s.t. } l + g - g \cos(\theta) - b \sin(\theta) = 0. \quad (4b)$$

This is an example of an OPF with a disconnected feasible space, since there are two discrete solutions to (4b) and we are asking for the one with lower cost.

To see how a NLP solver would approach this problem, we adopt a common practice [4, 23] and form a penalized version of (4).¹ The penalized unconstrained problem is given by

$$\begin{aligned} \mathcal{L}_{\rho} = & c(g - g \cos(\theta) + b \sin(\theta)) \\ & + \rho/2(l + g - g \cos(\theta) - b \sin(\theta))^2, \end{aligned} \quad (5)$$

where ρ is a penalty parameter. For large enough ρ , the solutions of (5) would coincide with those of (4) [4]. The function $\mathcal{L}_{\rho}$ is plotted in Fig. 2 (green line). We can see that there are two local minimas, with the left one being global. The strict local minimum (the right one) satisfies both first and second order optimality conditions. Therefore, if we initialize an NLP solver with a bad starting point, it would be stuck at the strict local solution. For this example, if the initial point is to the left of the maximum of the green curve, a solver would converge to the left solution; and if the initial point is to the right, a solver would find the right (suboptimal) solution. Hence, a flat start would lead to the global solution. However, for larger systems, flat starts are often not successful (e.g., see the 22-bus system in Section 6). Therefore, this 2-bus example is useful as it illustrates the geometry of the optimization landscape.

Now suppose μ is the multiplier corresponding to the equality constraint (4b) at the *strict local solution*. The partial Lagrangian of (4) by dualizing (4b) is:

$$\begin{aligned} \mathcal{L}_{\mu} = & c(g - g \cos(\theta) + b \sin(\theta)) \\ & + \mu(l + g - g \cos(\theta) - b \sin(\theta)). \end{aligned} \quad (6)$$

Since the sinusoidal functions are periodic with period 2π , let us consider the range $\theta \in [-\pi, \pi]$. It is interesting now to compare the solution of $\mathcal{L}_{\mu}$ and the original problem in (4) (or equivalently, $\mathcal{L}_{\rho}$). The blue curve in Fig. 2 plots $\mathcal{L}_{\mu}$. We observe two interesting facts. The first is that unlike $\mathcal{L}_{\rho}$, $\mathcal{L}_{\mu}$ in this 2-bus network only has a single minimum. Therefore, no matter where we initialize the NLP solver for $\mathcal{L}_{\mu}$, we would reach this minimum. The second fact is that the minimum of $\mathcal{L}_{\mu}$ is close to the global minimum of $\mathcal{L}_{\rho}$. Therefore, if we start a NLP solver for the ACOF at the solution of $\mathcal{L}_{\mu}$, we would reach the global solution. Interestingly, we are using the multiplier at the strict local solution. So even if a solution is not

¹We use quadratic penalties only as an analysis method that allows us to obtain cleaner theoretical results. This approach is standard in convergence analysis of nonlinear programming, for example, see [4] (Chap. 1), [23], and [27]. All simulation and numerical results in the paper are obtained using state-of-the-art solvers rather than quadratic penalty method.

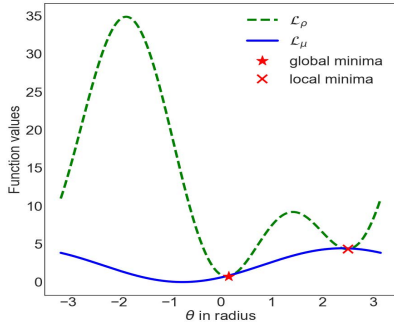


Figure 2: Geometry of the penalized objective functions $\mathcal{L}_\rho$ and the partial Lagrangian $\mathcal{L}_\mu$. The line admittance is $1 - j4$ and the penalty parameter is 2.

global, it is still very useful, since by solving $\mathcal{L}_\mu$ as an intermediate step, we would not be stuck at the strict local solution. We prove that this procedure is guaranteed to work for tree networks in the next section.

4.2 3-bus network

Now, let us show that the intuitions built in the 2-bus example still carry over into the 3-bus network. We again ignore the reactive power and set all voltage magnitudes to 1 p.u. to optimize over the angles. Suppose bus 1 is a generator and also the reference bus with an increasing cost function $c(\cdot)$, while bus 2 and bus 3 are load buses with angles $-\theta_2$ and $-\theta_3$, respectively. The load at bus 2 is l_2 and at bus 3 is l_3 . Then the ACOF in (1) can be simplified to

$$\min_{\theta_2, \theta_3} c\left(\sum_{j=2,3} g_{1j} - g_{1j} \cos(\theta_j) + b_{1j} \sin(\theta_j)\right) \quad (7a)$$

s.t.

$$l_2 + \sum_{j=1,3} (g_{2j} - g_{2j} \cos(\theta_{2j}) - b_{2j} \sin(\theta_{2j})) = 0 \quad (7b)$$

$$l_3 + \sum_{j=1,2} (g_{3j} - g_{3j} \cos(\theta_{3j}) - b_{3j} \sin(\theta_{3j})) = 0. \quad (7c)$$

As in the 2-bus case, to understand how a NLP solver may approach (7), we form its penalized version:

$$\mathcal{L}_\rho = c\left(\sum_{j=2,3} g_{1j} - g_{1j} \cos(\theta_j) + b_{1j} \sin(\theta_j)\right) \quad (8)$$

$$+ \frac{\rho}{2} \left(l_2 + \sum_{j=1,3} (g_{2j} - g_{2j} \cos(\theta_{2j}) - b_{2j} \sin(\theta_{2j})) \right)^2$$

$$+ \frac{\rho}{2} \left(l_3 + \sum_{j=1,2} (g_{3j} - g_{3j} \cos(\theta_{3j}) - b_{3j} \sin(\theta_{3j})) \right)^2.$$

It turns out that there are four local solutions (one of which is global) for (8).² All of these solutions satisfy both first order and second order conditions and they are listed in Table 1. At these solutions, the gradients $\nabla \mathcal{L}_\rho$ are zero and the Hessians $\nabla^2 \mathcal{L}_\rho$ are

²They are found via a grid search, i.e., we finely discretize the space and exhaustively check all points.

positive definite. This makes $\mathcal{L}_\rho$ look like valleys (convex) at all of the minimas. So it can be hard for an NLP solver to get out of being trapped at a strict local minima. The level sets around the solutions of $\mathcal{L}_\rho$ are plotted on the left of Fig. 3, where there is little difference between the local and the global minima.

Solution	Bus 2	Bus 3	Hessian matrix of $\mathcal{L}_\mu$
1st (global)	$\angle 0.52$	$\angle 0.52$	Positive definite
2nd	$\angle 0.7$	$\angle 2.2$	Indefinite
3rd	$\angle 2.2$	$\angle 0.7$	Indefinite
4th	$\angle 2.09$	$\angle 2.09$	Negative definite

Table 1: The four solutions to problem (7) through grid search. The Hessian of $\mathcal{L}_\rho$ is positive definite at all solutions. The definiteness of the Hessian of $\mathcal{L}_\mu$ is listed. The parameters are $g_{12} - jb_{12} = g_{13} - jb_{13} = 1 - j4$ and $g_{23} - jb_{23} = 0.1 - j0.4$.

Now we show that a partial Lagrangian behaves qualitatively differently. Suppose that we choose a strict local solution of (7). Let the multipliers corresponding to the equality constraints (7b) and (7c) be μ_1 and μ_2 , respectively. The partial Lagrangian for (7) is:

$$\mathcal{L}_\mu = c\left(\sum_{j=2,3} g_{1j} - g_{1j} \cos(\theta_j) + b_{1j} \sin(\theta_j)\right) \quad (9a)$$

$$+ \mu_1 \left(l_2 + \sum_{j=1,3} (g_{2j} - g_{2j} \cos(\theta_{2j}) - b_{2j} \sin(\theta_{2j})) \right) \quad (9b)$$

$$+ \mu_2 \left(l_3 + \sum_{j=1,2} (g_{3j} - g_{3j} \cos(\theta_{3j}) - b_{3j} \sin(\theta_{3j})) \right). \quad (9c)$$

In contrast to the penalized problem, there is only a single solution for $\mathcal{L}_\mu$ which satisfies both the first order and second order optimality conditions. It is close to the global solution of $\mathcal{L}_\rho$ (the black dot in Fig. 3(b)), even though the multipliers used in forming $\mathcal{L}_\mu$ are from a strict local solution.

If we look at the Hessian of $\mathcal{L}_\mu$, we see that the Hessian is either negative definite or indefinite at the strict local solutions of $\mathcal{L}_\rho$ (the definiteness of the Hessians for $\mathcal{L}_\mu$ at the local solutions of $\mathcal{L}_\rho$ are listed in Table 1). If the Hessian is not positive semidefinite, then there is always a direction to lower the objective value of a function. For example, these descent directions are shown as dotted arrows in Fig. 3(d) and Fig. 3(f).

All together, Fig. 3 shows how Algorithm 1 would get around the strict local solutions in $\mathcal{L}_\rho$. Suppose we solve the Lagrangian from a point around the global minimum $\mathbf{x}^*$. Since $\nabla^2 \mathcal{L}_\mu(\mathbf{x}^*)$ is positive definite, this means the starting point is at a valley of the Lagrangian surface. So solving the Lagrangian would return the global solution. Now let us use a point around the local solution, say $\bar{\mathbf{x}}$, as an initial point to solve the Lagrangian. As shown in Fig. 3(d) and 3(f), $\nabla^2 \mathcal{L}_\mu(\bar{\mathbf{x}})$ is negative definite or indefinite, so the surface of the Lagrangian is concave down or has a saddle. Then we can find at least one descent direction to get out of being trapped at the current point.

5 ANALYSIS OF ALGORITHM 1

In this section, we provide a rigorous analysis of Algorithm 1. As a first step, we focus our attention on systems with a tree topology and ignore the reactive power. For mesh networks with both active

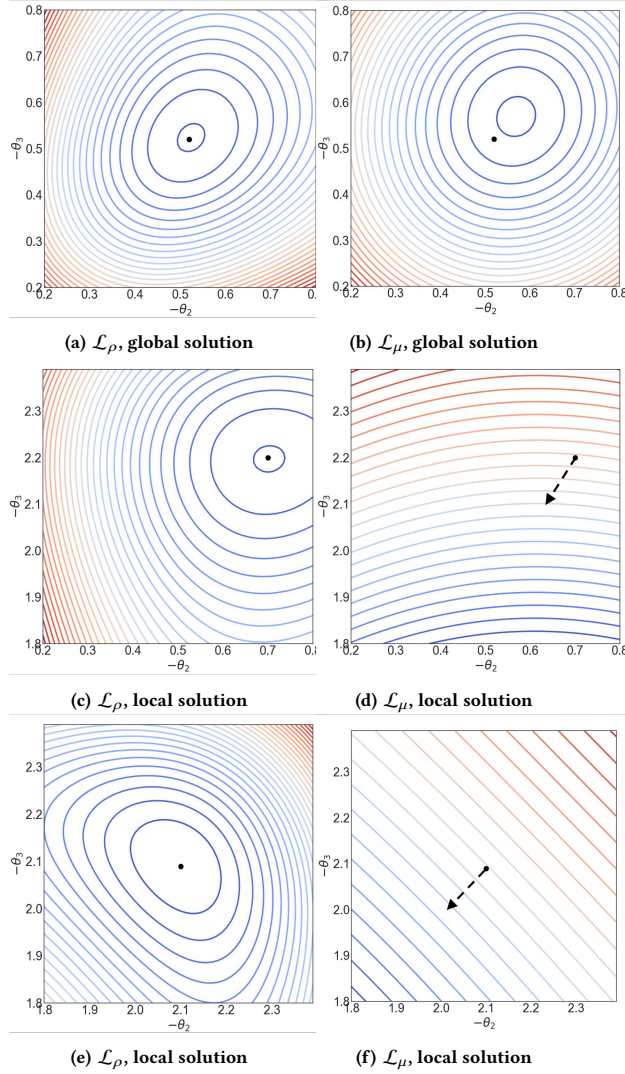


Figure 3: The contour plot of $\mathcal{L}_\rho$ and $\mathcal{L}_\mu$ nearby the 1st, 2nd and 4th solution. The Hessian matrix of $\mathcal{L}_\mu$ is positive definite in (b), indefinite in (d), and negative definite in (f). The black arrows in (d) and (f) indicates the descent directions of the function value.

and reactive power flows, we provide detailed simulation studies in Section 6 and show how the intuition from tree networks applies. Formal proofs for meshed systems is an important part of our future work.

We first consider a tree network with fixed voltage magnitudes and show that the minimizer of the Lagrangian falls into the attraction basin of the global minimum of the ACOPF problem, which generalizes the observations in Section 4.1. Then we optimize over both voltage magnitudes and angles for a 2-bus network, and look at the Hessian matrix of the Lagrangian as we do in Section 4.2. We prove that the Hessian matrix of the Lagrangian is positive definite at the global minimum and negative definite or indefinite at the local minimum.

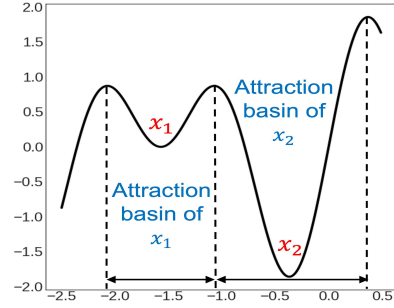


Figure 4: Illustrative figure of Definition 1. There are two local minimums, i.e., x_1 and x_2 . The attraction basins of each of them are marked as intervals and annotated with blue fonts.

5.1 Fixed voltage magnitudes

In this part, we consider a tree network with fixed voltage magnitudes. We note that the NLP solver we use in this paper, IPOPT, uses a barrier function to solve a sequence of unconstrained optimization problems using a mixture of gradient descent and Newton-type methods (with many different ways of tuning stepsizes). Without loss of generality, we assume that the NLP solver runs either a gradient descent or a Newton-type algorithm. For either algorithm, there is a theorem called the Capture Theorem (see [4], Prop. 1.2.4 for gradient-like algorithms and Prop. 1.4.1 for Newton-type algorithms) saying that once the algorithm enters the region of attraction around a local minimum it has to go to this local minimum. This means starting the solver from an initial point in the region of attraction of a solution would return this solution. Formally, the region of attraction of a solution is defined as follows [4]:

Definition 1. Let $\mathbf{x}^*$ be an unconstrained local minimum to $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Assume there exists a set X such that $f(\mathbf{x})$ is continuously differentiable on X and $\mathbf{x}^* \in X$. For every point $\mathbf{x} \neq \mathbf{x}^*$ and $\mathbf{x} \in X$, if the following inequality holds, then X is a subset of the region of attraction around $\mathbf{x}^*$:

$$\nabla f(\mathbf{x})^T (\mathbf{x}^* - \mathbf{x}) < 0, \forall \mathbf{x} \neq \mathbf{x}^*, \mathbf{x} \in X, \quad (10)$$

where $\nabla f(\mathbf{x})$ represents the gradient of $f(\cdot)$ at the point $\mathbf{x}$.

An illustrative figure of Definition 1 is given in Fig. 4. Intuitively, the inequality in (10) implies that the direction where the function value decreases (the descent direction) is aligned with the negative gradient. Now we give the following theorem about the performance of Algorithm 1 for a tree network with fixed voltage magnitudes:

THEOREM 2. Consider a N -bus radial network and keep the voltage magnitudes fixed to optimize over voltage angles. If the NLP solver gets stuck at a strictly local solution when it starts from an initialization point, then starting Algorithm 1 from the same initialization point is able to escape from this strictly local solution.

PROOF. Since Algorithm 1 uses the solution of the partial Lagrangian as a new initialization point to solve the primal problem again, we prove Theorem 2 by showing that the minimizer of the

partial Lagrangian falls into the region of attraction around the global minimum of the ACOF problem. As we assumed at the beginning of this part, the NLP solver is running either a gradient descent or a Newton-type algorithm. Both of the two follow the Capture Theorem in [4]. As a result, starting the solver from the minimizer of the Lagrangian would be able to find the global minimum.

We do the proof by induction starting with a 2-bus network. The ACOF problem for the 2-bus network is given in (4) and its Lagrangian is given in (6). We first study the solutions to (4) by looking at the equality constraint $h(\theta) = l + g - g \cos(\theta) - b \sin(\theta) = 0$. Its gradient can be written as

$$h'(\theta) = g \cos(\theta)(\tan(\theta) - b/g).$$

Suppose $\theta \in (-\pi/2, 3\pi/2)$, then the gradient h' is zero at $\theta = \tan^{-1}(b/g)$. We also have

$$h'(\theta) < 0, \forall \theta \in (-\pi/2, \tan^{-1}(b/g)) \quad (11a)$$

$$h'(\theta) > 0, \forall \theta \in (\tan^{-1}(b/g), \tan^{-1}(b/g) + \pi). \quad (11b)$$

This implies that $\theta = \tan^{-1}(b/g)$ is a minima of $h(\theta)$. Since for a feasible problem, the solution to $h(\theta) = 0$ must exist within $(-\pi/2, 3\pi/2)$, then by the intermediate value theorem, there are two solutions to (4), which satisfy the following inequalities:

$$-\pi/2 < \theta^* < \tan^{-1}(b/g) < \bar{\theta} < 3\pi/2, \quad (12)$$

where θ^* is the global minimum and $\bar{\theta}$ is the local minimum (see Appendix A for more details).

Now we use (10) to show that the interval $(-\pi/2, \tan^{-1}(b/g))$ is a subset of the attraction region of θ^* . For a sufficiently large penalty, the globally optimal solution θ^* can be very close to the global minimum of the unconstrained penalized problem in (5). Therefore, this is equivalent to showing

$$\mathcal{L}'_\rho(\theta)^T(\theta^* - \theta) < 0, \forall \theta \in (-\pi/2, \tan^{-1}(b/g)). \quad (13)$$

As ρ is sufficiently large, the sign of $\mathcal{L}'_\rho(\theta)$ is dominated by the gradient of the second term in (5), i.e.,

$$\begin{aligned} \mathcal{L}'_\rho(\theta) &\approx \rho(l + g - g \cos(\theta) - b \sin(\theta))(g \sin(\theta) - b \cos(\theta)) \\ &= \rho h(\theta) h'(\theta). \end{aligned}$$

For any $\theta \in (-\pi/2, \tan^{-1}(b/g))$, we have $h'(\theta) < 0$ from (11a), which means the function $h(\theta)$ is decreasing on the interval $(-\pi/2, \tan^{-1}(b/g))$. Also, the global minimum θ^* must satisfy $h(\theta^*) = 0$. Therefore we have

$$h(\theta) > 0, \forall \theta \in (-\pi/2, \theta^*)$$

$$h(\theta) < 0, \forall \theta \in (\theta^*, \tan^{-1}(b/g)).$$

Then the inequality in (13) follows from above. Therefore, by Definition 1, the interval $(-\pi/2, \tan^{-1}(b/g))$ is a subset of the attraction region of θ^* .

To obtain the minimizer of $\mathcal{L}_\mu$, we write out the optimality condition of (6) for the primal-dual optimal solution $(\hat{\theta}, \hat{\mu})$:

$$(c' + \hat{\mu})g \sin(\hat{\theta}) + (c' - \hat{\mu})b \cos(\hat{\theta}) = 0, \quad (16)$$

where c' is a shorthand for $c'(g - g \cos(\hat{\theta}) + b \sin(\hat{\theta}))$ and is the gradient of the cost function. Suppose $\hat{\theta} \in (-\pi/2, 3\pi/2)$, then $\hat{\theta}$

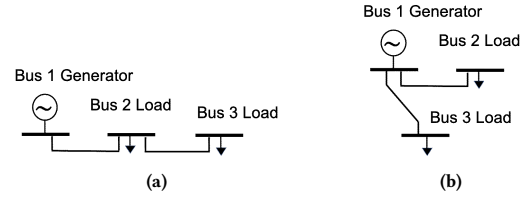


Figure 5: The two types of three bus networks with the tree structure.

solves

$$\tan^{-1}\left(\frac{\hat{\mu} - c'}{\hat{\mu} + c'} b/g\right) + k\pi, \quad k = 0, 1, \quad (17)$$

where the smaller value is the minimum of $\mathcal{L}_\mu$ and the larger one is the maximum (see Appendix B for more details). Let $\hat{\theta}$ be the minimum, which satisfies $-\pi/2 < \hat{\theta} < \tan^{-1}(b/g)$. Since the interval $(-\pi/2, \tan^{-1}(b/g))$ is a subset of the attraction region of θ^* , no matter what initial point we start Algorithm 1 from, solving the Lagrangian gives us a solution lying in the attraction region around the global minimum. Following from the Capture Theorem in [4], Algorithm 1 is able to get out of a strict local solution and reach the global minimum.

Now we induct from 2-bus to 3-bus networks. There are two types of tree topology for a 3-bus network, which are shown in Fig. 5. Since the topology in Fig. 5(b) is equivalent to two 2-bus networks, we focus on the 3-bus branch in Fig. 5(a), where bus 1 is the reference bus.

The ACOF problem for Fig. 5(a) can be written as follows:

$$\min_{\theta_{12}, \theta_{23}} c(P_1^G) \quad (18a)$$

$$\text{s.t. } l_2 + P_{21}^f + P_{23}^f = 0 \quad (18b)$$

$$l_3 + P_{32}^f = 0. \quad (18c)$$

Note that only bus 1 generates power and to deliver the power to bus 3 the power has to be delivered to bus 2 first. Since the cost function $c(\cdot)$ is increasing, given the load at bus 2 and bus 3, minimizing the power generation cost in (1) is equivalent to minimizing the power transfer cost on both lines. Therefore, (18) can be decomposed into two parts, and each part is nothing but solving the ACOF for a 2-bus network with voltage fixed and ignoring reactive power, i.e., the formulation in (4). To make this clear, we first rewrite (18) as follows:

$$\min_{\theta_{12}, \theta_{23}} c_1(P_{12}) + c_2(P_{23}) \quad (19a)$$

$$\text{s.t. } l_2 + P_{21} + P_{23} = 0 \quad (19b)$$

$$l_3 + P_{32} = 0. \quad (19c)$$

With l_2 fixed, for every given θ_{23} , we can always pick some θ_{12} to satisfy (19b). Therefore, we can regard θ_{23} as the optimization variable first. Then the problem (19) is reduced to

$$\begin{aligned} \min_{\theta_{23}} \tilde{c}(P_{23}) \\ \text{s.t. } l_3 + P_{32} = 0 \end{aligned}$$

where $\tilde{c}(\cdot)$ is some increasing cost function that takes into account the effect of θ_{23} on P_{12} . This problem has exactly the same formulation as the 2-bus network in (4). As we proved for the 2-bus

network, if we start Algorithm 1 from a point where θ_{23} is at the local minimum, then we still can get out of this local minimum.

Now we optimize θ_{12} for a given θ_{23} , then we can add P_{23} to l_2 . Therefore, (19) is reduced to

$$\begin{aligned} \min_{\theta_{12}} & c_1(P_{12}) \\ \text{s.t. } & \tilde{l}_2 + P_{21} = 0 \end{aligned}$$

where $\tilde{l}_2 = l_2 + P_{23}$. This problem also has the same formulation as the 2-bus network in (4). Therefore, if the initial point is a local minimum for θ_{12} , Algorithm 1 still can get out of this local minimum.

Let us assume Theorem 2 holds for a $(N-1)$ -bus radial network and consider a N -bus radial network. Similar to the proof for a 3-bus network, we can reduce the ACOPF problem for a N -bus network to the case of $(N-1)$ -bus. So by induction, Theorem 2 holds for the N -bus radial network. $\square$

5.2 Optimizing both voltage magnitudes and angles

In this part, we optimize both voltage magnitudes and angles for a 2-bus network. For simplicity, we ignore the reactive power. Suppose bus 1 is a generator and also the reference (slack) bus with linear cost \$1/MW, and bus 2 is the load bus with load l . The ACOPF in (1) can be simplified as

$$\min_{\theta, V_1, V_2} gV_1^2 - V_1V_2(g \cos(\theta) - b \sin(\theta)) \quad (22a)$$

$$\text{s.t. } l + V_2^2g - V_1V_2(g \cos(\theta) + b \sin(\theta)) = 0 \quad (22b)$$

$$V_{\min} \leq V_1, V_2 \leq V_{\max}. \quad (22c)$$

Let us collect all the variables into the vector $\mathbf{x} = (\theta, V_1, V_2)$. We denote the objective function by $f(\mathbf{x})$, and the equality constraint (22b) by $h(\mathbf{x}) = 0$. The following theorem gives the property of the Hessian matrix of the Lagrangian.

THEOREM 3. *Denote the global solution of (22) as $\mathbf{x}^*$ and the local solution as $\bar{\mathbf{x}}$. Then the Hessian matrix of a Lagrangian of (22), formed with multipliers at any of the local solutions, is positive definite at $\mathbf{x}^*$ and negative definite or indefinite at $\bar{\mathbf{x}}$.*

PROOF. To study the solution to (22), we look at the equality constraint (22b) directly. Its gradient with respect to θ can be written as

$$\partial h / \partial \theta = g \sin(\theta) - b \cos(\theta) = g \cos \theta (\tan(\theta) - b/g).$$

Suppose $\theta \in (-\pi/2, 3\pi/2)$, then $\partial h / \partial \theta$ is zero at $\tan^{-1}(b/g) + k\pi$, $k = 0, 1$, where the smaller value is located at the global minimum and the larger value is at the local minimum. Denote the global minimum as θ^* and at the local minimum as $\bar{\theta}$. They satisfy (see Appendix A for the details):

$$-\pi/2 < \theta^* < \tan^{-1}(b/g) < \bar{\theta} < 3\pi/2. \quad (23)$$

In Appendix C, we show that at least one of V_1 and V_2 need to be binding at a constraint, but both voltages cannot be binding at the same time. This allows us to consider the cases where V_1 is binding or V_2 is binding separately.

First, suppose V_1 is inactive and V_2 is binding. In this case, V_2 is a constant and the Lagrangian of (22) can be written as

$$\mathcal{L}_{\lambda, \mu} = f(\mathbf{x}) + \mu h(\mathbf{x}) + \bar{\lambda}_1 (V_1 - V_{\max}) + \lambda_1 (-V_1 + V_{\min}).$$

The multipliers are associated with some local solution, and μ is the Lagrange multiplier related to the equality constraint, and $\bar{\lambda}_1$ and λ_1 are the multipliers related to the inequality constraints of V_1 .

Denote the Hessian matrix of $\mathcal{L}_{\lambda, \mu}$ as $\nabla^2 \mathcal{L}_{\lambda, \mu}(\mathbf{x})$. To determine its definiteness, we write out all leading principal minors at a solution $\tilde{\mathbf{x}}$ (see Appendix D for the details):

$$D_1(\tilde{\mathbf{x}}) = \tilde{V}_1 \tilde{V}_2 \frac{-2gb}{g \cos(\tilde{\theta})(\tan(\tilde{\theta}) - b/g)} \quad (24a)$$

$$D_2(\tilde{\mathbf{x}}) = 2gD_1(\tilde{\mathbf{x}}). \quad (24b)$$

Following from the inequalities in (23), both leading principal minors in (24) are positive at the global minimum and negative at the local minimum. This means the Hessian matrix at $\mathbf{x}^*$ is positive definite. In contrast, the Hessian matrix at $\bar{\mathbf{x}}$ is negative definite.

Now we suppose V_2 is inactive and V_1 is binding. In this case, V_1 is a constant and the Lagrangian is:

$$\mathcal{L}_{\lambda, \mu} = f(\mathbf{x}) + \mu h(\mathbf{x}) + \bar{\lambda}_2 (V_2 - V_{\max}) + \lambda_2 (-V_2 + V_{\min}).$$

where the multipliers are associated with some local solution. Let us denote the Hessian matrix of the Lagrangian as $\nabla^2 \mathcal{L}_{\lambda, \mu}(\mathbf{x})$. Its leading principal minors at a feasible solution $\tilde{\mathbf{x}}$ are (see Appendix D for the details):

$$D_1(\tilde{\mathbf{x}}) = \tilde{V}_1 \tilde{V}_2 \frac{-2gb}{g \cos(\tilde{\theta})(\tan(\tilde{\theta}) - b/g)} \quad (25a)$$

$$\tilde{D}_2(\tilde{\mathbf{x}}) = 2g\tilde{\mu}D_1(\tilde{\mathbf{x}}). \quad (25b)$$

Since the multiplier μ represents the marginal price of consuming each additional unit of load, it is positive at the global minimum. This means $D_2(\tilde{\mathbf{x}})$ has the same sign as $D_1(\tilde{\mathbf{x}})$. For the global minimum $\mathbf{x}^*$, $D_1(\tilde{\mathbf{x}})$ is positive from (23), hence both leading principal minors in (25) are positive and the Hessian matrix is positive definite at $\mathbf{x}^*$. In contrast, at the local minimum $\bar{\mathbf{x}}$, $D_1(\tilde{\mathbf{x}})$ is negative following from (23). Then the Hessian matrix is either negative definite or indefinite at $\bar{\mathbf{x}}$. $\square$

The simulation results in the next section do not need to make any of the assumptions in Theorem 2 and 3. They are about mesh networks with all constraints included. Therefore, we suspect the theory can be made much stronger and would extend to larger meshed networks. However, analyzing these cases is challenging and is a future direction for us.

6 SIMULATION RESULTS

In this section we report the simulation results to validate the effectiveness of our algorithm. The NLP solver used here is IPOPT [29] and the convergence tolerance is set to 0.0001. It returns a feasible solution, which may or may not be a global optimum. We test our algorithm on IEEE meshed networks with 3, 9, 22, 39, 118 and 300 buses, and also on the IEEE radial network with 141 buses. For the 3-bus, 9-bus, 22-bus, 118-bus and 300-bus networks, the local and global solutions are known and listed in [5, 24]. We use the strict local solutions as starting points for the solver to demonstrate

the ability of Algorithm 1 of getting out of local solutions. For the 39-bus and 141-bus networks, we do an exhaustive search by discretizing each variable within their bounds to find the global solution. The simulation results show that for the 3, 9, 22, 118, 300 and 141-bus networks, Algorithm 1 finds the globally optimal solution in 1 iteration. For the 39-bus networks, it takes at most 3 iterations for Algorithm 1 to obtain the optimal solution.

6.1 3-Bus Mesh Network

The 3-bus network we use is shown in Fig. 5a and the voltage bounds are $[0.95, 1.05]$. Two solutions exist and they are listed in Table 2. This was an example used in [24] to show that multiple reasonably looking local solutions can exist, and contrary to conventional wisdom, the higher voltage one is the suboptimal one (although the cost difference is small).

If we start the nonlinear solver from an initial point near the second solution, then the solver cannot get out of the attraction basin and always returns the second solution. In contrast, if we launch Algorithm 1 using the second solution as a starting point, then the algorithm converges to the first solution (the global solution) after one iteration. Although the cost difference is small between the two solutions, larger networks will have bigger cost differences.

	Bus 1	Bus 2	Bus 3	Cost
Solution 1	$0.95 \angle 0$	$0.95 \angle -0.48$	$0.98 \angle -0.53$	1
Solution 2	$0.95 \angle 0$	$1.01 \angle -0.46$	$1.05 \angle -0.51$	1.0021

Table 2: The two local solutions for the 3-bus network in Fig. 5a. The cost is normalized to 1 for the global solution.

6.2 9-Bus Mesh Network

In the 9-bus network, there are 3 generators (bus 1, 2 and 3) and 9 transmission lines. The voltage bounds are $[0.9, 1.1]$. Four solutions exist. The cost of the worst local solution is 38% more than the cost at the global solution. We also find that the solutions at generators 2 and 3 and load buses 6, 7, and 8 are important to improve the cost. The power transfer along the lines between these buses tend to get stuck at a suboptimal solution, which leads to a cost more than 30% higher than the lowest one. For the nonlinear solver, we need to relaunch it using different initial points in order for these five nodes to get around the attraction basin. This requires many trials. In contrast, Algorithm 1 only requires one iteration to achieve the global solution, even starting from the local solution with the highest cost.

6.3 22-Bus Mesh Network

In the 22-bus network, the buses are connected in a loop. There are 11 generators and 22 transmission lines. The voltage bounds are $[0.95, 1.05]$. There exist two solutions, and the cost of the local solution is 30% higher than that of the global solution. The two solutions are quite different. We pick 5 buses that are evenly spaced and list their solutions in Table 3. Since the two solutions are very different, it is hard for a nonlinear solver to get around the local solution.

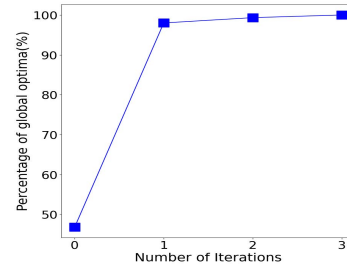


Figure 6: Percentage of globally optimal solutions for the 39-bus network after each iteration. Using a set of random starting points, 47% of them leads to the global optima after a direct call to IPOPT. The fraction of global optimal solutions increases to 98%, 99.93% and 100% after running one, two and three iterations of algorithm 1, respectively.

Particularly, if we initialize the solver with a flat start, we obtain the strict local solution. Furthermore, we generate 100 random points uniformly at random within the bounds of each variable. If these points are used to initiate the nonlinear solver, the local solution is always obtained and the global one cannot be reached. In comparison, Algorithm 1 can achieve the global solution after one iteration regardless of the initial point. This is an example where using random search is very computationally inefficient, and our deterministic algorithm turns out to be much more successful.

6.4 39-Bus Mesh Network

In the 39 bus network, there are 10 generators and 46 transmission lines. The voltage bounds are $[0.95, 1.05]$. Unlike the previous smaller networks, the number and the cost of the solutions are not previously known for this network. Therefore we conducted an exhaustive search to find the global solution. To evaluate the effectiveness of Algorithm 1, we choose 600 random points within the bounds of each variable using the uniform distribution. Then we start Algorithm 1 with these random points to observe the improvement of the solution quality.

In Fig. 6, we plot the fraction of global solutions in the set of all 600 results after each iteration. The x-axis represents the number of iterations that Algorithm 1 is ran, and y-axis represents the percentage of globally optimal solutions after each iteration. When we make a direct call to the solver, less than half of the solutions are globally optimal. One application of Algorithm 1 increases the percentage of globally optimal solutions to 98%. After two iterations, only four cases are not globally optimal. When we run Algorithm 1 for three iterations, all solutions are globally optimal.

We also calculate the average cost of the 600 solutions after each iteration of Algorithm 1 and plot the result in Fig. 7. The x-axis is the number of iterations of running Algorithm 1, and y-axis represents the average cost of 600 solutions, which is normalized using the optimal cost as the factor. After a direct call to the solver, the average cost is 30% higher than the optimal cost. As Algorithm 1 is ran, the average cost decreases quickly. After one iteration, the average cost is only 1.5% more than the globally optimal cost, and after three iterations all solution are at the global optimum.

	Bus 2	Bus 7	Bus 12	Bus 17	Bus 22	cost
Solution 1	1.0285 $\angle$ -0.045	1.05 $\angle$ 0	1.0285 $\angle$ -0.045	1.05 $\angle$ 0	1.0285 $\angle$ -0.045	1
Solution 2	0.95 $\angle$ -0.339	1.0145 $\angle$ 4.57	0.95 $\angle$ 3.089	1.0145 $\angle$ 1.714	0.95 $\angle$ 0.233	1.306

Table 3: The two solutions for the 22-bus network. We pick five buses and show their voltage and angles. The costs at the two solutions are normalized such that the globally optimal cost is 1.

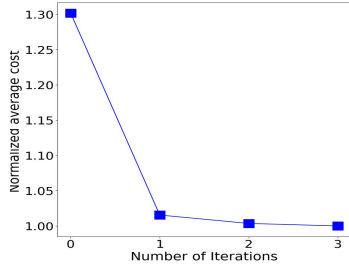


Figure 7: The average cost of all 600 solutions for the 39-bus network after each iteration. The cost is normalized such that the optimal cost is 1. After a direct call to IPOPT, the average cost is 30% higher than the optimal cost. Then the average cost reduces to 1.5%, 0.4% higher than the optimal value after running one and two iterations of Algorithm 1, respectively. After three iterations, the average cost is exactly the optimal cost.

6.5 118-Bus Mesh Network

The topology of the 118-bus meshed network can be found in [5]. There are 54 generators and 186 transmission lines. When the voltage bounds are $[0.94, 1.06]$ and power generation bounds are scaled by 4, two locally optimal solutions are known and listed in [5]. The strictly local minima has a cost that is 28.8% higher than the global minimum.

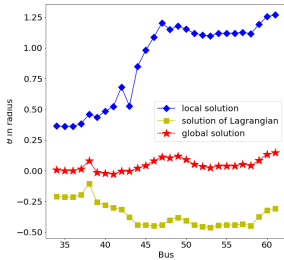


Figure 8: Angles of solutions for buses from 35 to 61.

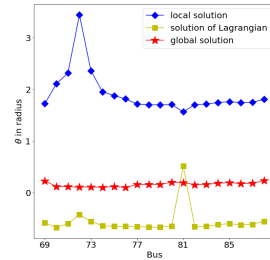


Figure 9: Angles of solutions for buses from 69 to 88.

Demonstration of how the proposed algorithm can escape from the strictly local minima for the 118-bus meshed system. Use two subsets of buses to show the relationship between the local minimum, the solution of the Lagrangian and the global minimum solution (Figures (a) and (b)). The solution to the partial Lagrangian (yellow squares) provides a good warm start for the solver to escape from the strictly local minima. Using the yellow squares as initialization points, the NLP solver is able to find the globally optimal solutions (red stars).

To show how the proposed algorithm can escape from the strictly local minima by using the solution to the partial Lagrangian as

a warm start, we choose two subsets of buses from the 118-bus meshed network and plot the angle of the voltage solutions for each subset. The angle solutions for the subset composed of buses from 34 to 61 are plotted in Fig. 8, and for the subset composed of buses from 69 to 88 plotted in Fig. 9.

When the NLP solver reaches the strictly local solution (marked as blue diamonds) from some initialization point, it terminates at this solution. To escape from this suboptimal solution, we call the solver to solve the partial Lagrangian starting from the same initialization point and get the solution marked as yellow square. We use this solution as the new warm start to solve the primal problem again, then we can obtain the global minima (marked as red stars) with a single run of Algorithm 1. As is shown in Fig. 8 and Fig. 9, the solution to the partial Lagrangian jumps quite far away from the local solution but stay close to the global solution. Therefore, it provides a good initialization point for the NLP solver to escape from the strictly local minima and get to the global minima.

6.6 300-Bus Mesh Network

The topology of the 300-bus meshed network can also be found in [5]. There are 69 generators and 411 transmission lines. When the voltage bounds are $[0.93, 1.07]$ and the reactive power generation lower bounds are tightened to -100 Mvar for all generator buses, two locally optimal solutions are known and listed in [5]. The strictly local minimum has a cost that is 0.62% higher than the global minimum. For the NLP solver to escape from the strictly local solution, only a single run of Algorithm 1 is needed.

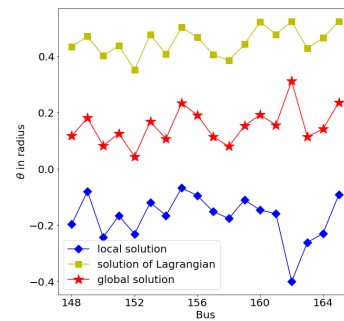


Figure 10: Demonstration of how the proposed algorithm can escape from the strictly local minima for the 300-bus meshed system and plot of angle solutions for the subset of buses from 148 to 166. The solution to the partial Lagrangian (yellow squares) jump far away from the strictly local minima (blue diamonds). As a result, starting from the yellow squares, the NLP solver is able to escape from the strictly local solution and find the globally optimal solution (red stars).

	Bus 1	Bus 30	Bus 50	Bus 80	Bus 121	relative cost
PG(MW) at solution 1	3.2578	5.7791	7.9975	5.0351	9.768	1
PG(MW) at solution 2	3.664	4.7109	8.2127	3.8811	13.8361	1.115
PG(MW) at solution 3	4.3365	5.1223	7.039	4.9725	9.5491	1.026

Table 4: The active power generation at a subset of generator buses at the three optimal solutions for the 141-bus radial network. The costs are normalized such that the global optimal cost is 1.

To show how the proposed algorithm can escape from the strictly local minimum by using the solution to partial Lagrangian as an initialization point, we choose a subset of buses from 148 to 166 and plot the angles of their voltage solutions in Fig. 10. As Fig. 10 shows, the solution to the partial Lagrangian (marked as yellow squares) jumps far away from the strictly local solution and thus act as a good warm start for the NLP solver to escape from the strictly local minima and find the global minima.

6.7 141-Bus Radial Network

The topology of the 141-bus radial network can be found in [18]. There are 9 generators and 140 transmission lines. When the voltage bounds are [0.83, 1.17] and the reactive power generation lower bounds are tightened for buses 2, 7 to 5 Mvar and bus 5 to 15 Mvar, 3 locally optimal solutions are found. The two strictly local minima has costs that is 11.5% and 2.6% higher than the global minimum, respectively. The active power generation at a subset of the generator buses for each of the solutions are shown in Table 4.

To test the performance of our algorithm and demonstrate how it escapes from local minima, we randomly generate 1000 initial points within the bounds of each variable using a uniform distribution, and call the NLP solver to solve the primal AC OPF problem starting from these initial points.

To demonstrate how the solution of partial Lagrangian can be a good warm start, we choose a subset of buses from the 141-bus radial network and plot the angle of the voltage solutions. First, we show how the proposed algorithm escapes from the first local minimum. In Fig. 11, the NLP solver reaches local solution-1 (marked as blue diamonds) from an initialization point. The NLP solver terminates at this solution. To escape from this suboptimal solution, we call the solver to solve the partial Lagrangian starting from the same initialization point and get the solution marked as yellow square. We use this solution as a new warm start to solve the primal problem again, then we can obtain the global solution (marked as red stars) with only a single run of Algorithm 1. As Fig 11 shows, the solution to the partial Lagrangian jumps quite far from the local minimum and provides a good starting point for the NLP solver to reach the global minimum.

In Fig. 12, we show how our proposed algorithm escapes from the second local minimum solution. For an initialization point where the NLP solver reaches local solution-2, we solve the partial Lagrangian. Initializing the NLP solver from the solution of the partial Lagrangian allows us to reach the global solution, and we show the voltage angles of different solutions for buses 60 to 69 in Fig. 12.³

³The reason why we choose different subsets is that the local and global solutions in the subset composed of buses 89 to 110 are very similar. To better illustrate the effectiveness of our method, we choose the subset composed of buses 60 to 69, where the local and global solutions are quite different.

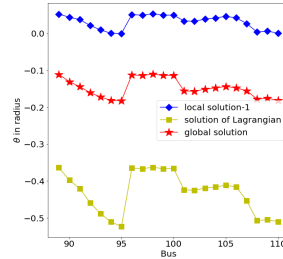


Figure 11: Angles of solutions for buses from 89 to 110.

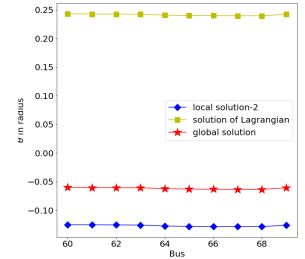


Figure 12: Angles of solutions for buses from 60 to 69.

Demonstration of how the proposed algorithm can escape from the strictly local minima for the 141-bus radial system. Specifically, Figure (a) shows how the solutions escape from the first local minimum and (b) shows how they escape from the second minimum. Using our proposed algorithm, the solution of the partial Lagrangian is able to escape the local minimum (blue diamonds). Using these (yellow squares) as initialization points, the NLP solver is able to find the globally optimal solution (red stars).

7 CONCLUSION

In this paper, we propose a simple algorithm to iteratively improve the solution quality of ACOFF problems. First, we solve the ACOFF problem using an existing nonlinear solver. From the solution and its associated dual variables, we construct a partial Lagrangian by dualizing the power balance equations. Optimizing this partial Lagrangian leads to a new solution. With this solution as an initial point, we again call the solver for the ACOFF problem. By repeating these steps, we can iteratively improve the solution quality, escaping from local solutions to find better ones. We illustrate the intuition behind our algorithm using 2 and 3-bus networks, which shows that the partial Lagrangian has a flatter optimization landscape compared to the original primal problem. We prove that the algorithm is guaranteed to work in tree networks. Theoretical analysis for more general networks is an important part of our future work. We validate the effectiveness of our algorithm on standard 9-bus, 22-bus, 39-bus, 118-bus and 300-bus mesh networks and also on the IEEE 141-bus radial network. Regardless of the initial points, our algorithm always finds the global optimum within at most three iterations.

8 ACKNOWLEDGMENTS

The authors are partially supported by the National Science Foundation through the grant ECCS-1807142.

REFERENCES

- [1] Mohammad A Abido. 2002. Optimal power flow using particle swarm optimization. *International Journal of Electrical Power & Energy Systems* 24, 7 (2002), 563–571.
- [2] Kyri Baker. 2021. Solutions of dc opf are never ac feasible. In *Proceedings of the Twelfth ACM International Conference on Future Energy Systems*. 264–268.
- [3] Anastasios G Bakirtzis, Pandel N Biskas, Christoforos E Zoumas, and Vasilios Petridis. 2002. Optimal power flow by enhanced genetic algorithm. *IEEE Trans. on Power Systems* 17, 2 (2002), 229–236.
- [4] Dimitri P Bertsekas. 1997. Nonlinear programming. *Journal of the Operational Research Society* 48, 3 (1997), 334–334.
- [5] Waqqas A Bukhsh, Andreas Grothey, Ken IM McKinnon, and Paul A Trodden. 2013. Local solutions of the optimal power flow problem. *IEEE Transactions on Power Systems* 28, 4 (2013), 4780–4788.
- [6] Mary B Cain, Richard P O'Neill, Anya Castillo, et al. 2012. History of optimal power flow and formulations. *FERC* 1 (2012), 1–36.
- [7] Florin Capitanescu. 2016. Critical review of recent advances and further developments needed in AC optimal power flow. *Electric Power Systems Research* 136 (2016), 57–68.
- [8] N. Costilla-Enriquez, Y. Weng, and B. Zhang. 2021. Combining Newton-Raphson and Stochastic Gradient Descent for Power Flow Analysis. *IEEE Transactions on Power Systems* 36, 1 (2021), 514–517. <https://doi.org/10.1109/TPWRS.2020.3029449>
- [9] Abhinav Dronamraju, Songyan Li, Qirui Li, Yuting Li, Daniel Tylavsky, Di Shi, and Zhiwei Wang. 2021. Implications of Stahl's theorems to holomorphic embedding Pt. II: Numerical convergence. *CSEE Journal of Power and Energy Systems* 7, 4 (2021), 773–784.
- [10] Ian A Hiskens and Robert J Davy. 2001. Exploring the power flow solution space boundary. *IEEE transactions on power systems* 16, 3 (2001), 389–395.
- [11] Javad Lavaei and Steven H Low. 2011. Zero duality gap in optimal power flow problem. *IEEE Transactions on Power Systems* 27, 1 (2011), 92–107.
- [12] Bernard Lesieutre, Julia Lindberg, Alisha Zachariah, and Nigel Boston. 2019. On the distribution of real-valued solutions to the power flow equations. In *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 165–170.
- [13] Bernard Lesieutre and Dan Wu. 2015. An efficient method to locate all the load flow solutions-revisited. In *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 381–388.
- [14] Songyan Li, Daniel Tylavsky, Di Shi, and Zhiwei Wang. 2021. Implications of Stahl's theorems to holomorphic embedding Pt. I: Theoretical convergence. *CSEE Journal of Power and Energy Systems* 7, 4 (2021), 761–772.
- [15] Julia Lindberg, Alisha Zachariah, Nigel Boston, and Bernard C Lesieutre. 2020. The Distribution of the Number of Real Solutions to the Power Flow Equations. *arXiv preprint arXiv:2010.03069* (2020).
- [16] Steven H Low. 2014. Convex relaxation of optimal power flow—Part I: Formulations and equivalence. *IEEE Transactions on Control of Network Systems* 1, 1 (2014), 15–27.
- [17] Weimin Ma and James S Thorp. 1993. An efficient algorithm to locate all the load flow solutions. *IEEE Transactions on Power Systems* 8, 3 (1993), 1077–1083.
- [18] Robert Mieth and Yuri Dvorkin. 2020. Distribution Electricity Pricing Under Uncertainty. *IEEE Transactions on Power Systems* 35, 3 (2020), 2325–2338. <https://doi.org/10.1109/TPWRS.2019.2954971>
- [19] Daniel K. Molzahn and Ian A. Hiskens. 2019. A Survey of Relaxations and Approximations of the Power Flow Equations. *Foundations and Trends in Electric Energy Systems* 4, 1-2 (2019), 1–221. <https://doi.org/10.1561/31000000012>
- [20] Daniel K Molzahn and Ian A Hiskens. 2019. A survey of relaxations and approximations of the power flow equations. *Now Publishers* (2019).
- [21] JA Momoh, RJ Koessler, MS Bond, B Stott, D Sun, A Papalexopoulos, and P Ristanovic. 1997. Challenges to optimal power flow. *IEEE Transactions on Power systems* 12, 1 (1997), 444–455.
- [22] James A Momoh, Rambabu Adapa, and ME El-Hawary. 1999. A review of selected optimal power flow literature to 1993. I. Nonlinear and quadratic programming approaches. *IEEE transactions on power systems* 14, 1 (1999), 96–104.
- [23] Julie Mulvaney-Kemp, Salar Fattahi, and Javad Lavaei. 2020. Load variation enables escaping poor solutions of time-varying optimal power flow. In *PESGM*.
- [24] Hung D Nguyen and Konstantin S Turitsyn. 2014. Appearance of multiple stable load flow solutions under power flow reversal conditions. In *2014 IEEE PES General Meeting| Conference & Exposition*. IEEE, 1–5.
- [25] HyungSeon Oh. 2019. A Unified and Efficient Approach to Power Flow Analysis. *Energies* 12, 12 (2019), 2425.
- [26] Zhifeng Qiu, Geert Deconinck, and Ronnie Belmans. 2009. A literature survey of optimal power flow problems in the electricity market context. In *IEEE/PES Power Systems Conference and Exposition*. 1–6.
- [27] R Tyrrell Rockafellar. 1993. Lagrange multipliers and optimality. *SIAM review* 35, 2 (1993), 183–238.
- [28] Yujie Tang and Steven Low. 2017. Distributed algorithm for time-varying optimal power flow. In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*. IEEE, 3264–3270.
- [29] Andreas Wächter and Lorenz T Biegler. 2006. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical programming* 106, 1 (2006), 25–57.
- [30] H. Wei, H. Sasaki, J. Kubokawa, and R. Yokoyama. 1998. An interior point nonlinear programming for optimal power flow problems with a novel data structure. *IEEE Transactions on Power Systems* 13, 3 (1998), 870–877. <https://doi.org/10.1109/59.708745>
- [31] Dan Wu, Daniel K Molzahn, Bernard C Lesieutre, and Krishnamurthy Dvijotham. 2017. A deterministic method to identify multiple local extrema for the ac optimal power flow problem. *IEEE Transactions on Power Systems* 33, 1 (2017), 654–668.
- [32] Baosen Zhang and David Tse. 2012. Geometry of injection regions of power networks. *IEEE Transactions on Power Systems* 28, 2 (2012), 788–797.
- [33] Ray Daniel Zimmerman, Carlos Edmundo Murillo-Sánchez, and Robert John Thomas. 2010. MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Transactions on power systems* 26, 1 (2010), 12–19.

A DETERMINE GLOBAL MINIMUM FOR ACOFP

In Section 5.1 and 5.2, we find two solutions to the supply/balance equality constraint, which satisfy the inequalities in (12) or (23). In this part, we give the reason why the smaller solution in (12) (or (23)) is the global minimum and the larger solution is the local minimum.

Let us subtract the power received at the load bus from the generation at the generator, then we have the transmission loss as follows:

$$\begin{aligned} \text{loss} &= g - g \cos(\theta) + b \sin(\theta) - (-g + g \cos(\theta) + b \sin(\theta)) \\ &= 2g(1 - \cos(\theta)). \end{aligned}$$

Due to the periodicity of arctangent function, the larger value $\bar{\theta}$ must be larger than $\pi/2$. Then the loss at θ^* is smaller than the loss at $\bar{\theta}$. So θ^* is an more optimal solution than $\bar{\theta}$. Since there are only two solutions for this example, θ^* must be the global minimum and $\bar{\theta}$ is the strict local minimum.

B DETERMINE GLOBAL MINIMUM FOR THE LAGRANGIAN

In this part, we determine the global minimum of the Lagrangian problem for the 2-bus network, where we fix the voltage magnitudes and optimize over the angles.

Let us denote the two solutions of the Lagrangian problem in (6) as $\hat{\theta}$ and $\bar{\theta}$, and the multipliers associated with them are $\hat{\mu}$ and $\bar{\mu}$, respectively. Then from (17), we have

$$-\frac{\pi}{2} < \hat{\theta} = \tan^{-1}\left(\frac{\hat{\mu} - c' b}{\hat{\mu} + c' g}\right) < \tan^{-1}\left(\frac{b}{g}\right) \quad (26a)$$

$$\frac{\pi}{2} < \bar{\theta} = \tan^{-1}\left(\frac{\bar{\mu} - c' b}{\bar{\mu} + c' g}\right) + \pi < \tan^{-1}\left(\frac{b}{g}\right) + \pi. \quad (26b)$$

Also we can represent the multiplier using θ by rearranging the terms in (16). We take $(\hat{\theta}, \hat{\mu})$ as an example, and $\bar{\mu}$ can be represented using $\bar{\theta}$ in a similar way. The expression of $\hat{\mu}$ in terms of $\hat{\theta}$ is

$$\hat{\mu} = -\frac{g \sin(\hat{\theta}) + b \cos(\hat{\theta})}{g \sin(\hat{\theta}) - b \cos(\hat{\theta})}. \quad (27)$$

Now let us write out the second-order derivative of the Lagrangian function, and plug (27) into it. Then we have:

$$\begin{aligned}\mathcal{L}_{\mu}''(\hat{\theta}) &= (1 + \hat{\mu})g \cos(\hat{\theta}) - (1 - \hat{\mu})b \sin(\hat{\theta}) \\ &= -\frac{2gb}{g \cos(\hat{\theta})(\tan(\hat{\theta}) - \frac{b}{g})}.\end{aligned}$$

Using the inequalities in (26), we have

$$\begin{aligned}\mathcal{L}_{\mu}''(\hat{\theta}) &> 0, \\ \mathcal{L}_{\mu}''(\tilde{\theta}) &< 0.\end{aligned}$$

Based on the first-order and second-order optimality conditions, $\hat{\theta}$ is the minimum of the Lagrangian problem, and $\tilde{\theta}$ is the maximum.

C VOLTAGE INEQUALITY CONSTRAINTS

In this part, we prove that not all inequality constraints in (22c) are inactive by contradiction. We first suppose all inequality constraints in (22c) are inactive, and convert (22) to the penalized unconstrained formulation:

$$\mathcal{L}_{\rho}(\mathbf{x}) = f(\mathbf{x}) + \rho/2[h(\mathbf{x})]^2. \quad (28)$$

Assume ρ is sufficiently large, then (28) can be viewed as being equivalent to the original problem (22). Let us take gradients of $\mathcal{L}_{\rho}(\mathbf{x})$ with respect to V_1 and V_2 at a feasible solution $\tilde{\mathbf{x}}$. Since $\tilde{\mathbf{x}}$ satisfies $h(\tilde{\mathbf{x}}) = 0$, the terms multiplied by ρh in the gradients can be ignored. So the gradients are given by

$$\frac{\partial \mathcal{L}_{\rho}}{\partial V_1} = 2g\tilde{V}_1 - \tilde{V}_2(g \cos(\tilde{\theta}) - b \sin(\tilde{\theta})) \quad (29a)$$

$$\frac{\partial \mathcal{L}_{\rho}}{\partial V_2} = -\tilde{V}_1(g \cos(\tilde{\theta}) - b \sin(\tilde{\theta})). \quad (29b)$$

1) If $\frac{\partial \mathcal{L}_{\rho}}{\partial V_1} = 0$, then we have

$$g \cos(\tilde{\theta}) - b \sin(\tilde{\theta}) = 2g\frac{\tilde{V}_1}{\tilde{V}_2} \quad (\tilde{V}_2 \neq 0). \quad (30)$$

Plug (30) into (29b) and we get

$$\frac{\partial \mathcal{L}_{\rho}}{\partial V_2} = -2g\frac{\tilde{V}_1^2}{\tilde{V}_2} < 0.$$

This means if V_1 is inactive, then V_2 must be on the boundary of the constraint set.

2) Suppose $\frac{\partial \mathcal{L}_{\rho}}{\partial V_2} = 0$. Since $\tilde{V}_1 \neq 0$, we have

$$g \cos(\tilde{\theta}) - b \sin(\tilde{\theta}) = 0. \quad (31)$$

If we plug (31) into (29a), then we have

$$\frac{\partial \mathcal{L}_{\rho}}{\partial V_1} = 2g\tilde{V}_1 > 0.$$

That is, if V_2 is inactive, then V_1 must be on the boundary of the constraint set. Therefore one of V_1 and V_2 must be binding, and (22) can be reduced to the bivariate optimization problem.

D HESSIAN MATRIX OF THE LAGRANGIAN

In this part, we derive the Hessian matrix of the Lagrangian function for problem (22), where we optimize both voltage magnitudes and angles for a 2-bus network. In Appendix C, we have shown that one of V_1 and V_2 must be binding, so here we consider the cases where V_1 is binding or V_2 is binding separately.

We first suppose V_1 is inactive (equivalently, V_2 is binding). Then the Hessian matrix of the Lagrangian is

$$\nabla^2 \mathcal{L}_{\lambda, \mu} = \begin{pmatrix} \frac{\partial^2 \mathcal{L}_{\lambda, \mu}}{\partial \theta^2} & \frac{\partial^2 \mathcal{L}_{\lambda, \mu}}{\partial \theta \partial V_1} \\ \frac{\partial^2 \mathcal{L}_{\lambda, \mu}}{\partial \theta \partial V_1} & \frac{\partial^2 \mathcal{L}_{\lambda, \mu}}{\partial V_1^2} \end{pmatrix}.$$

The two leading principal minors of $\nabla^2 \mathcal{L}_{\lambda, \mu}$ at a feasible solution $\tilde{\mathbf{x}}$ are

$$\begin{aligned}D_1(\tilde{\mathbf{x}}) &= \frac{\partial^2 \mathcal{L}_{\lambda, \mu}}{\partial \theta^2} \\ &= \tilde{V}_1 \tilde{V}_2 [(1 + \tilde{\mu})g \cos(\tilde{\theta}) - (1 - \tilde{\mu})b \sin(\tilde{\theta})] \\ D_2(\tilde{\mathbf{x}}) &= \nabla^2 \mathcal{L}_{\lambda, \mu} \\ &= 2gD_1(\tilde{\mathbf{x}}) - \tilde{V}_2^2 [(1 + \tilde{\mu})g \sin(\tilde{\theta}) + (1 - \tilde{\mu})b \cos(\tilde{\theta})]^2\end{aligned}$$

where $\tilde{\mu}$ is the dual solution associated with $\tilde{\mathbf{x}}$. If $(\tilde{\mathbf{x}}, \tilde{\mu})$ are the optimal primal-dual solutions, then we can write out the first-order optimality condition of the Lagrangian w.r.t. θ :

$$(1 + \tilde{\mu})g \sin(\tilde{\theta}) + (1 - \tilde{\mu})b \cos(\tilde{\theta}) = 0. \quad (33)$$

From (33), $D_2(\tilde{\mathbf{x}})$ can be simplified as

$$D_2(\tilde{\mathbf{x}}) = 2gD_1(\tilde{\mathbf{x}}).$$

Also, we can represent $\tilde{\mu}$ in terms of $\tilde{\theta}$:

$$\tilde{\mu} = -\frac{g \sin(\tilde{\theta}) + b \cos(\tilde{\theta})}{g \sin(\tilde{\theta}) - b \cos(\tilde{\theta})}. \quad (34)$$

If we plug (34) into $D_1(\tilde{\mathbf{x}})$, then we have

$$D_1(\tilde{\mathbf{x}}) = \tilde{V}_1 \tilde{V}_2 \frac{-2gb}{g \cos(\tilde{\theta})(\tan(\tilde{\theta}) - \frac{b}{g})}.$$

Following from the inequalities in (23), $D_1(\tilde{\mathbf{x}})$ and hence $D_2(\tilde{\mathbf{x}})$ are positive at the global minimum and negative at the local minimum. Therefore, Theorem 3 holds for the case where V_1 is inactive and V_2 is binding.

Now we suppose V_2 is inactive (equivalently, V_1 is binding). Then the Hessian matrix of the Lagrangian is

$$\tilde{\nabla}^2 \mathcal{L}_{\lambda, \mu} = \begin{pmatrix} \frac{\partial^2 \mathcal{L}_{\lambda, \mu}}{\partial \theta^2} & \frac{\partial^2 \mathcal{L}_{\lambda, \mu}}{\partial \theta \partial V_2} \\ \frac{\partial^2 \mathcal{L}_{\lambda, \mu}}{\partial \theta \partial V_2} & \frac{\partial^2 \mathcal{L}_{\lambda, \mu}}{\partial V_2^2} \end{pmatrix}. \quad (35)$$

The two leading principal minors of $\tilde{\nabla}^2 \mathcal{L}_{\lambda, \mu}$ at a feasible solution $\tilde{\mathbf{x}}$ are

$$\begin{aligned} D_1(\tilde{\mathbf{x}}) &= \frac{\partial^2 \mathcal{L}_{\lambda, \mu}}{\partial \theta^2} \\ &= \tilde{V}_1 \tilde{V}_2 \frac{-2gb}{g \cos(\tilde{\theta})(\tan(\tilde{\theta}) - \frac{b}{g})} \end{aligned} \quad (36a)$$

$$\begin{aligned} \tilde{D}_2(\tilde{\mathbf{x}}) &= \tilde{\nabla}^2 \mathcal{L}_{\lambda, \mu} \\ &= 2g\tilde{\mu}D_1(\tilde{\mathbf{x}}) - \tilde{V}_1^2 [(1 + \tilde{\mu})g \sin(\tilde{\theta}) + (1 - \tilde{\mu})b \cos(\tilde{\theta})]^2 \end{aligned} \quad (36b)$$

where $\tilde{\mu}$ is the dual solution associated with $\tilde{\mathbf{x}}$. If $(\tilde{\mathbf{x}}, \tilde{\mu})$ are the optimal primal-dual solutions, then the first-order optimality condition

w.r.t. θ takes the same form as in (33). Then $\tilde{D}_2(\tilde{\mathbf{x}})$ in (36b) can be simplified as

$$\tilde{D}_2(\tilde{\mathbf{x}}) = 2g\tilde{\mu}D_1(\tilde{\mathbf{x}}). \quad (37a)$$

Since the multiplier $\tilde{\mu}$ represents the marginal price and is positive at the global minimum, $\tilde{D}_2(\tilde{\mathbf{x}})$ has the same sign as $\tilde{D}_1(\tilde{\mathbf{x}})$. From the inequalities in (23), $\tilde{D}_1(\tilde{\mathbf{x}})$ is positive, hence the Hessian matrix $\tilde{\nabla}^2 \mathcal{L}_{\lambda, \mu}$ is positive definite at the global minimum. For the local minimum, since $\tilde{D}_1(\tilde{\mathbf{x}})$ is negative from (23), the Hessian matrix $\tilde{\nabla}^2 \mathcal{L}_{\lambda, \mu}$ cannot be positive definite. This means it is either negative definite or indefinite at the local minimum. Therefore, Theorem 3 also holds for the case where V_2 is inactive and V_1 is binding.