Empirical Study of Robust State Estimation for Power Systems

Reza Mohammadi-Ghazi and Javad Lavaei

***Abstract*—This paper provides insights into the existence of spurious local optima for the nonlinear ℓ_1-norm state estimator in power system applications. The linear least-absolute-value (LAV) estimator has attracted considerable attentions over the past few years, especially in the machine learning community, mainly due to its convexity and robustness with respect to sparse noise. Due to these properties, recent studies have attempted to apply the nonlinear version of the LAV to problems such as topological error detection in power systems. However, there has been no study so far that provides theoretical guarantees for finding the global solution of the nonlinear least-absolute-value (NLAV) estimator. In this study, we analyze the performance of NLAV on different real-world scenarios and compare it with the nonlinear least-squares (NLS), which is the common practice in power industry. In our study, we consider various factors, such as the number of measurements, available prior information, and the presence of bad data, on the performance of these estimators by performing more than 260000 simulations on several IEEE benchmark systems. Additionally, we use a recent result on robust principal component analysis and low-rank matrix completion to justify our observations and provide recommendations for real-world applications of these estimators.**

***Index Terms*—Robust state estimation, power systems, convergence, RIP, matrix completion**

I. Introduction

Supplying economical energy is the lifeblood of modern civilization, and is highly related to the reliability and efficiency of energy infrastructures, such as power systems that provide clean and convenient energy for industrial and individual uses. An important challenge in operating these infrastructures is their protection against progressive failures of stressed components, which may lead to blackouts [1], [2]. To prevent such events, the power system condition should be continuously monitored so that, if needed, required actions can be taken. This condition monitoring is performed through real-time state estimation that aims to recover the underlying voltage phasors of the system, given supervisory control and data acquisition (SCADA) measurements and a model that encompasses the system topology and specifications [3], [4]. Noting that bad data, sparse noise, and topological errors significantly affect the accuracy of state estimation, addressing these problems has received considerable attentions in the past few years.

A. State estimation problem for power systems

Given an N-bus power system and its associated vector of voltage phasors, denoted by $\mathbf{v} \in \mathbb{C}^N$, the state estimation problem is to recover $\mathbf{v}$ using a set of real-valued SCADA measurements $\{b_1, \ldots, b_m\}$, where m is the number of measurements. The i-th measurement, denoted by b_i, is

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

where $f_i(\cdot)$ is a nonlinear measurement function whose details will be explained in the preceding sections, and ϵ_i is an unknown measurement noise.

Given the above settings, recovering the true state of the systems can be formulated as different optimization problems [5]. In what follows, we will briefly review major state estimation techniques for power systems.

B. Nonlinear least-square state estimator

One of the most common approaches for solving the aforementioned state estimation problem is to employ the nonlinear-least-squares (NLS) technique that was first proposed in [6], [7]. In this approach, the error parameters ϵ_i are considered to be independent and identically distributed samples of a Gaussian distribution with zero mean and variances $1/w_1, \ldots, 1/w_m$. Then, the maximum likelihood estimation of $\mathbf{v}$ turns out to be the least-square problem:

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

This simple, but nonlinear, optimization problem benefits from the property that it can be solved using basic numerical methods, such as Gauss-Newton, with a guaranteed convergence if the initial guess for the solution is of high quality [7], [8].

C. Nonlinear least-absolute-value (NLAV) state estimator

Due to its differentiability, the NLS state estimator has nice features from both theoretical and computational perspectives. However, these elegant properties are at the expense of this estimator's lack of robustness with respect to outliers and sparse noise. To elaborate on this issue, consider a state estimation problem in presence of topological errors, which is an important example of such sparse noise in the power system model and has been studied extensively [9]–[13]. Recall that the measurement functions $f_i(\cdot)$ depend on the system's specifications. The challenge of solving the problem in presence of topological errors is that some of these functions do not match the actual model of the system because the operator does not have the full knowledge of the topology of the grid (namely, the ON/OFF status of the line breakers). In other words, the ground truth measurement function $f_i(\cdot)$ is not known for some indices i and, therefore, the operator uses a possibly different function $\tilde{f}_i(\cdot)$. In this case, the system operator would solve the following optimization problem:

$$\underset{\mathbf{u} \in \mathbb{C}^N}{\text{minimize}} \ \frac{1}{2} \sum_{i=1}^{m} w_i [\tilde{f}_i(\mathbf{u}) - b_i]^2$$

The authors are with the Department of Industrial Engineering and Operation Research, University of California Berkeley, CA 94710. Emails: {mohammadi, lavaei}@berkeley.edu

This work has been supported by grants from NSF, ARO, ONR and AFOSR.

It is known that even if a small subset of the available measurement functions $\tilde{f}_i(\cdot)$ differ from the true measurement functions $f_i(\cdot)$, the state estimation error would still propagate and affect the voltage estimation of many nodes that are far away from the lines involved in the topological errors. For more information about using the NLS for power systems in presence of topological errors, the reader is referred to [13]. The lack of robustness of the NLS with respect to outliers has been well studied in other fields such as statistical learning and inference (see [14] and references therein).

One solution for dealing with the sparse noise is to deploy the synchronized point-on-wave measurements provided by phasor measurement units (PMU), which makes the state estimation problem linear and enables the use of a least-absolute-value (LAV) estimator that is robust to bad measurements [11], [12], [15]–[17]. The caveat of this approach is the reliance on PMU measurements, which are not widely available. Hence, the LAV framework is inapplicable to the highly non-convex state estimation problem based on nonlinear power flow measurements or when combined with PMU measurements.

Another approach to addressing the above challenges is to consider the nonlinear least-absolute-value (NLAV) estimator, which aims to solve the following minimization problem:

$$\underset{\mathbf{u}\in\mathbb{C}^N}{\text{minimize}} \ \sum_{i=1}^{m} |\tilde{f}_i(\mathbf{u}) - b_i|. \tag{3}$$

By the virtue of *global functions* in [18] and under the assumption that the initial guess for the solution, i.e. $\mathbf{u}^0$, is close enough to the true solution, it is shown in [13] that (3) can be solved via local search algorithms to simultaneously detect sparse topological errors and recover the underlying state of the system.

D. Solving the state estimation problem

As it will be shown in the next section, the measurements functions $f_i(\cdot)$ are quadratic due to laws of physics and hence, the objective functions of (2) and (3) are highly nonlinear and non-convex. Therefore, even in the absence of measurement noise and outliers, local search algorithms may become stuck in *spurious* local solutions, i.e. non-global local solutions that do not correspond to the true state of the system. It is essential to find global minima of these optimization problems, but this cannot always be carried out efficiently due to their NP-hardness [8], [19].

One method for solving the above-mentioned nonlinear optimization problems is to convexify their objective functions through *lifting techniques*, which relax or reformulate a nonlinear polynomial optimization into a convex semi-definite program (SDP) [20]. However, the resulting SDP problems are high-dimensional matrix optimization, which cannot be solved fast, especially in cases where (quasi) real-time state estimation is needed. Therefore, solving the problem using local search algorithms is still the common practice in the power industry (in addition to heuristic techniques that linearize the problem at the expense of making the formulation less accurate). Nevertheless, there is no strong theoretical guarantee for the convergence of local search methods to the global optima of the estimator. Thus, providing such guarantees remains an open problem with a major impact.

In a recent study on the existence of spurious local solutions for state estimation in power systems, Zhang *et al.* have shown in [8], [19] that increasing the number of (seemingly redundant) measurements improves the convexity of the problem in the sense that non-global local solutions would disappear. Having a large number of measurements enables local search algorithms to find the global solution even with a random, but plausible, initialization. However, these studies have not provided any guarantees in presence of outliers where the NLS shows lack of robustness.

Due to its robustness with respect to outliers, the NLAV has gained a lot of attention in the past few years, but the related theoretical developments are far more limited than NLS due to the non-differentiability of NLAV's objective function. Some recent studies [18], [21], [22] have provided guarantees for finding the global optima of the NLAV estimator using local search algorithms in spacial cases for low-rank matrix completion problems, but the results have not yet been extended to power system applications. A thorough review of these studies will be presented in Section III. The objective of this paper is to analyze nonlinear estimators for power systems.

E. Contributions

Due to many theoretical results and empirical studies on the superiority of the convex ℓ_1-norm estimators, it is commonly believed that the non-convex NLAV estimator would also outperform the NLS due to the robustness induced by its ℓ_1-norm term to reject outliers. This has been recently verified in [18] for tensor completion problems, which is similar to state estimation, and used in [13] for topological error detection in power systems. However, the related results highly rely on the availability of some prior knowledge about the system, e.g., the closeness of the initial guess to the true solution, which may not be always possible in practice. In this study, we perform extensive empirical analyses on IEEE benchmark systems to determine the effect of the prior knowledge, the number of measurements, and the existence of sparse noise on the performance of NLS and NLAV estimators. We also use a mathematical equivalence between the rank-1 matrix completion problem and power state estimation to extend the theoretical guarantees for finding the global solutions of the former problem to the latter, and to justify our observations. The experimental simulations also include a sensitivity analysis on the effect of deviations from the above-mentioned conditions on the likelihood of finding the global minimizer of NLAV state estimator.

This work is organized as follows. First, we provide preliminaries on the power system state estimation problem followed by the problem formulation in Section II. The conversion of this problem to a matrix completion problem along with a review of the existing theoretical guarantees for recovering the exact solution of matrix completion is described in Section III. Section IV presents the empirical analysis results that support the theoretical discussions in Section 3. Finally, we conclude the paper with a summary of our findings.

F. Notations

Throughout this paper, boldface lower (resp. upper) case letters represent column vectors (resp. matrices), and calligraphic letters stand for sets, graphs, or operators. The symbols $\mathbb{R}$ and $\mathbb{C}$ denote the sets of real and complex numbers, respectively. $\mathbb{R}^d$ and $\mathbb{C}^d$ denote the spaces of d-dimensional real and complex vectors, respectively. The symbols $(\cdot)^T$ and $(\cdot)^*$ indicate the transpose and conjugate transpose of a vector/matrix. The imaginary unit is denoted by $\mathbf{i} = \sqrt{-1}$, and $\mathrm{Re}(\cdot)$, $\mathrm{Im}(\cdot)$ denote the real part and imaginary part of a given scalar or matrix. The notations $\|\mathbf{x}\|_1$, $\|\mathbf{x}\|_2$, and $\|\mathbf{X}\|_F$ show the ℓ_1-norm and ℓ_2-norm of vector $\mathbf{x}$ respectively, and the Frobenius norm of matrix $\mathbf{X}$. The relation $\mathbf{X} \succeq 0$ means that the matrix $\mathbf{X}$ is Hermitian positive semidefinite. The trace and the (i,j) entry of $\mathbf{X}$ are respectively denoted by $\mathrm{tr}(\mathbf{X})$ and $\mathbf{X}_{i,j}$. The notation $\mathbf{X}[S_1, \mathcal{S}_2]$ denotes the submatrix of $\mathbf{X}$ whose rows and columns are chosen from the given index sets $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively. $\{\mathbf{X}_\Omega\}$ denotes the set of those entries of the matrix $\mathbf{X}$ that are in the index set Ω.

II. Preliminaries

A. SCADA measurements

Consider an electric power network represented by a graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, where $\mathcal{V} := \{1, \ldots, N\}$ and $\mathcal{E} := \{1, \ldots, L\}$ denote the sets of buses and branches, respectively. Let $v_k \in \mathbb{C}$ denote the nodal complex voltage at bus $k \in \mathcal{V}$, whose magnitude and phase are shown as $|v_k|$ and $\measuredangle v_k$. Then, the current flowing from the k-th bus to the l-th bus can be written as

$$c_{k\to l} = Y_{k,l}(v_k - v_l) = [Y_{k,l}(\mathbf{e}_i - \mathbf{e}_l)]^T \mathbf{v}, \tag{4}$$

where $Y_{k,l} \in \mathbb{C}$ is the directional admittance of the line/transformer connecting the two buses, $\{\mathbf{e}_1, \ldots, \mathbf{e}_N\}$ are the canonical bases of $\mathbb{R}^N$, and $\mathbf{v} = [v_1, \ldots, v_N]^T$. The total current injection to the k-th bus is

$$c_k = \left[Y_k \mathbf{e}_k + \sum_{l \in \mathcal{N}_k} Y_{k,l}(\mathbf{e}_k - \mathbf{e}_l) \right]^T \mathbf{v}, \tag{5}$$

where Y_k is the shunt admittance at bus k and $\mathcal{N}_k$ is the set of neighboring buses of the k-th bus. Given that complex power is the product of voltage and conjugate current, and that current is a linear function of voltage, the power can be written as a quadratic function of $\mathbf{v}$. For instance, the complex power that is sent from the k-th bus to the l-th bus is

$$p_{k\to l} + q_{k\to l}\mathbf{i} = (\mathbf{v}^* \mathbf{P}_{k\to l} \mathbf{v}) + (\mathbf{v}^* \mathbf{Q}_{k\to l} \mathbf{v})\mathbf{i} \tag{6}$$

where $\mathbf{P}_{k\to l} = \frac{1}{2}(\mathbf{S}_{k\to l} + \mathbf{S}^*_{k\to l})$ and $\mathbf{Q}_{k\to l} = \frac{1}{2\mathbf{i}}(\mathbf{S}_{k\to l} - \mathbf{S}^*_{k\to l})$ are the Hermitian splitting for

$$\mathbf{S}_{k\to l} = Y^*_{k,l}(\mathbf{e}_k - \mathbf{e}_l)\mathbf{e}_l{}^T.$$

Similarly, the net complex power that is injected into or consumed at the k-th bus is

$$p_k + q_k \mathbf{i} = c_k^* v_k = (\mathbf{v}^* \mathbf{P}_k \mathbf{v}) + (\mathbf{v}^* \mathbf{Q}_k \mathbf{v})\mathbf{i} \tag{7}$$

where p_k and q_k are the active and reactive injected/consumed powers, respectively; $\mathbf{P}_k = \frac{1}{2}(\mathbf{S}_k + \mathbf{S}_k^*)$ and $\mathbf{Q}_k = \frac{1}{2\mathbf{i}}(\mathbf{S}_k - \mathbf{S}_k^*)$ are the Hermitian splitting for

$$\mathbf{S}_k = Y_k^* \mathbf{e}_k \mathbf{e}_k{}^T + \sum_{l \in \mathcal{N}_k} Y^*_{k,l}(\mathbf{e}_k - \mathbf{e}_l)\mathbf{e}_k{}^T.$$

Based on the above equations, the power at each bus is the summation of two quadratic functions $\mathbf{v}^*\mathbf{P}_k\mathbf{v}$ and $\mathbf{v}^*\mathbf{Q}_k\mathbf{v}$, which are used in (1). The nonlinearity of these functions is the reason for the non-convexity of the state estimation problems in (2) and (3).

B. Real-valued parameterization

As it was shown in the previous section, both the power and measurement matrices are in the complex domain. For the sake of convenience, we can map all these vectors and matrices into the real domain. In doing so, we define $\mathcal{V}_r = \mathcal{V} \setminus \{\text{slack bus}\}$, i.e., the set of all buses except the slack/reference bus. We introduce the following real-valued symmetrization operators for complex voltage vectors and measurement matrices:

$$\mathrm{RS}_m(\mathbf{X}) = \begin{bmatrix} \mathrm{Re}\{X[\mathcal{V}, \mathcal{V}]\} & -\mathrm{Im}\{X[\mathcal{V}, \mathcal{V}_r]\} \\ \mathrm{Im}\{X[\mathcal{V}_r, \mathcal{V}]\} & \mathrm{Re}\{X[\mathcal{V}_r, \mathcal{V}_r]\} \end{bmatrix}, \tag{8a}$$

$$\mathrm{RS}_v(\mathbf{x}) = \begin{bmatrix} \mathrm{Re}\{\mathbf{x}[\mathcal{V}]\}^T & \mathrm{Im}\{\mathbf{x}[\mathcal{V}_r]\}^T \end{bmatrix}^T. \tag{8b}$$

Basically, this operator maps a $n \times n$ Hermitian matrix into a $(2n-1) \times (2n-1)$ real-valued symmetric matrix, and a complex vector of length n into a real-valued vector of length $2n-1$ (by assuming that the slack bus serves as a reference for phases with its phase equal to zero). Considering these operators, we define:

$$\mathbf{M}_i \triangleq \mathrm{RS}_m(\widetilde{\mathbf{M}}_i), \tag{9a}$$

$$\mathbf{z} \triangleq \mathrm{RS}_v(\mathbf{v}) \tag{9b}$$

$$\mathbf{x} \triangleq \mathrm{RS}_v(\mathbf{u}) \tag{9c}$$

where $i \in \{1, \ldots, m\}$ is the index for the SCADA measurement, and $\widetilde{\mathbf{M}}_i$ can be any of the following measurement matrices: (i) $\mathbf{P}_k$ or $\mathbf{Q}_k$ for nodal measurements at bus k, (ii) $\mathbf{P}_{k\to l}$ or $\mathbf{Q}_{k\to l}$ for line measurements between buses k and $l \in \mathcal{N}_k$, (iii) $\mathbf{e}_k\mathbf{e}_k^T$ corresponding to the squared voltage magnitude at bus k. Note that $\mathbf{v}$ and $\mathbf{u}$ are the same as those defined in (1), (2), and (3).

Using the definitions in (9) along with the SCADA measurement equations (6) and (7), it is straightforward to observe that each classical measurement function can be written as

$$f_i(\mathbf{u}) = \mathbf{x}^T \mathbf{M}_i \mathbf{x} \tag{10}$$

and hence, the norm minimization problems (2) and (3) correspond to

$$\underset{\mathbf{x} \in \mathbb{R}^{2N-1}}{\text{minimize}} \ \sum_{i=1}^{m} \left(\mathbf{x}^T \mathbf{M}_i \mathbf{x} - b_i \right)^2, \tag{11a}$$

$$\underset{\mathbf{x} \in \mathbb{R}^{2N-1}}{\text{minimize}} \ \sum_{i=1}^{m} |\mathbf{x}^T \mathbf{M}_i \mathbf{x} - b_i|. \tag{11b}$$

Note that we ignore the PMU measurements in this study, but they can be easily incorporated in the above estimators using convex functions (namely, square or absolute value of linear terms).

III. Reformulation as matrix sensing

This section explains the reformulation of power systems' state estimation as a matrix completion problem, which helps us derive some theoretical guarantees to justify the empirical results we will present in Section IV.

A. Power system state estimation as a matrix completion

The task of filling the missing entries of a partially observed matrix is called matrix sensing, that is ubiquitous in machine learning, signal processing, and computer vision. This problem has been well studied in the past few years, especially for recovering low-rank matrices [18], [22]–[26]. The computational task in this problem is to recover a (usually low-rank) matrix $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}$ from partial linear measurements of the form

$$\mathbf{y} = \mathcal{A}(\mathbf{X}) \tag{12}$$

where $\mathcal{A}(\cdot)$ is a linear operator [27]. It is easy to show that the state estimation problem can be cast into this framework. To show this equivalence, notice that

$$f_i(\mathbf{u}) = \mathbf{x}^T \mathbf{M}_i \mathbf{x} = \text{tr}(\mathbf{M}_i \mathbf{x}\mathbf{x}^T) = \text{tr}(\mathbf{M}_i \mathbf{W}) \tag{13}$$

where $\mathbf{W} = \mathbf{x}\mathbf{x}^T$ is a rank-1 symmetric matrix. The measurement matrices $\mathbf{M}_i$ are highly sparse in power systems and $f_i(\mathbf{u})$ is a linear combination of the entries of $\mathbf{W}$. Thus, once $\mathbf{W}$ is recovered via matrix sensing techniques, it can be factorized to obtain $\mathbf{x}$.

A special form of the matrix sensing problem is called matrix completion where the measurement matrices are in the form of $\mathbf{e}_i \mathbf{e}_j^T$. This has the elegance of simplicity in terms of the linear operator, and the power system state estimation may be mapped into a matrix completion after solving a system of linear equations. Due to the above-mentioned equivalence, a brief review of the relevant results on the matrix completion is provided in the next section.

B. Matrix completion: Related work

An extensive amount of studies on the low-rank matrix recovery from partially observed entries rely on convex relaxations that achieve strong theoretical guarantees, but they suffer from extensive computational costs which make these techniques impractical for large problems (see [28], [29] and the references therein). Therefore, there has been growing interest in using the nonlinear optimization techniques for this problem in recent years. To introduce the non-convex matrix recovery problem, assume that $\mathbf{X} \in \mathbb{R}^{d \times d}$ is the target matrix, $\text{rank}(\mathbf{X}) = 1$, and $\Omega = \{(i,j) : \mathbf{X}_{i,j} \text{ is observed } \}$ is the set of indices for the observed entries of $\mathbf{X}$. The assumption of $\mathbf{X}$ being a rank-1 squared matrix is solely for simplicity and its similarity to the state estimation problem. Given the above information, most non-convex recovery methods apply nonlinear optimization algorithms to solve

$$\underset{\mathbf{y} \in \mathbb{R}^d}{\text{minimize}} \ \| \mathcal{P}_\Omega \left(\mathbf{X} - \mathbf{y}\mathbf{y}^T \right) \|_F^2, \tag{14}$$

where $\mathcal{P}_\Omega(\mathbf{X})$ is a projection onto the set of matrices with their supports in the measurement set Ω. This projection returns $\mathbf{X}_{i,j}$ if $(i,j) \in \Omega$, and 0 otherwise [26].

In a seminal work by Ge *et al.* [26], it has been shown that if a sufficiently large number of *random* observations are available, the objective function of (14) (after adding some regularization) does not have any spurious local minima, and therefore, all local minima are global. Thus, local search algorithms can efficiently find the global solution of the problem without requiring prior knowledge about the solution and concerns about initialization.

Although the objective function of (14) enjoys strong theoretical guarantees for convergence to global solution due to smoothness and differentiability, it is not robust with respect to outliers. Thus, methods like stochastic gradient descent (SGD) would fail to recover $\mathbf{X}$ in presence of large but sparse noise. Therefore, another line of research has recently focused on studying robust low-rank matrix recovery, which is to solve the following problem:

$$\underset{\mathbf{y} \in \mathbb{R}^d}{\text{minimize}} \ \| \mathcal{P}_\Omega \left(\mathbf{X} - \mathbf{y}\mathbf{y}^T \right) \|_1, \tag{15}$$

To analyze the above objective function, Josz *et al.* defined the notion of *global functions*, which are functions that may have multiple local minima, but all of them are global [18]. A simple and yet nontrivial such function is $f(x) = |x^2 - 1|$ that is shown in Figure 1. This function has two local minima at $x = \pm 1$, both of which are global. No matter how a local search solver is initialized, it can find a global solution efficiently. Then, in their study, they have proved that the objective function of the robust matrix completion (15) is a (weakly) global function if the number of measurements is sufficiently large and the projection operator is lossless in the sense that $\mathcal{P}_\Omega(\mathbf{X}) = \mathbf{X}$.

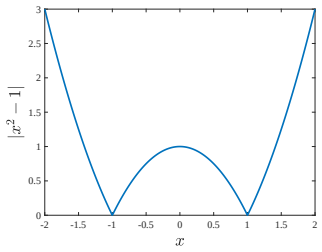


Figure 1: Simple example of a global function $|x^2 - 1|$

Li *et al.* in [22] have analyzed the geometry of the rank-r matrix completion problem using an approach similar to [26], and shown that the robust rank-r matrix completion does not have spurious local solution under the following conditions: (1) $\mathbf{X}$ is symmetric positive semidefinite, (2) the projection operator is a *Gaussian measurement operator*, (3) the number of observations is larger than a specified lower bound, (4) the fraction of outliers is less than a half (see Theorem 1 in [22]).

Despite the connection between the matrix completion formulation and power system state estimation described before, none of the above-mentioned results for the matrix completion problem is directly applicable to the robust state estimation for power system, because the measurement matrices are deterministic and highly sparse in power systems (i.e., far from being random). Moreover, nonlinear power measurements do not directly measure the entries of $\mathbf{W}$, except the diagonal

ones. Therefore, the guarantees provided in [18] are also not applicable to power system state estimation. To the best of our knowledge, the only study that has addressed deterministic projection operators for ℓ_1-norm estimators is the recent work by Fattahi and Sojoudi [21], where the authors proved that under some technical conditions, (15) does not have spurious local solutions and all local solutions are global. More specifically, consider the optimization problem

$$\underset{\mathbf{y}\in\mathbb{R}^d_+}{\text{minimize}} \ \|\mathcal{P}_\Omega\left(\mathbf{X}-\mathbf{y}\mathbf{y}^T\right)\|_1, \tag{16}$$

where $\mathbf{X}=\mathbf{y}_*\mathbf{y}_*{}^T+\mathbf{S}$ is the partial noisy observation of $\mathbf{y}_*\mathbf{y}_*{}^T$ that is the rank-1 ground truth with $\mathbf{y}_*$ to be the true solution, $\mathbf{S}$ the additive noise matrix, and $\mathbb{R}_+$ is the set of positive real numbers. Define the set of *bad* and *good* measurements as $B=\{(i,j):(i,j)\in\Omega \text{ and } \mathbf{S}_{i,j}\neq 0\}$ and $G=\{(i,j):(i,j)\in\Omega \text{ and } \mathbf{S}_{i,j}=0\}$, respectively. Also, define the *sparsity graph* induced by a set Ω, denoted by $\mathcal{G}(\Omega)$, as a graph with the vertex set $V=\{1,\dots,d\}$ that includes the edge (i,j) if $(i,j)\in\Omega$. Then, the following theorem is given in [21].

Theorem 1. *Assuming $\mathbf{y}_*>0$ and under appropriate regularization,* (16) *has no spurious local solution and has a unique local minimum, provided that $\mathcal{G}(G)$ has no bipartite component and*

$$\kappa(\mathbf{y}_*)^4\Delta(\mathcal{G}(B))=\mathcal{O}\left(\delta(\mathcal{G}(G))\right) \tag{17}$$

where

- $\kappa(\mathbf{y}_*)$ is the ratio of the largest to the smallest entry of $\mathbf{y}_*$.
- $\mathcal{G}(G)$ and $\mathcal{G}(B)$ are the sparsity graphs induced by the sets G and B (i.e., the good and bad measurements), respectively.
- $\Delta(\mathcal{G})$ and $\delta(\mathcal{G})$ denote the maximum and minimum degrees of the graph $\mathcal{G}$.

It is easy to observe the connection between the state estimation problem and (16) by substituting $\mathbf{X}$ with $\mathbf{W}$, except that the main assumption of Theorem 1 —the strict positivity of $\mathbf{x}$— is not satisfied in power systems due to the presence of complex conjugate of complex voltage vectors in (6) and (7); unless the voltage angles are all zero. Fortunately, these angles are usually small in practice, and in this study, we aim to empirically determine the likelihood of failure in finding the global solution of state estimation for various levels of violation from the main assumption of the above theorem. Note that the projection $\mathcal{P}_\Omega$ is slightly different in power systems; however, this is not an issue since the power measurements are linear combination of different entries of $\mathbf{W}$ as it was shown in (6) and (7); thus, the associated entries of $\mathbf{W}$ with the available measurements can, in theory, be deterministically determined/estimated by solving a system of linear equations. We will elaborate more on this in the next section.

IV. Empirical analysis

In this section, we develop empirical approaches to investigate the existence of spurious local solutions for nonlinear estimators. To this end, three case studies are conducted. The first study compares the performance of the NLS and NLAV estimators on different IEEE benchmark power systems. The second study quantifies the likelihood of finding the correct state by solving the NLAV via local search in presence of large, but sparse noise. Finally, the third study aims to propose necessary conditions for preventing the NLAV estimator to become stuck in spurious local solutions. We also perform a sensitivity analysis to quantify the effect of deviations from those conditions on the likelihood of recovering the true state via NLAV, and to assess the reliability of employing this estimator in real-world scenarios.

A. Comparison of the NLAV and NLS

1) Noiseless measurements: The first study aims to compare the performances of the NLAV and NLS state estimators in a noiseless scenario and to quantify the likelihood associated with each of these optimization problems to have a spurious local solution. The reason for disregarding the noise in this section is that the existence of spurious solutions is due to the nonlinearity of the problem rather than the measurement errors.

We conduct this empirical analysis using randomization, which is to perform an extensive number of simulations such that each one attempts to recover the system's state using both NLS and NLAV with random initializations and various numbers of measurements. By performing a large number of such simulations, we aim to achieve three main objectives. The first objective is to determine, through a Monte Carlo type of analysis, whether or not each of the above estimators has spurious local solutions. Secondly, we intend to assess the effect of prior knowledge, e.g. closeness of initial guess to the true solution, on recovering the true state. Finally, the third objective is to quantify the likelihood of true state recovery using these estimators.

This empirical study is designed as follows:

1) For an N-bus power system, we consider 20 different levels of measurement redundancy in the interval of $[3N, 3N+4L]$, where L is the number of branches. Note that we consider all nodal measurements, i.e. voltage magnitude and nodal power injections, as the minimum number of measurements being used. Then, we increase the measurement redundancy by randomly selecting the line measurements and adding them to the set of measurements.
2) We consider the distance between the initial guess and the global solution, i.e. $\|\mathbf{x}_0-\mathbf{z}\|_\infty$ with $\mathbf{x}_0$ to be the initial guess for $\mathbf{z}$, as the second parameter that may affect the performance of the estimators. We change this distance from 0.1 to 1 with 0.1 steps in our simulations, i.e. $\|\mathbf{x}_0-\mathbf{z}\|_\infty\in\{0.1,0.2,\dots,1\}$. Note that this distance represents the prior information about the global solution.
3) For each combination of the redundant measurement and the $\|\mathbf{x}_0-\mathbf{z}\|_\infty$, we perform 50 simulations. In each simulation, the magnitude and angle of each entry of the true solution $\mathbf{v}$ are randomly sampled from the intervals $[0.9,1.1]$ and $[-30^\circ,30^\circ]$, respectively; then,

$\mathbf{z}$ is obtained using (9). The initial guess $\mathbf{x}_0$ for each simulation is also randomly generated and normalized so that its distance to $\mathbf{z}$ matches the desired $\|\mathbf{x}_0 - \mathbf{z}\|_\infty$. Therefore, for each simulation, the direction along which we start approaching the global solution is totally random and the initial distance $\|\mathbf{x}_0 - \mathbf{z}\|_\infty$ is set ot be consistent with that scenario.

For the NLS estimator, we solve the optimization problem in (11a) using MATLAB's *lsqnonlin* function along with the Levenberg-Marquardt algorithm. However, this technique cannot be used for the NLAV estimator that is to solve (11b). For numerically solving the non-smooth ℓ_1-norm minimization, we use the famous technique discussed in [30] for convex optimization to reformulate the problem (11b) as a smooth non-convex quadratically-constrained quadratic program (QCQP). The reformulated problem is

$$\begin{aligned} \underset{\mathbf{t}\in\mathbb{R}^m}{\text{minimize}} \quad & \sum_{i=1}^{m} t_i \\ \text{subject to} \quad & \mathbf{x}^T\mathbf{M}_i\mathbf{x} - b_i \le t_i, \qquad i = 1, ..., m \\ & \mathbf{x}^T\mathbf{M}_i\mathbf{x} - b_i \ge -t_i, \qquad i = 1, ..., m \end{aligned} \tag{18}$$

Nonlinear problems of this form can be efficiently solved via different numerical algorithms, and some recent studies have provided theoretical guarantees for the convergence of such methods to a stationary point [31]. To obtain such a point, we use MATLAB's *fmincon* function along with the interior point method.

The above-mentioned procedure is based on 10000 simulations for each system, which are performed on the Amazon EC2 cluster with 60 CPUs and 225 gigabytes of RAM. The results of the simulations on the IEEE 14, 39, 57, 300, and 1354-bus systems are shown in Figure 2. m and $\overline{n}$ in these plots are, respectively, the number of measurements used and the degree of freedom, i.e. $m \in [3N, 3N + 4L]$ and $\overline{n} = 2N-1$. Note that by using the real-valued symmetrization operator (8a), the real and imaginary parts of the voltage at each bus is considered as an unknown, which gives $2N$ degrees of freedom in total. However, in order to have a unique solution, a reference needs to be defined for phase angles. This is accomplished by fixing the voltage angle at one of the buses (slack bus) and hence, the effective degree of freedom becomes $2N - 1$. Figures 2(a) to 2(e) illustrate the performance of the NLAV estimator for the above-mentioned IEEE systems, and Figures 2(f) to 2(j) are associated with the NLS. The average success rate of finding the global solution for all combinations of $m/\overline{n}$ and $\|\mathbf{x}_0 - \mathbf{z}\|_\infty$ is color coded in these plots. For instance, the success rate of the NLAV in finding the global solution for the 39-bus system with the minimum number of measurements, i.e. $m/\overline{n} = 1.5$, and $\|\mathbf{x}_0 - \mathbf{z}\|_\infty = 1$ is 0% and is shown with the black color at the top right corner block of Figure 2(b), while this success rate is 100% when full measurement set comprised of all voltage magnitude measurements and all line flow measurements is used, i.e. $m/\overline{n} = 3.9$, and $\|\mathbf{x}_0 - \mathbf{z}\|_\infty = 0.1$. This scenario corresponds to the white block at the bottom left corner of the same plot in Figure 2(b).

The results confirm that the success rate of recovering the true state is improved when more measurements are used and the initial guess is closer to the actual state. An important observation is that the NLS significantly outperforms the NLAV in these simulations by providing a considerably higher success rate when a large number of redundant measurements is used; however, the NLAV performs better on average for a small number of measurements. Also, there is always a specific level of measurement redundancy $m/\overline{n}$ after which the NLS can perfectly recover the state of the system. This threshold depends on the size and geometry of the system, e.g., it is around 2.7 for the 14-bus system (Figure 2(f)), which corresponds to using about one third of all possible line measurements, while we need $m/\overline{n} \simeq 4.4$ for both the 300 and 1354-bus system to obtain perfect recovery by using the NLS (Figure 2(i) and Figure 2(j)). Unlike the NLS, such a threshold does not exist for the NLAV estimator. Moreover, there are scenarios with full measurement set and relatively close initial guess to the global solution that the NLAV is unable to recover their true states. To verify that the undesirable recovered solutions are truly spurious minima and are not due to numerical instability, we perturb those solutions and use them as initial guesses, and re-solve the problem. We have repeated this for several cases and observed that all of them converged to the same local solutions. We have also checked the value of the objective function at 1e7 number of points that are randomly selected within a sphere with 0.1 radius around the recovered solution to ensure that it has the smallest value.

Notice that in Figures 2(e) and 2(j), the number $m/\overline{n}$ starts from 2.8 for the 1354-bus system. Due to the tremendous computational burden of performing all simulations for this system (more than 42 hours on Amazon cluster for each of these figures), we perform the analyses for $m/\overline{n} \ge 2.8$ that is comprised of 5500 simulations. Additionally, we use the quasi-Newton Hessian approximation for the interior point method and restrict the number of solver iterations to 3000 for consistency to real-world scenarios. The outcome of the analysis on the 1354-bus system is similar to the other systems, except that the likelihood of recovering the true states is lower compared to the smaller systems if the initial guess is far from the global solution. This could be due to the limitation enforced on the maximum number of iterations. Due to the similarity of the results for the 1354 and 300-bus systems, we will focus on the 14, 39, 57, and 300-bus systems in the remainder of this study.

To further investigate the properties of the spurious local solutions of the NLAV estimator, we visualize a few cases with a large number of measurements, i.e. $m/\overline{n} > 3.2$ and unsuccessful state recovery via the NLAV. The blocks associated with these measurements are shown with arrows in Figures 2(b) and 2(b). The positions of the above-mentioned cases on these plots show their associated $m/\overline{n}$ and $\|\mathbf{x}_0 - \mathbf{z}\|_\infty$. Note that there could be more than one unsuccessful case within one block, but we randomly select only one such case. Figures 3 and 4 visualize these cases for the 39 and 57-bus systems, respectively. The buses with inaccurate state recovery are shown with red color. The black-colored buses are those

for which the complex voltage is accurately recovered. An interesting observation about the spurious local solution in all cases with a large number of measurements is that the buses with an inaccurate state recovery are connected. Also, the inaccurate recovery usually occurs only at a small number of buses; however, there are a few cases where the NLAV fails to estimate the complex voltage at all buses.

2) Noisy measurements: The empirical analyses of the previous section show that the NLS outperforms the NLAV in the absence of (large) noise; however, the main purpose of using the NLAV is its robustness with respect to outliers as it was previously discussed. To compare the performance of these estimators in presence of bad data, we run some analyses similarly to the ones presented in previous section on the 57-bus system, but this time we corrupt 9 randomly selected measurements (equivalently 2% of the total number of possible measurements) with large noise. Figure 5 shows the result of this analysis. Comparing Figures 2(c) and 5(a) shows that the average success rate of the NLAV is reduced in presence of sparse noise and bad data, but this estimator can still recover the true states with overwhelming probability if the initial guess is relatively close to the true solution and the number of measurements is relatively large. However, the NLS estimator completely fails to recover the true solution in all cases. This shows the robustness of the NLAV with respect to outliers despite its lower performance compared to the NLS in noiseless scenario.

B. Effect of voltage angles in recovering the states in presence of bad data

The empirical analysis of this section aims to simultaneously study the following: (i) the effect of bad date in the likelihood of recovering the true state, (ii) the conditions that needs to be satisfied for perfect state recovery via NLAV.

For an N-bus power system we design this analysis as follows:

1) We consider 20 levels of measurement redundancy in the interval of $[3N, 3N + 4L]$ similar to what explained in Section IV-A1.
2) The number of bad data is the second parameter that we vary in this analysis. For a given level of measurement redundancy, i.e. $m/\overline{n}$, we increase the percentage of bad data from 0% to 10% of the total number of possible measurements with 1% step. Because the voltage magnitude sensors are highly accurate and well protected from attacks in practice, we assume that voltage magnitudes are exact and the bad data can affect the other measurements.
3) For each combination of measurement redundancy and bad data percentage we run 50 simulations, which leads to 11000 simulations in total for each system.
4) he ground truth voltage for each case is generated by randomly sampling the voltage magnitudes and angles from $[0.9, 1.1]$ and $[-\theta, \theta]$, respectively, where θ is the maximum voltage angle. A complete set of 11000 simulations are performed for $\theta \in \{0^\circ, 1^\circ, 3^\circ, 5^\circ, 7^\circ, 10^\circ\}$. The initial guess $\mathbf{x}_0$ is completely random with all entries in the interval [0,1].
5) In each simulation we solve (18).

Notice that for $\theta \neq 0$ the strict positivity of $\mathbf{z}$ is not satisfied due to the complex conjugate of voltage vectors in (6) and (7). Therefore, the entries of $\mathbf{W}$, which was defined in (13), are not strictly positive if we look at this problem from the matrix completion perspective. The existence of spurious local solution when $\mathbf{z}$ is not strictly positive has been proven in [21] by proposing a counterexample in which one diagonal element of matrix $\mathbf{X}$ is zero (see (16) and Theorem 1). Due to the assumption of exactness of voltage magnitudes, the diaginal entries of $\mathbf{W}$ are always non-zero in our problem. Therefore, we run this analysis for different values of θ to study the effect of deviations from the main condition in Theorem 1 on the average success rate in practice.

The results of the simulations on the IEEE 57-bus system are shown in Figure 6. The success rate of the NLAV in recovering the true solution is color coded in these plots. Similar to Figure 2, the y-axis is $m/\overline{n}$ where $\overline{n} = 2N - 1$, but the x-axis shows the number of bad data. It follows from these figures that the general pattern does not change for different values of θ, and the tiny differences between the plots are most likely related to different initializations. We do not show the results associated with $\theta = 0^\circ$ in this figure given the voltage magnitude measurements, the real parts of the states are trivially known. The results in this figure confirm that if the entries of $\mathbf{z}$ are not strictly positive, no matter how large their magnitudes are, the problem may have spurious local solution. However, the likelihood of finding the true solution is still significantly high (usually more than 90% except in a few cases with roughly 80% success rate) if the number of redundant measurements is large and number of bad data is less than 10%.

C. Analyzing the existence of spurious local solutions of NLAV via matrix completion

Based on previous sections, the state estimation can be reformulated as a matrix sensing problem that has been well studied under various conditions. In this section, we design our third empirical analysis within the framework of robust matrix completion to analyze the existence of spurious local solutions for the NLAV estimator and the possibility of avoiding such undesirable solutions in practice. We only focus on the NLAV and robust matrix completion due to the lack of robustness of the NLS in presence of bad data.

In the previous section, we observed that the non-positivity of the entries of $\mathbf{z}$, which corresponds to non-zero voltage angles, is one of the main reasons for the existence of spurious solutions. Basically, non-zero voltage angles is equivalent to violating the main assumptions of Theorem 1. This violation cannot be avoided in practice for AC power state estimation with power measurements, because the relative angles are non-zero. However, these angles are often small and hence, approximating each voltage magnitude with its real part may be acceptable. The objective of this section is to use robust matrix completion in order to recover the real parts of voltages when their angles are small. In doing so, we consider the exact same scenarios that were discussed in Section IV-B, but instead

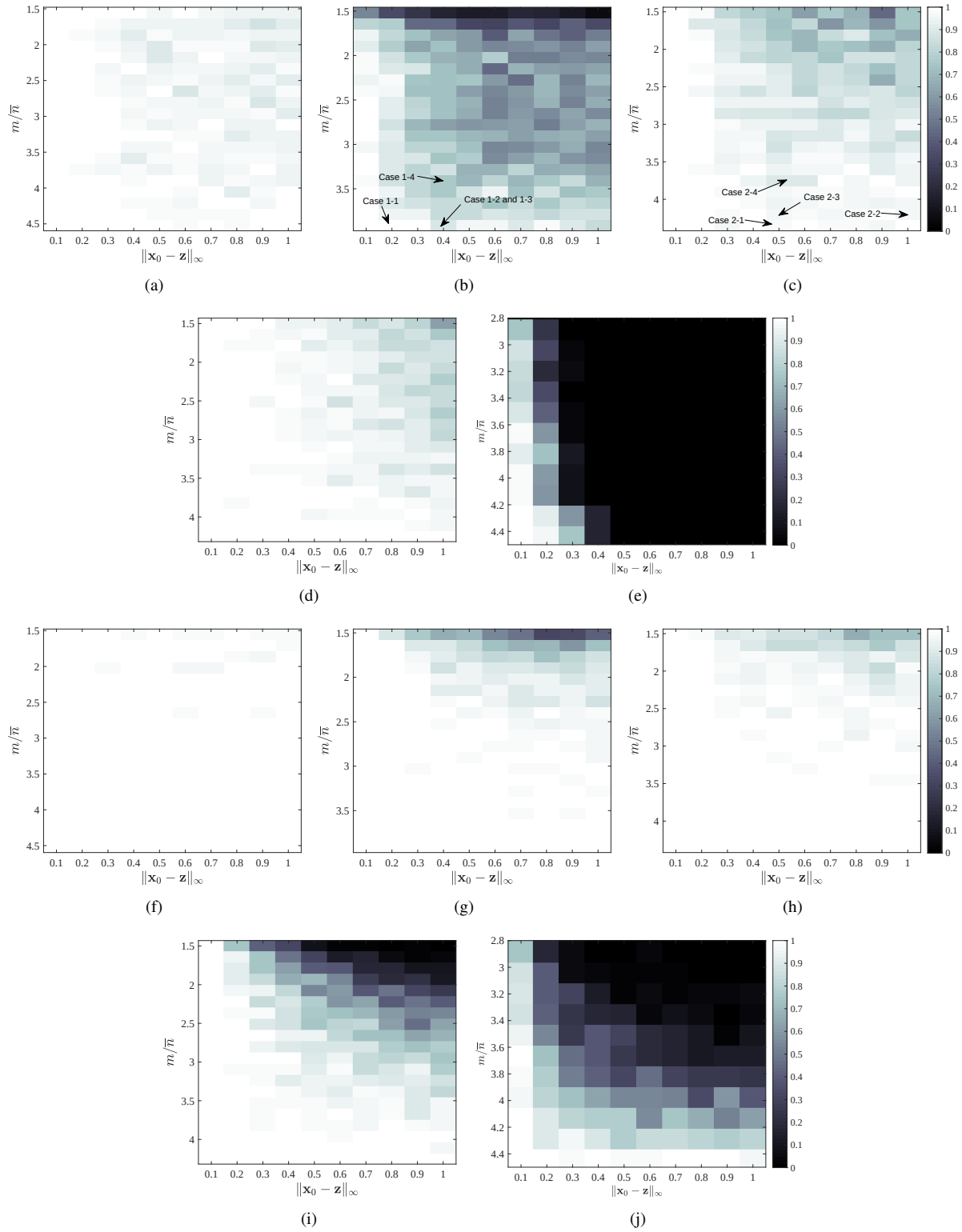
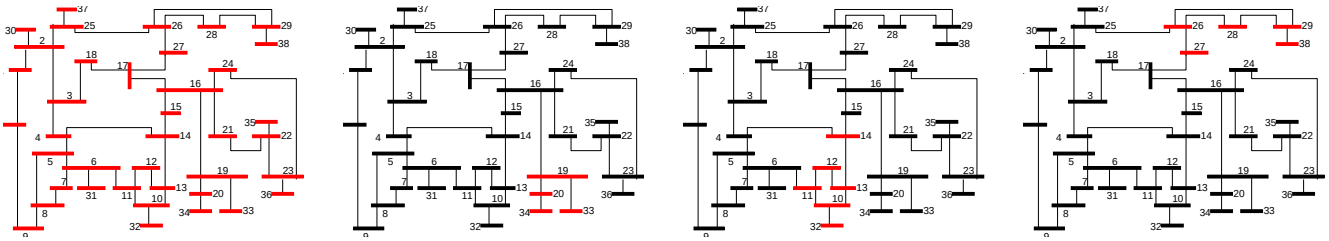


Figure 2: Comparing the performance of the NLAV and NLS in noiseless scenario. The results of the NLAV are shown in (a), (b), (c), (d), and (e), which are respectively for the IEEE 14, 39, 57, 300, and 1354-bus systems. The results associated with the NLS estimator are shown in (f) to (j) respectively for the same aforementioned systems.

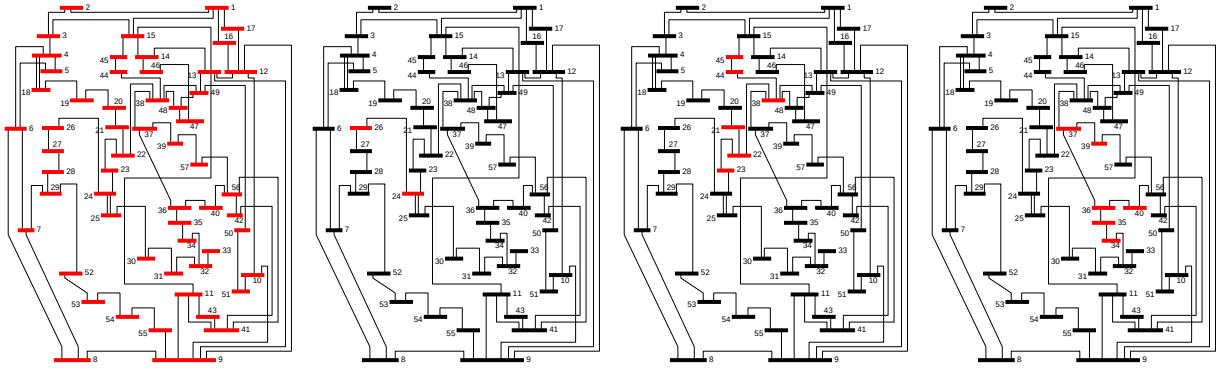
of solving (18), we solve the following optimization problem in this section:

$$\underset{\widetilde{\mathbf{x}} \in \mathbb{R}^d_+}{\text{minimize}} \ \| \mathcal{P}_\Omega \left(\widetilde{\mathbf{W}} - \widetilde{\mathbf{x}} \widetilde{\mathbf{x}}^T \right) \|_1, \tag{19}$$



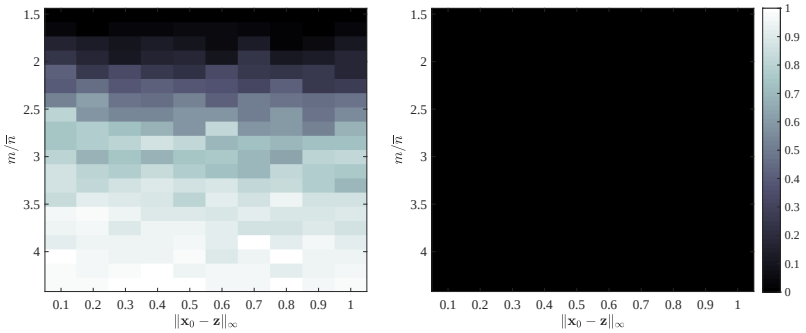
(a) (b) (c) (d)

Figure 3: Voltage mismatches for the cases shown in Figure 2(b). The red colored buses are those where the underlying states are not accurately estimated via the NLAV. (a) Case 1-1, (b) Case 1-2, (c) Case 1-3, (d) Case 1-4.



(a) (b) (c) (d)

Figure 4: Voltage mismatches for the cases shown in Figure 2(c). The red colored buses are those where the underlying states are not accurately estimated via the NLAV. (a) Case 2-1, (b) Case 2-2, (c) Case 2-3, (d) Case 2-4.



(a) (b)

Figure 5: Comparing the performance of the NLAV and NLS in recovering the true state of the 57-bus system when 2% of the measurements are corrupted with large, sparse noise. The plots show the average success rate of recovering the states for (a) NLAV, (b) NLS.

where $\widetilde{\mathbf{x}} = \mathrm{Re}(\mathbf{x})$ and $\widetilde{\mathbf{W}} = \mathrm{Re}(\mathbf{z})\mathrm{Re}(\mathbf{z}^T)$. Basically, we first ignore the imaginary part of $\mathbf{x}$ and then estimate those entries of $\widetilde{\mathbf{W}}$ that can be observed from the given set of power measurements via solving a linear system of equations. Note that this is different from DC approximation, because the real part of the voltage vector is not unity and ignoring the imaginary part acts as adding a noise to the estimation of the observed entries of $\mathbf{W}$. Then, we solve (19) to recover the real part of the voltage vector. Note that corrupted power measurements affect both diagonal and off-diagonal entries of $\{\widetilde{\mathbf{W}}_\Omega\}$. To satisfy the assumption we made about the exactness of voltage magnitude measurements, which corresponds to the exactness of the diagonal entries of $\widetilde{\mathbf{W}}$, we corrupt the off-diagonal entries of $\{\widetilde{\mathbf{W}}_\Omega\}$ that are associated with the bad noisy measurements.

Recall that the second condition in Theorem 1 for perfect recovery is that the sparsity graph of good data should not be bipartite. We also assumed that the diagonal entries of $\widetilde{\mathbf{W}}$ are

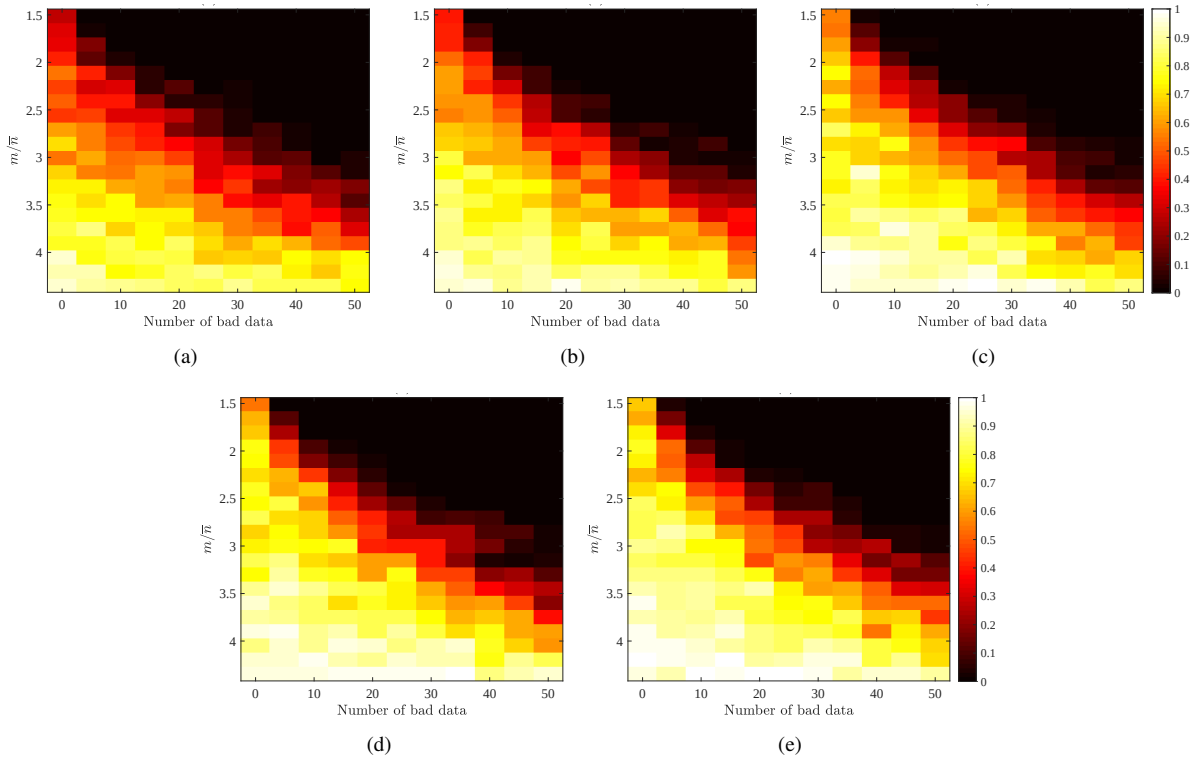


Figure 6: Results of the NLAV state estimator for different values of the maximum voltage angle. (a) $\theta = 1°$, (b) $\theta = 3°$, (c) $\theta = 5°$, (d) $\theta = 7°$, (e) $\theta = 10°$.

known exactly. These entries correspond to self loops in the sparsity graph of good measurements from $\widetilde{\mathbf{W}}$. Knowing that a graph is not bipartite if all nodes have a self loop, the above condition will be automatically satisfied.

Figure 7 shows the results of this set of analyses on the IEEE 57-bus system for different values of θ. m is the same as what defined before, but $n = N$ in these plots, because only the real part of the voltage vector is being recovered and hence, the degrees of freedom equals the number of buses. The success rate of the 50 simulations for each combination of m/n and various number of bad data is color coded in these plots. Note that the voltage vector cannot be exactly recovered when the angles are nonzero. In these simulations, we consider a successful recovery as the one that the voltage recovery error is less than or equal to $1\mathrm{e}{-6}$. It follows from Figure 7 that increasing the number of bad data decreases the likelihood of finding the global solution of (19). Furthermore, we have perfect recovery for $\theta \leq 3°$ when there is no bad data. Even in presence of a moderate amount of bad data, there is a high likelihood of accurately recovering the real parts of the complex voltages when the number of redundant measurements is sufficiently large. However, for $\theta \geq 5°$, the success rate drops to less than 60% even with full measurements and no bad data. Note that we do not show the results associate with $\theta = 0°$ in this figure because, given the voltage magnitude measurements, the real parts of the states are trivially known in that case.

D. Larger system

To study a larger system, we perform similar analyses on the IEEE 300-bus system, but only for $\theta = 3°$. Figure 8 illustrate the results that are similar to the 57-bus system and confirms that the NLAV state estimator may become stuck at spurious local solutions when it is applied to the complex state recovery problem. The existence of such spurious local solution is not related to the location of the non-positive entries of $\mathbf{W}$ nor the amount of redundant measurements. However, increasing the number of redundant measurements improves the likelihood of recovering the true solution.

V. Summary and Discussion

The ease of numerically solving the NLS problem with smooth and differentiable objective function as well as the existence of some theoretical guarantees for the high performance of this technique in noiseless scenario have made the NLS as the common practice in power industry. On the other hand, the NLAV is not widely used in this application due to the lack of theoretical guarantees for finding the global minimizer of its objective function. In the previous section, we presented the results of a comprehensive empirical analysis on the performance of the NLAV state estimator and its comparison with the NLS estimator in the power systems application. We have also shown, by working on the theoretical results of the robust matrix completion problems, that the NLAV suffers from the existence of spurious local solutions when applied to power system state estimation.

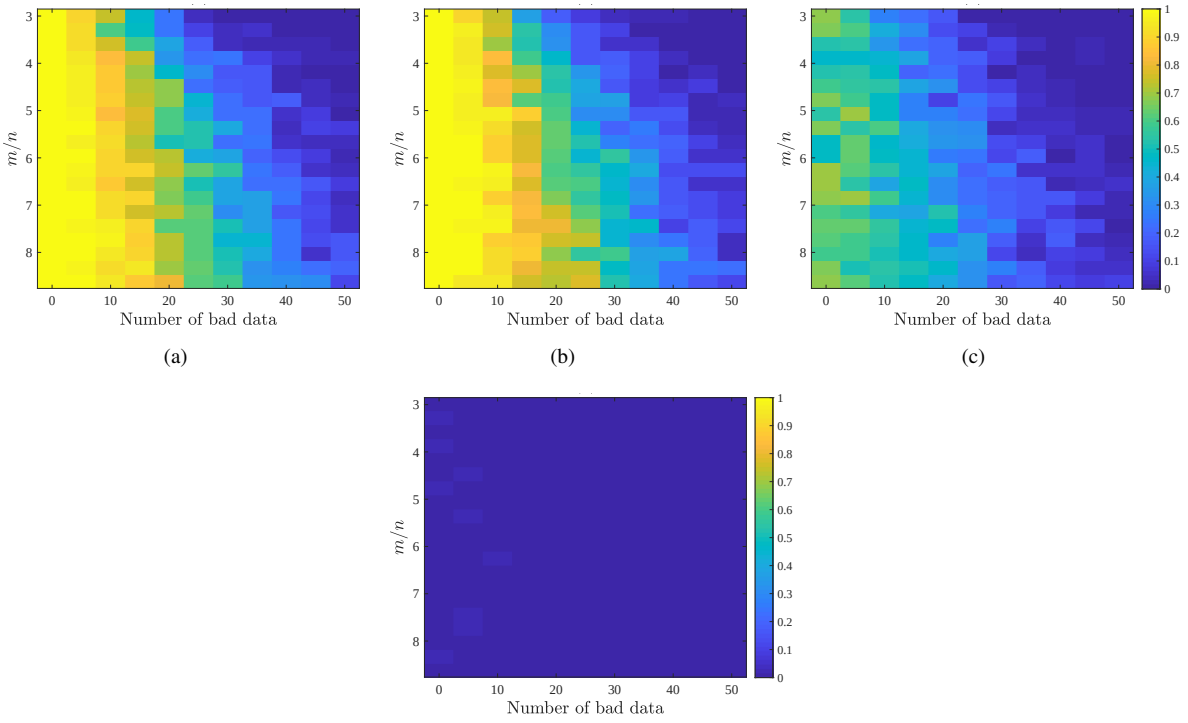


Figure 7: Results of using the matrix completion approach to recover voltage magnitudes for different values of voltage angle. (a) $\theta = 1°$, (b) $\theta = 3°$, (c) $\theta = 5°$, (d) $\theta = 7°$.

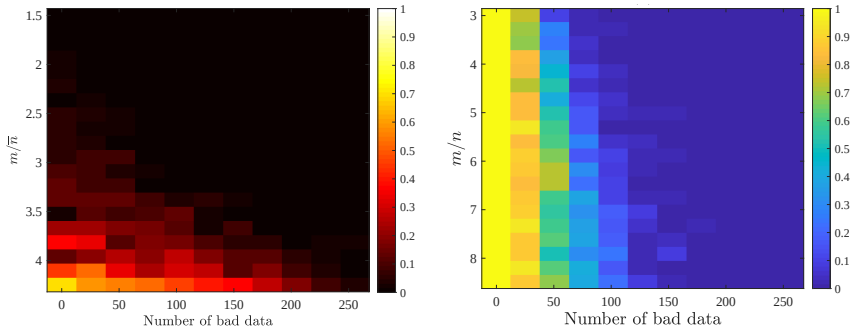


Figure 8: Empirical analyses on the 300-bus system with $\theta = 3°$. (a) state estimation results through the approach that was described in section IV-B, (b) recovering the real part of the states via matrix completion formulation as described in section IV-C.

However, the empirical analysis shows that the likelihood of becoming stuck in such undesirable solutions is very low if the number of measurements is relatively high and the initial point for the local search solver is somewhat close to the true state of the system. Note that the closeness of the initial guess is often satisfied under a normal condition if the flat starting point, i.e. all per unit voltage magnitudes are one and the angles are zero, is considered. Furthermore, the NLAV is capable of recovering the true state of the system with an overwhelming probability in presence of outliers, while the NLS completely fails to recover the true state in this scenario. This is particularly important for power system state estimation, because one main goal of this task is to monitor the system, which involves detecting topological errors, erroneous lines, etc. All these events manifest themselves in the form of outliers in the measurements that can completely corrupt the results of the NLS. Therefore, the findings of this study can be used to improve the commonly used methods of control and monitoring the power grids.

VI. Conclusion

It is known that the LAV estimator outperforms the least-squares estimator for linear systems in presence of sparse measurement errors of arbitrary magnitudes. However, the properties of the NLAV estimator in the nonlinear case are not known due to the high nonlinearity of the problem and the non-smoothness of this estimator. In this work, we compare the performances of the NLAV and NLS for real-world scenarios by estimating the success rate of these estimators with and without sparse noise, and by investigating the effect of prior information and number of measurements on the success rate of each estimator. We also follow a robust matrix completion approach to provide necessary conditions for successful recovery of NLAV and investigate the effect of deviating from those conditions.

Our results show that in the absence of noise, if the number of measurements is above a threshold, the NLS always recovers the true state of the system. However, such a threshold does not exist for the NLAV, and this estimator could always have spurious local solutions even if the full set of power measurements (all voltage magnitudes and line power flow measurements) is used. In contrast, the NLS completely fails to recover the true state when there is sparse noise in the measurements, while the NLAV recovers the underlying state with overwhelming probability in both noisy and noiseless scenarios. It is also showed that to avoid the NLAV estimator to become stuck in spurious local solutions, the voltage angles should be small at all buses; otherwise, the landscape of this estimator has local solutions that are different from the global solution, regardless of the magnitude of the voltage angles. However, the probability of recovering the true solution via the NLAV is significantly large if the measurement redundancy is large and the initial guess for the solver is relatively close to the true state.

References

[1] K. Clements and P. Davis, "Detection and identification of topology errors in electric power systems," *IEEE Transactions on Power Systems*, vol. 3, no. 4, pp. 1748–1753, 1988.

[2] Y. Lin and A. Abur, "Probabilistic load-dependent cascading failure with limited component interactions," in *Proceedings of the International Symposium on Circuits and Systems, Vancouver, BC, Canada, May*, 2004.

[3] A. Monticelli, "Electric power system state estimation," *Proceedings of the IEEE*, vol. 88, no. 2, pp. 262–282, 2000.

[4] R. Khorshidi, F. Shabaninia, and T. Niknam, "A new smart approach for state estimation of distribution grids considering renewable energy sources," *Energy*, vol. 94, pp. 29–37, 2016.

[5] V. Basetti, A. K. Chandel, and R. Chandel, "Power system dynamic state estimation using prediction based evolutionary technique," *Energy*, vol. 107, pp. 29–47, 2016.

[6] F. C. Schweppe and J. Wildes, "Power system static-state estimation, Part I: Exact model," *IEEE Transactions on Power Apparatus and Systems*, no. 1, pp. 120–125, 1970.

[7] F. C. Schweppe, "Power system static-state estimation, Part III: Implementation," *IEEE Transactions on Power Apparatus and Systems*, pp. 130–135, 1970.

[8] R. Zhang, J. Lavaei, and R. Baldick, "Spurious critical points in power system state estimation," in *51st Hawaii International Conference on System Sciences*, 2018.

[9] K. A. Clements and A. S. Costa, "Topology error identification using normalized lagrange multipliers," *IEEE Transactions on power systems*, vol. 13, no. 2, pp. 347–353, 1998.

[10] Y. Weng, M. D. Ilic, Q. Li, and R. Negi, "Convexification of bad data and topology error detection and identification problems in ac electric power systems," *IET Generation, Transmission & Distribution*, vol. 9, no. 16, pp. 2760–2767, 2015.

[11] M. Korkali and A. Abur, "Robust fault location using least-absolute-value estimator," *IEEE Transactions on power systems*, vol. 28, no. 4, pp. 4384–4392, 2013.

[12] Y. Lin and A. Abur, "Robust state estimation against measurement and network parameter errors," *IEEE Transactions on power systems*, vol. 33, no. 5, pp. 4751–4759, 2018.

[13] S. Park, R. Mohammadi-Ghazi, and J. Lavaei. (2018) Nonlinear least absolute value estimator for power systems state estimation and topological error detection. [Online]. Available: http://www.ieor.berkeley.edu/~lavaei/SE_TC_1_2018.pdf

[14] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Springer Series in Statistics, 2009.

[15] B. Donmez and A. Abur, "Sparse estimation based external system line outage detection," in *Power Systems Computation Conference (PSCC), Genoa, Italy, June*, 2016.

[16] M. Majidi, M. Etezadi-Amoli, and H. Livani, "Distribution system state estimation using compressive sensing," *Electrical Power and Energy Systems*, vol. 88, pp. 175–186, 2017.

[17] S. Goleijani and M. T. Ameli, "A multi-agent based approach to power system dynamic state estimation by considering algebraic and dynamic state variables," *Electric Power Systems Research*, vol. 163, pp. 470–481, 2018.

[18] C. Josz, Y. Ouyang, R. Y. Zhang, J. Lavaei, and S. Sojoudi, "A theory on the absence of spurious solutions for nonconvex and nonsmooth optimization," in *Thirty-second Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2018.

[19] R. Zhang, J. Lavaei, and R. Baldick, "Spurious local minima in power system state estimation," in *Proceedings of the 51st Hawaii International Conference on System Sciences*, 2018.

[20] R. Madani, J. Lavaei, and R. Baldick, "Convexification of power flow equations in the presence of noisy measurements," *IEEE Transactions on Automatic Control*, 2019.

[21] S. Fattahi and S. Sojoudi. (2018) Exact gurantees on the absence of spurious local minima for positive robust principal component analysis. [Online]. Available: https://arxiv.org/abs/1812.11466

[22] X. Li, Z. Zhu, A. M.-C. So, and R. Vidal. (2018) Nonconvex robust low-rank matrix recovery. [Online]. Available: https://arxiv.org/abs/1809.09237

[23] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, pp. 471–501, 2010.

[24] E. J. Canded, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM (JACM)*, vol. 58, 2011.

[25] Y. Chen, S. Bhojanapalli, S. Sanghavi, and RachelWard, "Coherent matrix completion," in *Proceedings of the 31-st International Conference on Machine Learning, Beijing, China*, 2014.

[26] R. Ge, J. D. Lee, and T. Ma, "Matrix completion has no spurious local minimum," in *Annual Conference on Neural Information Processing Systems (NIPS)*, 2016.

[27] Y. Chi, Y. M. Lu, and Y. Chen. (2018) Nonconvex optimization meets low-rank matrix factorization: An overview. [Online]. Available: https://arxiv.org/abs/1809.09573

[28] E. J. Candes and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, pp. 717–772, 2009.

[29] B. Recht, "A simpler approach to matrix completion," *The Journal of Machine Learning Research*, vol. 12, pp. 3413–3430, 2011.

[30] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2009.

[31] K. Khamaru and M. J. Wainwright. (2018) Convergence guarantees for a class of non-convex and non-smooth optimization problems. [Online]. Available: https://arxiv.org/pdf/1804.09629.pdf