Spurious Local Minima in Power System State Estimation

Richard Y. Zhang, Javad Lavaei and Ross Baldick, *Fellow, IEEE*

***Abstract*—The power systems state estimation problem computes the set of complex voltage phasors given quadratic measurements using nonlinear least squares (NLS). This is a nonconvex optimization problem, so even in the absence of measurement errors, local search algorithms like Newton / Gauss–Newton can become "stuck" at local minima, which correspond to nonsensical estimations. In this paper, we observe that local minima cease to be an issue as redundant measurements are added. Posing state estimation as an instance of the low-rank matrix recovery problem, we derive a bound for the distance between the true solution and the nearest spurious local minimum. We use the bound to show that spurious local minima of the nonconvex least squares objective become far-away from the true solution with the addition of redundant information.**

I. Introduction

In power systems, *state estimation* is the problem of recovering the underlying system voltage phasors, given possibly inaccurate SCADA (supervisory control and data acquisition) measurements, which are typically real and reactive power line flows and power injections, and voltage phasor amplitudes [2]. State estimation proves *situational awareness* by allowing the system operator to monitor and assess the condition of the power system at any given instant, and if needed, take action. Operators use state estimation to identify anomalous system conditions, to dispatch generation, and to avoid stability and thermal limits [3]. These functions are poised to become even more important as the penetration of wind and solar generation increases, due to the inherent variability and uncertainty of such resources [4].

On the other hand, a lack of situational awareness—particularly in observing the voltage phasor angles over a wide area—has been cited as a significant cause to a number of blackouts [5], [6]. A post-mortem analysis of the August 2003 Northeast blackout revealed that the voltage phasor angle difference between Cleveland and Michigan had been slowly diverging for nearly an hour before the start of the actual blackout [7]. Had the real-time state estimation been in service during the event, the operators would have had warning of the impending problem and an opportunity to take remedial action [8].

This work was supported by grants from NSF, DARPA, ONR and AFOSR. Parts of this work have appeared in *Proceedings of the Hawaii International Conference on System Sciences*, 2018 [1].

R. Y. Zhang was previously with the Department of Industrial Engineering and Operations Research, University of California, Berkeley. He is currently with the Department of Electrical and Computer Engineering, University of Illinois, Urbana–Champaign (email: ryz@alum.mit.edu).

J. Lavaei is with the Department of Industrial Engineering and Operations Research, University of California, Berkeley (email: lavaei@berkeley.edu).

R. Baldick is with the Department of Electrical and Computer Engineering, University of Texas at Austin (email: baldick@ece.utexas.edu).

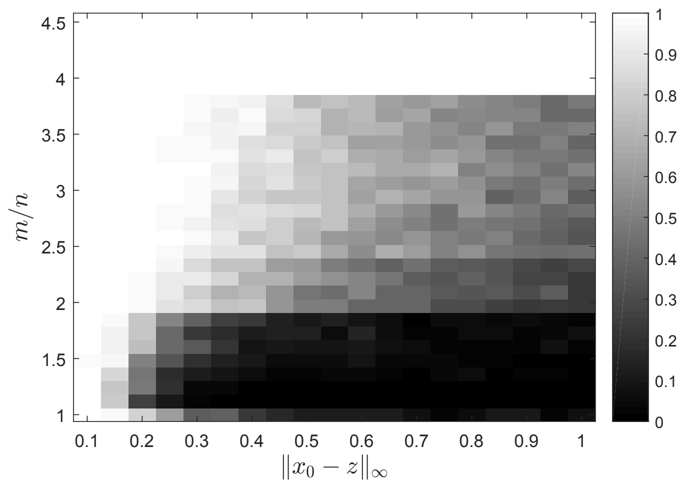

Figure 1: For a sufficiently large number of measurements, *Gauss–Newton always succeeds*, *even when starting from a random initial point*. The shade of each square represents the success rate of Gauss–Newton state estimation on the IEEE 14-bus system over 100 trials, with a lighter shade corresponding to a higher probability of success. The horizontal axis varies the quality of the initial point, from the true system state to a completely random vector. The vertical axis varies the number of measurements. (See Section VI for details.)

A. State estimation via nonlinear least squares

To this day, variants of the nonlinear least-squares procedure of Schweppe [2], [9] remain the most common approach for static AC state estimation. Given an N-bus power system with unknown voltage phasors $z \in \mathbb{C}^N$, the state estimation problem seeks to recover z from a set of SCADA measurements $b_1, \ldots, b_m \in \mathbb{R}$, where each i-th measurement

$$b_i = f_i(z) + \epsilon_i \tag{1}$$

comprises the output of a known "model" function $f_i(\cdot)$ and an unknown measurement noise ϵ_i with known variance $1/w_i$. The Schweppe algorithm estimates z by solving the weighted nonlinear least squares

$$\underset{u \in \mathbb{C}^N}{\text{minimize}} \ \frac{1}{2} \sum_{i=1}^{m} w_i [f_i(u) - b_i]^2 \tag{SEP}$$

using the Gauss-Newton method with a polar parameterization of u, starting from an initial guess $u^0 \in \mathbb{C}^N$. Convergence to a local minimum[1] is guaranteed by adjusting the step-sizes, or by adopting a trust-region strategy, as in the Levenberg–Marquardt algorithm; see e.g. [11, Sec.10.3].

B. Nonconvexity and the issue of spurious local minima

Existing software based on the Gauss–Newton method can produce spurious state estimations. When this occurs, the conventional wisdom is to conclude that the problem has been unduly biased by "bad data" [12]–[15], meaning that a small number of measurements have been tainted with a significant amount of noise. The bad data problem is well-studied; there are a number of mature techniques to systematically detect and eliminate bad measurements [12], [14], [15, Sec.7].

However, it is possible to obtain spurious estimations even in the absence of measurement noise, due to the inherent nonconvexity of (SEP). The core issue is the existence of *spurious local minima*, which can cause "greedy" local search algorithms to become trapped at a spurious estimation. The danger lies in spurious estimations that are physically meaningful but far from the true system state. In Section III, we illustrate this hazard on a two-bus example.

The question arises as to whether spurious local minima present an issue for practical state estimation. Performing a large number of numerical experiments, we obtain a surprising finding. Using a sufficiently large number of redundant, noise-free, randomly sampled measurements, *Gauss–Newton always succeeds* in recovering the true system state, even when starting from a *random initial point* (see Figure 1). To put in another way, local search never gets stuck at a spurious local minimum. It is as if they do not exist from the perspective of the Gauss–Newton algorithm.

C. Main results

The goal of this paper is to offer a theoretical explanation for the empirical success of local search for state estimation. State estimation is a specific instance of the *low-rank matrix recovery* problem in machine learning. Recently, this problem was shown to contain no spurious local minima under variants of the *restricted isometry property* (RIP) assumption [16]–[18]. In this case, local search is guaranteed to succeed starting from any arbitrary initial point.

Can we use RIP to establish a similar global recovery guarantee for power system state estimation? We investigate this question in Section IV. While power system state estimation does satisfy a version of RIP, its associated constant is too large for existing results to be applied. In fact, it was recently shown that large RIP constants cannot prevent spurious local minima from existing. In view of this negative result, establishing a global recovery guarantee for state estimation is likely very hard (or even impossible).

Instead, we focus our attention on establishing a local recovery guarantee. In Section V, we prove the following statement.

[1]More accurately, the line-search/trust-region Gauss-Newton method converges to a second-order local minimum using a suitable randomization strategy [10].

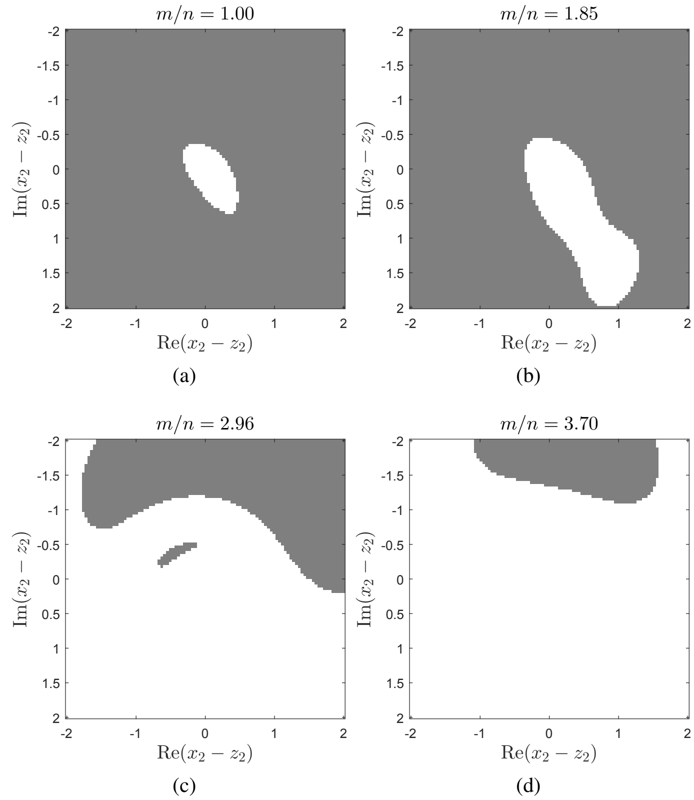

Figure 2: With an increasing number of measurements, our region with no spurious local minima (shown in white) grows to encompass a large part of the search space.

Theorem 1 (Informal)**.** *If a power system is **observable**, then the state estimation problem contains no spurious local minima within a **large** neighborhood of the true solution. Starting from a point within the neighborhood, gradient descent is guaranteed to converge at a linear rate to the global solution.*

A similar neighborhood of convergence can be established using a number of classical techniques, including the Banach and Kantorovich Theorems. The key feature of our region, however, is that it is large enough to offer a theoretical explanation for the empirical observations in Figure 1. For a sufficiently large number of measurements, our region with no spurious local minima grows to encompass a large part of the search space, as seen in Figure 2.

Our proof of Theorem 1 is based on a framework developed by Sun and Luo [19] for the general low-rank matrix recovery problem. Our results are much sharper, however, because state estimation is a symmetric rank-1 instance of matrix recovery. In particular, this allows us to use a number of bounds that are only valid in the rank-1 case.

D. Other approaches to state estimation

This paper considers the static AC formulation of state estimation as originally proposed by Schweppe [2]. Other formulations of the problem also exist, though these are not widely used in practice. For example, the so-called static "DC" formulation [20] linearizes the relationship between voltage and power, thereby making the resulting least squares problem convex and easy to solve. However, DC estimation is accurate only within a near-linear region of the underlying nonlinear model, and any inaccuracies can be greatly exacerbated in the presence of bad measurements and/or large modeling errors.

It is also possible to solve static AC state estimation to global optimality using *semidefinite programming* [21], [22]. The approach is known to enjoy a number of global recovery guarantees due to Madani et al. [23] and later Zhang et al. [24]. The primary disadvantage is the heavy computational and memory requirements, though these can be alleviated using chordal decomposition [25]–[28] and large-scale first-order algorithms like ADMM [29].

Recently, phase measurement units (PMUs) have been introduced into power systems, allowing for direct measurements of phase angles associated with bus voltage phasors. However, PMUs do not alleviate the quadratic nonconvexity inherent in (SEP). Spurious local minima can still exist with PMU measurements.

E. Related work on power flow

There is substantial literature and recent work on the power flow problem [30]–[35], which can be viewed as state estimation with the number of measurements m set to equal the number of degrees of freedom n. Here, spurious solutions exist because the equations are ambiguous: every solvable power equation admits a correct "high-voltage" solution and at least one spurious "low-voltage" solution [36], [37]. Recently, several lines of work have established a neighborhood around the true high-voltage power flow solution that contains no spurious "low-voltage" solutions [30], [31], [34]. In the other direction, conditions for the existence of solutions—either high- or low-voltage—have been completely characterized for certain classes of problems [35]. Viewing power flow as state estimation with $m = n$ measurements, however, both the high- and low-voltage solutions are globally optimal, as they would set the quadratic objective in (SEP) to zero. In this paper, we focus on a different issue: the presence of spurious local minima that are *not* globally optimal. An important future work is to understand the link between these two distinct but related notions. For example, it may be possible to adapt techniques for power flow solvability to give insights into how local search algorithms can fail, as opposed to this paper's emphasis on understanding when these algorithms succeed.

Notation

The sets $\mathbb{R}^n$, $\mathbb{C}^n$, $\mathbb{S}^n$, and $\mathbb{H}^n$ are the real and complex length-n vectors, and the real symmetric and complex Hermitian $n \times n$ matrices. Subscripts indicate element-wise indexing, the superscript "T" refers to the transpose, and the superscript "$*$" refers to the Hermitian transpose. We write $\mathbf{i} = \sqrt{-1}$ as the imaginary unit, and use $\operatorname{Re} z$ and $\operatorname{Im} z$ to refer to the real and imaginary parts, and $|z|$ to refer to the modulus. We use $X \succeq Y$ (resp. $X \succ Y$) to mean that $X - Y$ is positive semidefinite (resp. positive definite). We use $\lambda_{\max}(X)$ and $\lambda_{\min}(X)$ to refer to the most positive and least positive eigenvalues. The Euclidean norm of the vector x is $\|x\| = \left(\sum_{i=1}^n |x_i|^2\right)^{1/2}$. The spectral norm of a matrix M is $\|M\| = \max\{\|Mv\| : \|v\| = 1\}$ and its Frobenius norm is $\|M\|_F = \left(\sum_{i=1}^m \sum_{j=1}^n |M_{i,j}|^2\right)^{1/2}$.

II. FORMULATION

In this section, we review the classical formulation of the power system state estimation problem. We show that all classical SCADA measurements can be written in a standard quadratic form. Then, we characterize the local minima in the polar formulation of state estimation, and show that these have a one-to-one correspondence with their counterparts in the rectangular formulation.

A. SCADA measurements as quadratics

In state estimation, the classical SCADA measurements of nodal and branch powers, and voltage magnitudes, can all be posed as a homogeneous quadratic form

$$f_i(z) = z^* M_i z \qquad \text{where } M_i = M_i^* \tag{2}$$

with respect to a Hermitian measurement matrix M_i. It is this quadratic nature of $f_i(\cdot)$ that makes its corresponding least squares problem (SEP) nonconvex and strongly NP-hard in general.

Let us illustrate this on an N-bus power system with voltage phasors $z \in \mathbb{C}^N$. It is straightforward to see that any voltage magnitude measurement is a quadratic measurement

$$z_i^* z_i = z^*(e_i e_i^T) z \tag{3}$$

where e_j is the j-th column of the size-N identity matrix. Given that power is the product of voltage and current, and that current is generally assumed to be linear with respect to voltage, power must also be quadratic with respect to voltage.

To be more precise, let $Y_{i,j} \in \mathbb{C}$ be the admittance of the line or transformer from bus i to bus j, and let $Y_i \in \mathbb{C}$ be the shunt admittance at bus i. Then the current flowing from bus i to bus j is

$$c_{i\to j} = Y_{i,j}(z_i - z_j) = [Y_{i,j}(e_i - e_j)]^T z, \tag{4}$$

and the current injection at bus i is

$$c_i = \left[Y_i e_i + \sum_{j\in\mathcal{N}(i)} Y_{i,j}(e_i - e_j) \right]^T z, \tag{5}$$

where $\mathcal{N}(i)$ denotes the neighbors of bus i. The (complex) power injection at bus i is

$$p_i + \mathbf{i} q_i = c_i^* z_i = (z^* P_i z) + \mathbf{i}(z^* Q_i z), \tag{6}$$

where $P_i = \frac{1}{2}(S_i + S_i^*)$ and $Q_i = \frac{1}{2\mathbf{i}}(S_i - S_i^*)$ are the Hermitian splitting for

$$S_i = Y_i^* e_i e_i^T + \sum_{j\in\mathcal{N}(i)} Y_{i,j}^* (e_i - e_j) e_i^T.$$

Similarly, the power "sent" from the i-th bus to the j-th bus

$$p_{i\to j} + \mathbf{i} q_{i\to j} = c_{i\to j}^* z_i = (z^* P_{i\to j} z) + \mathbf{i}(z^* Q_{i\to j} z), \tag{7}$$

and the power "received" at the i-th bus due to the j-th bus

$$p_{i\leftarrow j} + \mathrm{i}q_{i\leftarrow j} = c^*_{j\rightarrow i} z_i = (z^* P_{i\leftarrow j} z) + \mathrm{i}(z^* Q_{i\leftarrow j} z), \quad (8)$$

can be written as quadratics where $P_{i\leftrightarrow j} = \frac{1}{2}(S_{i\leftrightarrow j} + S^*_{i\leftrightarrow j})$, $Q_{i\leftrightarrow j} = \frac{1}{2\mathrm{i}}(S_{i\leftrightarrow j} - S^*_{i\leftrightarrow j})$ are the Hermitian splitting for

$$S_{i\rightarrow j} = Y^*_{i,j}(e_i - e_j)e_i^T, \qquad S_{i\leftarrow j} = Y^*_{j,i}(e_j - e_i)e_i^T.$$

B. Polar parameterization

To solve the nonlinear least-squares problem (SEP) using the complex quadratic functions $f_i(\cdot)$ defined in (2), it is standard to express each complex variable in real polar form, as in

$$\underset{\substack{v\in\mathbb{R}^N \\ \theta\in 0\times\mathbb{R}^{N-1}}}{\text{minimize}} \ \frac{1}{2}\|W^{1/2}[\mathrm{pol}(v,\theta) - b]\|^2 \quad (9)$$

where $W = \mathrm{diag}(w_1, \ldots, w_m)$ is the diagonal weight matrix, $b = [b_i]_{i=1}^m$ is a vector, $\mathrm{pol}(v,\theta) = [\mathrm{pol}_i(v,\theta)]_{i=1}^m$ is a real vector-valued function of real variables

$$\mathrm{pol}_i(v,\theta) = \begin{bmatrix} v_1 e^{\mathrm{i}\theta_1} \\ \vdots \\ v_N e^{\mathrm{i}\theta_N} \end{bmatrix}^* M_i \begin{bmatrix} v_1 e^{\mathrm{i}\theta_1} \\ \vdots \\ v_N e^{\mathrm{i}\theta_N} \end{bmatrix}$$

and M_i is the Hermitian matrix from (2). Note that we force $\theta_1 = 0$ in (9) to remove the redundancy associated with absolute phase. Assuming noiseless measurements, any choice of (v, θ) satisfying

$$b = \mathrm{pol}(v,\theta) \quad (10)$$

is a ***global minimum*** with an objective value of zero. The true system state is obviously a global minimum, though in general it may not be the unique global minimum.

Applying the Gauss–Newton method to (9) yields the original Schweppe algorithm [9]. Adjusting the step-size using a back-tracking line search guarantees convergence [11] to the *first-order optimality conditions*

$$\nabla\mathrm{pol}(v,\theta)^T W[\mathrm{pol}(v,\theta) - b] = 0, \quad (11)$$

where $\nabla\mathrm{pol}(v,\theta)$ is the $m \times (2N-1)$ Jacobian matrix for $\mathrm{pol}(v,\theta)$. Adopting a suitable randomization [10] and/or trust region strategy [38], we can further guarantee convergence to the *second-order optimality conditions*

$$\nabla\mathrm{pol}(v,\theta)^T W \nabla\mathrm{pol}(v,\theta) + \sum_{i=1}^m w_i \nabla^2 \mathrm{pol}_i(v,\theta)[\mathrm{pol}_i(v,\theta) - b_i] \succeq 0. \quad (12)$$

A local minimum is guaranteed to satisfy (11) and (12). It is easy to verify that a global minimum satisfying (10) must also satisfy (11) and (12). We call a point (v, θ) satisfying (11) with $\mathrm{pol}(v,\theta) \neq b$ a *spurious first-order critical point*. We call a point (v,θ) satisfying (11) and (12) with $\mathrm{pol}(v,\theta) \neq b$ a *spurious second-order local minimum.*

C. Rectangular parameterization

We can also solve (SEP) by expressing each complex variable in rectangular form, as in

$$\underset{\substack{\xi\in\mathbb{R}^N \\ \eta\in 0\times\mathbb{R}^{N-1}}}{\text{minimize}} \ \frac{1}{2}\|W^{1/2}[\mathrm{rec}(\xi,\eta) - b]\|^2, \quad (13)$$

where W and b are the same as in (9), and $\mathrm{rec}(\xi,\eta) = [\mathrm{rec}_i(\xi,\eta)]_{i=1}^m$ is defined with respect to

$$\mathrm{rec}_i(\xi,\eta) = \begin{bmatrix} \xi \\ \eta \end{bmatrix}^T \begin{bmatrix} \mathrm{Re}\, M_i & -\mathrm{Im}\, M_i \\ \mathrm{Im}\, M_i & \mathrm{Re}\, M_i \end{bmatrix} \begin{bmatrix} \xi \\ \eta \end{bmatrix}. \quad (14)$$

Again, we force $\eta_1 = 0$ in (13) to remove the redundancy associated with absolute phase.

The critical points and local minima of the rectangular formulation are far easier to characterize than those of the polar formulation. We rewrite (13) as the following

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \ g(x) \equiv \frac{1}{2}\sum_{i=1}^m (x^T A_i x - c_i)^2,$$

where $x = [\xi_1, \ldots, \xi_N, \eta_2, \ldots, \eta_N]^T$ parameterizes the $n = 2N-1$ degrees of freedom in (13), and each $c_i = \sqrt{w_i} b_i$ and each A_i is defined

$$A_i = \sqrt{w_i} \begin{bmatrix} I_N & 0 \\ 0 & 0 \\ 0 & I_{N-1} \end{bmatrix}^T \begin{bmatrix} \mathrm{Re}\, M_i & -\mathrm{Im}\, M_i \\ \mathrm{Im}\, M_i & \mathrm{Re}\, M_i \end{bmatrix} \begin{bmatrix} I_N & 0 \\ 0 & 0 \\ 0 & I_{N-1} \end{bmatrix}.$$

A local minimum is guaranteed to satisfy the first- and second-order optimality conditions

$$\nabla g(x) = \sum_{i=1}^m A_i x (x^T A_i x - c_i) = 0, \quad (15)$$

$$\nabla^2 g(x) = \sum_{i=1}^m \left[A_i (x^T A_i x - c_i) + 2 A_i x x^T A_i\right] \succeq 0. \quad (16)$$

Again, any choice of x satisfying $x^T A_i x = c_i$ for all i is a *global minimum*. We call a point x satisfying $\nabla g(x) = 0$ with $x^T A_i x \neq c_i$ a *spurious first-order critical point*, and a point x satisfying $\nabla g(x) = 0$ and $\nabla^2 g(x) \succeq 0$ with $x^T A_i x \neq c_i$ a *spurious second-order local minimum.*

The critical points and local minima of the rectangular formulation have a one-to-one correspondence with their counterparts in the polar formulation (9).

Theorem 2. *Given any arbitrary $u \in \mathbb{R} \times \mathbb{C}^{N-1}$, define ξ, η and v, θ to satisfy $u = \xi + \mathrm{i}\eta = ve^{\mathrm{i}\theta}$. Then (v, θ) is a first-order critical point (resp. second-order local minimum) for the polar formulation (9) if and only if (ξ, η) is a first-order optimal critical point (resp. second-order local minimum) for the rectangular formulation (13).*

Proof: The proof is a straightforward application of the chain rule; it can be found in [1]. ■

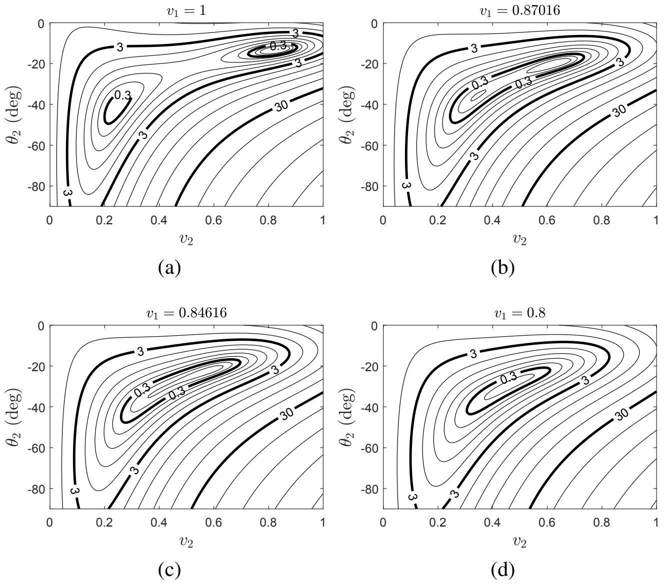

Figure 3: Contour plots of the least squares objective for the two-bus state estimation nonlinear least squares objective function.

III. Spurious local minima in a two-bus system

Even the simplest power systems with perfect, redundant measurements can suffer from spurious local minima. In this section, we consider a simple two-bus system whose state estimation problem admits two physically meaningful local minima. The problem has a unique, closed-form solution, so the potential for spurious estimations is entirely a limitation of local search. Some calculations reveal the spurious local minimum to be a distinct and unrelated phenomenon to the spurious "low-voltage" solution in the power flow problem. We first describe the system, then, in turn, discuss spurious solutions in state estimation and power flow.

A. Problem description

Consider a system with just two buses, connected by a single line with admittance

$$Y_{1,2} = Y_{2,1} = \frac{1}{0.01 + 0.1\mathbf{i}} \text{ per unit.}$$

To define a power flow problem on this system, we set bus 1 as the slack bus with voltage 1 p.u., and bus 2 as a PQ bus with a load of $2 + 1\mathbf{i}$ p.u. Solving this problem using Newton's method beginning with a flat start yields the following two phasors.

$$z_1 = 1, \qquad z_2 = 0.806 - 0.19\mathbf{i}.$$

To define a state estimation problem on this system, we make the following four noise-free measurements: 1) bus 1 voltage magnitude; 2) bus 2 real power injection; 3) bus 2 reactive power injection; 4) bus 1 real power injection. These are explicitly written

$$\begin{aligned} f_1(z) &= z_1^* z_1 &&= +1 \text{ p.u.},\\ f_2(z) &= \mathrm{Re}\,[(Y_{2,1}^*(z_2 - z_1)^* z_2] = p_2 &&= -2 \text{ p.u.},\\ f_3(z) &= \mathrm{Im}\,[Y_{2,1}^*(z_2 - z_1)^* z_2] = q_2 &&= -1 \text{ p.u.},\\ f_4(z) &= \mathrm{Re}\,[Y_{1,2}^*(z_1 - z_2)^* z_1] = p_1 &&\approx +2.07 \text{ p.u.}, \end{aligned}$$

using the formulas in (6)-(8). Note that the first three measurements coincide with the three equations in the power flow problem described above.

B. Spurious critical points and local minima in state estimation

The polar formulation of state estimation is the following problem

$$\underset{v_1, v_2, \theta_2 \in \mathbb{R}}{\text{minimize}} \; \|F(v_1, v_2 e^{\mathbf{i}\theta_2}) - b\|^2 \tag{17}$$

where $F(u) = [f_i(u)]_{i=1}^4$ and $b = F(z)$. Using a variable precision algebra toolbox to solve (15), we find the following critical points

$$\begin{bmatrix} v_1 \\ v_2 \\ \theta_2 \end{bmatrix} \in \left\{ \begin{bmatrix} 1 \\ 0.829 \\ -13.2^\circ \end{bmatrix}, \begin{bmatrix} 0.870 \\ 0.345 \\ -35.7^\circ \end{bmatrix}, \begin{bmatrix} 0.846 \\ 0.401 \\ -32.0^\circ \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \right\}, \tag{18}$$

that satisfy the first-order optimality condition (11) for the four measurements specified above. These residuals vector $r = F(v_1, v_2 e^{\mathbf{i}\theta_2}) - b$ at these four critical points have values

$$r \in \left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -0.24 \\ +0.14 \\ -0.06 \\ +0.17 \end{bmatrix}, \begin{bmatrix} -0.28 \\ +0.09 \\ -0.08 \\ +0.13 \end{bmatrix}, \begin{bmatrix} -1.00 \\ +2.00 \\ +1.00 \\ -2.07 \end{bmatrix} \right\},$$

with associated squared norms

$$\|r\|^2 \in \{0,\ 0.11183,\ 0.11299,\ 10.297\}.$$

The first point in (18) is clearly the global minimum, corresponding to the true system state. Sweeping the objective function as in Figure 3 reveals the second critical point to be a local minimum, the third to be a saddle point, and the last (the zero vector) to be a local maximum.

To highlight the hazards of spurious local minima, consider estimating v_2 and θ_2 using nonlinear least squares, while fixing the slack bus at $v_1 = z_1$. The objective function has contour plot shown in Figure 3a, and we see two local minima: the true system state at $v_2 \approx 0.8$ and $\theta_2 \approx -10^\circ$, and a spurious estimate at $v_2 \approx 0.2$ and $\theta_2 \approx -40^\circ$. A state estimator based on local refinement could converge to either estimates if the initial guess were set sufficiently close. Both local minima have physically meaningful (but unacceptable) values and small least squares residual values. Indeed, they would be virtually indistinguishable if the measurements were tainted with noise.

In this simple problem, state estimation has a unique, closed-form solution:

$$z_1 = \sqrt{f_1(z)},$$
$$\mathrm{Im}\, z_2 = \frac{f_2(z)\mathrm{Im}\, Y_{1,2} + f_3(z)\mathrm{Re}\, Y_{1,2}}{|Y_{1,2}|^2 z_1},$$
$$\mathrm{Re}\, z_2 = \frac{f_4(z)/z_1 + z_1 \mathrm{Re}\, Y_{1,2} + \mathrm{Im}\, z_2 \mathrm{Im}\, Y_{1,2}}{\mathrm{Re}\, Y_{1,2}}.$$

Hence, the potential for spurious estimations is entirely a limitation of local search approach. It is not a reflection of nonunique solutions, unobservable states, nor the inherent "hardness" of the underlying problem.

C. Relation to low-voltage power flow solutions

The power flow problem is well-known to admit spurious *low-voltage solutions*, in addition to the desired *high-voltage solution*. In the case of our two-bus model, these are the following

$$\begin{bmatrix} v_1 \\ v_2 \\ \theta_2 \end{bmatrix} \in \left\{ \begin{bmatrix} 1 \\ 0.829 \\ -13.2^\circ \end{bmatrix}, \begin{bmatrix} 1 \\ 0.271 \\ -44.5 \end{bmatrix} \right\}. \tag{19}$$

The first, high-voltage solution corresponds to the true system state, and as such satisfies all of our state estimation equations

$$F(v_1, v_2 e^{\mathbf{i}\theta_2}) = b,$$

where F and b are the same as in (17). The second, low-voltage solution only satisfies the first three state estimation equations corresponding to power flow

$$f_i(v_1, v_2 e^{\mathbf{i}\theta_2}) = b_i \qquad \forall i \in \{1, 2, 3\}.$$

This solution is spurious in the sense that it does not reflect the true system state. (Note that the fourth state estimation equation is not included as it would overspecify bus 1.)

The spurious local minima in state estimation are superficially similar to the spurious low-voltage solutions in power flow. Indeed, our two-bus example admits exactly one spurious local minimum in state estimation and one spurious low-voltage solution in power flow. Both spurious estimations have unacceptably low voltage magnitudes and large angle differences. Due to the overdetermined nature of state estimation, however, further numerical calculations find no concrete relations between the two concepts.

In one direction, let us write $\phi(v_1, v_2, \theta_2) \equiv \|F(v_1, v_2 e^{\mathbf{i}\theta_2}) - b\|^2$ as the least-squares objective function in (17). Then, the gradient of the objective $\nabla\phi$ at the low-voltage solution is very large:

$$\nabla\phi(v_1, v_2, \theta_2) = \begin{bmatrix} 4.4535 & 7.5595 & -2.5542 \end{bmatrix}^T.$$

Hence, the low-voltage solution is far from being a stationary point for state estimation. In the other direction, the power flow mismatch $r = [f_i(v_1, v_2 e^{\mathbf{i}\theta_2}) - b_i]_{i=1}^3$ at our three spurious critical points in (18) are also large:

$$r \in \left\{ \begin{bmatrix} -0.24 \\ +0.14 \\ -0.06 \end{bmatrix}, \begin{bmatrix} -0.28 \\ +0.09 \\ -0.08 \end{bmatrix}, \begin{bmatrix} -1.00 \\ +2.00 \\ +1.00 \end{bmatrix} \right\}.$$

Hence, the spurious critical points are poor solutions to power flow.

IV. Global recovery guarantees

The *low-rank matrix recovery* problem is one of the best understood nonconvex problems in machine learning. The simplest version seeks to recover a vector $z \in \mathbb{R}^n$, given measurement matrices $A_1, \ldots, A_m$ and quadratic measurements $b_i = z^T A_i z$, by solving the following

$$\underset{x \in \mathbb{R}^n}{\text{minimize}}\ g(x) \triangleq \frac{1}{2} \sum_{i=1}^m (x^T A_i x - b_i)^2 \tag{20}$$

using a local search algorithm starting from a *random initial point*. As previously shown in Section II-C, the rectangular formulation of state estimation is a specific instance of (20).

Recently, low-rank matrix recovery was shown to admit *no spurious local minima* [17] under a δ-restricted isometry property (δ-RIP) assumption

$$(1-\delta)\|X\|_F^2 \le C \cdot \sum_{i=1}^m [\mathrm{tr}\,(A_i X)]^2 \le (1+\delta)\|X\|_F^2$$
$$\text{for all } X \in \mathbb{R}^{n\times n} \text{ such that } \mathrm{rank}(X) \le 2 \tag{21}$$

with an arbitrary scaling $C > 0$ and a sufficiently small constant $\delta < 1/5$. Hence, local search is guaranteed to recover the true solution, starting from an arbitrary initial initial point, as if the function were convex. Similar guarantees are possible when (21) is only partially satisfied, as in the case of the matrix completion problem [16], [18]. If these guarantees are applicable to the rectangular formulation of state estimation, then they are automatically applicable to the polar formulation by virtue of Theorem 2.

While it is NP-hard to verify that a set of matrices $A_1, \ldots, A_m$ satisfy δ-RIP [39], computing a lower-bound on δ is relatively easy. Specifically, we perform a large number—say 100,000—of random trials. For the k-th trial, we sample an $n \times 2$ matrix U element-wise from the standard Gaussian, and evaluate the ratio $\gamma_k = \sum_i [\mathrm{tr}\,(A_i U U^T)]^2 / \|UU^T\|_F^2$. Then some algebra yields the lower-bound

$$\delta \ge (\gamma_{\max} - \gamma_{\min})/(\gamma_{\max} + \gamma_{\min})$$

where $\gamma_{\min}$ and $\gamma_{\max}$ are the smallest and largest ratios evaluated over all trials. In fact, this also gives a lower-bound for the "incoherent" version of RIP used to analyze sparse measurement matrices [16], [18].

Performing this experiment on power system matrices, we obtain much larger RIP constants than those typically found in machine learning. For example, the two-bus example in Section III has an RIP constant of $\delta \ge 0.9973$. In fact, it was recently shown that RIP constants of $\delta \ge 1/2$ are too large to prevent spurious local minima from existing [40]. Local search may still succeed with overwhelming probability if it is able to avoid and escape local minima. However, it is no longer possible to make global recovery guarantees under this regime.

V. Local recovery guarantees

Instead, we construct a *local region* about the true solution z that is guaranteed to contain no spurious local minima. More precisely, we enforce a strong-convexity–like inequality (due to Sun and Luo [19]) over this region:

$$\nabla g(x)^T(x-z) \geq \mu \cdot \|x-z\|^2 \tag{22}$$

with some constant $\mu > 0$. As such, the region must contain *no spurious critical points*, because any x that satisfies $\nabla g(x) = 0$ must set the left-hand side of (22) to zero. The following gives an inner approximation for the region in terms of a quadratic matrix inequality.

Proposition 3. *Define $H(x) = \sum_{i=1}^m A_i x(x+z)^T A_i$. Then (22) holds for some $\mu > 0$ if $H(x) + H(x)^T \succ 0$.*

Proof: Let $g(x) = \frac{1}{2}\sum_{i=1}^m (x^T A_i x - z^T A_i z)^2$. Then, some simple algebra shows that

$$\nabla g(x) = \sum_{i=1}^m A_i x(x+z)^T A_i (x-z) = H(x)(x-z)$$

$$\nabla g(x)^T(x-z) = \frac{1}{2}(x-z)^T[H(x)+H(x)^T](x-z).$$

Finally, note that if $H \succ 0$, then $x^T H x \geq \lambda_{\min}(H) x^T x$. ∎

The set of x satisfying $H(x) + H(x)^T \succ 0$ coincides with the set $\mathcal{S}_1$ in [1]. Their numerical results found this region to be highly nonconservative, growing eventually to encompass almost the entire search space.

Our main goal in this paper is to derive a *local neighorhood* $\|x-z\|^2 < \alpha$ that satisfies the strong-convexity–like inequality (22). Gradient descent is guaranteed to converge linearly to the global solution if initialized within the neighborhood.

Proposition 4. *Suppose that (22) holds for all $\|x-z\|^2 \leq \alpha$, and let x_0 satisfy $\|x_0 - z\|^2 \leq \alpha$. Then gradient descent $x_{k+1} = x_k - t\nabla g(x_k)$ with a sufficiently small step-size $t > 0$ converges at a linear rate to z.*

Proof: The function g is a quartic polynomial, and its gradient ∇g is a cubic polynomial, so there must exist a *fixed* Lipschitz constant L such that

$$\|\nabla g(x)\| \leq L\alpha\|x-z\| \text{ holds for all } \|x-z\|^2 \leq \alpha.$$

Then, standard arguments show

$$\begin{aligned}\|x_{k+1}-z\|^2 &= \|x_k - z\|^2 - 2t\nabla g(x_k)^T(x_k - z) \\ &\quad + t^2\|\nabla g(x_k)\|^2, \\ &\leq (1 - 2\mu t + \alpha^2 L^2 t^2)\|x_k - z\|^2.\end{aligned}$$

Hence, the sequence converges at a linear rate with step-size $t < 2\mu/(\alpha^2 L^2)$. ∎

Our first main result gives a *sharp* estimate for the largest local neighborhood satisfying (22). We begin with an assumption analogous to the full column rank assumption in linear least squares.

Assumption 1. *There exists a choice of $y \in \mathbb{R}^m$ such that $\sum_{i=1}^m y_i A_i \succ 0$.*

Without Assumption 1, there may exist $z \neq 0$ that gives identically zero measurements $b_i = z^T A_i z = 0$ for all i. Such a solution lies in the "null space" of the quadratic measurements, and can never be recovered using least-squares.

Theorem 5. *Given the symmetric measurement matrices $A_1, \ldots, A_m$ and the solution vector $z \in \mathbb{R}^n$, define*

$$\alpha = \min_{\|h\|=1} \frac{\sum_{i=1}^m (z^T A_i h)^2}{\sum_{i=1}^m (h^T A_i h)^2}. \tag{23}$$

Then, under Assumption 1, all x within the open neighborhood $\|x-z\|^2 < \alpha$ satisfy the strong-convexity–like inequality (22).

Proof: The proof is given in Appendix A. ∎

Theorem 5 describes a neighborhood containing no spurious local critical points:

$$\nabla g(x) = 0, \quad \|x-z\|^2 < \alpha \quad \Longrightarrow \quad x = z. \tag{24}$$

The bound is sharp in the sense that α cannot be improved for certain choices of $A_1, \ldots A_m$. To see this, let us adopt δ-RIP (21) to yield

$$\frac{\sum_{i=1}^m (z^T A_i h)^2}{\sum_{i=1}^m (h^T A_i h)^2} \geq \frac{(1-\delta)\|hz^T\|_F^2}{(1+\delta)\|hh^T\|_F^2} = \left(\frac{1-\delta}{1+\delta}\right)\|z\|^2.$$

In the perfect RIP case with $\delta = 0$, this yields $\alpha \geq \|z\|^2$. But $\alpha = \|z\|^2$ is the largest choice of α that still satisfies (24), because the origin is always a trivial critical point $\nabla g(0) = 0$. Hence, (23) attains $\alpha = \|z\|^2$, and is optimal in this regime.

Nevertheless, the exact value of α is defined in terms of a nonconvex optimization problem in (23). Below, we derive a lower-bound on α using a convex relaxation.

Theorem 6. *Let the Hessian matrix $\nabla^2 g(z) = \sum_{i=1}^m A_i z z^T A_i$ be positive definite at the solution. Then, we have $\alpha \geq \alpha_{\text{lb}}$, where*

$$\alpha_{\text{lb}} \triangleq 1/\lambda_{\max}\left[\sum_{i=1}^m A_i \left(\sum_{j=1}^m A_j z z^T A_j\right)^{-1} A_i\right]. \tag{25}$$

Proof: The proof is given in Appendix B. ∎

In power system state estimation, the matrix $\nabla^2 g(z) = \sum_{i=1}^m A_i z z^T A_i$ is known as the *gain matrix* [41], and coincides with the Gauss–Newton matrix at the solution. It is nonsingular (and hence positive definite) if and only if the system is fully *observable* at state z [41]–[43]. Power system measurements are placed specifically to maximize observability, so the assumption in Theorem 6 is mild and generally satisfied in practice.

Theorem 6 suggests prioritizing *observability* and *diversity* in choosing state estimation measurements. This becomes clearer in the following lower-bound:

$$\alpha \geq \alpha_{\text{lb}} \geq \frac{\lambda_{\min}\left[\nabla^2 g(z)\right]}{\lambda_{\max}\left[\sum_{i=1}^m A_i^2\right]}.$$

As described above, we can make the numerator $\lambda_{\min}[\nabla^2 g(z)]$ grow quickly by selecting measurements A_i to maximize observability. At the same time, we can make the denominator $\lambda_{\max}\left[\sum_{i=1}^m A_i^2\right]$ grow slowly by selecting diverse measurements. For example, if the measurements A_i are sampled uniformly at random, then the denominator grows as $O(m/n)$, which in power systems is bounded by a fixed constant.

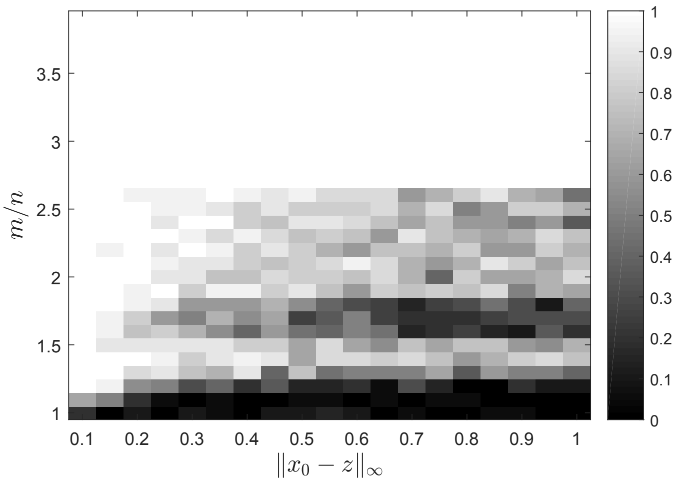

Figure 4: State estimation on the IEEE 39-bus system

VI. Numerical experiments

Finally, we perform study the empirical success rate of state estimation, and use the theoretical results developed throughout this paper to explain our findings. The systems we examine are the standard IEEE 14-bus and 39-bus test systems.

A. Problem description

Given a power flow case on an N-bus power system, we declare the high-voltage solution to be the true system state z, containing $n = 2N - 1$ degrees of freedom. We enumerate all SCADA measurements described in Section II-A: nodal voltage magnitude and real/reactive power injection, and branch real/reactive power injection/absorption. This results in 3 measurements for each node, and 4 measurements for each branch, for a total of $m_{\max} \approx 7N$ measurements. We reorder the list of measurements to make the first $n = 2N-1$ coincide with the original power flow problem, and then randomly shuffle the remaining $\approx 5N$ measurements.

B. Empirical success rate

We sweep the parameter pair (m, t) over $n \le m \le m_{\max}$ and $0 \le t \le 1$. For each fixed (m, t), we use the Levenberg–Marquardt variant of the Gauss–Newton method to solve state estimation on the first m measurements in our list, starting from the initial point

$$x_0 = z + t \cdot \frac{u + \mathbf{i}v}{\|u + \mathbf{i}v\|_\infty} \quad \text{where } u, v \sim \text{Gaussian}(0, I_n).$$

With $t = \|x_0 - z\|_\infty \approx 0$, the starting point is close to the true system state. This mimicks initializations used in practice, based on a flat start or the solution to a related problem. With $t = \|x_0 - z\|_\infty \approx 1$, the starting point becomes completely random. The trial is marked a "success" if the residual Euclidean norm drops below 10^{-9}, corresponding to an objective function value (and duality gap) of 10^{-18}. We repeat this trial 100 times and record the resulting success rate.

Results are shown in Figure 1 for the IEEE 14-bus system, and in Figure 4 for the IEEE 39-bus system. When the number of measurements m is small relative to the number of state variables n, the success rates rapidly drop to zero as the quality of the initial point deteriorates. As more redundant measurements are added, the success rates generally improve. Once a sufficiently large number of measurements are added, however, the success rate suddenly climbs to 100%. Despite the apparent nonconvexity, local optimization always converges to the true global optimum, even when starting from a random initial point.

C. Numerically evaluating the neighborhood

In Section IV, we derived a strongly-convex–like region about the true solution z that contains no spurious critical points. Proposition 3 gives an inner approximation for this region, based on a quadratic matrix inequality of the form $H(x) + H(x)^T \succ 0$. In this subsection, we compute 2D projections of this inner approximation while sweeping m over $n \le m \le m_{\max}$.

For each fixed m, we generate a 2D grid of test points x that match the true solution z except in two elements, which vary from -2 p.u. to $+2$ p.u. over 100 points in each direction. The exact elements we chose correspond to the real and imaginary parts of a particular voltage phasor. For each test point, we compute $H(x) + H(x)^T$ and attempt to compute its Cholesky factor. If it succeeds, then we mark the point as being inside the region; otherwise, we mark it as being outside.

Results are shown in Figure 2 for the IEEE 14-bus system and Figure 2 for the IEEE 39-bus system. As new measurements are added, the region grows to fill a vast portion of the entire space. In particular, it grows to encompass a ball of 2 p.u. radius, which is large enough to contain both a cold start as well as a random initial point of 1 p.u. magnitude.

D. Analytically bounding the neighborhood

Finally, we evaluate α_{lb} to lower-bound the radius of the neighorhood. The results are shown in Figure 6. We see that α_{lb} generally increases with the number of measurements m, though this is not guaranteed. Also, the actual values of α_{lb} are orders of magnitude smaller than the radii of the 2D projections computed in the previous subsection. This finding suggests conservatism in our derivation of α_{lb} in Theorem 6. Eliminating this conservatism is the subject of future work.

VII. Discussion

Overall, our empirical results indicate that spurious critical points in state estimation are made less likely by a diverse array of redundant measurements. Intuitively, it is very difficult for a large number of diverse observations to "conspire together" to point towards a spurious estimation. This intuition has been made precise in two special cases of the low-rank matrix recovery problem—matrix completion and matrix sensing [16], [17]. State estimation, however, is more complicated due to the presence of structure: the system topology is deterministic, and not all measurements

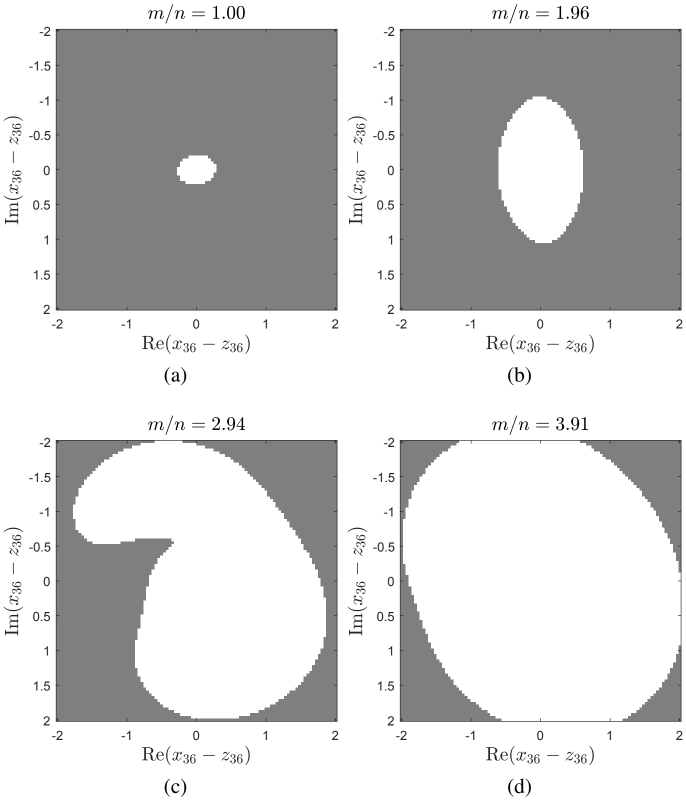

Figure 5: 2D projection of the region described Proposition 3 for the IEEE 39-bus test case.

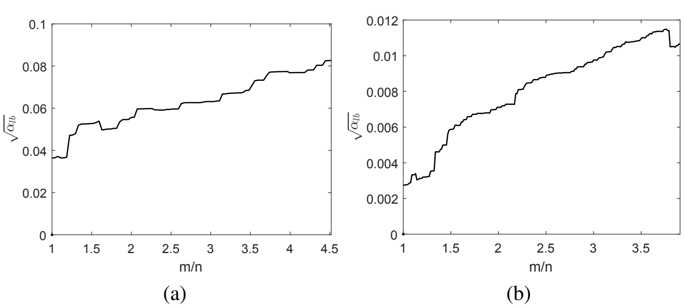

Figure 6: The value of $\sqrt{\alpha_{\text{lb}}}$ for four IEEE test cases: (a) 14-bus; (b) 39-bus.

are equally "good". Generalizing these prior arguments to the structured state estimation problem requires revisiting many mathematical concepts, and is left as future work.

For the most part, electric transmission systems are exhaustively measured, with a large number of measurements compared to unknowns. The results in Section VI seem to suggest that local convergence is not a significant issue for state estimation on real power systems. However, power system models are imprecise, with modeling errors hovering around 3%, and SCADA measurements are often noisy and spread out over a time interval. Measurement noise may create spurious critical points, though existing results for the matrix sensing problem suggest that these will not lie too far from the global minimum [17]. Another direction of future work is to extend our results in Section VI to the noisy case.

VIII. Conclusions

State estimation is a nonconvex, nonlinear least squares problem, that is NP-hard to solve in the general case. However, given a sufficiently large number of redundant, noise-free measurements, we observe that any local search algorithm is able to converge to the true solution, using an initial guess that is not necessarily close to the solution. In this paper, we develop a lower-bound on the distance between the true solution and the nearest spurious local minimum, and use it to numerically verify that critical points become increasing rare and far-away from the true solution with the addition of redundant information.

Appendix

A. Proof of Theorem 5

Our core approach is to convert the first- and second-order optimality conditions into matrix inequalities. We begin by proving three key lemmas.

Lemma 7. *Given the solution vector $z \in \mathbb{R}^n$, define the matrix-valued function F as the following*

$$F(h) = \begin{bmatrix} z^T \\ h^T \end{bmatrix} \left(\sum_{i=1}^m A_i h h^T A_i \right) \begin{bmatrix} z & h \end{bmatrix} \succeq 0. \qquad (26)$$

Then, the point $x \in \mathbb{R}^n$ is first-order optimal if it satisfies

$$2\nabla g(x)^T (x - z) = \begin{bmatrix} 4 & 3 \\ 3 & 2 \end{bmatrix} \bullet F(x - z) = 0, \qquad (27)$$

where $X \bullet Y \equiv \operatorname{tr}(XY)$ is the usual matrix inner product.

Proof: For simplicity, we write $h = x - z$. Some linear algebra yields

$$\begin{aligned} \nabla g(x)^T h &= \sum_{i=1}^m h^T A_i (z+h)(2z+h)^T A_i h \\ &= \operatorname{tr}(z+h)(2z+h)^T \underbrace{\left(\sum_{i=1}^m A_i h h^T A_i \right)}_{W}, \\ &= \frac{1}{2} \begin{bmatrix} 4 & 3 \\ 3 & 2 \end{bmatrix} \bullet \underbrace{\begin{bmatrix} z^T W z & z^T W h \\ h^T W z & h^T W h \end{bmatrix}}_{F(h)}. \end{aligned}$$

The third line notes that $W = W^T$ and $z^T W h = h^T W z$. ■

Lemma 8. *Under Assumption 1, there exists an absolute constant $c > 0$ such that $\sum_{i=1}^m (h^T A_i h)^2 \geq c \|h\|^4$ holds for all $h \in \mathbb{R}^n$.*

Proof: By the Cauchy–Schwarz inequality we have

$$\left(\sum_{i=1}^m y_i^2 \right) \left(\sum_{i=1}^m (h^T A_i h)^2 \right) \geq \left(\sum_{i=1}^m y_i h^T A_i h \right)^2 \geq \left(\gamma \|h\|^2 \right)^2$$

where $\gamma = \lambda_{\min} \left(\sum_{i=1}^m y_i A_i \right)$. Setting $c = \gamma^2 / \|y\|^2$ yields the desired bound. ■

Proof: Noting that $\left[\begin{smallmatrix}4&3\\3&2\end{smallmatrix}\right]-\left[\begin{smallmatrix}1&0\\0&-1\end{smallmatrix}\right]=\left[\begin{smallmatrix}3&3\\3&3\end{smallmatrix}\right]\succeq 0$, we have

$$\begin{aligned}2\nabla g(x)^T(x-z) &= \begin{bmatrix}4&3\\3&2\end{bmatrix}\bullet F(h) \geq \begin{bmatrix}1&0\\0&-1\end{bmatrix}\bullet F(h)\\ &= \sum_{i=1}^m (z^T A_i h)^2 - \sum_{i=1}^m (h^T A_i h)^2,\\ &\geq c\left(\alpha/\|h\|^2 - 1\right)\|h\|^4.\end{aligned}$$

The third line uses (23) and Lemma 8. Setting $\mu = \frac{1}{2}c(\alpha-\rho)$ yields the desired bound. ∎

B. Proof of Theorem 6

We will establish the following

$$\|h\|^2/\alpha_{\mathrm{lb}} \geq \min_{\rho\geq 0}\left\{\rho : \rho\sum_i A_i z z^T A_i \succeq \sum_i A_i h h^T A_i\right\},$$

which implies $\|h\|^2\sum_i (h^T A_i z)^2 \geq \alpha_{\mathrm{lb}}\sum_i (h^T A_i h)^2$ by multiplying the matrix inequality with hh^T. To do this, we factor $\sum_i A_i z z^T A_i = UU^T$ into its Cholesky factor U, which is invertible by hypothesis. Then, dividing by U on the left and U^T on the right yields

$$\rho I \succeq U^{-1}\sum_i A_i h h^T A_i U^{-T},$$

and the minimum value of ρ is

$$\begin{aligned}\rho &= \lambda_{\max}\left[U^{-1}\left(\sum_{i=1}^m A_i h h^T A_i\right)U^{-T}\right]\\ &= \|\begin{bmatrix}U^{-1}A_1 h & U^{-1}A_2 h & \cdots & U^{-1}A_m h\end{bmatrix}\|^2\\ &\leq \|\begin{bmatrix}U^{-1}A_1 h & U^{-1}A_2 h & \cdots & U^{-1}A_m h\end{bmatrix}\|_F^2\\ &= \|h^T\begin{bmatrix}A_1 U^{-1} & A_2 U^{-1} & \cdots & A_m U^{-1}\end{bmatrix}\|_2^2\\ &\leq \|h\|^2\|\begin{bmatrix}A_1 U^{-1} & A_2 U^{-1} & \cdots & A_m U^{-1}\end{bmatrix}\|^2\\ &= \|h\|^2/\alpha_{\mathrm{lb}}.\end{aligned}$$

The second line uses the definition of the spectral norm as the largest singular value. The third lind uses the Frobenius norm to bound the spectral norm. The fourth line vectorizes the matrix into a vector, while equating the matrix Frobenius norm with the vector 2-norm. The fifth line uses the spectral norm to bound the 2-norm.

References

[1] R. Y. Zhang, J. Lavaei, and R. Baldick, "Spurious critical points in power system state estimation," in *Hawaii Int. Conf. Syst. Sci.*, 2018.

[2] F. C. Schweppe and J. Wildes, "Power system static-state estimation, part i: Exact model," *IEEE Trans. Power. Ap. Syst.*, no. 1, pp. 120–125, 1970.

[3] J. G. Kassakian, R. Schmalensee, G. Desgroseilliers, T. D. Heidel, K. Afridi, A. Farid, J. Grochow, W. Hogan, H. Jacoby, J. Kirtley *et al.*, "The future of the electric grid," Massachusetts Institute of Technology, Tech. Rep., 2011.

[4] IVGTF Task 2.4, "Operating practices, procedures, and tools," North American Electric Reliability Corporation, Tech. Rep., 2011.

[5] Y. V. Makarov, V. I. Reshetov, A. Stroev, and I. Voropai, "Blackout prevention in the United States, Europe, and Russia," *Proc. IEEE*, vol. 93, no. 11, pp. 1942–1955, 2005.

[6] G. Andersson, P. Donalek, R. Farmer, N. Hatziargyriou, I. Kamwa, P. Kundur, N. Martins, J. Paserba, P. Pourbeik, J. Sanchez-Gasca *et al.*, "Causes of the 2003 major grid blackouts in North America and Europe, and recommended means to improve system dynamic performance," *IEEE Trans. Power Syst.*, vol. 20, no. 4, pp. 1922–1928, 2005.

[7] M. Patel, S. Aivaliotis, E. Ellen *et al.*, "Real-time application of synchrophasors for improving reliability," North American Electric Reliability Corporation, Tech. Rep., 2010.

[8] US-Canada Power System Outage Task Force, "Final report on the August 14, 2003 blackout in the United States and Canada: Causes and recommendations," US Department of Energy, Tech. Rep., 2004.

[9] F. C. Schweppe, "Power system static-state estimation, part iii: Implementation," *IEEE Trans. Power. Ap. Syst.*, no. 1, pp. 130–135, 1970.

[10] C. Jin, R. Ge, P. Netrapalli, S. M. Kakade, and M. I. Jordan, "How to escape saddle points efficiently," in *Int. Conf. Machine Learning*. PMLR, 2017, pp. 1724–1732.

[11] J. Nocedal and S. Wright, *Numerical optimization*, 2nd ed. Springer Science & Business Media, 2006.

[12] H. M. Merrill and F. C. Schweppe, "Bad data suppression in power system static state estimation," *IEEE Trans. Power. Ap. Syst.*, no. 6, pp. 2718–2725, 1971.

[13] F. C. Schweppe and E. J. Handschin, "Static state estimation in electric power systems," *Proc. IEEE*, vol. 62, no. 7, pp. 972–982, 1974.

[14] E. Handschin, F. C. Schweppe, J. Kohlas, and A. Fiechter, "Bad data analysis for power system state estimation," *IEEE Trans. Power. Ap. Syst.*, vol. 94, no. 2, pp. 329–337, 1975.

[15] A. Monticelli, "Electric power system state estimation," *Proc. IEEE*, vol. 88, no. 2, pp. 262–282, 2000.

[16] 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.

[17] S. Bhojanapalli, B. Neyshabur, and N. Srebro, "Global optimality of local search for low rank matrix recovery," in *Advances in Neural Information Processing Systems*, 2016, pp. 3873–3881.

[18] R. Ge, C. Jin, and Y. Zheng, "No spurious local minima in nonconvex low rank problems: A unified geometric analysis," in *Int. Conf. Machine Learning*. PMLR, 2017, pp. 1233–1242.

[19] R. Sun and Z.-Q. Luo, "Guaranteed matrix completion via non-convex factorization," *IEEE Trans. Inf. Theory*, vol. 62, no. 11, pp. 6535–6579, 2016.

[20] F. C. Schweppe and D. B. Rom, "Power system static-state estimation, part ii: Approximate model," *IEEE Trans. Power. Ap. Syst.*, no. 1, pp. 125–130, 1970.

[21] Y. Weng, Q. Li, R. Negi, and M. Ilić, "Semidefinite programming for power system state estimation," in *IEEE Power & Energy Society General Meeting*. IEEE, 2012, pp. 1–8.

[22] H. Zhu and G. B. Giannakis, "Robust power system state estimation for the nonlinear ac flow model," in *North American Power Symposium (NAPS)*. IEEE, 2012, pp. 1–6.

[23] R. Madani, M. Ashraphijuo, J. Lavaei, and R. Baldick, "Power system state estimation with a limited number of measurements," in *IEEE Conf. Decision & Control*. IEEE, 2016, pp. 672–679.

[24] Y. Zhang, R. Madani, and J. Lavaei, "Conic relaxations for power system state estimation with line measurements," *IEEE Trans. Control Network Syst.*, 2017.

[25] M. Fukuda, M. Kojima, K. Murota, and K. Nakata, "Exploiting sparsity in semidefinite programming via matrix completion I: General framework," *SIAM J. Optim.*, vol. 11, no. 3, pp. 647–674, 2001.

[26] M. S. Andersen, A. Hansson, and L. Vandenberghe, "Reduced-complexity semidefinite relaxations of optimal power flow problems," *IEEE Trans. Power Syst.*, vol. 29, no. 4, pp. 1855–1863, 2014.

[27] R. Y. Zhang and J. Lavaei, "Sparse semidefinite programs with guaranteed near-linear time complexity via dualized clique tree conversion," *arXiv:1710.03475*, 2017.

[28] ——, "Sparse semidefinite programs with near-linear time complexity," in *IEEE Conf. Decision & Control*. IEEE, 2018, pp. 1624–1631.

[29] R. Madani, A. Kalbat, and J. Lavaei, "ADMM for sparse semidefinite programming with applications to optimal power flow problem," in *IEEE Conf. Decision & Control*, 2015.

[30] K. Dvijotham, M. Chertkov, and S. Low, "A differential analysis of the power flow equations," in *IEEE Conf. Decision & Control*. IEEE, 2015, pp. 23–30.

[31] K. Dvijotham, S. Low, and M. Chertkov, "Solving the power flow equations: A monotone operator approach," *arXiv:1506.08472*, 2015.

[32] D. K. Molzahn, "Computing the feasible spaces of optimal power flow problems," *IEEE Trans. Power Syst.*, vol. 32, no. 6, pp. 4752–4763, 2017.

[33] K. Dvijotham, E. Mallada, and J. W. Simpson-Porco, "High-voltage solution in radial power networks: Existence, properties, and equivalent algorithms," *IEEE Control Syst. Lett.*, vol. 1, no. 2, pp. 322–327, 2017.

[34] K. Dvijotham, H. Nguyen, and K. Turitsyn, "Solvability regions of affinely parameterized quadratic equations," *IEEE Control Syst. Lett.*, vol. 2, no. 1, pp. 25–30, 2018.

[35] J. W. Simpson-Porco, "A theory of solvability for lossless power flow equations–Part II: Conditions for radial networks," *IEEE Trans. Control Netw. Syst.*, vol. 5, no. 3, pp. 1373–1385, 2018.

[36] F. Wu and S. Kumagai, "Steady-state security regions of power systems," *IEEE Trans. Circuits. Syst.*, vol. 29, no. 11, pp. 703–711, 1982.

[37] B. C. Lesieutre, P. W. Sauer, and M. Pai, "Existence of solutions for the network/load equations in power systems," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 46, no. 8, pp. 1003–1011, 1999.

[38] A. R. Conn, N. I. Gould, and P. L. Toint, *Trust region methods*. SIAM, 2000.

[39] A. M. Tillmann and M. E. Pfetsch, "The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 1248–1259, 2014.

[40] R. Y. Zhang, C. Josz, S. Sojoudi, and J. Lavaei, "How much restricted isometry is needed in nonconvex matrix recovery?" in *Advances in Neural Information Processing Systems*, 2018.

[41] F. F. Wu and A. Monticelli, "Network observability: theory," *IEEE Trans. Power. Ap. Syst.*, no. 5, pp. 1042–1048, 1985.

[42] G. Krumpholz, K. Clements, and P. Davis, "Power system observability: a practical algorithm using network topology," *IEEE Trans. Power. Ap. Syst.*, no. 4, pp. 1534–1542, 1980.

[43] A. Monticelli and F. F. Wu, "Network observability: Identification of observable islands and measurement placement," *IEEE Trans. Power. Ap. Syst.*, no. 5, pp. 1035–1041, 1985.

Richard Y. Zhang received the B.E. (hons) degree with first class honours in Electrical Engineering from the University of Canterbury, Christchurch, New Zealand, in 2009 and the S.M. and Ph.D. degrees in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology, Cambridge, MA, in 2012 and 2017 respectively. From 2017 to 2019, he was a Postdoctoral Scholar in the Department of Industrial Engineering and Operations Research at the University of California, Berkeley, CA. He is currently an Assistant Professor in the Department of Electrical and Computer Engineering at the University of Illinois, Urbana-Champaign.

His research interests are in computational mathematics and engineering, particularly in the solution of large-scale, nonconvex problems, and with applications in electric power systems, power electronics, data science, and machine learning.

Javad Lavaei is an Associate Professor in the Department of Industrial Engineering and Operations Research at University of California, Berkeley. He obtained his Ph.D. degree in Computing and Mathematical Sciences from California Institute of Technology in 2011. He has won multiple awards, including NSF CAREER Award, Office of Naval Research Young Investigator Award, AFOSR Young Faculty Award, DARPA Young Faculty Award, Google Faculty Research Award, Donald P. Eckman Award, Resonate Award, INFORMS Optimization Society Prize for Young Researchers, INFORMS ENRE Energy Best Publication Award, and SIAM Control and Systems Theory Prize. He is an associate editor of the IEEE TRANSACTIONS ON SMART GRID and of the IEEE CONTROL SYSTEMS LETTERS, and serves on the conference editorial boards of the IEEE Control Systems Society and European Control Association.

Ross Baldick (F'07) received the B.Sc. degree in mathematics and physics and the B.E. degree in electrical engineering from the University of Sydney, Sydney, Australia, and the M.S. and Ph.D. degrees in electrical engineering and computer sciences from the University of California, Berkeley, in 1988 and 1990, respectively. From 1991 to 1992, he was a Post-Doctoral Fellow with the Lawrence Berkeley Laboratory, Berkeley, CA. In 1992 and 1993, he was an Assistant Professor with the Worcester Polytechnic Institute, Worcester, MA. He is currently a Professor with the Department of Electrical and Computer Engineering, University of Texas, Austin.