Hybrid Methods in Solving Alternating-Current Optimal Power Flows

Jie Liu, Alan Claude Liddell, Jr., Jakub Mareček, and Martin Takáč

***Abstract*—Many steady-state problems in power systems, including rectangular power-voltage formulations of optimal power flows in the alternating-current model, can be cast as polynomial optimization problems (POP). For a POP, one can derive strong convex relaxations, or rather hierarchies of increasingly strong, but increasingly computationally challenging convex relaxations. We study means of switching from solving a convex relaxation to Newton's method working on a non-convex (augmented) Lagrangian of the POP.**

***Index Terms*—α-β theory, numerical analysis (mathematical programming), optimization, power system analysis computing.**

I. Introduction

The ALTERNATING-CURRENT optimal power flow problem (ACOPF) is one of the best known non-convex non-linear optimization problems, studied extensively since the 1960s [17], [30], [31], [39]. Early work focused on straightforward applications of Newton's method [43] to the non-convex problem, which produced exceptionally fast routines, albeit without any guarantees as to their global convergence. Inspired by Lavaei and Low [27], much recent work has focussed on iterative strengthening of convexifications of ACOPF by the iterated addition of variables and constraints [17], [21], [22], [36], [37]. Such iterative strengthening produces a hierarchy of relaxations that converges to the global optimum of the non-convex problem, asymptotically, under mild conditions, but at a considerable computational cost. It has not been clear how to obtain solutions fast, while benefitting from the convergence guarantees associated with the convexifications.

The shortcomings of the two approaches seem inherent in the non-convexity of the problem. Newton's method exhibits local quadratic convergence on non-convex problems. When one starts from an initial point outside of a neighbourhood of a stationary point, Newton's method may diverge and produce no feasible solution. Even within the neighbourhood, where Newton's method converges, the stationary point may turn out to be very far [6], [29] from the global optimum.

On the other hand, solving strong convexifications, such as the semidefinite programming (SDP) relaxations of [17], remains challenging, computationally. Leading second-order methods for solving the SDP relaxations, such as SeDuMi [42], often converge within dozens of iterations on SDP relaxations of even the largest available instances available, but the run-time and memory requirements of a single iteration may be prohibitively large. One may employ first-order methods [32], [34], whose memory requirements and per-iteration run-times are trivial, but whose rates of convergence are, unfortunately, linear or worse. Either way, as one progresses in the hierarchy of convexifications, the run-time to reach an acceptable accuracy grows fast.

To address this challenge, we introduce novel means of combining solvers working on a convexification and solvers working on the non-convex problem. We employ a first-order method [34] in solving a convexification, until we can guarantee local convergence of Newton's method on the non-convex Lagrangian of the problem, possibly considering some regularisation [34]. In particular, the guarantee considers points z_0 and z^*, such that when we start Newton's method or a similar algorithm at the point z_0, it will generate a sequence of points z_i converging to z^* with quadratic rate of convergence, i.e.,

$$\|z_i - z^*\| \leq (1/2)^{2^i - i} \|z_0 - z^*\|. \qquad (1)$$

The associated test requires the knowledge of the Lagrangian and its partial derivatives at z_0, but does not require the computation of z_i for $i > 0$, or solving of any additional optimization problems. This could be seen as means of on-the-fly choice of the solver, which preserves the global convergence guarantees associated with convexification, whilst improving upon the convergence rates of first-order methods.

The paper is organised as follows: Section II presents two key results from the past two decades of study of convergence properties of Newton's method on systems of polynomial equalities and illustrates them on the alternating-current power flows (ACPF). Section III presents our approach to polynomial optimization problems, in general, and alternating-current optimal power flows (ACOPF), in particular, with key results in Section III-C. Section IV describes our implementation and presents its computational performance on IEEE test cases, as well as a model of the transmission system of Poland. Section V puts this in the context of related work, while Section VI concludes the paper with suggestions for further research.

Manuscript received November 1, 2016; revised February 20, 2017 and May 22, 2017; accepted June 1, 2017. Date of publication June 13, 2017; date of current version October 19, 2017. The work of M. Takáč was supported by the National Science Foundation under Grant NSF:CCF:161871. Paper no. TSG-01508-2016. *(Corresponding author: Jakub Mareček.)*

J. Liu and M. Takáč are with Lehigh University, Bethlehem, PA 18015 USA.

A. C. Liddell, Jr., is with the University of Notre Dame, South Bend, IN 46556 USA.

J. Mareček is with IBM Research, D7 Dublin, Ireland (e-mail: jakub.marecek@ie.ibm.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSG.2017.2715282

II. A Brief Overview of α-β Theory

Our approach is based on the α-β theory of Smale [7], [10], [40], which is also known as the point estimation theory. We present some basic results of the theory first.

Consider a general real-valued polynomial system $f : \mathbb{R}^m \mapsto \mathbb{R}^n$, i.e., a system of polynomial equations $f := (f_1, \ldots, f_n)$ in variables $x := (x_1, \ldots, x_m) \in \mathbb{R}^m$. Let us define Newton's operator at $x \in \mathbb{R}^m$ as

$$N_f(x) := x - [\nabla f(x)]^\dagger f(x),$$

where $[\nabla f(x)]^\dagger \in \mathbb{R}^{m\times n}$ is the Moore-Penrose inverse of the Jacobian matrix of f at x. A sequence with initial point x_0 and iterates of Newton's method subsequently, $x_{i+1} := N_f(x_i)$ for $i \geq 0$, is well-defined if $[\nabla f(x_i)]^\dagger$ is well defined at all $x_i, i \geq 0$. We say that $x \in \mathbb{R}^m$ is an *approximate zero* of f if and only if

1) the sequence $\{x_i\}$ is well-defined; and
2) there exists $x' \in \mathbb{R}^m$ such that $f(x') = 0$ and $\|x_i - x'\| \leq (1/2)^{2^{i-1}-i}\|x_0 - x'\|$ for all $i \geq 0$.

We call $x' \in \mathbb{R}^m$ the *associated zero* of $x \in \mathbb{R}^m$ and say that x *represents* x'. The key result of α-β theory is as follows.

Proposition 1 [40]: Let $f : \mathbb{R}^m \mapsto \mathbb{R}^n$ be a system of polynomial equations and define functions $\alpha(f, x)$, $\beta(f, x)$, $\gamma(f, x)$ as:

$$\alpha(f, x) := \beta(f, x)\gamma(f, x), \tag{2a}$$

$$\beta(f, x) := \left\|[\nabla f(x)]^\dagger f(x)\right\| = \|x - N_f(x)\|, \tag{2b}$$

$$\gamma(f, x) := \sup_{k>1}\left\|\frac{[\nabla f(x)]^\dagger\left[\nabla^{(k)} f\right](x)}{k!}\right\|^{1/(k-1)}, \tag{2c}$$

where $[\nabla f(x)]^\dagger \in \mathbb{R}^{m\times n}$ is the Moore-Penrose inverse of the Jacobian matrix of f at x and $[\nabla^{(k)} f]$ is the symmetric tensor whose entries are the k-th partial derivatives of f at x. Then there is a universal constant $\alpha_0 \in \mathbb{R}$ such that if $\alpha(f, x) \leq \alpha_0$, then x is an approximate zero of f. Moreover, if x' denotes its associated zero, then $\|x - x'\| \leq 2\beta(f, x)$. It can be shown that $\alpha_0 = \frac{13-3\sqrt{17}}{4} \approx 0.157671$ satisfies this property.

We refer to [10] and [40] for the proof and a variety of extensions. Considering that [40] is somewhat difficult to read and a part of a five-paper series, we refer to the survey of Cucker and Smale [10] or the more recent survey of Beltrán and Pardo [3] for an overview.

Let us illustrate the approach on alternating-current power flows (ACPF), where the instance is defined by:

- a graph, where n vertices are partitioned into pv (representing buses with generators), pq (representing buses with loads), and $\{\mathcal{S}\}$ (representing the slack bus), and where adjacency of buses i and j is denoted $i \sim j$, and
- the admittance matrix $Y \in \mathbb{C}^{n\times n}$, with $G := \mathrm{Re}(Y)$, $B := \mathrm{Im}(Y)$,
- active loads and injections P_i at buses $i \in \mathrm{pq} \cup \mathrm{pv}$ and reactive loads Q_i at buses $i \in \mathrm{pq}$,
- voltage magnitude setpoints v_i at buses $i \in \mathrm{pv}$.

Following [14], we define the power-flow operator $F : \mathbb{R}^{2n} \mapsto \mathbb{R}^{2n}$ in terms of complex voltages $V_i = V_i^x + \imath V_i^y, i \in V$, with V^c stacked as $V_i^c = V_i^x, V_{n+i}^c = V_i^y$:

$$\begin{aligned}[F\{V^c\}]_i := {} & G_{ii}\left\{\left(V_i^x\right)^2 + \left(V_i^y\right)^2\right\} \\ & - \sum_{j\sim i} B_{ij}\left\{V_i^y V_j^x - V_i^x V_j^y\right\} \\ & - \sum_{j\sim i} G_{ij}\left\{V_i^x V_j^x + V_i^y V_j^y\right\} - P_i, i \in \mathrm{pv} \cup \mathrm{pq}\end{aligned} \tag{3a}$$

$$\begin{aligned}\left[F\{V^c\}\right]_{n+i} := {} & B_{ii}\left\{\left(V_i^x\right)^2 + \left(V_i^y\right)^2\right\} \\ & + \sum_{j\sim i} B_{ij}\left\{V_i^x V_j^x + V_i^y V_j^y\right\} \\ & + \sum_{j\sim i} G_{ij}\left\{V_i^y V_j^x - V_i^x V_j^y\right\} - Q_i, i \in \mathrm{pq}\end{aligned} \tag{3b}$$

$$\left[F\{V^c\}\right]_{n+i} := \left(V_i^x\right)^2 + \left(V_i^y\right)^2 - v_i^2, i \in \mathrm{pv}. \tag{3c}$$

Whether a point $x \in \mathbb{R}^m$ is in a domain of monotonicity can be tested by the simple comparison of α and α_0.

Proposition 2: For every instance of ACPF, there exists a universal constant $\alpha_0 \in \mathbb{R}$ and a function α of the instance of ACPF and a vector $x \in \mathbb{R}^m$ such that if $\alpha(F, x) \leq \alpha_0$, then x is an approximate zero of F.

Proof: One can either apply Proposition 1 to a problem in V^c, which stacks the real and imaginary parts of the complex-valued vector to obtain a real-valued problem, or one may apply an extension of the proposition to complex-valued polynomials, such as [13, Th. 4.3]. ■

Obviously, one needs to compute β (2b) and γ (2c) to compute α (2a). Because $\gamma(f, x)$ is difficult to compute in practice, we wish to establish a bound, e.g., when $m = n$. Let us first define some auxiliary quantities, which will be used in the following proposition. Define a pseudo-norm $\|\cdot\|_1$ on $\mathbb{R}^n$, along with an auxiliary diagonal matrix $\Delta_{(d)}$:

$$\|x\|_1 := \left(1 + \sum_{i=1}^n |x_i|^2\right)^{1/2}, \quad \Delta_{(d)}(x)_{i,i} := d_i^{1/2}\|x\|_1^{d_i-1},$$

where $d_i := \deg(f_i)$ is the degree of f_i in the system of polynomials $f = (f_1, \ldots, f_n)$. Let us consider the degree-d polynomial $g(x) := \sum_{|\nu|_p\leq d} g_\nu x^\nu$ where the coefficients $g_\nu \in \mathbb{R}$ and $x^\nu := x_1^{\nu_1}\cdots x_n^{\nu_n}$ with $|\nu|_p := \sum_{i=1}^n \nu_i$. We can define the following norm:

$$\|g\|_p^2 := \sum_{|\nu|_p\leq d} |g_\nu|^2 \frac{\nu!(d - |\nu|)!}{d!}, \tag{4}$$

where $\nu! := \prod_{i=1}^n \nu_i!$, which, in turn, makes it possible to define a norm on the polynomial system f:

$$\|f\|_p^2 := \sum_{i=1}^n \|f_i\|_p^2.$$

Finally, we define:

$$\mu(f, x) := \max\{1, \|f\|_p \cdot \|[\nabla f(x)]^\dagger \Delta_{(d)}(x)\|\}.$$

With these quantities, we arrive at the following proposition bounding $\gamma(f, x)$.

Proposition 3 [40]: Let $f : \mathbb{R}^n \mapsto \mathbb{R}^n$ be a polynomial system $f := (f_1, \dots, f_n)$ with degree $d_i := \deg(f_i), i \in \{1, 2, \dots, n\}$ and $D := \max_{i \in \{1,2,\dots,n\}}\{d_i\}$. If $x \in \mathbb{R}^n$ such that $[\nabla f(x)]$ is invertible, then

$$\gamma(f, x) \leq \frac{\mu(f, x)D^{3/2}}{2\|x\|_1}. \tag{5}$$

Notice that the proposition assumes a polynomial system, rather than a polynomial optimization problem.

III. The Theory

We extend the approach to polynomial optimization problems (POP). Considering the recent insights [26] into the availability and strength of certain Lagrangian relaxations of a POP, we derive a test, whereby knowing only the relaxation and its derivatives at a particular point, we can decide whether one can switch to Newton's method on the polynomial relaxation. Although there are many options for implementing the test, we suggest tracking the active set and waiting until it stabilises. Then, one may consider a polynomial, in whose construction inequalities in the active set are treated as equalities, while the remaining inequalities are disregarded. Notice that unless one runs Newton's method on that very polynomial, one may need to back-track, whenever the active set changes while running Newton's method.

A. The Preliminaries

In order to describe the approaches formally, we introduce some notation. Let us denote the polynomial ring over the reals by $\mathbb{R}[x]$ and consider the compact basic semi-algebraic set:

$$\begin{aligned}\mathbf{K} := \Big\{ x \in \mathbb{R}^m : g_j(x) \geq 0, \quad j = 1, \dots, p,\\ h_k(x) =, 0, \quad k = 1, \dots, q \Big\}\end{aligned} \tag{6}$$

for some $g_j \in \mathbb{R}[x]$, $j = 1, \dots, p$ in $x \in \mathbb{R}^m$, $h_k \in \mathbb{R}[x]$, $k = 1, \dots, q$. The corresponding polynomial optimization problem is:

$$POP: \quad f^* := \min_{x \in \mathbb{R}^m}\{f(x) : x \in \mathbf{K}\} \tag{7}$$

where $f \in \mathbb{R}[x]$ is the objective function. We use f^* to denote the value of the objective function f at the optimum of the POP (7); notice that there need not be a unique point at which f^* is attained. We use $\mathbb{P}^m$ to denote the space of all possible descriptions of a POP (7) in dimension m, and $T(x)$ to denote a measure of infeasibility of $\mathbb{R}^m$:

$$T(x) := \sum_{j=1}^{p} \min\{0, g_j(x)\}^2 + \sum_{k=1}^{q} h_k(x)^2 \tag{8}$$

in keeping with [34]. For additional background material on polynomial optimization, we refer to [2].

In a departure from the tradition, we use the term *Lagrangian* loosely, to mean a function $\tilde{L} : \mathbb{R}^{\tilde{m}} \mapsto \mathbb{R}, \tilde{m} > m$ associated with a particular instance of a POP (7) in $\mathbb{R}^m$. In the best-known example, one has $\tilde{m} = m+p+q$ and $\tilde{x} \in \mathbb{R}^{\tilde{m}}$ is the concatenation of the original variable $x \in \mathbb{R}^m$ and the so called Lagrangian coefficients λ, κ associated with the constraints:

$$L(x, \lambda, \kappa) := f(x) + \sum_{j=1}^{p} \lambda_j \min\{0, g_j(x)\} + \sum_{k=1}^{q} \kappa_k |h_k(x)| \tag{9}$$

The best-known version [5] of a Lagrangian relaxation is:

$$\rho_0 := \max_{\lambda \in \mathbb{R}^p, \kappa \in \mathbb{R}^q} \min_{x \in \mathbb{R}^m} L(x, \lambda, \kappa). \tag{10}$$

One often adds additional regularisation terms to the Lagrangian [34], which may improve the rate of convergence, but should not obscure the fact that ρ_0 of the Lagrangian relaxation may be far removed from f^*. One may also remove the min in (9) and add constraints on λ_j to be non-negative in (10), but either way, it is impossible to apply α-β theory directly.

Using the looser definition of the Lagrangian, we define the domain of monotonicity of a POP (7) with respect to a particular Lagrangian.

Definition 1 (Monotonicity Domain With Respect to $\tilde{L}$): For any $\tilde{x} \in \mathbb{R}^{\tilde{m}}$ and $\tilde{L} : \mathbb{R}^{\tilde{m}} \mapsto \mathbb{R}$, consider a sequence $\tilde{x}_0 := \tilde{x}$, $\tilde{x}_{i+1} := N_{\tilde{L}}(\tilde{x}_i)$ for $i > 0$. The point $\tilde{x}$ is within the domain of monotonicity with respect to $\tilde{L}$ if this sequence is well defined and there exists a point $\tilde{x}' \in \mathbb{R}^{\tilde{m}}$ such that $\tilde{L}(\tilde{x}') = 0$ and

$$\| \tilde{x}_i - \tilde{x}' \| \leq (1/2)^{2^i - i} \| \tilde{x}_0 - \tilde{x}' \| . \tag{11}$$

Then, we call $\tilde{x}'$ the associated stationary point of $\tilde{x}$ and say that $\tilde{x}$ represents $\tilde{x}'$.

Notice that we use tilde to stress the variable parts, such as the Lagrangian $\tilde{L}$ and its dimension $\tilde{m}$. Notice also that domains of monotonicity are known also as the region of attraction, the basin of attraction, etc.

B. The Assumptions

Recently, it has been realised that one can approximate the global optimum f^* as closely as possible, in case one applies the relaxation to a problem $\tilde{P}$ equivalent to P, which has sufficiently many redundant constraints. To state the result, we need some additional technical assumptions as follows.

Assumption 1: $\mathbf{K}$ is compact and $0 \leq g_j(x) \leq 1$ on $x \in \mathbf{K}$ for all $j = 1, \dots, p$, possibly after re-scaling. Moreover, the family of polynomials $\{g_j, 1 - g_j\}$ generates the algebra $\mathbb{R}[x]$.

Notice that if $\mathbf{K}$ is compact, one may always rescale variables x_i and add redundant constraints such that Assumption 1 holds. Further, we assume the following.

Assumption 2: There exists a unique point $x^* \in \mathbf{K}$, where f^* is attained.

Notice that one can easily construct an example with two generators with the same feed-in tariff and a single load bus connected to the two generators by branches so short that the losses are too low to measure, where this assumption is violated. At the same time, it is easy to see that an arbitrarily small perturbation to the cost function makes it possible to satisfy the assumption. Alternatively, one could replace Assumption 2 with an assumption on the separation of stationary points, as discussed in [9].

C. The Results

It is well-known that one can construct the following.

Lemma 1 (Lasserre Hierarchy): Let Assumption 1 hold for K (6) underlying a POP P with optimum f^*. For every $\epsilon > 0$, there exists $d_\epsilon \in \mathbb{N}$ such that for every $d \geq d_\epsilon$, there exists a *convex* Lagrangian relaxation of P, which yields a lower bound $f^* - \epsilon \leq \rho_d \leq f^*$.

Proof: There are several possible proofs. One follows that of [26, Corollary 2.1]. An alternative is based on [25] and considers a Lagrangian relaxation of semidefinite programming problems. There, the strong duality can be assured by a reformulation of the POP, see [20]. ∎

Notice, however, that these Lagrangians, while convex, are not polynomial, due to the presence of the semidefinite constraint. Moreover, for $d \geq d_\epsilon$, a single iteration of minimising the *convex* Lagrangian, even using a first-order method, can be computationally much more demanding than a single iteration of second-order methods for the basic Lagrangian ρ_0. We would hence like to study the domains of monotonicity with respect to variants of the basic Lagrangian, where there would be no inequalities.

At a particular point, one can check which inequalities are satisfied with equality, up to some tolerance, and replace such inequalities with equalities. As usual [5], we use $A(x, \epsilon) \subseteq \{1, 2, \dots, p\}$ to denote the index set of the so-called active inequalities of the POP (7) that are satisfied with equality, up to the tolerance of ϵ, at a point $x \in \mathbb{R}^m$:

$$A(x, \epsilon) := \big\{ j : |g_j(x)| \leq \epsilon, j = 1, 2, \dots, p \big\}, \tag{12}$$

At the point x, we can evaluate $A(x, \epsilon)$ and construct a locally valid, but polynomial, Lagrangian:

$$L'(x, \lambda, \kappa) := f(x) + \sum_{j=1}^{p} \big(\mathbb{1}_{j \in A} \lambda_j g_j(x) \big) + \sum_{k=1}^{q} \kappa_k h_k(x) \tag{13}$$

The following is clear:

Lemma 2: Let Assumptions 1 and 2 hold for $\mathbf{K}$ (6). For every $\epsilon > 0$, there exists $d_\epsilon \in \mathbb{N}$ such that for every $d \geq d_\epsilon$, the Lagrangian relaxation of $\tilde{P}_d$ yields a lower bound $f^* - \epsilon \leq \rho_d \leq f^*$ achieved at x_d^* and the active set $A(x_d^*, \epsilon)$ induces $L'(x, \lambda, \kappa)$ with optimum ρ_d.

Proof: The proof follows from the reasoning of [9, Propositions 7 and 8], as explained by Henrion and Lasserre [19]: Under Assumptions 1 and 2, the moment matrix for d makes it possible to extract a feasible solution by performing Schur decomposition [9], which in turn allows one to estimate the active set. ∎

This allows for the direct application of α-β theory.

Theorem 1: There exists a universal constant $\alpha_0 \in \mathbb{R}$, such that for all $m \in \mathbb{N}, P \in \mathbb{P}^m$, where Assumptions 1 and 2 hold for P, there exists a $d \in \mathbb{N}$, such that for every $\epsilon > 0$, there exists $d_\epsilon \in \mathbb{N}$ such that for every $d \geq d_\epsilon$, there is a Lagrangian relaxation $\tilde{L}_d$ in dimension $\tilde{m}$, and a function $\alpha: \mathbb{P}^{\tilde{m}} \times \mathbb{R}^{\tilde{m}} \mapsto \mathbb{R}$ such that if $\alpha(\nabla \tilde{L}_d, \tilde{x}) \leq \alpha_0$, then x is the domain of monotonicity of a solution with objective function value ρ_d such that $f^* - \epsilon \leq \rho_d \leq f^*$.

Proof: The proof follows from the observation that each *convex* Lagrangian of Lemma 1 is associated with a non-convex, but polynomial Lagrangian of Lemma 2, and that both Lagrangians will have a function value at their optima bounded from below by $f^* - \epsilon$. Formally, for all $m \in \mathbb{N}, p \in \mathbb{P}^m$, and $x \in \mathbb{R}^m$, there exists $\tilde{P}_d$, $d \in \mathbb{N}$, such that for every $\epsilon > 0$, there exists $d_\epsilon \in \mathbb{N}$ such that for every $d \geq d_\epsilon$, both the Lagrangian relaxation $\tilde{L}_d$ of $\tilde{P}_d$ and the new Lagrangian relaxation of the same problem L'_d yield a lower bound $f^* - \epsilon \leq \rho_d \leq f^*$. While minimising the convex Lagrangian of the polynomial optimization problem (7), we can apply Proposition 1 to the first-order conditions of the corresponding Lagrangian L'_d of Lemma 2. ∎

D. An Application to ACOPF

The alternating-current optimal power flow problem (ACOPF) extends the constraints of alternating-current power flow (ACPF) of Section II with a number of box constraints; an objective, which is the sum of quadratic functions of real powers; and the so-called thermal limits. We refer to [34] for the complete formulation. In terms of ACOPF, the theory can be summarised as follows.

Corollary 1: There exists a universal constant $\alpha_0 \in \mathbb{R}$, such that for every instance of ACOPF, there exists $\delta \in \mathbb{R}, \delta \geq 0$ and a function $\alpha : \mathbb{R}^m \mapsto \mathbb{R}$ specific to the instance of ACOPF, such that for any $\epsilon > \delta$ and vector $x \in \mathbb{R}^m$ if $\alpha(x) \leq \alpha_0$, x is in the domain of monotonicity of an optimum of the instance of ACOPF, which is no more than ϵ away from the value of the global optimum with respect to its objective function.

Proof: By Theorem 1. The δ accounts for the perturbation. ∎

In the hybridisation we propose, one starts by solving a convexification, followed by the estimation of the active set in the outer loop. Then, one may test the stability of the active set. Whenever the active set seems stable and the test of Proposition 1 applied to L' allows, we switch to Newton's method on the non-convex Lagrangian L'. Some back-tracking line search may be employed within Newton's method, until a sufficient decrease in L is observed. Although this algorithm may seem somewhat crude, it seems to perform well.

Alternatively, one may employ a variant, whose schematic overview is in Algorithm 1. There, we consider first-order optimality conditions of L' in the test on Line 6, but switch to Newton's method on the first-order optimality conditions of (9), while memorising the current value as S. While minimising (9), we check the active set; when it does change, we revert to solving the convexification with the memorised value S. Although this algorithm may seem even cruder than the above, it performs better still in practice.

IV. The Practice

In implementing a hybrid method for ACOPF, such as Algorithm 1, one encounters a number of challenges. One requires a solver for the convexification of ACOPF, a well-performing implementation of Newton's method for the non-convex Lagrangian L', and an implementation of Proposition 3. We will comment upon these in turn.

Algorithm 1 A Schema of the Hybrid Method

1: Initialise $x \in \mathbb{R}^m$, $\lambda \in \mathbb{R}^p$, $\kappa \in \mathbb{R}^q$, e.g., randomly
2: **for** $k \leftarrow 0, 1, 2, \ldots$ **do**
3: Update (x, λ, κ), e.g., using [34]
4: $A_k \leftarrow A(x, \epsilon)$, i.e., index-set (12) of inequalities satisfied up to ϵ-accuracy
5: Construct the polynomial Lagrangian function L' corresponding to A_k
6: **if** $k > K$ **and** $A_k = A_{k-1} = \ldots = A_{k-K}$ **and** $\alpha(\nabla L', x) \le \alpha_0$ **then**
7: $S \leftarrow (x, \lambda, \kappa)$
8: **for** $l \leftarrow 0, 1, 2, \ldots$ **do**
9: Update (x, λ, κ) using Newton's step, as discussed in Section IV-C.C
10: $A'_l \leftarrow A(x, \epsilon)$, i.e., index-set (12) of inequalities satisfied up to ϵ-accuracy
11: **if** $A'_l \neq A_k$ **then**
12: $(x, \lambda, \kappa) \leftarrow S$
13: **break**
14: **end if**
15: **if** infeasibility $T(x) < \epsilon$, cf. (8) **then**
16: Optionally, test sufficient conditions for global optimality, e.g., [35]
17: **break**
18: **end if**
19: **end for**
20: **if** infeasibility $T(x) < \epsilon$, cf. (8) **then**
21: **break**
22: **end if**
23: **end if**
24: **end for**

A. The Convexification

The convexification we use is based on of the Lagrangian of the relaxation of Lavaei and Low [27]. (As we have shown in [17], the relaxation of Lavaei and Low is the first level of the hierarchy of Lasserre [25], considered in Lemma 1.) In particular, we have used a variant introduced in [34].

To solve the convexification, we have used a problem-specific first-order method [34], which is based on a low-rank coordinate descent with a closed-form step. Outside of other advantages, this method maintains the feasible solution of ACOPF at least throughout the first iteration of the outer loop, which often suffices, and makes it unnecessary to extract the feasible solution of ACOPF, as suggested in the proof of Lemma 2.

B. The Test

A key contribution of ours is an implementation of Proposition 3 specific to ACOPF. There, one should observe that β is easy to obtain as $\beta(x, L) := \|d\|_2 = \|L_p\|_2$, where L_p is Newton's direction, and $f = \frac{\partial L}{\partial x}$. By observing $\frac{\partial L}{\partial x}$, we can use $d_i = 3\ \forall i$, thus $D = 3$ and $\Delta_{(d)}(x) = 3^{1/2}\|x\|_1^2 I_{2n\times 2n}$, where $I_{2n\times 2n}$ is a $2n \times 2n$ identity matrix, so

$$\mu(L, x) = \max\left\{1, \sqrt{3}\|x\|_1^2 \cdot \|\nabla L\| \cdot \|[\nabla^2 L(x)]^{-1}\|\right\},$$

where the spectral norm $\|[\nabla^2 L(x)]^{-1}\|$ can be computed as the inverse of the non-zero eigenvalue of $\nabla^2 L(x)$ whose absolute value is minimal.

A trivial implementation may run for days even on modest instances. In our implementation, we used about 2000 lines of algebraic manipulations in Python to generate considerable amounts of instance-specific, optimized C code employing Intel MKL Libraries. For example, the test for case2383wp involves about 30 MB of C code. This makes it possible to run the test within seconds even on case2383wp. Still, one may benefit from running the test, only when the active set has been constant for K iterations of the outer loop, as suggested in Line 6 of Algorithm 1.

C. The Newton Method

There are a number of options for implementing Newton's method in Line 9 of Algorithm 1. The straightforward option is to apply Newton's method to $\nabla L' = 0$, which has the local quadratic convergence rate [5, Proposition 4.4.3] and where the theory of the previous section holds. One can also use any other method with a quadratic rate of convergence for solving $\nabla L' = 0$ in order for the reasoning of the previous section to be applicable, see [7].

Further, a number of alternatives are possible:

1) One can smooth the non-smooth parts of the Lagrangian (9) and then apply Newton's method to solve (10), or consider projected Newton's method with box constraints. The implementation is non-trivial, considering the min-max structure, but standard. [4] details many practical suggestions for the former, while [24] presents the latter.
2) One can apply primal-dual interior-point methods to a variant of the problem with logarithmic barriers [5, Sec. 4.4.4] or similar [8]. The implementation is, again, non-trivial, but standard. The local rate of convergence is quadratic [8] or better, under mild conditions.
3) One can employ alternating-minimisation methods in solving (10), with Newton's step for some or all of the blocks. In particular, one can alternate between minimisation of primal variables (x) and maximisation of dual variables (λ, κ). Multiple Newton steps, each satisfying sufficient decrease, can be performed in each iteration of the loop, before a sufficient decrease in the convex Lagrangian is tested.

We will not provide theoretical results matching those of Section III for any of these three alternatives. Due to this fact, combined with the non-convexity of the Lagrangian, Newton's direction may turn out not to be a direction of descent, in which case one can multiply it by -1, as usual [16]. See Section IV-E below for some computational illustrations.

D. Set-Up of Computational Experiments

To validate the impact of our approach, we performed numerical experiments on a collection of well-known instances [48] and two variants of thermal limits. Whenever we mention "extended" next to the name of the instance, we use a formulation of thermal limits allowing for phase-shifting and tap-changing transformers as explained in [34, Sec. 5.2]. The experiments were performed on a computer with an Intel Xeon CPU E5-2620 clocked at 2.40GHz and 128 GB of RAM. Throughout, we compare the performance of the

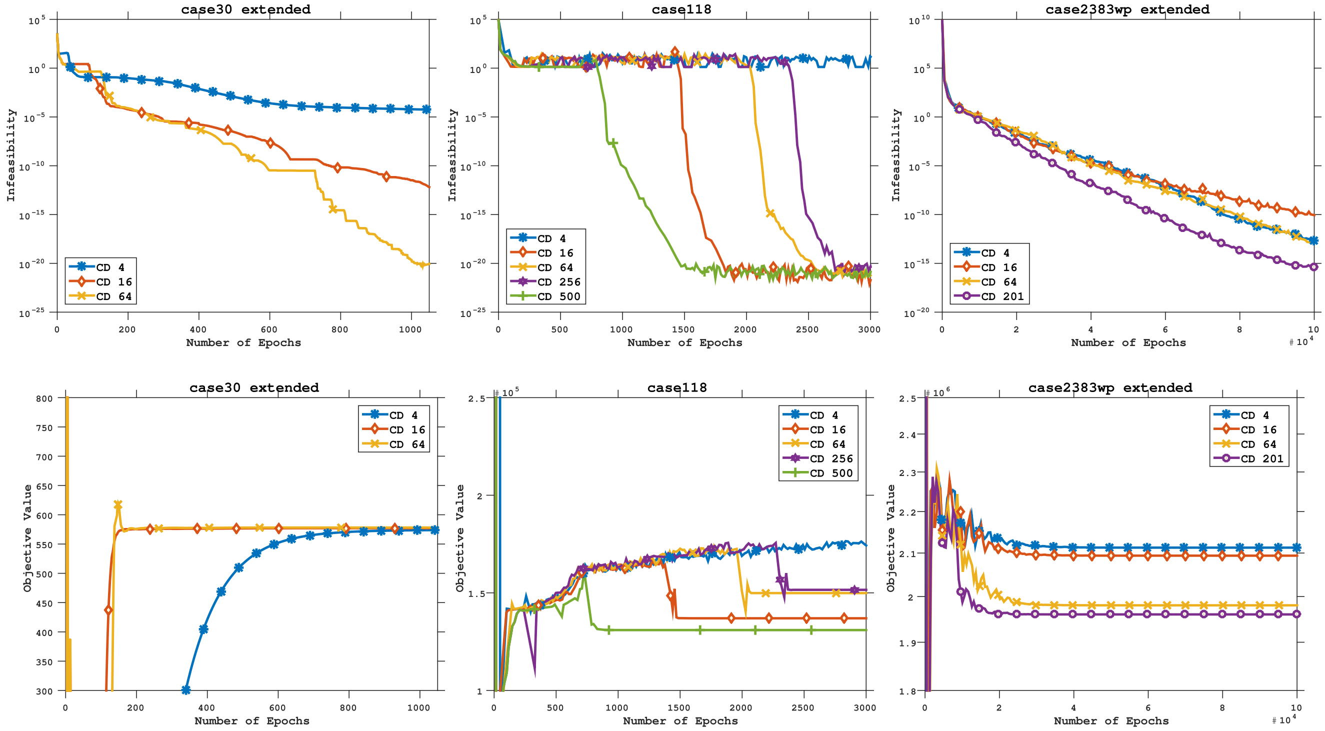


Fig. 1. The motivation: the evolution of a measure of infeasibility (top row) as defined in (8) and objective function (bottom row) when one switches from solving the convexification to Newton's method after a given number of steps on IEEE 30-bus test system (left), 118-bus test system (middle), and a snapshot of the Polish system (case2383wp; right).

coordinate descent of [34] on the Lavaei-Low SDP relaxation [27] (plotted in blue), against Newton's method on the non-convex Lagrangian (plotted in yellow), against the performance of a variant of the hybrid method (plotted in red), which switches from the coordinate descent on the convexification to to Newton's method on the non-convex Lagrangian, when the α-β test is satisfied. In particular, we plot the evolution of the infeasibility as defined in (8) and the evolution of the objective function over the number of epochs, where each epoch refers to either m iterations of coordinate descent, or m coordinate-wise Newton's steps, for an instance in dimension m.

We have used randomisation in sampling of coordinates, but we have used a fixed random seed for all runs of all methods. Unless stated otherwise, we have used voltage magnitudes uniformly at 1, phase angles uniformly at 0, and power generated uniformly at mid-points of the respective intervals as the initial point, in keeping with [48]. We discuss the stability of the methods in more detail below.

E. Results of Computational Experiments

In Figures 1–6, we present a sample of the results. First, we motivate the need for a hybrid method in Figure 1. There, each time series represents one run, where one starts by solving the convexification using coordinate descent, and switches over to Newton's method after a specified number of epochs. For example, series CD 4 is obtained by running 4 epochs of coordinate descent before switching to Newton's method. We chose series with 4^i epochs of coordinate descent,

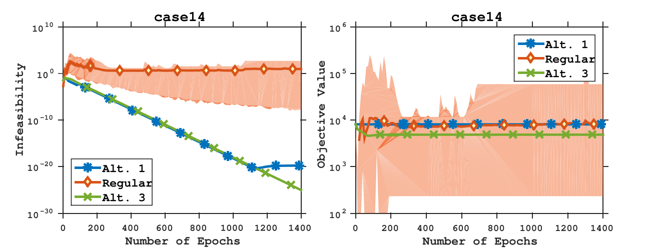


Fig. 2. An illustrative comparison of three variants of the Newton method, as discussed in Section IV-E.

$i = 1, 2, \ldots,$ up to the point where the α-β test is satisfied. Figure 1 shows that even after a number of iterations of coordinate descent, each of which decreases the value of the Lagrangian, Newton's method can diverge. See, for example, the series denoted CD 4 and CD 500 in the middle plots for the IEEE 118-bus test system, where one switches-over after 4 and 500 epochs of the coordinate descent on the convexification, respectively. In the middle plot in the top row, we see that T, the measure of infeasibility (8), for CD 4 does not seem to fall below 1, ever. In the middle plot in the bottom row, we see that a variety of stationary points can be reached, with the switch-over after 500 epochs (CD 500) yielding a considerably different stationary point compared to the switch-over after 16 epochs (CD 16) and 64 epochs (CD 64).

Next, we illustrate the performance of three variants of Newton's method in our own implementation in Figure 2, again in terms of the evolution of the objective and T, the measure of infeasibility (8). The shaded areas are the support of an empirical distribution obtained as 100 sample paths

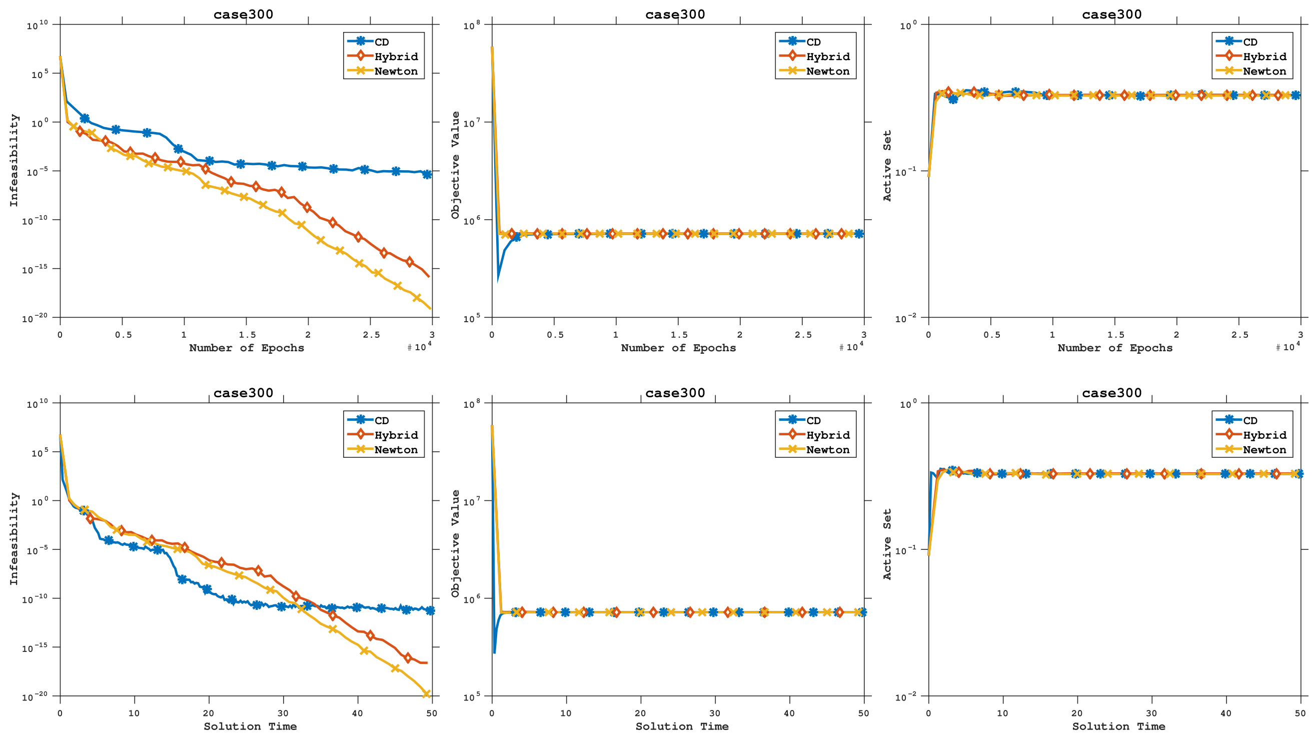


Fig. 3. The performance of the hybrid method on the IEEE 300-bus test system (case300).

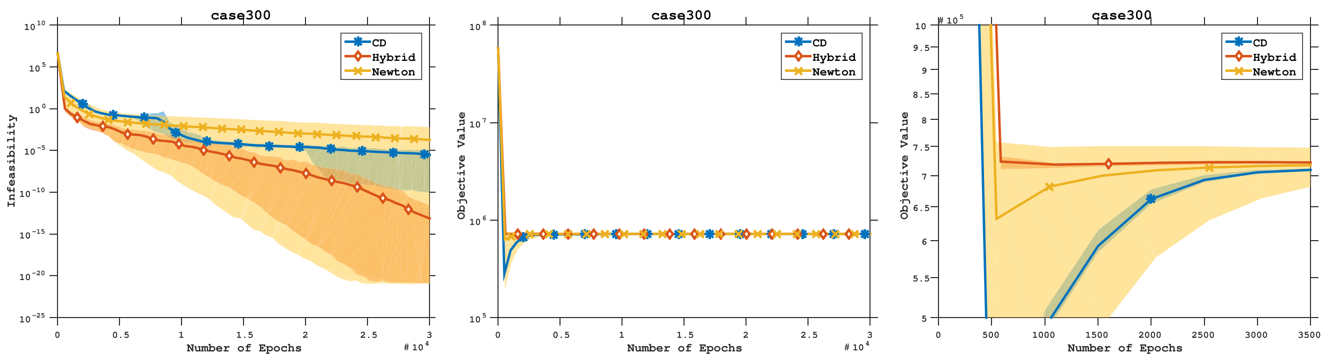


Fig. 4. The stability of the methods studied. Mean and support of an empirical distribution of the sample paths generated by 100 runs with 100 different random seeds, as described in Section IV-E.

where one adds Gaussian noise $\mathcal{N}(0, 10^{-7})$ to the initial point (see Section IV-D above). The time series are averages over the same 100 runs. For the sake of the comparison, we use the same perturbation of the initial point across each of the three methods and we use tuned and fixed penalty parameters and step sizes throughout. Each epoch refers to one Newton step. In the series labelled "Regular", we use the textbook Newton's method [5] on $\nabla L' = 0$, with Hessian obtained using symbolic differentiation. In the series labelled "Alt. 1)", we use Newton's method projected onto box constraints, as proposed by [24]. In the series labelled "Alt. 3)", we use alternating minimisation, where we solve the minimisation problem in primal variable (x) using Newton's method and where we use the gradient step for dual variables (λ, κ). As expected, regular Newton's method has quadratic rate of local convergence, but major issues with numerical stability; increasing the standard deviation of the noise applied to the initial point to 0.0001 or moving from case14 to a larger instance seems to make the method useless. Although we have no theoretical justification for this, Alternative 3 seems to exhibit local quadratic convergence and outperforms all other methods we have experimented with. In the subsequent results, we hence employ Alternative 3.

Next, we compare the hybrid method against the use of coordinate descent on the convexification on its own and the use of Newton's method on its own. For the first illustration, we chose the IEEE 300-bus test system. As above, we plot a measure T of infeasibility (left; see Equation (8) for the definition) and the objective function value (middle) against both wall-clock time (top row) and epochs (bottom row) in Figure 3. As can be seen by comparing the top and bottom row, the wall-clock time corresponding to one epoch across the three methods is similar. On the other hand, the convergence rates are visibly different, with the infeasibility

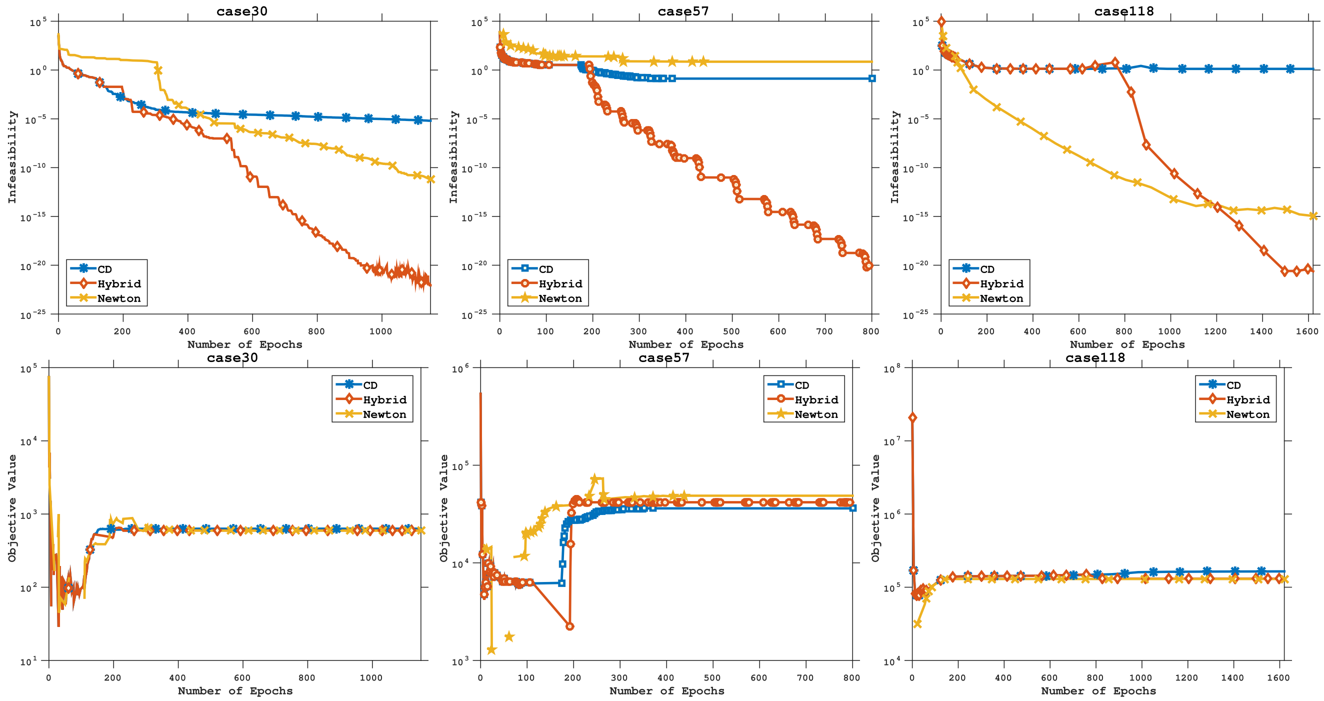


Fig. 5. The performance of the hybrid method on three IEEE test systems: 30-bus, 57-bus, and 118-bus.

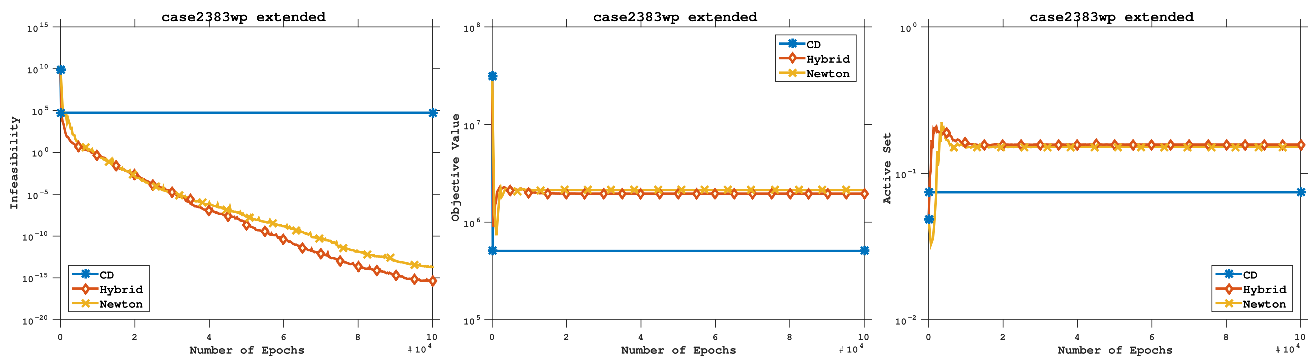


Fig. 6. The performance of the hybrid method on a snapshot of the Polish transmission system (case2383wp).

decreasing at a quadratic rate for Newton's and hybrid method.

On the IEEE 300-bus test system, we can also illustrate the stability of the methods with respect to the sampling of coordinates (in coordinate descent and its use in the hybrid method) and random perturbations to the initial point. Figure 4 presents the mean (dark-colored lines) and support (transparent regions) of an empirical distribution of the sample paths generated by 100 runs with 100 different random seeds. The empirical distribution has been sampled from a distribution generated as follows: we have added Gaussian noise $N(0, 0.01)$ to the initial point (see Section IV-D above) and the coordinates have been sampled uniformly at random. Note that the right-most plot is a close-up of the middle plot. In contrast to Figure 3, where Newton's method outperforms the hybrid method on the one sample, the average in Figure 4 suggests that Newton's method will not yield infeasibility less than 10^{-6} on average. This seems to demonstrate the benefits of the hybrid method compared to the coordinate descent and Newton's method, in terms of the rate of convergence and stability.

Further, we present the results on three more IEEE test systems in Figure 5 in a more concise form with only the evolution of T, the measure of infeasibility (top row), and the objective function (bottom row) over the number of epochs. The 30-bus (on the left) and 118-bus (on the right) test systems illustrate the typical performance: the evolution of T, the measure of infeasibility, of the hybrid method overlaps with the first-order method until the switch-over. Thenceforth, the quadratic rate of convergence resembles that of Newton's method, except with a better starting point. The 57-bus test system (in the middle) demonstrates the importance of the starting point: our implementation of Newton's method from the Matpower starting point does not converge.

Next, to illustrate the scalability of the approach, we present the results on a snapshot of the Polish system in Figure 6. There are 2383 buses in the snapshot and, more importantly, tap-changing and phase-shifting transformers, double-circuit transmission lines, and multiple generators at each bus, which complicate the formulation of the thermal limits, as explained in [34, Sec. 5.2]. Despite the preliminary nature of our implementation, compared to the established codes, developed over a decade or more [48], the convergence seems very robust.

Finally, in the right-most plots of Figures 3 and 6, we plot the ratio of the cardinality of the active set to the number of inequalities over the epochs or time. This provides an empirical justification for the choice of Algorithm 1: the active set clearly stabilises much earlier than the objective function value, and is only a small fraction of the count of the inequalities, which allows for the short run-time of the test implementing Proposition 3.

V. Related Work

Let us present a brief overview of the rich history of the study of the convergence of Newton's method. The best known result in the field is the theorem of Kantorovich [23], which formalises the assumptions under which whenever for a closed ball of radius t_* centered at x_0, it holds that $\nabla F(x)+\nabla F(x)^T \succ 0$ for all x in the ball, the ball is a domain of monotonicity for the function F. Traditionally, it has been assumed that testing the property across the closed ball is difficult.

Recently, Henrion and Korda [18] have shown that the domain of monotonicity of a polynomial system can be computed by solving an infinite-dimensional linear program over the space of measures, whose value can be approximated by a certain hierarchy of convex semidefinite optimization problems. See also the work of Valmórbida *et al.* [44]–[46] in the context of partial differential equations, and elsewhere [12]. Dvijotham *et al.* [14], [15] showed that it can also be cast as a certain non-convex semidefinite optimization problem. Notice, however, that this line of work [14], [15], [18] does not consider inequalities and may be rather computationally challenging. Similarly, the α-β theory [7], [10], [40], does not consider inequalities.

To summarise: traditionally, the convergence of Newton's method could be guaranteed only by the non-constructive arguments of the theorem of Kantorovich [23]. Alternatively, one could use recently developed approaches [14], [15], [18], albeit at a computational cost possibly higher than that of solving ACOPF. Our approach seems to improve upon these considerably.

Let us also highlight some of the recent advances in convexifications in power systems. Lavaei and Low [27] have shown that a convex relaxation employing optimization over positive semidefinite matrices (semidefinite programming, SDP) produces global optima in some cases. In particular, this is the case for tree-like network topologies [28], [47] and some IEEE benchmark instances [27]. For further classes of instances [33], [38], [41], minor changes to the instance make the feasible set convex as well. Still, there are instances where the SDP relaxation provides only unsatisfactory lower bounds [6], [29], [34]. There, the SDP relaxation can be strengthened by the iterated addition of further constraints [17] or variables and constraints [17], [21], [22], [36], [37]. Such iterative strengthening produces a hierarchy of relaxations that converges to the global optimum of the non-convex problem, asymptotically, under mild conditions, albeit at a considerable computational cost. Similarly, one can also derive a convergent hierarchy of upper bounds [11]. See [30], [31], [39] for further references. Our approach aims to make the strengthened convexifications practical.

VI. Conclusion

Newton-type methods can converge to particularly bad local optima of non-convex problems, when applied directly. On the other hand, even the fastest first-order methods for solving strong convexifications are rather slow. Hybrid methods combine the use of Newton-type methods on the non-convex problem with the use of (hierarchies of) strong convexifications so as to benefit from both from the guarantees of convergence associated with (hierarchies of) convex relaxations and the quadratic rates of convergence of Newton-type methods. Crucially, such hybrid methods can be implemented in a distributed fashion, as discussed in [34]. This improves upon [14] and [15] and opens up many novel directions for future research.

An important direction for further study is the trade-off between run-time to acceptable precision in solving convexifications, in practice, and the strength of their convergence guarantees, in theory. Throughout our computational tests, we have used a variant of the relaxation of Lavaei and Low [27] introduced by Marecek and Takac [34], which is exact on a variety of well-known instances. Hierarchies of relaxations, where one can show generically global convergence [17], [21], [22], [36], [37], are computationally more challenging. Our approach preserves the guarantees associated with whichever convexification is used.

Another important direction for futher study are the rates of local convergence of the variants of Newton's method used in power systems practice. In theory [23], it is clear that the quadratic rate of local convergence can be obtained when the full Newton step is applied to the Lagrangian of equality-constrained optimization problems, or after inequalities in the active set are converted to equalities and the remainder of inequalities are dropped. In practice, however, one may wish to use projection for the inequalities and block-wise updates, or even more complicated numerical methods [48], whose rates of convergence are not clear. Although it may remain unclear whether certain methods used in practical power systems analysis are actually locally quadratically convergent, our test applies to the cross-over to any locally quadratically convergent method.

Finally, one should like to extend the applicability of the approach to real-time applications with possibly time-varying constraints [1]. In the spirit of robust optimization, one could pre-compute a μ-like bound, which would allow for the variation of some of the coefficients in some of the constraints within specified uncertainty sets. The corresponding eigenvalue optimization problem could possibly also be cast as a semidefinite program. Alternatively, one could aim to update the bound based on the update to the coefficients, where the stability of eigenvalues is a well-studied subject in random matrix theory. In this respect, we have taken only the very first step.

References

[1] E. D. Anese and A. Simonettom, "Optimal power flow pursuit," *IEEE Trans. Smart Grid*, to be published.

[2] M. F. Anjos and J.-B. Lasserre, Eds., "Handbook on semidefinite, conic and polynomial optimization," in *International Series in Operations Research & Management Science*, vol. 166. New York, NY, USA: Springer, 2012.

[3] C. Beltrán and L. M. Pardo, "Efficient polynomial system-solving by numerical methods," *J. Fixed Point Theory Appl.*, vol. 6, no. 1, pp. 63–85, 2009.

[4] M. Benzi, G. H. Golub, and J. Liesen, "Numerical solution of saddle point problems," *Acta Numerica*, vol. 14, pp. 1–137, 2005.

[5] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA, USA: Athena Sci., 1999.

[6] W. A. Bukhsh, A. Grothey, K. I. M. McKinnon, and P. A. Trodden, "Local solutions of the optimal power flow problem," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4780–4788, Nov. 2013.

[7] P. Chen, "Approximate zeros of quadratically convergent algorithms," *Math. Comput.*, vol. 63, no. 207, pp. 247–270, 1994.

[8] T. F. Coleman and J. Liu, "An interior Newton method for quadratic programming," *Math. Program.*, vol. 85, no. 3, pp. 491–523, 1999.

[9] R. M. Corless, P. M. Gianni, and B. M. Trager, "A reordered Schur factorization method for zero-dimensional polynomial systems with multiple roots," in *Proc. Int. Symp. Symb. Algebr. Comput. (ISSAC)*, 1997, pp. 133–140.

[10] F. Cucker and S. Smale, "Complexity estimates depending on condition and round-off error," *J. ACM*, vol. 46, no. 1, pp. 113–184, Jan. 1999.

[11] E. de Klerk, R. Hess, and M. Laurent, "Improved convergence rates for Lasserre-type hierarchies of upper bounds for box-constrained polynomial optimization," *SIAM J. Optim.*, vol. 27, no. 1, pp. 347–367, 2017.

[12] M. Demenkov, "A MATLAB tool for regions of attraction estimation via numerical algebraic geometry," in *Proc. Int. Conf. Mech. 7th Polyakhov's Reading*, Feb. 2015, pp. 1–5.

[13] W. Deren and Z. Fengguang, "The theory of Smale's point estimation and its applications," *J. Comput. Appl. Math.*, vol. 60, nos. 1–2, pp. 253–269, 1995.

[14] K. Dvijotham, M. Chertkov, and S. Low, "A differential analysis of the power flow equations," in *Proc. 54th IEEE Conf. Decis. Control (CDC)*, Osaka, Japan, Dec. 2015, pp. 23–30, doi: 10.1109/CDC.2015.7402082.

[15] K. Dvijotham, M. Chertkov, and S. Low, "Monotone operator approach to power flow solutions," in *Proc. Amer. Control Conf. (ACC)*, Boston, MA, USA, Jul. 2016, pp. 1769–1769, doi: 10.1109/ACC.2016.7525174.

[16] A. Fiacco and G. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques* (Research Analysis Corporation. Research Series). New York, NY, USA: Wiley, 1968.

[17] B. Ghaddar, J. Marecek, and M. Mevissen, "Optimal power flow as a polynomial optimization problem," *IEEE Trans. Power Syst.*, vol. 31, no. 1, pp. 539–546, Jan. 2016.

[18] D. Henrion and M. Korda, "Convex computation of the region of attraction of polynomial control systems," *IEEE Trans. Autom. Control*, vol. 59, no. 2, pp. 297–312, Feb. 2014.

[19] D. Henrion and J.-B. Lasserre, *Detecting Global Optimality and Extracting Solutions in GloptiPoly*. Berlin, Germany: Springer, 2005, pp. 293–310.

[20] C. Josz and D. Henrion, "Strong duality in Lasserre's hierarchy for polynomial optimization," *Optim. Lett.*, vol. 10, no. 1, pp. 3–10, 2016.

[21] C. Josz, J. Maeght, P. Panciatici, and J. C. Gilbert, "Application of the moment-SOS approach to global optimization of the OPF problem," *IEEE Trans. Power Syst.*, vol. 30, no. 1, pp. 463–470, Jan. 2015.

[22] C. Josz and D. K. Molzahn, "Moment/sum-of-squares hierarchy for complex polynomial optimization," *SIAM J. Optim.* [Online]. Available: https://arxiv.org/abs/1508.02068

[23] L. Kantorovich, "On Newton's method for functional equations," *Doklady Akademii Nauk SSSR*, vol. 59, no. 7, pp. 1237–1240, 1948.

[24] D. Kim, S. Sra, and I. S. Dhillon, "Tackling box-constrained optimization via a new projected quasi-Newton approach," *SIAM J. Sci. Comput.*, vol. 32, no. 6, pp. 3548–3563, 2010.

[25] J. B. Lasserre, "Convergent SDP-relaxations in polynomial optimization with sparsity," *SIAM J. Optim.*, vol. 17, no. 3, pp. 822–843, 2006.

[26] J. B. Lasserre, "A Lagrangian relaxation view of linear and semidefinite hierarchies," *SIAM J. Optim.*, vol. 23, no. 3, pp. 1742–1756, 2013.

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

[28] J. Lavaei, D. Tse, and B. Zhang, "Geometry of power flows and optimization in distribution networks," *IEEE Trans. Power Systems*, vol. 29, no. 2, pp. 572–583, Mar. 2014.

[29] B. C. Lesieutre, D. K. Molzahn, A. R. Borden, and C. L. DeMarco, "Examining the limits of the application of semidefinite programming to power flow problems," in *Proc. 49th Annu. Allerton Conf. Commun. Control Comput. (Allerton)*, Sep. 2011, pp. 1492–1499.

[30] S. H. Low, "Convex relaxation of optimal power flow—Part I: Formulations and equivalence," *IEEE Trans. Control Netw. Syst.*, vol. 1, no. 1, pp. 15–27, Mar. 2014.

[31] S. H. Low, "Convex relaxation of optimal power flow—Part II: Exactness," *IEEE Trans. Control Netw. Syst.*, vol. 1, no. 2, pp. 177–189, Jun. 2014.

[32] R. Madani, A. Kalbat, and J. Lavaei, "ADMM for sparse semidefinite programming with applications to optimal power flow problem," in *Proc. 54th IEEE Conf. Decis. Control (CDC)*, Dec. 2015, pp. 5932–5939.

[33] R. Madani, S. Sojoudi, and J. Lavaei, "Convex relaxation for optimal power flow problem: Mesh networks," *IEEE Trans. Power Syst.*, vol. 30, no. 1, pp. 199–211, Jan. 2015.

[34] J. Marecek and M. Takac, "A low-rank coordinate-descent algorithm for semidefinite programming relaxations of optimal power flow," *Optim. Methods Softw.*, vol. 32, no. 4, pp. 849–871, 2017.

[35] D. K. Molzahn, B. C. Lesieutre, and C. L. DeMarco, "A sufficient condition for global optimality of solutions to the optimal power flow problem," *IEEE Trans. Power Syst.*, vol. 29, no. 2, pp. 978–979, Mar. 2014.

[36] D. K. Molzahn and I. A. Hiskens, "Sparsity-exploiting moment-based relaxations of the optimal power flow problem," *IEEE Trans. Power Syst.*, vol. 30, no. 6, pp. 3168–3180, Nov. 2015.

[37] D. K. Molzahn, I. A. Hiskens, C. Josz, and P. Panciatici, "Computational analysis of sparsity-exploiting moment relaxations of the OPF problem," in *Proc. Power Syst. Comput. Conf. (PSCC)*, Genoa, Italy, Jun. 2016, pp. 1–7.

[38] D. K. Molzahn, C. Josz, I. A. Hiskens, and P. Panciatici, "A Laplacian-based approach for finding near globally optimal solutions to OPF problems," *IEEE Trans. Power Syst.*, vol. 32, no. 1, pp. 305–315, Jan. 2017.

[39] P. Panciatici *et al.*, "Advanced optimization methods for power systems," in *Proc. Power Syst. Comput. Conf. (PSCC)*, Wrocław, Poland, Aug. 2014, pp. 1–18.

[40] M. Shub and S. Smale, "On the complexity of Bezout's theorem I—Geometric aspects," *J. Amer. Math. Soc.*, vol. 6, no. 2, pp. 459–501, 1993.

[41] S. Sojoudi and J. Lavaei, "Convexification of optimal power flow problem by means of phase shifters," in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Vancouver, BC, Canada, Oct. 2013, pp. 756–761.

[42] J. F. Sturm, "Using sedumi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optim. Methods Softw.*, vol. 11, no. 1–4, pp. 625–653, 1999.

[43] W. F. Tinney and C. E. Hart, "Power flow solution by Newton's method," *IEEE Trans. Power App. Syst.*, vol. PAS-86, no. 11, pp. 1449–1460, Nov. 1967.

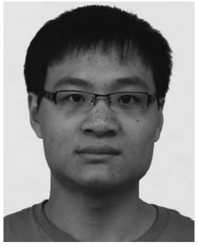
[44] G. Valmórbida, M. Ahmadi, and A. Papachristodoulou, "Stability analysis for a class of partial differential equations via semidefinite programming," *IEEE Trans. Autom. Control*, vol. 61, no. 6, pp. 1649–1654, Jun. 2016.

[45] G. Valmórbida and J. Anderson, "Region of attraction analysis via invariant sets," in *Proc. Amer. Control Conf.*, Portland, OR, USA, Jun. 2014, pp. 3591–3596.

[46] G. Valmórbida, S. Tarbouriech, and G. Garcia, "Region of attraction estimates for polynomial systems," in *Proc. 48th IEEE Conf. Decis. Control Held Jointly 28th Chin. Control Conf. (CDC/CCC)*, Dec. 2009, pp. 5947–5952.

[47] B. Zhang and D. Tse, "Geometry of injection regions of power networks," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 788–797, May 2013.

[48] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 12–19, Feb. 2011.



Jie Liu received the B.S. degree in applied mathematics from Nankai University, Tianjin, China, in 2011, and the M.A. degree in mathematics from the University at Buffalo, Buffalo, NY, USA, in 2013. He is currently pursuing the Ph.D. degree in industrial engineering with Lehigh University, Bethlehem, PA, USA. His contributions have been supported by RCEAS Dean's Fellowship and Gotshall Fellowship from Lehigh University over the summer of 2015, funding from IBM Ireland Ltd., for an internship in the summer of 2016, and by IBM Corporation via its IBM Ph.D. Fellowship programme (2017–2018). His research interest lies in large-scale optimization, including randomized first-order algorithms, parallel and distributed algorithms, with applications in power systems and machine learning.

Alan Claude Liddell, Jr., received the B.S. degree in mathematics from University of Mary Washington, Fredericksburg, VA, USA, in 2012. He is currently pursuing the Ph.D. degree in computational mathematics with the University of Notre Dame, South Bend, IN, USA. His research interests lie in numerical algebraic geometry. He contributed to the research while on an internship with IBM Research—Ireland, in the summer of 2016.

Jakub Mareček received the first two degrees from Masaryk University, Brno, Czech Republic, and the Ph.D. degree from the University of Nottingham, Nottingham, U.K., in 2006, 2009, and 2012, respectively, all in computer science, with a focus on mathematical optimization. In 2012, he joined IBM Research—Ireland, as a Research Staff Member. His research interests include polynomial optimization problems in power systems, in particular, and the design and development of solvers for real-world optimization and control problems, in general.

Martin Takáč received the B.Sc. and M.Sc. degrees from Comenius University, Bratislava, Slovakia, and the Ph.D. degree from the University of Edinburgh, Edinburgh, U.K., in 2008, 2010, and 2014, respectively. In 2014, he joined the Department of Industrial and Systems Engineering, Lehigh University, Bethlehem, PA, USA, as an Assistant Professor. His current research interests include convex and nonconvex optimization, design and analysis of algorithms, parallel and distributed computing, and machine learning.