Spurious Critical Points in Power System State Estimation

Richard Y. Zhang and Javad Lavaei[*]
University of California, Berkeley
{ryz, lavaei}@berkeley.edu

Ross Baldick
University of Texas at Austin
baldick@ece.utexas.edu

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 critical points of the nonconvex least squares objective become increasing rare and far-away from the true solution with the addition of redundant information.

1. 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 [1]. State estimation proves *situational awareness* to a system operator, by allowing them 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 [2]. These functions are posed to become even more important as the penetration of wind and solar generation increases, due to the inherent variability and uncertainty of such resources [3].

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 [4, 5]. 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 [6]. Had an accurate, real-time state estimation been in place, the operators would have had warning of the impending problem and an opportunity to take remedial action [7].

1.1. State estimation via nonlinear least squares

Nonlinear least squares (NLS) for state estimation as originally suggested by Schweppe [1, 8] remains the most common approach to this day. Given an N-bus power system, we define the voltage phasors as $z \in \mathbb{C}^N$. Then the state estimation problem is the problem of recovering z given a set of SCADA measurements $b_1, \ldots, b_m \in \mathbb{R}$, where each i-th measurement

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

is the sum of a known "measurement" function $f_i(\cdot)$ and an unknown measurement error ε_i. (Here, z^* denotes the Hermitian conjugate of the vector z, and $M_i = M_i^*$ is a complex Hermitian matrix.) If the errors $\varepsilon_1, \ldots, \varepsilon_m$ are selected i.i.d. from the zero-mean normal distribution with variances $1/w_1, \ldots, 1/w_m$, then the maximum likelihood estimator for z is the minimizer of the weighted nonlinear least-squares problem

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

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 [8]. 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. [9, Sec.10.3].

1.2. Spurious estimates

Existing software based on the Gauss–Newton method 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", meaning that some of the measurements may have error 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 [10, 11, Sec.7].

However, it is possible to obtain spurious estimates even in the absence of measurement errors, due to the nonconvexity of (SEP). The maximum likelihood estimator is defined to be the *global minimizer*, but only first-order (and second-order) *critical points* can be found in a reasonable amount of time using conventional algorithms, which may not even be local minimizers when the objective is nonconvex. Indeed, several of Karp's 21 NP-complete problems can be posed as the global minimization of (SEP), so we must conclude that perfect state estimation (without spurious estimates) is NP-hard in the general case.

1.3. Main results

This paper is motivated by a surprising observation: using a sufficiently large number of redundant, error-free measurements, convergence to the global minimizer almost always occurs, even when the initial guess is not particularly close to the true solution. To offer a theoretical explanation for this finding, we derive in Section 3 a lower-bound on the distance between the true solution and the nearest spurious local minimum (Theorem 5). The bound is the solution to a quasi-convex optimization problem, and can be evaluated in polynomial time using an interior-point method. Using the bound, we show in Section 4 that critical points of the nonconvex least squares objective become increasing rare and far-away from the true solution with the addition of redundant information. Put in another way, redundant measurements make the nonlinear least squares problem more convex.

1.4. Related work

State estimation can be viewed a special instance of a rank-constrained semidefinite program (SDP) [12]—a linear optimization over the positive semidefinite matrix variable $X \succeq 0$, subject to $\operatorname{rank} X = 1$. The rank constraint makes the overall problem nonconvex, and it is standard to "convexify" by replacing it with a convex surrogate. The resulting problem was shown by Madani *et al.* [13] and later Zhang *et al.* [14] to enjoy a number of global convergence guarantees. The primary disadvantage of the convex approach is its heavy computational and memory requirements, though these can be somewhat reduced using chordal decomposition [15, 16] and large-scale first-order optimization algorithms like ADMM [17].

Recently, equivalent versions of the nonconvex problem (SEP) have been studied in the greater context of rank-constrained SDPs. 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 [18, 19]. While these previous results do not directly apply to state estimation (as power systems are not random), it is interesting to note that they 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.

This paper uses "state estimation" to refer to the static, AC formulation of the problem, while noting that there also exists a formulation for the so-called "DC power flow model" [20]. The "DC" model linearizes the relationship between voltage and power, so the resulting least squares problem is 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, state estimation can be formulated as a dynamic problem [21]. By assuming that the voltage phasors are somewhat slowly-varying with time, past information can 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.

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. We write $\mathbf{i} = \sqrt{-1}$ as the imaginary unit. The calligraphic font is reserved for sets.

2. Critical Points in State Estimation

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

2.1. 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

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

where $Y_{i,j} \in \mathbb{C}$ is the directional 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 and flows to the neighboring buses $\mathcal{N}(i)$, as in

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

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

$$z_i^* z_i = z^*(e_i e_i^T)z. \quad (4)$$

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 the i-th bus can be written

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

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

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

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

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

can be written 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

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

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

2.2. Polar parameterization

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

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

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

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

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 [8]. Adopting a line search or a trust-region strategy guarantees convergence to a first-order optimal point $(\hat{v}, \hat{\theta})$ satisfying

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

$$\sum_{i=1}^m w_i [\mathsf{p}_i(\hat{v}, \hat{\theta}) - f_i(z)] \begin{bmatrix} \partial \mathsf{p}_i(\hat{v}, \hat{\theta})/\partial v_j \\ \partial \mathsf{p}_i(\hat{v}, \hat{\theta})/\partial \theta_j \end{bmatrix} = 0,$$

for all $j \in \{2, \ldots, N\}$.

If the point $(\hat{v}, \hat{\theta})$ matches the true solution z by satisfying $\hat{v}_i e^{\mathbf{i}\hat{\theta}_i} = z_i$, then (9) is trivially satisfied since

the residual $p_i(\hat{v}, \hat{\theta}) - f_i(z) = 0$ is zero. However, it is also possible for $(\hat{v}, \hat{\theta})$ to satisfy (9) without setting the residual to zero. Such a point is a *spurious* critical point, because it satisfies first-order optimality without being globally optimal.

2.3. 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 error-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 error-free measurements:

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

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

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

that satisfy the condition (9) 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 reveal the second critical point to be a local minimum, the third to be a saddle
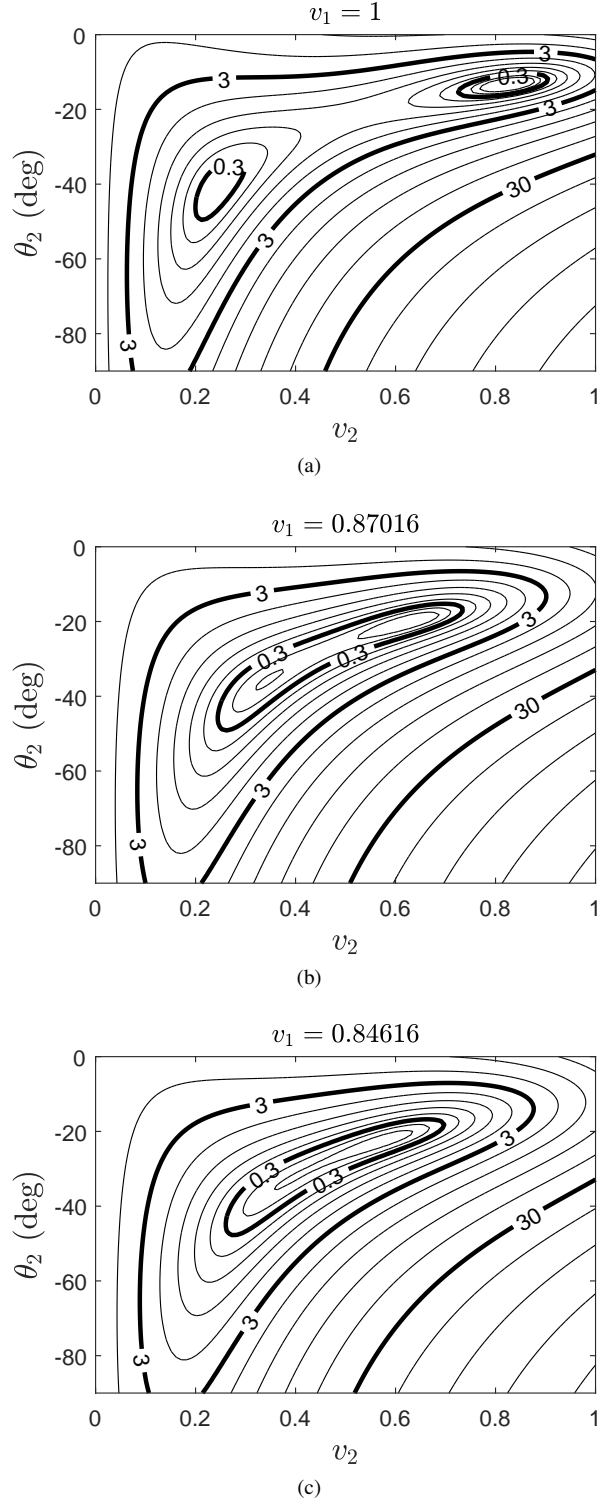

(a)

(b)

(c)

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

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 $\theta_2 \approx -10°$, and a spurious estimate at $v_2 \approx 0.2$ and $\theta_2 \approx -40°$. 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 values and small least-squares residual values. Indeed, they would be virtually indistinguishable if the measurements were tainted with error.

In this simple problem, the true system state has a unique, closed-form solution:

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

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 [14] is guaranteed to produce the true system state, even in the presence of some measurement error.

3. Critical Points in Quadratic Recovery

Our main theoretical tool for studying the critical points of state estimation is a mathematical insight on the *quadratic recovery problem*. Briefly, the quadratic recovery problem seeks to recover a fixed, unknown $z \in \mathbb{R}^n$, by solving a weighted nonlinear least-squares problem

$$\operatorname*{minimize}_{x \in \mathbb{R}^n} \sum_{i=1}^m w_i (x^T A_i x - b_i)^2 \qquad \text{(QRP)}$$

given error-free measurements $b_1, \ldots, b_m \in \mathbb{R}$, where

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

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

satisfying

$$\sum_{i=1}^m w_i A_i x (x^T A_i x - z^T A_i z) = 0, \qquad \text{(FOC)}$$

which we will refer to as a *critical point*. The true solutions $x = \pm z$ are obviously critical points, but given the nonconvexity of (QRP), there will generally exist a number of other choices of x that also satisfy (FOC). These correspond to local extrema and infliction points of the objective function, and we name them *spurious critical points*.

Now, consider the matrix-valued function

$$H(h, w) = \sum_{i=1}^m w_i A_i (z + h)(2z + h)^T A_i,$$

and note that (FOC) is equivalent to a null-space equation

$$[H(x - z, w)](x - z) = 0. \qquad (10)$$

We observe that any choice of $x \neq z$ that makes the matrix $H(x - z, w)$ nonsingular cannot be a critical point satisfying (FOC), because it is not possible for the equivalent statement (10) to be satisfied. Defining $\mathcal{S}_0$ as the set of all such choices of x yields a region within $\mathbb{R}^n$ that excludes all spurious critical points.

Proposition 1 (No spurious critical points). *Define*

$$\mathcal{S}_0 \triangleq \{x \in \mathbb{R}^n : H(x - z, w) \text{ is invertible}\}.$$

Then any $x \in \mathcal{S}_0 \backslash \{z\}$ is not a critical point, i.e. it does not satisfy (FOC).

The result can be conservative, meaning that $\mathcal{S}_0$ is not always the largest possible set without any spurious critical points. Nevertheless, our numerical experience found $\mathcal{S}_0$ to be very large, suggesting that Proposition 1 is reasonably nonconservative for our applications. Indeed, if we assume that the matrices $A_1, \ldots, A_m$ are generic (e.g. if all of their entries were selected independently from a probability distribution), then $\mathcal{S}_0$ fills the entire space $\mathbb{R}^n$ except for an exponential number of discrete points.

In practice, it is more convenient to work with a conservative version of Proposition 1.

Corollary 2. *Define*

$$\mathcal{S}_1 := \{x \in \mathbb{R}^n : H(x - z, w) + H(x - z, w)^T \succ 0\},$$

Then any $x \in \mathcal{S}_1 \backslash \{z\}$ is not a critical point, i.e. it does not satisfy (FOC).

Proof. Any real matrix H is invertible if its symmetric projection $H + H^T$ is positive definite, so we have $\mathcal{S}_1 \subseteq \mathcal{S}_0$, meaning that $\mathcal{S}_1$ inherits all of the properties of $\mathcal{S}_0$. □

3.1. Application to state estimation

State estimation can be viewed a special, complex case of quadratic recovery, so we may use Proposition 1 and Corollary 2 to study its critical points. To explain, let us take a rectangular parameterization of each complex voltage phasor $u_i = x_i + \mathbf{i} y_i$ in (SEP), and write the measurement equations (1) in the absence of measurement errors as

$$f_i(u) = r_i(x,y) = \begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} \operatorname{Re} M_i & -\operatorname{Im} M_i \\ \operatorname{Im} M_i & \operatorname{Re} M_i \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \quad (11)$$

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

$$\underset{\substack{x \in \mathbb{R}^N \\ y \in 0 \times \mathbb{R}^{N-1}}}{\text{minimize}} \frac{1}{2} \sum_{i=1}^m w_i [r_i(x,y) - f_i(z)]^2, \quad (12)$$

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

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

At the same time, the critical points associated with the rectangular parameterization, which satisfy

$$\sum_{i=1}^m w_i [r_i(\hat{x},\hat{y}) - f_i(z)] \frac{\partial r_i(\hat{x},\hat{y})}{\partial x_1} = 0, \quad (13)$$

$$\sum_{i=1}^m w_i [r_i(\hat{x},\hat{y}) - f_i(z)] \begin{bmatrix} \partial r_i(\hat{x},\hat{y})/\partial x_j \\ \partial r_i(\hat{x},\hat{y})/\partial y_j \end{bmatrix} = 0,$$

for all $j \in \{2,\ldots,N\}$, have a one-to-one correspondence with the critical points associated with the polar parameterization.

Theorem 3. *Given $z \in \mathbb{R} \times \mathbb{C}^{N-1}$, define x,y and v,θ to satisfy $z = x + \mathbf{i}y = ve^{\mathbf{i}\theta}$. Then (v,θ) is a critical point satisfying the polar optimality conditions (9) if and only if (x,y) is a critical point satisfying the rectangular first-order optimality conditions (13).*

Proof. The proof is a straightforward application of the chain rule; it is included for completeness in Appendix A. $\square$

Accordingly, we may use Proposition 1 to study the critical points associated with the rectangular parameterization of (SEP). Theorem 3 says that these are one and the same as the critical points for the conventional, polar parameterization of (SEP)

3.2. The nearest critical point

One application of Corollary 2 is to lower-bound the distance between the solution and its closest spurious critical point. Conceptually, this can be posed as a robust optimization problem for the radius of the largest "ball" that fits within $\mathcal{S}_1$, as in

$$\rho_{\text{opt}} \triangleq \text{maximize } \rho \text{ subject to } \mathcal{B}(\rho) \subseteq \mathcal{S}_1, \quad (14)$$

where

$$\mathcal{B}(\rho) \triangleq \{x \in \mathbb{R}^n : \|x - z\|_\ell \leq \rho\}.$$

Given that $\mathcal{B}(\rho^\star)$ contains no spurious critical point (Corollary 2), $\rho^\star$ must be a lower-bound for the distance to the nearest critical point:

$$\|x - z\|_\ell \geq \rho_{\text{opt}} \text{ holds for all } x \neq z \text{ satisfying (FOC).}$$

Problem (14) can be solved by a bisection method if we have an efficient way of validating the inclusion $\mathcal{B}(\rho) \subseteq \mathcal{S}_1$. To describe how this is done, we define the "residual Jacobian" function as

$$J(x) \triangleq \begin{bmatrix} A_1 x & \cdots & A_m x \end{bmatrix}^T,$$

and note that

$$H(h,w) = J(z+h)^T W J(2z+h)$$

$$= \begin{bmatrix} J(z) \\ J(h) \end{bmatrix}^T \begin{bmatrix} 2W & 3W \\ 0 & W \end{bmatrix} \begin{bmatrix} J(z) \\ J(h) \end{bmatrix}$$

where $W = \operatorname{diag}(w)$. Accordingly, the inclusion $\mathcal{B}(\rho) \subseteq \mathcal{S}_1$ can be written

$$\begin{bmatrix} J(z) \\ J(h) \end{bmatrix}^T \begin{bmatrix} 4W & 3W \\ 3W & 2W \end{bmatrix} \begin{bmatrix} J(z) \\ J(h) \end{bmatrix} \succ 0$$

$$\text{holds for all } \|h\|_\ell \leq \rho. \quad (15)$$

Now, let us recall a classic result from control theory [22, Lem.1].

Lemma 4. *Given A,B,C,D,G and the uncertainty set $\mathcal{U}$, define $F(\Delta) = D + C\Delta(I - A\Delta)^{-1}B$ as a linear fractional transform. Then*

$$\begin{bmatrix} G \\ F(\Delta) \end{bmatrix}^T J \begin{bmatrix} G \\ F(\Delta) \end{bmatrix} \prec 0 \text{ holds for all } \Delta \in \mathcal{U}$$

if there exists a Hermitian multiplier P satisfying

$$\begin{bmatrix} I & 0 \\ A & B \end{bmatrix}^T P \begin{bmatrix} I & 0 \\ A & B \end{bmatrix} + \begin{bmatrix} 0 & G \\ C & D \end{bmatrix}^T J \begin{bmatrix} 0 & G \\ C & D \end{bmatrix} \prec 0 \quad (16)$$

while relating to $\mathcal{U}$ by

$$\begin{bmatrix} \Delta \\ I \end{bmatrix}^T P \begin{bmatrix} \Delta \\ I \end{bmatrix} \succeq 0 \text{ holds for all } \Delta \in \mathcal{U} \setminus \{\infty\},$$

$$\begin{bmatrix} I \\ 0 \end{bmatrix}^T P \begin{bmatrix} I \\ 0 \end{bmatrix} \succeq 0 \text{ holds for } \infty \in \mathcal{U}. \quad (17)$$

To give an example application of the lemma, suppose that we chose $\|\cdot\|$ to refer to the usual infinity norm. Then we use the singular value decomposition to obtain an LFR representation

$$J(h) = \sum_{j=1}^{n} h_j \left[A_1 e_j \quad \cdots \quad A_m e_j \right]^T = \sum_{j=1}^{n} h_j C_j B_j$$

$$= \underbrace{\left[C_1 \quad \cdots \quad C_n \right]}_{C} \underbrace{\begin{bmatrix} h_1 I_{v_1} & & \\ & \ddots & \\ & & h_1 I_{v_1} \end{bmatrix}}_{\Delta(h)} \underbrace{\begin{bmatrix} B_1 \\ \vdots \\ B_n \end{bmatrix}}_{B}.$$

The condition $\|h\|_\infty \le \rho$ is equivalent to $\Delta \in \mathcal{U}$, where

$$\mathcal{U} \triangleq \{ \Delta(h) : \sigma_{\max}[\Delta(h)] \le \rho \},$$

and the standard choice of multipliers P to enforce (17) is

$$P = \begin{bmatrix} -\text{diag}(Q_1, \ldots, Q_n) & \text{diag}(S_1, \ldots, S_n) \\ \text{diag}(S_1, \ldots, S_n)^T & \rho^2 \text{diag}(Q_1, \ldots, Q_n) \end{bmatrix}$$

in which each $Q_j = Q_j^T \succeq 0$ is positive semidefinite and each $S_j = -S_j^T$ is skew symmetric; see [23, Sec.3.4.3]. Applying Lemma 4, a conservative approximation for $\mathcal{B}(\rho) \subseteq \mathcal{S}_1$ is the following linear matrix inequality feasibility problem

$$\begin{bmatrix} 0 & J(z) \\ C & 0 \end{bmatrix}^T \begin{bmatrix} 4W & 3W \\ 3W & 2W \end{bmatrix} \begin{bmatrix} 0 & J(z) \\ C & 0 \end{bmatrix}$$
$$+ \begin{bmatrix} Q & -SB \\ -B^T S^T & -\rho^2 B^T Q B \end{bmatrix} \succeq 0, \quad (18)$$

where

$$Q = \text{diag}(Q_1, \ldots, Q_n) \succeq 0, \quad (19)$$
$$S = \text{diag}(S_1, \ldots, S_n) = S^T. \quad (20)$$

are multiplier variables.

Let us write $\mathcal{B}(\rho) \subseteq \mathcal{S}_2$ to mean that there exist Q, S to satisfy (18)-(20) for that choice of ρ, and $\mathcal{B}(\rho) \not\subseteq \mathcal{S}_2$ to mean otherwise. Lemma 4 says that the condition $\mathcal{B}(\rho) \subseteq \mathcal{S}_2$ implies $\mathcal{B}(\rho) \subseteq \mathcal{S}_1$, so we may replace the latter by the former in (14) to yield a lower-bound on the infinity-norm distance between the solution and the closest critical point.

Theorem 5. *Define ρ_{opt} as the optimal objective for*

$$\rho_{\text{opt}} \triangleq \text{maximize } \rho \text{ subject to } \mathcal{B}(\rho) \subseteq \mathcal{S}_2,$$

Then $\|x - z\|_\infty \ge \rho_{\text{opt}}$ holds for any critical point $x \ne z$ satisfying (FOC).

Since (18)-(20) are just a set of convex linear matrix inequalities, the existence of Q, S can be efficiently checked using an interior-point method as a convex feasibility problem. Then, a bisection method can be used to optimize over the precise value of ρ.

Algorithm 1 (Bisection). Inputs: Lower-bound $\underline{\rho}_1 = 0$, Upper-bound $\overline{\rho}_1 > 0$ satisfying $\mathcal{B}(\overline{\rho}_1) \not\subseteq \mathcal{S}_2$; Accuracy $\varepsilon > 0$,
   Do: For $k = 1, 2, \ldots$

1. Set $\rho = (\underline{\rho}_k + \overline{\rho}_k)/2$.

2. If the gap $\overline{\rho}_k - \underline{\rho}_k < \varepsilon$, then return ρ as an ε-approximation of ρ_{opt}. Otherwise, go to Step 3.

3. Check the condition $\mathcal{B}(\rho) \subseteq \mathcal{S}_2$. If the condition holds true, then update the lower-bound $\underline{\rho}_{k+1} = \rho$ and keep the upper-bound $\overline{\rho}_{k+1} = \overline{\rho}_k$. Otherwise, update the upper-bound $\overline{\rho}_{k+1} = \rho$ and keep the lower-bound $\underline{\rho}_{k+1} = \underline{\rho}_k$.

4. Go to Step 1.

Remark 6. The initial upper-bound $\overline{\rho}_1$ can be computed using an exponential line search, e.g. by incrementing β until $\mathcal{B}(2^\beta) \not\subseteq \mathcal{S}_2$ and setting $\overline{\rho}_1 \leftarrow 2^\beta$.

4. 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 4.1 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 4.2, we use the bounds derived in Section 3 to show that spurious critical points become increasing rare and far-away from the true solution with the addition of redundant information.

4.1. Numerical experiments

Figure 2 illustrates the success rate of quadratic recovery for the IEEE 14-bus and 39-bus systems, using error-free measurements. Each square represents
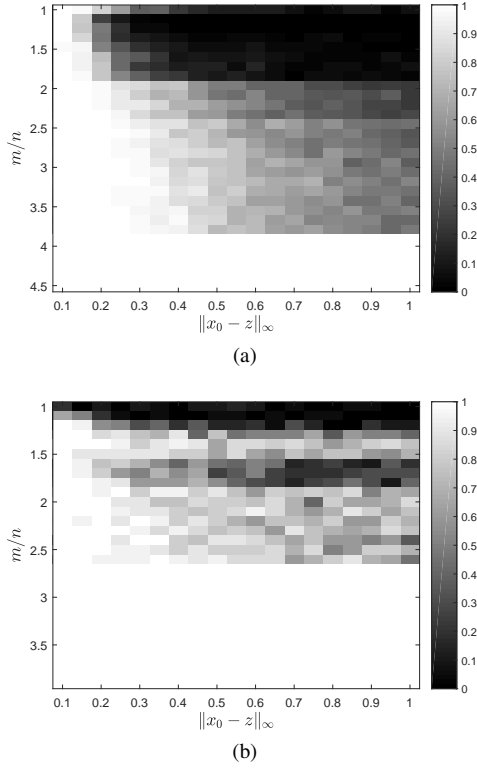
(a)

(b)

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

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 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 measure-
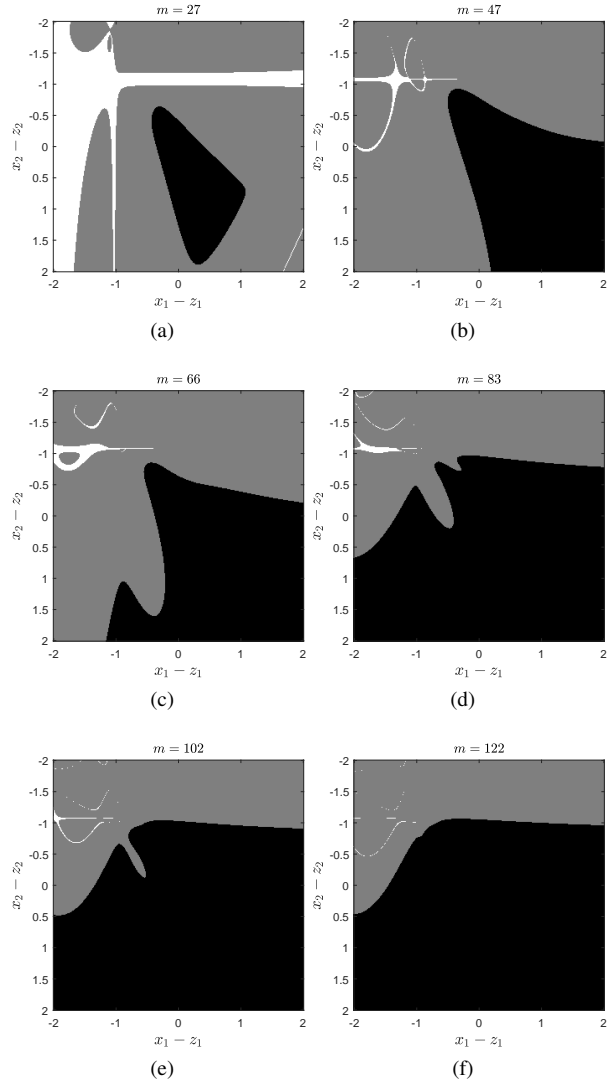

(a)   (b)

(c)   (d)

(e)   (f)

Figure 3: The regions (gray & black) $\mathcal{S}_0$ and (black) $\mathcal{S}_1$ projected onto the first two coordinates of x.
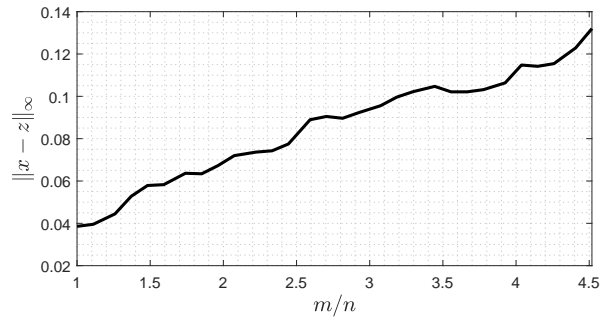

Figure 4: The bound between the solution and the closest critical point from Theorem 5.

ments defined earlier in Section 2.1.

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.

4.2. Comparison with theory

In Section 3, we had defined two regions in $\mathbb{R}^n$, denoted $\mathcal{S}_0$ (Proposition 1) and $\mathcal{S}_1$ (Corollary 2), that are guaranteed to contain no spurious critical points. Figure 3 plots a 2-D projection of these two regions for the IEEE 14-bus system, over the same sets of measurements previously considered above. More specifically, we generate a grid of test points, each identical to z except the first two components, and numerically verify whether each is contained within $\mathcal{S}_0$ and $\mathcal{S}_1$. (In the case of $\mathcal{S}_0$, the invertibility of a matrix M is verified by computing its condition number and checking if it is $< 10^6$).

As redundant measurements are added, the sets $\mathcal{S}_0$ and $\mathcal{S}_1$ grow to fill a vast portion of the entire space. Any spurious critical point must lie outside of either sets, and hence become increasingly rare and far-away from the solution. To confirm this suspicion, we evaluate the lower-bound in Theorem 5, which is itself a lower-bound on the size of the largest hypercube that can be contained within $\mathcal{S}_1$. The results, plotted in Figure 4, found the lower-bound to be increasing with the number of measurements m, with value ranging between 0.04 and 0.13. This suggests that an initial guess accurate to around 0.1 p.u. should be enough to allow Gauss–Newton to recover the true system system state,

in spite of any nonconvexity in the problem.

5. 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 [18, 19]. 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 4 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 that 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 [19]. Another direction of future work is to extend our results in Section 4 to the noisy case.

6. 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, error-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.

A. Proof of Theorem 3

Let us begin by stating a straightforward consequence of the partial differential chain rule.

Lemma 7. *Given a continuously differentiable function*

of two variables $\mathsf{r} : \mathbb{R}^2 \to \mathbb{R}$*, we define the polar param-*
eterization $x(v,\theta) = v\cos\theta$ *and* $y(v,\theta) = v\sin\theta$*, and*
define

$$\mathsf{p}(v,\theta) = \mathsf{r}(x(v,\theta), y(v,\theta)).$$

Then the partial derivatives satisfy

$$\begin{bmatrix} \partial\mathsf{p}/\partial v \\ \partial\mathsf{p}/\partial\theta \end{bmatrix} = \underbrace{\begin{bmatrix} \cos\theta & \sin\theta \\ -v\sin\theta & v\cos\theta \end{bmatrix}}_{U(v,\theta)} \begin{bmatrix} \partial\mathsf{r}/\partial x \\ \partial\mathsf{r}/\partial y \end{bmatrix} \quad (21)$$

Proof of Theorem 3. Applying (21) to (13) with $j \in \{2,\ldots,N\}$ reveals

$$\sum_{i=1}^{m} w_i[\mathsf{p}_i(v,\theta) - f_i(z)] \begin{bmatrix} \partial\mathsf{p}_i(v,\theta)/\partial v_j \\ \partial\mathsf{p}_i(v,\theta)/\partial\theta_j \end{bmatrix} =$$

$$U(v_j,\theta_j) \sum_{i=1}^{m} w_i[\mathsf{r}_i(x,y) - f_i(z)] \begin{bmatrix} \partial\mathsf{r}_i(x,y)/\partial x_j \\ \partial\mathsf{r}_i(x,y)/\partial y_j \end{bmatrix}. \quad (22)$$

In the case that $v_j \neq 0$, the matrix $U(v_j,\theta_j)$ is invertible, so the j-th instance of (13) is satisfied if and only if the corresponding instance of (9) is also satisfied. On the other hand, $v_j = 0$ occurs if and only if $x_j = y_j = 0$, and both j-th instances of (13) and (9) are zero because $\partial\mathsf{r}_i(0,0)/\partial x_j = \partial\mathsf{r}_i(0,0)/\partial y_j = 0$ holds for every i. Repeating these arguments for each j proves the desired statement. $\qquad\square$

References

[1] 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.

[2] 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.

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

[4] 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.

[5] 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.

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

[7] 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.

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

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

[10] 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.

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

[12] 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.

[13] 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.

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

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] C. W. Scherer, "LMI relaxations in robust control," *European Journal of Control*, vol. 12, no. 1, pp. 3–29, 2006.

[23] C. Scherer and S. Weiland, "Linear matrix inequalities in control," Dutch Institute for Systems and Control, Delft, The Netherlands, Tech. Rep. 2, 1999, lecture Notes.