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 quadratic 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].

A. State estimation via nonlinear least squares

Nonlinear least squares (NLS) for state estimation as originally suggested by Schweppe [2], [9] remains the most common approach to this day. Given an N-bus power system, we define the voltage phasors[1] as $z \in \mathbb{C}^N$. Then the state estimation problem is the problem of recovering z given a set of SCADA measurements $b_1, \dots, b_m \in \mathbb{R}$, where each i-th measurement

$$b_i = f_i(z) + \epsilon_i \text{ where } f_i(z) = z^* M_i z \tag{1}$$

is the sum of a known "measurement" function $f_i(\cdot)$ and an unknown measurement noise ϵ_i. (Here, z^* denotes the Hermitian conjugate of the vector z, and $M_i = M_i^*$ is a complex Hermitian matrix.) If the noises $\epsilon_1, \dots, \epsilon_m$ are selected independently and identically from the zero-mean normal distribution with variances $1/w_1, \dots, 1/w_m$, then the maximum likelihood estimator for z is the minimizer of the weighted nonlinear least squares problem

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

Starting from an initial guess $u^0 \in \mathbb{C}^N$, solving (SEP) using the Gauss–Newton method with a polar parameterization of u yields the classic Schweppe algorithm [9]. Convergence can be guaranteed (under mild assumptions) by adjusting the Gauss–Newton step sizes using backtracking line search, or by adopting a trust-region strategy, as in the Levenberg–Marquardt algorithm; see e.g. [10, Sec.10.3].

Schweppe's method solves the static, AC formulation of the power system state estimation problem, which we will simply refer to as "state estimation" for brevity. Other formulations of the problem also exist, though these are not widely used in practice. For example, the so-called static "DC" formulation [11] 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. Also, dynamic formulations of state estimation [12] assume that the voltage phasors are somewhat slowly-varying with time, so that past information may be used to inform future estimations. In reality, voltage phasors can change rapidly, and dynamic estimators can be susceptible to spurious estimations during transient events, like the tripping of a line or the loss of a generator.

This work was supported by ONR N00014-17-1-2933, DARPA Young Faculty Award, and AFOSR YIP Award. Parts of this work have appeared in *Proceedings of the Hawaii International Conference on System Sciences*, 2018 [1].

[1] Schweppe's formulation used a polar parameterization for the complex voltage phasors. We begin by formulating the problem in complex phasors directly, and then specialize to polar parameterization in Section II-B and rectangular parameterization in Section III-B.

B. Spurious estimates

Existing software for state estimation based on nonlinear least squares can produce spurious, nonsensical estimations, particularly during unusual or emergency conditions when accurate estimates are needed the most. When this occurs, the conventional wisdom is to conclude that the estimation has been unduly biased by "bad data" [13]–[16], meaning that some of the measurements may have noise variances considerably larger than expected. The influence of bad data can be reduced by reweighing the measurements according to their reliability; considerable work exists on the problem of iteratively refining these weights to improve the estimates [13], [15], [16, Sec.7].

However, it is possible to obtain spurious estimates even in the absence of measurement noise, due to the nonconvexity of (SEP). The maximum likelihood estimator is defined to be the *global minimizer*, but only first-order critical points (and second-order local minima) can be found in a reasonable amount of time using conventional algorithms. Indeed, several of Karp's 21 NP-complete problems can be posed as the global minimization of (SEP), so perfect state estimation without spurious estimates must be NP-hard in the worst case.

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), so spurious estimates are still possible. The state estimation problem with PMU measurements remains NP-hard.

C. Main results

This paper is motivated by a surprising observation: using a sufficiently large number of redundant, noise-free measurements, convergence to the global minimizer almost always occurs, even when the initial guess is not particularly close to the true solution. To put in another way, redundant measurements make the nonlinear least squares problem more convex.

Our main results seek to offer a theoretical explanation for this observation. State estimation (with a rectangular parameterization) is a specific instance of a nonconvex least squares problem over quadratic measurements, which we refer to as *quadratic recovery*. In Section III-A, we show that the nonconvex quadratic recovery problem becomes "effectively" strongly-convex within a neighborhood of the true solution (Theorem 4). If an initial point can be chosen to lie within the neighborhood, then gradient descent (taking sufficiently small steps) is guaranteed to converge onto the global solution.

In practice, state estimation is commonly performed using a polar parameterization. In Section III-B, we show that the critical points associated with the rectangular parameterization have a one-to-one correspondence with the critical points associated with the polar parameterization (Theorem 6). We also derive a lower-bound on the Euclidean radius of this neighborhood (Theorem 7). Using the bound, we show in Section IV that critical points of the nonconvex least squares objective become increasingly far-away from the true solution with the addition of redundant information.

D. Related work

State estimation is an instance of *quadratic recovery*, a class of problems that is ubiquitous in statistical and machine learning, with applications in matrix completion, matrix sensing, phase retrieval, and quadratic equations. While quadratic recovery is NP-hard in the worst-case, a line of seminal results from the 2000s [17]–[19] showed that it can be often solved to global optimiality using an approach known as *convexification*. The essential idea is to relax the nonconvex problem into a convex problem on a higher dimension, and to solve the resulting convex problem using a local search algorithm. For the state estimation problem, the convexification approach is well-known [20] and has been shown by Madani *et al.* [21] and later Zhang *et al.* [22] to enjoy a number of global optimality guarantees. The primary disadvantage of convexification is its heavy computational and memory requirements, though these can be somewhat reduced using chordal decomposition [23], [24] and large-scale first-order optimization algorithms like ADMM [25].

Nevertheless, the local optimization approach to quadratic recovery has remained highly popular, despite the apparent risk of getting stuck at a spurious critical point. A line of results developed for the matrix completion and matrix sensing problems have found that, given a sufficiently large number of *random* measurements (i.e. with matrices $M_1, \ldots, M_m$ random), the objective function in (SEP) admits no spurious local minima [26], [27]. Indeed, our Theorem 4 can be viewed as a generalization of a result by Sun and Luo [28] to *deterministic* measurements. It is interesting to note that these theoretical results all arrive at a similar conclusion—the nonconvex problem (SEP) can be made more convex by introducing a sufficiently large amount of redundant information. In this regard, state estimation is not a special case.

Finally, there is substantial literature on the spurious solutions of the power flow equations, which can be viewed as state estimation in which the number of measurements m is set to equal the number of degrees of freedom n. In this special case, spurious solutions exist because the equations are ambiguous: every power equation admits a correct "high-voltage" solution and at least one spurious "low-voltage" solution [29], [30]. A recent line of work establishes a neighborhood around the true "high-voltage" power flow solution that contains no spurious "low-voltage" solutions [31]–[33]. From the state estimation perspective, however, both the "high-voltage" and "low-voltage" solutions would be considered globally optimal, as they would all set the quadratic objective in (SEP) to zero. In this paper, we focuses on a separate and orthogonal issue: the presence of spurious critical points that are locally optimal but *not* globally optimal.

Notation

Upper-case letters denote matrices and lower-case letters denote vectors and scalars. Subscripts are used for element-wise indexing. The superscript "T" refers to the transpose, and the superscript "$*$" refers to the Hermitian transpose.

The sets $\mathbb{R}$ and $\mathbb{C}$ are the real and complex numbers. We write $\mathbf{i} = \sqrt{-1}$ as the imaginary unit. Given a complex number

$z \in \mathbb{C}$, we use $\operatorname{Re} z$ and $\operatorname{Im} z$ to refer to its real and imaginary parts, and $|z|$ to refer to its modulus. Given a vector $x \in \mathbb{R}^n$ or $x \in \mathbb{C}^n$, the norm notation $\|x\|$ refers to its Euclidean norm, i.e. $\|x\|^2 = \sum_{i=1}^n |x_i|^2$.

The sets $\mathbb{S}^n$ and $\mathbb{H}^n$ are the $n \times n$ real symmetric and complex Hermitian matrices. Given a matrix $X \in \mathbb{S}^n$ or $X \in \mathbb{H}^n$, we use $X \succeq 0$ and $X \succ 0$ to mean that the matrix X is positive semidefinite and positive definite respectively. We use $\lambda_{\max}(X)$ and $\lambda_{\min}(X)$ to refer to its most positive and least positive eigenvalues (or in the case that $X \succeq 0$, the largest and smallest eigenvalues). Given a compatible Y, we use $X \succeq Y$ and $X \succ Y$ to mean $X - Y \succeq 0$ and $X - Y \succ 0$ respectively.

For an arbitrary matrix $M \in \mathbb{R}^{m\times n}$ or $M \in \mathbb{C}^{m\times n}$, the norms $\|M\|$ and $\|M\|_F$ refer to the spectral norm and the Frobenius norm respectively, i.e. $\|M\| = \max\{\|Mv\| : \|v\| = 1\}$ and $\|M\|_F^2 = \sum_{i=1}^m \sum_{j=1}^n |M_{i,j}|^2$.

II. Critical Points in State Estimation

Even in the absence of measurement noise, state estimation can produce spurious estimations, due to the nonconvexity of the underlying least squares problem. In this section, we review the classical formulation for the state estimation problem in power systems. We give explicit examples of spurious estimations from first-order critical points on a simple two-bus example.

A. SCADA measurements as quadratics

In state estimation, the classical SCADA measurements of nodal and branch powers, and voltage magnitudes, can all be posed in the quadratic form shown in (1). Let us illustrate this on an N-bus power system. Writing the vector of voltage phasors as $z \in \mathbb{C}^N$ and the j-th column of the size-N identity matrix as e_j, the current phasor flowing from the i-th bus to the j-th bus is

$$c_{i\to j} = Y_{i,j}(z_i - z_j) = [Y_{i,j}(e_i - e_j)]^T z, \tag{2}$$

where $Y_{i,j} \in \mathbb{C}$ is the admittance of the corresponding line or transformer. The current injection at the i-th bus is the total current flowing from the node, which comprises a shunt-current together with current flows to the neighboring buses $\mathcal{N}(i)$, as in

$$c_i = \left[Y_i e_i + \sum_{j\in\mathcal{N}(i)} Y_{i,j}(e_i - e_j) \right]^T z, \tag{3}$$

where Y_i is the corresponding shunt-admittance at bus i. It is straightforward to see that any voltage magnitude measurement is a quadratic measurement

$$z_i^* z_i = z^*(e_i e_i^T) z. \tag{4}$$

Given that power is the product of voltage and current, and that the currents are themselves linear to voltage, the power measurements can also be written as quadratics with respect to z. For example, the (complex) power consumed at or injected into the i-th bus can be written

$$p_i + \mathbf{i} q_i = c_i^* z_i = (z^* P_i z) + \mathbf{i}(z^* Q_i z), \tag{5}$$

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

$$p_{i\to j} + \mathbf{i} q_{i\to j} = c_{i\to j}^* z_i = (z^* P_{i\to j} z) + \mathbf{i}(z^* Q_{i\to j} z), \tag{6}$$

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

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

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

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

The quadratic nature of these measurement functions makes the least squares problem (SEP) nonconvex.

B. Polar parameterization

It is standard to parameterize each complex voltage phasor candidate into its polar form $u_i = v_i e^{\mathbf{i}\theta_i}$. In the absence of measurement noise, this rewrites each measurement equation (1) into

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

and the nonlinear least squares problem (SEP) into an optimization over real variables

$$\underset{\substack{v\in\mathbb{R}^N \\ \theta\in 0\times\mathbb{R}^{N-1}}}{\text{minimize}} \ \frac{1}{2}\sum_{i=1}^m w_i[\operatorname{pol}_i(v,\theta) - f_i(z)]^2. \tag{8}$$

We force the angle of the first element of u to be zero in order to remove the redundancy associated with absolute phase. We do this because the measurements remain identical $f_i(u) = f_i(e^{\mathbf{i}\varphi}u)$ after an absolute phase shift by $\varphi \in [0, 2\pi)$ radians.

Applying the Gauss–Newton method to (8) yields the original Schweppe algorithm [9]. Adjusting the step-size using a back-tracking line search guarantees convergence [10] to a first-order optimal point $(\hat{v}, \hat{\theta})$ satisfying

$$\sum_{i=1}^m w_i[\operatorname{pol}_i(\hat{v},\hat{\theta}) - f_i(z)]\frac{\partial \operatorname{pol}_i(\hat{v},\hat{\theta})}{\partial v_1} = 0, \tag{9}$$

$$\sum_{i=1}^m w_i[\operatorname{pol}_i(\hat{v},\hat{\theta}) - f_i(z)]\begin{bmatrix}\partial \operatorname{pol}_i(\hat{v},\hat{\theta})/\partial v_j \\ \partial \operatorname{pol}_i(\hat{v},\hat{\theta})/\partial \theta_j\end{bmatrix} = 0, \quad \text{for all } j \in \{2,\dots,N\}. \tag{10}$$

If the point $(\hat{v},\hat{\theta})$ sets the residual $\operatorname{pol}_i(\hat{v},\hat{\theta}) - f_i(z) = 0$ to zero, then we say that it is *globally optimal*. The true solution $(\hat{v},\hat{\theta})$ that satisfies $\hat{v}_i e^{\mathbf{i}\hat{\theta}_i} = z_i$ is globally optimal, though other globally optimal solutions may also exist. It is also possible for $(\hat{v},\hat{\theta})$ to satisfy (9)-(10) without setting the residual to zero. Such a point is a *spurious* critical point, because it satisfies first-order optimality without being globally optimal.

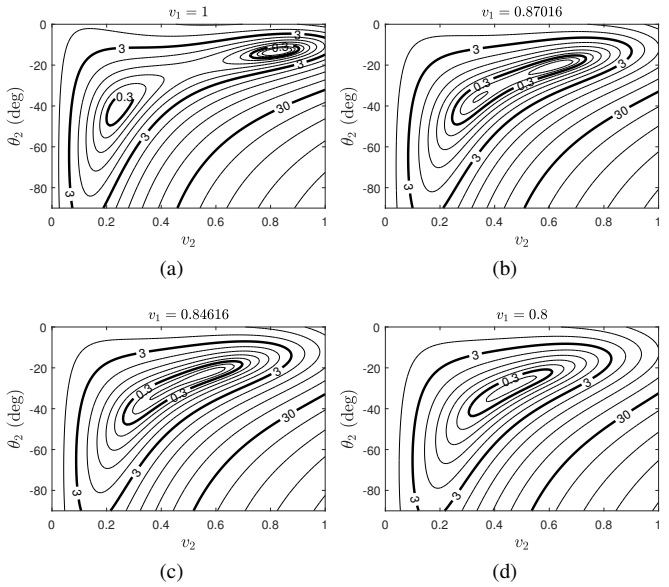

Figure 1: Contour plots of the least squares objective for the two-bus state estimation nonlinear least squares objective function.

It is worth emphasizing that the basic Gauss–Newton method does not even guarantee convergence to a local minimum. In practice, the algorithm can get stuck at saddle-points, which are also first-order optimal. Furthermore, the algorithm can break down altogether if the linear system of equations that defines the Gauss–Newton direction ever becomes singular. Both issues are addressed in the Levenberg–Marquardt algorithm, which reduces the Gauss–Newton method down to gradient descent if these issues arise; see [10, Sec.10.3] and also recent work on the ability of gradient descent to escape saddle points [26], [34], [35].

C. Example: Two-bus system

Even the simplest power systems with perfect, redundant measurements can suffer from spurious critical points. To give an illustration, consider a system with just two buses, connected by a single line with admittance

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

Setting bus-1 as the slack bus and loading bus-2 with a complex power load of $2 + 1\mathbf{i}$ per unit yields the two voltage phasors

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

The vector $z = [z_1; z_2]$ is the true system state. By construction, it is a global minimizer for any version of the nonlinear least squares problem (SEP) with noise-free measurements, and always has an objective value of zero.

To estimate z, let us define $v_1, v_2 \in \mathbb{R}$ as the two voltage magnitude estimates, $\theta_2 \in (-\pi, \pi]$ as the estimate for the angle of the second bus (keeping the angle of the first bus at zero). Consider making the following four noise-free measurements:

$$\begin{aligned} f_1(z) &= z_1^* z_1, \\ f_2(z) &= \operatorname{Re}\left[(Y_{2,1}^*(z_2 - z_1)^* z_2\right] = p_2, \\ f_3(z) &= \operatorname{Im}\left[Y_{2,1}^*(z_2 - z_1)^* z_2\right] = q_2, \\ f_4(z) &= \operatorname{Re}\left[Y_{1,2}^*(z_1 - z_2)^* z_1\right] = p_1, \end{aligned}$$

defined according to (5)-(7). Using a symbolic algebra toolbox, we discover the following critical points

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

that satisfy the condition (9)-(10) for the four measurements specified above. These four critical points have least squares objective values of

$$\{0,\ 0.11183,\ 0.11299,\ 10.297\}.$$

The first point is clearly the global minimum, corresponding to the true system state. Sweeping the objective function as in Figure 1 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 $v_1 = 1$. The objective function has contour plot shown in Figure 1a, 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, the true system state has a unique, closed-form solution:

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

Consequently, our inability to reliably estimate the true system state should be viewed as a shortcoming of the solution approach, rather than a reflection of the inherent "hardness" or "ambiguity" of the underlying estimation problem. For this specific example, the semidefinite relaxation approach of [22] is guaranteed to produce the true system state, even in the presence of some measurement noise.

III. Local Minima in Quadratic Recovery

In this section, we will consider a more general problem that we call *quadratic recovery*. Briefly, the problem seeks to

recover a fixed, unknown $z \in \mathbb{R}^n$, by solving a nonlinear least squares problem

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

given real symmetric measurement matrices $A_1, \dots, A_m \in \mathbb{S}^n$ and noise-free quadratic measurements $b_1, \dots, b_m \in \mathbb{R}$ satisfying

$$b_i \triangleq z^T A_i z.$$

Many important problems in statistical and machine learning are special instances of (QRP). As we explain in Section III-B below, state estimation is also a special case of (QRP).

Quadratic recovery is a nonconvex optimization problem. Conventional optimization algorithms based on local refinement (including the Newton, Gauss–Newton, gradient descent and their variants) are only guaranteed to converge onto a first-order optimal point satisfying

$$\nabla g(x) = \sum_{i=1}^{m} A_i x(x^T A_i x - b_i) = 0 \qquad \text{(FOC)}$$

called a *critical point* [10]. Trust region algorithms [36] (including the Nesterov's cubic regularization [37]) are additionally able to guarantee second-order optimality

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

A recent line of results showed that stochastic gradient descent [38] and regular gradient descent (using a random initial point) [34] also achieve second-order optimality with probability one. This result can be extended to algorithms that reduce to gradient descent at saddle-points, including the Levenberg–Marquardt algorithm.

If a first- and second-order optimal point x also sets the residual $x^T A_i x - b_i = 0$ to zero, then we say that it is *globally optimal*. The true solutions $x = \pm z$ are globally optimal, though other globally optimal solutions may also exist. Given the nonconvexity of (QRP), however, there will generally exist a number of other choices of x that are first- and second-order optimal without setting the residual to zero. These points are *spurious* local minima and saddle points.

Local algorithms have remained highly popular for quadratic recovery, despite the risk of getting stuck at a spurious local minimum or saddle point. Much of the initial interest was due to their speed, particularly in large-scale "big-data" applications where even a solution of modest accuracy is deemed sufficient [39]. However, later empirical evidence found that they are also accurate, in that they are able to consistently converge onto globally optimal solutions, regardless of the apparent nonconvexity [28], [40]–[45]. Local optimization remains the best performing approach in a wide range of applications, including rank-constrained semidefinite programming, matrix completion, matrix sensing, phase retrieval, and quadratic equations.

A. No spurious local minima in quadratic recovery

The strong practical performance of local algorithms for solving the nonconvex problem (QRP) has motivated significant theoretical interest in recent years. A number of authors have derived global optimality guarantees [40], [43], [46], or specific initializations that allow local algorithms to converge onto global solutions [28], [41], [42], [44], [45].

A particularly useful line of results was developed by Sun and Luo [28] for the matrix completion problem, which is a special instance of (QRP) with randomly chosen $A_1, \dots, A_m$. They showed that every x within a neighborhood of the solution $\|x - z\|^2 \le \rho$ satisfies a strong-convexity–like inequality

$$\nabla g(x)^T(x - z) \ge c\|x - z\|^2.$$

As a consequence, the neighborhood $\|x-z\|^2 \le \rho$ contains *no spurious critical points*—any first-order optimal point x within the neighborhood is guaranteed to be the global solution z. Moreover, if an initial point can be chosen to lie within the neighborhood, then gradient descent (taking sufficiently small steps) is guaranteed to converge onto the global solution.

Our main result in this subsection is a generalization of the Sun and Luo result for the general problem (QRP), which we use to analyze the spurious critical points and local minima of state estimation in Section III-B. We begin with the following restriction on the measurement matrices $A_1, \dots, A_m$, which is 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 the assumption, there may exist $z \neq 0$ that gives identically zero measurements $b_i = z^T A_i z = 0$ without being identically zero. Such a solution lies in the "null-space" of the quadratic measurements; it would be indistinguishable from the zero vector $x = 0$, which is also globally optimal for this special case.

Lemma 1. *Under Assumption 1, there exists an absolute constant $c > 0$ such that*

$$\sum_{i=1}^{m}(h^T A_i h)^2 \ge c\|h\|^4$$

holds for all $h \in \mathbb{R}^n$.

Proof: By the Cauchy–Schwarz inequality we have

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

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

Our core approach is to convert the first- and second-order optimality conditions into matrix inequalities.

Lemma 2. *Given the solution vector $z \in \mathbb{R}^n$, define the matrix-valued function F as the following*

$$F(h) = \begin{bmatrix} z^T \\ h^T \end{bmatrix}\left(\sum_{i=1}^{m} A_i h h^T A_i\right)\begin{bmatrix} z & h \end{bmatrix} \succeq 0. \qquad (11)$$

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

$$2\nabla g(x)^T(x-z) = \begin{bmatrix} 4 & 3 \\ 3 & 2 \end{bmatrix} \bullet F(x-z) = 0, \qquad (12)$$

and second-order optimal if it satisfies

$$(x-z)^T\nabla^2 g(x)(x-z) = \begin{bmatrix} 2 & 3 \\ 3 & 3 \end{bmatrix} \bullet F(x-z) \geq 0, \quad (13)$$

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

Proof: The proof is given in Appendix A. ∎

Lemma 3. *Under Assumption 1, the matrix-valued function F defined in (11) satisfies $F(h) = 0$ if and only if $h = 0$.*

Proof: If $h = 0$, then $F(h) = 0$. For the converse, note that the $(2,2)$ block of $F(h)$ requires $0 = \sum_i (h^T A_i h)^2 \geq c\|h\|^4$ via Lemma 1. ∎

Viewing the first- and second-order optimality conditions (FOC) and (SOC) as linear inequalities (12) and (13) imposed upon a semidefinite matrix variable $F(x-z) \equiv F \succeq 0$, we use Lagrangian duality to prove that $F = 0$ is the only feasible point that satisfies the inequalities. Equivalently, this is to show all first- and second-order optimal choices of x satisfy $F(x-z) = 0$, which under Assumption 1 is precisely $x = z$.

Theorem 4. *Given the symmetric measurement matrices $A_1, \ldots, A_m$ and the solution vector $z \in \mathbb{R}^n$, let*

$$\alpha = \min_{\|h\|=1} \frac{\sum_{i=1}^m (z^T A_i h)^2}{\sum_{i=1}^m (h^T A_i h)^2}. \qquad (14)$$

Then, under Assumption 1, we have

$$\|x-z\|^2 \geq 2\alpha,$$

for every spurious local minimum $x \neq z$ satisfying (FOC) and (SOC), and

$$\|x-z\|^2 \geq \alpha,$$

for every spurious critical minimum $x \neq z$ satisfying (FOC).

Proof: If x lies within the radius $\|x-z\|^2 \leq \rho$, then it must satisfy

$$\begin{bmatrix} \rho & 0 \\ 0 & -\alpha \end{bmatrix} \bullet F(x-z) \geq 0 \qquad (15)$$

due to (14). Hence, for every local minimum x satisfying (12), (13), and (15), the following must hold

$$G_\rho(y_1, y_2, y_3) \bullet F(x-z) \leq 0 \qquad (16)$$

for all y_1 and $y_2 \geq 0$, $y_3 \geq 0$, where

$$G_\rho(y_1, y_2, y_3) = y_1 \begin{bmatrix} 4 & 3 \\ 3 & 2 \end{bmatrix} - y_2 \begin{bmatrix} 2 & 3 \\ 3 & 3 \end{bmatrix} - y_3 \begin{bmatrix} \rho & 0 \\ 0 & -\alpha \end{bmatrix}.$$

Conversely, a choice of y_1, y_2, y_3 that satisfies

$$G_\rho(y_1, y_2, y_3) \succ 0, \quad y_2 \geq 0, \quad y_3 \geq 0 \qquad (17)$$

would certify the inexistence of a spurious local minimum $x \neq z$ lying within the radius $\|x-z\|^2 \leq \rho$, because it contradicts (16) for every choice of x such that $F(x-z) \neq 0$, and that under Assumption 1, we have $F(x-z) = 0$ if and only if $x = z$. Computing the largest radius is a quasi-convex maximization

$$\rho^\star = \sup_{y_1, y_2 \geq 0, y_3 \geq 0} \{\rho : G_\rho(y_1, y_2, y_3) \succ 0\},$$

with solution $\rho^\star = 2\alpha$ and $y_1^\star = y_2^\star = 1$ and $y_3^\star = 1/\alpha$. To prove $\|x-z\|^2 \geq \alpha$ for a spurious critical point x, we repeat the same arguments above but fix $y_2 = 0$, and obtain $\rho^\star = \alpha$ and $y_1^\star = 1$ and $y_3^\star = 1/\alpha$. ∎

Theorem 5. *Under the same conditions as Theorem 4, there exists a fixed $\mu > 0$ such that*

$$\nabla g(x)^T(x-z) \geq \mu\|x-z\|^2$$

holds for every x satisfying $\|x-z\|^2 \leq \rho < \alpha$.

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

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

The third line uses (14) and Lemma 1. Setting $\mu = \frac{1}{2}c(\alpha - \rho)$ yields the desired bound. ∎

It is helpful to view α as a measure of *local strong-convexity* for a given instance of (QRP). Within a neighborhood of the solution

$$\mathcal{B}_\alpha = \{x \in \mathbb{R}^n : \|x-z\|^2 < \alpha\},$$

the problem is essentially strongly convex. If we define a gradient Lipschitz constant L for this neighorhood

$$\|\nabla g(x)\| \leq L\|x-z\| \text{ holds for all } x \in \mathcal{B}_\alpha$$

then gradient descent with the fixed step-size t

$$x_{k+1} = x_k - t\nabla g(x_k)$$

satisfies the following

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

Hence, gradient descent is guaranteed to converge to z starting from an initial point $x_0 \in \mathcal{B}_\alpha$, so long as the step-size is small enough to satisfy $t < 2\mu/L^2$.

In one variant of quadratic recovery known as matrix completion, the measurement matrices $A_1, \ldots, A_m$ are randomly selected to satisfy a version of the restricted isometry property with overwhelming probability

$$(1-\delta) \leq \frac{\sum_{i=1}^m (v^T A_i u)^2}{\mathbf{E}\left[\sum_{i=1}^m (v^T A_i u)^2\right]} \leq (1+\delta), \qquad (18)$$

where δ is small, and the expectation (denoted by the symbol $\mathbf{E}$) in the denominator has value

$$\mathbf{E}\left[\sum_{i=1}^m (v^T A_i u)^2\right] = \tau\|uv^T + vu^T\|_F^2 \qquad (19)$$

for some $\tau > 0$. Accordingly, we have $\alpha \approx \|z\|^2/2$, so the problem contains no spurious local minima within neighborhood $\|x - z\|^2 \leq \|z\|^2$. Following this result, Ge, Lee and Ma [26] showed that (18) and (19) imply that any local minimum must satisfy $\|x - z\|^2 \leq O(\delta)\|z\|^2$. Since this neighborhood contains no spurious local minima, we must conclude that matrix completion contains no spurious local minima anywhere.

B. Application to state estimation

The *rectangular* parameterization of state estimation is an instance of quadratic recovery. We write each complex voltage phasor in (SEP) as $u_i = \xi_i + \mathbf{i}\eta_i$, and each measurement equation (1) in the absence of measurement noise as

$$f_i(u) = \mathrm{rec}_i(\xi, \eta) = \begin{bmatrix}\xi \\ \eta\end{bmatrix}^T \begin{bmatrix}\mathrm{Re}\, M_i & -\mathrm{Im}\, M_i \\ \mathrm{Im}\, M_i & \mathrm{Re}\, M_i\end{bmatrix} \begin{bmatrix}\xi \\ \eta\end{bmatrix}. \quad (20)$$

Then, the nonlinear least squares problem (SEP) is written

$$\underset{\substack{x\in\mathbb{R}^N \\ y\in 0\times\mathbb{R}^{N-1}}}{\text{minimize}} \ \frac{1}{2}\sum_{i=1}^m w_i[\mathrm{rec}_i(\xi,\eta) - f_i(z)]^2, \quad (21)$$

which we immediately recognize as an instance of (QRP), with $n = 2N - 1$ decision variables, and highly sparse data matrices

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

At the same time, the critical points associated with the rectangular parameterization have a one-to-one correspondence with the critical points associated with the polar parameterization.

Theorem 6. *Given any arbitrary $u \in \mathbb{R} \times \mathbb{C}^{N-1}$, define ξ, η and v, θ to satisfy $u = \xi + \mathbf{i}\eta = ve^{\mathbf{i}\theta}$. Then (v, θ) is a first-order optimal critical point for the polar parameterization of state estimation (8) if and only if (ξ, η) is a first-order optimal critical point for the rectangular parameterization of state estimation (21).*

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

Accordingly, we may adopt existing results on the quadratic recovery problem to make conclusions about state estimation in rectangular coordinates. Theorem 6 says that these are one and the same as the critical points for state estimation in the conventional, polar coordinates.

For deterministic choices of $A_1, \ldots, A_m$, the restricted isometry conditions (19) and (18) usually do not hold. In this case, computing α directly via its formula (14) is not much easier than solving (QRP) directly. Instead, we can lower-bound α using a technique from [1].

Theorem 7. *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 $\alpha \geq \alpha_{\mathrm{lb}}$, where*

$$\alpha_{\mathrm{lb}} \triangleq 1/\lambda_{\max}\left[\sum_{i=1}^m A_i \left(\sum_{j=1}^m A_j z z^T A_j\right)^{-1} A_i\right]. \quad (22)$$

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* [47], and is nonsingular (and hence positive definite) if and only if the system is fully *observable* at state z [47]–[49]. Power system measurements are placed specifically to achieve observability, so the assumption in Theorem 7 is mild and generally satisfied in practice.

When the matrices $A_1, \ldots, A_m$ are highly sparse, the bound α_{lb} is related to a property of the solution vector z known as *incoherence*. To explain, suppose that $A_1, \ldots, A_m$ were perfectly sparse, each containing exactly one nonzero element, as in

$$A_i = \beta_i e_{k_i} e_{k_i}^T \qquad i \in \{1, \ldots, m\},$$

where $k_i \in \{1, \ldots, n\}$ is the column/row number of the nonzero element, and $\beta_i \neq 0$ is the associated numerical value. (Recall that e_k refers to the k-th column of the size-n identity matrix) Then, so long as the matrix $\sum_{j=1}^m A_j z z^T A_j$ is nonsingular, we have

$$\sum_{i=1}^m A_i \left(\sum_{j=1}^m A_j z z^T A_j\right)^{-1} A_i = \sum_{i=1}^n z_i^{-2} e_i e_i^T,$$

and hence

$$\alpha_{\mathrm{lb}} = \min_i z_i^2.$$

In this special case, α_{lb} is completely independent of the measurement matrices $A_1, \ldots, A_m$. We see that α_{lb} is largest when the individual elements of z are spread-out, as in

$$|z_1| \approx |z_2| \approx \cdots \approx |z_n|.$$

In the literature, such a vector is said to be incoherent [18], and is known to be "easy" to recover, using either convex or nonconvex optimization.

In power system state estimation, the solution vector z contains the bus voltage phasors. These are always incoherent under normal conditions, because the voltage magnitude at each bus is always close to its nominal value of 1 per unit. This gives a heuristic explanation for why state estimation tends to work so well in practice, using sparse measurements that only directly infer a few voltages.

IV. The effect of redundant measurements

Consider adding additional redundant measurements to a given estimation problem. Intuitively, this introduces new information to the problem, so we would generally expect the "hardness" of the problem to be decreased. At the same time, it is certainly possible for the new measurements to make the nonlinear least squares problem (SEP) more nonconvex, thereby increasing the likelihood for the solution algorithm to get "stuck" at a local minimum.

Our numerical results in Section IV-A below suggest that (SEP) generally becomes less nonconvex with redundant measurements. In fact, once a sufficiently large number of measurements are added, the algorithm stops converging onto local minima altogether. In Section IV-B, we use the bounds derived in Section III to show that spurious critical points

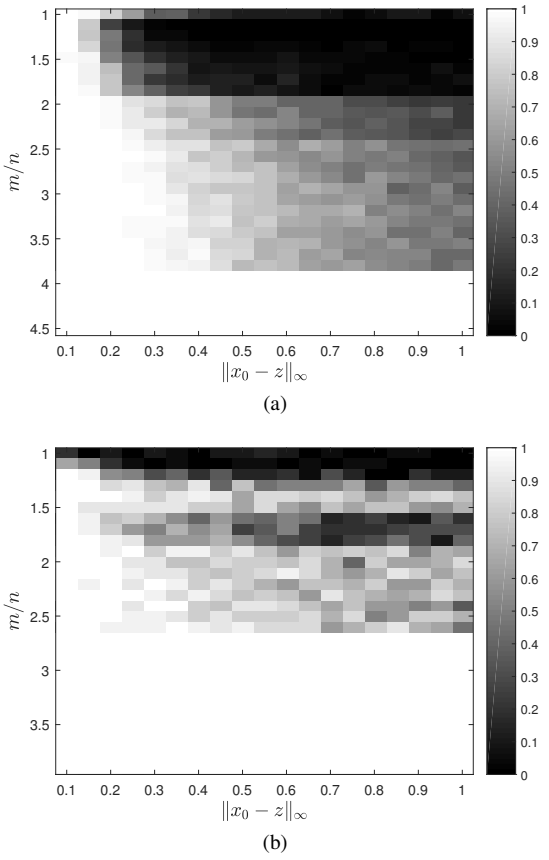

Figure 2: State estimation success rate on: (a) 14-bus system; (b) 39-bus system.

become increasing rare and far-away from the true solution with the addition of redundant information.

A. Numerical experiments

Figure 2 illustrates the success rate of quadratic recovery for the IEEE 14-bus and 39-bus systems, using noise-free measurements. Each square represents a single experiment over 100 trials (14-bus) or 20 trials (39-bus), and each trial attempts to solve the rectangular parameterization of (SEP) using the Levenberg–Marquardt algorithm (a trust-region variant of the Gauss–Newton algorithm), starting from a random initial guess. 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}. The color of each square represents the success rate of the corresponding experiment.

The experiments vary column-wise according to the infinity norm error $\|x_0 - z\|_\infty$ of the random initial guess x_0, which is fixed per experiment. To do this, we chose each x_0 as follows: (a) select a random direction $h \in \mathbb{R}^n$ on the $n+1$ dimensional sphere; (b) rescale h to the desired infinity norm; (c) output $x_0 = z + h$. The experiments vary row-wise according to the number of measurements m considered. The first row of experiments have a baseline $n = 2N - 1$ measurements that coincide with the underlying powerflow problem. In each

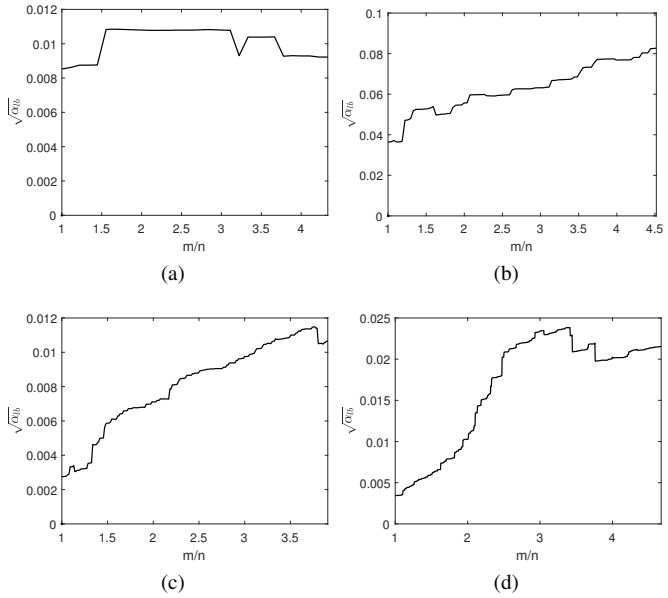

Figure 3: The value of $\sqrt{\alpha_{\mathrm{lb}}}$ for four IEEE test cases: (a) 5-bus; (b) 14-bus; (c) 39-bus; (d) 118-bus.

subsequent row, redundant measurements are added, one batch at a time, selected uniformly at random from the remaining measurements. We repeat this until reaching the final row, containing all measurements defined earlier in Section II-A.

Let us make a few observations about the results. First, when the number of measurements are small, the success rates rapidly drop to zero as the initial point is pulled further away from the solution. In the special case of $m \approx n$, the success rate is nonzero, even when the initial point is far from the solution. Upon closer examination, however, we find that the solution algorithm is not actually converging to the true solution z, but alternative solutions to the power flow equations.

As more redundant measurements are added, the success rates generally improve, although this is not guaranteed. Indeed, in the case of the 39-bus problem at around $m \approx 1.5n$, the success rates actually worsen once new measurements have been introduced. Also, some measurements seem to affect the success rate more than others. In particular, the measurements added at $m/n \approx$2, 2.5 and 4 for the 14-bus problem and $m/n \approx 1.5$, 2, 2.5 for the 39-bus problem appear to have the most effect.

Finally, with many redundant measurements, the success rates climb dramatically to 100%. Despite the apparent nonconvexity, local optimization always converges to the true global optimum. We always recover the true system state z.

B. Comparison with theory

In Section III, we showed that there is an "essentially strongly-convex" neighborhood surrounding the true solution z that contains no spurious critical points. If an initial point can be chosen to lie within the neighborhood, then gradient descent (taking sufficiently small steps) is guaranteed to converge onto

the global solution. In Theorem 7, we derived a lower-bound $\alpha_{\rm lb}$ on the size of this neighborhood.

As redundant measurements are added, our numerical results in the previous section suggests that the "essentially strongly-convex" neighborhood should generally become bigger. Any spurious critical point must become increasingly far-away from the solution. To confirm this suspicion, we evaluated $\alpha_{\rm lb}$ for the examples considered in the previous section. The results, plotted in Figure 7, found that $\alpha_{\rm lb}$ generally increases with the number of measurements m, though this is not guaranteed. The actual value of $\alpha_{\rm lb}$ suggests that an initial guess x_0 accurate to around $\|x_0 - z\| \approx 0.01$ should be enough to allow gradient descent to recover the true system system state, in spite of any nonconvexity in the problem.

V. 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 quadratic recovery problem—matrix completion and matrix sensing [26], [27]. State estimation, however, is more complicated due to the presence of structure: the system topology is deterministic, and not all measurements 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, power systems are exhaustively measured, with a large number of measurements compared to unknowns. The results in Section IV 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 [27]. Another direction of future work is to extend our results in Section IV to the noisy case.

VI. 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 Lemma 2

For simplicity, we write $h = x - z$. Then, completing the square yields

$$x^T A_i x - z^T A_i z = (x+z)^T A_i (x-z) = (2z+h)^T A_i h.$$

Substitituting $b_i = z^T A_i z$ and the above identity into (FOC) yields

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

Noting that $W = W^T$ and $z^T W h = h^T W z$, we have

$$\begin{aligned}2\nabla g(x)^T h =& 4z^T W z + 6 z^T W h + 2 h^T W h \\ =& \begin{bmatrix}4 & 3\\ 3 & 2\end{bmatrix} \bullet \underbrace{\begin{bmatrix} z^T W z & z^T W h \\ h^T W z & h^T W h\end{bmatrix}}_{F(h)},\end{aligned}$$

which is precisely (12). Applying the same technique with the same definition of W to (SOC) yields

$$\begin{aligned}h^T \nabla^2 g(x) h =& \operatorname{tr}\left[h(2z+h)^T + 2(z+h)(z+h)^T\right] W \\ =& 2z^T W z + 6 z^T W h + 3 h^T W h \\ =& \begin{bmatrix}2 & 3\\ 3 & 3\end{bmatrix} \bullet \begin{bmatrix} z^T \\ h^T\end{bmatrix} \left(\sum_{i=1}^m A_i h h^T A_i\right) \begin{bmatrix} z & h\end{bmatrix},\end{aligned}$$

which is (13).

B. Proof of Theorem 7

We will establish the following

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

which implies $\|h\|^2 \sum_i (h^T A_i z)^2 \geq \alpha_{\rm lb} \sum_i (h^T A_i h)^2$ by multiplying the matrix inequality with hh^T. To do this, we factor $\sum_i A_i z z^T A_i = U U^T$ 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

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

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 *Proceedings of the 51st Annual Hawaii International Conference on System Sciences*, 2018.

[2] F. C. Schweppe and J. Wildes, "Power system static-state estimation, part i: Exact model," *IEEE T. 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 T. Power. Ap. Syst.*, no. 1, pp. 130–135, 1970.

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

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

[12] A. S. Debs and R. E. Larson, "A dynamic estimator for tracking the state of a power system," *IEEE T. Power. Ap. Syst.*, no. 7, pp. 1670–1678, 1970.

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

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

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

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

[17] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[18] E. Candès and B. Recht, "Exact matrix completion via convex optimization," *Commun. ACM*, vol. 55, no. 6, pp. 111–119, 2012.

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

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

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

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

[23] 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.

[24] 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.

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

[26] 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.

[27] 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.

[28] 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.

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

[30] 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.

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

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

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

[34] J. D. Lee, M. Simchowitz, M. I. Jordan, and B. Recht, "Gradient descent only converges to minimizers," in *Conference on Learning Theory*, 2016, pp. 1246–1257.

[35] A. Anandkumar and R. Ge, "Efficient approaches for escaping higher order saddle points in non-convex optimization," in *Conference on Learning Theory*, 2016, pp. 81–102.

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

[37] Y. Nesterov and B. T. Polyak, "Cubic regularization of newton method and its global performance," *Math. Program.*, vol. 108, no. 1, pp. 177–205, 2006.

[38] R. Ge, F. Huang, C. Jin, and Y. Yuan, "Escaping from saddle points—online stochastic gradient for tensor decomposition," in *Conference on Learning Theory*, 2015, pp. 797–842.

[39] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, Aug 2009.

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

[41] R. H. Keshavan, A. Montanari, and S. Oh, "Matrix completion from a few entries," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2980–2998, 2010.

[42] ——, "Matrix completion from noisy entries," *Journal of Machine Learning Research*, vol. 11, no. Jul, pp. 2057–2078, 2010.

[43] M. Journée, F. Bach, P.-A. Absil, and R. Sepulchre, "Low-rank optimization on the cone of positive semidefinite matrices," *SIAM J. Optim.*, vol. 20, no. 5, pp. 2327–2351, 2010.

[44] P. Jain, P. Netrapalli, and S. Sanghavi, "Low-rank matrix completion using alternating minimization," in *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*. ACM, 2013, pp. 665–674.

[45] M. Hardt, "Understanding alternating minimization for matrix completion," in *IEEE 55th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2014, pp. 651–660.

[46] N. Boumal, V. Voroninski, and A. Bandeira, "The non-convex burer-monteiro approach works on smooth semidefinite programs," in *Advances in Neural Information Processing Systems*, 2016, pp. 2757–2765.

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

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

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