Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Review:

Distribution system state estimation: an overview of recent developments*

Gang WANG[1], Georgios B. GIANNAKIS[1], Jie CHEN[2,3], Jian SUN‡[2,3]

[1]*Department of Electrical and Computer Engineering and Digital Technology Center, University of Minnesota, Minneapolis, MN 55455, USA*

[2]*School of Automation, Beijing Institute of Technology, Beijing 100081, China*

[3]*Key Laboratory of Intelligent Control and Decision of Complex Systems, Beijing 100081, China*

E-mail: gangwang@umn.edu; georgios@umn.edu; chenjie@bit.edu.cn; sunjian@bit.edu.cn

Received Sept. 23, 2018; Revision accepted Nov. 27, 2018; Crosschecked Jan. 8, 2019

Abstract: In the envisioned smart grid, high penetration of uncertain renewables, unpredictable participation of (industrial) customers, and purposeful manipulation of smart meter readings, all highlight the need for accurate, fast, and robust power system state estimation (PSSE). Nonetheless, most real-time data available in the current and upcoming transmission/distribution systems are nonlinear in power system states (i.e., nodal voltage phasors). Scalable approaches to dealing with PSSE tasks undergo a paradigm shift toward addressing the unique modeling and computational challenges associated with those nonlinear measurements. In this study, we provide a contemporary overview of PSSE and describe the current state of the art in the nonlinear weighted least-squares and least-absolute-value PSSE. To benchmark the performance of unbiased estimators, the Cramér-Rao lower bound is developed. Accounting for cyber attacks, new corruption models are introduced, and robust PSSE approaches are outlined as well. Finally, distribution system state estimation is discussed along with its current challenges. Simulation tests corroborate the effectiveness of the developed algorithms as well as the practical merits of the theory.

Key words: State estimation; Cramér-Rao bound; Feasible point pursuit; Semidefinite relaxation; Proximal linear algorithm; Composite optimization; Cyber attack; Bad data detection

https://doi.org/10.1631/FITEE.1800590 **CLC number:** TP311

1 Introduction

The electric power grid, arguably the largest complex system on Earth, is recognized as the greatest engineering achievement of the 20$^{\text{th}}$ century (Wulf, 2000): thousands of miles of transmission lines and millions of miles of distribution lines, connecting millions of power generators to millions of factories and homes. To maintain grid efficiency, reliability, and sustainability, power system operators have to constantly monitor the operating conditions of the system (Schweppe et al., 1970; Abur and Gómez-Expósito, 2004; Giannakis et al., 2013). In the early 1960s, system operators tried to compute the voltages at a few selected buses based upon manually collected meter readings from geographically distributed current and potential transformers. Unfortunately, due in part to timing, model uncertainties, and metering errors, the alternate current (AC) power flow equations were never feasible.

With the seminal contributions of Schweppe

‡ Corresponding author

* Wang G and Giannakis GB were supported by the National Natural Science Foundation of China (NSFC) (Nos. 1514056, 1505970, and 1711471). Chen J and Sun J were supported by the NSFC (Nos. 61621063 and 61522303), the NSFC-Zhejiang Joint Fund for the Integration of Industrialization and Informatization (No. 61720106011), the Projects of Major International (Regional) Joint Research Program NSFC (No. 61720106011), and the Program for Changjiang Scholars and Innovative Research Team in University (No. IRT1208)

ORCID: Gang WANG, http://orcid.org/0000-0002-7266-2412

et al. (1970), the statistical foundations were laid for (static/dynamic) power system state estimation (PSSE). Ever since their pioneering work, a number of solutions and generalizations of PSSE have been proposed and worked out. Furthermore, many articles, chapters, and books have nicely reviewed the progress in this area (Wood and Wollenberg, 1996; Monticelli, 2000; Baran, 2001; Abur and Gómez-Expósito, 2004; Caro and Conejo, 2012; Huang et al., 2012; Giannakis et al., 2013; Della Giustina et al., 2014; Kekatos et al., 2017; Ahmad et al., 2018). In this study, we provide a contemporary overview of PSSE, linking the relevant physics to the signal processing and (nonconvex) optimization methods and algorithms. Our main goal is to describe some of the recent advances in this area, identify current challenges, and suggest directions for future research. Admittedly, our collection is by no means exhaustive, but it gives an indication of the related research taking place.

The notations are explained as follows: Matrices (column vectors) are denoted by upper- (lower-) case boldface letters; in particular, $\mathbf{1}$ is an all-one vector of suitable dimension. Sets are represented using calligraphic letters. Symbol "$^{\mathsf{T}}$" represents the transpose and symbol "$^{\mathsf{H}}$" represents the Hermitian transpose. $\overline{(\cdot)}$ denotes complex conjugate, while $\Re(\cdot)$ ($\Im(\cdot)$) takes the real (imaginary) part of a complex number.

2 Grid modeling preliminaries

In this section, we briefly review Kirchhoff's and Ohm's laws as well as the power flow equations. An electric power grid comprising N nodes (i.e., buses) and L edges (i.e., lines) can be 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. Since the focus of this study is on AC circuits, steady-state voltages and currents will be represented by their single-phase equivalent phasors per unit.

A transmission line $(n, n') \in \mathcal{L}$ connecting nodes n and $n' \in \mathcal{N}$ can be modeled by its line series admittance $y_{nn'} = g_{nn'} + \mathrm{j}b_{nn'}$, and total shunt susceptance $\mathrm{j}b^{\mathrm{s}}_{nn'}$. Letting $v_n \in \mathbb{C}$ be the complex voltage at bus n, the current $i_{nn'} \in \mathbb{C}$ flowing from node n to node n' across line (n, n') is given by

$$i_{nn'} = (y_{nn'} + \mathrm{j}b^{\mathrm{s}}_{nn'}/2)v_n - y_{nn'}v_{n'}. \tag{1}$$

The reverse-direction current $i_{n'n}$ flowing from node n' to node n can be expressed symmetrically. According to Eq. (1), unless $b_{nn'} = 0$, it holds that $i_{nn'} \neq i_{n'n}$. According to Kirchhoff's current law, the complex current injected into node n is equal to the sum of currents on the lines incident to node n; that is,

$$i_n = \sum_{n' \in \mathcal{N}_n} i_{n'n}, \tag{2}$$

where $\mathcal{N}_n \subseteq \mathcal{N}$ is the set of nodes directly connected to node n.

Upon stacking up all nodal voltages (currents) to form the $N \times 1$ vector $\boldsymbol{v}$ ($\boldsymbol{i}$), Ohm's law dictates that

$$\boldsymbol{i} = \boldsymbol{Y}\boldsymbol{v}, \tag{3}$$

where $\boldsymbol{Y}$ is the so-called bus admittance matrix, whose $(n, n)^{\mathrm{th}}$ entry is $-\sum_{n' \in \mathcal{N}_n} y_{n'n}$ and $(n, n')^{\mathrm{th}}$ entry is $\sum_{n' \in \mathcal{N}_n}(y_{n'n} + \mathrm{j}b^{\mathrm{s}}_{n'n}/2)$. Note that $\boldsymbol{Y}$ is symmetric and sparse. Similarly, one can collect all line currents in the $2L \times 1$ vector $\boldsymbol{i}_{\mathrm{f}}$ and write

$$\boldsymbol{i}_{\mathrm{f}} = \boldsymbol{Y}_{\mathrm{f}}\boldsymbol{v}, \tag{4}$$

for a properly defined $2L \times N$ complex matrix $\boldsymbol{Y}_{\mathrm{f}}$.

Let $s_n := p_n + \mathrm{j}q_n$ denote the complex power injected into node n. Using the definition $s_n = v_n\overline{i}_n$, the vector collecting all complex power injections $\boldsymbol{s} = \boldsymbol{p} + \mathrm{j}\boldsymbol{q}$ can be compactly expressed as

$$\boldsymbol{s} = \mathrm{diag}(\boldsymbol{v})\overline{\boldsymbol{i}} = \mathrm{diag}(\boldsymbol{v})\overline{\boldsymbol{Y}}\,\overline{\boldsymbol{v}}. \tag{5}$$

The power flowing from node n to node n' over line (n, n') seen from node n is similarly given by $S_{nn'} = v_n\overline{i}_{nn'}$, or in a matrix-vector representation by

$$\boldsymbol{S} = \mathrm{diag}(\boldsymbol{v})\overline{\boldsymbol{Y}_{\mathrm{f}}}\,\overline{\boldsymbol{v}}. \tag{6}$$

Matrix $\boldsymbol{Y}$ is often given in rectangular coordinates as $\boldsymbol{Y} = \boldsymbol{G} + \mathrm{j}\boldsymbol{B}$. Depending on whether the complex voltages $\boldsymbol{v}$ are expressed in polar or rectangular coordinates, the AC power flow equations admit two options. Specifically, if voltages are $v_n = V_n \exp(\mathrm{j}\theta_n)$, the real and imaginary parts of the power flow equations (Eq. (5)) can be written as

$$\begin{cases} p_n = \sum\limits_{n'=1}^{N} V_n V_{n'}(G_{nn'}\cos\theta_{nn'} + B_{nn'}\sin\theta_{nn'}), \\ q_n = \sum\limits_{n'=1}^{N} V_n V_{n'}(G_{nn'}\sin\theta_{nn'} - B_{nn'}\cos\theta_{nn'}), \end{cases} \tag{7}$$

where for notational brevity, we use $\theta_{nn'} := \theta_n - \theta_{n'}$ for all $n \in \mathcal{N}$. Observe that both $\{p_n\}$ and $\{q_n\}$

are solely functions of the phase differences $\{\theta_{nn'}\}$; the power injections $\{s_n\}$ are invariant if all nodal voltages are shifted by a common angle. This justifies the adoption of a reference bus (i.e., slack bus) in the power system analysis that is assumed to have a zero voltage phase without loss of generality.

If, on the other hand, voltages are given in rectangular coordinates as $v_n = V_{r,n} + \mathrm{j}V_{i,n}$, the power injections become quadratically related to voltages:

$$\begin{cases} p_n = V_{r,n} \sum\limits_{n'=1}^{N} (V_{r,n}G_{nn'} - V_{i,n}B_{nn'}) \\ \qquad + V_{i,n} \sum\limits_{n'=1}^{N} (V_{i,n}G_{nn'} + V_{r,n}B_{nn'}), \\ q_n = V_{i,n} \sum\limits_{n'=1}^{N} (V_{i,n}G_{nn'} - V_{r,n}B_{nn'}) \\ \qquad - V_{r,n} \sum\limits_{n'=1}^{N} (V_{r,n}G_{nn'} + V_{i,n}B_{nn'}). \end{cases} \tag{8}$$

Using the fact that $\overline{s}_n = \overline{v}_n i_n = (\boldsymbol{v}^{\mathsf{H}}\boldsymbol{e}_n)(\boldsymbol{e}_n^{\mathsf{T}}\boldsymbol{i}) = \boldsymbol{v}^{\mathsf{H}}\boldsymbol{e}_n\boldsymbol{e}_n^{\mathsf{T}}\boldsymbol{v}$, Eq. (8) can be compactly expressed as

$$\begin{cases} p_n = \boldsymbol{v}^{\mathsf{H}}\boldsymbol{H}_n^p\boldsymbol{v}, \quad \text{with} \quad \boldsymbol{H}_n^p := \dfrac{\boldsymbol{Y}_n^{\mathsf{H}} + \boldsymbol{Y}_n}{2}, \\ q_n = \boldsymbol{v}^{\mathsf{H}}\boldsymbol{H}_n^q\boldsymbol{v}, \quad \text{with} \quad \boldsymbol{H}_n^q := \dfrac{\boldsymbol{Y}_n^{\mathsf{H}} - \boldsymbol{Y}_n}{2\mathrm{j}}, \end{cases} \tag{9}$$

where $\boldsymbol{Y}_n := \boldsymbol{e}_n\boldsymbol{e}_n^{\mathsf{T}}\boldsymbol{Y}$ for all $n \in \mathcal{N}$. With regard to power flows, since a line current can also be expressed as $i_{nn'} = \boldsymbol{e}_{nn'}^{\mathsf{T}}\boldsymbol{i}_{\mathrm{f}}$, it holds that $\overline{S}_{nn'} = \overline{v}_n i_{nn'} = (\boldsymbol{v}^{\mathsf{H}}\boldsymbol{e}_n)(\boldsymbol{e}_{nn'}^{\mathsf{T}}\boldsymbol{i}_{\mathrm{f}}) = \boldsymbol{v}^{\mathsf{H}}\boldsymbol{e}_n\boldsymbol{e}_{nn'}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{f}}\boldsymbol{v}$, thus yielding

$$\begin{cases} P_{nn'} = \boldsymbol{v}^{\mathsf{H}}\boldsymbol{H}_{nn'}^P\boldsymbol{v}, \quad \text{with} \quad \boldsymbol{H}_{nn'}^P := \dfrac{\boldsymbol{Y}_{nn'}^{\mathsf{H}} + \boldsymbol{Y}_{nn'}}{2}, \\ Q_{nn'} = \boldsymbol{v}^{\mathsf{H}}\boldsymbol{H}_{nn'}^Q\boldsymbol{v}, \quad \text{with} \quad \boldsymbol{H}_{nn'}^Q := \dfrac{\boldsymbol{Y}_{nn'}^{\mathsf{H}} - \boldsymbol{Y}_{nn'}}{2\mathrm{j}}, \end{cases} \tag{10}$$

where $\boldsymbol{Y}_{nn'} := \boldsymbol{e}_n\boldsymbol{e}_{nn'}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{f}}$ for all lines $(n, n') \in \mathcal{L}$. Similar expressions can be obtained for the voltage magnitude squares; that is,

$$V_n^2 = \boldsymbol{v}^{\mathsf{H}}\boldsymbol{H}_n^v\boldsymbol{v}, \quad \text{with} \quad \boldsymbol{H}_n^v := \boldsymbol{e}_n\boldsymbol{e}_n^{\mathsf{T}}. \tag{11}$$

In a nutshell, Eqs. (9)–(11) imply that power injections, flows, and voltage magnitude squares are quadratic functions of the voltage vector, which can be collectively described by $\{\boldsymbol{v}^{\mathsf{H}}\boldsymbol{H}_m\boldsymbol{v}\}$ for Hermitian matrices $\boldsymbol{H}_m = \boldsymbol{H}_m^{\mathsf{H}} \in \mathbb{C}^{N\times N}$ given in Eqs. (9)–(11).

3 Weighted least-squares estimation

As explained in Section 2, given system parameters collected in $\boldsymbol{Y}$ and $\boldsymbol{Y}_{\mathrm{f}}$, all power system quantities can be expressed in terms of voltage vector $\boldsymbol{v}$, thus justifying its term as the system state. In practice, the supervision control and data acquisition (SCADA) system hosting geographically distributed metering devices measures a subset of electric quantities every few seconds and forwards the readings via remote terminal units (RTUs) to a control center for grid monitoring. The task of PSSE entails recovering the voltage vector given the available measurements and grid parameters. In this section, we start with the (weighted) (W) least-squares (LS) formulation of PSSE, provide the Cramér-Rao bound (CRB) on the variance of any unbiased LS estimator, and outline several WLS-based PSSE solvers.

3.1 Problem formulation

Suppose that we have a total of M SCADA measurements $\{z_m\}_{m=1}^M$ collectively denoted by $\boldsymbol{z} \in \mathbb{R}^M$, which relates to $\boldsymbol{v}$ via the model:

$$\boldsymbol{z} = \boldsymbol{h}(\boldsymbol{v}) + \boldsymbol{\epsilon}, \tag{12}$$

for some properly defined vector-valued nonlinear function $\boldsymbol{h}(\cdot)$, where $\boldsymbol{\epsilon} \in \mathbb{R}^M$ captures the model mismatches and measurement noise.

Traditionally, the system state $\boldsymbol{v}$ is expressed in polar coordinates, which is a $(2N-1)\times 1$ vector that comprises the real and imaginary parts of $\boldsymbol{v}$ after excluding the imaginary part of the reference bus. In that case, function $\boldsymbol{h}(\boldsymbol{v})$ maps the real-valued state vector to SCADA measurements through the nonlinear power flow (Eq. (7)). The motivation behind expressing the states in polar coordinates is twofold. First, voltage magnitude measurements are directly related to states. In addition, the Jacobian matrix of $\boldsymbol{h}(\boldsymbol{v})$ required in the Gauss-Newton iterations is amenable to approximations. Nonetheless, when iterative optimization solvers are employed, working directly with the N-dimensional complex voltage vector has in general lower computational complexity than that in the real case. This is primarily due to the compact quadratic representations of SCADA measurements in the complex voltage state vector, enabling us to exploit the inherent sparsity of quadratic measurement matrices (Eqs. (9)–(11)). Refer to recent computational reformulations of PSSE in complex variables (Wang G et al., 2017, 2018b; Wang Z et al., 2017; Džafić et al., 2018a, 2018b). In a different but related context, where the goal is to reconstruct complex signal vectors from their intensity-only measurements, Candès

et al. (2015) and Wang et al. (2018a, 2018d) carefully justified the computational advantages of optimizing over complex variables rather than over their real expansions.

Upon pre-whitening, vector $\boldsymbol{\epsilon}$ can be assumed to follow a standardized Gaussian distribution. The maximum likelihood estimate (MLE) of $\boldsymbol{v}$ coincides with the following nonlinear LS estimate (Kay, 1993):

$$\hat{\boldsymbol{v}} := \arg\min_{\boldsymbol{v}\in\mathbb{C}^n} \ \|\boldsymbol{z}-\boldsymbol{h}(\boldsymbol{v})\|_2^2 . \tag{13}$$

Due to nonlinear $\boldsymbol{h}(\boldsymbol{v})$, the LS at hand is nonconvex, and its general instance is non-deterministic polynomial (NP) hard (Pardalos and Vavasis, 1991). Hence, solving the LS-based PSSE problem is indeed challenging.

3.2 Cramér-Rao bound analysis

In estimation theory (Kay, 1993, Chapter 3), the CRB provides a universal lower bound on the variance of any unbiased estimator. Appreciating its central role as a performance benchmark, we establish the CRB for the LS PSSE. Evidently, from Eq. (13), the CRB analysis of PSSE entails finding derivatives of a real-valued function $\boldsymbol{h}(\boldsymbol{v})$ with respect to complex-valued variables in $\boldsymbol{v}$. Addressing this challenge calls for the so-called Wirtinger's derivatives and calculus for complex analysis, the basics of which are provided in Appendix A. The following result provides a closed-form CRB for the LS PSSE under the additive white Gaussian noise model (12), whose proof can be found in Wang et al. (2018b).

Theorem 1 Consider estimating the unknown vector $\boldsymbol{v}\in\mathbb{C}^N$ from noisy observations $\{z_m\in\mathbb{R}\}_{m=1}^M$ obeying model (12), where the noise $\boldsymbol{\epsilon}$ follows the standardized Gaussian distribution $\mathcal{N}(\boldsymbol{0},\boldsymbol{I})$. Then it holds for any unbiased estimator $\hat{\boldsymbol{v}}$ that

$$\mathrm{cov}(\hat{\boldsymbol{v}}) \succeq \left[\boldsymbol{F}^{\dagger}(\boldsymbol{v},\overline{\boldsymbol{v}})\right]_{1:N,1:N}, \tag{14}$$

where $\mathrm{cov}(\cdot)$ denotes the covariance matrix of the argument, "$\dagger$" denotes the pseudo-inverse operator, and the Fisher information matrix (FIM) is given as

$$\boldsymbol{F}(\boldsymbol{v},\overline{\boldsymbol{v}}) = \sum_{m=1}^{M} \boldsymbol{F}_m(\boldsymbol{v},\overline{\boldsymbol{v}}), \tag{15}$$

$$\boldsymbol{F}_m(\boldsymbol{v},\overline{\boldsymbol{v}}) := \begin{bmatrix} (\boldsymbol{H}_m\boldsymbol{v})(\boldsymbol{H}_m\boldsymbol{v})^{\mathsf{H}} & (\boldsymbol{H}_m\boldsymbol{v})(\overline{\boldsymbol{H}}_m\overline{\boldsymbol{v}})^{\mathsf{H}} \\ (\overline{\boldsymbol{H}}_m\overline{\boldsymbol{v}})(\boldsymbol{H}_m\boldsymbol{v})^{\mathsf{H}} & (\overline{\boldsymbol{H}}_m\overline{\boldsymbol{v}})(\overline{\boldsymbol{H}}_m\overline{\boldsymbol{v}})^{\mathsf{H}} \end{bmatrix}. \tag{16}$$

Furthermore, matrix $\boldsymbol{F}(\boldsymbol{v},\overline{\boldsymbol{v}})$ has at least rank-one deficiency even when all SCADA quantities are measured.

The proof of Theorem 1 can be found in Wang et al. (2018b). Regarding Theorem 1, a couple of remarks are of interest. The rank deficiency of the FIM stems from the inherent voltage ambiguity; that is, all SCADA observations are invariant even if all nodal voltages are shifted by a uni-modular phase constant. This issue can be fixed by selecting a reference bus and setting its phase to be zero or any constant. Although it is rank-deficient, the pseudo-inverse of $\boldsymbol{F}(\boldsymbol{v},\overline{\boldsymbol{v}})$ qualifies itself as a valid lower bound on the mean square error (MSE) of any unbiased estimator (Stoica and Marzetta, 2001). In addition, this lower bound is often attainable in practice, and is predictive of the optimal estimator performance (Stoica and Marzetta, 2001). Having derived the CRB for PSSE, we will deal with LS PSSE solvers in the next subsection.

3.3 Gauss-Newton method

Consider the nonlinear LS problem (13) again, for which the Gauss-Newton iterations are widely known to be the "workhorse" solution (Bertsekas, 1999, Chapter 1.5; Abur and Gómez-Expósito, 2004, Chapter 2). Starting with an initial guess or a flat-voltage profile vector (e.g., the all-one vector, the voltage vector obtained by solving the linearized power flow equations), denoted by $\boldsymbol{v}_0\in\mathbb{C}^N$, the Gauss-Newton method successively approximates the nonlinear LS fit in Eq. (13) using the linear one of the first-order Taylor expansion of $\boldsymbol{h}(\boldsymbol{v}_i)$, and relies on its minimizer to yield the next iteration $\boldsymbol{v}_{i+1}$. Specifically, according to Wirtinger's calculus in Appendix A, the first-order Taylor expansion of $\boldsymbol{h}(\boldsymbol{v})$ around the current iteration $\boldsymbol{v}_i$ is (refer to Eq. (A6))

$$\boldsymbol{h}(\boldsymbol{v}) \approx \boldsymbol{h}(\boldsymbol{v}_i) + \nabla_{\mathrm{c}}^{\mathsf{H}}\boldsymbol{h}(\boldsymbol{v}_i) \begin{bmatrix} \boldsymbol{v}-\boldsymbol{v}_i \\ \overline{\boldsymbol{v}}-\overline{\boldsymbol{v}}_i \end{bmatrix}, \tag{17}$$

where the complex Jacobian $\nabla_{\mathrm{c}}\boldsymbol{h}(\boldsymbol{v}_i)$ of $\boldsymbol{h}(\boldsymbol{v},\overline{\boldsymbol{v}})$ with respect to $[\boldsymbol{v}^{\mathsf{T}}\ \overline{\boldsymbol{v}}^{\mathsf{T}}]^{\mathsf{T}}$ evaluated using $\boldsymbol{v}_i$ is given in Eq. (A5). For notational brevity, let $\boldsymbol{J}_i := \nabla_{\mathrm{c}}\boldsymbol{h}(\boldsymbol{v}_i)\in\mathbb{C}^{2N\times M}$, and assume that $\boldsymbol{J}_i\boldsymbol{J}_i^{\mathsf{H}}$ is invertible, which is often true when SCADA data are measured across the network and $M\geq N$.

The next iteration $\boldsymbol{v}_{i+1}$ can then be obtained by

solving the ensuing LS

$$\min_{\boldsymbol{v}\in\mathbb{C}^N} \left\| \boldsymbol{z}-\boldsymbol{h}(\boldsymbol{v}_i,\overline{\boldsymbol{v}}_i)-\boldsymbol{J}_i^{\mathsf{H}}\begin{bmatrix}\boldsymbol{v}-\boldsymbol{v}_i\\ \overline{\boldsymbol{v}}-\overline{\boldsymbol{v}}_i\end{bmatrix}\right\|_2^2, \tag{18}$$

or equivalently, solving the linear system

$$\boldsymbol{J}_i\boldsymbol{J}_i^{\mathsf{H}}\begin{bmatrix}\boldsymbol{v}-\boldsymbol{v}_i\\ \overline{\boldsymbol{v}}-\overline{\boldsymbol{v}}_i\end{bmatrix}=\boldsymbol{J}_i\left[\boldsymbol{z}-\boldsymbol{h}(\boldsymbol{v}_i,\overline{\boldsymbol{v}}_i)\right], \tag{19}$$

for which celebrated efficient solvers that can exploit the sparsity of $\boldsymbol{J}_i$ exist (Saad, 2003). Succinctly, the state estimate is iteratively updated using the following until some stopping criteria are met:

$$\begin{bmatrix}\boldsymbol{v}_{i+1}\\ \overline{\boldsymbol{v}}_{i+1}\end{bmatrix}=\begin{bmatrix}\boldsymbol{v}_i\\ \overline{\boldsymbol{v}}_i\end{bmatrix}+\left(\boldsymbol{J}_i\boldsymbol{J}_i^{\mathsf{H}}\right)^{-1}\boldsymbol{J}_i\left[\boldsymbol{z}-\boldsymbol{h}(\boldsymbol{v}_i)\right]. \tag{20}$$

Refer to some generalizations to cope with complex-valued measurements and phasor measurement units (PMU) data in Džafić et al. (2018b).

On the other hand, if $\boldsymbol{\epsilon}\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{\Sigma})$ for some covariance matrix $\boldsymbol{\Sigma}\succ\boldsymbol{0}$ (instead of $\boldsymbol{\epsilon}\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{I})$), the maximum likelihood estimation can be expressed as the WLS estimation:

$$\min_{\boldsymbol{v}\in\mathbb{C}^N} \left\|\boldsymbol{\Sigma}^{-1/2}\left[\boldsymbol{z}-\boldsymbol{h}(\boldsymbol{v})\right]\right\|_2^2. \tag{21}$$

The Gauss-Newton iterations can be similarly derived by treating $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{z}$ as $\boldsymbol{z}$ and $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{h}(\boldsymbol{v})$ as $\boldsymbol{h}(\boldsymbol{v})$ in Eq. (20), yielding

$$\begin{bmatrix}\boldsymbol{v}_{i+1}\\ \overline{\boldsymbol{v}}_{i+1}\end{bmatrix}=\begin{bmatrix}\boldsymbol{v}_i\\ \overline{\boldsymbol{v}}_i\end{bmatrix}+\left(\boldsymbol{J}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{J}_i^{\mathsf{H}}\right)^{-1}\boldsymbol{J}_i\boldsymbol{\Sigma}^{-1}\left[\boldsymbol{z}-\boldsymbol{h}(\boldsymbol{v}_i)\right]. \tag{22}$$

The Gauss-Newton iterations in Eqs. (20) and (22) may not guarantee convergence; however, they rely heavily on $\boldsymbol{v}_0$. To improve convergence and ensure descent of the (W)LS cost function, the Gauss-Newton method is implemented in the modified form (Bertsekas, 1999, Chapter 1.5):

$$\begin{bmatrix}\boldsymbol{v}_{i+1}\\ \overline{\boldsymbol{v}}_{i+1}\end{bmatrix}=\begin{bmatrix}\boldsymbol{v}_i\\ \overline{\boldsymbol{v}}_i\end{bmatrix}+\mu_i\left(\boldsymbol{J}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{J}_i^{\mathsf{H}}\right)^{-1}\boldsymbol{J}_i\boldsymbol{\Sigma}^{-1}\left[\boldsymbol{z}-\boldsymbol{h}(\boldsymbol{v}_i)\right], \tag{23}$$

where $\mu_i>0$ is a step size chosen by means of a backtracking line search (Bertsekas, 1999, Chapter 1.2). Furthermore, it is known that the Gauss-Newton procedure for nonconvex optimization can get stuck at local solutions (Bertsekas, 1999, Chapter 1.5). All in all, the main challenge lies in developing PSSE solvers that can attain or approximate the global optimum at affordable computational complexity. Toward this objective, we review a few recent interesting proposals along this line next.

3.4 Semidefinite programming relaxation

As explained in Section 2, the challenge of the LS PSSE Eq. (13) arises from the quadratic functions $\{h_m(\boldsymbol{v})=\boldsymbol{v}^{\mathsf{H}}\boldsymbol{H}_m\boldsymbol{v}\}$. A recent line of research to tackle the nonlinear measurements expresses $\{h_m\}$ as linear functions of the outer-product $\boldsymbol{V}:=\boldsymbol{v}\boldsymbol{v}^{\mathsf{H}}$ and subsequently solves a semidefinite program (SDP) over $\boldsymbol{V}\in\mathbb{C}^{N\times N}$ after relaxing the nonconvex rank constraint $\mathrm{rank}(\boldsymbol{V})=1$ (Zhu and Giannakis, 2011, 2014; Kim et al., 2014; Wang et al., 2014). Precisely, problem (13) can be equivalently rewritten as

$$\begin{aligned}\hat{\boldsymbol{V}}:=\arg\min_{\boldsymbol{V}\in\mathbb{C}^{N\times N}}\ &\sum_{m=1}^{M}\left[z_m-\mathrm{tr}(\boldsymbol{H}_m\boldsymbol{V})\right]^2\\ \text{s.t.}\quad &\boldsymbol{V}\succeq\boldsymbol{0},\ \mathrm{rank}(\boldsymbol{V})=1,\end{aligned} \tag{24}$$

where the constraints jointly ensure that any solution $\hat{\boldsymbol{V}}$ can be uniquely expressed as $\hat{\boldsymbol{V}}=\hat{\boldsymbol{v}}\hat{\boldsymbol{v}}^{\mathsf{H}}$ for some $\hat{\boldsymbol{v}}\in\mathbb{C}^N$. Problem (24) can be readily transformed into a convex SDP after dropping the rank-one constraint, yielding

$$\begin{aligned}\min_{\substack{\boldsymbol{V}\in\mathbb{C}^{N\times N},\\ \boldsymbol{\beta}\in\mathbb{R}^M}}\ &\boldsymbol{1}^{\mathsf{T}}\boldsymbol{\beta}\\ \text{s.t.}\quad &\boldsymbol{V}\succeq\boldsymbol{0},\\ &\begin{bmatrix}\beta_m & z_m-\mathrm{tr}(\boldsymbol{H}_m\boldsymbol{V})\\ z_m-\mathrm{tr}(\boldsymbol{H}_m\boldsymbol{V}) & 1\end{bmatrix}\succeq\boldsymbol{0},\\ &\forall\, m=1,2,\ldots,M,\end{aligned} \tag{25}$$

for which off-the-shelf convex programming solvers can be employed. An estimate of $\boldsymbol{v}$ can be recovered as the principal component of the $\boldsymbol{V}$-solution of problem (25) via eigenvalue decomposition, or through randomization techniques (Zhu and Giannakis, 2014).

In terms of performance, it can be shown that under appropriate assumptions, solving the convex SDP (25) attains the global optimum of the LS PSSE problem (13) when the SCADA data are noise-free; refer to Zhu and Giannakis (2014) for details. In practice, the SDP relaxation approach approximates the global optimum well even in the presence of noise. Nevertheless, solving SDPs often calls for interior-point solvers, whose computational complexity grows at least cubically with the matrix size N (Park and Boyd, 2017). This complexity can be a burden for real-time power system operation, which motivates lightweight alternatives.

3.5 Feasible point pursuit

Casting the LS in problem (13) as a nonconvex quadratically constrained quadratic program (QCQP), the feasible point pursuit (FPP) method investigated in Wang et al. (2018b) offers a computationally affordable alternative for approximating the global optimum of PSSE. The idea of FPP is to solve a sequence of convexified QCQPs obtained by successively forming convex inner-restrictions of the original nonconvex feasibility set, to approximate the feasible solutions of the original nonconvex QCQP (Mehanna et al., 2015; Park and Boyd, 2017).

Toward this end, the FPP-PSSE solver starts with reformulating problem (13) into a QCQP (Wang et al., 2016):

$$\min_{\substack{\boldsymbol{v}\in\mathbb{C}^N,\\ \boldsymbol{\chi}\in\mathbb{R}^M}} \ \|\boldsymbol{\chi}\|_2^2 \tag{26a}$$

$$\text{s.t.} \quad \boldsymbol{v}^{\mathsf{H}}\boldsymbol{H}_m\boldsymbol{v} \le z_m+\chi_m, \tag{26b}$$

$$\boldsymbol{v}^{\mathsf{H}}\boldsymbol{H}_m\boldsymbol{v} \ge z_m-\chi_m, \tag{26c}$$

$$\forall\, m=1,2,\ldots,M,$$

where vector $\boldsymbol{\chi}\in\mathbb{R}^M$ consists of M auxiliary variables $\{\chi_m\ge 0\}_{m=1}^M$, capturing the residuals of fitting the M SCADA observations.

Evidently, if a certain $\boldsymbol{H}_m$ is neither positive nor negative semidefinite, constraints (26b) and (26c) are both nonconvex. Yet, if $\boldsymbol{H}_m$ is (positive or negative) semidefinite, one and only one of constraints (26b) and (26c) is nonconvex. Without loss of generality, consider decomposing every $\boldsymbol{H}_m$ into its positive and negative definite parts as $\boldsymbol{H}_m:=\boldsymbol{H}_m^+ + \boldsymbol{H}_m^-$ using eigenvalue decomposition. Equivalently, constraints (26b) and (26c) can be rewritten as

$$\boldsymbol{v}^{\mathsf{H}}\boldsymbol{H}_m^+\boldsymbol{v}+\boldsymbol{v}^{\mathsf{H}}\boldsymbol{H}_m^-\boldsymbol{v} \le z_m+\chi_m, \tag{27a}$$

$$\boldsymbol{v}^{\mathsf{H}}\boldsymbol{H}_m^+\boldsymbol{v}+\boldsymbol{v}^{\mathsf{H}}\boldsymbol{H}_m^-\boldsymbol{v} \ge z_m-\chi_m. \tag{27b}$$

It is clear that only the negative component $\boldsymbol{v}^{\mathsf{H}}\boldsymbol{H}_m^-\boldsymbol{v}$ in inequality (27a) is nonconvex; it is similar for the positive part in inequality (27b). By definition, we find that the following holds for any point $\boldsymbol{v}_i\in\mathbb{C}^N$:

$$\boldsymbol{v}^{\mathsf{H}}\boldsymbol{H}_m^-\boldsymbol{v} \le 2\Re(\boldsymbol{v}_i^{\mathsf{H}}\boldsymbol{H}_m^-\boldsymbol{v})-\boldsymbol{v}_i^{\mathsf{H}}\boldsymbol{H}_m^-\boldsymbol{v}_i, \tag{28a}$$

$$\boldsymbol{v}^{\mathsf{H}}\boldsymbol{H}_m^+\boldsymbol{v} \ge 2\Re(\boldsymbol{v}_i^{\mathsf{H}}\boldsymbol{H}_m^-\boldsymbol{v})-\boldsymbol{v}_i^{\mathsf{H}}\boldsymbol{H}_m^-\boldsymbol{v}_i. \tag{28b}$$

Starting with some $\boldsymbol{v}_0\in\mathbb{C}^N$, each iteration of FPP first replaces the nonconvex sources in inequalities (27a) and (27b) with the corresponding linear bounds in inequalities (28a) and (28b) evaluated at the current iteration $\boldsymbol{v}_i$, and subsequently solves the resulting convex QCQP to obtain the next iteration $\boldsymbol{v}_{i+1}$ as the $\boldsymbol{v}$-solution of

$$\min_{\substack{\boldsymbol{v}\in\mathbb{C}^N,\\ \boldsymbol{\chi}\in\mathbb{R}^M}} \ \|\boldsymbol{\chi}\|_2^2$$

$$\begin{aligned}\text{s.t.} \quad & \boldsymbol{v}^{\mathsf{H}}\boldsymbol{H}_m^+\boldsymbol{v}+2\Re(\boldsymbol{v}_i^{\mathsf{H}}\boldsymbol{H}_m^-\boldsymbol{v}) \le z_m+\boldsymbol{v}_i^{\mathsf{H}}\boldsymbol{H}_m^-\boldsymbol{v}_i+\chi_m,\\ & \boldsymbol{v}^{\mathsf{H}}\boldsymbol{H}_m^-\boldsymbol{v}+2\Re(\boldsymbol{v}_i^{\mathsf{H}}\boldsymbol{H}_m^-\boldsymbol{v}) \ge z_m+\boldsymbol{v}_i^{\mathsf{H}}\boldsymbol{H}_m^-\boldsymbol{v}_i-\chi_m,\\ & \forall\, m=1,2,\ldots,M,\end{aligned} \tag{29}$$

for which standard convex programming methods can be used. The FPP procedure has been shown to converge to a stationary point of the LS Eq. (13). Yet, extensive numerical tests demonstrate its capability of attaining (near-)optimal solutions. Generalizations to handle bad data as well as PMU measurements are possible, and related discussion can be found in Wang et al. (2018b).

3.6 Numerical tests

To summarize this section, we provide numerical tests comparing the Gauss-Newton method, semidefinite relaxation (SDR) based, and FPP-based PSSE solvers, using the IEEE 14-bus and 30-bus benchmark systems (Christie, 1999). The true voltage profile was generated with magnitude uniformly sampled from $[0.9, 1.1]$ in per unit system and angles from $[-0.4\pi, 0.4\pi]$. Independent zero-mean Gaussian additive noise with standard deviation 0.02 for voltage meters and 0.05 for power meters was assumed. All reported results below were averaged over 100 independent trials.

The first experiment assessed the mean square estimation error (MSE) performance of the three schemes against the CRB benchmark in Theorem 1 using the IEEE 14-bus system. A varying number of measurements were simulated. At first, all voltage magnitudes and all active power flows at both the sending- and receiving-ends were measured, which correspond to the base case 1 on the x axis of Fig. 1. Increasing one on the x axis implies including a new type of measurements from $\{Q_{nn'}^f, Q_{nn'}^t, P_n, Q_n\}$ at all buses or over all lines. In other words, when the axis value equals five, all measurements were used. In the second experiment on the IEEE 30-bus system, we simulated all nodal voltage magnitudes as well as all the active power flows at both sending- and receiving-ends. Fig. 2 describes the estimated

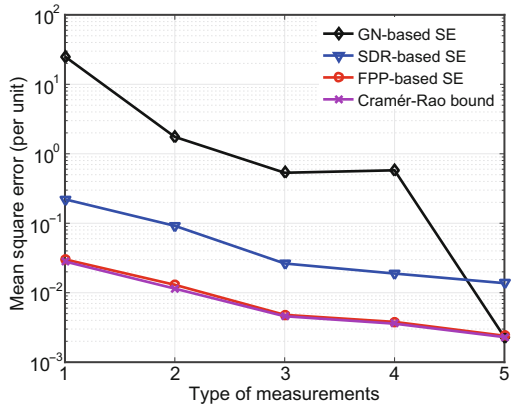

Fig. 1 MSE and Cramér-Rao bound versus the type of measurements used for the IEEE 14-bus system

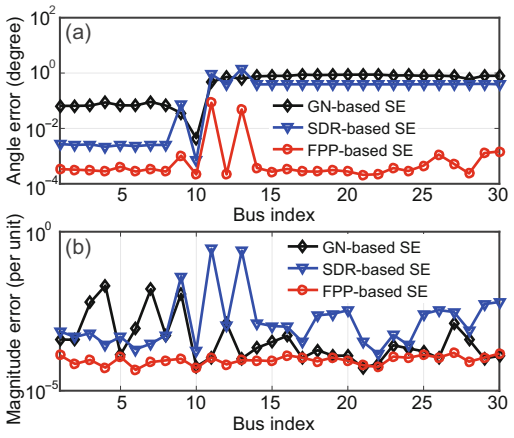

Fig. 2 Angle estimation errors (a) and voltage magnitudes (b) per bus for the IEEE 30-bus system

voltage angles and magnitudes across buses obtained using different schemes. Figs. 1 and 2 corroborate the near-optimal performance and robustness of the FPP-based PSSE solver in the simulated settings.

4 Robust power system state estimation

With utilities increasingly shifting toward smart grid technology as well as other upgrades with inherent cyber vulnerabilities, the power grid has seen correlative threats from adversarial cyber attacks growing in form and frequency (Fairley, 2016). Traditional PSSE approaches, particularly the (W)LS SE, are being challenged by these new issues concerning data integrity and uninformed model changes (Liu et al., 2011; Zhu and Giannakis, 2012; Kekatos and Giannakis, 2013). These concerns strongly motivate the development of robust PSSE approaches against anomalous (i.e., bad) data and model inaccuracies. In addition to cyber attacks, bad data (i.e., outliers) may be due to communication delays, system parameter uncertainties, and/or meter mis-calibration. In this section, we review some of the recent advances in robust PSSE.

4.1 Attack models

In the presence of bad data, the following corruption model is considered (Duchi and Ruan, 2017a; Wang G et al., 2017): letting $\{a_m \in \mathbb{R}\}_{m=1}^M$ model an arbitrary attack sequence, we observe

$$z_m \approx \begin{cases} \boldsymbol{v}^{\mathsf{H}}\boldsymbol{H}_m\boldsymbol{v}, & \text{if } m \in \mathcal{I}^{\mathrm{n}}, \\ a_m, & \text{if } m \in \mathcal{I}^{\mathrm{a}}, \end{cases} \tag{30}$$

for $m = 1, 2, \ldots, M$, where additive measurement noise can be included if "$\approx$" is replaced with "$=$" and the set $\mathcal{I}^{\mathrm{n}} \subseteq \{1, 2, \ldots, M\}$ ($\mathcal{I}^{\mathrm{a}}$) collects the indices of nominal (outlying) data. Furthermore, elements of $\mathcal{I}^{\mathrm{a}}$ are assumed to be randomly chosen from $\{1, 2, \ldots, M\}$. Relying on whether $\{a_m\}$ is independent of $\{\boldsymbol{H}_m\}$, we consider the following two models for the attacks:

M1: Attacks $\{a_m\}_{m=1}^M$ are independent of $\{\boldsymbol{H}_m\}_{m=1}^M$.

M2: Attacks $\{a_m\}_{m\in\mathcal{I}^{\mathrm{a}}}$ are independent of nominal measurement matrices $\{\boldsymbol{H}_m\}_{m\in\mathcal{I}^{\mathrm{n}}}$.

Note that M1 requires full independence between the corruption and the measurements. In other words, the attacker may solely corrupt $\{a_m\}$ without any knowledge of $\boldsymbol{H}_m$ and $\boldsymbol{v}$. On the other hand, M2 allows completely arbitrary dependence between a_m and $(\boldsymbol{v}, \boldsymbol{H}_m)$ for $m \in \mathcal{I}^{\mathrm{a}}$. This is practical since the type of corruption may depend on the individual measurement $\boldsymbol{v}^{\mathsf{H}}\boldsymbol{H}_m\boldsymbol{v}$ being recorded.

4.2 Problem formulation

Having elaborated on the data corruption model along with the system model in Section 2, the problem of robust PSSE is formally stated next.

Expressed simply, the goal of robust PSSE is to recover all bus voltages $\boldsymbol{v} \in \mathbb{C}^N$ given network parameters $(\boldsymbol{Y}, \boldsymbol{Y}_{\mathrm{f}})$ and the available measurements $\boldsymbol{z} \in \mathbb{R}^M$, whose entries as shown in Eq. (30) satisfy M1 or M2. The first attempt may be seeking the (W)LS estimate using one of the solvers discussed in Section 3. However, it is well known that the (W)LS criterion is sensitive to bad data and can give rise to very bad solutions even if only a few meters are

comprised. On the other hand, the ℓ_1 (least absolute value (LAV)) based losses yielding median-based estimators, have been well documented in statistics and optimization for their ability in handling gross errors in the measurements $\boldsymbol{z}$ (Huber, 2011). This prompts us to consider minimizing the ℓ_1 loss of the residuals, which yields the so-called LAV estimate (Kotiuga and Vidyasagar, 1982; Abur and Celik, 1991; Jabr and Pal, 2004):

$$\min_{\boldsymbol{v}\in\mathbb{C}^N} \ f(\boldsymbol{v}) = \frac{1}{M}\sum_{m=1}^{M} \big|\boldsymbol{v}^{\mathsf{H}}\boldsymbol{H}_m\boldsymbol{v} - z_m\big|. \tag{31}$$

It can be easily checked that the loss function $f(\boldsymbol{v})$ is nonconvex and nonsmooth due to the nonlinear measurements $\{\boldsymbol{v}^{\mathsf{H}}\boldsymbol{H}_m\boldsymbol{v}\}$ and the absolute-value operator $|\cdot|$, respectively. In fact, $f(\boldsymbol{v})$ is not even locally convex near the global optima $\pm\boldsymbol{v}^*$ even in the absence of noise. This is evident from the simplification $f(v) = |v^{\mathsf{H}}v - 1|$ of a scalar variable $v \in \mathbb{C}$. Therefore, a local analysis based on smoothness and convexity is nearly impossible, suggesting that the Gauss-Newton method presented in Section 3.3 is not applicable for minimizing Eq. (31). Due to these reasons, tackling problem (31) is challenging. Upon linearizing $\{\boldsymbol{v}^{\mathsf{H}}\boldsymbol{H}_m\boldsymbol{v}\}$ at the most recent iteration, a sequence of linear programs were solved (Kotiuga and Vidyasagar, 1982). Strategies for improving linear programming by leveraging the system's structure (Abur and Celik, 1991) or via iterative reweighting (Jabr and Pal, 2003, 2004) have been discussed. Despite these efforts, LAV estimators have not been widely employed yet in today's power networks mostly due to their computational inefficiency (Göl and Abur, 2014). However, the criterion $f(\boldsymbol{v})$ exhibits several unique structural properties, which are amenable to developing efficient algorithms as we elaborate on next.

4.3 Deterministic proximal linear algorithm

We start by rewriting the function $f(\boldsymbol{v})$ in Eq. (31) as a composition of the convex $c(\boldsymbol{u}) := \|\boldsymbol{u}\|_1/M$ and the smooth $\boldsymbol{s}(\boldsymbol{v}) : \mathbb{C}^N \to \mathbb{R}^M$ whose per m^{th} entry is $s_m(\boldsymbol{v}) := \boldsymbol{v}^{\mathsf{H}}\boldsymbol{H}_m\boldsymbol{v} - z_m$. This compositional structure lends itself favorably to the proximal linear (prox-linear) algorithms (Fletcher and Watson, 1980; Wang G et al., 2017, 2018c). For interested readers, we provide a brief introduction to the so-called composite optimization and the prox-linear algorithm in Appendix B.

Following Eqs. (B2) and (B3), starting with $\boldsymbol{v}_0 = \mathbf{1}$, the deterministic prox-linear algorithm minimizes Eq. (31) by iteratively solving (Wang G et al., 2017):

$$\boldsymbol{v}_{i+1} = \arg\min_{\boldsymbol{v}\in\mathbb{C}^N} \big\|\Re\big(\boldsymbol{B}_i(\boldsymbol{v}-\boldsymbol{v}_i)\big) - \boldsymbol{c}_i\big\|_1 + \frac{1}{2}\big\|\boldsymbol{v}-\boldsymbol{v}_i\big\|_2^2, \tag{32}$$

where the coefficients are given by

$$\boldsymbol{B}_i := \frac{2\mu_i}{M}\big[\boldsymbol{H}_1\boldsymbol{v}_i,\, \boldsymbol{H}_2\boldsymbol{v}_i, \ldots,\, \boldsymbol{H}_M\boldsymbol{v}_i\big]^{\mathsf{H}}, \tag{33}$$

$$\boldsymbol{c}_i := \frac{\mu_i}{M}\big[z_1 - \boldsymbol{v}_i^{\mathsf{H}}\boldsymbol{H}_1\boldsymbol{v}_i,\, z_2 - \boldsymbol{v}_i^{\mathsf{H}}\boldsymbol{H}_2\boldsymbol{v}_i, \ldots, z_M - \boldsymbol{v}_i^{\mathsf{H}}\boldsymbol{H}_M\boldsymbol{v}_i\big]^{\mathsf{H}}. \tag{34}$$

Observe that problem (32) is a convex quadratic program, which can be solved efficiently by means of standard convex programming methods, including subgradient-type methods (Ben-Tal and Nemirovski, 2001). Under appropriate conditions, the deterministic prox-linear procedure (32) converges quadratically fast to optimum $\boldsymbol{v}^*$ or $-\boldsymbol{v}^*$, meaning that we have to solve only about $\log_2(\log_2(1/\epsilon))$ such quadratic programs to find an ϵ-optimal estimate. This number in practice amounts to 5–8 or so. Alternatively, an alternating direction method of multipliers (ADMM) based solver was developed for iteratively coping with problem (32) (Wang G et al., 2017).

4.4 Stochastic proximal linear alternative

With microgrids becoming increasingly interconnected, seeking the exact minimizer of Eq. (32) per iteration of the deterministic prox-linear scheme may be computationally expensive, or can be intractable. This discourages the applicability of the deterministic prox-linear scheme to robust PSSE of large-scale power networks. In this context, we present an inexpensive stochastic alternative of Eq. (32) for minimizing problem (31) next. Advantages of the stochastic prox-linear approach over its deterministic counterpart Eq. (32) include simple closed-form updates as well as fast convergence to find an (approximately) optimal solution; refer to Duchi and Ruan (2017b) for discussion on stochastic composite optimization.

Instead of relying on all data to obtain the next iteration $\boldsymbol{v}_{i+1}$ by solving the quadratic subproblem (32), the stochastic prox-linear approach samples a

single datum $m_i \in \{1, 2, \ldots, M\}$ uniformly at random per iteration, and inductively constructs $\boldsymbol{v}_{i+1}$ as the optimizer of (Wang G et al., 2017)

$$\min_{\boldsymbol{v}\in\mathbb{C}^N} \left|c_{m_i,i} - \Re\big(\boldsymbol{b}_{m_i,i}^{\mathsf{H}}(\boldsymbol{v}-\boldsymbol{v}_i)\big)\right| + \frac{1}{2\mu_i}\left\|\boldsymbol{v}-\boldsymbol{v}_i\right\|_2^2, \tag{35}$$

where $|c_{m_i,i} - \Re(\boldsymbol{b}_{m_i,i}^{\mathsf{H}}(\boldsymbol{v}-\boldsymbol{v}_i))|$ can be understood as the "linearization" of the ℓ_1 loss $|z_{m_i} - \boldsymbol{v}^{\mathsf{H}}\boldsymbol{H}_{m_i}\boldsymbol{v}|$ of sampled datum $(\boldsymbol{H}_{m_i}, z_{m_i})$ around the current iteration $\boldsymbol{v}_i$, whose coefficients are given by

$$\begin{cases} \boldsymbol{b}_{m_i} := 2\boldsymbol{H}_{m_i}\boldsymbol{v}_i, \\ c_{m_i} := z_{m_i} - \boldsymbol{v}_i^{\mathsf{H}}\boldsymbol{H}_{m_i}\boldsymbol{v}. \end{cases} \tag{36}$$

Again, we obtain a convex quadratic program (35). Compared to the quadratic program (32) encountered in the deterministic prox-linear approach, fortunately, the solution to problem (35) can be provided in simple closed form. To this end, define the projection operator $\mathrm{proj}_\mu(x) : \mathbb{R}\times\mathbb{R}_+ \to \mathbb{R}$ that returns the real number in interval $[-\mu, \mu]$ closest to the given number $x \in \mathbb{R}$. Then we have the solution to problem (35), given by

$$\boldsymbol{v}_{i+1} = \boldsymbol{v}_i + \mathrm{proj}_{\mu_i}\left(\frac{c_{m_i,i} - \Re\big(\boldsymbol{b}_{m_i,i}^{\mathsf{H}}\boldsymbol{v}_i\big)}{\|\boldsymbol{b}_{m_i,i}\|_2^2}\right)\boldsymbol{b}_{m_i,i}, \tag{37}$$

which is repeated until some convergence criteria are met. Intuitively, measurements with a relatively small residual, i.e., $|c_{m_i,i} - \Re(\boldsymbol{b}_{m_i,i}^{\mathsf{H}}\boldsymbol{v}_i)|/\|\boldsymbol{b}_{m_i,i}\|_2^2 \le \mu_i$, are deemed "nominal," for which $\boldsymbol{v}_i$ is updated along the direction of $\boldsymbol{b}_{m_i}$ with a step size of $|c_{m_i,i} - \Re(\boldsymbol{b}_{m_i,i}^{\mathsf{H}}\boldsymbol{v}_i)|/\|\boldsymbol{b}_{m_i,i}\|_2^2$. The measurements of larger residuals obeying $|c_{m_i,i} - \Re(\boldsymbol{b}_{m_i,i}^{\mathsf{H}}\boldsymbol{v}_i)|/\|\boldsymbol{b}_{m_i,i}\|_2^2 \ge \mu_i$, on the other hand, are likely to be outliers, or corrupted by outliers; thus, $\boldsymbol{v}_i$ is updated using a thresholded step size of μ_i.

Regarding the stochastic prox-linear solver (37) for minimizing problem (31), some observations are of interest. In terms of computational complexity, the number of complex scalar operations (e.g., additions and multiplications) can be estimated for Eq. (37). To this end, it is instrumental to note from Eqs. (9)–(11) that the matrices $\{\boldsymbol{H}_m\}_{m=1}^M$ are highly sparse, because both $\boldsymbol{Y}$ and $\boldsymbol{Y}_{\mathrm{f}}$ have only few nonzero entries. Indeed, for power flow or voltage magnitude square measurements, one can verify that the corresponding matrices $\{\boldsymbol{H}_m\}$ have one nonzero entry or three nonzero entries, respectively. As such, evaluating their corresponding coefficients $\{\boldsymbol{b}_m\}$ and $\{c_m\}$ (Eq. (36)) and performing the updates $\boldsymbol{v}_{i+1}$ (Eq. (37)) entail just at most, say, 10 scalar operations. In this case, the stochastic prox-linear procedure (37) incurs per-iteration complexity of $\mathcal{O}(1)$. Surprisingly, this $\mathcal{O}(1)$ complexity holds regardless of the network size N. If the power injection measurements are used too, the number of operations per iteration increases to the order of the number of neighbors, which typically remains much smaller than N in real-world networks.

For convergence, a diminishing step size is usually required for a stochastic optimization algorithm. Similar to other stochastic gradient-type methods, the step size sequence $\{\mu_i\}$ of the stochastic prox-linear procedure should be square summable but not summable (Duchi and Ruan, 2017b); that is,

$$\sum_{i=0}^{+\infty}\mu_i = +\infty, \quad \sum_{i=0}^{+\infty}\mu_i^2 < +\infty. \tag{38}$$

As an example, it suffices to take $\mu_i = \alpha i^{-\gamma}$ with appropriately chosen constants $\alpha > 0$ and $0.5 < \gamma \le 1$. Using the results in Theorem 1 of Duchi and Ruan (2017b), one can establish that the iteration sequence $\{\boldsymbol{v}_i\}$ in Eq. (37) with step sizes satisfying Eq. (38) converges to a stationary point of Eq. (31) almost surely. Fig. 3 illustrates the convergence performance of the Gauss-Newton and deterministic/stochastic prox-linear algorithms on the IEEE 14-bus test system, where the simulated noise-free measurements include all active and reactive power flows, as well as all squared voltage magnitudes.

Finally, PMU measurements, if available, can be readily incorporated in Eqs. (13) and (31). The presented LS and LAV PSSE solvers (mentioned in

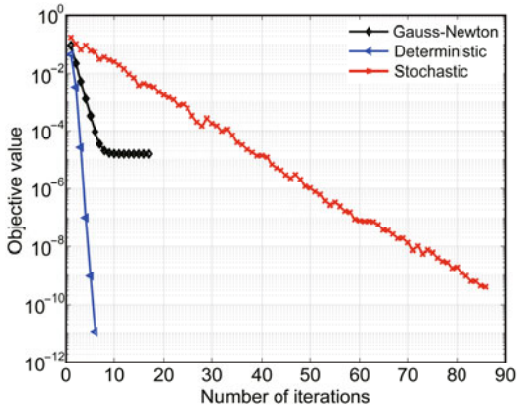

Fig. 3 Convergence performance of Gauss-Newton, deterministic, and stochastic prox-linear algorithms for the IEEE 14-bus system

Sections 3 and 4) apply with no or minimal modifications. Refer to Zamzam et al. (2018) and Zhang et al. (2018a, 2018b, 2019) for generalizations to real-time estimation and forecasting based on deep neural networks.

4.5 Bad data identification

An alternative to coping with bad data is to augment the measurement model (12) entry-wise with an attack variable as (Zhu and Giannakis, 2012)

$$z_m = \boldsymbol{v}^{\mathsf{H}}\boldsymbol{H}_m\boldsymbol{v} + a_m + \epsilon_m,\ \forall\, m = 1, 2, \ldots, M, \quad (39)$$

where $\boldsymbol{a} := [a_1, a_2, \ldots, a_M]^{\mathsf{T}} \in \mathbb{R}^M$ is an unknown vector whose m^{th} entry is deterministically nonzero, only if the m^{th} measurement z_m is attacked (Kosut et al., 2011; Liu et al., 2011; Kekatos and Giannakis, 2013). In practice, the attack vector $\boldsymbol{a}$ is very sparse; namely, most of its entries are zero.

Seeking both $\boldsymbol{v} \in \mathbb{C}^N$ and $\boldsymbol{a} \in \mathbb{R}^M$ from only M measurements in Eq. (39) may appear impossible, as the number of unknowns exceeds the number of equations. Recent results have shown that this is possible by leveraging the parsimony of $\boldsymbol{a}$ (Zhu and Giannakis, 2012; Kekatos and Giannakis, 2013; Aghamolki et al., 2018). If the number $1 \leq k \ll M$ of bad data is known a priori, one would ideally wish to solve

$$\begin{aligned} \min_{\boldsymbol{v}\in\mathbb{C}^N,\, \boldsymbol{a}\in\mathbb{R}^M}\ & \frac{1}{2M}\sum_{m=1}^{M}\left(z_m - \boldsymbol{v}^{\mathsf{H}}\boldsymbol{H}_m\boldsymbol{v} - a_m\right)^2 \\ \text{s.t.}\ & \|\boldsymbol{a}\|_0 \leq k, \end{aligned} \quad (40)$$

where the ℓ_0 (pseudo-)norm $\|\cdot\|_0$ counts the number of nonzero entries in the argument. Due to the combinatorial nature of the ℓ_0 norm, problem (40) is NP-hard in general (Nesterov, 2013). For the special case of $k = 1$, it can be efficiently handled by means of the largest normalized residual (LNR) test; refer to details in Section 5.7 of Abur and Gómez-Expósito (2004) and Mili et al. (1994).

An alternative to the constrained formulation (40) is to replace the ℓ_0 norm with its convex surrogate ℓ_1 norm (Kekatos and Giannakis, 2013):

$$\begin{aligned} \min_{\boldsymbol{v}\in\mathbb{C}^N,\, \boldsymbol{a}\in\mathbb{R}^M}\ & \frac{1}{2M}\sum_{m=1}^{M}\left(z_m - \boldsymbol{v}^{\mathsf{H}}\boldsymbol{H}_m\boldsymbol{v} - a_m\right)^2 \\ \text{s.t.}\ & \|\boldsymbol{a}\|_1 \leq \hat{k}, \end{aligned} \quad (41)$$

where $\hat{k}$ is an estimate (or upper bound) of the practically unknown k. Equivalently, problem (41) can be given in its Lagrangian form:

$$\min_{\substack{\boldsymbol{v}\in\mathbb{C}^N,\\ \boldsymbol{a}\in\mathbb{R}^M}}\ \frac{1}{2M}\sum_{m=1}^{M}\left(z_m - \boldsymbol{v}^{\mathsf{H}}\boldsymbol{H}_m\boldsymbol{v} - a_m\right)^2 + \lambda\|\boldsymbol{a}\|_1, \quad (42)$$

for an appropriately selected regularization parameter $\lambda > 0$. Solving problem (42) offers simultaneously the state estimate $\hat{\boldsymbol{v}}$ and outlier vector $\hat{\boldsymbol{a}}$. In other words, it jointly performs state estimation and bad data identification.

Nonetheless, due to the nonlinear measurements $\{\boldsymbol{v}^*\boldsymbol{H}_m\boldsymbol{v}\}$, problem (42) is still nonconvex. Similar to the LS PSSE in Section 3.4, semidefinite programming relaxation can be invoked for handling problem (42). Specifically, upon introducing $\boldsymbol{V} := \boldsymbol{v}\boldsymbol{v}^{\mathsf{H}}$ and dropping the rank constraint, one can express problem (42) as a convex program over $(\boldsymbol{V}, \boldsymbol{a})$, giving rise to (Zhu and Giannakis, 2012)

$$\begin{aligned} \min_{\substack{\boldsymbol{V}\in\mathbb{C}^{N\times N},\\ \boldsymbol{a}\in\mathbb{R}^M}}\ & \frac{1}{2M}\sum_{m=1}^{M}\left[z_m - \mathrm{tr}\left(\boldsymbol{H}_m\boldsymbol{V}\right) - a_m\right]^2 + \lambda\|\boldsymbol{a}\|_1 \\ \text{s.t.}\ & \boldsymbol{V} \succeq \boldsymbol{0}, \end{aligned} \quad (43)$$

which can be efficiently handled again using standard convex programming approaches. Upon finding its $\boldsymbol{V}$-optimizer $\hat{\boldsymbol{V}}$, state estimate $\hat{\boldsymbol{v}}$ can be recovered using eigen-decomposition or randomization (Zhu and Giannakis, 2012, 2014). Refer to Kekatos and Giannakis (2013) for generalizations using PMU measurements as well as decentralized implementations.

5 Distribution system state estimation

Unlike transmission networks where metering devices are installed at almost all buses, low-voltage distribution grids have partial observability due to limited instrumentation, low industrial investment interest, and the sheer scale (Lu et al., 1995; Baran, 2001). Utilities have been implementing distribution automation systems (DAS) at the substation, and a few on the feeders, which provide not only control and monitoring of devices such as switches and capacitor banks, but also measurements including voltages, currents, and power flows (Baran, 2001). Real-time measurements of distribution networks obtained by DAS offer the possibility of performing SE at a distribution level. Due to limited installation of DAS and a scarcity of real-time measurements, the distribution system state estimation

(DSSE) task that involves estimating the voltage phasors of all buses, now becomes particularly challenging.

We identify the following challenges that must be addressed to design PSSE approaches that are applicable in distribution systems:

1. Unbalanced phases. Distribution systems are multi-phase, consisting mainly of feeders. Feeders are mostly radial, but they have laterals that can be single- or two-phase in general. Furthermore, most loads in residential electricity networks are either single- or two-phase, rather than three-phase. Due to these reasons, distribution systems are unbalanced in nature.

2. High r/x ratios. Due to low-voltage levels as well as relatively short connecting lines, distribution systems possess higher r/x ratios than transmission grids. This situation challenges the convergence of iterative SE algorithms such as Newton-Raphson (Ahmad et al., 2018).

3. Limited availability of real-time data. Real-time data of distribution systems are available only through DAS at very few locations. The most common measurement point is the substation, and probably few on the feeders. The available real-time measurements, however, are presently insufficient to recover the network state, which justifies the partial observability of distribution systems (Bhela et al., 2018; Zamzam et al., 2018).

4. Complex measurement functions. The DAS can provide both phasor measurements and real-valued measurements (Džafić et al., 2018b; Zamzam et al., 2018). The former consists of complex voltages or current flows at selected buses or lines, which are linear functions of the voltage state vector provided by the PMUs or μPMUs. Real-valued measurements, on the other hand, include voltage magnitudes, current flow magnitudes, as well as real and reactive power flows. Similar to the SCADA data in transmission networks, these real-valued data are nonlinear functions of the state vector.

5. Network model uncertainty. In the DSSE framework, network topology and line parameters are typically assumed to be perfectly known. However, due to network infrastructure aging and lack of real-time monitoring of switches (thus uninformed topology changes), the knowledge of a distribution system model has a large uncertainty (Della Giustina et al., 2014; Zhang et al., 2017).

To improve observability in distribution grids, real-time measurements have to be augmented with a high number of pseudo-measurements when performing DSSE (Baran, 2001). Pseudo-measurements comprise predictions of energy consumption or generation, obtained using load and generation forecast procedures based on historical data (Clements, 2011). They are much less accurate than the real-time measurements; thus, the noise variance information should be accounted for in the WLS- and LAV-based DSSE (Sections 3 and 4). In terms of solvers, the algorithmic advances outlined in previous sections can be used for or generalized to DSSE too. Refer to Singh et al. (2009) and Ahmad et al. (2018) for other approaches.

6 Conclusions and future work

In this paper, we have outlined some of the recent advances in PSSE, with a focus on solvers that can efficiently attain (near-)optimal solutions to the nonconvex SE tasks. After developing the Cramér-Rao bound for benchmarking performance of any unbiased estimator, the WLS-based SE has been reviewed. Three efficient solvers have been discussed, including the Gauss-Newton iterations, semidefinite programming relaxation, and feasible point pursuit, all of which were efficiently implemented in the complex domain. To cope with the cyber attacks in the envisioned smart grid, robust PSSE was enabled using the ℓ_1-based losses, for which prox-linear algorithms using composite optimization were advocated. Finally, DSSE along with its current challenges was outlined.

The perspective of this overview opens up a number of exciting directions for future research to realize the vision of smarter power grids, including: (1) generalizing the presented nonconvex optimization approaches to enable dynamic state estimation and tracking; (2) exploring more efficient solvers through, e.g., stochastic, online, parallel, and distributed implementations for PSSE of large-scale networks; (3) leveraging advances in signal processing over networks (graphs) as well as deep learning (e.g., deep neural networks) to provide novel paths to address challenges related to the nonconvexity of PSSE and the partial observability of distribution systems.

References

Abur A, Celik MK, 1991. A fast algorithm for the weighted least-absolute-value state estimation (for power systems). *IEEE Trans Power Syst*, 6(1):1-8. https://doi.org/10.1109/59.131040

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

Aghamolki HG, Miao Z, Fan L, 2018. SOCP convex relaxation-based simultaneous state estimation and bad data identification. https://arxiv.org/abs/1804.05130

Ahmad F, Rasool A, Ozsoy E, et al., 2018. Distribution system state estimation—a step towards smart grid. *Renew Sust Energ Rev*, 81:2659-2671. https://doi.org/10.1016/j.rser.2017.06.071

Baran ME, 2001. Challenges in state estimation on distribution systems. Power Engineering Society Summer Meeting, p.429-433. https://doi.org/10.1109/PESS.2001.970062

Ben-Tal A, Nemirovski A, 2001. Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications. SIAM, Philadelphia, USA.

Bertsekas DP, 1999. Nonlinear Programming. Athena Scientific, Belmont, Massachusetts, USA.

Bhela S, Kekatos V, Veeramachaneni S, 2018. Enhancing observability in distribution grids using smart meter data. *IEEE Trans Smart Grid*, 9(6):5953-5961. https://doi.org/10.1109/TSG.2017.2699939

Burke JV, Ferris MC, 1995. A Gauss-Newton method for convex composite optimization. *Math Programm*, 71(2):179-194. https://doi.org/10.1007/BF01585997

Candès EJ, Li X, Soltanolkotabi M, 2015. Phase retrieval via Wirtinger flow: theory and algorithms. *IEEE Trans Inform Theory*, 61(4):1985-2007. https://doi.org/10.1109/TIT.2015.2399924

Caro E, Conejo A, 2012. State estimation via mathematical programming: a comparison of different estimation algorithms. *IET Gener Transm Distrib*, 6(6):545-553. https://doi.org/10.1049/iet-gtd.2011.0663

Christie RD, 1999. Power Systems Test Case Archive. University of Washington. https://labs.ece.uw.edu/pstca/

Clements KA, 2011. The impact of pseudo-measurements on state estimator accuracy. IEEE Power and Energy Society General Meeting, p.1-4. https://doi.org/10.1109/PES.2011.6039370

Della Giustina D, Pau M, Pegoraro PA, et al., 2014. Electrical distribution system state estimation: measurement issues and challenges. *IEEE Instrum Meas Mag*, 17(6):36-42. https://doi.org/10.1109/MIM.2014.6968929

Duchi JC, Ruan F, 2017a. Solving (most) of a set of quadratic equalities: composite optimization for robust phase retrieval. *Inform Infer J IMA*, iay015. https://doi.org/10.1093/imaiai/iay015

Duchi JC, Ruan F, 2017b. Stochastic methods for composite optimization problems. https://arxiv.org/abs/1703.08570

Džafić I, Jabr RA, Hrnjić T, 2018a. High performance distribution network power flow using Wirtinger calculus. *IEEE Trans Smart Grid*, in press. https://doi.org/10.1109/TSG.2018.2824018

Džafić I, Jabr RA, Hrnjić T, 2018b. Hybrid state estimation in complex variables. *IEEE Trans Power Syst*, 33(5):5288-5296. https://doi.org/10.1109/TPWRS.2018.2794401

Fairley P, 2016. Cybersecurity at US utilities due for an upgrade: tech to detect intrusions into industrial control systems will be mandatory. *IEEE Spectr*, 53(5):11-13. https://doi.org/10.1109/MSPEC.2016.7459104

Fletcher R, Watson GA, 1980. First and second order conditions for a class of nondifferentiable optimization problems. *Math Programm*, 18(1):291-307. https://doi.org/10.1007/BF01588325

Giannakis GB, Kekatos V, Gatsis N, et al., 2013. Monitoring and optimization for power grids: a signal processing perspective. *IEEE Signal Process Mag*, 30(5):107-128. https://doi.org/10.1109/MSP.2013.2245726

Göl M, Abur A, 2014. LAV based robust state estimation for systems measured by PMUs. *IEEE Trans Smart Grid*, 5(4):1808-1814. https://doi.org/10.1109/TSG.2014.2302213

Huang YF, Werner S, Huang J, et al., 2012. State estimation in electric power grids: meeting new challenges presented by the requirements of the future grid. *IEEE Signal Process Mag*, 29(5):33-43. https://doi.org/10.1109/MSP.2012.2187037

Huber PJ, 2011. Robust Statistics. In: Lovric M (Ed.), International Encyclopedia of Statistical Science. Springer, Berlin, p.1248-1251.

Jabr R, Pal B, 2003. Iteratively re-weighted least-absolute-value method for state estimation. *IET Gener Transm Distrib*, 150(4):385-391. https://doi.org/10.1049/ip-gtd:20030462

Jabr R, Pal B, 2004. Iteratively reweighted least-squares implementation of the WLAV state-estimation method. *IET Gener Transm Distrib*, 151(1):103-108. https://doi.org/10.1049/ip-gtd:20040030

Kay SM, 1993. Fundamentals of Statistical Signal Processing, Vol. I: Estimation Theory. Prentice Hall, Englewood Cliffs, USA.

Kekatos V, Giannakis GB, 2013. Distributed robust power system state estimation. *IEEE Trans Power Syst*, 28(2):1617-1626. https://doi.org/10.1109/TPWRS.2012.2219629

Kekatos V, Wang G, Zhu H, et al., 2017. PSSE redux: convex relaxation, decentralized, robust, and dynamic approaches. https://arxiv.org/abs/1708.03981

Kim SJ, Wang G, Giannakis GB, 2014. Online semidefinite programming for power system state estimation. IEEE Conf on Acoustics, Speech, and Signal Process, p.6024-6027. https://doi.org/10.1109/ICASSP.2014.6854760

Kosut O, Jia L, Thomas J, et al., 2011. Malicious data attacks on the smart grid. *IEEE Trans Smart Grid*, 2(4):645-658. https://doi.org/10.1109/TSG.2011.2163807

Kotiuga WW, Vidyasagar M, 1982. Bad data rejection properties of weighted least-absolute-value techniques applied to static state estimation. *IEEE Trans Power Appar Syst*, 101(4):844-853. https://doi.org/10.1109/TPAS.1982.317150

Kreutz-Delgado K, 2009. The complex gradient operator and the CR-calculus. https://arxiv.org/abs/0906.4835

Lewis AS, Wright SJ, 2016. A proximal method for composite minimization. *Math Programm*, 158(1-2):501-546. https://doi.org/10.1007/s10107-015-0943-9

Liu Y, Ning P, Reiter MK, 2011. False data injection attacks against state estimation in electric power grids. *ACM Trans Inform Syst Sec*, 14(1):1-33. https://doi.org/10.1145/1952982.1952995

Lu C, Teng J, Liu WH, 1995. Distribution system state estimation. *IEEE Trans Power Syst*, 10(1):229-240. https://doi.org/10.1109/59.373946

Mehanna O, Huang K, Gopalakrishnan B, et al., 2015. Feasible point pursuit and successive approximation of nonconvex QCQPs. *IEEE Signal Process Lett*, 22(7):804-808. https://doi.org/10.1109/LSP.2014.2370033

Mili L, Cheniae MG, Rousseeuw PJ, 1994. Robust state estimation of electric power systems. *IEEE Trans Circ Syst I Fundam Theory Appl*, 41(5):349-358. https://doi.org/10.1109/81.296336

Monticelli A, 2000. Electric power system state estimation. *Proc IEEE*, 88(2):262-282. https://doi.org/10.1109/5.824004

Nesterov Y, 2013. Introductory Lectures on Convex Optimization: a Basic Course. Springer Science & Business Media, Boston, USA.

Pardalos PM, Vavasis SA, 1991. Quadratic programming with one negative eigenvalue is NP-hard. *J Glob Optim*, 1(1):15-22. https://doi.org/10.1007/BF00120662

Park J, Boyd S, 2017. General heuristics for nonconvex quadratically constrained quadratic programming. https://arxiv.org/abs/1703.07870

Saad Y, 2003. Iterative Methods for Sparse Linear Systems (2nd Ed.). Society for Industrial and Applied Mathematics, Philadelphia, USA.

Schweppe FC, Wildes J, Rom D, 1970. Power system static state estimation: parts I, II, and III. *IEEE Trans Power Appar Syst*, 89(1):120-135.

Singh R, Pal B, Jabr R, 2009. Choice of estimator for distribution system state estimation. *IET Gener Transm Distrib*, 3(7):666-678. https://doi.org/10.1049/iet-gtd.2008.0485

Stoica P, Marzetta TL, 2001. Parameter estimation problems with singular information matrices. *IEEE Trans Signal Process*, 49(1):87-90. https://doi.org/10.1109/78.890346

Wang G, Kim SJ, Giannakis GB, 2014. Moving-horizon dynamic power system state estimation using semidefinite relaxation. IEEE PES General Meeting & Conf Exposition, p.1-5. https://doi.org/10.1109/PESGM.2014.6939925

Wang G, Zamzam AS, Giannakis GB, et al., 2016. Power system state estimation via feasible point pursuit. IEEE Global Conf Signal and Information Process, p.773-777. https://doi.org/10.1109/GlobalSIP.2016.7905947

Wang G, Giannakis GB, Chen J, 2017. Robust and scalable power system state estimation via composite optimization. https://arxiv.org/abs/1708.06013

Wang G, Giannakis GB, Saad Y, et al., 2018a. Phase retrieval via reweighted amplitude flow. *IEEE Trans Signal Process*, 66(11):2818-2833. https://doi.org/10.1109/TSP.2018.2818077

Wang G, Zamzam AS, Giannakis GB, et al., 2018b. Power system state estimation via feasible point pursuit: algorithms and Cramér-Rao bound. *IEEE Trans Signal Process*, 66(6):1649-1658. https://doi.org/10.1109/TSP.2018.2791977

Wang G, Zhu H, Giannakis GB, et al., 2018c. Robust power system state estimation from rank-one measurements. *IEEE Trans Contr Netw Syst*, in press. https://doi.org/10.1109/TCNS.2019.2890954

Wang G, Giannakis GB, Eldar YC, 2018d. Solving systems of random quadratic equations via truncated amplitude flow. *IEEE Trans Inform Theory*, 64(2):773-794. https://doi.org/10.1109/TIT.2017.2756858

Wang Z, Cui B, Wang J, 2017. A necessary condition for power flow insolvability in power distribution systems with distributed generators. *IEEE Trans Power Syst*, 32(2):1440-1450. https://doi.org/10.1109/TPWRS.2016.2588341

Wood AJ, Wollenberg BF, 1996. Power Generation, Operation, and Control (2nd Ed.). Wiley & Sons, New York, USA.

Wulf WA, 2000. Great achievements and grand challenges. *Nat Acad Eng*, 30(1):5-10.

Zamzam AS, Fu X, Sidiropoulos ND, 2018. Data-driven learning-based optimization for distribution system state estimation. https://arxiv.org/abs/1807.01671

Zhang L, Wang G, Giannakis GB, 2017. Going beyond linear dependencies to unveil connectivity of meshed grids. IEEE 7th Workshop on Computational Advances in Multi-sensor Adaptive Processing, p.1-5. https://doi.org/10.1109/CAMSAP.2017.8313078

Zhang L, Wang G, Giannakis GB, 2018a. Real-time power system state estimation via deep unrolled neural networks. IEEE Global Conf on Signal and Information Processing, in press.

Zhang L, Wang G, Giannakis GB, 2018b. Real-time power system state estimation and forecasting via deep neural networks. https://arxiv.org/abs/1811.06146

Zhang L, Wang G, Giannakis GB, 2019. Power system state forecasting via deep recurrent neural networks. IEEE Conf on Acoustics, Speech, and Signal Process, in press.

Zhu H, Giannakis GB, 2011. Estimating the state of AC power systems using semidefinite programming. North American Power Symp, p.1-7. https://doi.org/10.1109/NAPS.2011.6024862

Zhu H, Giannakis GB, 2012. Robust power system state estimation for the nonlinear AC flow model. North American Power Symp, p.1-6. https://doi.org/10.1109/NAPS.2012.6336405

Zhu H, Giannakis GB, 2014. Power system nonlinear state estimation using distributed semidefinite programming. *IEEE J Sel Top Signal Process*, 8(6):1039-1050. https://doi.org/10.1109/JSTSP.2014.2331033

Appendix A: Wirtinger's calculus

Introducing the complex conjugate coordinates $[\boldsymbol{v}^{\mathsf{T}}\ \overline{\boldsymbol{v}}^{\mathsf{T}}]^{\mathsf{T}} \in \mathbb{C}^{2N}$, one can express $\boldsymbol{h}(\boldsymbol{v}) = \boldsymbol{h}(\boldsymbol{v}, \overline{\boldsymbol{v}}) \in \mathbb{R}^{M}$. It now becomes evident that $\boldsymbol{h}(\boldsymbol{v}, \overline{\boldsymbol{v}})$ is holomorphic (i.e., complex differentiable) in $\boldsymbol{v}$ for a fixed

$\overline{\boldsymbol{v}}$, and vice versa. Following the convention, (partial) derivatives will be denoted by row vectors, while (sub)gradients are denoted by column vectors. The partial Wirtinger derivatives are given by (Kreutz-Delgado, 2009)

$$\frac{\partial h_m}{\partial \boldsymbol{v}} := \frac{\partial h_m(\boldsymbol{v},\overline{\boldsymbol{v}})}{\partial \boldsymbol{v}}\bigg|_{\overline{\boldsymbol{v}}=\text{const.}} = \left[\frac{\partial h_m}{\partial v_1}, \frac{\partial h_m}{\partial v_2}, \ldots, \frac{\partial h_m}{\partial v_N}\right], \tag{A1}$$

$$\frac{\partial h_m}{\partial \overline{\boldsymbol{v}}} := \frac{\partial h_m(\boldsymbol{v},\overline{\boldsymbol{v}})}{\partial \overline{\boldsymbol{v}}}\bigg|_{\boldsymbol{v}=\text{const.}} = \left[\frac{\partial h_m}{\partial \overline{v}_1}, \frac{\partial h_m}{\partial \overline{v}_2}, \ldots, \frac{\partial h_m}{\partial \overline{v}_N}\right], \tag{A2}$$

for $m = 1, 2, \ldots, M$, where the partial derivative with respect to $\boldsymbol{v}$ ($\overline{\boldsymbol{v}}$) treats $\overline{\boldsymbol{v}}$ ($\boldsymbol{v}$) as a constant in h_m. The complex gradient of $h_m(\boldsymbol{v},\overline{\boldsymbol{v}})$ with respect to $\boldsymbol{v}$ or $\overline{\boldsymbol{v}}$ can be defined by

$$\nabla_{\boldsymbol{v}} h_m := \left(\frac{\partial h_m}{\partial \boldsymbol{v}}\right)^{\mathsf{H}}, \quad \nabla_{\overline{\boldsymbol{v}}} h_m := \left(\frac{\partial h_m}{\partial \overline{\boldsymbol{v}}}\right)^{\mathsf{H}}, \tag{A3}$$

which gives rise to the complex gradient of h_m in the conjugate coordinate system:

$$\nabla_{\mathrm{c}} h_m := \left[\nabla_{\boldsymbol{v}}^{\mathsf{T}} h_m \ \nabla_{\overline{\boldsymbol{v}}}^{\mathsf{T}} h_m\right]^{\mathsf{T}} = \left[\frac{\partial h_m}{\partial \boldsymbol{v}} \ \frac{\partial h_m}{\partial \overline{\boldsymbol{v}}}\right]^{\mathsf{H}}. \tag{A4}$$

Upon introducing the so-called complex Jacobian matrix:

$$\nabla_{\mathrm{c}} \boldsymbol{h} := [\nabla_{\mathrm{c}} h_1, \nabla_{\mathrm{c}} h_2, \ldots, \nabla_{\mathrm{c}} h_M] \in \mathbb{C}^{2N\times M}, \tag{A5}$$

the first-order Taylor expansion of $\boldsymbol{h}(\boldsymbol{v}+\Delta\boldsymbol{v})$ for given vectors $\boldsymbol{v}$ and $\Delta\boldsymbol{v} \in \mathbb{C}^N$ is defined as

$$\begin{aligned}\boldsymbol{h}(\boldsymbol{v}+\Delta\boldsymbol{v}) &\approx \boldsymbol{h}(\boldsymbol{v}) + \nabla_{\mathrm{c}}^{\mathsf{H}} \boldsymbol{h}(\boldsymbol{v}) \begin{bmatrix} \Delta\boldsymbol{v} \\ \Delta\boldsymbol{v} \end{bmatrix} \\ &= \boldsymbol{h}(\boldsymbol{v}) + 2\Re\big(\nabla_{\boldsymbol{v}}^{\mathsf{H}} \boldsymbol{h}(\boldsymbol{v})\Delta\boldsymbol{v}\big).\end{aligned} \tag{A6}$$

Appendix B: Composite optimization

Consider minimizing functions of the form:

$$f(\boldsymbol{v}) := c(\boldsymbol{s}(\boldsymbol{v})) \quad \text{s.t.} \quad \boldsymbol{v} \in \mathcal{V}, \tag{B1}$$

where $c(\cdot) : \mathbb{R}^M \in \mathbb{R}$ is convex, $\boldsymbol{s}(\cdot) : \mathbb{C}^N \to \mathbb{R}^M$ is smooth, and $\mathcal{V}$ is some convex set (or $\mathcal{V} = \mathbb{C}^n$ if there is no constraint on $\boldsymbol{v}$). Since $f(\cdot)$ ($\boldsymbol{s}(\cdot)$) is a real-valued function of complex arguments, real analysis is not applicable. Yet, this is not an issue, as Wirtinger's calculus in Appendix A can be employed. This compositional structure lends itself favorably to the prox-linear algorithm (Fletcher and Watson, 1980; Burke and Ferris, 1995), which we develop next.

Similar to other iterative schemes such as gradient descent, trust-region, and Newton's method, the prox-linear algorithm builds a local model of the loss function and repeatedly minimizes this model. However, thanks to the compositional structure, the local model is obtained by linearizing only $\boldsymbol{s}(\cdot)$. Specifically, we first define close to any given point $\boldsymbol{w} \in \mathbb{C}^N$ a local "linearization" of $f(\cdot)$ as

$$f_{\boldsymbol{w}}(\boldsymbol{v}) := c\big(\boldsymbol{s}(\boldsymbol{w}) + 2\Re\big(\nabla_{\boldsymbol{v}}^{\mathsf{H}} \boldsymbol{s}(\boldsymbol{w})(\boldsymbol{v}-\boldsymbol{w})\big)\big), \tag{B2}$$

where $\nabla_{\boldsymbol{v}} \boldsymbol{s}(\boldsymbol{v}) \in \mathbb{C}^{N\times M}$ is the complex Jacobian matrix of $\boldsymbol{s}(\cdot)$ evaluated at point $\boldsymbol{w}$ (Eqs. (A5) and (A6) in Appendix A).

In contrast to the originally nonconvex $f(\boldsymbol{v})$, the linearization $f_{\boldsymbol{w}}(\boldsymbol{v})$ in Eq. (B2) becomes convex, which is indeed the key behind the prox-linear procedure. Starting with some point $\boldsymbol{v}_0 \in \mathbb{C}^N$, the (deterministic) prox-linear algorithm proceeds inductively to obtain iterations $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots\}$ by minimizing the quadratically regularized models (Burke and Ferris, 1995; Lewis and Wright, 2016):

$$\boldsymbol{v}_{i+1} = \arg\min_{\boldsymbol{v}\in\mathcal{X}} \left\{ f_{\boldsymbol{v}_i}(\boldsymbol{v}) + \frac{1}{2\mu_i}\big\|\boldsymbol{v}-\boldsymbol{v}_i\big\|_2^2 \right\}, \tag{B3}$$

where $\mu_i > 0$ is a step size that can be fixed a priori to some constant, or be determined "on-the-fly" by a line search (Burke and Ferris, 1995).

Observe that $f_{\boldsymbol{v}_i}(\boldsymbol{v})$ is convex in $\boldsymbol{v}$, so is problem (B3). It has been shown that if $c(\cdot)$ is L-Lipschitz and $\nabla\boldsymbol{s}$ is κ-Lipschitz, choosing any constant step size $0 < \mu < 1/(\kappa L)$ ensures that the algorithm (Eq. (B3)) (Lewis and Wright, 2016): (1) is a descent method (i.e., the iterations $\{\boldsymbol{v}_i\}$ monotonically decrease the function value of $f(\boldsymbol{v})$); (2) finds an (approximate) stationary point of Eq. (B1). Interested readers can refer to Lewis and Wright (2016) for a contemporary review on composite optimization.