A Sufficient Condition for Local Optima to be Globally Optimal

Fengyu Zhou and Steven H. Low
Engineering & Applied Science, Caltech, Pasadena CA

***Abstract*— Consider an optimization problem with a convex cost function but a non-convex compact feasible set $\mathcal{X}$, and its relaxation with a compact and convex feasible set $\hat{\mathcal{X}} \supset \mathcal{X}$. We prove that if from any point $x \in \hat{\mathcal{X}} \backslash \mathcal{X}$ there is a path connecting x to $\mathcal{X}$ along which both the cost function and a Lyapunov-like function are improvable, then any local optimum in $\mathcal{X}$ for the original non-convex problem is a global optimum. We use this result to show that, for AC optimal power flow problems, a well-known sufficient condition for exact relaxation also guarantees that all its local optima are globally optimal. This helps explain the widespread empirical experience that local algorithms for optimal power flow problems often work extremely well.**

I. Introduction

Motivation. Optimal power flow (OPF) is a class of constrained optimization problems that minimizes certain cost subject to nonlinear physical laws and operational constraints. OPF is fundamental in power systems as it underlies numerous power systems applications. It is non-convex and NP-hard [1], [2]. Traditionally OPF problems have been solved mostly using local algorithms such as Newton-Raphson or interior-point methods (e.g. [3]). Over the last decade various convex relaxations have been developed for solving OPF; see [4] and references therein. Empirically, convex relaxation methods often produce globally optimal solutions. More significantly, they provide a way to check the quality of the solutions produced by local algorithms. When verified against convex relaxations, these local algorithms turn out to often produce globally optimal solutions. This is useful as local algorithms are much more scalable than semidefinite relaxations. While sufficient conditions are known that guarantee the exactness of semidefinite relaxations (see surveys in [5], [4]), to the best of our knowledge, no analytical result is known that explains the remarkable performance of local algorithms on OPF problems. In this paper we provide the first sufficient condition for local optima of OPF to be globally optimal.

Summary. Specifically consider an optimization problem with a convex cost function but a non-convex compact feasible set $\mathcal{X}$, and its relaxation with a compact and convex feasible set $\hat{\mathcal{X}} \supset \mathcal{X}$. We prove that if there is a Lyapunov-like function that only vanishes over $\mathcal{X}$, and from any point $x \in \hat{\mathcal{X}} \setminus \mathcal{X}$ there is a path connecting x to $\mathcal{X}$ along which both the cost function and the Lyapunov-like function are improvable, then any local optimum in $\mathcal{X}$ for the original non-convex problem is a global optimum. We use this result to show that, for AC OPF over radial networks, a well-known sufficient condition for exact relaxation also guarantees that all its local optima are globally optimal. This helps explain the widespread empirical experience that local algorithms for OPF often work extremely well.

Beyond OPF. Though motivated by OPF, our main results (Theorems 1 and 2) are applicable to general non-convex optimization problems. These problems frequently arise in applications. Many cyber-physical systems, for instance, are governed by nonlinear physical laws that render their optimization non-convex. Most machine learning problems are non-convex problems since nonlinear models such as neural networks have a powerful representability of real data [6]. For some classes of non-convex problems, even solving them approximately is NP-hard [7]. Yet for many non-convex problems in signal processing and machine learning (e.g., dictionary learning, phase retrieval, sparse coding, matrix completion, low-rank semidefinite programs), simple local algorithms such as gradient descent or alternating minimization often produce globally optimal solutions. Some of the theoretical explanations for this phenomenon are summarized in [8], [9], [10], [11] and references therein. A common method is to study the gradient and curvature of the cost function over the feasible set and show that (i) all local optima are globally optimal, and (ii) any local maximum or saddle point always has a negative curvature in an eigen direction of the Hessian. This implies that a stochastic gradient descent or a well designed local algorithm can easily escape local maxima or saddle points to produce a global solution [12], [13], [8], [9]. Another technique, developed in [10], treats the non-convex sparse decoding problem as a convex problem with an unknown gradient and proves that alternative minimization is a gradient descent algorthim with approximate gradients that approach the true gradient.

These methods analyze the optimization landscape through the gradient and curvature of the cost function and usually require that the feasible set $\mathcal{X}$ has a simple structure, e.g., polytopes or spherical surface. When the feasible set $\mathcal{X}$ is a high dimensional manifold with highly non-convex features, it may be difficult to compute the gradient or curvature of the cost function over $\mathcal{X}$.

In contrast, our method is applicable to such a problem with a highly non-convex feasible set $\mathcal{X}$ because we leverage its convex relaxation $\hat{\mathcal{X}}$. An interesting feature is that our

Fengyu Zhou is with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA, 91125. Email: `f.zhou@caltech.edu`

Steven H. Low is with the Department of Electrical Engineering and the Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA, 91125. Email: `slow@caltech.edu`

conditions only involve properties of $\hat{\mathcal{X}} \setminus \mathcal{X}$ *outside* the feasible set, yet they imply an important (global optimality) property of local optimal points *in* $\mathcal{X}$.

The rest of the paper is organized as follows. We formulate our problem in Section II, prove our main results in Sections III and IV, apply these results to OPF in Section V, and conclude in Section VI.

II. Preliminaries

In this paper, we will use $\mathbb{K}$ to denote the set $\mathbb{R}$ of real numbers or the set $\mathbb{C}$ of complex numbers. For any finite interger n, $\mathbb{K}^n$ is a Banach space with a norm $\|\cdot\|$.

Consider a (potentially non-convex) optimization problem

$$\underset{x}{\text{minimize}} \quad f(x) \tag{1a}$$
$$\text{subject to} \quad x \in \mathcal{X} \tag{1b}$$

and its convex relaxation

$$\underset{x}{\text{minimize}} \quad f(x) \tag{2a}$$
$$\text{subject to} \quad x \in \hat{\mathcal{X}}. \tag{2b}$$

Here $\mathcal{X}$ is a nonempty compact subset of $\mathbb{K}^n$, not necessarily convex, while $\hat{\mathcal{X}} \subseteq \mathbb{K}^n$ is an arbitrary compact and convex superset of $\mathcal{X}$. The cost function $f : \hat{\mathcal{X}} \to \mathbb{R}$ is convex and continuous over $\hat{\mathcal{X}}$. We do not require the relaxation $\hat{\mathcal{X}}$ to be efficiently represented.

Definition 1: A point $x^{\text{lo}} \in \mathcal{X}$ is called a *local optimum* of (1) if there exists a $\delta > 0$ such that $f(x^{\text{lo}}) \le f(x)$ for all $x \in \mathcal{X}$ with $\|x - x^{\text{lo}}\| < \delta$.

Definition 2: We say the relaxation (2) is *exact* with respect to (1) if any optimal point of (2) is feasible, and hence globally optimal, for (1).

Definition 2 implies in particular that, if (2) is exact, then for any $\hat{x} \in \hat{\mathcal{X}} \setminus \mathcal{X}$, $f(\hat{x}) > \min_{x \in \hat{\mathcal{X}}} f(x)$.

Definition 3: A *path* in $\mathcal{S} \subseteq \mathbb{K}^n$ connecting point a to point b is a continuous function $h : [0,1] \to \mathcal{S}$ such that $h(0) = a$ and $h(1) = b$.

We may refer to a path by the corresponding function h in the remainder of the paper.

Lemma 1: The following are equivalent:

(A) Problems (2) is exact with respect to (1).

(B) For any $x \in \hat{\mathcal{X}} \setminus \mathcal{X}$, there is a path h in $\hat{\mathcal{X}}$ such that $h(0) = x, h(1) \in \mathcal{X}$, $f(h(t))$ is non-increasing for $t \in [0,1]$ and $f(h(0)) > f(h(1))$.

Proof: (A) $\Longrightarrow$ (B): Let x^* be any optimal point of (2). By (A), $x^* \in \mathcal{X}$, thus for $x \in \hat{\mathcal{X}} \setminus \mathcal{X}$, we could choose the path as the line segment from x to x^* since $\hat{\mathcal{X}}$ is convex.

(B) $\Longrightarrow$ (A): Condition (B) implies that no point $x \in \hat{\mathcal{X}} \setminus \mathcal{X}$ can be optimal for (2). ∎

Lemma 1 is not surprising, and in fact many works in the literature proving exact relaxations of optimal power flow problems can be interpreted as using (B) to prove (A) by implicitly finding such a path h for each $x \in \hat{\mathcal{X}} \setminus \mathcal{X}$ (see also Section V).

Condition (B) does not say anything about the local optima in $\mathcal{X}$ for (1). In the next section we will strengthen (B) by equipping the path with a Lyapunov-like function and show that the stronger condition implies that all local optima of (1) are globally optimal. We start by classifying local minima.

Definition 4: We classify each local optimum x^{lo} of (1) into three disjoint classes: x^{lo} is a

- *Global optimum* if $f(x^{\text{lo}}) \le f(x)$ for all the feasible $x \in \mathcal{X}$.
- *Pseudo local optimum* if there is a path $h : [0,1] \to \mathcal{X}$ such that $h(0) = x^{\text{lo}}$, $f(h(t)) \equiv f(x^{\text{lo}})$ for all $t \in [0,1]$ and $h(1)$ is not a local optimum.
- *Genuine local optimum* if it is neither a global optimum nor a pseudo local optimum.

Examples of all three classes are shown in Fig. 1.

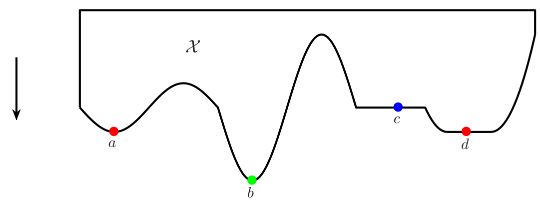

Fig. 1. Examples for three classes of local optima. The arrow indicates the direction along which the cost function linearly decreases. Point b (in green) is a global optimum, point c (in blue) is a pseudo local optimum, while points a and d (in red) are genuine local optima.

Definition 5: A point x is *improvable* in $\mathcal{X}$ if there is a path $h : [0,1] \to \mathcal{X}$ such that

- $h(0) = x$;
- $f(h(t))$ is non-increasing for $t \in [0,1]$;
- $h(1)$ is not a local optimum or $f(h(1)) < f(x)$.

Remark 1: If a local optimum is improvable in $\mathcal{X}$, then it must be a pseudo local optimum.

Finally we need the following properties about a collection of paths.

Definition 6: A set $\{h_i : i \in \mathcal{I}\}$ of paths indexed by i is said to be *uniformly bounded* if there is a finite number M such that $\|h_i(t)\| \le M$ for every $i \in \mathcal{I}$ and $t \in [0,1]$.

Definition 7: A set $\{h_i : i \in \mathcal{I}\}$ of paths is said to be *uniformly equicontinuous* if for any $\epsilon > 0$, there exists a $\delta > 0$ such that $\|h_i(t_1) - h_i(t_2)\| < \epsilon$ for every $i \in \mathcal{I}$ whenever $|t_1 - t_2| < \delta$.

Remark 2: The index set $\mathcal{I}$ could be empty or uncountably infinite. An empty path set (i.e., when $\mathcal{I} = \emptyset$) is considered to be both uniformly bounded and uniformly equicontinuous.

Remark 3: If $\mathcal{S}$ is compact in $\mathbb{K}^n$ and all paths in a set $\mathcal{H} = \{h_i : i \in \mathcal{I}\}$ are linear and $[0,1] \to \mathcal{S}$, then $\mathcal{H}$ must be both uniformly bounded and uniformly equicontinuous.

III. Main Result

Definition 8: A *Lyapunov-like function* [1] associated with (1) and (2) is a continuous function $V : \hat{\mathcal{X}} \to \mathbb{R}^+$ such that $V(x) = 0$ for $x \in \mathcal{X}$ and $V(x) > 0$ for $x \in \hat{\mathcal{X}} \setminus \mathcal{X}$.

The strengthened version of (B) is as follows.

(C) There exists a Lyapunov-like function V associated with (1) and (2) such that:

[1]In contrast to a standard Lyapunov function, we do not require V to be differentiable here.

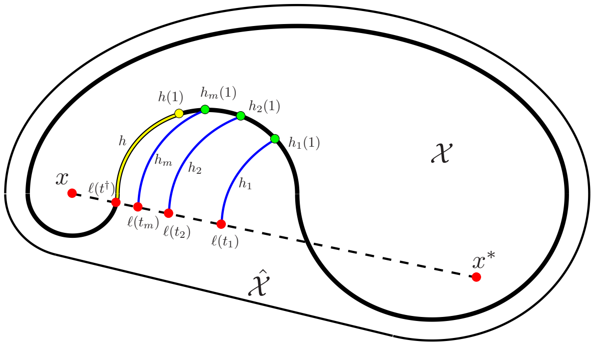

Fig. 2. Sketch of notations for the proof of Theorem 1. Point x and $\ell(t^\dagger)$ will be later proved to be identical to each other.

a) For any $x \in \hat{\mathcal{X}} \setminus \mathcal{X}$, there is a path h_x in $\hat{\mathcal{X}}$ such that $h_x(0) = x, h_x(1) \in \mathcal{X}$, both $f(h_x(t))$ and $V(h_x(t))$ are non-increasing for $t \in [0,1]$ and $f(h_x(0)) > f(h_x(1))$.
b) The set $\{h_x\}_{x\in\hat{\mathcal{X}}\setminus\mathcal{X}}$ is uniformly bounded and uniformly equicontinuous.

Remark 4: As (C) is stronger than (B), Lemma 1 implies that (C) $\Longrightarrow$ (A).

Theorem 1: If (C) holds, then any local optimum in $\mathcal{X}$ for (1) is either a global optimum or a pseudo local optimum.

Proof: An illustrative sketch of the notations used in this proof is in Fig. 2. Suppose $x \in \mathcal{X}$ is a local but not global optimum for (1). We will prove that x must be improvable in $\mathcal{X}$ (and thus a pseudo local optimum).

Let $x^* \neq x$ be a global optimum of (1), i.e., $f(x^*) < f(x)$. Let $\ell : [0,1] \to \hat{\mathcal{X}}$ be the linear function characterizing the line segment from x to x^*, i.e., $\ell(t) = (1-t)x + tx^*$ with $f(\ell(1)) = f(x^*) < f(x)$. Note that $f(\ell(t))$ is non-increasing in t. To see this, consider any $t \geq 0, \epsilon > 0$ with $t + \epsilon \leq 1$, $x_1 = \ell(t)$, $x_2 = \ell(t+\epsilon)$. Setting $s := \epsilon/(1-t)$, we have $x_2 = (1-s)x_1 + sx^*$. Since f is convex and x^* is also a global optimum of (2) over $\hat{\mathcal{X}}$ by Remark 4, we have

$$f(x_2) \;\leq\; (1-s)f(x_1) + sf(x^*) \;\leq\; f(x_1).$$

Define

$$t^\dagger := \sup_{t\in[0,1]} t \quad \text{s. t. } \ell(\tau) \in \mathcal{X} \quad \forall \tau \leq t.$$

As $\mathcal{X}$ is closed, $\ell(t^\dagger)$ is also in $\mathcal{X}$. We first prove $\ell(t^\dagger)$ must be x (i.e., $t^\dagger = 0$). Otherwise, as x is a local optimum, we could find $\delta \in (0, t^\dagger)$ such that $f(\ell(t)) \geq f(\ell(0)) = f(x)$ for all $t \in [0,\delta)$. Since $f(\ell(t))$ is non-increasing in t, we must have $f(\ell(t)) \equiv f(\ell(0)) = f(x)$ for all $t \in [0,\delta)$. It contradicts the fact that $f(\ell(t))$ is convex and $f(\ell(1)) = f(x^*) < f(x) = f(\ell(0))$ for the same reason f is non-increasing in t.

Therefore $\ell(t^\dagger) = x$ and $f(\ell(t^\dagger)) = f(x)$. It is sufficient to show $\ell(t^\dagger)$ is improvable in $\mathcal{X}$. That is to say, it is sufficient to find some function $h : [0,1] \to \mathcal{X}$ such that $h(0) = \ell(t^\dagger)$, $f(h(t))$ is non-increasing in $t \in [0,1]$ and either $f(h(1)) < f(\ell(t^\dagger))$ or $h(1)$ is not a local optimum in $\mathcal{X}$ for (1).

By the definition of $t^\dagger$, there is a decreasing sequence $t_m \to t^\dagger$ such that $t_m \in (t^\dagger, 1]$ and $\ell(t_m) \in \hat{\mathcal{X}} \setminus \mathcal{X}$ for all m. Since $f(\ell(t))$ is non-increasing in t, the sequence $f(\ell(t_m))$ is non-decreasing in m and $f(\ell(t_m)) \leq f(\ell(t^\dagger))$. For each $\ell(t_m)$ we take the function $h_m : [0,1] \to \hat{\mathcal{X}}$ guaranteed by Condition (C). As the sequence h_m is uniformly bounded and uniformly equicontinuous, a subsequence must uniformly converge to a limit h by Arzelà-Ascoli theorem. Without loss of generality, we denote this subsequence as h_m as well. Next we prove this h satisfies all the properties in Definition 5, implying the improvability of x.

To show $h(t) \in \mathcal{X}$ for any fixed $t \in [0,1]$, we consider the sequence $(V(h_m(t)) : m \in \mathbb{Z})$. As V is continuous, we have

$$\lim_{m\to\infty} V(h_m(t)) = V(h(t)) \geq 0.$$

On the other hand, we have $V(h_m(t)) \leq V(h_m(0))$, thus

$$\begin{aligned}\lim_{m\to\infty} V(h_m(t)) &\leq \lim_{m\to\infty} V(h_m(0)) \\ &= \lim_{m\to\infty} V(\ell(t_m)) = V(\ell(t^\dagger)) = 0.\end{aligned}$$

Hence $V(h(t)) = 0$ and $h(t) \in \mathcal{X}$.

To show $h(0) = \ell(t^\dagger)$, we consider

$$h(0) = \lim_{m\to\infty} h_m(0) = \lim_{m\to\infty} \ell(t_m) = \ell(t^\dagger).$$

To show $f(h(t))$ is non-increasing, we take any $s, t \in [0,1]$ such that $s < t$. As f is continuous, we have

$$\begin{aligned}f(h(s)) &= \lim_{m\to\infty} f(h_m(s)) \\ f(h(t)) &= \lim_{m\to\infty} f(h_m(t))\end{aligned}$$

and by Condition (C) we have $f(h_m(s)) \geq f(h_m(t))$ for each m. Therefore $f(h(s)) \geq f(h(t))$.

Finally, we will show if $f(h(1)) = f(\ell(t^\dagger))$ then $h(1)$ must not be a local minimal in $\mathcal{X}$ for (1). For each m,

$$f(h_m(1)) < f(h_m(0)) = f(\ell(t_m)) \leq f(\ell(t^\dagger)) = f(h(1))$$

and $h_m(1) \in \mathcal{X}$. Since the sequence $h_m(1)$ converges to $h(1)$ as $m \to \infty$, within any open neighborhood of $h(1)$ in $\mathcal{X}$, we could always find some $h_m(1)$ with strictly smaller cost value. Thus $h(1)$ cannot be a local minimum in $\mathcal{X}$. ■

IV. Corollaries

A corollary of Theorem 1 is as follows.

Corollary 1: Condition (C) implies that the feasible set of (1) is connected.

Proof: If $\mathcal{X}$ is not connected, then by definition $\mathcal{X}$ can be partitioned into two disjoint non-empty closed sets $\mathcal{X}_1$ and $\mathcal{X}_2$ with $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2$, which are hence both compact. Further we let x_i be any global optimum of $\min_{x\in\mathcal{X}_i} f(x)$ for $i = 1, 2$. Clearly $x_1 \neq x_2$ and they are both local optima of (1).

If $f(x_1) = f(x_2)$, then any convex combination of x_1, x_2 must be a global optimum to (2). Since x_1 and x_2 are not connected, there must some convex combination that is outside $\mathcal{X}$. This contradicts the exactness of relaxation.

If $f(x_1) \neq f(x_2)$, without loss of generality we assume $f(x_1) < f(x_2)$, i.e., x_2 is not a global optimum of (1). But x_2 is not a pseudo local optimum of (1) either, contradicting Theorem 1. To see this, note that any point $x' \in \mathcal{X}$ which is connected to x_2 via a path in $\mathcal{X}$ must also be a point in $\mathcal{X}_2$ and if $f(x') = f(x_2)$ then x' must be a local optimum of (1) as well. ∎

Theorem 1 guarantees that any local optimum is either global optimum or pseudo local optimum. It is therefore possible for a gradient-based algorithm to be trapped at a pseudo local optimum. In the following, we discuss some conditions that rule out pseudo local optima and therefore guarantee that any local optimum must be a global optimum.

Corollary 2: If all local optima of (1) are isolated, then Condition (C) implies that any local optimum of (1) is a global optimum.

Here, local optima being isolated means any local optimum of (1) has an open neighborhood which contains no other local optimum. The proof is straightforward as by definition isolated local optimum could not be pseudo local optimum.

Another way to get rid of pseudo local optima is by strengthening the monotonicity of $f(h_x(t))$ in Condition (C). Consider the following condition which is slightly stronger than (C).

(C') Condition (C) holds, and there exists $k > 0$ such that $\forall x \in \hat{\mathcal{X}} \setminus \mathcal{X}$, $\forall 0 \leq t < s \leq 1$ we have

$$f(h_x(t)) - f(h_x(s)) \geq k\|h_x(t) - h_x(s)\|. \qquad (3)$$

In Condition (C'), $\|\cdot\|$ could be any norm on $\mathbb{K}^n$. As a caveat, ℓ_0-"norm" is *not* allowed here as it is not a norm since it does not satisfy $\|\alpha x\| = |\alpha|\|x\|$. Note that Condition (C) already implies $f(h_x(t)) - f(h_x(s)) \geq 0$, while (C') strengthens this condition by enforcing a positive lower bound depending on h_x.

Theorem 2: If (C') holds, then any local optimum of (1) must be a global optimum.

Proof: Following the proof of Theorem 1, suppose $x \in \mathcal{X}$ is a local but not global optimum for (1). Then we have $x = \ell(t^\dagger)$ and could obtain a limit point of the sequence h_m, denoted as h. Since both sides of (3) are continuous in $h_m(t)$ and $h_m(s)$, and the limits of $h_m(t)$ and $h_m(s)$ are $h(t)$ and $h(s)$, we must have whenever $h(t) \neq h(s)$,

$$f(h(t)) - f(h(s)) \geq k\|h(t) - h(s)\| > 0.$$

Taking $t = 0$ we can conclude that $h(0)$ (i.e. $\ell(t^\dagger)$, which is the same point as x) is not a local optimum of (1). ∎

V. Application: Optimal Power Flow

In this section, we apply the main results of Sections III and IV to non-convex AC Optimal Power Flow problems (OPFs) to show that a well-known sufficient condition for exact Second-Order Cone Program (SOCP) relaxation also guarantees that any local optimum of the original non-convex problem is a global optimum. This sufficient condition for exactness is proposed in [14]. It is applicable to a radial network and requires that no lower bound for real or reactive power injections ever be tight at any optimal point of the SOCP relaxation. Under the same condition, our result explains why many local search algorithms such as primal-dual interior-point methods tend to converge to a global optimum, even though in general OPF is non-convex and NP-hard.

A. System Model

Consider a power network with an underlying connected directed graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$. Let $\mathcal{V} := \{0, 1, \cdots, N-1\}$ be the set of buses (i.e., nodes), and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ be the set of power lines (i.e., edges). We will refer to a power line from bus j to bus k by $j \to k$ or (j,k) interchangeably. For each power line (j,k), its series admittance is denoted by $y_{jk} \in \mathbb{C}$, and its series impedance is hence $z_{jk} := y_{jk}^{-1}$. Both the real and imaginary parts of z_{jk} are assumed to be positive.

We adopt the Branch Flow Model (BFM) to formulate power flow equations. For each bus j, let $V_j \in \mathbb{C}, s_j = p_j + \mathbf{i}q_j \in \mathbb{C}$ denote its voltage and bus injection respectively. For line (j,k), let S_{jk} and $I_{jk} \in \mathbb{C}$ denote the branch power flow and current from bus j to k, both at the sending end. We will denote the conjugate of a complex number a by a^{H}. The power flow equations are:

$$V_j - V_k = z_{jk} I_{jk}, \qquad \forall (j,k) \in \mathcal{E} \quad (4a)$$

$$S_{jk} = V_j I_{jk}^{\mathsf{H}}, \qquad \forall (j,k) \in \mathcal{E} \quad (4b)$$

$$\sum_{i:i\to j} (S_{ij} - z_{ij}|I_{ij}|^2) + s_j = \sum_{k:j\to k} S_{jk}, \qquad \forall j \in \mathcal{V} \quad (4c)$$

If the underlying graph $\mathcal{G}$ is a tree, then we can introduce $v_j := |V_j|^2 \in \mathbb{R}$ and $\ell_{jk} := |I_{jk}|^2 \in \mathbb{R}$. The power flow equations (4) can be equivalently represented in terms of new variables as:

$$v_j = v_k + 2\mathrm{Re}(z_{jk} S_{jk}^{\mathsf{H}}) - |z_{jk}|^2 \ell_{jk}, \quad \forall (j,k) \in \mathcal{E} \quad (5a)$$

$$v_j = \frac{|S_{jk}|^2}{\ell_{jk}}, \qquad \forall (j,k) \in \mathcal{E} \quad (5b)$$

$$s_j = \sum_{k:j\to k} S_{jk} - \sum_{i:i\to j} (S_{ij} - z_{ij}\ell_{ij}), \qquad \forall j \in \mathcal{V} \quad (5c)$$

These are the DistFlow equations introduced in [15], [16].

Given a cost function $f(s) : \mathbb{C}^N \to \mathbb{R}$, we are interested in the following OPF problem:

$$\begin{aligned}
\underset{x=(s,v,\ell,S)}{\text{minimize}} \quad & f(s) && (6a)\\
\text{subject to} \quad & (5) && (6b)\\
& \underline{v}_j \leq v_j \leq \overline{v}_j && (6c)\\
& \underline{s}_j \leq s_j \leq \overline{s}_j && (6d)\\
& \ell_{jk} \leq \overline{\ell}_{jk} && (6e)
\end{aligned}$$

All the inequalities for complex numbers in this section are enforced for both the real and imaginary parts.

B. Optimality

Consider a stronger version of monotonicity.

Definition 9: A function $g : \mathbb{R} \to \mathbb{R}$ is *strongly increasing* if there exists real $c > 0$ such that for any $a > b$, we have

$$g(a) - g(b) \geq c(a - b).$$

We now make the following assumptions on OPF:

(i) The underlying graph $\mathcal{G}$ is a tree.
(ii) The cost function f is convex, and is strongly increasing in $\mathrm{Re}(s_j)$ (or $\mathrm{Im}(s_j)$) for each $j \in \mathcal{V}$ and non-decreasing in $\mathrm{Im}(s_j)$ (or $\mathrm{Re}(s_j)$).
(iii) The problem (6) is feasible.
(iv) The line current limit satisfies $\overline{\ell}_{jk} \leq \underline{v}_j |y_{jk}|^2$.

Assumption (i) is generally true for distribution networks and assumption (iii) is typically mild. As for (ii), f is commonly assumed to be convex and increasing in $\mathrm{Re}(s_j)$ and $\mathrm{Im}(s_j)$ in the literature (e.g., [17], [18], [19]). Assumption (ii) is only slightly stronger since one could always perturb any increasing function by an arbitrarily small linear term to achieve strong monotonicity. Assumption (iv) is not common in the literature but is also mild because of the following reason. Typically $V_j = (1 + \epsilon_j)e^{\mathbf{i}\theta_j}$ in per unit where $\epsilon \in [-0.1, 0.1]$ and the angle difference $\theta_{jk} := \theta_j - \theta_k$ between two neighboring buses j, k typically has a small magnitude. Thus the maximum value of $|V_j - V_k|^2 = |(1 + \epsilon_j)e^{\mathbf{i}\theta_{jk}} - (1+\epsilon_k)|^2$, which is equivalent to $\overline{\ell}_{jk}/|y_{jk}|^2$, should be much smaller than $\underline{v}_j$ which is ≈ 1 per unit.

Problem (6) is non-convex, as constraint (5b) is not convex. Denote by $\mathcal{X}$ the set of (s, v, ℓ, S) that satisfy (6b)-(6e), so (6) is in the form of (1). We can relax (6) by convexifying (5b) into a second-order cone [14]:

$$\begin{aligned} &\underset{x=(s,v,\ell,S)}{\text{minimize}} && f(s) && \text{(7a)}\\ &\text{subject to} && \text{(5a), (5c), (6c)} - \text{(6e)} && \text{(7b)}\\ &&& |S_{jk}|^2 \leq v_j \ell_{jk} && \text{(7c)} \end{aligned}$$

One can similarly regard $\hat{\mathcal{X}}$ as the set of (s, v, ℓ, S) that satisfy (7b), (7c). It is proved in [14] that if $\underline{s}_j = -\infty - \mathbf{i}\infty$ for all $j \in \mathcal{V}$, then (7) is exact, meaning any optimal solution of (7) is also feasible and hence globally optimal for (6). Now we show that the same condition also guarantees that any local optimum of (6) is also globally optimal. This implies that a local search algorithm such as the primal-dual interior point method can produce a global optimum as long as it converges.

Theorem 3: If $\underline{s}_j = -\infty - \mathbf{i}\infty$ for all $j \in \mathcal{V}$, then any local optimum of (6) is a global optimum.

To prove Theorem 3, our strategy is to find a Lyapunov-like function such that Condition (C') holds. Theorem 2 then directly implies the theorem. Let

$$V(x) := \sum_{(j,k)\in\mathcal{E}} v_j \ell_{jk} - |S_{jk}|^2. \qquad (8)$$

Clearly, V is a valid Lyapunov-like function satisfying Definition 8.

Proof: We first prove that for V defined in (8), Condition (C) holds. For each $x = (s, v, \ell, S) \in \hat{\mathcal{X}} \setminus \mathcal{X}$, let $\mathcal{M}$ be the set of $(j, k) \in \mathcal{E}$ such that $|S_{jk}|^2 < v_j \ell_{jk}$. For $(j, k) \in \mathcal{M}$, the quadratic function

$$\phi_{jk}(a) := \frac{|z_{jk}|^2}{4} a^2 + \big(v_j - \mathrm{Re}(z_{jk} S_{jk}^{\mathsf{H}})\big)a + |S_{jk}|^2 - v_j \ell_{jk}$$

must have a unique positive zero as $|S_{jk}|^2 - v_j \ell_{jk} < 0$. We define Δ_{jk} to be this positive zero if $(j, k) \in \mathcal{M}$ and 0 otherwise.

Assumption (iv) implies $\ell_{jk} \leq v_j |y_{jk}|^2$, and therefore

$$\begin{aligned} & v_j - \mathrm{Re}(z_{jk} S_{jk}^{\mathsf{H}}) \;\geq\; v_j - |z_{jk}||S_{jk}| \\ \geq\; & v_j - |z_{jk}|\sqrt{v_j \ell_{jk}} \;\geq\; v_j - |z_{jk}|\sqrt{v_j^2 |y_{jk}|^2} \;=\; 0. \end{aligned}$$

It further implies $\phi_{jk}(a)$ is strictly increasing for $a \in [0, \Delta_{jk}]$.

Now consider the path $h_x(t) := (\tilde{s}(t), \tilde{v}(t), \tilde{\ell}(t), \tilde{S}(t))$ for $t \in [0, 1]$, where

$$\tilde{s}_j(t) = s_j - \frac{t}{2} \sum_{i:i\to j} z_{ij}\Delta_{ij} - \frac{t}{2} \sum_{k:j\to k} z_{jk}\Delta_{jk}, \qquad (9a)$$

$$\tilde{v}_j(t) = v_j, \qquad (9b)$$

$$\tilde{\ell}_{jk}(t) = \ell_{jk} - t\Delta_{jk}, \qquad (9c)$$

$$\tilde{S}_{jk}(t) = S_{jk} - \frac{t}{2} z_{jk}\Delta_{jk}. \qquad (9d)$$

Clearly we have $h_x(0) = x \in \hat{\mathcal{X}} \setminus \mathcal{X}$. We show in Appendix A that $h_x(t)$ is feasible for (7) for $t \in [0, 1]$ and $h_x(1)$ is feasible for (6). Therefore, h_x is indeed $[0, 1] \to \hat{\mathcal{X}}$ and $h_x(1) \in \mathcal{X}$.

Since $z_{jk} > 0$, both real and imaginary parts of $\tilde{s}_j(t)$ are strictly decreasing for $(j, k) \in \mathcal{M}$ and stay unchanged otherwise. By assumption (ii), $f(\tilde{s}(t))$ is also strictly decreasing. To show $V(h_x(t))$ is also decreasing, we notice that $V(h_x(t))$ equals to

$$\begin{aligned} & \sum_{(j,k)\in\mathcal{E}} \tilde{v}_j(t)\tilde{\ell}_{jk}(t) - |\tilde{S}_{jk}(t)|^2 \\ = & \sum_{(j,k)\in\mathcal{M}^c} v_j\ell_{jk} - |S_{jk}|^2 + \sum_{(j,k)\in\mathcal{M}} \tilde{v}_j(t)\tilde{\ell}_{jk}(t) - |\tilde{S}_{jk}(t)|^2 \\ = & \sum_{(j,k)\in\mathcal{M}^c} v_j\ell_{jk} - |S_{jk}|^2 - \sum_{(j,k)\in\mathcal{M}} \phi_{jk}(t\Delta_{jk}). \end{aligned}$$

As $\phi_{jk}(a)$ is strictly increasing for $a \in [0, \Delta_{jk}]$, we conclude that $V(h_x(t))$ is strictly decreasing for $t \in [0, 1]$.

By Remark 3, the set $\{h_x\}_{x\in\hat{\mathcal{X}}\setminus\mathcal{X}}$ is uniformly bounded and uniformly equicontinuous as all $h_x(t)$ are linear functions in t. In summary, Condition (C) is satisfied.

Finally, we show Condition (C') also holds. By assumption (ii), there exists some real $c > 0$ independent of x such that for any $0 \leq a < b \leq 1$,

$$\begin{aligned} & f(\tilde{s}(a)) - f(\tilde{s}(b)) \\ \geq\; & c \sum_{j\in\mathcal{V}} \mathrm{Re}(\tilde{s}_j(a) - \tilde{s}_j(b)) + \mathrm{Im}(\tilde{s}_j(a) - \tilde{s}_j(b)) \\ =\; & c\|\tilde{s}(a) - \tilde{s}(b)\|_{\mathrm{m}} \end{aligned}$$

where $\|\cdot\|_{\mathrm{m}}$ is defined as $\|\mathbf{a}\|_{\mathrm{m}} := \sum_i |\mathrm{Re}(a_i)| + |\mathrm{Im}(a_i)|$ over the complex vector space. It is easy to check $\|\cdot\|_{\mathrm{m}}$ is a valid norm.

On the other hand, by (9) we have $\|\tilde{v}(a) - \tilde{v}(b)\|_{\mathrm{m}} \equiv 0$ and

$$\|\tilde{\ell}(a) - \tilde{\ell}(b)\|_{\mathrm{m}} \le \frac{1}{\max\limits_{(j,k)\in\mathcal{E}}\{|z_{jk}|\}}\|\tilde{s}(a) - \tilde{s}(b)\|_{\mathrm{m}},$$

$$\|\tilde{S}(a) - \tilde{S}(b)\|_{\mathrm{m}} \le \frac{1}{2}\|\tilde{s}(a) - \tilde{s}(b)\|_{\mathrm{m}}.$$

Therefore,

$$\|h_x(a) - h_x(b)\|_{\mathrm{m}} \le \Big(\frac{3}{2} + \frac{1}{\max\limits_{(j,k)\in\mathcal{E}}}\Big)\|\tilde{s}(a) - \tilde{s}(b)\|_{\mathrm{m}}$$

and there exists $\hat{c} > 0$ independent of x, a, b such that

$$f(\tilde{s}(a)) - f(\tilde{s}(b)) \ge \hat{c}\|h_x(a) - h_x(b)\|_{\mathrm{m}}.$$

Therefore Condition (C') is also satisfied and by Theorem 2, any local optimum of (6) is a global optimum. ■

VI. Conclusion

Our main results (Theorems 1 and 2) provide new conditions to guarantee global optimality for non-convex optimization problems. Specifically we show that if from any point $x \in \hat{\mathcal{X}} \setminus \mathcal{X}$ there is a path connecting x to $\mathcal{X}$ along which both the cost function and a Lyapunov-like function are improvable, then any local optimum in $\mathcal{X}$ is globally optimal. This implies that, for AC OPF over radial networks, a well-known sufficient condition for exact relaxation also guarantees that there are no spurious local optimum. Hence local algorithms are not only much more scalable than semidefiniite relaxations, but also tend to perform well.

Most works in the literature prove global optimality of non-convex problems by studying the gradient and curvature of the cost function and are applicable for problems with a nonconvex cost function but a tractable feasible set. We propose a new approach that is applicable to problems with highly non-convex feasible sets by leveraging their convex relaxations. In general constructing a Lyapunov-like function function V and a path h_x for each $x \in \hat{\mathcal{X}} \setminus \mathcal{X}$ that satisfy condition (C) is difficult. As an example, the feasible set of OPF is characterized by nonlinear power flow equations and exhibit a complicated structure (see [20] for a visualization). Fortunately, its SOCP relaxation suggests a natural Lyapunov-like function. Our results suggest that other conditions for exact relaxations might also be leveraged to guarantee global optimality of local algorithms.

Besides OPF, there are other problems, e.g., matrix completion [21], [13], that are generally NP-hard, but in practice can be solved by local algorithms or convex relaxations. Our condition is sufficient for both exact relaxation and global optimality of local solutions. This suggests a potentially deeper connection between those two properties. In particular our condition may characterize the intersection of problem instances that have exact relaxations and instances whose local optima are always globally optimal.

References

[1] A. Verma, "Power grid security analysis : An optimization approach," Ph.D. dissertation, Columbia University, 2009.

[2] K. Lehmann, A. Grastien, and P. Van Hentenryck, "AC-feasibility on tree networks is NP-hard," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 798–801, 2016.

[3] R. A. Jabr, A. H. Coonick, and B. J. Cory, "A primal-dual interior point method for optimal power flow dispatching," *IEEE Transactions on Power Systems*, vol. 17, no. 3, pp. 654–662, 2002.

[4] D. K. Molzahn, I. A. Hiskens *et al.*, "A survey of relaxations and approximations of the power flow equations," *Foundations and Trends® in Electric Energy Systems*, vol. 4, no. 1-2, pp. 1–221, 2019.

[5] S. H. Low, "Convex relaxation of optimal power flow–part II: Exactness," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 2, pp. 177–189, 2014.

[6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[7] P. Jain, P. Kar *et al.*, "Non-convex optimization for machine learning," *Foundations and Trends® in Machine Learning*, vol. 10, no. 3-4, pp. 142–336, 2017.

[8] J. Sun, Q. Qu, and J. Wright, "When are nonconvex problems not scary?" *arXiv preprint arXiv:1510.06096*, 2015.

[9] R. Ge, C. Jin, and Y. Zheng, "No spurious local minima in nonconvex low rank problems: A unified geometric analysis," in *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 2017, pp. 1233–1242.

[10] S. Arora, R. Ge, T. Ma, and A. Moitra, "Simple, efficient, and neural algorithms for sparse coding," *Proceedings of Machine Learning Research*, vol. 40, January 2015.

[11] S. Burer and R. D. Monteiro, "Local minima and convergence in low-rank semidefinite programming," *Mathematical Programming*, vol. 103, no. 3, pp. 427–444, 2005.

[12] J. Sun, Q. Qu, and J. Wright, "Complete dictionary recovery over the sphere i: Overview and the geometric picture," *IEEE Transactions on Information Theory*, vol. 63, no. 2, pp. 853–884, 2016.

[13] R. Ge, J. D. Lee, and T. Ma, "Matrix completion has no spurious local minimum," in *Advances in Neural Information Processing Systems*, 2016, pp. 2973–2981.

[14] M. Farivar and S. H. Low, "Branch flow model: Relaxations and convexification–part I," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2554–2564, 2013.

[15] M. E. Baran and F. F. Wu, "Optimal Capacitor Placement on radial distribution systems," *IEEE Trans. Power Delivery*, vol. 4, no. 1, pp. 725–734, 1989.

[16] ——, "Optimal Sizing of Capacitors Placed on A Radial Distribution System," *IEEE Trans. Power Delivery*, vol. 4, no. 1, pp. 735–743, 1989.

[17] B. Zhang and D. Tse, "Geometry of injection regions of power networks," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 788–797, 2013.

[18] L. Gan, N. Li, U. Topcu, and S. H. Low, "Exact convex relaxation of optimal power flow in radial networks," *IEEE Transactions on Automatic Control*, vol. 60, no. 1, pp. 72–87, 2015.

[19] M. Nick, R. Cherkaoui, J.-Y. Le Boudec, and M. Paolone, "An exact convex formulation of the optimal power flow in radial distribution networks including transverse components," *IEEE Transactions on Automatic Control*, vol. 63, no. 3, pp. 682–697, 2017.

[20] I. A. Hiskens and R. J. Davy, "Exploring the power flow solution space boundary," *IEEE transactions on power systems*, vol. 16, no. 3, pp. 389–395, 2001.

[21] E. J. Candès and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2053–2080, 2010.

Appendix

A. Feasibility of $h_x(t)$

We first show that for $t \in [0, 1]$, $h_x(t)$ is feasible for (7). That is to say, constraints (5a), (5c), (6c)-(6e) and (7c) are satisfied.

To show (5a) holds, consider

$$\begin{aligned}
&\tilde{v}_k(t) + 2\mathrm{Re}(z_{jk}\tilde{S}_{jk}^{\mathsf{H}}(t)) - |z_{jk}|^2\tilde{\ell}_{jk}(t) \\
=\ & v_k + 2\mathrm{Re}\Big(z_{jk}\big(S_{jk}^{\mathsf{H}} - \frac{t}{2}z_{jk}^{\mathsf{H}}\Delta_{jk}\big)\Big) - |z_{jk}|^2(\ell_{jk} - t\Delta_{jk}) \\
=\ & v_k + 2\mathrm{Re}(z_{jk}S_{jk}^{\mathsf{H}}) - |z_{jk}|^2\ell_{jk} \quad = \quad v_j \quad = \quad \tilde{v}_j(t).
\end{aligned}$$

To show (5c) holds, consider

$$\begin{aligned}
&\tilde{S}_{ij}(t) - z_{ij}\tilde{\ell}_{ij}(t) = S_{ij} - \frac{t}{2}z_{ij}\Delta_{ij} - z_{ij}(\ell_{ij} - t\Delta_{ij}) \\
=&S_{ij} - z_{ij}\ell_{ij} + \frac{t}{2}z_{ij}\Delta_{ij}
\end{aligned}$$

and hence

$$\begin{aligned}
&\sum_{k:j\to k}\tilde{S}_{jk}(t) - \sum_{i:i\to j}(\tilde{S}_{ij}(t) - z_{ij}\tilde{\ell}_{ij}(t)) \\
=&\sum_{k:j\to k}\big(S_{jk} - \frac{t}{2}z_{jk}\Delta_{jk}\big) - \sum_{i:i\to j}\big(S_{ij} - z_{ij}\ell_{ij} + \frac{t}{2}z_{ij}\Delta_{ij}\big) \\
=&s_j - \frac{t}{2}\sum_{i:i\to j}z_{ij}\Delta_{ij} - \frac{t}{2}\sum_{k:j\to k}z_{jk}\Delta_{jk} = \tilde{s}_j(t).
\end{aligned}$$

Constraints (6c)-(6e) are satisfied since $\tilde{v} = v$ and $(\tilde{s}, \tilde{\ell})$ are decreased compared to (s, ℓ) for both the real and imaginary parts for each coordinate.

To show (7c) holds, we only consider those $(j, k) \in \mathcal{M}$ since the constraints for other lines are not affected. We have

$$\begin{aligned}
&|\tilde{S}_{jk}(t)|^2 - \tilde{v}_j(t)\tilde{\ell}_{jk}(t) \\
=&\big|S_{jk} - \frac{t}{2}z_{jk}\Delta_{jk}\big|^2 - v_j(\ell_{jk} - t\Delta_{jk}) \\
=&|S_{jk}|^2 - \mathrm{Re}(z_{jk}S_{jk}^{\mathsf{H}})t\Delta_{jk} + \frac{|z_{jk}|^2}{4}t^2\Delta_{jk}^2 - v_j(\ell_{jk} - t\Delta_{jk}) \\
=&\phi_{jk}(t\Delta_{jk}) \ \leq \ \phi_{jk}(\Delta_{jk}) \ = \ 0. \qquad (10)
\end{aligned}$$

where the last inequality follows because $\phi_{jk}(a)$ is increasing for $a \in [0, \Delta_{jk}]$ and the last equality follows because, by definition, Δ_{jk} is the unique positive zero of ϕ_{jk}.

Moreover, when $t = 1$, equality is achieved in (10) and hence $h_x(1)$ satisfies (5b). As a result, $h_x(1)$ is feasible for (6), as desired.