Conditions for Exact Convex Relaxation and No Spurious Local Optima

Fengyu Zhou, *Student Member, IEEE,* and Steven H. Low, *Fellow, IEEE*

Abstract—Non-convex optimization problems can be approximately solved via relaxation or local algorithms. For many practical problems such as optimal power flow (OPF) problems, both approaches tend to succeed in the sense that relaxation is usually exact and local algorithms usually converge to a global optimum. In this paper, we study conditions which are sufficient or necessary for such non-convex problems to simultaneously have exact relaxation and no spurious local optima. Those conditions help us explain the widespread empirical experience that local algorithms for OPF problems often work extremely well.

Index Terms—Convex relaxation, local optimum, optimal power flow, semidefinite program.

I. INTRODUCTION

NON-CONVEX optimization problems in general are computationally challenging. However, many heuristics tend to work well for real-world problems. Those approaches include convex relaxations and local algorithms. It is usually hoped that relaxations yields exact solutions and local optima are also globally optimal. In this paper, we derive the conditions, sufficient or necessary, for these two properties to hold simultaneously. Our focus is specifically on the optimization formulation with convex cost and non-convex constraints.

A. Related Works

Many problems have been proved to have exact relaxation and no spurious local optima (such as matrix completion [2], [3], [4], low rank semidefinite program [5], [6], [7]), the proofs for those two properties are usually based on different types of certificates. In this subsection, we review some widely-used certificates for each property.

One type of certificates to exhibit relaxation exactness is by showing that any relaxed (and infeasible) point maps to a feasible solution with lower cost. This asserts that relaxed points cannot be the optimal solution. For instance, [8], [9] prove that optimal power flow problems can be solved via second-order cone relaxation under certain conditions, using the argument that any solution in the interior of the second-order cone can always be moved towards the boundary to further reduce the cost. In [5], [6], it is proved that if a semidefinite program has a solution with sufficiently large rank, then one can always reduce the rank without increasing the cost or violating the constraints. Another type of certificates involve studying the dual variables and KKT conditions. The

underlying idea is a pair of primal and dual solutions satisfying KKT conditions that certify their optimality for both the primal and dual problems. Thus constructing dual variables with certain structures can also certify the optimality of primal solutions. In [2] for instance, the dual variable is related to the subgradient of the cost function at a desired matrix and therefore it helps certify the optimality of that desired matrix. Another example is [10], which proves the primal matrix should be of rank 1 through the argument that the null space of its dual matrix has dimension at most 1. Similar techniques are also used in [11], [12], [13].

There are also considerable literature establishing the global optimality of local optima. We refer to [14], [15] and references therein. In [14], the authors focus on a class of problems with twice continuously differentiable function as the cost and Riemannian manifold as the feasible set. The values of Riemannian gradient and Hessian at a certain point then help certify properties such as *strong gradient*, *negative curvature* or *local convexity* in its neighborhood. It then eliminates spurious local optima and saddle points, where local algorithms can be trapped. This technique or idea were also used in [16] for dictionary recovery problem and [17] for phase synchronization. In both problems, the Riemannian manifold is some n-sphere or the Cartesian product of n-spheres. The framework summarized in [15] also leverages the landscape of the cost function, and the problem is usually reformulated into an unconstrained form. Instead of explicitly computing the gradient and Hessian matrix, the paper shows it suffices to find a single direction of improvement. For certain symmetric positive definite problems, the paper shows that the decision variable will always get closer to the global optimizer when the cost is reduced. A similar idea was also applied in [18], where the main result is built upon a correlation condition which states that the gradient (or any updating rule) is correlated with the direction from the current location towards the global optimizer. Therefore the underlying algorithm such as gradient descent can always produce a solution closer to the global optimizer as the algorithm progresses.

B. Contribution

The brief review above shows most works study exact relaxation and local optimality separately. It is unclear what might be the common feature of non-convex problems that possess both properties. Many real-world non-convex problems, however, seem to possess both properties, either provably or empirically, and it is hard to explain why these nice properties, though seemingly different, often occur simultaneously. Besides, most literature on local optimality focuses on

problems without constraints or with tractable constraints. This is usually the case for problems in the learning area. However, for problems arising in cyber-physical systems, the constraints could include non-convex functions enforced by physical laws, as we will see in power systems. In these cases, either the feasible set is not a Riemannian manifold, or the Riemannian gradient and Hessian are very hard to derive. These questions motivate us to study conditions, sufficient or necessary, for problems to simultaneously have exact relaxation and no spurious local optima. These conditions also help us study local optimality using properties of its relaxation, instead of its landscape.

Our conditions have two parts. The first part, which also appeared in [1], is on the sufficient condition. Roughly, if for any relaxed point, there exists a path connecting it to the non-convex feasible set and the path satisfies the following:

- along the path the cost is non-increasing,
- along the path the 'distance' to the non-convex feasible set is non-increasing,

then the problem must have exact relaxation and no spurious local optima simultaneously. Here the 'distance' can be any properly constructed function, as we will define later as a Lyapunov-like function (Definition 10). The second part is on the necessary condition, which says that if a problem does have exact relaxation and no spurious local optima simultaneously, then there must exist such a Lyapunov-like function and paths satisfying the requirements above. [1]

Though Lyapunov-like functions and paths are guaranteed to exist, for specific problems it could still be difficult to construct them. We then derive certain rules to construct a Lyapunov-like function and paths of a new problem from primitive problems with known Lyapunov-like functions and paths. This process allows us to reuse and extend known results as the problem changes and grows. Finally, we apply the proposed approach to two specific problems, optimal power flow (OPF) and low rank SDP. Our work proves the first known condition (that can be checked *a priori*) for OPF to have no spurious local optima, and it helps explain the widespread empirical experience that local algorithms for OPF problems often work extremely well.

C. Background for Power Systems

As one of the applications and main motivation of this work, OPF is a core problem in power systems. First proposed in [19], OPF is a class of optimization problems that minimizes a certain cost subject to nonlinear physical laws and operational constraints. It is known to be non-convex and NP-hard in its AC formulation [20], [10], [21]. Therefore, there is no known efficient algorithm that can solve all problem instances in polynomial time. Traditional approaches to solving OPF are usually based on local algorithms such as Newton-Raphson, see [22], [23], [24] for examples. Over the past decade, techniques on convex relaxation have also been introduced to solve OPF [25], [26]. A surprising empirical finding in the literature shows that despite the non-convexity, both local algorithms and convex relaxations very often yield global

---

---

optimum of the original non-convex problem [25], [26], [10], [27]. In recent years, there have been considerable analytical works on provable conditions for the relaxation exactness, which are summarized in the reviews [28], [29] and references therein. However, few analytical results are known on the performance guarantee of local algorithms. In this paper, we show that a known sufficient condition for exact relaxation is also sufficient for local optima to be globally optimal. To the best of our knowledge, this is the first analytical result of its kind, and we hope that the approaches developed in this paper can help derive more sufficient conditions along this direction.

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 positive integer n, $\mathbb{K}^n$ is a Banach space.

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 $f(x^{\text{lo}}) \leq f(x)$ for all $x \in \mathcal{X}$ with $\|x - x^{\text{lo}}\| < \delta$.*

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

Unless otherwise specified, we will always use the term *exact* to refer to such strong exactness. Definition 2 implies in particular that, if (2) is exact, then $\forall \hat{x} \in \hat{\mathcal{X}} \setminus \mathcal{X}$, $f(\hat{x}) > \min_{x \in \hat{\mathcal{X}}} f(x)$.

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 as the corresponding function h in the remainder of the paper.

Lemma 1. *The following are equivalent:*

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

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 $x \in \hat{\mathcal{X}} \setminus \mathcal{X}$ [28].

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

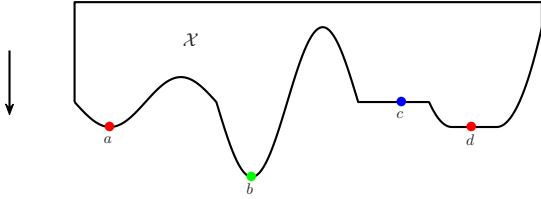Examples of all three classes are shown in Fig. 1.

Fig. 1. Examples for three classes of local optima. The arrow indicates the direction along which the cost function decreases. Point b is a global optimum, point c is a pseudo local optimum, while points a and d 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. *A local optimum is a pseudo local optimum if and only if it is improvable in $\mathcal{X}$.*

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 $\|h_i(t)\|_\infty \leq M$ for every $i \in \mathcal{I}$ and $t \in [0,1]$.*

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

Let $\Pi|_a^b$ be the family of all the finite ordered subsets of $[a,b]$. We use Π as a shorthand for $\Pi|_0^1$. For $\pi = (t_0, \cdots, t_N) \in \Pi$ and a path h, define

$$L_\pi(h) := \sum_{i=1}^{N} \|h(t_{i-1}) - h(t_i)\|_{\ell_2}.$$

Clearly, $L_\pi(h)$ is always finite for given π and h.

Definition 8 ([30]). *For path h, define the function $L(h) := \sup_{\pi \in \Pi} L_\pi(h)$. We say h is rectifiable iff $L(h)$ is finite. When h is rectifiable, $L(h)$ is also referred to as its length.*

Definition 9 ([30]). *For a rectifiable path $h : [0,1] \to \mathbb{K}^n$, let its arc-length reparameterization be $\bar{h} : [0,1] \to \mathbb{K}^n$ and*

$$\begin{cases} \bar{h}\left(\frac{1}{L(h)} \sup_{\pi \in \Pi|_0^t} L_\pi(h)\right) := h(t), & \text{if } L(h) > 0 \\ \bar{h} := h, & \text{if } L(h) = 0 \end{cases}$$

One could see $L(\bar{h}) = L(h) < \infty$ and they have the same function image, i.e., $\{\bar{h}(t)|t \in [0,1]\} = \{h(t)|t \in [0,1]\}$. For $0 \leq t_1 \leq t_2 \leq 1$, $\bar{h}$ has the property that $\sup_{\pi \in \Pi|_{t_1}^{t_2}} L_\pi(\bar{h}) = (t_2 - t_1)L(\bar{h})$.

Lemma 2. *For a set of rectifiable paths $h_i, i \in \mathcal{I}$, if the values of $L(h_i)$ are uniformly bounded, then the set of $\bar{h}_i, i \in \mathcal{I}$ is uniformly equicontinuous.*

Proof. Assume $L(h_i) \leq M$ for all $i \in \mathcal{I}$, then for any $0 \leq t_1 \leq t_2 \leq 1$, we have for any i,

$$\|\bar{h}_i(t_1) - \bar{h}_i(t_2)\|_\infty \leq \|\bar{h}_i(t_1) - \bar{h}_i(t_2)\|_{\ell_2}$$
$$\leq \sup_{\pi \in \Pi|_{t_1}^{t_2}} L_\pi(\bar{h}_i) = (t_2 - t_1)L(h_i) \leq M|t_1 - t_2|.$$

Setting $\delta = \epsilon/M$, the equicontinuity is proved. $\qquad \square$

Corollary 1. *If $\mathcal{S}$ is compact in $\mathbb{K}^n$ and all paths in a set $\mathcal{H} = \{h_i : i \in \mathcal{I}\}$ are $[0,1] \to \mathcal{S}$ and consist of at most N linear segments, then $\{\bar{h}_i : i \in \mathcal{I}\}$ must be both uniformly bounded and uniformly equicontinuous. Here N is a finite constant for all paths in $\mathcal{H}$.*

III. SUFFICIENT CONDITIONS

In this section, we first study the sufficient conditions under which (2) is exact w.r.t. (1) and all the local optima of (1) are also globally optimal. Those sufficient conditions will be proposed by strengthening Condition (B). Note that (B) has already implied (2) is exact w.r.t. (1), so our strategy is to strengthen (B) in order to rule out the possibility of genuine local optima and pseudo local optima.

A. Ruling Out Genuine Local Optima

Definition 10. *A* Lyapunov-like function [2] *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 $x \in \hat{\mathcal{X}} \setminus \mathcal{X}$.*

A strengthened version of (B) is as follows.

---

[2] In contrast to a standard Lyapunov function, we do not require V to be differentiable here.
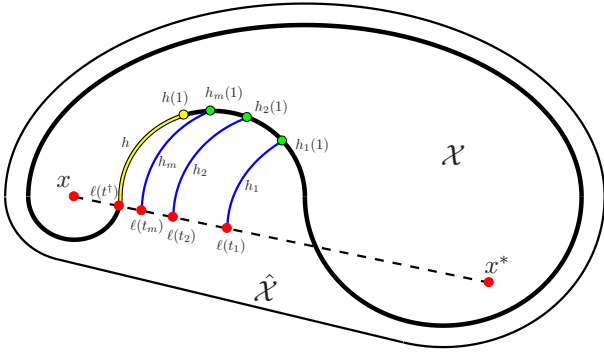
Fig. 2. Sketch of notations for the proof of Theorem 1. Point x and $\ell(t^\dagger)$ will be later proved to be identical to each other.

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

(C1) 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 $f(h_x(0)) > f(h_x(1))$.

(C2) The set $\{h_x\}_{x \in \hat{\mathcal{X}} \setminus \mathcal{X}}$ is uniformly bounded and uniformly equicontinuous.

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

Proof. (C) $\implies$ (A) is because (C) is stronger than (B). As for the second part of the argument, we include an illustrative sketch of the notations 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), so $f(x^*) < f(x)$. 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 $\hat{\mathcal{X}}$, 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 $f(\ell(t_m)) < f(\ell(t^\dagger))$. [3] 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 $\hat{\mathcal{X}}$ is closed, we have $h(t) = \lim_{m \to \infty} h_m(t) \in \hat{\mathcal{X}}$. Further consider V is continuous and $V(h_m(t)) \leq V(h_m(0))$, thus

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

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

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

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,

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

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 $\mathcal{X}$. $\square$

B. Ruling Out Pseudo Local Optima

So far, Condition (C) has eliminated the possibility of having genuine local optima, and in this subsection we further strengthen the condition to also rule out pseudo local optima. Consider the following lemma and its corollaries.

Lemma 3. *If (1) is exact with respect to (2) and (1) has no genuine local optima, then 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).

---

[3]The strict inequality is due to the convexity of $f(\ell(t))$ and the fact that $f(\ell(1)) < f(\ell(t^\dagger))$.

If $f(x_1) = f(x_2)$, then any convex combination of x_1, x_2 must be a global optimum to (2). Since there is no path in $\mathcal{X}$ that connects x_1 and x_2, there must be 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. $\square$

Corollary 2. *Condition (C) implies that the feasible set of* (1) *is connected.*

Now we are in a good position to discuss some conditions that rule out pseudo local optima and therefore guarantee that any local optimum must be a global optimum.

Corollary 3. *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. In fact, in this case the optimum can be proved to be also unique.

Another way to eliminate 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

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

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 $h(0)$ (which is the same point as x) is not a local optimum of (1). $\square$

IV. NECESSARY CONDITIONS

In this section we will study the necessary conditions for a non-convex problem to have exact relaxation and no spurious local optima simultaneously. It turns out the results are not exactly the converses of Theorem 1 or Theorem 2, but in a slightly weaker sense. Specifically, we show that if a non-convex problem is known to have exact relaxation and no spurious local optima simultaneously, then the Lyapunov-like function and paths satisfying Condition (C) are guaranteed to exist. However, it still may or may not be easy to find those functions or paths in practice for a specific problem.

A. Results

Assumption 1. *The feasible set $\mathcal{X}$ is semianalytic and the cost function f is analytic.*

We refer to [31] for more detailed definitions and properties of semianalytic sets. This assumption is not restrictive for most engineering problems. If $\mathbb{K}$ is chosen as $\mathbb{C}$, then we suggest to view all the complex functions as functions of real variables by separating the real and imaginary parts, and the space of $\mathbb{C}^n$ can be viewed as a shorthand for $\mathbb{R}^{2n}$ in this section.

Theorem 3 (necessary condition)**.** *If* (2) *is exact with respect to* (1) *and any local optimum of* (1) *is globally optimal, then there exists a Lyapunov-like function V and a corresponding family of paths $\{h_x\}_{x \in \hat{\mathcal{X}} \setminus \mathcal{X}}$ satisfying (C1) and (C2).*

Remark 3. *Note that Theorem 3 is NOT the converse of Theorem 1 in a strict sense. There are a few differences in their settings.*

- *Theorem 1 allows pseudo local optimum (in the conclusion) of the theorem, while Theorem 3 disallows it (in the premise).*
- *Theorem 3 relies on Assumption 1 while Theorem 1 does not.*

B. Proof Setup

In the rest of the section, we will prove Theorem 3. From now on, we assume (2) is exact with respect to (1) and any local optimum of (1) is also globally optimal. We first have the following definition and lemmas, which are the main reasons we introduced Assumption 1.

Definition 11 (Whitney regularity [31], [32], [33])**.** *For a compact set $\mathcal{U} \subset \mathbb{K}^n$ and a positive integer p, we say $\mathcal{U}$ is p-regular if there exists $C > 0$ such that $\forall x, y \in \mathcal{U}$, x, y can be joined by a rectifiable curve h in $\mathcal{U}$ satisfying $L(h) \leq C \|x - y\|^{1/p}$.*

Lemma 4 (Theorem 6.10 in [31])**.** *If $\mathcal{U}$ is a compact connected subanalytic subset of $\mathbb{K}^n$, then there is a positive integer p such that $\mathcal{U}$ is p-regular and the curves can always be chosen semianalytic.*

The proof of Lemma 4 can be found in [31]. Note that any semianalytic set is also subanalytic.

Lemma 5. *For any $x_0 \in \mathcal{X}$ that is not a local optimum of* (1) *and for any $\epsilon > 0$, there exists a path h in $\mathcal{X}$ such that*

$h(0) = x_0$, $f(h(t))$ is non-increasing in t, $f(h(1)) < f(h(0))$ and $L(h) < \epsilon$.

Proof. Consider the set $\mathcal{U} := \{x \in \mathcal{X} : f(x) \le f(x_0)\}$, which by definition is also semi-analytic. Since $x_0 \in \mathcal{X}$ is not an optimum of (1), the problem $\min_{x \in \mathcal{U}} f(x)$ must also be exact with respect to (2) and it does not introduce new local optima compared to (1). By Lemma 3, $\mathcal{U}$ must be connected.

According to Lemma 4, there is a rectifiable and semi-analytic curve h_0 in $\mathcal{U}$ such that $h_0(0) = x_0$, $L(h_0) < \epsilon$, $f(h_0(1)) < f(x_0)$ and $f(h_0(t)) \le f(x_0)$ for all $t \in [0, 1]$. Here $f(h_0(1))$ can be chosen as any point in $\mathcal{U}$ which has a strictly smaller cost value than x_0 and is sufficiently close to x_0 in Euclidean distance. [4] It is known that a semianalytic curve is analytic except for a finite number of points [34]. Assume $h_0(t)$ is not analytic at $0 = a_0 < a_1 < \cdots < a_k = 1$ where $k \ge 1$. By Theorem on the parametrization of a semi-analytic arc in [35] and the assumption that f is analytic, the value of $f(h_0(t))$ within any interval $[a_{\ell-1}, a_\ell]$ should be equal to some analytic function defined over an open superset of $[a_{\ell-1}, a_\ell]$. Since $f(h_0(1)) < f(h_0(0))$, the function $f(h_0(t))$ cannot be a constant function over $[0, 1]$. Let $[a_{\ell-1}, a_\ell]$ be the first interval within which $f(h_0(t))$ is not constant, then $f(h_0(a_{\ell-1})) = f(h_0(0))$. As $f(h_0(t))$ within $[a_{\ell-1}, a_\ell]$ equals to a analytic function defined over an open superset of $[a_{\ell-1}, a_\ell]$, there must be a small subinterval $[a_{\ell-1}, a_{\ell-1} + \delta)$ for some $\delta > 0$ within which we always have

$$f(h_0(t)) = f(h_0(a_{\ell-1})) + \sum_{i=0}^{\infty} c_i (t - a_{\ell-1})^i,$$

where the right hand side is the Taylor expansion of $f(h_0(t))$ at $a_{\ell-1}$. Since $f(h_0(t))$ is not constant over $[a_{\ell-1}, a_\ell]$, the coefficients c_i cannot all be zeros by the identity theorem. Suppose c_{i_0} is the first nonzero coefficient in the sequence $\{c_i\}_{i=0}^{\infty}$, we have two cases. If $c_{i_0} > 0$, then $f(h_0(t))$ is strictly increasing within $[a_{\ell-1}, a_{\ell-1} + \delta')$ for some small positive $\delta' < \delta$. It contradicts to the facts that $f(h_0(a_{\ell-1})) = f(x_0)$ and $f(h_0(t)) \le f(x_0)$ for all $t \in [0, 1]$. If $c_{i_0} < 0$, then $f(h_0(t))$ is strictly decreasing within $[a_{\ell-1}, a_{\ell-1} + \delta')$ for some small positive $\delta' < \delta$. Then we can construct a new path h such that $h(t) = h_0(t \cdot (a_{\ell-1} + \delta'))$ for all $t \in [0, 1]$. It is easy to check such h satisfies all the requirements in Lemma 5. $\square$

Now we consider weaker versions of (C1) and (C2).

(C3) 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 $t \in [0, 1]$.

(C4) All the $\{L(h_x)\}_{x \in \hat{\mathcal{X}} \setminus \mathcal{X}}$ are finite and uniformly bounded.

Compared to (C1), (C3) does not require $f(h_x(0)) > f(h_x(1))$ to strictly hold. Then we have a weaker version of Theorem 3 as follows.

Lemma 6 (weaker necessary condition). *If* (2) *is exact to* (1) *and any local optimum of* (1) *is also globally optimal, then there always exists a Lyapunov-like function V and a*

corresponding family of paths $\{h_x\}_{x \in \hat{\mathcal{X}} \setminus \mathcal{X}}$ satisfying (C3) and (C4).

We now show that Lemma 6, though weaker in its statement, actually implies Theorem 3, so later on we will only focus on the proof of Lemma 6. To see this, we suppose $V^\dagger$ and $\{h_x^\dagger\}_{x \in \hat{\mathcal{X}} \setminus \mathcal{X}}$ are the Lyapunov-like function and paths guaranteed by Lemma 6.

For each $x \in \hat{\mathcal{X}} \setminus \mathcal{X}$, if $h_x^\dagger(1)$ is a local optimum (so it is also a global optimum) of (1) then we must have $f(h_x(1)) < f(h_x(0))$ since the relaxation is exact. We construct $h_x^\ddagger = h_x^\dagger$.

If $h_x^\dagger(1)$ is not a local optimum, then by Lemma 5, there exists a path h_x^l, which satisfies $h_x^l(0) = h_x^\dagger(1)$, $h_x^l(t) \in \mathcal{X}$, $f(h_x^l(t))$ is non-increasing, $f(h_x^l(1)) < f(h_x^l(0))$ and $L(h_x^l) < \epsilon$. Here we choose ϵ as a fixed positive value for all x. We then construct

$$h_x^\ddagger(t) := \begin{cases} h_x^\dagger(2t), & \text{if } t \in [0, 1/2] \\ h_x^l(2t - 1), & \text{if } t \in [1/2, 1]. \end{cases}$$

We now let $V = V^\dagger$ and $h_x = \overline{h_x^\ddagger}$ for all $x \in \hat{\mathcal{X}} \setminus \mathcal{X}$. Recall that $\overline{h_x^\ddagger}$ is the arc-length reparameterization of $h_x^\ddagger$. Clearly, such construction satisfies (C1) as we strictly reduce the cost at the end of each path unless the path has already reached the global optimum. Besides, $\{h_x\}_{x \in \hat{\mathcal{X}} \setminus \mathcal{X}}$ is uniformly bounded as $\hat{\mathcal{X}}$ is compact, and $\{L(h_x)\}_{x \in \hat{\mathcal{X}} \setminus \mathcal{X}}$ also has the uniform upperbound as both $L(h_x^\dagger)$ and $L(h_x^l)$ are uniformly bounded for all x. By Lemma 2 $\{h_x\}_{x \in \hat{\mathcal{X}} \setminus \mathcal{X}}$ is uniformly equicontinuous, so (C2) is also satisfied. To summarize, when Lemma 6 is correct, one can always revise the Lyapunov-like function and paths provided by Lemma 6 to make Theorem 3 hold as well. Therefore, in the rest of the section, we will only prove Lemma 6 and the correctness of Theorem 3 just follows.

Let x^* be any global optimum of (1), then it is also an optimum of (2). Define

$$\mathcal{H} := \{h \mid h : [0, 1] \to \hat{\mathcal{X}} \text{ is continuous and } L(h) < \infty\}$$
$$\bar{\mathcal{H}} := \{\bar{h} \mid h \in \mathcal{H}\}$$
$$\hat{\mathcal{H}} := \{h \mid h \in \bar{\mathcal{H}}, \ f(h(t)) \ge f(h(1)) \text{ for all } t \in [0, 1]\}.$$

An immediate observation is if a continuous function $h : [0, 1] \to \hat{\mathcal{X}}$ satisfies $L(h) < \infty$ and $f(h(t)) \ge f(h(1))$ for all t, then $\bar{h} \in \hat{\mathcal{H}}$.

C. Construction

We construct V as

$$V(x) = \inf_{\substack{h \in \hat{\mathcal{H}} \\ h(0) = x \\ h(1) \in \mathcal{X}}} L(h) \qquad (4)$$

Lemma 7. *For a sequence $(h_i)_{i=1}^{\infty}$ where $h_i \in \hat{\mathcal{H}}$, if both $(h_i)_{i=1}^{\infty}$ and $(L(h_i))_{i=1}^{\infty}$ are uniformly bounded, then there must be a subsequence which uniformly converges to some h^* such that its arc-length reparameterization, denoted as $\overline{h^*}$, is in $\hat{\mathcal{H}}$. Furthermore, $L(h^*) = L(\overline{h^*}) \le \limsup_i L(h_i)$.*

Proof. By Lemma 2, $(h_i)_{i=1}^{\infty}$ is both uniformly bounded and uniformly equicontinuous. By Arzelà-Ascoli theorem, a subsequence of $(h_i)_{i=1}^{\infty}$ uniformly converges to a limit h^*. Without

loss of generalization, we denote this subsequence as $(h_i)_{i=1}^\infty$ as well. By uniform limit theorem and the compactness of $\hat{\mathcal{X}}$, $\overline{h^*}$ is a continuous function mapping from $[0,1]$ to $\hat{\mathcal{X}}$. To show $\overline{h^*} \in \hat{\mathcal{H}}$, it is sufficient to show $f(h^*(t)) \geq f(h^*(1))$ for all $t \in [0,1]$ and $L(h^*) < \infty$. If $f(h^*(t)) = f(h^*(1)) - \epsilon$ for some $t \in [0,1]$ and $\epsilon > 0$, then for sufficiently large i, we would have $|f(h_i(t)) - f(h^*(t))| < \epsilon/3$ and $|f(h_i(1)) - f(h^*(1))| < \epsilon/3$. Thus $f(h_i(t)) \leq f(h_i(1)) - \epsilon/3$ and it contradicts to $h_i \in \hat{\mathcal{H}}$.

Instead of showing $L(h^*) < \infty$, we directly prove $L(h^*) \leq \limsup_i L(h_i)$. Otherwise, there exists $\pi = (t_1, \cdots, t_N) \in \Pi$ such that $L_\pi(h^*) = \limsup_i L(h_i) + \epsilon$ for $\epsilon > 0$. For sufficiently large i, we have

$$|L_\pi(h_i) - L_\pi(h^*)|$$
$$= \Big| \sum_{j=1}^N \|h_i(t_{j-1}) - h(t_j)\|_{\ell_2} - \sum_{j=1}^N \|h^*(t_{j-1}) - h^*(t_j)\|_{\ell_2} \Big|$$
$$\leq \sum_{j=1}^N \Big(\|h_i(t_{j-1}) - h^*(t_{j-1})\|_{\ell_2} + \|h_i(t_j) - h^*(t_j)\|_{\ell_2} \Big) \leq \frac{\epsilon}{2}$$

Thus, $L(h_i) \geq L_\pi(h_i) \geq \limsup_i L(h_i) + \epsilon/2$ holds for sufficiently large i. It contradicts to the definition of $\limsup$. As a result, we must have $L(h^*) \leq \limsup_i L(h_i)$. $\square$

Lemma 8. *The optimization in* (4) *is feasible and the optimal cost can be achieved.*

Proof. We fix some $x \in \hat{\mathcal{X}}$. To show the feasibility, consider $h_{\text{fea}}(t) := (1-t)x + tx^*$, which is feasible to (4). Let $L_{\text{fea}} = L(h_{\text{fea}})$. Since L_{fea} is finite and $L(h)$ is non-negative, $V(x)$ must be finite. To show the achievability of the optimal cost, if not, then there must be a sequence of feasible $(h_i)_{i=1}^\infty$ such that

$$L_{\text{fea}} > L(h_i) \geq L(h_{i+1}) > V(x) \text{ for all } i \geq 1$$
$$\lim_{i\to\infty} L(h_i) = V(x).$$

The compactness of $\hat{\mathcal{X}}$ implies $(h_i)_{i=1}^\infty$ is uniformly bounded as well. By Lemma 7, a subsequence of $(h_i)_{i=1}^\infty$, denoted as $(h_i)_{i=1}^\infty$ as well, uniformly converges to a limit h^* and $L(h^*) = L(\overline{h^*}) \leq V(x)$. Moreover,

$$\overline{h^*}(0) = h^*(0) = \lim_{i\to\infty} h_i(0) = x$$
$$\overline{h^*}(1) = h^*(1) = \lim_{i\to\infty} h_i(1) \in \mathcal{X}.$$

Above all, we proved $\overline{h^*}$ is feasible to (4), and the cost $L(\overline{h^*})$ is not worse than $V(x)$. It contradicts to the non-achievability assumption. $\square$

For each $x \in \hat{\mathcal{X}} \setminus \mathcal{X}$, we construct h_x as

$$h_x = \underset{\substack{h \in \hat{\mathcal{H}} \\ h(0)=x \\ h(1)\in\mathcal{X}}}{\arg\min} L(h) \qquad (5)$$

If there are multiple minimizers then h_x can be chosen as any one of them.

Lemma 9. *For $x \in \hat{\mathcal{X}} \setminus \mathcal{X}$, the function h_x is injective.*

Proof. Otherwise, for some x, there exist $t_1 < t_2$ such that $h_x(t_1) = h_x(t_2)$. Since $h_x \in \hat{\mathcal{H}} \subseteq \bar{\mathcal{H}}$, we have

$$\sup_{\pi\in\Pi|_{t_1}^{t_2}} L_\pi(h_x) = (t_2 - t_1)L(h_x) = (t_2 - t_1)V(x) > 0$$

Consider a new path defined as

$$h^*(t) := \begin{cases} h_x(t), & \text{if } t \in [0,1] \setminus [t_1, t_2] \\ h_x(t_1), & \text{if } t \in [t_1, t_2] \end{cases}$$

It is easy to check h^* is continuous and entirely within $\hat{\mathcal{X}}$. For any $t \in [0,1]$, $h_x \in \hat{\mathcal{H}}$ implies $f(h^*(t)) \geq f(h_x(1)) = f(h^*(1))$. Further, we have

$$L(h^*) = \sup_{\pi\in\Pi} L_\pi(h^*)$$
$$= \sup_{\pi\in\Pi|_0^{t_1}} L_\pi(h^*) + \sup_{\pi\in\Pi|_{t_1}^{t_2}} L_\pi(h^*) + \sup_{\pi\in\Pi|_{t_2}^1} L_\pi(h^*)$$
$$= \sup_{\pi\in\Pi|_0^{t_1}} L_\pi(h_x) + 0 + \sup_{\pi\in\Pi|_{t_2}^1} L_\pi(h_x)$$
$$< \sup_{\pi\in\Pi|_0^{t_1}} L_\pi(h_x) + \sup_{\pi\in\Pi|_{t_1}^{t_2}} L_\pi(h_x) + \sup_{\pi\in\Pi|_{t_2}^1} L_\pi(h_x)$$
$$= \sup_{\pi\in\Pi} L_\pi(h_x) = L(h_x).$$

Above all, the arc-length reparameterization of h^*, denoted as $\overline{h^*}$, is feasible to (5) but achieves a strictly lower cost than h_x. This contradicts to the optimality of h_x. $\square$

Corollary 4. *For distinctive $t_1, t_2, t_3 \in [0,1]$, if $f(h_x(t_1)) \geq f(h_x(t_2))$ and $f(h_x(t_1)) > f(h_x(t_3))$, then $\|h_x(t_1) - h_x(t_2)\|_{\ell_2} + \|h_x(t_1) - h_x(t_3)\|_{\ell_2} > \|h_x(t_2) - h_x(t_3)\|_{\ell_2}$.*

Proof. It is sufficient to show $h_x(t_1)$ is not the convex combination of $h_x(t_2)$ and $h_x(t_3)$. Otherwise, we assume $h_x(t_1) = \lambda h_x(t_2) + (1-\lambda)h_x(t_3)$ for some $\lambda \in [0,1]$. First, Lemma 9 implies $\lambda \neq 0, 1$. For $\lambda \in (0,1)$, the convexity of f implies

$$f(h_x(t_1)) = f(\lambda h_x(t_2) + (1-\lambda)h_x(t_3))$$
$$\leq \lambda f(h_x(t_2)) + (1-\lambda)f(h_x(t_3))$$
$$< \lambda f(h_x(t_1)) + (1-\lambda)f(h_x(t_1)) = f(h_x(t_1)).$$

This contradiction shows $h_x(t_1)$ is not the convex combination of $h_x(t_2)$ and $h_x(t_3)$. Then the triangle inequality implies this corollary. $\square$

Lemma 10. *For each h_x defined in* (5)*, we have $f(h_x(t))$ is non-increasing in t for $t \in [0,1]$.*

Proof. We fix an $x_0 \in \hat{\mathcal{X}} \setminus \mathcal{X}$ and prove the result for h_{x_0} defined above. If not, then there exist $0 \leq t_1 < t_2 \leq 1$ such that $f(h_{x_0}(t_1)) < f(h_{x_0}(t_2))$. Now define

$$t^\dagger = \underset{t\in[t_1,1]}{\arg\max} f(h_{x_0}(t)), \quad t^\ddagger = \underset{\substack{t\in[t^\dagger,1] \\ f(h_{x_0}(t))=f(h_{x_0}(t^\dagger))}}{\max} t.$$

In other words, $t^\dagger$ is an arbitrary maximizer of $f(h_{x_0}(t))$, while $t^\ddagger$ is the largest maximizer. Clearly, both $t^\dagger$ and $t^\ddagger$ are well defined (due to the continuity and closedness) and are strictly between t_1 and 1. We also have $f(h_{x_0}(t^\ddagger)) > f(h_{x_0}(1))$ strictly holds. By the continuity of $h_{x_0}(\cdot)$ and

$f(h_{x_0}(\cdot))$, there exist $r, \delta > 0$ such that $[t^{\ddagger} - \delta, t^{\ddagger} + \delta] \subseteq (t_1, 1)$ and

- for $x \in \mathcal{B}(h_{x_0}(t^{\ddagger}), r) \cap \hat{\mathcal{X}}$, $f(h_{x_0}(1)) \leq f(x)$.
- for $t \in [t^{\ddagger} - \delta, t^{\ddagger})$, $h_{x_0}(t) \in \mathcal{B}(h_{x_0}(t^{\ddagger}), r)$. Therefore $f(h_{x_0}(1)) \leq f(h_{x_0}(t)) \leq f(h_{x_0}(t^{\ddagger}))$.
- for $t \in (t^{\ddagger}, t^{\ddagger} + \delta]$, $h_{x_0}(t) \in \mathcal{B}(h_{x_0}(t^{\ddagger}), r)$. Therefore $f(h_{x_0}(1)) \leq f(h_{x_0}(t)) < f(h_{x_0}(t^{\ddagger}))$.

Now we construct another h^* as

$$h^*(t) = \begin{cases} h_{x_0}(t), & \text{if } t \in [0,1] \setminus [t^{\ddagger} - \delta, t^{\ddagger} + \delta] \\ \frac{t^{\ddagger} + \delta - t}{2\delta} h_{x_0}(t^{\ddagger} - \delta) + \frac{t - t^{\ddagger} + \delta}{2\delta} h_{x_0}(t^{\ddagger} + \delta), \\ \quad \text{if } t \in [t^{\ddagger} - \delta, t^{\ddagger} + \delta]. \end{cases}$$

It is easy to verify h^* is continuous. For $t \in [t^{\ddagger} - \delta, t^{\ddagger} + \delta]$, $h^*(t)$ is the convex combination of $h_{x_0}(t^{\ddagger} - \delta)$ and $h_{x_0}(t^{\ddagger} + \delta)$, and must be within $\mathcal{B}(h_{x_0}(t^{\ddagger}), r) \cap \hat{\mathcal{X}}$, which is convex. Therefore, h^* is entirely within $\hat{\mathcal{X}}$ and $f(h^*(t)) \geq f(h_{x_0}(1)) = f(h^*(1))$ holds for all t.

Next, we are showing $L(h^*) < L(h_{x_0})$ by (6). The strict inequality in (6) is because of Corollary 4.

Above all, the arc-length reparameterization of h^*, denoted as $\overline{h^*}$, is feasible to (5) but achieves a strictly lower cost than h_{x_0}. This contradicts to the optimality of h_{x_0}. $\qquad \square$

D. Verification

1) To show V satisfies Definition 10: It is sufficient to show V is continuous in x. The proof is twofold. To abuse the notations a little bit, we let $h_x(t) \equiv x$ for $x \in \mathcal{X}$, so such h_x is the unique minimizer of (4) and $L(h_x) = V(x) = 0$ for $x \in \mathcal{X}$.

First we show for $x_0 \in \hat{\mathcal{X}}$ and $\epsilon > 0$, there exists $\delta_+ > 0$ such that $\forall x \in \mathcal{B}(x_0, \delta_+) \cap \hat{\mathcal{X}}$, $V(x) \leq V(x_0) + \epsilon$. There are two scenarios. If $h_{x_0}(1)$ is a global optimum of (1), then we could set $\delta_+ = \epsilon$. For any $x \in \mathcal{B}(x_0, \delta_+) \cap \hat{\mathcal{X}}$, construct

$$h^*(t) = \begin{cases} (1 - 2t)x + 2tx_0, & t \in [0, \frac{1}{2}] \\ h_{x_0}(2t - 1), & t \in (\frac{1}{2}, 1]. \end{cases}$$

Its arc-length reparameterization $\overline{h^*}$ is feasible to (4) (w.r.t. x) and $V(x) \leq L(\overline{h^*}) = |x - x_0| + L(h_{x_0}) \leq V(x_0) + \epsilon$. Next we focus on the scenario that $h_{x_0}(1)$ is not a global optimum of (1), so it is not a local optimum neither. By Lemma 5, there is a path $h^\natural$ in $\mathcal{X}$ such that $h^\natural(0) = h_{x_0}(1)$, $f(h^\natural(t))$ is non-increasing in t, $f(h^\natural(1)) < f(h^\natural(0))$ and $L(h^\natural) < \epsilon/2$. Suppose $f(h^\natural(0)) - f(h^\natural(1)) = \tau > 0$. Since f is continuous, there must be some $\gamma > 0$ such that for any $x \in \mathcal{B}(x_0, \gamma) \cap \hat{\mathcal{X}}$, we have $|f(x) - f(x_0)| < \tau$. Now we choose δ_+ as $\min(\gamma, \epsilon/2)$. For any $x \in \mathcal{B}(x_0, \delta_+) \cap \hat{\mathcal{X}}$, construct

$$h^*(t) = \begin{cases} (1 - 3t)x + 3tx_0, & t \in [0, \frac{1}{3}] \\ h_{x_0}(3t - 1), & t \in (\frac{1}{3}, \frac{2}{3}] \\ h^\natural(3t - 2), & t \in (\frac{2}{3}, 1]. \end{cases}$$

Its arc-length reparameterization $\overline{h^*}$ is feasible to (4) (w.r.t. x) and $V(x) \leq L(\overline{h^*}) = |x - x_0| + L(h_{x_0}) + L(h^\natural) \leq \delta_+ + V(x_0) + \epsilon/2 \leq V(x_0) + \epsilon$.

Second we show for $x_0 \in \mathcal{X}$ and $\epsilon > 0$, there exists $\delta_- > 0$ such that $\forall x \in \mathcal{B}(x_0, \delta_-) \cap \hat{\mathcal{X}}$, $V(x) \geq V(x_0) - \epsilon$. If not, then there must be a sequence $(x_i)_{i=1}^{\infty}$ such that $\lim_{i \to \infty} x_i = x_0$

but $V(x_i) < V(x_0) - \epsilon$ for all $i \geq 1$. Let $h_i := h_{x_i}$ for $i \geq 0$, then both $(h_i)_{i=1}^{\infty}$ and $(L(h_i))_{i=1}^{\infty}$ are uniformly bounded. By Lemma 7, a subsequence of $(h_i)_{i=1}^{\infty}$ uniformly converges to a limit h^* and

$$L(h^*) = L(\overline{h^*}) \leq \limsup_i L(h_i)$$
$$= \limsup_i V(x_i) \leq V(x_0) - \epsilon.$$

Lemma 7 also indicates $\overline{h^*} \in \hat{\mathcal{H}}$ and $h^*(0) = \lim_{i \to \infty} h_i(0) = \lim_{i \to \infty} x_i = x_0$, $h^*(1) = \lim_{i \to \infty} h_i(1) \in \mathcal{X}$ (as $\mathcal{X}$ is closed). Therefore, $\overline{h^*}$ is feasible to (4) but its cost is strictly lower than $V(x_0)$. It leads to the contradiction.

2) To show (C3) holds: By our construction (5), h_x is entirely within $\hat{\mathcal{X}}$, and $h_x(0) = x$, $h_x(1) \in \mathcal{X}$. Lemma 10 shows $f(h_x(t))$ is non-increasing for $t \in [0,1]$. It is sufficient to show $V(h_x(t))$ is also non-increasing for $t \in [0,1]$. Consider the following lemma.

Lemma 11. *For fixed $x_0 \in \hat{\mathcal{X}}$ and $t_0 \in [0,1]$,*

$$V(h_{x_0}(t_0)) = \sup_{\pi \in \Pi|_{t_0}^1} L_\pi(h_{x_0})$$

Proof. Let $x_1 = h_{x_0}(t_0)$. We have $L(h_{x_1}) = V(h_{x_0}(t_0))$. If the lemma does not hold, then we have two cases.

First, if $L(h_{x_1}) < \sup_{\pi \in \Pi|_{t_0}^1} L_\pi(h_{x_0})$, then let

$$h^*(t) = \begin{cases} h_{x_0}(2t_0 t), & t \in [0, \frac{1}{2}] \\ h_{x_1}(2t - 1), & t \in (\frac{1}{2}, 1]. \end{cases}$$

It is easy to check h^* is continuous and entirely within $\hat{\mathcal{X}}$, and $h^*(0) = x_0, h^*(1) = h_{x_1}(1) \in \mathcal{X}$. For $t \in [0, 1/2]$,

$$f(h^*(t)) = f(h_{x_0}(2t_0 t)) \geq f(h_{x_0}(t_0)) = f(x_1)$$
$$= f(h_{x_1}(0)) \geq f(h_{x_1}(1)) = f(h^*(1)).$$

For $t \in [1/2, 1]$,

$$f(h^*(t)) = f(h_{x_1}(2t - 1)) \geq f(h_{x_1}(1)) = f(h^*(1)).$$

Further, we have

$$L(h^*) = \sup_{\pi \in \Pi} L_\pi(h^*)$$
$$= \sup_{\pi \in \Pi|_0^{0.5}} L_\pi(h^*) + \sup_{\pi \in \Pi|_{0.5}^1} L_\pi(h^*)$$
$$= \sup_{\pi \in \Pi|_0^{t_0}} L_\pi(h_{x_0}) + \sup_{\pi \in \Pi} L_\pi(h_{x_1})$$
$$= \sup_{\pi \in \Pi|_0^{t_0}} L_\pi(h_{x_0}) + L(h_{x_1})$$
$$< \sup_{\pi \in \Pi|_0^{t_0}} L_\pi(h_{x_0}) + \sup_{\pi \in \Pi|_{t_0}^1} L_\pi(h_{x_0}) = L(h_{x_0}).$$

Above all, the arc-length reparameterization of h^*, denoted as $\overline{h^*}$, is feasible to (5) (w.r.t. x_0) but achieves a strictly lower cost than h_{x_0}. This contradicts to the optimality of h_{x_0}.

Second, if $L(h_{x_1}) > \sup_{\pi \in \Pi|_{t_0}^1} L_\pi(h_{x_0})$, then let

$$h^*(t) = \begin{cases} h_{x_0}(t_0), & t \in [0, t_0] \\ h_{x_0}(t), & t \in (t_0, 1]. \end{cases}$$

It is easy to check h^* is continuous and entirely within $\hat{\mathcal{X}}$, and $h^*(0) = h_{x_0}(t_0) = x_1, h^*(1) = h_{x_0}(1) \in \mathcal{X}$. For $t \in [0,1]$,

$$L(h^*) = \sup_{\pi \in \Pi} L_\pi(h^*) = \sup_{\pi \in \Pi|_0^{t^\ddagger - \delta}} L_\pi(h^*) + \sup_{\pi \in \Pi|_{t^\ddagger - \delta}^{t^\ddagger + \delta}} L_\pi(h^*) + \sup_{\pi \in \Pi|_{t^\ddagger + \delta}^1} L_\pi(h^*)$$

$$= \sup_{\pi \in \Pi|_0^{t^\ddagger - \delta}} L_\pi(h_{x_0}) + \|h_{x_0}(t^\ddagger - \delta) - h_{x_0}(t^\ddagger + \delta)\|_{\ell_2} + \sup_{\pi \in \Pi|_{t^\ddagger + \delta}^1} L_\pi(h_{x_0})$$

$$< \sup_{\pi \in \Pi|_0^{t^\ddagger - \delta}} L_\pi(h_{x_0}) + \|h_{x_0}(t^\ddagger - \delta) - h_{x_0}(t^\ddagger)\|_{\ell_2} + \|h_{x_0}(t^\ddagger) - h_{x_0}(t^\ddagger + \delta)\|_{\ell_2} + \sup_{\pi \in \Pi|_{t^\ddagger + \delta}^1} L_\pi(h_{x_0})$$

$$\leq \sup_{\pi \in \Pi|_0^{t^\ddagger - \delta}} L_\pi(h_{x_0}) + \sup_{\pi \in \Pi|_{t^\ddagger - \delta}^{t^\ddagger + \delta}} L_\pi(h_{x_0}) + \sup_{\pi \in \Pi|_{t^\ddagger + \delta}^1} L_\pi(h_{x_0}) = \sup_{\pi \in \Pi} L_\pi(h_{x_0}) = L(h_{x_0}). \tag{6}$$

$f(h_{x_0}(t)) \geq f(h_{x_0}(1))$ implies $f(h^*(t)) \geq f(h_{x_0}(1)) = f(h^*(1))$. Further, we have

$$L(h^*) = \sup_{\pi \in \Pi} L_\pi(h^*) = \sup_{\pi \in \Pi|_0^{t_0}} L_\pi(h^*) + \sup_{\pi \in \Pi|_{t_0}^1} L_\pi(h^*)$$

$$= 0 + \sup_{\pi \in \Pi|_{t_0}} L_\pi(h_{x_0}) < L(h_{x_1}).$$

Above all, the arc-length reparameterization of h^*, denoted as $\overline{h^*}$, is feasible to (5) (w.r.t. x_1) but achieves a strictly lower cost than h_{x_1}. This contradicts to the optimality of h_{x_1}. $\square$

Using this lemma, we are in a good position to show $V(h_x(t))$ is non-increasing for $t \in [0, 1]$. For any $t_1 < t_2$, we have

$$V(h_x(t_1)) = \sup_{\pi \in \Pi|_{t_1}^1} L_\pi(h_x) = \sup_{\pi \in \Pi|_{t_1}^{t_2}} L_\pi(h_x) + \sup_{\pi \in \Pi|_{t_2}^1} L_\pi(h_x)$$

$$\geq \sup_{\pi \in \Pi|_{t_2}^1} L_\pi(h_x) = V(h_x(t_2)).$$

3) To show (C4) holds: The set $\{L(h_x)\}_{x \in \hat{\mathcal{X}} \setminus \mathcal{X}}$ is uniformly bounded by $\max_{x \in \hat{\mathcal{X}}} |x - x^*|$, which is finite.

To summarize, we have verified that such construction is well defined and satisfies both (C3) and (C4), so Lemma 6 is proved. Since we have shown that Lemma 6 implies Theorem 3, the latter is also proved.

V. OTHER PROPERTIES

A. Constructing from Primitives

Though the previous section guarantees the existence of the Lyapunov-like function and paths under certain conditions, it is not clear how to systematically find or construct them. In this subsection, we show that if one can find the Lyapunov-like function and paths for some primitive problems, then there are natural ways to construct the Lyapunov-like function and paths for new problems built up from those primitives in certain ways. To streamline the notations, we will use the tuple $(f, \mathcal{X})$ to refer to (1) and the tuple $(f, \mathcal{X}, \hat{\mathcal{X}})$ to refer to the problem pair (1), (2). Assume $(V, \{h_x\}_{x \in \hat{\mathcal{X}} \setminus \mathcal{X}})$ is a valid construction of Lyapunov-like function and paths for $(f, \mathcal{X}, \hat{\mathcal{X}})$. In this subsection, when we say V and h_x are valid, it means they not only are valid by definition, but also satisfy (C1) and (C2).

1) Function Composition: Suppose $g : \mathbb{R} \to \mathbb{R}$ is non-decreasing and convex. Then $(V, \{h_x\}_{x \in \hat{\mathcal{X}} \setminus \mathcal{X}})$ is also a valid construction of Lyapunov-like function and paths for $(g \circ f, \mathcal{X}, \hat{\mathcal{X}})$. This result is trivial as $g \circ f$ preserves the convexity over $\hat{\mathcal{X}}$ and monotonicity over any path.

2) Union of Feasible Sets: Suppose for two pairs of problems $(f_1, \mathcal{X}_1, \hat{\mathcal{X}}_1)$, for which $(V^1, \{h_x^1\}_{x \in \hat{\mathcal{X}}_1 \setminus \mathcal{X}_1})$ is valid, and $(f_2, \mathcal{X}_2, \hat{\mathcal{X}}_2)$, for which $(V^2, \{h_x^2\}_{x \in \hat{\mathcal{X}}_2 \setminus \mathcal{X}_2})$ is valid. We consider a new problem $(f, \mathcal{X}, \hat{\mathcal{X}})$ where $\mathcal{X} := (\mathcal{X}_1 \cup \mathcal{X}_2) \cap \hat{\mathcal{X}}_1 \cap \hat{\mathcal{X}}_2$ and $\hat{\mathcal{X}} := \hat{\mathcal{X}}_1 \cap \hat{\mathcal{X}}_2$. The formulation of f will be provided later. If for any $x \in \hat{\mathcal{X}} \setminus \mathcal{X}$, we have $h_x^1 \equiv h_x^2$, then construct $\tilde{V} : \hat{\mathcal{X}} \to \mathbb{R}$ such that $\tilde{V}(x) := V^1(x) \cdot V^2(x)$ and $\tilde{h}_x = h_x^1$ for all $x \in \hat{\mathcal{X}} \setminus \mathcal{X}$. We have the following two results.

Corollary 5. *For any $\lambda \in (0, 1)$, define $f : \hat{\mathcal{X}} \to \mathbb{R}$ as $f(x) := \lambda f_1(x) + (1 - \lambda) f_2(x)$. Then $(\tilde{V}, \{\tilde{h}_x\}_{x \in \hat{\mathcal{X}} \setminus \mathcal{X}})$ is valid for $(f, \mathcal{X}, \hat{\mathcal{X}})$.*

Corollary 6. *Define function $f : \hat{\mathcal{X}} \to \mathbb{R}$ as $f(x) := \max(f_1(x), f_2(x))$. Then $(\tilde{V}, \{\tilde{h}_x\}_{x \in \hat{\mathcal{X}} \setminus \mathcal{X}})$ is valid for $(f, \mathcal{X}, \hat{\mathcal{X}})$.*

Proof for Corollary 5 and Corollary 6. The function $\tilde{V}$ is still continuous and vanishes if and only if $x \in \mathcal{X}$ (since $V(x) = 0 \Leftrightarrow V^1(x) = 0$ or $V^2(x) = 0$). By construction, $\{h_x\}_{x \in \hat{\mathcal{X}} \setminus \mathcal{X}}$ is a subset of $\{h_x^1\}_{x \in \hat{\mathcal{X}}_1 \setminus \mathcal{X}_1}$ so (C2) is naturally satisfied. To see (C1) holds, we fix any $x \in \hat{\mathcal{X}} \setminus \mathcal{X}$. Then $h_x(0) = h_x^1(x) = x$ and $h_x(1) = h_x^1(1) \in \mathcal{X}_1 \subseteq \mathcal{X}$. Further, $V(h_x(t)) = V^1(h_x(t)) V^2(h_x(t)) = V^1(h_x^1(t)) V^2(h_x^2(t))$ as h_x^1 and h_x^2 conincide when $x \in \hat{\mathcal{X}} \setminus \mathcal{X}$. Because both $V^1(h_x^1(t))$ and $V^2(h_x^2(t))$ are non-negative and non-increasing, so is $V(h_x(t))$. Finally, as $f_1(h_x(t))$ and $f_2(h_x(t))$ are both non-increasing over $[0, 1]$, their convex-combination or maximum (i.e., $f(h_x(t))$) must be non-increasing as well. A similar argument can also be applied to show $f(h_x(1)) < f(h_x(0))$. Thus (C1) holds and it completes the proof. $\square$

3) Intersection of Feasible Sets: We still consider two pairs of problems $(f_1, \mathcal{X}_1, \hat{\mathcal{X}})$, for which $(V^1, \{h_x^1\}_{x \in \hat{\mathcal{X}}_1 \setminus \mathcal{X}_1})$ is valid, and $(f_2, \mathcal{X}_2, \hat{\mathcal{X}})$, for which $(V^2, \{h_x^2\}_{x \in \hat{\mathcal{X}}_2 \setminus \mathcal{X}_2})$ is valid. Different from the previous setting, two pairs are required to share the same relaxation set $\hat{\mathcal{X}}$. Further, we view each $x \in \hat{\mathcal{X}}$ as a tuple with two parts $x := (u, v)$. Define $\mathcal{P}_1$ and $\mathcal{P}_2$ as two projection operators such that $\mathcal{P}_1 x = u$ and $\mathcal{P}_2 x = v$.

We consider a new problem $(f, \mathcal{X}, \hat{\mathcal{X}})$ where $\mathcal{X} := \mathcal{X}_1 \cap \mathcal{X}_2$. The formulation of f will be provided later.

If f_i, V^i and h_x^i are completely separated with respect to u and v in the sense that for $i = 1, 2$, $f_i(x), V^i(x)$ depend on $\mathcal{P}_i x$ only and $\mathcal{P}_{1-i}(h_x^i(t))$ is constant, then we can construct $\tilde{V}$ as $\tilde{V}(x) := V^1(x) + V^2(x)$. For $x \in \hat{\mathcal{X}} \setminus \mathcal{X}$, the path $\tilde{h}_x$ is

TABLE I
A SUMMARY ON CONSTRUCTING V AND h_x FROM PRIMITIVES.

Operation	Primitive problem	New problem	V, h_x for new problem	Additional requirements
Function Composition	$(f, \mathcal{X}, \hat{\mathcal{X}})$: V, h_x	$(g \circ f, \mathcal{X}, \hat{\mathcal{X}})$	$\tilde{V} := V$ $\tilde{h}_x := h_x$	g is non-decreasing and convex
Union of feasible sets (with cost as the sum)	$(f_1, \mathcal{X}_1, \hat{\mathcal{X}}_1)$: V^1, h_x^1 $(f_2, \mathcal{X}_2, \hat{\mathcal{X}}_2)$: V^2, h_x^2	$(f, \mathcal{X}, \hat{\mathcal{X}})$ where $f := \lambda f_1 + (1 - \lambda) f_2$ $\mathcal{X} := (\mathcal{X}_1 \cup \mathcal{X}_2) \cap \hat{\mathcal{X}}_1 \cap \hat{\mathcal{X}}_2$ $\hat{\mathcal{X}} := \hat{\mathcal{X}}_1 \cap \hat{\mathcal{X}}_2$	$\tilde{V} := V^1 \times V^2$ $\tilde{h}_x := h_x^1$	h_x^1 and h_x^2 coincide for $x \in \hat{\mathcal{X}} \setminus \mathcal{X}$
Union of feasible sets (with cost as the maximum)	$(f_1, \mathcal{X}_1, \hat{\mathcal{X}}_1)$: V^1, h_x^1 $(f_2, \mathcal{X}_2, \hat{\mathcal{X}}_2)$: V^2, h_x^2	$(f, \mathcal{X}, \hat{\mathcal{X}})$ where $f := \max(f_1, f_2)$ $\mathcal{X} := (\mathcal{X}_1 \cup \mathcal{X}_2) \cap \hat{\mathcal{X}}_1 \cap \hat{\mathcal{X}}_2$ $\hat{\mathcal{X}} := \hat{\mathcal{X}}_1 \cap \hat{\mathcal{X}}_2$	$\tilde{V} := V^1 \times V^2$ $\tilde{h}_x := h_x^1$	h_x^1 and h_x^2 coincide for $x \in \hat{\mathcal{X}} \setminus \mathcal{X}$
Intersection of feasible sets (with cost as the sum)	$(f_1, \mathcal{X}_1, \hat{\mathcal{X}})$: V^1, h_x^1 $(f_2, \mathcal{X}_2, \hat{\mathcal{X}})$: V^2, h_x^2 $(u, v) =: x \in \hat{\mathcal{X}}$	$(f, \mathcal{X}, \hat{\mathcal{X}})$ where $f := \lambda f_1 + (1 - \lambda) f_2$ $\mathcal{X} := \mathcal{X}_1 \cap \mathcal{X}_2$	$\tilde{V} := V^1 + V^2$ $\tilde{h}_x$ is defined as in (7)	$f_i(x), V^i(x)$ depend on $\mathcal{P}_i x$ only and $\mathcal{P}_{1-i} h_x^i$ is constant
Intersection of feasible sets (with cost as the maximum)	$(f_1, \mathcal{X}_1, \hat{\mathcal{X}})$: V^1, h_x^1 $(f_2, \mathcal{X}_2, \hat{\mathcal{X}})$: V^2, h_x^2 $(u, v) =: x \in \hat{\mathcal{X}}$	$(f, \mathcal{X}, \hat{\mathcal{X}})$ where $f := \max(f_1, f_2)$ $\mathcal{X} := \mathcal{X}_1 \cap \mathcal{X}_2$	$\tilde{V} := V^1 + V^2$ $\tilde{h}_x$ is defined as in (7)	$f_i(x), V^i(x)$ depend on $\mathcal{P}_i x$ only and $\mathcal{P}_{1-i} h_x^i$ is constant

constructed in three ways depending on the values of $V^1(x)$ and $V^2(x)$.

If $V^1(x) = 0$ then $\tilde{h}_x := h_x^2$, (7a)

If $V^2(x) = 0$ then $\tilde{h}_x := h_x^1$, (7b)

If $V^1(x), V^2(x) > 0$ then

$$\tilde{h}_x(t) := \begin{cases} h_x^1(2t), & t \in [0, \frac{1}{2}) \\ h_{h_x^1(1)}^2(2t - 1), & t \in [\frac{1}{2}, 1] \end{cases} . \quad (7c)$$

Corollary 7. *For any* $\lambda \in (0, 1)$, *define* $f : \hat{\mathcal{X}} \to \mathbb{R}$ *as* $f(x) := \lambda f_1(x) + (1 - \lambda) f_2(x)$. *Then* $(\tilde{V}, \{\tilde{h}_x\}_{x \in \hat{\mathcal{X}} \setminus \mathcal{X}})$ *is valid for* $(f, \mathcal{X}, \hat{\mathcal{X}})$.

Corollary 8. *Define function* $f : \hat{\mathcal{X}} \to \mathbb{R}$ *as* $f(x) := \max(f_1(x), f_2(x))$. *Then* $(\tilde{V}, \{\tilde{h}_x\}_{x \in \hat{\mathcal{X}} \setminus \mathcal{X}})$ *is valid for* $(f, \mathcal{X}, \hat{\mathcal{X}})$.

Proof for Corollary 7 and Corollary 8. The function $\tilde{V}$ is still continuous and vanishes if and only if $x \in \mathcal{X}$ (since $V(x) = 0 \Leftrightarrow V^1(x) = 0$ *and* $V^2(x) = 0$). The set $\{\tilde{h}_x\}_{x \in \hat{\mathcal{X}} \setminus \mathcal{X}}$ satisfies (C2) as each path is constructed either as h_x^i or the concatenation of h_x^1 and $h_{h_x^1(1)}^2$. Next we are showing $\tilde{h}_x(1)$ is in $\mathcal{X}$. If $\tilde{h}_x$ is constructed by (7a), then we have $\tilde{V}(\tilde{h}_x(1)) = V^1(h_x^2(1)) + V^2(h_x^2(1)) = V^1(h_x^2(1))$. Since $V^1(h_x^2(1))$ only depends on $\mathcal{P}_1 h_x^2(1)$ and $\mathcal{P}_1 h_x^2(1) = \mathcal{P}_1 h_x^2(0) = \mathcal{P}_1 x$, there must be $V^1(h_x^2(1)) = V^1(x) = 0$. Thus $\tilde{V}(\tilde{h}_x(1)) = 0$ and $\tilde{h}_x(1) \in \mathcal{X}$. It is similar if $\tilde{h}_x$ is constructed by (7b). When $\tilde{h}_x$ is constructed by (7c), then

$$\begin{aligned} \tilde{V}(\tilde{h}_x(1)) &= V^1(h_{h_x^1(1)}^2(1)) + V^2(h_{h_x^1(1)}^2(1)) \\ &= V^1(h_{h_x^1(1)}^2(1)) = V^1(h_{h_x^1(1)}^2(0)) \\ &= V^1(h_x^1(1)) = 0, \end{aligned}$$

so $\tilde{h}_x(1) \in \mathcal{X}$ as well. The monotonicity properties of $\tilde{V}(\tilde{h}_x(t))$ and $f(\tilde{h}_x(t))$ are also the direct consequence of the fact that f_i, V^i and h_x^i are completely separated. $\square$

A summary of this subsection has been provided in Table I.

B. Weak Exactness

One observation from the proof of Theorem 1 is we do not actually need $f(h_x(0)) > f(h_x(1))$ to eliminate genuine local optima. However, such strict inequality is required to show the exactness. We can consider a weaker version of exactness defined as follows.

Definition 12 (Weak Exactness). *We say the relaxation* (2) *is weakly exact with respect to* (1) *if at least one optimum of* (2) *is feasible, and hence globally optimal, for* (1).

Theorem 4. *If there exists a Lyapunov-like function* V *associated with* (1) *and* (2) *such that (C3) and (C2) hold, then* (2) *is weakly exact with respect to* (1) *and any local optimum in* $\mathcal{X}$ *for* (1) *is either a global optimum or a pseudo local optimum.*

The argument on weak exactness follows from the fact that the path connects any global optimum of (2) must determine a endpoint in $\mathcal{X}$ with the same cost, which by definition must be a global optimum as well. The argument on local optimality follows directly from the proof of Theorem 1.

VI. APPLICATIONS

In this sections we will use two examples to show for specific problems, what V and $\{h_x\}$ might look like. The first example is Optimal Power Flow (OPF) problem in power systems with tree structres, which is also the motivating problem for us to develop this theory. By finding the Lyapunov-like function and paths, we show in [1] the first known condition (that can be checked *a priori*) for OPF to have no spurious local optima. The same condition was only known to guarantee exact relaxation before our work.

In the second example, we study the Low Rank Semidefinite Program (LRSDP) problem, which was known to have weakly exact relaxation [5], [6] and no spurious local optima [7] in existing literatures. Specifically, we show that part of the results proved in [7] can also be proved by finding appropriate V and $\{h_x\}$. They exemplify the usage of Theorem 1, Theorem 2 and Theorem 4 in practice.

A. Optimal Power Flow

Consider a radial 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.

As we assume $\mathcal{G}$ is a tree, we can adopt the DistFlow Model [36], [37] 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. Let $v_j := |V_j|^2 \in \mathbb{R}$ and $\ell_{jk} := |I_{jk}|^2 \in \mathbb{R}$. We will denote the conjugate of a complex number a by a^{H}.

The power flow equations are:

$$v_j = v_k + 2\mathrm{Re}(z_{jk} S_{jk}^{\mathsf{H}}) - |z_{jk}|^2 \ell_{jk}, \quad \forall (j, k) \in \mathcal{E} \quad (8a)$$

$$v_j = \frac{|S_{jk}|^2}{\ell_{jk}}, \quad \forall (j, k) \in \mathcal{E} \quad (8b)$$

$$s_j = \sum_{k:j \to k} S_{jk} - \sum_{i:i \to j} (S_{ij} - z_{ij} \ell_{ij}), \quad \forall j \in \mathcal{V}. \quad (8c)$$

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

$$\begin{aligned} &\underset{x=(s,v,\ell,S)}{\text{minimize}} && f(s) && (9a)\\ &\text{subject to} && (8) && (9b)\\ &&& \underline{v}_j \le v_j \le \overline{v}_j && (9c)\\ &&& \underline{s}_j \le s_j \le \overline{s}_j && (9d)\\ &&& \ell_{jk} \le \overline{\ell}_{jk} && (9e) \end{aligned}$$

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

Definition 13. *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*

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

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 (9) is feasible.
(iv) The line current limit satisfies $\overline{\ell}_{jk} \le \underline{v}_j |y_{jk}|^2$.

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., [38], [9]). 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 (9) is non-convex, as constraint (8b) is not convex. Denote by $\mathcal{X}$ the set of (s, v, ℓ, S) that satisfy (9b)-(9e), so (9) is in the form of (1). We can relax (9) by convexifying (8b) into a second-order cone [8]:

$$\begin{aligned} &\underset{x=(s,v,\ell,S)}{\text{minimize}} && f(s) && (10a)\\ &\text{subject to} && (8a), (8c), (9c) - (9e) && (10b)\\ &&& |S_{jk}|^2 \le v_j \ell_{jk} && (10c) \end{aligned}$$

One can similarly regard $\hat{\mathcal{X}}$ as the set of (s, v, ℓ, S) that satisfy (10b), (10c). It is proved in [8] that if $\underline{s}_j = -\infty - \mathbf{i}\infty$ for all $j \in \mathcal{V}$, then (10) is exact, meaning any optimal solution of (10) is also feasible and hence globally optimal for (9). Now we show that the same condition also guarantees that any local optimum of (9) 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 5. *If* $\underline{s}_j = -\infty - \mathbf{i}\infty$ *for all* $j \in \mathcal{V}$, *then any local optimum of* (9) *is a global optimum.*

Proof. Our strategy is to construct appropriate V and $\{h_x\}$ and then prove such construction satsify both Condition (C) and Condition (C'). Let

$$V(x) := \sum_{(j,k) \in \mathcal{E}} v_j \ell_{jk} - |S_{jk}|^2. \quad (11)$$

Clearly, V is a valid Lyapunov-like function satisfying Definition 10.

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

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

must have a unique positive root as $\phi_{jk}(0) < 0$. We define Δ_{jk} to be this positive root if $(j, k) \in \mathcal{M}$ and 0 otherwise.

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

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

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

$$\tilde{s}_j(t) = s_j - \frac{t}{2} \sum_{i:i \to j} z_{ij} \Delta_{ij} - \frac{t}{2} \sum_{k:j \to k} z_{jk} \Delta_{jk}, \quad (12a)$$

$$\tilde{v}_j(t) = v_j, \quad (12b)$$

$$\tilde{\ell}_{jk}(t) = \ell_{jk} - t\Delta_{jk}, \quad (12c)$$

$$\tilde{S}_{jk}(t) = S_{jk} - \frac{t}{2} z_{jk} \Delta_{jk}. \quad (12d)$$

Clearly we have $h_x(0) = x$. It can be easily checked that $h_x(t)$ is feasible for (10) for $t \in [0,1]$ and $h_x(1)$ is feasible for (9) (see [1]). 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

$$\sum_{(j,k)\in\mathcal{E}} \tilde{v}_j(t)\tilde{\ell}_{jk}(t) - |\tilde{S}_{jk}(t)|^2$$

$$= \sum_{(j,k)\in\mathcal{M}^c} v_j\ell_{jk} - |S_{jk}|^2 + \sum_{(j,k)\in\mathcal{M}} \tilde{v}_j(t)\tilde{\ell}_{jk}(t) - |\tilde{S}_{jk}(t)|^2$$

$$= \sum_{(j,k)\in\mathcal{M}^c} v_j\ell_{jk} - |S_{jk}|^2 - \sum_{(j,k)\in\mathcal{M}} \phi_{jk}(t\Delta_{jk}).$$

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 Corollary 1, 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 $0 \le a < b \le 1$,

$$f(\tilde{s}(a)) - f(\tilde{s}(b))$$

$$\ge c \sum_{j\in\mathcal{V}} \text{Re}(\tilde{s}_j(a) - \tilde{s}_j(b)) + \text{Im}(\tilde{s}_j(a) - \tilde{s}_j(b))$$

$$= c\|\tilde{s}(a) - \tilde{s}(b)\|_{\text{m}}$$

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

On the other hand, by (12) we have $\|\tilde{v}(a) - \tilde{v}(b)\|_{\text{m}} \equiv 0$ and

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

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

Therefore,

$$\|h_x(a) - h_x(b)\|_{\text{m}} \le \left(\frac{3}{2} + \frac{1}{\min_{(j,k)\in\mathcal{E}}\{\|z_{jk}\|_{\text{m}}\}}\right)\|\tilde{s}(a) - \tilde{s}(b)\|_{\text{m}}$$

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

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

Therefore Condition (C') is also satisfied and by Theorem 2, any local optimum of (9) is a global optimum. $\square$

The results in this subsection only apply to radial networks, which serve as the underlying network for balanced distribution power systems. For transmission systems and unbalanced distribution systems, networks are usually highly meshed. It has been found that for most of meshed networks, both convex relaxation and local search algorithms can also yield the global optimum for most of testcases [39], [40]. Thus Theorem 3

suggests that there may also exist similar Lyapunov-like function and paths for meshed networks. Finding such Lyapunov-like function and paths would be an interesting future work to extend our results in this paper.

B. Low Rank Semidefinite Program

This subsection proves a known result in [7] but using a different approach. Adopting the same notations as [7], we have the following problem.

$$\begin{align} \underset{X\ge 0}{\text{minimize}} \quad & \text{tr}(CX) \tag{13a}\\ \text{subject to} \quad & \text{tr}(A_iX) = b_i, \quad i = 1,\cdots,m \tag{13b}\\ & \text{rank}(X) \le r \tag{13c} \end{align}$$

Here, C, A_i, X are all n-by-n matrices. We assume the problem is feasible and $\{X \ge 0 \mid (13b)\}$ is compact.

Theorem 6. *If $(r+1)(r+2)/2 > m+1$, then any local optimum of* (13) *is either a global optimum or a pseudo local optimum.*

Before proving Theorem 6, we consider the convex relaxation of (13) as

$$\begin{align} \underset{X\ge 0}{\text{minimize}} \quad & \text{tr}(CX) \tag{14a}\\ \text{subject to} \quad & \text{tr}(A_iX) = b_i, \quad i = 1,\cdots,m \tag{14b} \end{align}$$

As a side note, the results in [5], [6] show that if $(r+1)(r+2)/2 > m$, then (14) is weakly exact to (13). While our theorem is the same as in [7], some insights to find V and $\{h_X\}$ are also from the structures first raised in [5], [6].

Proof. Clearly, (13) can be reformulated in the form of (1) by setting $f(X) = \text{tr}(CX)$, $\mathcal{X} = \{X \ge 0 \mid (13b),(13c)\}$ and $\hat{\mathcal{X}} = \{X \ge 0 \mid (13b)\}$. Define V as

$$V(X) := \sum_{i=r+1}^{n} \lambda_i(X)$$

where $\lambda_i(X)$ is the ith eigenvalue of X (in decreasing order). This function V satisfies Definition 10 and is concave.

For fixed $X \in \hat{\mathcal{X}} \setminus \mathcal{X}$, we denote $\text{rank}(X)$ as $r_0 > r$. We first construct $r_0 - r$ paths labeled as $h_1, h_2, \cdots, h_{r_0-r}$. When we construct h_i, if $i > 1$ then we assume path h_{i-1} has already been constructed and let $X_{i-1} := h_{i-1}(1)$. We let $X_0 = X$. For $i \ge 1$, if $\text{rank}(X_{i-1}) \le r_0 - i$ then we let $h_i(t) \equiv X_{i-1}$ for $t \in [0,1]$. Otherwise, we decompose X_{i-1} as $U\Sigma U^{\text{H}}$ where Σ is a k-by-k positive definite diagonal matrix with $k = \text{rank}(X_{i-1}) > r_0 - i$. The linear system

$$\begin{align} \text{tr}(CUYU^{\text{H}}) &= 0\\ \text{tr}(A_iUYU^{\text{H}}) &= 0, \quad i = 1,\cdots,m \tag{15} \end{align}$$

must have a non-zero solution for Hermitian matrix $Y \in \mathbb{C}^{k\times k}$. To see this, we have $k \ge r_0 - i + 1 \ge r+1$, and thus $k(k+1)/2 \ge (r+1)(r+2)/2 > m+1$. As a result, (15) has more unknown variables than equations. We simply denote this non-zero solution as Y and for any $\alpha \in \mathbb{R}$, αY is also a solution to (15). The concavity of V also implies that

$V(U(\Sigma + \alpha Y)U^{\mathsf{H}})$ is concave in α when U and Σ are fixed. Since $\Sigma > 0$, one of the following two scenarios must be true.

- $\exists a < 0$ such that $V(U(\Sigma + \alpha Y)U^{\mathsf{H}})$ is non-decreasing, $\mathrm{rank}(U(\Sigma + \alpha Y)U^{\mathsf{H}}) \leq k$ for $\alpha \in [a, 0]$ and $\mathrm{rank}(U(\Sigma + aY)U^{\mathsf{H}}) \leq k - 1$.
- $\exists b > 0$ such that $V(U(\Sigma + \alpha Y)U^{\mathsf{H}})$ is non-increasing, $\mathrm{rank}(U(\Sigma + \alpha Y)U^{\mathsf{H}}) \leq k$ for $\alpha \in [0, b]$ and $\mathrm{rank}(U(\Sigma + bY)U^{\mathsf{H}}) \leq k - 1$.

Without loss of generality, we suppose $V(U(\Sigma + \alpha Y)U^{\mathsf{H}})$ is non-increasing for $\alpha \in [0, b]$ (otherwise we take $-Y$ instead). We then construct h_i as $h_i(t) = U(\Sigma + tbY)U^{\mathsf{H}}$ for $t \in [0, 1]$. By construction, $V(h_i(t))$ is non-increasing and $f(h_i(t))$ stays a constant.

Finally, we construct h_X as the concatenation of paths $h_1, \cdots, h_{r_0-r}$. That is to say,

$$h_X(t) := h_i((r_0 - r)t - i + 1) \text{ for } t \in \left[\frac{i-1}{r_0 - r}, \frac{i}{r_0 - r}\right].$$

It is easy to see h_X is continuous and $h_X(0) = h_1(0) = X$. To see $h_X(1) \in \mathcal{X}$, we prove that $\mathrm{rank}(X_i) \leq r_0 - i$. We first have $\mathrm{rank}(X_0) = \mathrm{rank}(X) = r_0$. For $i \geq 1$, we have $\mathrm{rank}(X_i) = \mathrm{rank}(X_{i-1})$ if $\mathrm{rank}(X_{i-1}) \leq r_0 - i$ and $\mathrm{rank}(X_i) \leq \mathrm{rank}(X_{i-1}) - 1$ otherwise. By induction, we can prove $\mathrm{rank}(X_i) \leq r_0 - i$ always holds. As a result, $\mathrm{rank}(h_X(1)) = \mathrm{rank}(h_{r_0-r}(1)) \leq r$ and thus $h_X(1) \in \mathcal{X}$. By construction, $h_i(t)$ never violates (13b) and thus is in $\hat{\mathcal{X}}$, so is $h_X(t)$ for all t. Functions $V(h_i(t))$ and $f(h_i(t))$ being non-increasing implies that $V(h_X(t))$ and $f(h_X(t))$ are also non-increasing. Therefore, (C3) is satisfied. By Corollary 1 (C2) also holds for $\{\overline{h_X}\}$. It completes the proof (by Theorem 4). $\square$

Remark 4. *In [7], Theorem 3.4 claims that any local optimum of* (13) *should also be globally optimal, unless it is harbored in some positive-dimensional face of SDP. The result in our paper further asserts that if it is indeed harbored in such a face, then there must be some point on the edge of this face whose cost can be further reduced in its neighborhood (i.e., the local optimum is in the same situation as point c rather than d as in Fig. 1).*

VII. Conclusion and Discussion

TABLE II
Sufficient and necessary conditions

Condition	Relaxation exactness	Local optimality
Sufficient conditions: $\Rightarrow$		
(C1), (C2)	Strong exactness	l.o. is p.l.o. or g.o.
(C3), (C2)	Weak exactness	l.o. is p.l.o. or g.o.
(C')	Strong exactness	l.o. is g.o.
Necessary condition: $\Leftarrow$		
(C1), (C2)	Strong exactness	l.o. is g.o.

Table II summaries both sufficient and necessary conditions for non-convex problem (1) to simultaneously have exact (weak or strong) relaxation and no spurious local optima (allowing or not allowing pseudo local optima). The necessary condition relies on Assumption 1, which is usually true for

real-world problems. Those results provide a new perspective to certify a non-convex problem is computationally easy to solve. Furthermore, whenever the problem is indeed computationally easy, the certificates (Lyapunov-like functions and paths) are guaranteed to exist. We also provide a hierarchical framework which shows how such certificates for a complicated problem can be constructed from primitive problems. Our results have been applied to OPF and LRSDP problems.

Based on the examples shown in Section VI, a natural way to apply this approach is to first look at existing results on exact relaxation, and then construct V and $\{h_x\}$ according to the hidden structure underlying the exactness. Once V and $\{h_x\}$ are appropriately constructed, our result can help extend existing results on relaxation exactness to new results on local optimality.

Compared to some existing techniques to study local optimality, our results do not require differentiating or analyzing the curvature of feasible sets. It allows the feasible sets to incorporate more complicated and possibly non-convex constraints. Those non-convex constraints are common for problems arising in cyber physical systems which are generally governed by physical laws.

References

[1] F. Zhou and S. H. Low, "A sufficient condition for local optima to be globally optimal," in *To appear in Proc. of the 2020 Conference on Decision and Control*. IEEE, 2020.

[2] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, no. 6, p. 717, 2009.

[3] 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.

[4] 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.

[5] A. I. Barvinok, "Problems of distance geometry and convex properties of quadratic maps," *Discrete & Computational Geometry*, vol. 13, no. 2, pp. 189–202, 1995.

[6] G. Pataki, "On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues," *Mathematics of operations research*, vol. 23, no. 2, pp. 339–358, 1998.

[7] 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.

[8] 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.

[9] 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.

[10] J. Lavaei and S. H. Low, "Zero duality gap in optimal power flow problem," *IEEE Transactions on Power Systems*, vol. 27, no. 1, pp. 92–107, 2012.

[11] J. Jaldén, C. Martin, and B. Ottersten, "Semidefinite programming for detection in linear systems-optimality conditions and space-time decoding," in *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03).*, vol. 4. IEEE, 2003, pp. IV–9.

[12] C. Lu, Y.-F. Liu, W.-Q. Zhang, and S. Zhang, "Tightness of a new and enhanced semidefinite relaxation for mimo detection," *SIAM Journal on Optimization*, vol. 29, no. 1, pp. 719–742, 2019.

[13] Z. Li, Q. Guo, H. Sun, and J. Wang, "Sufficient conditions for exact relaxation of complementarity constraints for storage-concerned economic dispatch," *IEEE Transactions on Power Systems*, vol. 31, no. 2, pp. 1653–1654, 2015.

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

[15] 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.

[16] 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.

[17] N. Boumal, "Nonconvex phase synchronization," *SIAM Journal on Optimization*, vol. 26, no. 4, pp. 2355–2377, 2016.

[18] 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.

[19] J. Carpentier, "Contribution to the economic dispatch problem," *Bulletin de la Societe Francoise des Electriciens*, vol. 3, no. 8, pp. 431–447, 1962.

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

[21] 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.

[22] J. A. Momoh, R. Adapa, and M. El-Hawary, "A review of selected optimal power flow literature to 1993. I. nonlinear and quadratic programming approaches," *IEEE transactions on power systems*, vol. 14, no. 1, pp. 96–104, 1999.

[23] J. A. Momoh, M. El-Hawary, and R. Adapa, "A review of selected optimal power flow literature to 1993. II. newton, linear programming and interior point methods," *IEEE Transactions on Power Systems*, vol. 14, no. 1, pp. 105–111, 1999.

[24] 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.

[25] R. A. Jabr, "Radial distribution load flow using conic programming," *IEEE transactions on power systems*, vol. 21, no. 3, pp. 1458–1459, 2006.

[26] X. Bai, H. Wei, K. Fujisawa, and Y. Wang, "Semidefinite programming for optimal power flow problems," *Int'l J. of Electrical Power & Energy Systems*, vol. 30, no. 6-7, pp. 383–392, 2008.

[27] S. Bose, D. F. Gayme, K. M. Chandy, and S. H. Low, "Quadratically constrained quadratic programs on acyclic graphs with application to power flow," *IEEE Transactions on Control of Network Systems*, vol. 2, no. 3, pp. 278–287, 2015.

[28] 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.

[29] D. K. Molzahn and I. A. Hiskens, "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.

[30] V. A. Toponogov, *Differential geometry of curves and surfaces*. Springer, 2006.

[31] E. Bierstone and P. D. Milman, "Semianalytic and subanalytic sets," *Publications Mathématiques de l'Institut des Hautes Études Scientifiques*, vol. 67, no. 1, pp. 5–42, 1988.

[32] E. Bierstone, "Differentiable functions," *Boletim da Sociedade Brasileira de Matemática-Bulletin/Brazilian Mathematical Society*, vol. 11, no. 2, pp. 139–189, 1980.

[33] R. Hardt, "Some analytic bounds for subanalytic sets," *Differential geometric control theory, Progress in Math*, vol. 27, pp. 259–267, 1983.

[34] A. M. Gabrièlov, "Projections of semi-analytic sets," *Functional Analysis and its applications*, vol. 2, no. 4, pp. 282–291, 1968.

[35] S. Łojasiewicz, "On semi-analytic and subanalytic geometry," *Banach Center Publications*, vol. 34, no. 1, pp. 89–104, 1995.

[36] 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.

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

[38] 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.

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

[40] S. Gopinath, H. Hijazi, T. Weißer, H. Nagarajan, M. Yetkin, K. Sundar, and R. Bent, "Proving global optimality of acopf solutions," *Electric Power Systems Research*, vol. 189, p. 106688, 2020.