GENERAL INSTRUCTION

- **Authors: We cannot accept new source files as corrections for your paper. If possible, please annotate the PDF proof we have sent you with your corrections and upload it via the Author Gateway. Alternatively, you may send us your corrections in list format. You may also upload revised graphics via the Author Gateway.**

Queries

Q1. Author: Please confirm or add details for any funding or financial support for the research of this article.
Q2. Author: Please provide the expansion of LS.
Q3. Author: We have renumbered the equations after equation (16) in order to maintain the sequential order, as per the IEEE style. Please check.
Q4. Author: Please provide the report number in Ref. [9].
Q5. Author: Please update Refs. [13], [21], [29], and [30].
Q6. Author: Please provide the year in Ref. [32].

Robust Power System State Estimation From Rank-One Measurements

Gang Wang, *Member, IEEE*, Hao Zhu, *Member, IEEE*, Georgios B. Giannakis, *Fellow, IEEE*, and Jian Sun, *Member, IEEE*

***Abstract*—The unique features of current and upcoming energy systems, namely high penetration of uncertain renewables, unpredictable customer participation, and purposeful manipulation of meter readings, all highlight the need for fast and robust power system state estimation (PSSE). In the absence of noise, PSSE is equivalent to solving a system of quadratic equations, which, also related to power flow analysis, is NP-hard in general. Assuming the availability of all power flow and voltage magnitude measurements, this paper first suggests a simple algebraic technique to transform the power flows into rank-one measurements, for which the ℓ_1-based misfit is minimized. To uniquely cope with the nonconvexity and nonsmoothness of ℓ_1-based PSSE, a deterministic proximal-linear solver is developed based on composite optimization, whose generalization using stochastic gradients is discussed too. This paper also develops conditions on the ℓ_1-based loss function such that exact recovery and quadratic convergence of the proposed scheme are guaranteed. Simulated tests using several IEEE benchmark test systems under different settings corroborate our theoretical findings, as well as the fast convergence and robustness of the proposed approaches.**

***Index Terms*—Bad data analysis, composite optimization, least-absolute-value ~~(LAV)~~ estimator, proximal-linear algorithm, supervisory control and data acquisition ~~(SCADA)~~ measurement.**

I. Introduction

THE North American power grid is praised as the greatest engineering achievement of the 20th century [1]. To maintain grid efficiency, reliability, and sustainability, system operators constantly monitor the operating conditions of electricity networks [2], [3]. In the 1960s, power system engineers tried to compute voltages at critical buses based on meter readings manually collected from geometrically distributed current and potential transformers. Due to timing, model mismatches, and metering errors, however, the exact (ac) power flow equations were never infeasible.

With the development of supervisory control and data acquisition (SCADA) systems, a wealth of improved data metered from across the network became available. In the seminal work of Schweppe *et al.* [4], the modern statistical foundation for power system state estimation (PSSE) was laid. Given a collection of SCADA data along with corresponding measurement matrices, the goal of PSSE is to compute the complex voltages (or, the voltage magnitudes and angles if polar coordinates are used) at all network buses. Since then, substantial contributions have been devoted to PSSE. Interested readers can refer to [3] for a review of recent developments on PSSE.

Based on the weighted least-squares (WLS) estimation criterion, the Gauss–Newton solver is arguably the "workhorse" for PSSE, and it is also employed in practice [2]. Yet, the nonconvex nature of WLS poses challenges on the Gauss–Newton method, including sensitivity to initialization and outliers, as well as no convergence guarantee [5]. To address these challenges, semidefinite programming (SDP) relaxation approaches have been pursued [6]–[8]. However, SDP incurs computational complexity that does not scale well with problem dimension, discouraging its use in practical settings.

With utilities increasingly shifting toward smart grid technology and other upgrades with inherent cyber vulnerabilities, correlative threats from adversarial cyberattacks on the North American power grid continue to grow in frequency and form [9]. These introduce new yet critical challenges to PSSE, particularly to the WLS-based SE solvers, concerning data integrity and uninformed model changes [10]–[14]. Such concerns motivate well the development of accurate and robust approaches to endow PSSE with resilience to anomalous (i.e., bad) data and model inaccuracies.

A. Related Work

In this context, robust PSSE has recently received renewed interest. To cope with the malicious data, the largest normalized residual ~~(LNR)~~ test was incorporated while performing PSSE [10]. The least-median-squares and least-trimmed-squares

Manuscript received September 17, 2018; revised September 18, 2018, November 10, 2018, and November 30, 2018; accepted December 5, 2018. Date of publication; date of current version. The work of G. Wang and G. B. Giannakis was supported by the NSF under Grant 1514056, Grant 1505970, and Grant 1711471. The work of H. Zhu was supported by the NSF under Grant 1802319. The work of J. Sun was supported in part by ~~the NSFC under Grant U1509215 and Grant 61621063, and in part by the Program for Changjiang Scholars and Innovative Research~~ Team in University (IRT1208). Recommended by Associate Editor Y. Hong. *(Corresponding author: Jian Sun.)*

G. Wang and G. B. Giannakis are with the Digital Technology Center and the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: gangwang@umn.edu; georgios@umn.edu).

H. Zhu is with the Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: haozhu@utexas.edu).

J. Sun is with the State Key Laboratory of Intelligent Control and Decision of Complex Systems and the School of Automation, Beijing Institute of Technology, Beijing 100081, China (e-mail: sunjian@bit.edu.cn).

Digital Object Identifier 10.1109/TCNS.2019.2890954

based alternatives were pursued [15]. Unfortunately, the aforementioned robust PSSE proposals incur unfavorable computational complexities and/or stringent storage requirements, which limit their practical uses in real-world power networks.

On the other hand, the ℓ_1-based criterion has been well known in optimization and statistics for its robustness to outliers [16], [2, Ch. 6]. In addition to being robust, the ℓ_1-based estimator is also statistically optimal in the maximum likelihood sense, when the independent additive noise follows a Laplacian distribution. In the PSSE literature, the ℓ_1-based criterion was advocated for bad data identification and rejection in [17]. Research focus has shifted toward devising efficient and user-friendly algorithms to handle the nonconvexity and nonsmoothness issues of ℓ_1-loss function; see, e.g., [18] and [19].

B. This Paper

This paper revisits the ℓ_1-based robust PSSE with a focus on development of efficient algorithms and theory on exact state recovery in the noiseless case. Leveraging recent advances in solving rank-one quadratic equations (i.e., phase retrieval) [20], [21], we first suggest a simple algebraic procedure to transform the power flows into rank-one measurements, namely with corresponding transformed measurement matrices being rank one. Subsequently, we develop two efficient and easy-to-implement algorithms to optimize the ℓ_1-loss of the obtained rank-one measurements. With appropriate conditions on the ℓ_1-loss function, we establish exact recovery as well as quadratic convergence for our approach. Simulated tests using three IEEE benchmark systems showcase the robustness and computational efficiency of our proposed scheme relative to competing Gauss–Newton method.

The rest of this paper is organized as follows. System modeling and problem formulation are given in Section II. The procedure to obtain rank-one measurements is presented in Section III, followed by two algorithms in Section IV. Exact state recovery and convergence are established in Section V. Numerical tests are provided in Section VI, and this paper is concluded in Section VII.

Notation: Matrices (column vectors) are denoted by upper- (lower-) case boldface letters; e.g., $\mathbf{A}$ ($\mathbf{a}$). Sets are denoted using calligraphic letters. Symbol $^{\mathcal{T}}$ ($^{\mathcal{H}}$) represents (Hermitian) transpose, and $\overline{(\cdot)}$ complex conjugate, whereas $\Re(\cdot)$ ($\Im(\cdot)$) takes the real (imaginary) part of a complex number.

II. System Modeling and Problem Formulation

A. System Modeling

Consider an electric power grid modeled as a graph $\mathcal{G} = (\mathcal{N}, \mathcal{L})$, whose nodes $\mathcal{N} := \{1, 2, \ldots, N\}$ correspond to buses and whose edges $\mathcal{L} := \{(n, n')\} \subseteq \mathcal{N} \times \mathcal{N}$ correspond to lines. Throughout this paper, all analysis pertains to the per unit (p.u.) system. The complex voltage per bus $n \in \mathcal{N}$ can be given in rectangular coordinates as $v_n = \Re(v_n) + j\Im(v_n)$. For brevity, all nodal voltages are stacked up to form the vector $\mathbf{v} := [v_1 \cdots v_N]^{\mathcal{T}} \in \mathbb{C}^N$. In the ac-based SE literature, a subset of following system variables can be measured by SCADA [2, Ch. 2]:

1) $|v_n|$*:* the voltage magnitude at bus n;
2) $P_{nn'}$ *(*$Q_{nn'}$*):* the active (reactive) power flow from buses n to n' at the sending terminal;
3) P_n *(*Q_n*):* the active (reactive) power injection into bus n.

Compliant with the ac power flow model [2], these system variables can be expressed as quadratic functions of $\mathbf{v}$. This justifies why the voltage vector $\mathbf{v}$ is referred to as the system state. To this end, observe that the squared voltage magnitudes $V_n := |v_n|^2 = [\Re(v_n)]^2 + [\Im(v_n)]^2$ can be written as

$$V_n = \mathbf{v}^{\mathcal{H}}\mathbf{H}_n^V\mathbf{v}, \text{ with } \mathbf{H}_n^V := \mathbf{h}_n\mathbf{h}_n^{\mathcal{T}} \tag{1}$$

for all $n \in \mathcal{N}$, where we have introduced the measurement vector $\mathbf{h}_n := \mathbf{e}_n$ with $\mathbf{e}_n$ being the nth canonical vector in $\mathbb{R}^N$. To express power injections as functions of $\mathbf{v}$, introduce the bus admittance matrix $\mathbf{Y} = \mathbf{G} + j\mathbf{B} \in \mathbb{C}^N$, where $\mathbf{G}$ and $\mathbf{B} \in \mathbb{R}^{N\times N}$ are the real and imaginary parts of $\mathbf{Y}$ [2], respectively. In rectangular coordinates, the active and reactive powers P_n and Q_n injected into bus n can be expressed as

$$\begin{aligned} P_n &= \Re(v_n)\sum_{n'=1}^{N}\left[\Re(v_{n'})G_{nn'} - \Im(v_n)B_{nn'}\right] \\ &\quad + \Im(v_n)\sum_{n'=1}^{N}\left[\Im(v_n)G_{nn'} + \Re(v_n)B_{nn'}\right] \end{aligned} \tag{2}$$

$$\begin{aligned} Q_n &= \Im(v_n)\sum_{n'=1}^{N}\left[\Re(v_n)G_{nn'} - \Im(v_n)B_{nn'}\right] \\ &\quad - \Re(v_n)\sum_{n'=1}^{N}\left[\Im(v_n)G_{nn'} + \Re(v_n)B_{nn'}\right] \end{aligned} \tag{3}$$

which can be compactly expressed as

$$P_n = \mathbf{v}^{\mathcal{H}}\mathbf{H}_n^P\mathbf{v}, \text{ with } \mathbf{H}_n^P := \frac{\mathbf{Y}_n^{\mathcal{H}} + \mathbf{Y}_n}{2} \tag{4a}$$

$$Q_n = \mathbf{v}^{\mathcal{H}}\mathbf{H}_n^Q\mathbf{v}, \text{ with } \mathbf{H}_n^Q := \frac{\mathbf{Y}_n^{\mathcal{H}} - \mathbf{Y}_n}{2j} \tag{4b}$$

with $\mathbf{Y}_n := \mathbf{e}_n\mathbf{e}_n^{\mathcal{T}}\mathbf{Y}$ for all $n \in \mathcal{N}$.

With regards to power flows, Kirchhoff's current law dictates that the complex current over the line (n, n') at the "sending" end is $i_{nn'} = y_{nn'}^s v_n + y_{nn'}(v_n - v_n')$, where $y_{nn'}^s$ is the shunt admittance at bus n' associated with the line (n, n'). The ac power flow model, in conjunction with Ohm's law, further asserts that the complex power flowing over line (n, n') at the "sending" end can be expressed as

$$\begin{aligned} S_{nn'} &= P_{nn'} + jQ_{nn'} = v_n\bar{i}_{nn'} \\ &= |v_n|^2(\bar{y}_{nn'}^s + \bar{y}_{nn'}) - v_n\bar{v}_n\bar{y}_{nn'}. \end{aligned} \tag{5}$$

It is worth pointing out that the complex power flow over line $(n, n') \in \mathcal{L}$ at the "receiving" end is captured by that over line $(n', n) \in \mathcal{L}$ at the "sending" end. Upon defining the following matrices for all lines $(n, n') \in \mathcal{L}$:

$$\mathbf{Y}_{nn'} := (y_{nn'}^s + y_{nn'})\mathbf{e}_n\mathbf{e}_n^{\mathcal{T}} - \bar{y}_{nn'}\mathbf{e}_{n'}\mathbf{e}_{n'}^{\mathcal{T}} \tag{6}$$

the active and reactive power flows at the "sending" terminal, namely the real and imaginary parts of $S_{nn'}$ in (5) can be given in a compact representation as

$$P_{nn'} = \mathbf{v}^{\mathcal{H}} \mathbf{H}^{P}_{nn'} \mathbf{v}, \text{ with } \mathbf{H}^{P}_{nn'} := \frac{\mathbf{Y}^{\mathcal{H}}_{nn'} + \mathbf{Y}_{nn'}}{2} \tag{7a}$$

$$Q_{nn'} = \mathbf{v}^{\mathcal{H}} \mathbf{H}^{Q}_{nn'} \mathbf{v}, \text{ with } \mathbf{H}^{Q}_{nn'} := \frac{\mathbf{Y}^{\mathcal{H}}_{nn'} - \mathbf{Y}_{nn'}}{2j}. \tag{7b}$$

Having expressed all SCADA measurements as functions of $\mathbf{v}$, the PSSE problem can be presented next.

B. Power System State Estimation

In practice, the SCADA system measures a subset of the system variables specified in (1), (4), and (7). Suppose now a total of M such variables are measured, which are stacked up to form the following $M \times 1$ measurement vector:

$$\mathbf{z} := \big[\{\check{V}_n\}_{n\in\mathcal{N}_V}, \{\check{P}_n\}_{n\in\mathcal{N}_P}, \{\check{Q}_n\}_{n\in\mathcal{N}_Q} \\ \{\check{P}_{nn'}\}_{(n,n')\in\mathcal{L}_P}, \{\check{Q}_{nn'}\}_{(n,n')\in\mathcal{L}_Q}\big]^{\mathcal{T}} \in \mathbb{R}^M \tag{8}$$

where the check-marked terms represent possibly noisy observations of the corresponding error-free variables. The subsets $\mathcal{N}_V, \mathcal{N}_P, \mathcal{N}_Q \subseteq \mathcal{N}$, and $\mathcal{L}_P, \mathcal{L}_Q \subseteq \mathcal{L}$ specify the locations where meters are installed and the associated type of variables are measured. Succinctly, per mth measurement in $\mathbf{z}$ can be equivalently rewritten as

$$z_m := \mathbf{v}^{\mathcal{H}} \mathbf{H}_m \mathbf{v} + \epsilon_m \tag{9}$$

for all $m \in \{1, 2, \ldots, M\}$, where the terms $\epsilon_m \in \mathbb{R}$ capture the metering errors and modeling inaccuracies, and the Hermitian measurement matrices $\mathbf{H}_m \in \mathbb{C}^{N\times N}$ can correspond to any subset of the matrices defined in (1), (4), and (7). The critical goal of PSSE is to obtain $\mathbf{v} \in \mathbb{C}^N$ based on the available data $\{(z_m; \mathbf{H}_m)\}_{m=1}^M$.

Without loss of generality, adopting the LS error objective, which coincides with the maximum likelihood criterion assuming additive white Gaussian noise, PSSE pursues problem[1]

$$\underset{\mathbf{x}\in\mathbb{C}^N}{\text{minimize}}\ \ell(\mathbf{x}) := \frac{1}{2M} \sum_{m=1}^{M} \left(z_m - \mathbf{x}^{\mathcal{H}} \mathbf{H}_m \mathbf{x}\right)^2. \tag{10}$$

Because of the quadratic terms inside the squares, the *quartic* function $\ell(\mathbf{x})$ is nonconvex, whose general instance is NP-hard [22]. Hence, it is computationally intractable to compute the LS or maximum likelihood estimate of $\mathbf{v}$ in general.

Q2

C. Prior Contributions

Minimizing the nonlinear LS loss in (10), the Gauss–Newton method is the "workhorse" [2, Ch. 2]. Upon linearizing all quadratic terms $\mathbf{x}^{\mathcal{H}} \mathbf{H}_m \mathbf{x}$ around a given point using Taylor's expansion, the Gauss–Newton subsequently approximates the nonlinear LS fit in (10) using a linear one per iteration, and relies on its resultant minimizer to obtain the next iterate [3]. It typically converges in a few (≤ 10) iterations, very fast for small- or medium-size problems. However, it is known that the Gauss–Newton iterations for nonconvex LS are sensitive to the initial point, and they may diverge in certain cases; see e.g., [5, Ch. 5].

On the other hand, several numerical polynomial-time SE algorithms have been pursued based on convex programming [6]. By means of matrix lifting, such convex approaches start expressing all quadratic measurements $\mathbf{x}^{\mathcal{H}} \mathbf{H}_m \mathbf{x}$ as linear functions $\mathrm{Tr}(\mathbf{H}_m \mathbf{X})$ of the rank-one matrix variable $\mathbf{X} := \mathbf{x}\mathbf{x}^{\mathcal{H}} \succeq \mathbf{0}$. Upon discarding the nonconvex rank constraint, the nonlinear LS in (10) boils down to (or can be converted into) a convex SDP. In terms of computational efficiency, such convex schemes entail solving for an $N \times N$ positive semidefinite matrix from M SDP constraints, whose worst case computational complexity is $\mathcal{O}(M^4 N^{1/2} \log(1/\epsilon))$ for any given solution accuracy $\epsilon > 0$ [23, Sec. 6.6.2]. This complexity and the resultant storage requirement evidently do not scale nicely to the increasingly interconnected large power networks.

III. Rank-One Measurement Approach

Drawing from advances in nonconvex optimization, this section presents a new framework for scalable, accurate, and robust PSSE. Specifically, our proposed approach reformulates the rank-two power flows in (7) into rank-one quadratic measurements (namely with corresponding measurement matrices having rank one), followed by two efficient algorithms for minimizing the ℓ_1-based misfit of the transformed measurements. In contrast to previous SE approaches that minimize the *quartic* polynomial in (10), a novel *quadratic* objective functional is obtained and subsequently minimized. With more complicated algebraic manipulations, the power injections can also be accounted for in our proposed framework. In the presence of additive noise, near-optimal statistical performance of the developed approach is numerically demonstrated.

A. Measurement Transformation

In this paper, we focus on the following types of measurements: first, $\{|v_m|\}_{m=1}^N$ the voltage magnitudes at all buses, and second, $\{P_{nn'}\}_{(m,n)\in\mathcal{L}_P}$ and/or $\{Q_{nn'}\}_{(m,n)\in\mathcal{L}_Q}$ the active and/or reactive power flows on a selected subset of lines, namely $\mathcal{L}_P, \mathcal{L}_Q \subseteq \mathcal{L}$. Consider first the noise-free case, where all available measurements can be described as

$$z_m = \mathbf{v}^{\mathcal{H}} \mathbf{H}_m \mathbf{v} \quad \forall m = 1, \ldots, M. \tag{11}$$

Without loss of generality, let the first N measurements be the squared voltage magnitudes at the N buses, namely $z_m = |v_m|^2$ for $m = 1, 2, \ldots, N$, and the remaining ones be the (active or reactive) power flows. It is clear from (1) that the squared voltage magnitude measurements are given by

$$z_m = \left|\mathbf{h}_m^{\mathcal{H}} \mathbf{v}\right|^2 \quad \forall m = 1, \ldots, N \tag{12}$$

whose corresponding measurement matrices have rank one; that is, $\{\mathbf{H}_m = \mathbf{h}_m \mathbf{h}_m^{\mathcal{H}}\}_{m=1}^N$.

Now let us consider the power flow data pairs $(P_{nn'}; \mathbf{H}^P_{nn'})$ and $(Q_{nn'}; \mathbf{H}^Q_{nn'})$ in (7). Upon substituting $\mathbf{Y}_{nn'}$ in (6) into (7),

[1]Throughout, $\mathbf{v}$ is fixed for the actual system state, whereas $\mathbf{x}$ is used for the optimization variable and the state estimate.

one can rewrite for all lines $(n, n') \in \mathcal{L}_P$

$$\mathbf{H}^P_{nn'} = \frac{1}{2}\left(\alpha^P_{nn'}\mathbf{e}_n\mathbf{e}_n^{\mathcal{T}} - \beta^P_{nn'}\mathbf{e}_n\mathbf{e}_{n'}^{\mathcal{T}} - \bar{\beta}^P_{nn'}\mathbf{e}_{n'}\mathbf{e}_n^{\mathcal{T}}\right) \tag{13}$$

and similarly for all lines $(n, n') \in \mathcal{L}_Q$

$$\mathbf{H}^Q_{nn'} = \frac{1}{2}\left(\alpha^Q_{nn'}\mathbf{e}_n\mathbf{e}_n^{\mathcal{T}} - \beta^Q_{nn'}\mathbf{e}_n\mathbf{e}_{n'}^{\mathcal{T}} - \bar{\beta}^Q_{nn'}\mathbf{e}_{n'}\mathbf{e}_n^{\mathcal{T}}\right) \tag{14}$$

where the four coefficients are given by

$$\alpha^P_{nn'} := 2\Re(y^s_{nn'} + y_{nn'}), \qquad \beta^P_{nn'} := y_{nn'} \tag{15a}$$

$$\alpha^Q_{nn'} := -2\Im(y^s_{nn'} + y_{nn'}), \qquad \beta^Q_{nn'} := jy_{nn'}. \tag{15b}$$

It is worth pointing out that $\alpha^P_{nn'}$ and $\alpha^Q_{nn'}$ are both real, whereas $\beta^P_{nn'}$ and $\beta^Q_{nn'}$ are in general complex.

It is self-evident from (13) and (14) that each of the active or reactive power flow measurement matrices $\{\mathbf{H}^P_{nn'}\}_{(n,n')\in\mathcal{L}_P}$ or $\{\mathbf{H}^Q_{nn'}\}_{(n,n')\in\mathcal{L}_Q}$ is of *at most* rank-two, with three nonzero matrix entries at the (n, n)-, (n, n')-, and (n', n)th positions depending on whether the involved coefficients are nonzero or not. Yet, these power flow matrices are generally indefinite. Upon neglecting the superscripts for notational brevity, both $\mathbf{H}^P_{nn'}$ and $\mathbf{H}^Q_{nn'}$ are in the following form:

$$\mathbf{H}_{nn'} = \frac{1}{2}\left(\alpha_{nn'}\mathbf{e}_n\mathbf{e}_n^{\mathcal{T}} - \beta_{nn'}\mathbf{e}_n\mathbf{e}_{n'}^{\mathcal{T}} - \bar{\beta}_{nn'}\mathbf{e}_{n'}\mathbf{e}_n^{\mathcal{T}}\right). \tag{16}$$

We establish the following result for power flow data $\{z_m\}$.

Proposition 1 (Rank-one measurements): Suppose that the voltage magnitudes are measured at all buses. Then, one can construct equivalently a new measurement $\check{z}_m$ for each power flow z_m so that its corresponding measurement matrix $\check{\mathbf{H}}_{nn'}$ is of rank one, and positive semidefinite.

Proof: Depending on whether the coefficient $\alpha_{nn'}$ or $\beta_{nn'}$ is zero or not, we discuss separately the ensuing three cases:

c1) $\alpha_{nn'} \neq 0, \beta_{nn'} \neq 0$;

c2) $\alpha_{nn'} = 0, \beta_{nn'} \neq 0$;

c3) $\alpha_{nn'} \neq 0, \beta_{nn'} = 0$

where the trivial case of $\alpha_{nn'} = \beta_{nn'} = 0$ is excluded because all SCADA measurements are assumed nonzero. The goal of the following section is to transform all power flow measurement matrices into *rank-one (symmetric) positive semidefinite* matrices. The three cases are individually discussed next . Q3

Consider first the case c1). When both $\alpha_{nn'}$ and $\beta_{nn'}$ are nonzero, the matrix $\mathbf{H}_{nn'}$ has exactly rank two with three nonzero entries. Its two nonzero eigenvalues can be obtained by solving the quadratic equation

$$\lambda\left(\lambda - \frac{\alpha_{nn'}}{2}\right) - \frac{|\beta_{nn'}|^2}{4} = 0$$

which is derived by setting the determinant of $(\lambda\mathbf{I}_N - \mathbf{H}_{nn'})$ to zero. Its closed-form solutions are given by

$$\lambda_1 = \frac{\alpha_{nn'} + \sqrt{\alpha^2_{nn'} + 4|\beta_{nn'}|^2}}{4} > 0$$

$$\lambda_2 = \frac{\alpha_{nn'} - \sqrt{\alpha^2_{nn'} + 4|\beta_{nn'}|^2}}{4} < 0.$$

Let $\mathbf{u}_1, \mathbf{u}_2 \in \mathbb{C}^N$ be the unit eigenvectors of $\mathbf{H}_{nn'}$ associated with the eigenvalues λ_1, λ_2, respectively. Hence, one can write $\mathbf{H}_{nn'} := \lambda_1\mathbf{u}_1\mathbf{u}_1^{\mathcal{H}} + \lambda_2\mathbf{u}_2\mathbf{u}_2^{\mathcal{H}}$. To obtain a rank-one positive semidefinite matrix, the first attempt would be to compensate for the negative eigenvalue λ_2 and make it zero. This is tantamount to adding $-\lambda_2\mathbf{u}_2\mathbf{u}_2^{\mathcal{H}}$ to $\mathbf{H}_{nn'}$, and accordingly adding $-\lambda_2\mathbf{v}^{\mathcal{H}}(\mathbf{u}_2\mathbf{u}_2^{\mathcal{H}})\mathbf{v}$ to the measurement $z_{nn'}$; that is

$$\check{\mathbf{H}}_{nn'} := \mathbf{H}_{nn'} - \lambda_2\mathbf{u}_2\mathbf{u}_2^{\mathcal{H}} \tag{17a}$$

$$\check{z}_{nn'} := z_{nn'} - \lambda_2\mathbf{v}^{\mathcal{H}}(\mathbf{u}_2\mathbf{u}_2^{\mathcal{H}})\mathbf{v} \tag{17b}$$

in which the transformed measurement matrix $\check{\mathbf{H}}_{nn'}$ is rank-one and symmetric positive semidefinite, and $\check{z}_{nn'}$ is the resultant transformed measurement. To realize this however, entails, evaluating the term $\mathbf{v}^{\mathcal{H}}(\mathbf{u}_2\mathbf{u}_2^{\mathcal{H}})\mathbf{v}$, which requires knowledge of the true state vector $\mathbf{v}$. This procedure is thus not feasible, and one has to develop a new twist to bypass this hurdle.

Recall from our working assumption that we have access to all squared voltage magnitudes $\{|v_n|^2\}_{n=1}^N$. Based on this fact, we show in the following that it is sufficient to add a matrix of the form $(\delta_{nn'}/2)\mathbf{e}_{n'}\mathbf{e}_{n'}^{\mathcal{T}}$ to $\mathbf{H}_{nn'}$ such that the resulting sum, denoted by

$$\check{\mathbf{H}}_{nn'} := \mathbf{H}_{nn'} + (\delta_{nn'}/2)\mathbf{e}_{n'}\mathbf{e}_{n'}^{\mathcal{T}} \tag{18}$$

can be rendered rank-one. Here, $\delta_{nn'}$ is an unknown coefficient to be determined next. Toward this end, setting the determinant of $(\lambda\mathbf{I}_N - \check{\mathbf{H}}_{nn'})$ to zero leads to

$$\left(\lambda - \frac{\alpha_{nn'}}{2}\right)\left(\lambda - \frac{\delta_{nn'}}{2}\right) - \frac{|\beta_{nn'}|^2}{4} = 0. \tag{19}$$

To yield a rank-one matrix $\check{\mathbf{H}}_{nn'}$, it is sufficient for the quadratic equation (19) to have exactly one nonzero solution. By basic linear algebra, this is equivalent to having a zero constant term in (19), giving rise to

$$\alpha_{nn'}\delta_{nn'} - |\beta_{nn'}|^2 = 0$$

or alternatively

$$\delta_{nn'} := |\beta_{nn'}|^2/\alpha_{nn'}.$$

It can be verified that the transformed measurement matrix

$$\check{\mathbf{H}}_{nn'} := \mathbf{H}_{nn'} + (|\beta_{nn'}|^2/(2\alpha_{nn'}))\mathbf{e}_{n'}\mathbf{e}_{n'}^{\mathcal{T}} \tag{20}$$

is rank-one. In addition, if $\alpha_{nn'} > 0$, then $\check{\mathbf{H}}_{nn'}$ is positive semidefinite. Therefore, one can write

$$\check{\mathbf{H}}_{nn'} := \mathbf{h}_{nn'}\mathbf{h}_{nn'}^{\mathcal{H}} \tag{21}$$

with the equivalent measurement vector being

$$\mathbf{h}_{nn'} := \sqrt{\frac{\alpha_{nn'}}{2}}\mathbf{e}_n + \frac{\bar{\beta}_{nn'}}{\sqrt{2\alpha_{nn'}}}\mathbf{e}_{n'}. \tag{22}$$

The transformed measurement $\check{z}_{nn'}$ corresponding to $\check{\mathbf{H}}_{nn'}$ can be given by

$$\begin{aligned}\check{z}_{nn'} &:= \mathbf{v}^{\mathcal{H}}\check{\mathbf{H}}_{nn'}\mathbf{v} = \mathbf{v}^{\mathcal{H}}\mathbf{H}_{nn'}\mathbf{v} + \mathbf{v}^{\mathcal{H}}\left(\delta_{nn'}\mathbf{e}_{n'}\mathbf{e}_{n'}^{\mathcal{T}}/2\right)\mathbf{v}\\ &= z_{nn'} + (|\beta_{nn'}|^2/2\alpha_{nn'})|v_{n'}|^2\end{aligned} \tag{23}$$

for which the required quantity $(|\beta_{nn'}|^2/2\alpha_{nn'})|v_{n'}|^2$ is available, or can be obtained as long as $|v_n|^2$ is available.

If $\alpha_{nn'} < 0$, one can instead define

$$\check{\mathbf{H}}_{nn'} := -\mathbf{H}_{nn'} + (|\beta_{nn'}|^2/(2\alpha_{nn'}))\mathbf{e}_{n'}\mathbf{e}_{n'}^{\mathcal{T}} \tag{24}$$

and write the equivalent measurement vector as

$$\mathbf{h}_{nn'} := \sqrt{\frac{-\alpha_{nn'}}{2}}\mathbf{e}_n + \frac{\bar{\beta}_{nn'}}{\sqrt{-2\alpha_{nn'}}}\mathbf{e}_{n'} \tag{25}$$

for which the corresponding measurement becomes

$$\begin{aligned}\check{z}_{nn'} &:= \mathbf{v}^{\mathcal{H}}\check{\mathbf{H}}_{nn'}\mathbf{v} = -\mathbf{v}^{\mathcal{H}}\mathbf{H}_{nn'}\mathbf{v} - \mathbf{v}^{\mathcal{H}}\left(\delta_{nn'}\mathbf{e}_{n'}\mathbf{e}_{n'}^{\mathcal{T}}/2\right)\mathbf{v}\\ &= -z_{nn'} - (|\beta_{nn'}|^2/2\alpha_{nn'})|v_n|^2. \end{aligned}\tag{26}$$

Likewise, the quantity $-(|\beta_{nn'}|^2/2\alpha_{nn'})|v_n|^2$ is also available under our working assumption.

Let us now focus on the case c2). To transform $\mathbf{H}_{nn'}$ into a rank-one positive semidefinite matrix, it is sufficient to add a matrix of the form $(\gamma_{nn'}/2)\mathbf{e}_n\mathbf{e}_n^{\mathcal{T}} + (\delta_{nn'}/2)\mathbf{e}_{n'}\mathbf{e}_{n'}^{\mathcal{T}}$, to yield

$$\check{\mathbf{H}}_{nn'} := \mathbf{H}_{nn'} + (\gamma_{nn'}/2)\mathbf{e}_n\mathbf{e}_n^{\mathcal{T}} + (\delta_{nn'}/2)\mathbf{e}_{n'}\mathbf{e}_{n'}^{\mathcal{T}} \tag{27}$$

for some coefficients $\gamma_{nn'} > 0$ and $\delta_{nn'} > 0$ to be determined. Similar to the discussion for case c1), to find $\gamma_{nn'}$ and $\delta_{nn'}$, one sets the determinant of $(\lambda\mathbf{I}_N - \check{\mathbf{H}}_{nn'})$ to zero, leading to

$$\left(\lambda - \frac{\gamma_{nn'}}{2}\right)\left(\lambda - \frac{\delta_{nn'}}{2}\right) - \frac{|\beta_{nn'}|^2}{4} = 0.$$

The fact that $\check{\mathbf{H}}_{nn'}$ is rank-one implies that

$$\gamma_{nn'}\delta_{nn'} - |\beta_{nn'}|^2 = 0.$$

Without loss of generality, one can take

$$\gamma_{nn'} := 1, \quad \text{and} \quad \delta_{nn'} := |\beta_{nn'}|^2$$

and $\check{\mathbf{H}}_{nn'}$ in (27) becomes rank-one and can be written as

$$\check{\mathbf{H}}_{nn'} := \mathbf{h}_{nn'}\mathbf{h}_{nn'}^{\mathcal{H}} \tag{28}$$

with

$$\mathbf{h}_{nn'} := \frac{1}{\sqrt{2}}\mathbf{e}_n - \frac{\beta_{nn'}}{\sqrt{2}}\mathbf{e}_{n'}. \tag{29}$$

The transformed measurement associated with $\check{\mathbf{H}}_{nn'}$ can be found as follows:

$$\begin{aligned}\check{z}_{nn'} &:= \mathbf{v}^{\mathcal{H}}\mathbf{H}_{nn'}\mathbf{v}\\ &= \mathbf{v}^{\mathcal{H}}\mathbf{H}_{nn'}\mathbf{v} + \mathbf{v}^{\mathcal{H}}\left(\beta_{nn'}\mathbf{e}_n\mathbf{e}_n^{\mathcal{T}}/2 + \delta_{nn'}\mathbf{e}_{n'}\mathbf{e}_{n'}/2\right)\mathbf{v}\\ &= z_{nn'} + (1/2)|v_n|^2 + (|\beta_{nn'}|^2/2)|v_{n'}|^2 \end{aligned}\tag{30}$$

for which the required quantities $|v_n|^2$ and $|\beta_{nn'}|^2|v_{n'}|^2$ can be computed when $|v_n|^2$ and $|v_{n'}|^2$ are available.

Let us now turn to the last case c3). Since $\beta_{nn'} = \bar{\beta}_{nn'} = 0$, one can write $\mathbf{H}_{nn'} = \alpha_{nn'}\mathbf{e}_n\mathbf{e}_n^{\mathcal{T}}/2$, which is already rank-one. To make it positive semidefinite, it suffices to take the absolute value, and define

$$\check{\mathbf{H}}_{nn'} := \mathbf{h}_{nn'}\mathbf{h}_{nn'}^{\mathcal{H}} = \left(\sqrt{\frac{|\alpha_{nn'}|}{2}}\mathbf{e}_n\right)\left(\sqrt{\frac{|\alpha_{nn'}|}{2}}\mathbf{e}_n\right)^{\mathcal{T}}. \tag{31}$$

The transformed measurement is given by

$$\check{z}_{nn'} = |z_{nn'}|. \tag{32}$$

To summarize, for any active or reactive power flow data $(z_{nn'}; \mathbf{H}_{nn'})$, we have developed a strategy to obtain a new measurement pair $(\check{z}_{nn'}; \check{\mathbf{H}}_{nn'})$, in which $\check{\mathbf{H}}_{nn'}$ becomes rank-one and positive semidefinite. Specifically, this is accomplished through steps in (21)–(32), by depending upon the values of coefficients $\alpha_{nn'}$ and $\beta_{nn'}$, provided that the voltage magnitudes at all buses are available. ■

The assumption on full voltage measurements can be relaxed. Indeed, it is possible to build up rank-one measurements via linear combinations, so long as two of the following SCADA quantities $\{|v_n|, |v_{n'}|, p_{nn'}, p_{n'n}, q_{nn'}, q_{n'n}\}$ are measured on every line $(n, n') \in \mathcal{L}$. In a nutshell, one can readily rewrite all measurements as intensities of some known and deterministic linear transforms of the state vector, namely

$$\check{z}_m = \left|\mathbf{h}_m^{\mathcal{H}}\mathbf{v}\right|^2 \quad \forall m = 1, \ldots, M \tag{33}$$

where the measurement vectors $\mathbf{h}_m \in \mathbb{C}^N$ are given in (1), (22), (25), (29), and (31), whereas the corresponding transformed measurements $\check{z}_m > 0$ are defined in (1), (23), (26), (30), and (32). Moreover, all vectors $\mathbf{h}_m$ are highly sparse, each having at most two nonzero entries independent of the system size N. This feature can be carefully exploited to endow the iterative PSSE solvers with computational efficiency and scalability.

IV. Prox-Linear SE Solvers

In general, given a set of (consistent) quadratic equations, there may exist multiple solutions even after excluding trivial ambiguities. In the context of phase retrieval, in which the measurement matrices are rank-one positive semidefinite (i.e., $\mathbf{H}_m = \mathbf{h}_m\mathbf{h}_m^{\mathcal{H}}$), a number $M \geq 4N - 4$ of random quadratic equations suffice for uniqueness of the solution [24]. In the power systems literature though, it remains an open question that how many quadratic measurements as in (1), (4), and (7) are required for the uniqueness of PSSE solution. For concreteness, this contribution assumes that a large enough number of measurements are available, and they collectively determine a unique solution, namely the underlying true system state.

Leveraging the rank-one measurement model, this section presents two algorithms for scalable and exact power system state recovery based on nonconvex optimization. Specifically, we focus on the ℓ_1-loss to fit the intensity measurements $\check{z}_m$ in (33) instead of z_m in (11). Despite the nonconvexity and nonsmoothness of the resulting loss function, we first develop a deterministic prox-linear algorithm. When the initialization is sufficiently close to the underlying true voltage state vector, and the loss function satisfies a certain local "stability condition," we show that our first deterministic approach recovers the true voltage vector at a quadratic convergence rate. It entails solving a quadratic program per iteration, for which off-the-shelf convex programming toolboxes are widely available, but the resulting complexity does not scale well with the system size. To endow the algorithm with scalability, a stochastic generalization is pursued, which processes a single measurement per iteration.

It is well known in statistics and power systems literature that ℓ_1-based loss functions yield median-based estimators, and they can cope with gross errors in the measurements $\check{z}_m$ in a relatively benign way [21]. This prompts us to consider the ℓ_1-loss [i.e., least-absolute-value (LAV)] formulation

$$\underset{\mathbf{x}\in\mathbb{C}^N}{\text{minimize}}\ \ell(\mathbf{x}) := \frac{1}{M}\sum_{m=1}^{M}\left|\check{z}_m - |\mathbf{h}_m^{\mathcal{H}}\mathbf{x}|^2\right|. \tag{34}$$

Evidently, this loss function is nonsmooth and nonconvex, which is not even locally convex near $\mathbf{v}$. This can be understood from the scalar case $\ell(x) = |1 - x^2|$. As such, it is unclear how to efficiently minimize such functions.

However, the function exhibits several appealing structural properties that we explore in the following to develop iterative algorithms to locally solve the problem (34) efficiently [21], [25], [26]. To this end, consider first expressing the loss function as the composition $\ell(\mathbf{x}) := c(\mathbf{s}(\mathbf{x}))$, of the convex function $c(\cdot) := \|\cdot\|_1$ and the smooth one $\mathbf{s}(\mathbf{x}) := \frac{1}{M}(\check{\mathbf{z}} - |\mathbf{H}\mathbf{x}|^2)$, where $\check{\mathbf{z}} := [\check{z}_1\ \cdots\ \check{z}_M]^{\mathcal{T}}$, $\mathbf{H} := [\mathbf{h}_1\ \cdots\ \mathbf{h}_M]^{\mathcal{H}}$, and the modulus operator $|\cdot|$ is understood elementwise when applied to a vector. Such a compositional structure lends itself nicely to iterative procedures that are referred to as proximal-linear (prox-linear) algorithms, which we are described in detail ahead.

Consider a real-valued $\mathbf{x} \in \mathbb{R}^N$ for now. The (deterministic) prox-linear method for minimizing $\ell(\mathbf{x}) = c(\mathbf{s}(\mathbf{x}))$ is to linearize $\mathbf{s}$ only, followed by successively minimizing a sequence of locally regularized models. Specifically, starting with some initialization $\mathbf{x}_0 \in \mathbb{R}^N$, the prox-linear method defines a local "linearization" of ℓ around the current iterate $\mathbf{x}_t \in \mathbb{R}^N$ as

$$\ell_{\mathbf{x}_t}(\mathbf{x}) := c\big(\mathbf{s}(\mathbf{x}_t) + \nabla^{\mathcal{T}}\mathbf{s}(\mathbf{x}_t)(\mathbf{x} - \mathbf{x}_t)\big) \tag{35}$$

with $\nabla\mathbf{s}(\mathbf{x}_t) \in \mathbb{R}^{N\times M}$ representing the Jacobian matrix of $\mathbf{s}$ evaluated at point $\mathbf{x}_t$; and subsequently, it proceeds inductively to obtain iterates $\mathbf{x}_1, \mathbf{x}_2, \ldots$by minimizing the quadratically regularized models [25], [26]

$$\mathbf{x}_{t+1} = \arg\min_{\mathbf{x}\in\mathbb{R}^N}\left\{\ell_{\mathbf{x}_t}(\mathbf{x}) + \frac{1}{2\mu_t}\|\mathbf{x} - \mathbf{x}_t\|_2^2\right\} \tag{36}$$

where $\mu_t > 0$ is a step size that can be fixed *a priori* to some constant, or be determined "on-the-fly" through a line search [25], [26]. Furthermore, observing that the linearization $\ell_{\mathbf{x}_t}(\mathbf{x})$ is convex in $\mathbf{x}$, so problem (36) is convex in $\mathbf{x}$ as well. It has been shown in [26] that when c is L-Lipschitz and $\nabla\mathbf{s}$ is κ-Lipschitz, choosing any step size $0 < \mu < \frac{1}{\kappa L}$ guarantees that the algorithm (36) is a descent method; that is, the iterates $\{\mathbf{x}_k\}$ monotonically decrease the function value of $\ell(\mathbf{x})$; and finds an (approximate) stationary point of (34).

Nevertheless, the PSSE problem (34) involves optimization over complex-valued variables in $\mathbf{x} \in \mathbb{C}^N$. It can be checked that the functions ℓ and $\mathbf{s}$ do not satisfy the Cauchy–Riemann (CR) equations; see e.g., [27, Th. 7.2] for the definition of CR equations. Hence, functions ℓ and $\mathbf{s}$ are not holomorphic (i.e., complex-differentiable) in $\mathbf{x}$. As such, the "linearization," or the first-order Taylor's expansion of $\mathbf{s}(\mathbf{x})$ in $\mathbf{x} \in \mathbb{C}^N$ alone [cf. (35)] does not exist. To address this challenge, we invoke Wirtinger's calculus to generalize prox-linear algorithms to optimization over complex-valued arguments in the sequel. Please refer to [28] for basics of Wirtinger's calculus.

Algorithm 1: Deterministic Prox-linear SE Solver.

1: **Input** data $\{(z_m, \mathbf{H}_m)\}_{m=1}^M$, step size $\mu > 0$, initialization $\mathbf{v}_0 \in \mathbb{C}^N$, solution accuracy $\epsilon > 0$, and set $t = 0$.
2: **Prepare** the power flow data $\{(z_m, \mathbf{H}_m)\}_{m=N+1}^M$ according to (1), (22)–(23), (25)–(26), (29)–(30), and (31)–(32) to obtain $\{(\check{z}_m, \mathbf{h}_m)\}_{m=N+1}^M$ based on $\{z_m = |v_m|^2\}_{m=1}^N$.
3: **Repeat**
4: **Evaluate** $\mathbf{a}_{m,t}$ and $b_{m,t}$ in (38).
5: **Solve** (37) to yield $\mathbf{x}_{t+1}$.
6: $t = t + 1$.
7: **Until** $\|\mathbf{x}_t - \mathbf{x}_{t-1}\|_2 \le \epsilon\sqrt{N}$.
8: **Return** $\mathbf{x}_t$.

A. Deterministic Prox-Linear SE Solver

Our first deterministic prox-linear approach to (34) is simply stated: begin with initialization $\mathbf{x}_0 := \mathbf{1} \in \mathbb{R}^N$, and proceed by successively minimizing quadratically regularized functions around the current iterate $\mathbf{x}_t \in \mathbb{C}^N$ to yield the next iterate

$$\mathbf{x}_{t+1} = \arg\min_{\mathbf{x}\in\mathbb{C}^N}\frac{1}{M}\sum_{m=1}^{M}\left|b_{m,t} - 2\Re\big(\mathbf{a}_{m,t}^{\mathcal{H}}\mathbf{x}\big)\right| + \frac{1}{2\mu_t}\|\mathbf{x} - \mathbf{x}_t\|_2^2 \tag{37}$$

where the term $b_{m,t} - 2\Re(\mathbf{a}_{m,t}^{\mathcal{H}}\mathbf{x})$ can be interpreted as the first-order Taylor's approximation of the nonholomophic function $\check{z}_m - |\mathbf{h}_m^{\mathcal{H}}\mathbf{x}|^2$ at $\mathbf{x}_t$ based upon the Wirtinger derivatives; see Appendix A for the rigorous derivation. The coefficients $\mathbf{a}_{m,t}$ and $b_{m,t}$ are given by

$$\mathbf{a}_{m,t} := \big(\mathbf{h}_m^{\mathcal{H}}\mathbf{x}_t\big)\,\mathbf{h}_m \tag{38a}$$

$$b_{m,t} := \check{z}_m + \big|\mathbf{h}_m^{\mathcal{H}}\mathbf{x}_t\big|^2. \tag{38b}$$

Observe that the problem (37) to be tackled per iteration of our deterministic prox-linear SE solver is a convex quadratic program, which can be efficiently solved with standard convex programming methods. Under appropriate conditions, our scheme converges quadratically fast to the true state vector $\mathbf{v}$, meaning that we have to solve only about $\log_2\log_2(1/\epsilon)$ such quadratic programs to obtain an estimate $\mathbf{x}$ of $\mathbf{v}$ satisfying $\mathrm{dist}(\mathbf{x}, \mathbf{v}) \le \epsilon\|\mathbf{v}\|_2$. As will be corroborated by our numerical tests in Section VI, this boils down to $5 \sim 8$ convex quadratic programs in practice. Moreover, our approach applies both in the noiseless setting, and when a constant (random) portion of the measurements are even adversarially corrupted.

For implementation, our deterministic prox-linear solver is summarized in Algorithm 1. Regarding computational complexity, preparing the data in Step 2 can be performed within $\mathcal{O}(M)$ operations. Exploiting the sparsity of $\mathbf{h}_m$'s, evaluating the coefficients $\{(b_{m,t}, \mathbf{a}_{m,t})\}_{m=1}^M$ in Step 5 can also be done with $\mathcal{O}(M)$ operations. The overall complexity of Algorithm 1 is indeed dominated by solving the quadratic program of (37) in Step 6. With standard convex programming solvers, the resultant complexity is often $\mathcal{O}(MN^2)$. Iterative procedures depending on the alternating direction method of multipliers can reduce this number to $\mathcal{O}(MN\log(1/\epsilon))$ [21], [29]. The latter complexity, however, may still become unfavorable for large-size power networks. Furthermore, even though $\{\mathbf{h}_m\}_{m=1}^M$ have at most two nonzero entries, this property cannot be fully exploited to speed up computations for solving the quadratic program of (37). To address these issues, we advocate a stochastic alternative of (37) ahead for solving problem (34).

B. Stochastic Prox-Linear SE Solver

The stochastic prox-linear method deals with a single measurement per iteration. Initialized with $\mathbf{x}_0$, our (stochastic) prox-linear SE solver operates by first sampling uniformly a loss function via randomly picking $m_t \in \{1, 2, \ldots, M\}$, and relies on minimizing its quadratically regularized "linearization" around $\mathbf{x}_t$ to yield $\mathbf{x}_{t+1}$ [30]; that is, define inductively for $t = 0, 1, 2, \ldots$ that

$$\mathbf{x}_{t+1} = \arg\min_{\mathbf{x}\in\mathbb{C}^N} \left| b_{m_t,t} - 2\Re\left(\mathbf{a}_{m_t,t}^{\mathcal{H}}\mathbf{x}\right)\right| + \frac{1}{2\mu_t}\|\mathbf{x}-\mathbf{x}_t\|_2^2 \tag{39}$$

where the coefficients $b_{m_t,t}$ and $\mathbf{a}_{m_t,t}$ are given in (38), with $b_{m_t,t} - 2\Re(\mathbf{a}_{m_t,t}^{\mathcal{H}}\mathbf{x})$ being the first-order Taylor's expansion of the m_tth error function $\check{z}_{m_t} - |\mathbf{h}_{m_t}^{\mathcal{H}}\mathbf{x}|^2$ around $\mathbf{x}_t$. Evidently, problem (39) is again a quadratic program too. Compared with the first quadratic program in (36), fortunately, the solution to (39) can be found in simple closed form.

To that end, we invoke an earlier result in [29, Prop. 3], which is included in Appendix B for completeness. Upon defining $\mathbf{w} := \mathbf{x} - \mathbf{x}_t$, $\mathbf{a} := \mathbf{a}_{m_t,t}$, and $b := b_{m_t,t} - 2\Re(\mathbf{a}_{m_t,t}^{\mathcal{H}}\mathbf{x}_t)$ in Proposition 2, one can readily find the solution to (39) as follows:

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \mathrm{proj}_{\mu_t}\left(\frac{b_{m_t,t} - 2\Re\left(\mathbf{a}_{m_t,t}^{\mathcal{H}}\mathbf{x}_t\right)}{\|\mathbf{a}_{m_t,t}\|_2^2}\right)\mathbf{a}_{m_t,t}.$$

Substituting $\mathbf{a}_{m_t}$ and b_{m_t} of (38) into the last equality leads to our stochastic prox-linear SE solver

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \mathrm{proj}_{\mu_t}\left(\frac{\check{z}_{m_t} - \left|\mathbf{h}_{m_t}^{\mathcal{H}}\mathbf{x}_t\right|^2}{4\left|\mathbf{h}_{m_t}^{\mathcal{H}}\mathbf{x}_t\right|^2\cdot\|\mathbf{h}_{m_t}\|_2^2}\right)\cdot 2\left(\mathbf{h}_{m_t}^{\mathcal{H}}\mathbf{x}_t\right)\mathbf{h}_{m_t}. \tag{40}$$

We summarize our second (stochastic) prox-linear SE solver in Algorithm 2 for further reference. In terms of computational complexity, we report the exact number of complex scalar operations (e.g., additions, multiplications) needed per stochastic prox-linear SE iteration of (40) next. Relying on whether $\mathbf{h}_{m_t}$ has 1 or 2 nonzero entries, the following statements hold true. If $\mathbf{h}_{m_t}$ has 1 (2) nonzero entries, evaluating $|\mathbf{h}_{m_t}^{\mathcal{H}}\mathbf{x}_t|^2$ requires 2 (4) operations, and $\|\mathbf{h}_{m_t}\|_2^2$ requires 1 (3) operations, plus another 5 operations for the remaining, all summing to a total of 8 (12) operations. In other words, per iteration of Algorithm 2 (cf. Steps 4–6) must perform only 12 complex scalar operations or so. Interestingly, this per-iteration complexity of $\mathcal{O}(1)$ holds regardless of the power network under investigation, or more precisely, the system size N. It is self-evident that this $\mathcal{O}(1)$ per-iteration complexity scales nicely to large- and even massive-size power networks.

Algorithm 2: Stochastic Prox-linear SE Solver.

1: **Input** data $\{(z_m, \mathbf{H}_m)\}_{m=1}^M$, step size $\mu > 0$, initialization $\mathbf{v}_0 \in \mathbb{C}^N$, solution accuracy $\epsilon > 0$, and set $t = 0$.
2: **Prepare** the power flow data $\{(z_m, \mathbf{H}_m)\}_{m=N+1}^M$ according to (1), (22)–(23), (25)–(26), (29)–(30), and (31)–(32) to obtain $\{(\check{z}_m, \mathbf{h}_m)\}_{m=N+1}^M$ based on $\{z_m = |v_m|^2\}_{m=1}^N$.
3: **Repeat**
4:     **Draw** $m_t \in \{1, 2, \ldots, M\}$ uniformly at random.
5:     **Evaluate** Evaluate $\mathbf{a}_{m_t,t}$ and $b_{m_t,t}$ in (38).
6:     **Update** $\mathbf{x}_{t+1}$ via (40).
7:     $t = t + 1$.
8: **Until** $\|\mathbf{x}_t - \mathbf{x}_{t-1}\|_2 \le \epsilon\sqrt{N}$.
9: **Return** $\mathbf{x}_t$.

V. Convergence Analysis and Exact Recovery

In this section, we begin our development by providing convergence guarantees for the proposed prox-linear SE solvers. For concreteness, we will focus on the deterministic prox-linear Algorithm 1, whereas convergence can be also established for Algorithm 2 in a probabilistic sense. Interested readers are referred to [30]. Under certain conditions on the loss function ℓ, exact recovery results are also established.

Recall our loss function $c(\mathbf{s}(\mathbf{x})) = \frac{1}{M}\|\check{\mathbf{z}} - |\mathbf{H}\mathbf{x}|^2\|_1$, where $c(\mathbf{u}) = \|\mathbf{u}\|_1 : \mathbb{R}^M \to \mathbb{R}$, and $\mathbf{s}(\mathbf{x}) = \frac{1}{M}(\check{\mathbf{z}} - |\mathbf{H}\mathbf{x}|^2) : \mathbb{C}^N \to \mathbb{R}^M$. It is easy to verify for any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^N$ that it holds

$$|c(\mathbf{u}) - c(\mathbf{v})| \le \sum_{n=1}^M |u_m - v_m| = \|\mathbf{u}-\mathbf{v}\|_1 \le \sqrt{M}\|\mathbf{u}-\mathbf{v}\|_2$$

where the last inequality arises from the equivalence of norms. By definition, this asserts that c is $\sqrt{M}$-Lipschitz continuous. Focusing now on the complex Jacobian $\nabla_{\mathrm{x}}\mathbf{s}$ for any $\mathbf{x}$ and $\mathbf{y} \in \mathbb{C}^N$, we deduce

$$\|\nabla_{\mathrm{x}}\mathbf{s}(\mathbf{x}) - \nabla_{\mathrm{x}}\mathbf{s}(\mathbf{y})\|_2 = \frac{1}{M}\left\|\mathbf{H}^{\mathcal{H}}\mathbf{H}(\mathbf{x}-\mathbf{y})\right\|_2 \le L\|\mathbf{x}-\mathbf{y}\|_2$$

with $L := (2/M)\lambda_{\max}(\mathbf{H}^{\mathcal{H}}\mathbf{H})$, which confirms that $\nabla_{\mathrm{x}}\mathbf{s}$ is L-Lipschitz continuous.

Appealing to the results in [26, Th. 5.3], one can conclude that our deterministic prox-linear SE solver with constant step size $\mu \le 1/(L\sqrt{M}) = \sqrt{M}/(2\lambda_{\max}(\mathbf{H}^{\mathcal{H}}\mathbf{H}))$ converges to a stationary point of $\ell(\mathbf{x})$ in (34).

We provide in the sequel conditions on the function ℓ such that exact recovery of power system states by our prox-linear SE solvers is guaranteed. Going beyond [21], which is limited to optimization over real-valued variables, we introduce two complimentary conditions on $\ell(\mathbf{x})$ and its linearization $\ell_{\mathbf{x}}(\mathbf{y})$ of complex-valued variables $\mathbf{x}, \mathbf{y} \in \mathbb{C}^N$.

If ℓ is defined over real-valued vectors, namely $\ell(\mathbf{x}): \mathbb{R}^N \to \mathbb{R}$, we can readily have the following stability condition for establishing fast convergence: for any given $\mathbf{v} \in \mathbb{R}^N$, there exists a constant $\rho > 0$ such that $\ell(\mathbf{x}) - \ell(\mathbf{v}) \geq \rho\|\mathbf{x}-\mathbf{v}\|_2\,\|\mathbf{x}+\mathbf{v}\|_2$ holds for all $\mathbf{x} \in \mathbb{R}^N$, which is included in Appendix B for completeness. This condition, though crucial for establishing fast convergence, does not generalize to functions of complex-valued variables. It is obvious in the complex case that if $\mathbf{v} \in \mathbb{C}^N$ is an optimal solution to (34), then $e^{j\phi}\mathbf{v}$ with any $\phi \in [0,\, 2\pi)$ is also an optimal solution. This ambiguity thus prompts us to define the Euclidean distance of any estimate $\mathbf{x} \in \mathbb{C}^N$ to $\mathbf{v} \in \mathbb{C}^N$ as $\mathrm{dist}(\mathbf{x},\, \mathbf{v}) := \min_{\phi\in[0,2\pi)} \|\mathbf{x} - e^{j\phi}\mathbf{v}\|_2$. This definition can be thought of as enforcing zero-phase angle at the reference bus, a standard procedure in power flow problems to eliminate the phase ambiguity. Invoking a result for stable phase retrieval in [31, Th. 3.1], we define the following stability condition for functions $\ell(\mathbf{x}): \mathbb{C}^N \to \mathbb{R}$, which also applies to practical settings where the measurements may contain noise or be even adversarially corrupted.

Condition 1: There exists some constant $\rho > 0$ such that the inequality holds for all $\mathbf{x} \in \mathbb{C}^N$

$$\ell(\mathbf{x}) - \ell(\mathbf{v}) \geq \rho\sqrt{\|\mathbf{x}-\mathbf{v}\|_2^2\,\|\mathbf{x}+\mathbf{v}\|_2^2 - 4|\Im(\mathbf{x}^{\mathcal{H}}\mathbf{v})|^2}. \quad (41)$$

Evidently, when the measurements $\check{\mathbf{z}} = |\mathbf{H}\mathbf{v}|^2$ are noiseless, it holds that $\ell(\mathbf{v}) = 0$. Similar to the real-valued case studied in [21], our Condition 1 is instrumental in establishing fast convergence of the prox-linear algorithm for optimizing functions of complex-valued variables. Besides Condition 1, we require a condition on the linearization $\ell_{\mathbf{x}}(\mathbf{y}): \mathbb{C}^N \to \mathbb{R}$ of $\ell(\mathbf{x}) = c(\mathbf{s}(\mathbf{x}))$ around $\mathbf{x}$ defined by

$$\ell_{\mathbf{x}}(\mathbf{y}) := c\big(\mathbf{s}(\mathbf{x}) + 2\Re\big(\nabla_{\mathbf{x}}^{\mathcal{H}}\mathbf{s}(\mathbf{x})(\mathbf{y}-\mathbf{x})\big)\big). \quad (42)$$

Condition 2: There exists a constant $L < +\infty$ such that the inequality holds for all $\mathbf{x},\, \mathbf{y} \in \mathbb{C}^N$

$$|\ell(\mathbf{y}) - \ell_{\mathbf{x}}(\mathbf{y})| \leq \frac{L}{2}\|\mathbf{x}-\mathbf{y}\|_2^2. \quad (43)$$

This condition basically requires that the locally linearized convex approximation $\ell_{\mathbf{x}}(\mathbf{y})$ is quadratically close [cf. (43)] to the nonconvex function $\ell(\mathbf{y})$. Indeed, the ℓ_1-based PSSE cost function ℓ in (34) automatically satisfies Condition 2 globally. To show this, let us first express $\ell_{\mathbf{x}}(\mathbf{y})$ according to the definition of (42)

$$\ell_{\mathbf{x}}(\mathbf{y}) = \frac{1}{M}\sum_{m=1}^{M}\Big||\mathbf{h}_m^{\mathcal{H}}\mathbf{x}|^2 - \check{z}_m + 2\Re\big(\mathbf{x}^{\mathcal{H}}\mathbf{h}_m\mathbf{h}_m^{\mathcal{H}}(\mathbf{y}-\mathbf{x})\big)\Big|.$$

On the other hand, for any $m \in \{1,\, 2,\, \ldots,\, M\}$ and $\mathbf{y} \in \mathbb{C}^N$, the following holds true

$$\big|\mathbf{h}_m^{\mathcal{H}}\mathbf{y}\big|^2 = \big|\mathbf{h}_m^{\mathcal{H}}\mathbf{x}\big|^2 + 2\Re\big(\mathbf{x}^{\mathcal{H}}\mathbf{h}_m\mathbf{h}_m^{\mathcal{H}}(\mathbf{y}-\mathbf{x})\big) + \big|\mathbf{h}_m^{\mathcal{H}}(\mathbf{y}-\mathbf{x})\big|^2.$$

Subtracting $\check{z}_m$ from both sides, summing from $m = 1$ to M, and leveraging the triangle inequality, we have that

$$\ell(\mathbf{y}) = \frac{1}{M}\sum_{m=1}^{M}\big||\mathbf{h}_m^{\mathcal{H}}\mathbf{y}|^2 - \check{z}_m\big| \leq \ell_{\mathbf{x}}(\mathbf{y}) + \frac{1}{M}\sum_{m=1}^{M}\big|\mathbf{h}_m^{\mathcal{H}}(\mathbf{y}-\mathbf{x})\big|^2$$

$$\ell(\mathbf{y}) = \frac{1}{M}\sum_{m=1}^{M}\big||\mathbf{h}_m^{\mathcal{H}}\mathbf{y}|^2 - \check{z}_m\big| \geq \ell_{\mathbf{x}}(\mathbf{y}) - \frac{1}{M}\sum_{m=1}^{M}\big|\mathbf{h}_m^{\mathcal{H}}(\mathbf{y}-\mathbf{x})\big|^2.$$

Rewriting the last terms in matrix-vector form yields

$$|\ell(\mathbf{y}) - \ell_{\mathbf{x}}(\mathbf{y})| \leq (\mathbf{y}-\mathbf{x})^{\mathcal{H}}\Big(\frac{1}{M}\mathbf{H}^{\mathcal{H}}\mathbf{H}\Big)(\mathbf{y}-\mathbf{x}) \leq \frac{L}{2}\|\mathbf{y}-\mathbf{x}\|_2^2 \quad (44)$$

which proves that Condition 2 is satisfied globally by the ℓ_1-loss in (34). Moreover, one can take $L = \lambda_{\max}(\mathbf{H}^{\mathcal{H}}\mathbf{H}/M)$, namely the largest eigenvalue of matrix $\mathbf{H}^{\mathcal{H}}\mathbf{H}/M$.

Under Conditions 1 and 2, we now devote to exact recovery guarantees for the deterministic prox-linear SE solver in Algorithm 1. The following result implies exact recovery of power system states at a quadratic rate by our proposed prox-linear SE solver in Algorithm 1 under suitable conditions.

Theorem 1: Let Conditions 1 and 2 hold. Assuming that the quadratic program (37) is solved exactly per iteration of Algorithm 1, the successive prox-linear SE iterates $\mathbf{x}_t$ satisfy

$$\frac{\mathrm{dist}(\mathbf{x}_t,\mathbf{v})}{\|\mathbf{v}\|_2} \leq \frac{\rho}{L}\left(\frac{L}{\rho}\cdot\frac{\mathrm{dist}(\mathbf{x}_0,\mathbf{v})}{\|\mathbf{v}\|_2}\right)^{2^t}. \quad (45)$$

For readability, the proof of Theorem 1 is postponed to Appendix C. Regarding Theorem 1, three observations come in order.

Remark 1 (Exact recovery): If the initialization $\mathbf{x}_0$ of the iterations is accurate enough, meaning that it satisfies the condition $\mathrm{dist}(\mathbf{x}_0,\, \mathbf{v}) < (\rho/L)\|\mathbf{v}\|_2$, the prox-linear SE solver recovers exactly the true state vector $\mathbf{v} \in \mathbb{C}^N$. In terms of initializations, there are several approaches for this desideratum, three of which are discussed next. Since power systems are typically operating close to the flat voltage profile $\mathbf{1}$, it is reasonable to initialize the algorithm with $\mathbf{x}_0 = \mathbf{1}$. Moreover, as the voltage magnitudes at all buses are assumed available, one can use the voltage magnitude vector $\mathbf{x}_0 = |\mathbf{v}|$ as the initializer. Alternatively, it is also feasible to initialize with the estimate found by solving the linearized dc power flow equations.

Remark 2 (Quadratic convergence rate): When $\mathrm{dist}(\mathbf{x}_0,\, \mathbf{v}) < (\rho/L)\|\mathbf{v}\|_2$ holds true, our prox-linear SE algorithm converges quadratically fast to the globally optimal solution of the nonconvex and nonsmooth optimization problem (34). Expressed differently, to obtain a solution $\mathbf{x}_t$ of (at most) ϵ-relative error, namely $\mathrm{dist}(\mathbf{x}_t,\, \mathbf{v})/\|\mathbf{v}\|_2 \leq \epsilon$, we must only run Algorithm 1 for about $\log_2\log_2(1/\epsilon)$ iterations, or equivalently, solve $\log_2\log_2(1/\epsilon)$ convex quadratic programs as in (37). This, in practice, amounts to about $5 \sim 10$ such quadratic programs.

Remark 3: Under the condition $\mathrm{dist}(\mathbf{x}_0,\, \mathbf{v}) < (\rho/L)\|\mathbf{v}\|_2$, it is worth pointing out that Condition 1 can be replaced by a condition requiring only the function $\ell(\mathbf{x})$ to satisfy the inequality (49) locally for all $\mathbf{x}$ within the neighborhood of $\mathbf{v}$ defined by $\mathrm{dist}(\mathbf{x},\, \mathbf{v}) < (\rho/L)\|\mathbf{v}\|_2$.

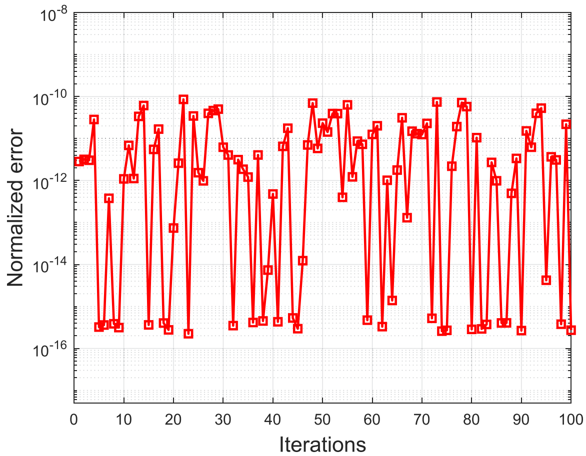

Fig. 1. Exact recovery performance of Algorithm 1 for the IEEE 14-bus system (noiseless case).

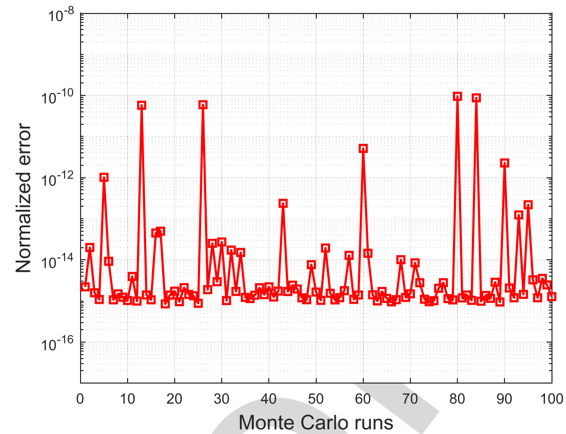

Fig. 2. Exact recovery performance of Algorithm 1 for the IEEE 118-bus system (noiseless case).

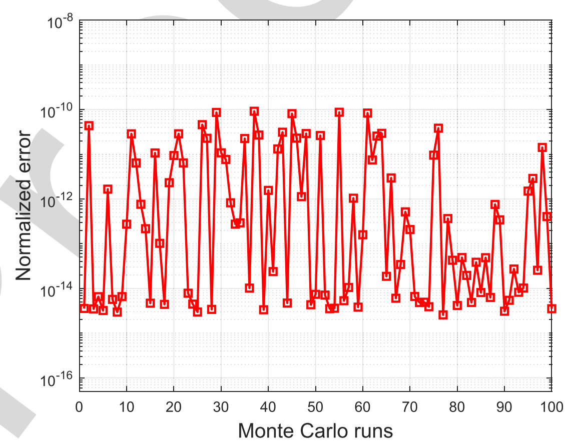

Fig. 3. Exact recovery performance of Algorithm 1 for the IEEE 300-bus system (noiseless case).

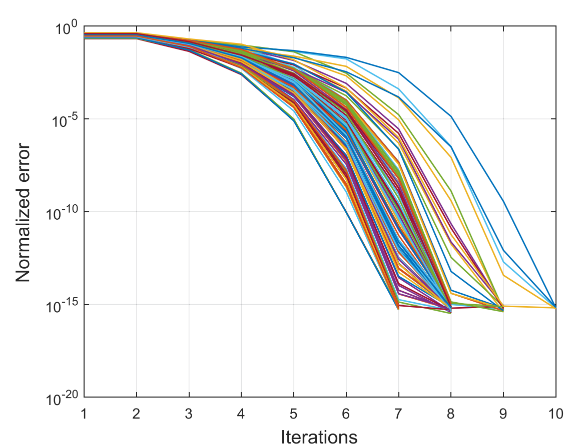

Fig. 4. Quadratic convergence of Algorithm 1 in 100 runs for the IEEE 14-bus system (noiseless case).

VI. Numerical Tests

In this section, we perform a number of numerical tests to evaluate our approach and compare with the “workhorse” LS-based Gauss–Newton method. Several power network benchmarks including the IEEE 14-, 118-, and 300-bus systems were simulated, following the MATLAB-based toolbox MATPOWER [32], [33]. The Gauss–Newton iterations were implemented by using the embedded SE function “doSE.m” in MATPOWER. To carefully isolate the relative performance of the iterative algorithms, rather than initialization employed, all simulated schemes were initialized with the flat voltage vector (i.e., the all-one vector) in all reported experiments.

A. Tests With Zero Noise

The first experiment examines the exact recovery and convergence performance of Algorithm 1 from noiseless data on the IEEE 14-, 118-, and 300-bus test systems. The actual voltage magnitude (in p.u.) and angle (in radians) of each bus were uniformly distributed over $[0.9, 1.1]$, and over $[-0.1\pi, 0.1\pi]$. The voltage magnitude squares at all buses as well as the active power flows across all lines were measured. The quadratic programs in Step 5 of Algorithm 1 were solved by the standard convex programming solver SeDuMi [34] with a constant step size of $\mu_t = 1,000$. Algorithm 1 terminates either when a maximum number 20 of iterations are simulated, or when the normalized distance between two consecutive iterates becomes smaller than 10^{-10}, namely $\text{dist}(\mathbf{x}_t, \mathbf{x}_{t-1})/\sqrt{N} \le 10^{-10}$. A total of 100 Monte Carlo (MC) runs were carried out. Figs. 1–3 plot the normalized estimation errors $\text{dist}(\mathbf{x}_t, \mathbf{v})/\sqrt{N}$ for the 100 MC realizations on the simulated three systems, whose corresponding L values are 0.9980, 3.0201, and 6.3102. Furthermore, Fig. 4 depicts the convergence of the normalized estimation error of Algorithm 1 for the 100 runs on the 14-bus system. Evidently, Algorithm 1 achieves exact recovery over the 100 runs, and enjoys quadratic convergence in this noiseless setting, validating our theoretical findings in Theorem 1.

B. Tests With Outlying Measurements

One of the claimed advantages of the ℓ_1-based loss function in (34) is its robustness to outliers. In this test, we evaluate the robustness of Algorithm 1 to measurements with outliers

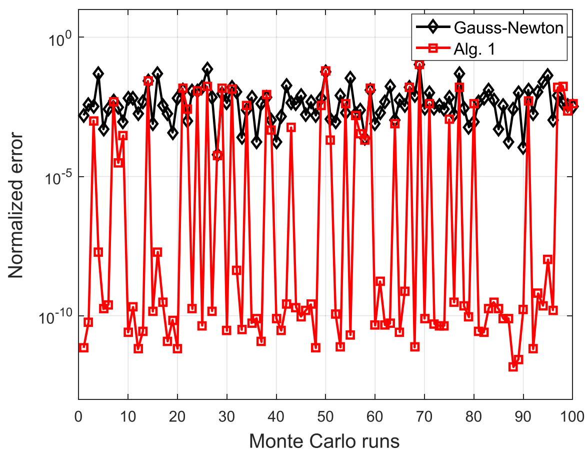

Fig. 5. Recovery performance for the IEEE 118-bus system with 1 outlying measurement (noiseless case).

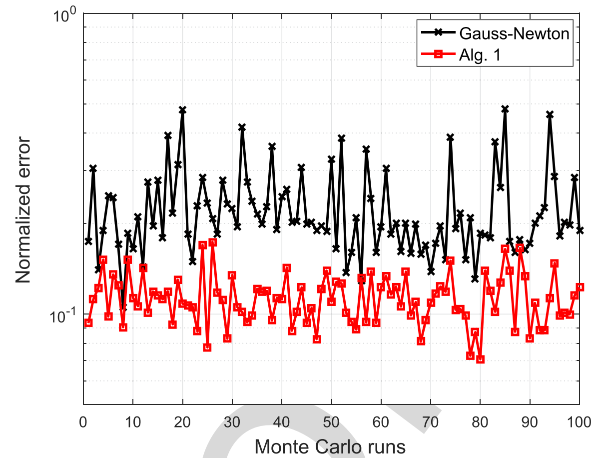

Fig. 6. Estimation performance for the IEEE 300-bus system under additive noise and 5% outlying measurements.

in terms of the exact recovery. Concretely, the IEEE 118-bus system with its default voltage profile was simulated. The voltage magnitudes at all buses along with the active and reactive power flows across all lines were measured. Considering the fact that the nodal voltage magnitudes in transmission networks are maintained close to one, we assume that only the power meters can be compromised. A total of 100 MC runs were performed. Per run, one power flow meter was randomly compromised, whose measurement was purposefully manipulated and amplified to five times its original value.

Both the Gauss–Newton and Algorithm 1 were simulated in this experiment, whose corresponding normalized estimation errors for the 100 MC realizations are presented in Fig. 5. It is self-evident from the plots that the Gauss–Newton method is not robust to outlying measurements, whereas our proposed prox-linear scheme in Algorithm 1 can identify and automatically reject the bad data, yielding exact recovery of the true states in most cases even under adversarial attacks.

C. Tests With Additive Noise and Outliers

The third experiment assesses the robust estimation performance of Algorithm 1 relative to Gauss–Newton, in a setting where both additive noises and outliers are present. The large IEEE 300-bus benchmark system with its default voltage profile was simulated. All active and reactive power flows as well as all voltage magnitudes were measured. Additive noise was independently generated from normal distributions having zero-mean and standard deviations 0.004 and 0.008 for the voltage magnitude and line flow measurements, respectively [18]. In addition to additive noise, 5% of the entire measured power flows were corrupted uniformly at random with “outliers” drawn independently from a Gaussian distribution with zero-mean and standard deviation 5. The normalized estimation errors obtained by the Gauss–Newton method and Algorithm 1 for 100 MC independent realizations are reported in Fig. 6. Evidently, our developed algorithm consistently exhibits more robust estimation performance than LS-based Gauss–Newton against additive noise and outlying measurements.

VII. Conclusion

Robust PSSE is approached by minimizing the ℓ_1-based loss function from the vantage point of composite optimization. To enable efficient algorithms and exact state recovery, the power quantities were first transformed into rank-one measurements. Building on advances in nonconvex and nonsmooth composite optimization, two algorithms were put forth for minimizing the ℓ_1-based loss of the transformed rank-one measurements. Our algorithms require no tuning of parameters, except for a step size. We also developed “stability conditions” on the ℓ_1-based loss function such that exact state recovery and quadratic convergence are guaranteed by our approach in the noiseless case. Simulated tests using three IEEE benchmark networks under different settings validate our theoretical findings, and showcase the efficacy of our approach.

Appendix

A. Wirtinger's Calculus

Introducing the complex conjugate coordinates $[\mathbf{x}^{\mathcal{T}}\ \overline{\mathbf{x}}^{\mathcal{T}}]^{\mathcal{T}} \in \mathbb{C}^{2N}$, one can rewrite $\mathbf{s}(\mathbf{x}) = \mathbf{s}(\mathbf{x}, \overline{\mathbf{x}}) \in \mathbb{C}^{M}$. It is obvious now that $\mathbf{s}(\mathbf{x}, \overline{\mathbf{x}})$ becomes holomorphic in $\mathbf{x}$ for a fixed $\overline{\mathbf{x}}$, and vice versa. This leads to the partial Wirtinger derivatives [28]

$$\frac{\partial s_m}{\partial \mathbf{x}} := \left.\frac{\partial s_m(\mathbf{x}, \overline{\mathbf{x}})}{\partial \mathbf{x}}\right|_{\overline{\mathbf{x}}=\text{constant}} = \left[\frac{\partial s_m}{\partial x_1}\ \frac{\partial s_m}{\partial x_2}\ \cdots\ \frac{\partial s_m}{\partial x_N}\right]$$

$$\frac{\partial s_m}{\partial \overline{\mathbf{x}}} := \left.\frac{\partial s_m(\mathbf{x}, \overline{\mathbf{x}})}{\partial \overline{\mathbf{x}}}\right|_{\mathbf{x}=\text{constant}} = \left[\frac{\partial s_m}{\partial \overline{x}_1}\ \frac{\partial s_m}{\partial \overline{x}_2}\ \cdots\ \frac{\partial s_m}{\partial \overline{x}_N}\right]$$

where the partial derivative with respect to $\mathbf{x}$ ($\overline{\mathbf{x}}$) treats $\overline{\mathbf{x}}$ ($\mathbf{x}$) as a constant in s_m. The complex gradient of $s_m(\mathbf{x}, \overline{\mathbf{x}})$ with respect to $\mathbf{x}$ or $\overline{\mathbf{x}}$ can be defined by

$$\nabla_{\mathbf{x}} s_m := \left(\frac{\partial s_m}{\partial \mathbf{x}}\right)^{\mathcal{H}}, \ \text{ and }\ \nabla_{\overline{\mathbf{x}}} s_m := \left(\frac{\partial s_m}{\partial \overline{\mathbf{x}}}\right)^{\mathcal{H}}$$

yielding the complex gradient of s_m in new coordinate system

$$\nabla_c s_m := \left[\nabla_{\mathbf{x}}^{\mathcal{T}} s_m \; \nabla_{\overline{\mathbf{x}}}^{\mathcal{T}} s_m\right]^{\mathcal{T}} = \left[\frac{\partial s_m}{\partial \mathbf{x}} \; \frac{\partial s_m}{\partial \overline{\mathbf{x}}}\right]^{\mathcal{H}}.$$

Upon introducing the complex Jacobian

$$\nabla_c \mathbf{s} := [\nabla_c s_1 \; \nabla_c s_2 \; \cdots \; \nabla_c s_M] \in \mathbb{C}^{2N \times M}$$

we can define for given vectors $\mathbf{x}$ and $\Delta\mathbf{x} \in \mathbb{C}^N$ the following first-order Taylor's expansion:

$$\begin{aligned}\mathbf{s}(\mathbf{x}+\Delta\mathbf{x}) &\approx \mathbf{s}(\mathbf{x}) + \nabla_c^{\mathcal{H}}\mathbf{s}(\mathbf{x})\begin{bmatrix}\Delta\mathbf{x}\\ \overline{\Delta\mathbf{x}}\end{bmatrix}\\ &= \mathbf{s}(\mathbf{x}) + 2\Re\big(\nabla^{\mathcal{H}}\mathbf{s}_{\mathbf{x}}(\mathbf{x})\Delta\mathbf{x}\big) \in \mathbb{R}^{M\times N}. \end{aligned} \quad (46)$$

B. Supporting Results

Proposition 2 ([29, Prop. 3]): Given $\mathbf{a} \in \mathbb{C}^N$ and $b \in \mathbb{R}$, the solution of

$$\underset{\mathbf{w}\in\mathbb{C}^N}{\text{minimize}} \; \left|b - \Re\left(\mathbf{a}^{\mathcal{H}}\mathbf{w}\right)\right| + \frac{1}{2\mu}\|\mathbf{w}\|_2^2 \quad (47)$$

can be obtained as $\hat{\mathbf{w}} := \text{proj}_\mu(b/\|\mathbf{a}\|_2^2)\cdot\mathbf{a}$, where $\text{proj}_\mu(x) : \mathbb{R}\times\mathbb{R}_+ \to \mathbb{R}$ is the projection operator that returns the real number in interval $[-\tau, \tau]$ closest to any given $x \in \mathbb{R}$.

Condition 3 [21, Condition 1]: For any given $\mathbf{v} \in \mathbb{R}^N$, there exists a parameter $\rho > 0$ such that function $\ell(\mathbf{x})$ satisfies the following for all $\mathbf{x} \in \mathbb{R}^n$:

$$\ell(\mathbf{x}) - \ell(\mathbf{v}) \geq \rho\|\mathbf{x}-\mathbf{v}\|_2\|\mathbf{x}+\mathbf{v}\|_2. \quad (48)$$

Concerning the lower bound in (41), we have the next result.

Lemma 1: For any fixed $\mathbf{v} \in \mathbb{C}^N$, the inequality holds for all $\mathbf{x}$, $\mathbf{v} \in \mathbb{C}^N$

$$\|\mathbf{x}-\mathbf{v}\|_2^2\,\|\mathbf{x}+\mathbf{v}\|_2^2 - 4\left|\Im(\mathbf{x}^{\mathcal{H}}\mathbf{v})\right|^2 \geq \|\mathbf{v}\|_2^2\,\text{dist}^2(\mathbf{x},\mathbf{v}). \quad (49)$$

Proof: The left-hand-side term of (49) can be rewritten as

$$\begin{aligned}&\|\mathbf{x}-\mathbf{v}\|_2^2\,\|\mathbf{x}+\mathbf{v}\|_2^2 - 4\Im^2(\mathbf{x}^{\mathcal{H}}\mathbf{v})\\ &= \left(\|\mathbf{x}\|_2^2 + \|\mathbf{v}\|_2^2 - 2\Re(\mathbf{x}^{\mathcal{H}}\mathbf{v})\right)\left(\|\mathbf{x}\|_2^2 + \|\mathbf{v}\|_2^2 + 2\Re(\mathbf{x}^{\mathcal{H}}\mathbf{v})\right)\\ &\quad - 4\Im^2(\mathbf{x}^{\mathcal{H}}\mathbf{v})\\ &= \left(\|\mathbf{x}\|_2^2 + \|\mathbf{v}\|_2^2\right)^2 - 4\left[\Re^2(\mathbf{x}^{\mathcal{H}}\mathbf{v}) + \Im^2(\mathbf{x}^{\mathcal{H}}\mathbf{v})\right]\\ &= \left(\|\mathbf{x}\|_2^2 + \|\mathbf{v}\|_2^2\right)^2 - 4|\mathbf{x}^{\mathcal{H}}\mathbf{v}|^2\\ &= \left(\|\mathbf{x}\|_2^2 + \|\mathbf{v}\|_2^2 + 2|\mathbf{x}^{\mathcal{H}}\mathbf{v}|\right)\left(\|\mathbf{x}\|_2^2 + \|\mathbf{v}\|_2^2 - 2|\mathbf{x}^{\mathcal{H}}\mathbf{v}|\right)\\ &\geq \|\mathbf{v}\|_2^2\left(\|\mathbf{x}\|_2^2 + \|\mathbf{v}\|_2^2 - 2|\mathbf{x}^{\mathcal{H}}\mathbf{v}|\right)\\ &= \|\mathbf{v}\|_2^2\,\text{dist}^2(\mathbf{x},\mathbf{v}).\end{aligned}$$

Taking the square root from both sides of the inequality yields the statement of Lemma 1. ■

C. Proof of Theorem 1

The proof is based on that of [21, Th. 1], but we here generalize its results to function optimization over complex domains. Observe that the regularized function $g(\mathbf{x}) := \ell_{\mathbf{x}_t}(\mathbf{x}) + \frac{L}{2}\|\mathbf{x}-\mathbf{x}_t\|_2^2$ is L-strongly convex in $\mathbf{x} \in \mathbb{C}^N$, and its minimum is attained at $\mathbf{x}_{t+1}$ [cf. (37)]. The standard optimality conditions for strongly convex minimization confirms that $g(\mathbf{x}_{t+1}) \leq g(\mathrm{e}^{j\angle\mathbf{x}_t^{\mathcal{H}}\mathbf{v}}\mathbf{v}) - \frac{L}{2}\|\mathrm{e}^{j\angle\mathbf{x}_t^{\mathcal{H}}\mathbf{v}}\mathbf{v} - \mathbf{x}_{t+1}\|_2^2$; that is,

$$\begin{aligned}&\ell_{\mathbf{x}_t}(\mathbf{x}_{t+1}) + \frac{L}{2}\|\mathbf{x}_{t+1}-\mathbf{x}_t\|_2^2 \leq \ell_{\mathbf{x}_t}(\mathrm{e}^{j\angle\mathbf{x}_t^{\mathcal{H}}\mathbf{v}}\mathbf{v})\\ &\quad + \frac{L}{2}\left\|\mathrm{e}^{j\angle\mathbf{x}_t^{\mathcal{H}}\mathbf{v}}\mathbf{v} - \mathbf{x}_t\right\|_2^2 - \frac{L}{2}\left\|\mathrm{e}^{j\angle\mathbf{x}_t^{\mathcal{H}}\mathbf{v}}\mathbf{v} - \mathbf{x}_{t+1}\right\|_2^2. \quad (50)\end{aligned}$$

Recalling now Condition 2, we have that

$$\ell(\mathbf{x}_{t+1}) \leq \ell_{\mathbf{x}_t}(\mathbf{x}_{t+1}) + \frac{L}{2}\|\mathbf{x}_{t+1}-\mathbf{x}_t\|_2^2 \quad (51a)$$

$$\ell_{\mathbf{x}_t}(\mathrm{e}^{j\angle\mathbf{x}_t^{\mathcal{H}}\mathbf{v}}\mathbf{v}) \leq \ell(\mathrm{e}^{j\angle\mathbf{x}_t^{\mathcal{H}}\mathbf{v}}\mathbf{v}) + \frac{L}{2}\left\|\mathrm{e}^{j\angle\mathbf{x}_t^{\mathcal{H}}\mathbf{v}}\mathbf{v} - \mathbf{x}_t\right\|_2^2. \quad (51b)$$

Substituting (51a) and (51b) into (50) gives rise to

$$\begin{aligned}\ell(\mathbf{x}_{t+1}) &\leq \ell_{\mathbf{x}_t}(\mathrm{e}^{j\angle\mathbf{x}_t^{\mathcal{H}}\mathbf{v}}\mathbf{v}) + \frac{L}{2}\left\|\mathrm{e}^{j\angle\mathbf{x}_t^{\mathcal{H}}\mathbf{v}}\mathbf{v} - \mathbf{x}_t\right\|_2^2\\ &\quad - \frac{L}{2}\left\|\mathrm{e}^{j\angle\mathbf{x}_t^{\mathcal{H}}\mathbf{v}}\mathbf{v} - \mathbf{x}_{t+1}\right\|_2^2\\ &\leq \ell(\mathrm{e}^{j\angle\mathbf{x}_t^{\mathcal{H}}\mathbf{v}}\mathbf{v}) + L\left\|\mathrm{e}^{j\angle\mathbf{x}_t^{\mathcal{H}}\mathbf{v}}\mathbf{v} - \mathbf{x}_t\right\|_2^2 - \frac{L}{2}\left\|\mathrm{e}^{j\angle\mathbf{x}_t^{\mathcal{H}}\mathbf{v}}\mathbf{v} - \mathbf{x}_{t+1}\right\|_2^2\end{aligned}$$

which, in conjunction with $\ell(\mathrm{e}^{j\angle\mathbf{x}_t^{\mathcal{H}}\mathbf{v}}\mathbf{v}) = \ell(\mathbf{v})$ in (34), yields

$$\begin{aligned}L\left\|\mathrm{e}^{j\angle\mathbf{x}_t^{\mathcal{H}}\mathbf{v}}\mathbf{v} - \mathbf{x}_t\right\|_2^2 &\geq \ell(\mathbf{x}_{t+1}) - \ell(\mathbf{v})\\ &\quad + \frac{L}{2}\left\|\mathrm{e}^{j\angle\mathbf{x}_t^{\mathcal{H}}\mathbf{v}}\mathbf{v} - \mathbf{x}_{t+1}\right\|_2^2. \quad (52)\end{aligned}$$

Invoking further Condition 1 in (52), we have that

$$\begin{aligned}L\left\|\mathrm{e}^{j\angle\mathbf{x}_t^{\mathcal{H}}\mathbf{v}}\mathbf{v} - \mathbf{x}_t\right\|_2^2 &\geq \frac{L}{2}\left\|\mathrm{e}^{j\angle\mathbf{x}_t^{\mathcal{H}}\mathbf{v}}\mathbf{v} - \mathbf{x}_{t+1}\right\|_2^2\\ &\quad + \rho\sqrt{\|\mathbf{x}_{t+1}-\mathbf{v}\|_2^2\,\|\mathbf{x}_{t+1}+\mathbf{v}\|_2^2 - 4\left|\Im(\mathbf{x}_{t+1}^{\mathcal{H}}\mathbf{v})\right|^2} \quad (53)\end{aligned}$$

in which the last term can be replaced with its lower bound $\rho\|\mathbf{v}\|_2\,\text{dist}(\mathbf{x}_{t+1},\mathbf{v})$ established in Lemma 1. Upon dropping the nonnegative term $\frac{L}{2}\|\mathrm{e}^{j\angle\mathbf{x}_t^{\mathcal{H}}\mathbf{v}}\mathbf{v} - \mathbf{x}_{t+1}\|_2^2$, and recalling the definition of $\text{dist}(\mathbf{x}_t,\mathbf{v})$, we immediately have

$$\rho\|\mathbf{v}\|_2\,\text{dist}(\mathbf{x}_{t+1},\mathbf{v}) \leq L\,\text{dist}^2(\mathbf{x}_t,\mathbf{v})$$

dividing both sides of which by $\rho\|\mathbf{v}\|_2^2$ yields

$$\frac{\text{dist}(\mathbf{x}_{t+1},\mathbf{v})}{\|\mathbf{v}\|_2} \leq \frac{L}{\rho}\cdot\frac{\text{dist}^2(\mathbf{x}_t,\mathbf{v})}{\|\mathbf{v}\|_2^2} = \frac{\rho}{L}\left(\frac{L}{\rho}\cdot\frac{\text{dist}(\mathbf{x}_t,\mathbf{v})}{\|\mathbf{v}\|_2}\right)^2.$$

Applying the above-mentioned inequality successively from the initialization $\mathbf{x}_0$ for t iterations through $\mathbf{x}_t$ gives rise to (45), concluding the proof.

References

[1] W. A. Wulf, "Great achievements and grand challenges," *Bridge*, vol. 30, no. 3/4, pp. 5–10, Fall 2010. [Online]. Available: http://www.greatachievements.org/

[2] A. Abur and A. Gómez-Expósito, *Power System State Estimation: Theory and Implementation*. New York, NY, USA: Marcel Dekker, 2004.

[3] G. Wang, G. B. Giannakis, J. Chen, and J. Sun, "Distribution system state estimation: An overview of recent developments," *Front. Inf. Technol. Electron. Eng.*, vol. 20, no. 1, pp. 1–14, Jan. 2019.

[4] F. C. Schweppe, J. Wildes, and D. Rom, "Power system static state estimation: Parts I, II, and III," *IEEE Trans. Power App. Syst.*, vol. PAS-89, no. 1, pp. 120–135, Jan. 1970.

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

[6] H. Zhu and G. B. Giannakis, "Power system nonlinear state estimation using distributed semidefinite programming," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 6, pp. 1039–1050, Dec. 2014.

[7] Y. Zhang, R. Madani, and J. Lavaei, "Conic relaxations for power system state estimation with line measurements," *IEEE Trans. Control Netw. Syst.*, vol. 5, no. 3, pp. 1193–1205, Sep. 2018.

[8] S. Bhela, V. Kekatos, and S. Veeramachaneni, "Enhancing observability in distribution grids using smart meter data," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 5953–5961, Nov. 2018.

[9] *Mission Support Center*, "Cyber threat and vulnerability analysis of the U.S. electric sector," Idaho Nat. Lab., Idaho Falls, ID, USA, Tech. Rep., 2016. Q4

[10] H. M. Merrill and F. C. Schweppe, "Bad data suppression in power system static state estimation," *IEEE Trans. Power App. Syst.*, vol. PAS-90, no. 6, pp. 2718–2725, Nov. 1971.

[11] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Trans. Inf. Syst. Secur.*, vol. 14, no. 1, May 2011, Art. no. 13.

[12] S. Soltan, M. Yannakakis, and G. Zussman, "Power grid state estimation following a joint cyber and physical attack," *IEEE Trans. Control Netw. Syst.*, vol. 5, no. 1, pp. 499–512, Mar. 2018.

[13] H. G. Aghamolki, Z. Miao, and L. Fan, "SOCP convex relaxation-based simultaneous state estimation and bad data identification," 2018, arXiv:1804.05130. Q5

[14] S. Akhlaghi, N. Zhou, and Z. Huang, "A multi-step adaptive interpolation approach to mitigating the impact of nonlinearity on dynamic state estimation," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3102–3111, Jul. 2018.

[15] L. Mili, M. Cheniae, N. Vichare, and P. Rousseeuw, "Robustification of the least absolute value estimator by means of projection statistics [power system state estimation]," *IEEE Trans. Power Syst.*, vol. 11, no. 1, pp. 216–225, Feb. 1996.

[16] P. J. Huber, "Robust Statistics," in *International Encyclopedia of Statistical Science*. New York, NY, USA: Springer, 2011, pp. 1248–1251.

[17] W. W. Kotiuga and M. Vidyasagar, "Bad data rejection properties of weighted least-absolute-value techniques applied to static state estimation," *IEEE Trans. Power App. Syst.*, vol. PAS-101, no. 4, pp. 844–853, Apr. 1982.

[18] A. Abur and M. K. Celik, "A fast algorithm for the weighted least-absolute-value state estimation (for power systems)," *IEEE Trans. Power Syst.*, vol. 6, no. 1, pp. 1–8, Feb. 1991.

[19] R. Jabr and B. Pal, "Iteratively re-weighted least-absolute-value method for state estimation," *IET Gener., Transmiss., Distrib.*, vol. 150, no. 4, pp. 385–391, Jul. 2003.

[20] G. Wang, G. B. Giannakis, and Y. C. Eldar, "Solving systems of random quadratic equations via truncated amplitude flow," *IEEE Trans. Inf. Theory*, vol. 64, no. 2, pp. 773–794, Feb. 2018.

[21] J. C. Duchi and F. Ruan, "Solving (most) of a set of quadratic equalities: Composite optimization for robust phase retrieval," 2018, arXiv:1705.02356.

[22] P. M. Pardalos and S. A. Vavasis, "Quadratic programming with one negative eigenvalue is NP-hard," *J. Global Optim.*, vol. 1, no. 1, pp. 15–22, 1991.

[23] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*, vol. 2. Philadelphia, PA, USA: SIAM, 2001.

[24] A. Conca, D. Edidin, M. Hering, and C. Vinzant, "An algebraic characterization of injectivity in phase retrieval," *Appl. Comput. Harmon. Anal.*, vol. 38, no. 2, pp. 346–356, Mar. 2015.

[25] J. V. Burke and M. C. Ferris, "A Gauss-Newton method for convex composite optimization," *Math. Program.*, vol. 71, no. 2, pp. 179–194, Dec. 1995.

[26] A. S. Lewis and S. J. Wright, "A proximal method for composite minimization," *Math. Program.*, vol. 158, no. 1/2, pp. 501–546, Jul. 2016.

[27] A. I. Markushevich, *Theory of Functions of A Complex Variable*, 2nd ed. New York, NY, USA: Chelsea, 1977.

[28] D. H. Brandwood, "A complex gradient operator and its application in adaptive array theory," *IEE ~~Proc. F~~ Commun., Radar Signal Process.*, vol. 130, no. 1, pp. 11–16, Feb. 1983.

[29] G. Wang, G. B. Giannakis, and J. Chen, "Robust and scalable power system state estimation via composite optimization," ~~2017, arXiv:1708.06013.~~

[30] J. C. Duchi and F. Ruan, "Stochastic methods for composite optimization problems," 2017, arXiv:1703.08570.

[31] R. Balan, "Reconstruction of signals from magnitudes of redundant representations: The complex case," *Found. Comput. Math.*, vol. 16, no. 3, pp. 677–721, Jun. 2016.

[32] *Power Systems Test Case Archive*, Univ. Washington, Seattle, WA, USA. [Online]. Available: http://www.ee.washington.edu/research/pstca Q6

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

[34] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optim. Method Softw.*, vol. 11, no. 1-4, pp. 625–653, Jan. 1999.

Gang Wang (M'18) received the B.Eng. degree in electrical engineering and automation from the Beijing Institute of Technology, Beijing, China, in 2011, and the Ph.D. degree in electrical and computer engineering from the University of Minnesota, Minneapolis, MN, USA, in 2018.

He is currently a Postdoctoral Associate with the Department of Electrical and Computer Engineering, University of Minnesota. His research interests focus on the areas of statistical signal processing, stochastic and (non-)convex optimization with applications to autonomous energy grids, and deep learning.

Dr. Wang was a recipient of a National Scholarship (2014), a Guo Rui Scholarship (2017), and an Innovation Scholarship (first place in 2017), all from China, as well as a Best Student Paper Award at the 2017 European Signal Processing Conference.

Hao Zhu (M'12) received the B.Eng. degree in electrical engineering from Tsinghua University, Beijing, China in 2006, and the M.Sc. (minor: Math) and Ph.D. degrees in electrical engineering from the University of Minnesota, Minneapolis, MN, USA, in 2009 and 2012, respectively.

She is currently an Assistant Professor in electrical and computer engineering (ECE) with The University of Texas at Austin, Austin, TX, USA. She was a Postdoctoral Researcher, from 2012 to 2013, and then an Assistant Professor in ECE from 2014–2017 with the University of Illinois at Urbana-Champaign. Her current research interests include power system monitoring and operations considering its cyber-physical coupling, as well as energy data analytics.

Dr. Zhu was a recipient of the NSF CAREER Award in 2017, the Siebel Energy Institute Seed Grant, and the U.S. AFRL Summer Faculty Fellowship in 2016. She is the faculty advisor and/or coauthor of two best papers at the North American Power Symposium. She is currently a member of the steering committee of the IEEE Smart Grid representing the IEEE Signal Processing Society.

Georgios B. Giannakis (F'97) received the Diploma degree in electrical engineering from the National Technical University of Athens, Athens, Greece, in 1981, the M.Sc. degree in electrical engineering in 1983, the M.Sc. degree in mathematics, and the Ph.D. degree in electrical engineering both in 1986 from the University of Southern California, Los Angeles, CA, USA.

He was with the University of Virginia, from 1987 to 1998 and since 1999, he has been a Professor with the University of Minnesota, Minneapolis, MN, USA, where he holds an Endowed Chair in Wireless Telecommunications, University of Minnesota McKnight Presidential Chair in Electrical and Computer Engineering, and serves as the Director of the Digital Technology Center. His general interests include the areas of communications, networking, and statistical signal processing—subjects on which he has authored/coauthored more than 430 journal papers, 730 conference papers, 25 book chapters, two edited books, and two research monographs (h-index ~~134~~). His current research focuses on learning from big data, wireless cognitive radios, and network science with applications to social, brain, and power networks with renewables. He is the (co-)inventor of 30 patents issued.

Dr. Giannakis was the (co-)recipient of nine best paper awards from the IEEE Signal Processing (SP) and Communications Societies, including the G. Marconi Prize Paper Award in Wireless Communications. He is also a recipient of the Technical Achievement Awards from the SP Society (2000), from the EURASIP (2005), a Young Faculty Teaching Award, the G. W. Taylor Award for Distinguished Research from the University of Minnesota, and the IEEE Fourier Technical Field Award (inaugural recipient in 2015). He is a Fellow of the EURASIP, and has served the IEEE in a number of posts, including that of a Distinguished Lecturer for the IEEE-SP Society.

Jian Sun (M'10) was born in Jilin Province, China, in 1978. He received the Bachelor's degree ~~from the Department of Automation and Electric Engineering,~~ Jilin Institute of Technology, Changchun, China, in 2001, the Master's degree from the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences (CAS), Changchun, China, in 2004, and the Ph.D. degree from the Institute of Automation, CAS, Beijing, China, in 2007.

He was a Research Fellow with the Faculty of Advanced Technology, University of Glamorgan, U.K., from April 2008 to October 2009. He was a Postdoctoral Research Fellow with the Beijing Institute of Technology, Beijing, China, from December 2007 to May 2010. In May 2010, he joined the School of Automation, Beijing Institute of Technology, where he has been a Professor since 2013. His research interests include control, security, and learning of networked control systems and cyber-physical systems.

Dr. Sun is currently on the Editorial Boards of *Journal of Systems Science and Complexity* and *Acta Automatica Sinica*.