Power System State Estimation via Feasible Point Pursuit: Algorithms and Cramér-Rao Bound

Gang Wang, *Student Member, IEEE,* Ahmed S. Zamzam, *Student Member, IEEE,*
Georgios B. Giannakis, *Fellow, IEEE,* and Nicholas D. Sidiropoulos, *Fellow, IEEE*

***Abstract*—Accurately monitoring the system's operating point is central to the reliable and economic operation of an autonomous energy grid. Power system state estimation (PSSE) aims to obtain complete voltage magnitude and angle information at each bus given a number of system variables at selected buses and lines. Power flow analysis is a special case of PSSE, and amounts to solving a set of noise-free power flow equations. Physical laws dictate quadratic relationships between available quantities and unknown voltages, rendering general instances of power flow and PSSE nonconvex and NP-hard. Past approaches are largely based on gradient-type iterative procedures or semidefinite relaxation (SDR). Due to nonconvexity, the solution obtained via gradient-type schemes depends on initialization, while SDR methods do not perform as desired in challenging scenarios. This paper puts forth novel *feasible point pursuit* (FPP)-based solvers for power flow and PSSE, which iteratively seek feasible solutions for a nonconvex quadratically constrained quadratic programming (QCQP) reformulation of the weighted least-squares (WLS) problem. Relative to the prior art, the developed solvers offer superior performance at the cost of higher complexity. Furthermore, they converge to a stationary point of the WLS problem. As a baseline for comparing different estimators, the Cramér-Rao lower bound (CRLB) is derived for the fundamental PSSE problem in this paper. Judicious numerical tests on several IEEE benchmark systems showcase markedly improved performance of our FPP-based solvers for both power flow and PSSE tasks over popular WLS-based Gauss-Newton iterations and SDR approaches.**

***Index terms*—** Power flow analysis, state estimation, nonconvex QCQP, feasible point pursuit, Cramér-Rao lower bound, autonomous energy grid.

Manuscript received May 11, 2017; revised September 10, 2017 and November 19, 2017; accepted December 25, 2017. Date of publication November 8, 2017; date of current version xxx, 2018. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Ami Wiesel. The work of G. Wang and G. B. Giannakis was supported in part by NSF grants 1423316, 1442686, 1508993, and 1509040. The work of A. S. Zamzam and N. D. Sidiropoulos was supported in part by NSF grants 1231504 and 1525194. This paper was presented in part at the 2016 IEEE Global Conference on Signal and Information Processing, Washington, DC, USA, December 79, 2016. (*Corresponding author: Georgios B. Giannakis.*)

G. Wang is with the Digital Technology Center and the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA, and also with the State Key Laboratory of Intelligent Control and Decision of Complex Systems, Beijing Institute of Technology, Beijing 100081, China (e-mail: gangwang@umn.edu).

A. S. Zamzam 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: ahmedz@umn.edu; georgios@umn.edu).

N. D. Sidiropoulos was with the Digital Technology Center and the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455, USA, and is now with the Department of Electrical and Computer Engineering, University of Virginia, Charlottesville, Virginia 22904. (e-mail: nikos@virginia.edu).

I. Introduction

Recognized as the greatest engineering achievement of the twentieth century [1], the electric power grid is a complex cyber-physical system comprising multiple subsystems, each with a transmission infrastructure to deliver electricity from power generators to distribution networks to customers. Accurately monitoring the operational condition of a power grid is crucial to various system control and optimization tasks, which include unit commitment, optimal power flow (OPF), and economic dispatch [2], [3]. To enable such an accurate monitoring, a set of system variables are specified (and enforced) or measured at selected buses and lines for determining or estimating the system's operating point, namely complex voltages at all buses of the grid. These two tasks correspond to the so-termed power flow analysis and power system state estimation (PSSE), respectively. Both are central to monitoring, control, and future planning of electricity networks.

In power engineering, power flow analysis is a numerical analysis of the normal steady-state flow of electric power over the grid, that is crucial for planning future power system expansions (e.g., designing components such as generators, lines, transformers, and capacitors), as well as in determining the best operation of the existing systems [4]. The goal of power flow analysis is to obtain complete voltage magnitude and angle information at each bus for specified or enforced load and generator active power and voltage conditions [4]. Once this information is available, other system variables including active and reactive power flows as well as generator reactive power outputs can be analytically obtained.

Power flow analysis amounts to solving a set of quadratic equations given by the nonlinear AC power flow model obeying Ohm's and Kirchhoff's laws. Solving power flow equations for both transmission and distribution systems is known to be NP-hard [5]. Due to the nonlinear nature, several numerical solvers have been developed to obtain a solution that is within an acceptable tolerance. Past solvers include the Gauss-Seidel and Newton-Raphson iterative algorithms [4], a non-iterative moment-based approach [6], and the semidefinite relaxation (SDR) [7]. The Gauss-Seidel method is reported as the earliest devised power flow solver [4]. On the other hand, the Newton-Raphson algorithm iteratively seeks improved approximations to the zeros of real-valued functions, featuring quadratic convergence whenever the initial point lands within a small neighborhood of the zeros [8]. As convergence of both algorithms relies heavily on the initial point, they may diverge

if the initialization is not reliable [7]. With a carefully designed objective function and sufficiently small angle differences across lines, the SDR approaches have been shown capable of recovering the true power flow solution provided that the set of available specifications includes all voltage magnitudes, and the active power flows over a spanning tree of the network [7].

The task of PSSE can be described as estimating the voltage magnitudes and angles at all buses across the network from a subset of supervisory control and data acquisition (SCADA) measurements including active and reactive power injections and flows (at both the sending and receiving ends), as well as squared voltage magnitudes. Since its appearance in the 1970s [9], PSSE has become a prerequisite for supervisory control, system planning, and economic dispatch [9]. Nonlinear SCADA measurements however, render the PSSE problem nonconvex and NP-hard in general [10].

PSSE solvers so far are largely based on Gauss-Newton iterations and SDR heuristics. The "workhorse" Gauss-Newton method for nonconvex optimization has two limitations [11, Sec. 1.5], i.e., sensitivity to the initial guess, and lack of convergence guarantees. SDR-based approaches on the other hand solve first for a matrix variable that can be computationally expensive [7], [10], [12]–[14]. SDR's performance degrades when the data-size is relatively small, or when the data do not include all voltage magnitudes [7]. For PSSE of large-scale networks, robust and distributed Gauss-Newton and SDR implementations have been reported in [10], [14], [15], [16], [17], [18].

Solving power flow equations and the PSSE can be shown equivalent to solving nonconvex QCQPs, which in its general form is NP-hard [19]. Many heuristics have recently been put forward [20], [21], [22], [23], [24], [25]. A feasible point pursuit (FPP) algorithm developed in [21], which can be viewed as a special case of the general convex-concave procedure in [23]–[25], was demonstrated to enjoy improved numerical performance over the SDR-based methods. The FPP heuristic has been employed for solving OPF [26], where the resulting solver was empirically shown more effective for multi-phase transmission networks than popular SDR- and moment relaxation-based ones [26].

Building on our precursors [21], [27] and inspired by the inherent nonconvex challenge, the objective of this work is to develop power flow and PSSE solvers capable of attaining or approximating the global optimum at manageable computational complexity. Starting with the WLS formulation, both power flow and PSSE tasks are reformulated as nonconvex QCQPs, which are tackled by FPP. We show that every KKT point of the resulting QCQP is a stationary point of the WLS problem, and further that our FPP-based solvers converge to a stationary point of the WLS. As a baseline for comparing different SE approaches, the Cramér-Rao lower bound (CRLB) is for the first time derived for the fundamental PSSE problem assuming additive white Gaussian noise (AWGN). This is achieved by means of Wirtinger's calculus for functional analysis over complex domains. Finally, numerical experiments using several IEEE benchmark systems corroborate the superior performance of FPP-based solvers over existing methods for both power flow and PSSE tasks.

Regarding notation, matrices (vectors) are denoted by upper- (lower-) case boldface letters, and $\overline{(\cdot)}$, $(\cdot)^{\mathcal{T}}$, and $(\cdot)^{\mathcal{H}}$ stand for complex conjugate, transpose, and conjugate-transpose, respectively. Calligraphic letters are reserved for sets, e.g., $\mathcal{N}$. Symbol $\Re\{\cdot\}$ ($\Im\{\cdot\}$) takes the real (imaginary) part of a complex-valued object, and $\mathrm{diag}(\boldsymbol{x})$ is a diagonal matrix holding in order entries of $\boldsymbol{x}$ on its diagonal.

II. System Modeling and Problem Statement

An electric transmission network having N nodes (buses) and E edges (lines) can be represented by a graph $\mathcal{G} := \{\mathcal{N}, \mathcal{E}\}$, whose nodes $\mathcal{N} := \{1, 2, \ldots, N\}$ correspond to buses, and whose edges $\mathcal{E} := \{(m, n)\} \subseteq \mathcal{N} \times \mathcal{N}$ correspond to transmission lines. For every bus $n \in \mathcal{N}$, let $V_n := |V_n|e^{j\theta_n}$ be the nodal complex voltage, whose magnitude and phase are given by $|V_n|$ and θ_n, respectively; likewise for the complex current injection $I_n := |I_n|e^{j\phi_n}$. Let also $S_n := P_n + jQ_n$ be the corresponding complex power injection, in which P_n and Q_n are the active and reactive power injection, respectively. For every line $(m, n) \in \mathcal{E}$, let I_{mn} denote the complex current flowing from bus m to n, and $S^f_{mn} := P^f_{mn} + jQ^f_{mn}$ the complex power flow from bus m to n seen at the sending end, where P^f_{mn} and Q^f_{mn} are the active and reactive power flow, respectively; and likewise for the receiving-end (active and reactive) power flow P^t_{mn} and Q^t_{mn}.

The AC power flow model dictates that system variables $\{P_n\}$, $\{Q_n\}$, $\{P^f_{mn}\}$, $\{Q^f_{mn}\}$, $\{P^{\mathcal{T}}_{mn}\}$, $\{Q^{\mathcal{T}}_{mn}\}$, and $\{|V_n|^2\}$ are quadratic functions of the state vector $\boldsymbol{v}$. Clearly, this holds true for the squared voltage magnitude understood as $|V_n|^2 = V_n\overline{V}_n$. To specify the relationship between power quantities and $\boldsymbol{v}$, introduce $\boldsymbol{Y} \in \mathbb{C}^{N\times N}$ to represent the bus admittance matrix, which is in general symmetric. Ohm's law in conjunction with Kirchhoff's law reads as

$$\boldsymbol{i} = \boldsymbol{Y}\boldsymbol{v}. \tag{1}$$

It is worth mentioning that $\boldsymbol{Y}$ is sparse, thus enabling efficient computations in large-size power networks, and its (m, n)-th entry is given by

$$Y_{mn} := \begin{cases} -y_{mn}, & (m,n) \in \mathcal{E} \\ y^g_{nn} + \sum_{k\in\mathcal{N}_n} y_{nk}, & m = n \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

where y_{mn} denotes the admittance of line $(m, n) \in \mathcal{E}$, y^g_{nn} the admittance to the ground at bus $n \in \mathcal{N}$, and $\mathcal{N}_n$ the set of neighboring buses directly connected to bus n. For $m \neq n$, let y^s_{mn} be the shunt admittance at bus m associated with line (m, n). Recall from Ohm's and Kirchhoff's laws that the current flowing from bus m to n can be expressed as

$$I_{mn} = y^s_{mn}V_m + y_{mn}(V_m - V_n) \tag{3}$$

whereby the reverse-direction current I_{nm} can be given symmetrically. Due to $y^s_{mn} \neq 0$ in general, it holds $I_{mn} \neq -I_{nm}$.

The AC model also asserts $P_n + jQ_n = V_n\overline{I}_n$, $\forall n \in \mathcal{N}$. Appealing again to (1) leads to the next matrix-vector form

$$\boldsymbol{p} + j\boldsymbol{q} = \mathrm{diag}(\boldsymbol{v})\overline{\boldsymbol{i}} = \mathrm{diag}(\boldsymbol{v})\overline{\boldsymbol{Y}}\,\overline{\boldsymbol{v}} \tag{4}$$

where both active and reactive power injections are quadratically related to $\boldsymbol{v}$. Likewise, the sending-end active and reactive power flow over line $(m,n)\in\mathcal{E}$ can be written as

$$\begin{aligned} P^f_{mn}+jQ^f_{mn} &= V_m\overline{I}_{mn} \\ &= (y^s_{mn}+\overline{y}^s_{mn})\,V_m\overline{V}_m-\overline{y}_{mn}V_m\overline{V}_n \end{aligned} \tag{5}$$

where the second equality is obtained by substituting I_{mn} in (3) into the first. Hence, P^f_{mn} and Q^f_{mn} can also be expressible as quadratic functions of $\boldsymbol{v}$. By symmetry, this quadratic relationship also holds for P^t_{mn} and Q^t_{mn}.

To perform either power flow analysis or PSSE, a total of L system variables are specified or measured by the system operator. The nonlinear AC networks have available the next seven types of quantities: $|V_n|^2$, P_n, Q_n, P^f_{mn}, Q^f_{mn}, P^t_{mn}, and Q^t_{mn}. If $\mathcal{N}_V$, $\mathcal{N}_P$, $\mathcal{N}_Q$, $\mathcal{E}^f_P$ ($\mathcal{E}^f_Q$), and $\mathcal{E}^t_P$ ($\mathcal{E}^t_Q$) denote the selected sets of buses/lines over which actual quantities of the corresponding type are available, the elaborated quadratic relationships prompt us to define the $L\times 1$ data vector $\boldsymbol{z}:=\big[\{|V_n|^2\}_{n\in\mathcal{N}_V},\ \{P_n\}_{n\in\mathcal{N}_P},\ \{Q_n\}_{n\in\mathcal{N}_Q},\ \{P^f_{mn}\}_{(m,n)\in\mathcal{E}^f_P},\ \{Q^f_{mn}\}_{(m,n)\in\mathcal{E}^f_Q},\ \{P^t_{mn}\}_{(m,n)\in\mathcal{E}^t_P},\ \{Q^t_{mn}\}_{(m,n)\in\mathcal{E}^t_Q}\big]^{\mathcal{T}}\in\mathbb{R}^L$, whose entries can be succinctly given by

$$z_\ell=\boldsymbol{v}^{\mathcal{H}}\boldsymbol{H}_\ell\boldsymbol{v},\quad 1\le\ell\le L \tag{6}$$

where $\{\boldsymbol{H}_\ell\}_{\ell=1}^L$ are some coefficient matrices to be specified. For this purpose, let $\{\boldsymbol{e}_n\in\mathbb{R}^N\}_{n=1}^N$ be the canonical basis of $\mathbb{R}^N$, and introduce also the admittance-dependent matrices

$$\begin{aligned} \boldsymbol{Y}_n &:= \boldsymbol{e}_n\boldsymbol{e}_n^{\mathcal{T}}\boldsymbol{Y}, && \forall n\in\mathcal{N},\\ \boldsymbol{Y}^f_{mn} &:= (\overline{y}_{mn}+y_{mn})\boldsymbol{e}_m\boldsymbol{e}_m^{\mathcal{T}}-y_{mn}\boldsymbol{e}_m\boldsymbol{e}_n^{\mathcal{T}}, && \forall(m,n)\in\mathcal{E},\\ \boldsymbol{Y}^{\mathcal{T}}_{mn} &:= (\overline{y}_{nm}+y_{nm})\boldsymbol{e}_m\boldsymbol{e}_m^{\mathcal{T}}-y_{nm}\boldsymbol{e}_m\boldsymbol{e}_n^{\mathcal{T}}, && \forall(m,n)\in\mathcal{E}. \end{aligned}$$

For $|V_n|^2=V_n\overline{V}_n=\boldsymbol{v}^{\mathcal{H}}\boldsymbol{e}_n\boldsymbol{e}_n^{\mathcal{T}}\boldsymbol{v}$, it is clear that the corresponding $\boldsymbol{H}_n$ in (6) is

$$\boldsymbol{H}_{V,n}:=\boldsymbol{e}_n\boldsymbol{e}_n^{\mathcal{T}}\succeq\boldsymbol{0},\quad\forall n\in\mathcal{N} \tag{7}$$

which are rank-1. By taking separately the real and imaginary parts of (4) and (5), we obtain the $\{\boldsymbol{H}_\ell\}$ associated with the active and reactive power injections for all buses $n\in\mathcal{N}$

$$\boldsymbol{H}_{P,n}:=\frac{1}{2}\left(\boldsymbol{Y}_n+\boldsymbol{Y}_n^{\mathcal{H}}\right),\quad \boldsymbol{H}_{Q,n}:=\frac{j}{2}\left(\boldsymbol{Y}_n-\boldsymbol{Y}_n^{\mathcal{H}}\right) \tag{8}$$

and with sending-end and receiving-end active and reactive power flow at all lines $(m,n)\in\mathcal{E}$

$$\boldsymbol{H}^f_{P,mn}:=\frac{1}{2}\left(\boldsymbol{Y}^f_{mn}+\left(\boldsymbol{Y}^f_{mn}\right)^{\mathcal{H}}\right) \tag{9a}$$

$$\boldsymbol{H}^f_{Q,mn}:=\frac{j}{2}\left(\boldsymbol{Y}^f_{mn}-\left(\boldsymbol{Y}^f_{mn}\right)^{\mathcal{H}}\right) \tag{9b}$$

$$\boldsymbol{H}^t_{P,mn}:=\frac{1}{2}\left(\boldsymbol{Y}^{\mathcal{T}}_{mn}+\left(\boldsymbol{Y}^{\mathcal{T}}_{mn}\right)^{\mathcal{H}}\right) \tag{9c}$$

$$\boldsymbol{H}^t_{Q,mn}:=\frac{j}{2}\left(\boldsymbol{Y}^t_{mn}-\left(\boldsymbol{Y}^{\mathcal{T}}_{mn}\right)^{\mathcal{H}}\right). \tag{9d}$$

It is worth stressing that all $\{\boldsymbol{H}_\ell\}$ in (8) and (9) are sparse, low-rank, and Hermitian, but they are non-definite in general. The power flow and PSSE problems are formulated in order next.

A. Power flow analysis

Power flow analysis deals with specified power quantities, which are enforced for optimally operating an electric power grid. Specifically, given L perfectly known specifications $\{z_\ell\}_{\ell=1}^L$ and valid network parameters $\{\boldsymbol{H}_\ell\}_{\ell=1}^L$ as in (6), the goal of power flow analysis is to decide the state vector $\boldsymbol{v}\in\mathbb{C}^N$ that satisfies all specifications, namely,

$$\text{find}\quad \boldsymbol{v}\in\mathbb{C}^N \tag{10a}$$

$$\text{subject to}\quad \boldsymbol{v}^{\mathcal{H}}\boldsymbol{H}_\ell\boldsymbol{v}=z_\ell,\quad 1\le\ell\le L. \tag{10b}$$

Recall that each bus in a power system is classified as a PQ, PV, or slack (reference) bus based on the constraints imposed per bus. PQ buses, which often correspond to loads, specify and enforce only active and reactive power injection P_n and Q_n on bus n. On the other hand, the PV buses, which are typically associated with generators, enforce active power injection P_n and voltage magnitude $|V_n|$. For the slack bus, its voltage phase is fixed at $\theta_n=0$, by convention. With $\theta_n=0$, the power flow problem in (10) is equivalent to solving for $2N-1$ real-valued unknowns from L quadratic equations. The classical power flow problem considers the particular case where the $L=2N-1$ specifications are enforced only at the PV, PQ, and slack buses as opposed to a combination of buses and lines.

B. Power system state estimation

PSSE on the other hand deals with noisy observations acquired by the SCADA system adhering to

$$z_\ell=\boldsymbol{v}^{\mathcal{H}}\boldsymbol{H}_\ell\boldsymbol{v}+\eta_\ell,\quad 1\le\ell\le L \tag{11}$$

where η_ℓ accounts for the zero-mean distributed measurement error with known variance σ^2_ℓ, henceforth assumed independent across meters. The goal of PSSE is, given SCADA measurements $\{z_\ell\in\mathbb{R}\}_{\ell=1}^L$ and also parameters $\{\boldsymbol{H}_\ell\}_{\ell=1}^L$, estimate the state vector $\boldsymbol{v}\in\mathbb{C}^N$.

Adopting the WLS criterion, the SE task can be cast as that of solving the following nonlinear LS problem

$$\hat{\boldsymbol{v}}:=\arg\min_{\boldsymbol{v}\in\mathbb{C}^N}\sum_{\ell=1}^L w_\ell\big(z_\ell-\boldsymbol{v}^{\mathcal{H}}\boldsymbol{H}_\ell\boldsymbol{v}\big)^2 \tag{12}$$

where entries of the weight vector $\boldsymbol{w}:=[w_1\ \cdots\ w_L]^{\mathcal{T}}$ are often taken as $w_\ell:=1/\sigma^2_\ell$ for known σ^2_ℓ values. The WLS estimate $\hat{\boldsymbol{v}}$ coincides with the maximum likelihood one when the error vector $\boldsymbol{\eta}:=[\eta_1\ \cdots\ \eta_L]^{\mathcal{T}}$ obeys the multivariate Gaussian distribution $\mathcal{N}(\boldsymbol{0},\mathrm{diag}(\boldsymbol{\sigma}^2))$ with $\boldsymbol{\sigma}^2:=[\sigma^2_1\ \cdots\ \sigma^2_L]^{\mathcal{T}}$. Unfortunately, due to the quadratic terms $\{\boldsymbol{v}^{\mathcal{H}}\boldsymbol{H}_\ell\boldsymbol{v}\}$ inside the squares, the WLS SE problem is nonconvex. Minimizing nonconvex objectives, which typically exhibit many stationary points, is *NP-hard* in general [19]. Hence, solving the problem in (12) is indeed challenging.

PSSE approaches so far can be grouped as convex and nonconvex ones. The latter includes the "workhorse" Gauss-Newton method, which is also typically employed in practice: Upon linearizing the error function in the LS cost around a given estimate, the minimizer of the norm of the resulting

linearized approximation is used to initialize the next iteration [11, Sec. 1.5]. Minimizing nonconvex functions, Gauss-Newton iterations can be problematic due to: i) its sensitivity to the initial point; and, ii) lack of convergence guarantee to even a stationary point [11]. Convex approaches via SDR [7], [10] express all data $\{z_\ell\}$ as linear functions of the outer-product $\boldsymbol{V} := \boldsymbol{v}\boldsymbol{v}^{\mathcal{H}} \in \mathbb{C}^{N\times N}$. Problem (12) is then convexified by dropping the nonconvex constraint $\mathrm{rank}(\boldsymbol{V}) = 1$. SDR-based methods seldom yield solutions of rank-1 in the noisy case. Further eigen-decomposition or randomization procedures are required to recover the estimator $\hat{\boldsymbol{v}}$ from the SDR solution $\hat{\boldsymbol{V}}$. Performance of SDR solutions degrades when the data size is small, or when the set of measurements does not include the voltage magnitude at all buses, as will be demonstrated by our numerical results in Sec. V.

III. Feasible Point Pursuit based Solvers

In this section, the FPP-based power flow and PSSE solvers will be developed based on procedures distinct from existing iterative optimization and SDR-based SE approaches. To this end, some basics of FPP are first reviewed. For nonconvex QCQPs, FPP iteratively solves a series of convexified QCQPs obtained with successive convex inner-restrictions of the original nonconvex feasibility set, and with additive slacks to approximate the feasible solutions of the original nonconvex QCQP [21]. Specifically, starting with an initial guess, FPP first decomposes the quadratic terms in all nonconvex constraints into their convex and nonconvex parts by means of eigen-decomposition, which can be efficiently carried out offline; then it linearizes the nonconvex parts around the current iterate to obtain a restricted convex QCQP. Due to restriction of the feasibility set, the convexified QCQP may be infeasible. To sustain feasibility, a slack variable is introduced for each relaxed constraint, with a convex penalty on the slack variables added to the cost function, which can enforce sparing use of slacks to produce solutions of minimal constraint violation. The minimizer of the regularized convex QCQP subproblem is taken as the next iterate, which will be used as the linearization point of the nonconvex components at the next iteration. This successive convex approximation and feasibility-restoring procedure is repeated until a certain stopping criterion is met. Further details of FPP can be found in [21], [26].

Note that the power flow problem (10) consists of quadratic equality constraints, which are not in the standard QCQP form. To apply FPP, equalities are relaxed to inequalities, while penalizing the slack variables $\boldsymbol{s} := \{s_\ell \geq 0\}_{\ell=1}^L$, yielding

$$\underset{\boldsymbol{v}\in\mathbb{C}^N,\,\{s_\ell\}_{l=1}^L}{\text{minimize}} \quad f(\boldsymbol{s}) = \sum_{l=1}^{L} s_\ell^2 \tag{13a}$$

$$\text{subject to} \quad \left|z_\ell - \boldsymbol{v}^{\mathcal{H}}\boldsymbol{H}_\ell\boldsymbol{v}\right| \leq s_\ell, \quad 1\leq \ell\leq L \tag{13b}$$

where other choices of the convex penalty function $f(\cdot)$ include the (weighted) ℓ_1 or ℓ_∞ norm. Problem (13) is equivalent to the original power flow formulation (10) when the latter is feasible. To see this, assume that the set of power flow equations in (10b) admits (possibly more than one) feasible solutions. Clearly at the optimum of (13), the objective reduces to zero, the slack variables $\{s_\ell\}_{\ell=1}^L$ take zero values, and all equalities in (13b) are achieved, thus yielding a feasible solution to the set of power flow equations in (10).

Similarly, our PSSE formulation in (12) minimizes a quartic polynomial of $\boldsymbol{v}$. To use FPP, problem (12) is reformulated as

$$\underset{\boldsymbol{v}\in\mathbb{C}^N,\,\{s_\ell\}_{l=1}^L}{\text{minimize}} \quad f(\boldsymbol{s}) = \sum_{l=1}^{L} w_\ell s_\ell^2 \tag{14a}$$

$$\text{subject to} \quad \left|z_\ell - \boldsymbol{v}^{\mathcal{H}}\boldsymbol{H}_\ell\boldsymbol{v}\right| \leq s_\ell, \quad 1\leq \ell\leq L \tag{14b}$$

where the slack variables $\boldsymbol{s} := \{s_\ell \geq 0\}_{\ell=1}^L$ in this case relate to the deviations between noisy measurements $\{z_\ell\}_{\ell=1}^L$ and the actual quantities $\{\boldsymbol{v}^{\mathcal{H}}\boldsymbol{H}_\ell\boldsymbol{v}\}_{\ell=1}^L$. Problem (14) can be similarly shown equivalent to (12). Other convex penalty functions $f(\cdot)$ in (14a) can also be selected. In particular, if the error vector follows the multivariate Laplace distribution, i.e., $\boldsymbol{\eta} \sim \mathrm{Laplace}(\boldsymbol{0}, \boldsymbol{b})$ with $\boldsymbol{b} := [b_1 \ \cdots \ b_L]^{\mathcal{T}}$ collecting all scaling parameters, minimizing the ℓ_1-based function $f(\boldsymbol{s}) = \sum_{\ell=1}^L w_\ell s_\ell$ with $w_\ell = 1/b_\ell$ in (14) produces the maximum likelihood estimate [3], [7].

Apparently, the reformulated power flow and PSSE problems are of the same form [cf. (13) and (14)], except for a minor difference in the cost functions. Setting unit weights $w_\ell = 1$ in (14) reduces problem (14) to (13). Without loss of generality, we will hereafter focus on the PSSE formulation (14), and develop the novel FPP solver. The power flow problem can be readily handled with all weights being $w_\ell = 1$.

In this direction, let us first convert problem (14) into a standard QCQP. Note that constraints (14b) can be replaced by two sets of inequalities to arrive at

$$\underset{\boldsymbol{v}\in\mathbb{C}^N,\,\boldsymbol{s}\in\mathbb{R}^L}{\text{minimize}} \quad \sum_{l=1}^{L} w_\ell s_\ell^2 \tag{15a}$$

$$\text{subject to} \quad \boldsymbol{v}^{\mathcal{H}}\boldsymbol{H}_\ell\boldsymbol{v} \leq z_\ell + s_\ell, \qquad 1\leq \ell\leq L \tag{15b}$$

$$\boldsymbol{v}^{\mathcal{H}}\left(-\boldsymbol{H}_\ell\right)\boldsymbol{v} \leq -z_\ell + s_\ell, \quad 1\leq \ell\leq L. \tag{15c}$$

It is shown in Proposition 3 of the Appendix that any KKT point of (15) is also a stationary point of (12). Evidently, problem (15) is nonconvex even for (semi)definite coefficient matrices $\{\boldsymbol{H}_\ell\}_{\ell=1}^L$. Next we demonstrate how to take advantage of FFP to solve the problem at hand in detail.

As discussed in Sec. II, there are two types of $\{\boldsymbol{H}_\ell\}$ matrices, one corresponding to the squared voltage magnitude, and the other to power quantities. Type-I $\{\boldsymbol{H}_\ell\}$ are positive semidefinite [cf. (7)], while Type-II are non-definite [cf. (8) and (9)]. For ease of exposition, let us introduce the FPP constraint convexification procedure using one nonconvex quadratic constraint in (15). Along the lines of FPP, consider the term $\boldsymbol{v}^{\mathcal{H}}\boldsymbol{H}_\ell\boldsymbol{v}$ in (15b) for some $\boldsymbol{H}_\ell$ in (8), which can be decomposed into its convex and nonconvex components as

$$\boldsymbol{v}^{\mathcal{H}}\boldsymbol{H}_\ell^{(+)}\boldsymbol{v} + \boldsymbol{v}^{\mathcal{H}}\boldsymbol{H}_\ell^{(-)}\boldsymbol{v} \leq z_\ell + s_\ell \tag{16}$$

where $\boldsymbol{H}_\ell^{(+)}$ and $\boldsymbol{H}_\ell^{(-)}$ represent the positive semidefinite (convex) and negative semidefinite (nonconvex) parts of $\boldsymbol{H}_\ell$ in (16), respectively. For the nonconvex source $\boldsymbol{v}^{\mathcal{H}}\boldsymbol{H}_\ell^{(-)}\boldsymbol{v}$ in (16), an inner linear restriction will be derived next.

The following inequality holds for any $\boldsymbol{y} \in \mathbb{C}^N$ due to the negative semidefiniteness of $\boldsymbol{H}_\ell^{(-)}$

$$(\boldsymbol{v}-\boldsymbol{y})^{\mathcal{H}}\boldsymbol{H}_\ell^{(-)}(\boldsymbol{v}-\boldsymbol{y}) \leq 0. \tag{17}$$

Upon expanding the left-hand-side and rearranging terms, one arrives at

$$\boldsymbol{v}^{\mathcal{H}}\boldsymbol{H}_\ell^{(-)}\boldsymbol{v} \leq 2\Re\Big\{\boldsymbol{y}^{\mathcal{H}}\boldsymbol{H}_\ell^{(-)}\boldsymbol{v}\Big\} - \boldsymbol{y}^{\mathcal{H}}\boldsymbol{H}_\ell^{(-)}\boldsymbol{y}.$$

Key to the FPP algorithm is replacing the nonconvexity stemming from $\boldsymbol{H}_\ell^{(-)}$ in (16) or (15b) by its inner linear approximation at some given point $\boldsymbol{y}$ to yield

$$\boldsymbol{v}^{\mathcal{H}}\boldsymbol{H}_\ell^{(+)}\boldsymbol{v} + 2\Re\Big\{\boldsymbol{y}^{\mathcal{H}}\boldsymbol{H}_\ell^{(-)}\boldsymbol{v}\Big\} \leq z_\ell + \boldsymbol{y}^{\mathcal{H}}\boldsymbol{H}_\ell^{(-)}\boldsymbol{y} + s_\ell. \tag{18}$$

The strategy in selecting the linearization point $\boldsymbol{y}$ will be discussed shortly. In the same fashion, the nonconvex quadratic constraints in (15c) can be replaced by

$$\boldsymbol{v}^{\mathcal{H}}\big(-\boldsymbol{H}_\ell^{(-)}\big)\boldsymbol{v} - 2\Re\Big\{\boldsymbol{y}^{\mathcal{H}}\boldsymbol{H}_\ell^{(+)}\boldsymbol{v}\Big\} \leq -z_\ell - \boldsymbol{y}^{\mathcal{H}}\boldsymbol{H}_\ell^{(+)}\boldsymbol{y} + s_\ell. \tag{19}$$

Heed that the flexibility introduced by the slacks $\{s_\ell\}_{\ell=1}^L$ always restores the feasibility of the relaxed constraints, which contributes to improved performance of FPP over other convexification approaches [21]. In the presence of noise, the minimum values required for $\{s_\ell \geq 0\}_\ell^L$ to satisfy (18) and (19) depend on the measurement error contained in $\{z_\ell\}_{\ell=1}^L$.

The FPP method replaces all nonconvex constraints in (15b) by their convex restriction (18), and those in (15c) by (19) to derive a convexified QCQP regularized with slack variables to ensure feasibility. Minimizing some convex penalty function of the slacks $\{s_\ell\}_{\ell=1}^L$ not only minimizes the fitting error between $\{z_\ell\}$ and $\{\boldsymbol{v}^{\mathcal{H}}\boldsymbol{H}_\ell\boldsymbol{v}\}$, but also enforces sparing use of slacks and promotes solutions of minimal constraint violation.

In a nutshell, the developed FPP-based PSSE solver can be understood as follows. Starting with an initial point $\boldsymbol{v}_0$ (typically the flat voltage profile point, i.e., all-ones vector), our FPP-based solver successively tackles a sequence of convexified QCQPs with the linearization point being the current iterate $\boldsymbol{v}_k$, which is the $\boldsymbol{v}$-minimizer obtained by solving a convexified QCQP at the previous iteration. Hence, assuming available the $\boldsymbol{v}$-minimizer $\boldsymbol{v}_k$ at the $(k+1)$-st iteration, our FPP-based solver boils down to solving the following convexified QCQP subproblem

$$\{\boldsymbol{v}_{k+1}, \boldsymbol{s}_{k+1}\} := \arg\min_{\boldsymbol{v},\,\boldsymbol{s}} \sum_{l=1}^{L} w_\ell s_\ell^2 \tag{20a}$$

subject to

$$\boldsymbol{v}^{\mathcal{H}}\boldsymbol{H}_\ell^{(+)}\boldsymbol{v} + 2\Re\Big\{\boldsymbol{y}^{\mathcal{H}}\boldsymbol{H}_\ell^{(-)}\boldsymbol{v}\Big\} \leq z_\ell + \boldsymbol{y}^{\mathcal{H}}\boldsymbol{H}_\ell^{(-)}\boldsymbol{y} + s_\ell \tag{20b}$$

$$\boldsymbol{v}^{\mathcal{H}}\boldsymbol{H}_\ell^{(-)}\boldsymbol{v} + 2\Re\Big\{\boldsymbol{y}^{\mathcal{H}}\boldsymbol{H}_\ell^{(+)}\boldsymbol{v}\Big\} \geq z_\ell + \boldsymbol{y}^{\mathcal{H}}\boldsymbol{H}_\ell^{(+)}\boldsymbol{y} - s_\ell \tag{20c}$$

$$\forall \ell = 1, 2, \ldots, L$$

where $\boldsymbol{y} := \boldsymbol{v}_k$ is the $\boldsymbol{v}$-minimizer of (20) at the k-th iteration. The QCQP in (20) is convex, which can be solved in polynomial time using off-the-shelf solvers [28]

The FPP-based PSSE solver is summarized in Algorithm 1. The following three properties of our FPP-based solver are worth highlighting.

Algorithm 1: FPP-based power flow and PSSE Solvers

Input: Data $\{(z_\ell, \boldsymbol{H}_\ell)\}$; weights $\{w_\ell = 1\}$ for power flow, and $\{w_\ell = 1/\sigma_\ell^2\}$ for PSSE; solution accuracy $\epsilon > 0$.
Initialization: set $k = 0$ and $\boldsymbol{y} = [1 \ \cdots \ 1]^{\mathcal{T}}$.
repeat
  $\{\boldsymbol{v}_k, \boldsymbol{s}_k\} \leftarrow$ minimizer of problem (20)
  $\boldsymbol{y} \leftarrow \boldsymbol{v}_k$
  $k \leftarrow k + 1$
until $\|\boldsymbol{v}_k - \boldsymbol{v}_{k-1}\|_2 \leq \epsilon$.
Output: $\hat{\boldsymbol{v}} \leftarrow \boldsymbol{v}_k$.

Remark 1 (*Power flow analysis).* Cast as a special instance of PSSE, the power flow problem in (10) can be solved by our developed FPP-based PSSE solver with unit weights $w_\ell = 1$.

Remark 2 (*Bad data removal).* Besides the ℓ_2-norm in (20a), other convex penalty functions can be used to fit different (noisy) data models. In particular, adopting the weighted ℓ_1-norm (i.e., replacing s_ℓ^2 with $|s_\ell|$) yields the weighted least-absolute-value estimator known for bad data cleansing [29].

Remark 3 (*Synchrophasors).* Synchrophasors, if available, can be easily incorporated into the developed PSSE formulation (20). To see this, letting $\boldsymbol{\zeta}_n = \boldsymbol{\Phi}_n\boldsymbol{v} + \boldsymbol{\epsilon}_n$ collect the noisy PMU data at bus n, hybrid estimation exploiting both nonlinear SCADA measurements and linear PMU ones can be achieved [30] with an additional data-fitting term for the PMU data in (20a), namely, $\sum_{n\in\mathcal{P}} \|\boldsymbol{\zeta}_n - \boldsymbol{\Phi}_n\boldsymbol{v}\|_2^2$, where $\mathcal{P}$ denotes the subset of the PMU-instrumented buses.

In terms of computational complexity, Algorithm 1 involves solving a convex QCQP of the form (20) per iteration, which can be easily formulated as a second-order cone program. The worst-case complexity is $\mathcal{O}((2N+3L)^{3.5})$ [31], that is clearly lower than $\mathcal{O}((2N+2L)^{6.5})$ incurred by SDR [21]. Algorithm 1 usually takes a few iterations to converge. On the theoretical side, the next result establishes convergence of our developed FPP-based solvers to a stationary point of the WLS formulation.

Proposition 1 (Global convergence of FPP-based solvers)**.** *Let $\{\boldsymbol{v}_k\}_{k=0}^\infty$ be any sequence generated by the FPP-based solver in Algorithm 1. Then, all limit points of $\{\boldsymbol{v}_k\}_{k=0}^\infty$ are stationary points of the WLS problem in* (12)*.*

Proof. As elaborated in Sec. III, solving problem (15) is equivalent to solving problem (12). The nonconvex QCQP of complex-valued vector $\boldsymbol{v} \in \mathbb{C}^N$ in (15) can be equivalently posed as a QCQP of the expanded real-valued vector $\boldsymbol{u} := [\Re(\boldsymbol{v})^{\mathcal{T}} \ \Im(\boldsymbol{v})^{\mathcal{T}}]^{\mathcal{T}} \in \mathbb{R}^{2N}$, where the associated quadratic matrices $\{\overline{\boldsymbol{H}}_\ell\}$ are given as

$$\overline{\boldsymbol{H}}_\ell := \begin{bmatrix} \Re(\boldsymbol{H}_\ell) & -\Im(\boldsymbol{H}_\ell) \\ \Im(\boldsymbol{H}_\ell) & \Re(\boldsymbol{H}_\ell) \end{bmatrix} \in \mathbb{R}^{2N\times 2N}, \quad 1 \leq \ell \leq L.$$

Accordingly, each constraint in (15) can be re-expressed as the difference between two convex functions. To see this, consider e.g. constraint (15b), which can be rewritten as

$$\big(\boldsymbol{u}^{\mathcal{T}}\overline{\boldsymbol{H}}_\ell^{(+)}\boldsymbol{u} - s_\ell\big) - \big(\boldsymbol{u}^{\mathcal{T}}(-\overline{\boldsymbol{H}}_\ell^{(-)})\boldsymbol{u}\big) \leq z_\ell \tag{21}$$

where $\overline{\boldsymbol{H}}_\ell^{(+)}$ and $\overline{\boldsymbol{H}}_\ell^{(-)}$ are the positive and negative semidefinite parts of $\overline{\boldsymbol{H}}_\ell$, hence rendering terms $\boldsymbol{u}^\mathcal{T}\overline{\boldsymbol{H}}_\ell^{(+)}\boldsymbol{u} - s_\ell$ and $\boldsymbol{u}^\mathcal{T}(-\overline{\boldsymbol{H}}_\ell^{(-)})\boldsymbol{u}$ both convex. Algorithm 1 is tantamount to an application of the convex-concave procedure [23], [24] to the reformulated QCQP in the real domain. Hence, the sequence generated by Algorithm 1 converges to a stationary point of (12) by invoking the results in [25, Thm. 10]. □

IV. Cramér-Rao Bound for PSSE

According to standard results from estimation theory [32], the variance of any unbiased estimator is lower bounded by the Cramér-Rao lower bound (CRLB). Appreciating its key role as a performance benchmark across different estimators, this section establishes the CRLB for the fundamental PSSE problem. The CRLB analysis of PSSE however, entails finding derivatives (gradient and Hessian) of a real-valued function with respect to multiple complex-valued variables. To address this challenge, we call for advanced complex analysis tools based on the so-termed Wirtinger derivative and Wirtinger's calculus, which are detailed in the Appendix. The following result provides a closed-form CRLB for any unbiased PSSE solver under the AWGN model in (11), which can be directly used to assess the performance of other PSSE solvers.

Proposition 2. *Consider estimating the unknown state vector $\boldsymbol{v} \in \mathbb{C}^N$ from noisy data $\{z_\ell\}_{\ell=1}^L$ obeying the model in* (11)*, where the noise η_ℓ is assumed Gaussian distributed with mean zero and variance σ_ℓ^2, and is also independent across meters. Then the covariance matrix of any unbiased estimator $\hat{\boldsymbol{v}}$ obeys*

$$\mathrm{Cov}(\hat{\boldsymbol{v}}) \succeq [\boldsymbol{F}^\dagger(\boldsymbol{v}, \overline{\boldsymbol{v}})]_{1:N,1:N} \tag{22}$$

where the Fisher information matrix is given by

$$\boldsymbol{F}(\boldsymbol{v}, \overline{\boldsymbol{v}}) = \begin{bmatrix} \sum_{\ell=1}^L \frac{1}{\sigma_\ell^2}(\boldsymbol{H}_\ell \boldsymbol{v})(\boldsymbol{H}_\ell \boldsymbol{v})^\mathcal{H} & \sum_{\ell=1}^L \frac{1}{\sigma_\ell^2}(\boldsymbol{H}_\ell \boldsymbol{v})(\overline{\boldsymbol{H}}_\ell \overline{\boldsymbol{v}})^\mathcal{H} \\ \sum_{\ell=1}^L \frac{1}{\sigma_\ell^2}(\overline{\boldsymbol{H}}_\ell \overline{\boldsymbol{v}})(\boldsymbol{H}_\ell \boldsymbol{v})^\mathcal{H} & \sum_{\ell=1}^L \frac{1}{\sigma_\ell^2}(\overline{\boldsymbol{H}}_\ell \overline{\boldsymbol{v}})(\overline{\boldsymbol{H}}_\ell \overline{\boldsymbol{v}})^\mathcal{H} \end{bmatrix} \tag{23}$$

Furthermore, $\boldsymbol{F}$ has at least rank-1 deficiency even when all possible SCADA measurements are available.

The proof of Proposition 2 is deferred to the Appendix. Even though the Fisher information matrix (FIM) in (23) is rank deficient, the pseudo-inverse of FIM qualifies itself as a valid yet generally looser lower bound on the mean-square error (MSE) of any unbiased estimator [33]. This lower bound is often attainable in practice, and is predictive of optimal estimator performance [33], as will be demonstrated by our numerical tests in Sec. V. The derived CRLB in (22) will be employed to benchmark and compare performance of different PSSE solvers next.

V. Simulated Tests

In the section, we compare the proposed FPP-based solvers in Algorithm 1 with existing alternatives including the WLS via Gauss-Newton iterations (GN-based), and the SDR-based solver (SDR-based) [10], [14] for both power flow and PSSE tasks on several IEEE benchmark systems [34]. Throughout, all reported numerical results were obtained by averaging over 100 independent Monte Carlo realizations. The three

TABLE I
Empirical success rate on IEEE test systems with $\theta = 0.1\pi$.

Test case	5-bus	9-bus	14-bus	24-bus	30-bus	39-bus
FPP-based	100%	100%	100%	100%	100%	100%
SDR-based	0	29%	40%	2%	94%	97%
GN-based	100%	100%	100%	100%	87%	64%

TABLE II
Empirical success rate on IEEE test systems with $\theta = 0.3\pi$.

Test case	5-bus	9-bus	14-bus	24-bus	30-bus	39-bus
FPP-based	100%	100%	100%	100%	100%	100%
SDR-based	4%	33%	15%	10%	0	0
GN-based	100%	55%	33%	21%	0	5%

PSSE solvers from noisy measurements are compared in terms of the mean-square error $\sum_{i=1}^{100} \|\hat{\boldsymbol{v}}_i - \boldsymbol{v}\|_2^2/100$, where $\hat{\boldsymbol{v}}_i$ is the returned estimate at the i-th realization, and $\boldsymbol{v}$ the actual voltage profile. In the absence of noise, performance of the power flow solvers is assessed through the empirical success rate over 100 trials. A success is declared for a trial if the returned power flow solution $\hat{\boldsymbol{v}}$ incurs a relative violation on the given set of L power flow equations, given by $\sum_{\ell=1}^L (z_\ell - \hat{\boldsymbol{v}}^\mathcal{H}\boldsymbol{H}_\ell\hat{\boldsymbol{v}})^2/\sum_{\ell=1}^L z_\ell^2$ less than 10^{-3}. (The reason why $\|\boldsymbol{v} - \hat{\boldsymbol{v}}\|_2^2$ is not used is due to existence of possibly multiple solutions $\boldsymbol{v}$ satisfying the set of power flow equations.)

Different system quantities and voltage profiles were generated via the MATLAB-based toolbox MATPOWER [35]. The Gauss-Newton method was implemented using the SE function 'doSE.m' in MATPOWER, which was modified to terminate either upon convergence, or, when the condition number of the approximate linearization exceeds 10^5 flagging explosion of the iterates [10]. The SDR- and FPP-based solvers were realized via the optimization modeling package YALMIP [36], as well as the interior-point solver SeDuMi [28]. Furthermore, the flat-voltage profile point was used as the initial guess for the Gauss-Newton and FPP approaches. In order to fix the phase ambiguity, the phase generated at the reference bus is set to 0 in all tests. The FPP solver stops either when a maximum number 100 of iterations are reached, or when the objective value improvement between two consecutive iterations becomes smaller 10^{-5}. All experiments were conducted on an Intel CPU @ 3.4 GHz (32 GB RAM) computer.

To evaluate the performance of the FPP-based solver for power flow analysis, the first experiment simulates noiseless data corresponding to the classical power flow problem. That is, a total of $L = 2N - 1$ system variables were specified at the PV, PQ, and slack buses to solve for $2N - 1$ real-valued unknowns in $\boldsymbol{v} \in \mathbb{C}^N$ with the reference bus's phase fixed at 0. The actual voltage magnitude of each bus was uniformly distributed over $[0.9, 1.1]$, and its angle over $[-\theta, \theta]$ with $\theta = 0.1\pi$ and 0.3π. Empirical success rate results on several IEEE benchmark systems were reported in Tables I and II for $\theta = 0.1\pi$ and 0.3π, respectively. Apparently, our developed FPP-based power flow solver solves exactly the classical power flow problem in all simulated tests, while the SDR-based one fails with high probability. The Gauss-Newton method performs well when the initial point lies close to the actual solution due

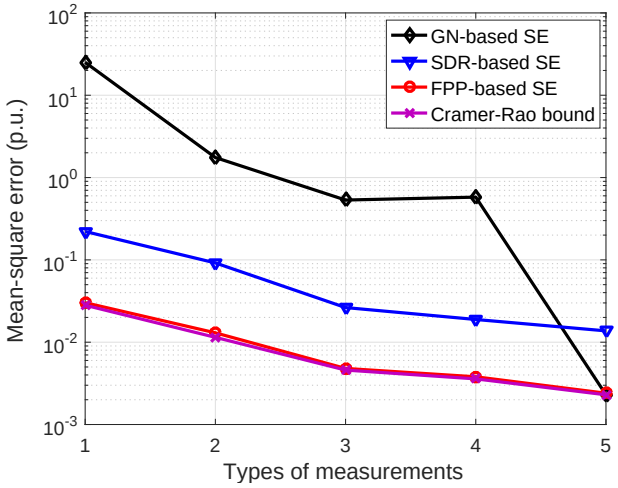


Fig. 1. MSEs as well as CRLB versus types of measurements used on the IEEE 14-bus test system using: i) Gauss-Newton based SE; ii) SDR-based SE; and iii) FPP-based SE.

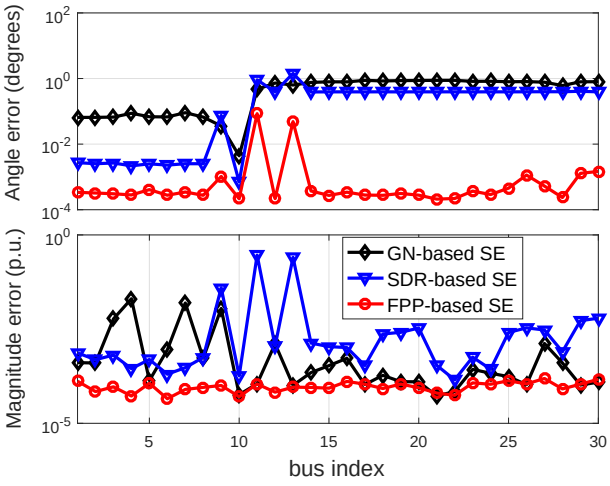


Fig. 2. Magnitude and angle estimation errors at each bus on the IEEE 30-bus benchmark system using: i) Gauss-Newton based SE; ii) SDR-based SE; and iii) FPP-based SE.

to small θ in Table I, while it diverges frequently for large θ values in Table II.

The second experiment compares the MSE performance of various approaches relative to the analytical Cramér-Rao bound in (22) on the IEEE 14-bus test system [34]. The actual voltage magnitude and angle of each bus were generated uniformly over $[0.9,\ 1.1]$, and $[-0.4\pi,\ 0.4\pi]$, respectively. To demonstrate the SE performance evolution of various approaches with respect to the increasing number of measurements, we performed 5 tests denoted by the x-axis values $\{1,\ 2,\ 3,\ 4,\ 5\}$. Recall that the SCADA system can measure seven types of power quantities, namely, $\{|V_k|^2,\ P^f_{mn},\ P^t_{mn},\ Q^f_{mn},\ Q^t_{mn},\ P_n,\ Q_n\}$ in (7)-(9). The first test simulated the three algorithms using all measurements of the first three types $\{|V_k|^2,\ P^f_{mn},\ P^t_{mn}\}$, whose mean-square error performance averaged over 100 independent realizations were given by the y-values at the x-axis value of 1 in Fig. 1. The second to the fifth tests were implemented by including in order an additional type of measurements from $\{Q^f_{mn},\ Q^t_{mn},\ P_n,\ Q_n\}$, which correspond to the x-values of 2 to 5 in Fig. 1.

Measurement noise was randomly and independently generated from Gaussian distribution having zero-mean and standard deviation 0.1. The SDR estimator was recovered from the SDR solution by picking the minimum-cost vector over the eigenvector and $5,000$ zero-mean Gaussian randomizations with covariance matrix being the SDR solution. The MSE as well as the CRLB versus the types of measurements available are shown in Fig. 1, corroborating the near-optimal performance relative to the CRLB and robustness of our developed FPP-based PSSE solver.

The last experiment on the IEEE 30-bus benchmark system simulates a high signal-to-noise ratio and complete-data scenario, where all voltage magnitude as well as all active power flow at both sending- and receiving-ends were measured to be advantageous to the SDR-based method [7]. Independent zero-mean Gaussian noise was assumed to have standard deviations 0.05 for power measurements and 0.02 for voltage measurements. The actual voltage magnitude and angle of each bus were generated uniformly at random over $[0.9,\ 1.1]$, and $[-0.4\pi,\ 0.4\pi]$, respectively. Figure 2 depicts the average magnitude and angle estimation errors of three PSSE schemes across buses. The curves in Fig. 2 demonstrate the merits of the FPP-based PSSE solver in this scenario. Regarding running time, the Gauss-Newton scheme converges in 0.2 seconds typically, while both the SDR-based and the FPP-based methods take about 10 seconds on average.

VI. Conclusions

Motivated by the inherent nonconvexity of the power flow and PSSE tasks and leveraging recent advances in handling nonconvex QCQPs, this work first reformulated power flow and PSSE as a nonconvex QCQP. The resulting nonconvex QCQP was subsequently solved by the FPP algorithm. The novel FPP-based solvers were shown to converge to a stationary point of the WLS formulation. To fairly compare different PSSE solvers from noisy data, the CRLB for PSSE assuming an AWGN model was derived based on Wirtinger's calculus for functions over complex domains. Extensive numerical tests showed markedly improved performance of our FPP-based solver for both power flow and PSSE tasks at the price of increased runtime over competing Gauss-Newton- and SDR-based alternatives on a variety of IEEE test systems.

Pertinent future research directions include developing distributed implementations for large-scale power networks by exploiting the natural low-rank and sparsity structure present in the coefficient matrices $\{\boldsymbol{H}_\ell\}$. Another possibility consists of leveraging state-of-the-art approaches for tackling random quadratic systems of equations to solve the power flow and PSSE problems [37]. Generalizing feasible point pursuit algorithms to other nonconvex power grid control tasks such as stochastic energy management, and distribution system-level power flow and PSSE constitute meaningful directions for future research as well.

VII. Appendix

Proof of Proposition 2: For the AWGN model in (11) with $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \mathrm{diag}(\boldsymbol{\sigma^2}))$, the data likelihood can be written as

$$p(\boldsymbol{z};\boldsymbol{v}) = \prod_{\ell=1}^{L} \frac{1}{\sqrt{2\pi\sigma_\ell^2}} \exp\left[-\frac{\left(z_\ell - \boldsymbol{v}^{\mathcal{H}}\boldsymbol{H}_\ell\boldsymbol{v}\right)^2}{2\sigma_\ell^2}\right]$$

and the negative log-likelihood $f(\boldsymbol{v}) = -\ln p(\boldsymbol{z};\boldsymbol{v})$ is

$$f(\boldsymbol{v}) = \sum_{\ell=1}^{L}\left[\frac{1}{2\sigma_\ell^2}\left(z_\ell - \boldsymbol{v}^{\mathcal{H}}\boldsymbol{H}_\ell\boldsymbol{v}\right)^2 + \frac{1}{2}\ln\left(2\pi\sigma_\ell^2\right)\right]. \quad (24)$$

The Fisher information matrix is defined as the Hessian of the objective function $f(\boldsymbol{v}) \in \mathbb{R}$ with respect to the variable vector $\boldsymbol{v} \in \mathbb{C}^N$. So the task of deriving the Cramér-Rao bound amounts to finding the Hessian of a real-valued function with respect to a complex-valued vector. Recall from *Wirtinger's calculus* that $f(\boldsymbol{v})$ can be equivalently rewritten as $f(\boldsymbol{v},\overline{\boldsymbol{v}})$ [38]. Upon introducing the conjugate coordinates $[\boldsymbol{v}^{\mathcal{T}}\ \overline{\boldsymbol{v}}^{\mathcal{T}}]^{\mathcal{T}} \in \mathbb{C}^{2N}$, the so-called *Wirtinger derivative* is [38]

$$\frac{\partial f}{\partial \boldsymbol{v}} := \left.\frac{\partial f(\boldsymbol{v},\overline{\boldsymbol{v}})}{\partial \boldsymbol{v}^{\mathcal{T}}}\right|_{\overline{\boldsymbol{v}}=\text{constant}} = \left[\frac{\partial f}{\partial v_1}\ \cdots\ \frac{\partial f}{\partial v_N}\right]\Bigg|_{\overline{\boldsymbol{v}}=\text{constant}}$$
$$\frac{\partial f}{\partial \overline{\boldsymbol{v}}} := \left.\frac{\partial f(\boldsymbol{v},\overline{\boldsymbol{v}})}{\partial \overline{\boldsymbol{v}}^{\mathcal{T}}}\right|_{\boldsymbol{v}=\text{constant}} = \left[\frac{\partial f}{\partial \overline{v}_1}\ \cdots\ \frac{\partial f}{\partial \overline{v}_N}\right]\Bigg|_{\boldsymbol{v}=\text{constant}}.$$

Our definitions here follow the convention in multivariate calculus that derivatives are denoted by row vectors, and gradients by column vectors. For brevity, let $\phi_\ell(\boldsymbol{v},\overline{\boldsymbol{v}}) := z_\ell - \overline{\boldsymbol{v}}^{\mathcal{T}}\boldsymbol{H}_\ell\boldsymbol{v}$. Accordingly, the derivatives of f in (24) can be obtained as

$$\frac{\partial f}{\partial \boldsymbol{v}} = \sum_{\ell=1}^{L}\frac{1}{\sigma_\ell^2}\phi_\ell(\boldsymbol{v},\overline{\boldsymbol{v}})\frac{\partial \phi_\ell(\boldsymbol{v},\overline{\boldsymbol{v}})}{\partial \boldsymbol{v}^{\mathcal{T}}} \quad (26a)$$

$$\frac{\partial f}{\partial \overline{\boldsymbol{v}}} = \sum_{\ell=1}^{L}\frac{1}{\sigma_\ell^2}\phi_\ell(\boldsymbol{v},\overline{\boldsymbol{v}})\frac{\partial \phi_\ell(\boldsymbol{v},\overline{\boldsymbol{v}})}{\partial \overline{\boldsymbol{v}}^{\mathcal{T}}} \quad (26b)$$

where the partial derivatives of ϕ_ℓ can be found as

$$\frac{\partial \phi_\ell(\boldsymbol{v},\overline{\boldsymbol{v}})}{\partial \boldsymbol{v}^{\mathcal{T}}} = -\overline{\boldsymbol{v}}^{\mathcal{T}}\boldsymbol{H}_\ell = -(\boldsymbol{H}_\ell\boldsymbol{v})^{\mathcal{H}} \quad (27a)$$

$$\frac{\partial \phi_\ell(\boldsymbol{v},\overline{\boldsymbol{v}})}{\partial \overline{\boldsymbol{v}}^{\mathcal{T}}} = -\boldsymbol{v}^{\mathcal{T}}\boldsymbol{H}_\ell^{\mathcal{T}} = -(\overline{\boldsymbol{H}}_\ell\overline{\boldsymbol{v}})^{\mathcal{H}}. \quad (27b)$$

In the conjugate coordinate system, the complex Hessian is defined as

$$\boldsymbol{\mathcal{H}} := \nabla^2 f = \begin{bmatrix} \boldsymbol{\mathcal{H}}_{\boldsymbol{vv}} & \boldsymbol{\mathcal{H}}_{\boldsymbol{\overline{v}v}} \\ \boldsymbol{\mathcal{H}}_{\boldsymbol{v\overline{v}}} & \boldsymbol{\mathcal{H}}_{\boldsymbol{\overline{vv}}} \end{bmatrix} \quad (28)$$

whose blocks are given by

$$\boldsymbol{\mathcal{H}}_{\boldsymbol{vv}} := \frac{\partial}{\partial \boldsymbol{v}^{\mathcal{T}}}\left(\frac{\partial f}{\partial \boldsymbol{v}}\right)^{\mathcal{H}}, \quad \boldsymbol{\mathcal{H}}_{\boldsymbol{\overline{v}v}} := \frac{\partial}{\partial \overline{\boldsymbol{v}}^{\mathcal{T}}}\left(\frac{\partial f}{\partial \boldsymbol{v}}\right)^{\mathcal{H}}$$
$$\boldsymbol{\mathcal{H}}_{\boldsymbol{v\overline{v}}} := \frac{\partial}{\partial \boldsymbol{v}^{\mathcal{T}}}\left(\frac{\partial f}{\partial \overline{\boldsymbol{v}}}\right)^{\mathcal{H}}, \quad \boldsymbol{\mathcal{H}}_{\boldsymbol{\overline{vv}}} := \frac{\partial}{\partial \overline{\boldsymbol{v}}^{\mathcal{T}}}\left(\frac{\partial f}{\partial \overline{\boldsymbol{v}}}\right)^{\mathcal{H}}.$$

After substituting (26) and (27) into the last equations, and with some tedious algebraic manipulations, the first block of $\boldsymbol{\mathcal{H}}$ can be obtained as

$$\begin{aligned}\boldsymbol{\mathcal{H}}_{\boldsymbol{vv}} =& \frac{\partial}{\partial \boldsymbol{v}^{\mathcal{T}}}\Big(\sum_{\ell=1}^{L}\frac{-1}{\sigma_\ell^2}\phi_\ell(\boldsymbol{v},\overline{\boldsymbol{v}})\boldsymbol{H}_\ell\boldsymbol{v}\Big)\\ =& \sum_{\ell=1}^{L}\frac{1}{\sigma_\ell^2}\Big(\boldsymbol{H}_\ell\boldsymbol{v}(\boldsymbol{H}_\ell\boldsymbol{v})^{\mathcal{H}} - \phi_\ell(\boldsymbol{v},\overline{\boldsymbol{v}})\boldsymbol{H}_\ell\Big). \end{aligned} \quad (29)$$

The other blocks can be derived in a similar fashion. Upon omitting algebraic details, the remaining three blocks can be obtained as follows

$$\boldsymbol{\mathcal{H}}_{\boldsymbol{\overline{v}v}} = \sum_{\ell=1}^{L}\frac{1}{\sigma_\ell^2}\boldsymbol{H}_\ell\boldsymbol{v}(\overline{\boldsymbol{H}}_\ell\overline{\boldsymbol{v}})^{\mathcal{H}} \quad (30)$$

$$\boldsymbol{\mathcal{H}}_{\boldsymbol{v\overline{v}}} = \sum_{\ell=1}^{L}\frac{1}{\sigma_\ell^2}\overline{\boldsymbol{H}}_\ell\overline{\boldsymbol{v}}(\boldsymbol{H}_\ell\boldsymbol{v})^{\mathcal{H}} \quad (31)$$

$$\boldsymbol{\mathcal{H}}_{\boldsymbol{\overline{vv}}} = \sum_{\ell=1}^{L}\frac{1}{\sigma_\ell^2}\Big(\overline{\boldsymbol{H}}_\ell\overline{\boldsymbol{v}}(\overline{\boldsymbol{H}}_\ell\overline{\boldsymbol{v}})^{\mathcal{H}} - \phi_\ell(\boldsymbol{v},\overline{\boldsymbol{v}})\overline{\boldsymbol{H}}_\ell\Big). \quad (32)$$

Evaluating the Hessian $\boldsymbol{\mathcal{H}}$ in (28) [and its blocks in (29)-(32)] at the true value of $\boldsymbol{v}$, and taking the expectation with respect to the noise vector $\boldsymbol{\eta}$, it is easy to verify that $\mathbb{E}\left[\phi_\ell(\boldsymbol{v},\overline{\boldsymbol{v}})\right] = 0$. Hence, the ϕ_ℓ-related terms disappear, so the FIM $\boldsymbol{F} := \mathbb{E}[\boldsymbol{\mathcal{H}}] \in \mathbb{C}^{2N\times 2N}$ can be expressed as [39]

$$\begin{aligned}\boldsymbol{F} =& \begin{bmatrix} \sum_{\ell=1}^{L}\boldsymbol{H}_\ell\boldsymbol{v}(\boldsymbol{H}_\ell\boldsymbol{v})^{\mathcal{H}}/\sigma_\ell^2 & \sum_{\ell=1}^{L}\boldsymbol{H}_\ell\boldsymbol{v}(\overline{\boldsymbol{H}}_\ell\overline{\boldsymbol{v}})^{\mathcal{H}}/\sigma_\ell^2 \\ \sum_{\ell=1}^{L}\overline{\boldsymbol{H}}_\ell\overline{\boldsymbol{v}}(\boldsymbol{H}_\ell\boldsymbol{v})^{\mathcal{H}}/\sigma_\ell^2 & \sum_{\ell=1}^{L}\overline{\boldsymbol{H}}_\ell\overline{\boldsymbol{v}}(\overline{\boldsymbol{H}}_\ell\overline{\boldsymbol{v}})^{\mathcal{H}}/\sigma_\ell^2 \end{bmatrix}\\ =& \sum_{\ell=1}^{L}\boldsymbol{g}_\ell\boldsymbol{g}_\ell^{\mathcal{H}} \triangleq \boldsymbol{G}\boldsymbol{G}^{\mathcal{H}} \end{aligned} \quad (33)$$

where $\boldsymbol{G} := [\boldsymbol{g}_1\ \cdots\ \boldsymbol{g}_L] \in \mathbb{C}^{2N\times L}$ is introduced to show the rank-deficiency of $\boldsymbol{F}$, whose ℓ-th column is given as

$$\boldsymbol{g}_\ell := \begin{bmatrix} \boldsymbol{H}_\ell\boldsymbol{v}/\sigma_\ell \\ \overline{\boldsymbol{H}}_\ell\overline{\boldsymbol{v}}/\sigma_\ell \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}_\ell/\sigma_\ell & \boldsymbol{0} \\ \boldsymbol{0} & \overline{\boldsymbol{H}}_\ell/\sigma_\ell \end{bmatrix}\begin{bmatrix} \boldsymbol{v} \\ \overline{\boldsymbol{v}} \end{bmatrix}. \quad (34)$$

To demonstrate the rank-1 deficiency of $\boldsymbol{F}$, it suffices to find a nonzero vector $\boldsymbol{d} \in \mathbb{C}^{2N}$ such that $\boldsymbol{F}\boldsymbol{d} = \boldsymbol{0}$. To this end, consider the vector $\boldsymbol{d} := \left[\boldsymbol{v}^{\mathcal{T}}\ -\overline{\boldsymbol{v}}^{\mathcal{T}}\right]^{\mathcal{T}} \neq \boldsymbol{0}$. It is straightforward to check that for all $\ell = 1, 2, \ldots, L$

$$\boldsymbol{g}_\ell^{\mathcal{H}}\boldsymbol{d} = \begin{bmatrix}\boldsymbol{v}^{\mathcal{H}}\boldsymbol{H}_\ell/\sigma_\ell & \overline{\boldsymbol{v}}^{\mathcal{H}}\overline{\boldsymbol{H}}_\ell/\sigma_\ell\end{bmatrix}\begin{bmatrix}\boldsymbol{v} \\ -\overline{\boldsymbol{v}}\end{bmatrix} = 0$$

therefore giving rise to $\boldsymbol{F}\boldsymbol{d} = \sum_{\ell=1}^{L}\boldsymbol{g}_\ell\left(\boldsymbol{g}_\ell^{\mathcal{H}}\boldsymbol{d}\right) = \boldsymbol{0}$. That is, for any nonzero $\boldsymbol{v}$, there always exists a nonzero vector $\boldsymbol{d} = \left[\boldsymbol{v}^{\mathcal{T}}\ -\overline{\boldsymbol{v}}^{\mathcal{T}}\right]^{\mathcal{T}}$ lying in the null space of $\boldsymbol{F}$, hence verifying the rank-1 deficiency of $\boldsymbol{F}$. This concludes the proof.

Proposition 3. *Any KKT point of problem* (15) *is a stationary point of problem* (12).

Proof of Proposition 3: Assume without loss of generality that $w_\ell = 1$ for all $1 \le \ell \le L$. It is clear that all stationary points $\boldsymbol{v}^*$ of (12) satisfy the first-order optimality condition [11, Sec. 1.1], namely,

$$\sum_{\ell=1}^{L}\left[(\boldsymbol{v}^*)^{\mathcal{H}}\boldsymbol{H}_\ell\boldsymbol{v}^* - z_\ell\right]\boldsymbol{H}_\ell\boldsymbol{v}^* = \boldsymbol{0}. \quad (35)$$

Upon introducing $\boldsymbol{\lambda} := \{\lambda_\ell \ge 0\}_{\ell=1}^{L}$ and letting $\boldsymbol{\mu} := \{\mu_\ell \ge 0\}_{\ell=1}^{L}$ denote the dual variables associated with

constraints (15b) and (15c), respectively, one can write the Lagrangian of (15) as

$$\mathcal{L}(\boldsymbol{v},\boldsymbol{s};\boldsymbol{\lambda},\boldsymbol{\mu}) := \sum_{\ell=1}^{L} s_\ell^2 + \sum_{\ell=1}^{L} \lambda_\ell(\boldsymbol{v}^{\mathcal{H}}\boldsymbol{H}_\ell\boldsymbol{v} - z_\ell - s_\ell) + \sum_{\ell=1}^{L} \mu_\ell(-\boldsymbol{v}^{\mathcal{H}}\boldsymbol{H}_\ell\boldsymbol{v} - s_\ell + z_\ell). \quad (36)$$

Any pair of primal and dual optimal points $(\boldsymbol{v}^*,\boldsymbol{s}^*,\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*)$ of (15) obeys the KKT conditions [11, Sec. 5.1]

$$2s_\ell^* - \lambda_\ell^* - \mu_\ell^* = 0, \quad 1 \le \ell \le L \quad (37)$$

$$\sum_{\ell=1}^{L} \lambda_\ell^* \boldsymbol{H}_\ell\boldsymbol{v}^* - \sum_{\ell=1}^{L} \mu_\ell^* \boldsymbol{H}_\ell\boldsymbol{v}_\ell^* = \boldsymbol{0} \quad (38)$$

$$\lambda_\ell^*((\boldsymbol{v}^*)^{\mathcal{H}}\boldsymbol{H}_\ell\boldsymbol{v}^* - z_\ell - s_\ell^*) = 0, \quad 1 \le \ell \le L \quad (39)$$

$$\mu_\ell^*((\boldsymbol{v}^*)^{\mathcal{H}}\boldsymbol{H}_\ell\boldsymbol{v}^* - z_\ell + s_\ell^*) = 0, \quad 1 \le \ell \le L \quad (40)$$

$$\lambda_\ell^* \ge 0,\ \mu_\ell^* \ge 0, \quad 1 \le \ell \le L \quad (41)$$

$$-s_\ell^* \le (\boldsymbol{v}^*)^{\mathcal{H}}\boldsymbol{H}_\ell\boldsymbol{v}^* - z_\ell \le s_\ell^*, \quad 1 \le \ell \le L. \quad (42)$$

Consider first the trivial case, where $s_\ell^* = 0$ for all $1 \le \ell \le L$. Using (42), all KKT points of the QCQP problem in (15) obeying $(\boldsymbol{v}^*)^{\mathcal{H}}\boldsymbol{H}_\ell\boldsymbol{v}^* - z_\ell = 0$ for all $1 \le \ell \le L$, and thus satisfying (35), are stationary points of the WLS problem of (12) too.

On the other hand, if there exists $s_\ell^* > 0$ for certain $\ell = 1, 2, \ldots, L$, collected in the set $\mathcal{S} \subseteq \{1, 2, \ldots, L\}$, then exactly one of λ_ℓ^* and μ_ℓ^* is 0, which can be deduced by combining conditions (37), (39), and (40). As such, assume again without loss of generality that $\mu_\ell^* = 0$ for all $\ell \in \mathcal{S}$. Appealing to (37), it holds that $\lambda_\ell^* = 2s_\ell^* > 0$ for all $\ell \in \mathcal{S}$, while all other λ_ℓ's for $\ell \notin \mathcal{S}$ as well as all μ_ℓ's for $1 \le \ell \le L$ are 0. Hence, by means of (39), we have $(\boldsymbol{v}^*)^{\mathcal{H}}\boldsymbol{H}_\ell\boldsymbol{v}^* - z_\ell = s_\ell^*$ for all $\ell \in \mathcal{S}$, which in conjunction with (38) yields

$$\sum_{\ell\in\mathcal{S}} \lambda_\ell^* \boldsymbol{H}_\ell\boldsymbol{v}^* = \sum_{\ell\in\mathcal{S}} 2s_\ell^* \boldsymbol{H}_\ell\boldsymbol{v}^* = \sum_{\ell\in\mathcal{S}} 2((\boldsymbol{v}^*)^{\mathcal{H}}\boldsymbol{H}_\ell\boldsymbol{v}^* - z_\ell)\boldsymbol{H}_\ell\boldsymbol{v}^* = \boldsymbol{0}$$

confirming $\sum_{\ell=1}^{L}((\boldsymbol{v}^*)^{\mathcal{H}}\boldsymbol{H}_\ell\boldsymbol{v}^* - z_\ell)\boldsymbol{H}_\ell\boldsymbol{v}^* = \boldsymbol{0}$, due to the fact that $(\boldsymbol{v}^*)^{\mathcal{H}}\boldsymbol{H}_\ell\boldsymbol{v}^* - z_\ell = s_\ell = 0$ for all $\ell \notin \mathcal{S} \subseteq \{1, 2, \ldots, L\}$. Therefore, any KKT point of the QCQP problem (15) is a stationary point of the WLS problem (12) as well.

REFERENCES

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

[2] V. Kekatos, G. Wang, H. Zhu, and G. B. Giannakis, "PSSE redux: Convex relaxation, decentralized, robust, and dynamic approaches," *arXiv:1708.03981*, 2017.

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

[4] A. Gómez-Expósito, A. J. Conejo, and C. Cañizares, *Electric Energy Systems: Analysis and Operation*. CRC Press, 2016.

[5] K. Lehmann, A. Grastien, and P. Van Hentenryck, "AC-feasibility on tree networks is NP-hard," *IEEE Trans. Power Syst.*, vol. 31, no. 1, pp. 798–801, Jan. 2016.

[6] X. T. Jiang, B. Fardanesh, D. Maragal, G. Stefopoulos, J. H. Chow, and M. Razanousky, "Improving performance of the non-iterative direct state estimation method," in *Power and Energy Conference at Illinois*, Champaign, IL, Feb. 2014, pp. 1–6.

[7] Y. Zhang, R. Madani, and J. Lavaei, "Conic relaxations for power system state estimation with line measurements," *IEEE Trans. Control Netw. Syst.*, 2017 (to appear).

[8] W. F. Tinney and C. E. Hart, "Power flow solution by Newton's method," *IEEE Trans. Power App. Syst.*, no. 11, pp. 1449–1460, Nov. 1967.

[9] F. C. Schweppe, J. Wildes, and D. Rom, "Power system static state estimation: Parts I, II, and III," vol. 89, pp. 120–135, Jan. 1970.

[10] H. Zhu and G. B. Giannakis, "Power system nonlinear state estimation using distributed semidefinite programming," vol. 8, no. 6, pp. 1039–1050, Dec. 2014.

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

[12] S.-J. Kim, G. Wang, and G. B. Giannakis, "Online semidefinite programming for power system state estimation," in *Proc. IEEE Conf. on Acoustics, Speech and Signal Process.*, Florence, Italy, May 2014, pp. 6024–6027.

[13] G. Wang, S.-J. Kim, and G. B. Giannakis, "Moving-horizon dynamic power system state estimation using semidefinite relaxation," in *Proc. IEEE PES General Meeting*, Washington, DC, July 2014, pp. 1–5.

[14] H. Zhu and G. B. Giannakis, "Estimating the state of AC power systems using semidefinite programming," in *Proc. North American Power Symposium*, Boston, MA, Aug. 2011, pp. 1–7.

[15] A. Minot, Y. M. Lu, and N. Li, "A distributed Gauss-Newton method for power system state estimation," *IEEE Trans. Power Syst.*, vol. 31, no. 5, pp. 3804–3815, Sept. 2016.

[16] P. Chavali and A. Nehorai, "Distributed power system state estimation using factor graphs," *IEEE Trans. Signal Process.*, vol. 63, no. 11, pp. 2864–2876, Jun. 2015.

[17] G. Wang, G. B. Giannakis, and J. Chen, "Robust and scalable power system state estimation via composite optimization," *preprint*, 2017.

[18] V. Kekatos and G. B. Giannakis, "Distributed robust power system state estimation," vol. 28, no. 2, pp. 1617–1626, May 2013.

[19] K. G. Murty and S. N. Kabadi, "Some NP-complete problems in quadratic and nonlinear programming," *Math. Program.*, vol. 39, no. 2, pp. 117–129, Jun. 1987.

[20] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Oper. Res.*, vol. 26, no. 4, pp. 681–683, Aug. 1978.

[21] O. Mehanna, K. Huang, B. Gopalakrishnan, A. Konar, and N. D. Sidiropoulos, "Feasible point pursuit and successive approximation of non-convex QCQPs," vol. 22, no. 7, pp. 804–808, Nov. 2015.

[22] F. Facchinei, L. Lampariello, and G. Scutari, "Feasible methods for nonconvex nonsmooth problems with applications in green communications," *Math. Program., Series A and B*, vol. 164, no. 1-2, pp. 55–90, July 2017.

[23] A. L. Yuille and A. Rangarajan, "The concave-convex procedure," *Neural Comput.*, vol. 15, no. 4, pp. 915–936, Apr. 2003.

[24] J. Park and S. Boyd, "General heuristics for nonconvex quadratically constrained quadratic programming," *arXiv:1703.07870*, 2017.

[25] G. R. Lanckriet and B. K. Sriperumbudur, "On the convergence of the concave-convex procedure," in *Adv. Neural Inform. Process. Syst.*, Vancouver, B.C., Canada, Dec. 2009, pp. 1759–1767.

[26] A. S. Zamzam, N. D. Sidiropoulos, and E. Dall'Anese, "Beyond relaxation and Newton-Raphson: Solving AC OPF for multi-phase systems with renewables," *IEEE Trans. Smart Grid*, 2017 (to appear).

[27] G. Wang, A. S. Zamzam, G. B. Giannakis, and N. D. Sidiropoulos, "Power system state estimation via feasible point pursuit," in *IEEE Global Conf. Signal and Inf. Process.*, Washington, D.C., USA, 2016.

[28] 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.

[29] M. K. Celik and A. Abur, "A robust WLAV state estimator using transformations," vol. 7, no. 1, pp. 106–113, Feb. 1992.

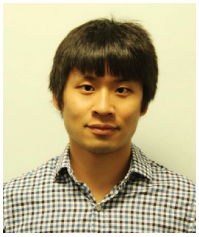
[30] A. Gomez-Exposito, A. Abur, P. Rousseaux, A. de la Villa Jaen, and C. Gomez-Quiles, "On the use of PMUs in power system state estimation," in *Proc. Power Syst. Computat. Conf.*, vol. 22, 2011.

[31] F. Alizadeh and D. Goldfarb, "Second-order cone programming," *Math. Program., Ser. B*, vol. 95, no. 1, pp. 3–51, Jan. 2003.

[32] S. M. Kay, *Fundamentals of Statistical Signal Processing, Vol. I: Estimation Theory*. Prentice Hall, 1993.

[33] P. Stoica and T. L. Marzetta, "Parameter estimation problems with singular information matrices," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 87–90, Jan. 2001.

[34] Power systems test case archive. Univ. of Washington. [Online]. Available: http://www.ee.washington.edu/research/pstca.
[35] 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," vol. 26, no. 1, pp. 12–19, Feb. 2011.
[36] J. Lofberg, "A toolbox for modeling and optimization in MATLAB," in *Proc. of the CACSD Conf.*, 2004. [Online]. Available: http://users.isy.liu.se/johanl/yalmip/.
[37] G. Wang, G. B. Giannakis, and Y. C. Eldar, "Solving systems of random quadratic equations via truncated amplitude flow," *IEEE Trans. Inf. Theory,* 2017 (to appear)*;* see also *arXiv:1605.08285,* 2016.
[38] K. Kreutz-Delgado, "The complex gradient operator and the CR-calculus," *arXiv:0906.4835*, 2009.
[39] A. Van den Bos, "A Cramér-Rao lower bound for complex parameters," *IEEE Trans. Signal Process.*, vol. 40, no. 10, Oct. 1994.



Gang Wang (S'12) received the B.Eng. degree in Electrical Engineering and Automation from the Beijing Institute of Technology, Beijing, China, in 2011. He is currently working toward his Ph.D. degree in the Department of Electrical and Computer Engineering at the University of Minnesota, Minneapolis, MN, USA.

His research interests focus on the areas of high-dimensional statistical signal processing, stochastic and nonconvex optimization with applications to autonomous energy grids, and deep learning. He received a National Scholarship (2014), a Guo Rui Scholarship (2017), and twice the Ministry of Industry and Information Technology Innovation Scholarship (2016, 2017), all from China, a Best Student Paper Award at the 2017 European Signal Processing Conference, and the Student Travel Awards from the NSF (2016) and NIPS (2017).

Ahmed S. Zamzam (S'14) received the B.Sc. (with highest honors) from Cairo University, Giza, Egypt, in 2013, and the M.Sc. degree in wireless technologies from Nile University, Giza, Egypt, in 2015. Since September 2015, he has been working towards his Ph.D. degree in the Department of Electrical and Computer Engineering at the University of Minnesota. His research interests include power systems, optimization, and signal processing.

Georgios B. Giannakis (F'97) received his Diploma in Electrical Engr. from the Ntl. Tech. Univ. of Athens, Greece, 1981. From 1982 to 1986 he was with the Univ. of Southern California (USC), where he received his MSc. in Electrical Engineering, 1983, MSc. in Mathematics, 1986, and Ph.D. in Electrical Engr., 1986. He was with the University of Virginia from 1987 to 1998, and since 1999 he has been a professor with the Univ. of Minnesota, where he holds an Endowed Chair in Wireless Telecommunications, a University of Minnesota McKnight Presidential Chair in ECE, and serves as director of the Digital Technology Center.

His general interests span the areas of communications, networking and statistical signal processing - subjects on which he has published more than 400 journal papers, 680 conference papers, 25 book chapters, two edited books and two research monographs (h-index 128). 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 28 patents issued, and the (co-) recipient of 8 best paper awards from the IEEE Signal Processing (SP) and Communications Societies, including the G. Marconi Prize Paper Award in Wireless Communications. He also received Technical Achievement Awards from the SP Society (2000), from 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 (2015). He is a Fellow of EURASIP, and has served the IEEE in a number of posts, including that of a Distinguished Lecturer for the IEEE-SP Society.

Nicholas D. Sidiropoulos (F'09) received the Diploma degree in electrical engineering from Aristotelian University of Thessaloniki, Thessaloniki, Greece, and the M.S. and Ph.D. degrees in electrical engineering from the University of Maryland-College Park, College Park, MD, USA, in 1988, 1990, and 1992, respectively. He has served on the faculty of the University of Virginia (UVA), University of Minnesota, and the Technical University of Crete, Greece, prior to his current appointment as Chair of ECE at UVA.

His research interests are in signal processing, communications, optimization, tensor decomposition, and factor analysis, with applications in machine learning and communications. He received the NSF/CAREER award in 1998, the IEEE Signal Processing Society (SPS) Best Paper Award in 2001, 2007, and 2011, served as IEEE SPS Distinguished Lecturer (2008-2009), and currently serves as Vice President - Membership of IEEE SPS. He received the 2010 IEEE Signal Processing Society Meritorious Service Award, and the 2013 Distinguished Alumni Award from the University of Maryland, Dept. of ECE. He is a Fellow of IEEE (2009) and a Fellow of EURASIP (2014).