Robust and Scalable Power System State Estimation via Composite Optimization

Gang Wang, *Student Member, IEEE*, Georgios B. Giannakis, *Fellow, IEEE*, and Jie Chen, *Senior Member, IEEE*

***Abstract*—In today's cyber-enabled smart grids, high penetration of uncertain renewables, purposeful manipulation of meter readings, and the need for wide-area situational awareness, call for fast, accurate, and robust power system state estimation. The least-absolute-value (LAV) estimator is known for its robustness relative to the weighted least-squares (WLS) one. However, due to nonconvexity and nonsmoothness, existing LAV solvers based on linear programming are typically slow, hence inadequate for real-time system monitoring. This paper develops two novel algorithms for efficient LAV estimation, which draw from recent advances in composite optimization. The first is a deterministic linear proximal scheme that handles a sequence of convex quadratic problems, each efficiently solvable either via off-the-shelf algorithms or through the alternating direction method of multipliers. Leveraging the sparse connectivity inherent to power networks, the second scheme is stochastic, and updates only *a few* entries of the complex voltage state vector per iteration. In particular, when voltage magnitude and (re)active power flow measurements are used only, this number reduces to one or two, *regardless of* the number of buses in the network. This computational complexity evidently scales well to large-size power systems. Furthermore, by carefully *mini-batching* the voltage and power flow measurements, accelerated implementation of the stochastic iterations becomes possible. The developed algorithms are numerically evaluated using a variety of benchmark power networks. Simulated tests corroborate that improved robustness can be attained at comparable or markedly reduced computation times for medium- or large-size networks relative to the "workhorse" WLS-based Gauss-Newton iterations.**

***Index terms*— SCADA measurements, cyberattacks, alternating direction method of multipliers, prox-linear algorithms.**

I. Introduction

The North American electric grid, the largest machine on earth, is recognized as the greatest engineering achievement of the 20th century [1]: thousands of miles of transmission lines and millions of miles of distribution lines, linking thousands of power plants to millions of factories and homes. Accurately monitoring the grid's operating condition is critical to several control and optimization tasks, including optimal power flow, reliability analysis, attack detection, and future network expansion planning [2].

The work was supported in part by NSF grants 1423316, 1442686, 1508993, and 1509040. G. Wang and G. B. Giannakis are with the Digital Technology Center and the Electrical and Computer Engineering Department, University of Minnesota, Minneapolis, MN 55455, USA. G. Wang is also with the State Key Lab of Intelligent Control and Decision of Complex Systems, Beijing Institute of Technology, Beijing 100081, P. R. China. J. Chen is with the State Key Lab of Intelligent Control and Decision of Complex Systems, Beijing Institute of Technology, Beijing 100081, P. R. China. E-mails: {gangwang, georgios}@umn.edu; chenjie@bit.edu.cn.

To enable grid-wide monitoring, power system engineers in the 1960s pursued voltages at critical buses based on readings collected from current and potential transformers. But the power flow equations were never feasible due to timing and modeling inaccuracies. In a seminal work [3], the statistical foundation was laid for power system state estimation (PSSE), whose central task is to obtain the voltage magnitude and angle information at all buses given network parameters and measurements acquired from across the grid. Since then, there have been numerous PSSE contributions; see [2] for a recent review of PSSE advances, some of which are outlined below.

Power grids are primarily monitored by supervisory control and data acquisition (SCADA) systems. Parameter uncertainty, instrument mis-calibration, and unmonitored topology changes can however, yield grossly corrupted SCADA measurements (a.k.a. 'bad data') [4]. Designed for functionality and efficiency with little attention paid to security, today's SCADA systems are vulnerable to cyberattacks [5]. Bad data also come in the form of purposeful manipulation of smart meter readings, as asserted by the first hacker-caused power outage: the 2015 Ukraine blackout [6]. Any of these events can happen which will cause a given data collection to be much more inaccurate than is assumed by popular mathematical models. Efficient robust PSSE approaches against cyber threats are thus well motivated in the smart grid context [7].

Commonly used PSSE criteria include the weighted least-squares (WLS) and the least-absolute value (LAV) [8]. The former coincides with the maximum likelihood criterion when additive white Gaussian noise is assumed. Unfortunately, WLS comes with a number of limitations. First, obtaining the WLS estimate based on SCADA measurements amounts to minimizing a nonconvex quartic polynomial. As a result, the quality of the resultant iterative estimators (e.g., the Gauss-Newton one) relies heavily on the initialization. Further, the convergence of Gauss-Newton iterations for nonconvex objectives is hardly guaranteed in general [9]. As least as important, WLS estimators are sensitive to bad data [4]. They may yield very poor estimates in the presence of outliers. These issues were somewhat mitigated by incorporating the largest normalized residual (LNR) test for bad data removal [4], or, via reformulating the (possibly regularized) WLS into a semidefinite program (SDP) via convex relaxation [10], [11], [12], [13], [14]. The former alternates between the LNR test and the estimation, while the latter solves SDPs. The least-median-squares and the least-trimmed-squares estimators have provably improved performance under certain conditions [15]. Unfortunately, their computational complexities scale unfavorably with the number of buses in the network [2].

On the other hand, LAV estimators simultaneously identify and reject bad data while acquiring an accurate estimate of the state [16]. Recent research efforts have focused on dealing with the nonconvexity and nonsmoothness in LAV estimation. Upon linearizing the nonlinear measurement functions at the most recent iterate, a series of linear programs was solved [16]. Techniques for improving the linear programming by exploiting the system's structure [17], or via iterative reweighting [18] have also been reported. LAV estimation based only on PMU data was studied [19], [20], in which a strategic scaling was suggested to eliminate the effect of leverage measurements [15]. Despite these efforts, LAV estimators have not been widely employed yet in today's power networks due mostly to their computational inefficiency [19].

The LAV-based PSSE is revisited in this work from the viewpoint of composite optimization [21], [22], which considers minimizing functions $f(\mathbf{v}) = h(c(\mathbf{v}))$ that are compositions of a convex function h, and a smooth vector function c. Two novel proximal linear (prox-linear) procedures are developed based upon minimizing a sequence of convex quadratic subproblems. The first deterministic LAV solver minimizes functions constructed from a linearized approximation to the original objective and a quadratic regularization, each efficiently implementable using either off-the-shelf solvers, or, the alternating direction method of multipliers (ADMM). The convergence of such deterministic prox-linear algorithms has been documented [22], [23].

The second LAV solver builds on a stochastic prox-linear algorithm, and has each iteration minimizing the summation of a linearized approximation to the LAV loss of a single measurement and the regularization term [24]. Interestingly, each iteration of the stochastic LAV solver has a closed-form update. Thanks to the sparse connectivity inherent to power networks, this amounts to updating very few entries of the state vector. In particular, if the available measurements include voltage magnitudes and power flows only, each stochastic iteration performs about 10 scalar operations, which is *independent of* the network size. If power injections are included too, the number of operations increases to the order of the number of neighboring buses. This complexity evidently scales favorably to large-size power grids. Moreover, even faster implementation of the stochastic solver is realized by means of judiciously mini-batching the measurements. Finally, the novel algorithms are numerically tested using the IEEE 14-, 118-bus, and the PEGASE $9,241$-bus benchmark networks. Simulations corroborate their merits relative to the WLS-based Gauss-Newton iterations.

Outline. Grid modeling and problem formulation are given in Section II. Upon reviewing the basics of composite optimization, Section III presents the deterministic LAV solver, followed by its stochastic alternative in Section IV. Numerical tests are presented in Section V, while the paper is concluded in Section VI.

Notation. Matrices (column vectors) are denoted by upper- (lower-) case boldface letters. Symbols $(\cdot)^{\mathcal{T}}$ and $(\cdot)^{\mathcal{H}}$ represent (Hermitian) transpose, and $\overline{(\cdot)}$ complex conjugate. Sets are denoted using calligraphic letters. Symbol $\Re(\cdot)$ ($\Im(\cdot)$) takes the real (imaginary) part of a complex number. Operator $\mathrm{dg}(\mathbf{x}_i)$ defines a diagonal matrix holding entries of $\mathbf{x}_i$ on its diagonal, while $[\mathbf{x}_i]_{1\leq i\leq N}$ returns a matrix with $\mathbf{x}_i^{\mathcal{H}}$ being its i-th row.

II. Grid Modeling and Problem Formulation

An electric grid having N buses and L lines is modeled as a graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, whose nodes $\mathcal{N} := \{1, 2, \ldots, N\}$ correspond to buses and whose edges $\mathcal{E} := \{(n, n')\} \subseteq \mathcal{N}\times\mathcal{E}$ to lines. The complex voltage per bus $n \in \mathcal{N}$ is expressed in rectangular coordinates as $v_n = \Re(v_n) + j\Im(v_n)$, with all nodal voltages forming the vector $\mathbf{v} := [v_1 \ \cdots \ v_N]^{\mathcal{H}} \in \mathbb{C}^N$.

The voltage magnitude square $V_n := |v_n|^2 = \Re^2(v_n) + \Im^2(v_n)$ can be compactly expressed as a quadratic function of $\mathbf{v}$, namely

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

where $\mathbf{e}_n$ denotes the n-th canonical vector in $\mathbb{R}^N$. To express power injections as functions of $\mathbf{v}$, introduce the so-termed bus admittance matrix $\mathbf{Y} = \mathbf{G} + j\mathbf{B} \in \mathbb{C}^N$ [2]. In rectangular coordinates, the active and reactive power injections p_n and q_n at bus n are given by

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

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

which admits a compact representation as

$$p_n = \mathbf{v}^{\mathcal{H}}\mathbf{H}_n^p\mathbf{v}, \ \ \text{where} \ \ \mathbf{H}_n^p := \frac{\mathbf{Y}^{\mathcal{H}}\mathbf{e}_n\mathbf{e}_n^{\mathcal{T}} + \mathbf{e}_n\mathbf{e}_n^{\mathcal{T}}\mathbf{Y}}{2} \tag{4a}$$

$$q_n = \mathbf{v}^{\mathcal{H}}\mathbf{H}_n^q\mathbf{v}, \ \ \text{where} \ \ \mathbf{H}_n^q := \frac{\mathbf{Y}^{\mathcal{H}}\mathbf{e}_n\mathbf{e}_n^{\mathcal{T}} - \mathbf{e}_n\mathbf{e}_n^{\mathcal{T}}\mathbf{Y}}{2j}. \tag{4b}$$

Recognize that the line current from bus n to n' obeys $I_{nn'} = \mathbf{e}_{nn'}^{\mathcal{T}}\mathbf{i}_f = \mathbf{e}_{nn'}^{\mathcal{T}}\mathbf{Y}_f\mathbf{v}$, where $\mathbf{i}_f \in \mathbb{C}^{|\mathcal{E}|}$ collects all line currents, and $\mathbf{Y}_f \in \mathbb{C}^{|\mathcal{E}|\times N}$ relates the bus voltages to all line currents at the 'from' (sending) end. Ohm's and Kirchhoff's laws assert that the 'from-end' power flow over line (n, n') can be expressed as $\overline{S}_{nn'}^f = P_{nn'}^f - jQ_{nn'}^f = v_n^* i_{nn'}^f = (\mathbf{v}^{\mathcal{H}}\mathbf{e}_n)(\mathbf{e}_{nn'}^{\mathcal{T}}\mathbf{i}_f) = \mathbf{v}^{\mathcal{H}}\mathbf{e}_n\mathbf{e}_{nn'}^{\mathcal{T}}\mathbf{Y}_f\mathbf{v}$, yielding

$$P_{nn'}^f = \mathbf{v}^{\mathcal{H}}\mathbf{H}_{nn'}^P\mathbf{v}, \ \text{where} \ \mathbf{H}_{nn'}^P := \frac{\mathbf{Y}_f^{\mathcal{H}}\mathbf{e}_{nn'}\mathbf{e}_n + \mathbf{e}_n\mathbf{e}_{nn'}^{\mathcal{T}}\mathbf{Y}_f}{2} \tag{5a}$$

$$Q_{nn'}^f = \mathbf{v}^{\mathcal{H}}\mathbf{H}_{nn'}^Q\mathbf{v}, \ \text{where} \ \mathbf{H}_{nn'}^Q := \frac{\mathbf{Y}_f^{\mathcal{H}}\mathbf{e}_{nn'}\mathbf{e}_n - \mathbf{e}_n\mathbf{e}_{nn'}^{\mathcal{T}}\mathbf{Y}_f}{2j}. \tag{5b}$$

The active and reactive power flows measured at the 'to' (receiving) ends $P_{nn'}^t$ and $Q_{nn'}^t$ can be written symmetrically to $P_{nn'}^f$ and $Q_{nn'}^f$; and hence, they are omitted here for brevity.

Given line parameters collected in $\mathbf{Y}$ and $\mathbf{Y}_f$, all SCADA measurements including squared voltage magnitudes as well

as active and reactive power injections and flows can be expressed as quadratic functions of the voltages $\mathbf{v} \in \mathbb{C}^N$. This justifies why $\mathbf{v}$ is referred to as the system state. If $\mathcal{S}_V$, $\mathcal{S}_p$, $\mathcal{S}_q$, $\mathcal{S}_P^f$, $\mathcal{S}_Q^f$, $\mathcal{S}_P^t$, and $\mathcal{S}_Q^t$ signify the smart meter locations of the corresponding type, we have available the following (possibly noisy or even corrupted) measurements: $\{\check{V}_n\}_{n\in\mathcal{S}_V}$, $\{\check{p}_n\}_{n\in\mathcal{S}_p}$, $\{\check{q}_n\}_{n\in\mathcal{S}_q}$, $\{\check{P}^f_{nn'}\}_{(n,n')\in\mathcal{S}_P^f}$, $\{\check{Q}^f_{nn'}\}_{(n,n')\in\mathcal{S}_Q^f}$, $\{\check{P}^t_{nn'}\}_{(n,n')\in\mathcal{S}_P^t}$, and $\{\check{Q}^t_{nn'}\}_{(n,n')\in\mathcal{S}_Q^t}$, henceforth concatenated in the vector $\mathbf{z} \in \mathbb{R}^M$, where M denotes the total number of measurements.

In this paper, the following corruption model is considered [25]: If $\{\xi_i\} \subseteq \mathbb{R}$ models an arbitrary attack (or outlier) sequence, given the measurement matrices $\{\mathbf{H}_m\}_{m=1}^M$ in (1)-(5a), we observe for $1 \leq m \leq M$ the samples

$$z_m \approx \begin{cases} \mathbf{v}^{\mathcal{H}}\mathbf{H}_m\mathbf{v} & \text{if } m \in \mathcal{I}^{nom} \\ \xi_m & \text{if } m \in \mathcal{I}^{out} \end{cases} \tag{6}$$

where additive measurement noise can be included if $\approx$ is replaced by equality, and $\mathcal{I}^{nom}, \mathcal{I}^{out} \subseteq \{1, 2, \ldots, M\}$ collect the indices of nominal data and outliers, respectively. In other words, $\mathcal{I}^{out}$ is the set of meter indices that can be compromised. The indices in $\mathcal{I}^{out}$ are assumed chosen randomly from $\{1, 2, \ldots, M\}$. Instrument failures occur at random, although the attack sequence $\{\xi_m\}$ may rely on $\{\mathbf{H}_m\}$ (even adversarially). Specifically, two models will be considered for the attacks.

M1 Matrices $\{\mathbf{H}_m\}_{m=1}^M$ are independent of $\{\xi_m\}_{m=1}^M$.

M2 Nominal measurement matrices $\{\mathbf{H}_m\}_{m\in\mathcal{I}^{nom}}$ are independent of $\{\xi_m\}_{m\in\mathcal{I}^{out}}$.

It is worth highlighting that M1 requires full independence between the corruption and measurements. That is, the attacker may only corrupt ξ_m without knowing $\mathbf{H}_m$. On the contrary, M2 allows completely arbitrary dependence between ξ_m and $\mathbf{H}_m$ for $m \in \mathcal{I}^{out}$, which is natural as the type of corruption may also rely on the individual measurement $\mathbf{H}_m$ being taken.

Having elaborated on the system and corruption models, the PSSE problem can be stated as follows: Given matrices $\mathbf{Y}$, $\mathbf{Y}_f$, and the available measurements $\mathbf{z} \in \mathbb{R}^M$, with entries as in (6) obeying M1 or M2, recover the voltage vector $\mathbf{v} \in \mathbb{C}^N$. The first attempt may be seeking the WLS estimate, or the ML one when assuming independent Gaussian noise [3]. It is known however that the WLS criterion is sensitive to outliers, and may yield very bad estimates even if there are few grossly corrupted measurements [4]. As is well documented in statistics and optimization, the ℓ_1-based losses yield median-based estimators [26], and handle gross errors in the measurements $\mathbf{z}$ in a relatively benign way. Prompted by this, we will consider here minimizing the ℓ_1 loss of the residuals, which leads to the so-called LAV estimate [16]

$$\underset{\mathbf{v}\in\mathbb{C}^N}{\text{minimize}} \quad f(\mathbf{v}) := \frac{1}{M}\sum_{m=1}^{M} \left|\mathbf{v}^{\mathcal{H}}\mathbf{H}_m\mathbf{v} - z_m\right|. \tag{7}$$

Because of $\{\mathbf{v}^{\mathcal{H}}\mathbf{H}_m\mathbf{v}\}_{m=1}^M$ and the absolute-value operation, the LAV objective in (7) is nonsmooth, nonconvex, and not even locally convex near the optima $\pm\mathbf{v}^*$. This is clear from the scalar case $f(v) = |v^*v - 1|$, where $v \in \mathbb{R}$. A local analysis based on convexity and smoothness is thus impossible, and $f(\mathbf{v})$ is in general difficult to minimize. For this reason, Gauss-Newton is not applicable to minimize (7). Nevertheless, the criterion $f(\mathbf{v})$ possesses several unique structural properties, which we exploit next to develop efficient algorithms.

III. Deterministic Prox-Linear LAV Solver

In this section, we will develop a deterministic solver of (7). To that end, let us start rewriting the objective in (7) as

$$\underset{\mathbf{v}\in\mathbb{C}^N}{\text{minimize}} \quad f(\mathbf{v}) := h(c(\mathbf{v})) \tag{8}$$

the composition of a convex function $h : \mathbb{R}^M \to \mathbb{R}$, and a smooth vector function $c : \mathbb{C}^N \to \mathbb{R}^M$, a structure that is known to be amenable to efficient algorithms [21], [22]. It is clear that this general form subsumes (7) as a special case, for which we can take $h(\mathbf{u}) = (1/M)\|\mathbf{u}\|_1$ and $c(\mathbf{v}) = \left[\mathbf{v}^{\mathcal{H}}\mathbf{H}_m\mathbf{v} - z_m\right]_{1\leq m\leq M}$. The compositional structure lends itself well to the proximal linear (prox-linear) algorithm, which is a variant of the Gauss-Newton iterations [21]. Specifically, define close to a given $\mathbf{v}$ the local "linearization" of f as

$$f_{\mathbf{v}}(\mathbf{w}) := h\big(c(\mathbf{v}) + \nabla^{\mathcal{H}}c(\mathbf{v})(\mathbf{w} - \mathbf{v})\big) \tag{9}$$

where $\nabla c(\mathbf{v}) \in \mathbb{C}^{N\times M}$ denotes the Jacobian matrix of c at $\mathbf{v}$ based on Wirtinger derivatives for functions of complex-valued variables [27, Appendix]. In contrast to the nonconvex $f(\mathbf{v})$, function $f_{\mathbf{v}}(\mathbf{w})$ in (9) is convex in $\mathbf{w}$, which is the key behind the prox-linear method. Starting with some point $\mathbf{v}_0$, which can be the flat-voltage profile point, namely the all-one vector, construct the iteration

$$\mathbf{v}_{t+1} := \underset{\mathbf{v}\in\mathbb{C}^N}{\arg\min}\left\{f_{\mathbf{v}_t}(\mathbf{v}) + \frac{1}{2\mu_t}\|\mathbf{v} - \mathbf{v}_t\|_2^2\right\} \tag{10}$$

where $\mu_t > 0$ is a stepsize that can be fixed in advance, or be determined by a line search [22].

Evidently, the subproblem (10) to be solved at every iteration of the prox-linear algorithm is convex, and can be handled by off-the-shelf solvers such as CVX [28]. However, these interior-point based solvers may not scale well when the matrices $\{\mathbf{H}_m\}$ are large. For this reason, we derive next a more efficient iterative procedure using ADMM iterations [29].

When specifying f to be the LAV objective of (7), the minimization in (10) becomes

$$\mathbf{v}_{t+1} = \arg\min_{\mathbf{v}\in\mathbb{C}^N} \ \|\mathbf{A}_t(\mathbf{v} - \mathbf{v}_t) - \mathbf{c}_t\|_1 + \frac{1}{2}\|\mathbf{v} - \mathbf{v}_t\|_2^2 \tag{11}$$

with coefficients given by

$$\mathbf{A}_t := \left[(2\mu_t/M)\mathbf{v}_t^{\mathcal{H}}\mathbf{H}_m\right]_{1\leq m\leq M} \tag{12a}$$

$$\mathbf{c}_t := \left[(\mu_t/M)(z_m - \mathbf{v}_t^{\mathcal{H}}\mathbf{H}_m\mathbf{v}_t)\right]_{1\leq m\leq M}. \tag{12b}$$

For brevity, let $\mathbf{w} := \mathbf{v} - \mathbf{v}_t$, and rewrite (11) equivalently as a constrained optimization problem

$$\underset{\mathbf{u}\in\mathbb{C}^M,\ \mathbf{w}\in\mathbb{C}^N}{\text{minimize}} \quad \|\mathbf{u} - \mathbf{c}_t\|_1 + \frac{1}{2}\|\mathbf{w}\|_2^2 \tag{13a}$$

$$\text{subject to} \quad \mathbf{A}_t\mathbf{w} = \mathbf{u}. \tag{13b}$$

To decouple constraints and also facilitate the implementation of ADMM, introduce an auxiliary copy $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{w}}$ for $\mathbf{u}$ and $\mathbf{w}$ accordingly, and rewrite (13) into

$$\underset{\tilde{\mathbf{u}},\,\tilde{\mathbf{w}},\,\mathbf{u},\,\mathbf{w}}{\text{minimize}} \quad \|\tilde{\mathbf{u}}-\mathbf{c}_t\|_1+\frac{1}{2}\|\tilde{\mathbf{w}}\|_2^2 \tag{14a}$$

$$\text{subject to} \quad \tilde{\mathbf{u}}=\mathbf{u},\ \tilde{\mathbf{w}}=\mathbf{w},\ \mathbf{A}_t\mathbf{w}=\mathbf{u}. \tag{14b}$$

Letting $\boldsymbol{\lambda}\in\mathbb{C}^N$ and $\boldsymbol{\nu}\in\mathbb{C}^M$ be the Lagrange multipliers corresponding to the $\mathbf{w}$- and $\mathbf{u}$-consensus constraints, respectively, the augmented Lagrangian after leaving out the last equality in (14b) can be expressed as

$$\begin{aligned}\mathcal{L}(\tilde{\mathbf{w}},\tilde{\mathbf{u}},\mathbf{w},\mathbf{u};\boldsymbol{\lambda},\boldsymbol{\nu}) :=\ & \|\tilde{\mathbf{u}}-\mathbf{c}_t\|_1+\frac{1}{2}\|\tilde{\mathbf{w}}\|_2^2+\boldsymbol{\lambda}^{\mathcal{H}}(\tilde{\mathbf{w}}-\mathbf{w})\\ &+\boldsymbol{\nu}^{\mathcal{H}}(\tilde{\mathbf{u}}-\mathbf{u})+\frac{\rho}{2}\|\tilde{\mathbf{w}}-\mathbf{w}\|_2^2+\frac{\rho}{2}\|\tilde{\mathbf{u}}-\mathbf{u}\|_2^2\end{aligned} \tag{15}$$

where $\rho>0$ is a predefined learning rate. With $k\in\mathbb{N}$ denoting the iteration index for solving (13), or equivalently (10), ADMM cycles through the following recursions

$$\tilde{\mathbf{w}}^{k+1}:=\arg\min_{\tilde{\mathbf{w}}}\left\{\frac{1}{2}\|\tilde{\mathbf{w}}\|_2^2+\frac{\rho}{2}\left\|\tilde{\mathbf{w}}-(\mathbf{w}^k-\boldsymbol{\lambda}^k)\right\|_2^2\right\} \tag{16a}$$

$$\tilde{\mathbf{u}}^{k+1}:=\arg\min_{\tilde{\mathbf{u}}}\left\{\frac{1}{2}\|\tilde{\mathbf{u}}-\mathbf{c}_t\|_1+\frac{\rho}{2}\left\|\tilde{\mathbf{u}}-(\mathbf{u}^k-\boldsymbol{\nu}^k)\right\|_2^2\right\} \tag{16b}$$

$$\begin{aligned}&\left\{\mathbf{w}^{k+1},\mathbf{u}^{k+1}\right\}:=\\ &\quad\underset{\mathbf{w},\,\mathbf{u}}{\arg\min}\ \left\|\mathbf{w}-(\tilde{\mathbf{w}}^{k+1}+\boldsymbol{\lambda}^k)\right\|_2^2+\left\|\mathbf{u}-(\tilde{\mathbf{u}}^{k+1}+\boldsymbol{\nu}^k)\right\|_2^2\\ &\quad\text{s. to}\quad \mathbf{A}_t\mathbf{w}=\mathbf{u}\end{aligned} \tag{16c}$$

$$\begin{bmatrix}\boldsymbol{\lambda}^{k+1}\\ \boldsymbol{\nu}^{k+1}\end{bmatrix}=\begin{bmatrix}\boldsymbol{\lambda}^k+(\tilde{\mathbf{w}}^{k+1}-\mathbf{w}^{k+1})\\ \boldsymbol{\nu}^k+(\tilde{\mathbf{u}}^{k+1}-\mathbf{u}^{k+1})\end{bmatrix} \tag{16d}$$

where all the dual variables have been scaled by the factor $\rho>0$ [29]. Interestingly enough, the solutions of (16a)-(16c) can be provided in closed form, as we elaborate next.

Proposition 1. *The solutions of problems* (16a) *and* (16b) *are*

$$\tilde{\mathbf{w}}^{k+1}:=\frac{\rho}{1+\rho}\left(\mathbf{w}^k-\boldsymbol{\lambda}^k\right) \tag{17a}$$

$$\tilde{\mathbf{u}}^{k+1}:=\mathbf{c}_t+\mathcal{S}_{1/\rho}\left(\mathbf{u}^k-\boldsymbol{\nu}^k-\mathbf{c}_t\right) \tag{17b}$$

where the shrinkage operator $\mathcal{S}_\tau(\mathbf{x}):\mathbb{C}^N\times\mathbb{R}_+\to\mathbb{R}^N$ *is* $\mathcal{S}_\tau(\mathbf{x}):=\text{sign}(\mathbf{x})\odot\max(|\mathbf{x}|-\tau\mathbf{1},0)$ *[30], with* $\odot$ *and* $|\cdot|$ *denoting the entrywise multiplication and modulus operators, respectively, and* $\text{sign}(x):=\begin{cases}x/|x|, & x\neq 0\\ 0, & x=0\end{cases}$ *provides an entrywise definition of* $\text{sign}(\mathbf{x})$.

The constrained minimization of (16c) essentially projects the pair $(\tilde{\mathbf{w}}^{k+1}+\boldsymbol{\lambda}^k,\tilde{\mathbf{u}}^{k+1}+\boldsymbol{\nu}^k)$ onto the convex set specified by the linear equality constraint, namely $\{(\mathbf{w},\mathbf{u}):\mathbf{A}_t\mathbf{w}=\mathbf{u}\}$. Its solution is derived in a simple closed form next (see the Appendix for a proof).

Proposition 2. *Given* $\mathbf{b}\in\mathbb{C}^N$ *and* $\mathbf{d}\in\mathbb{C}^M$*, the solution of*

$$\begin{aligned}\underset{\mathbf{w}\in\mathbb{C}^N,\,\mathbf{u}\in\mathbb{C}^M}{\text{minimize}}\quad & \frac{1}{2}\|\mathbf{w}-\mathbf{b}\|_2^2+\frac{1}{2}\|\mathbf{u}-\mathbf{d}\|_2^2\\ \text{subject to}\quad & \mathbf{A}\mathbf{w}=\mathbf{u}\end{aligned}$$

TABLE I: Deterministic LAV Solver Using ADMM

1: Input data $\{(z_m,\mathbf{H}_m)\}_{m=1}^M$, stepsizes μ, $\rho>0$, and initialization $\mathbf{v}_0=\mathbf{1}$.
2: **For** $t=0,1,\ldots,$ **do**
3: Initialize $\mathbf{w}^0$, $\mathbf{u}^0$, $\boldsymbol{\lambda}^0$, and $\boldsymbol{\nu}^0$ to zero.
4: Evaluate $\mathbf{A}_t$ and $\mathbf{c}_t$ as in (12).
5: **For** $k=0,1,\ldots,$ **do**
6: Update $(\tilde{\mathbf{w}}^{k+1},\tilde{\mathbf{u}}^{k+1})$, $(\mathbf{w}^{k+1},\mathbf{u}^{k+1})$, and $(\boldsymbol{\mu}^{k+1},\boldsymbol{\nu}^{k+1})$ using (17), (19), and (16d), respectively.
7: **End for**
8: Update $\mathbf{v}_{t+1}$ via (20).
9: **End for**

is given as

$$\mathbf{w}^*:=\left(\mathbf{I}+\mathbf{A}^{\mathcal{H}}\mathbf{A}\right)^{-1}\left(\mathbf{b}+\mathbf{A}^{\mathcal{H}}\mathbf{d}\right) \tag{18a}$$

$$\mathbf{u}^*:=\mathbf{A}\mathbf{w}^*. \tag{18b}$$

Using Proposition 2, the minimizer of (16c) is found as

$$\mathbf{w}^{k+1}:=(\mathbf{I}+\mathbf{A}_t^{\mathcal{H}}\mathbf{A}_t)^{-1}\left[\tilde{\mathbf{w}}^{k+1}+\boldsymbol{\lambda}^k+\mathbf{A}^{\mathcal{H}}(\tilde{\mathbf{u}}^{k+1}+\boldsymbol{\nu}^k)\right] \tag{19a}$$

$$\mathbf{u}^{k+1}:=\mathbf{A}\mathbf{w}^{k+1}. \tag{19b}$$

The four updates in (16) are computationally simple except for the matrix inversion of (19a), which nevertheless can be cached once computed during the first iteration. In addition, variables $\tilde{\mathbf{u}}^0$, $\boldsymbol{\lambda}^0$, and $\boldsymbol{\nu}^0$ of ADMM can be initialized to zero. Finally, the solution of (11) can be obtained as

$$\mathbf{v}_{t+1}:=\mathbf{v}_t+\mathbf{w}^* \tag{20}$$

where $\mathbf{w}^*$ is the converged $\mathbf{w}$-iterate of the ADMM iterations in (17), (19), and (16d).

The novel deterministic LAV solver based on ADMM is summarized in Table I, in which the inner loop consisting of Steps 3-8 can be replaced with off-the-shelf solvers to solve (11) for $\mathbf{v}_{t+1}$.

As far as performance is concerned, if h is L-Lipschitz and ∇c is β-Lipschitz, then taking a constant stepsize $\mu\leq 1/(L\beta)$ in (10) guarantees that [23, Sec. 5]:

i) the proposed solver in Table I is a descent method; and,

ii) the iterate sequence $\{\mathbf{v}_t\}$ converges to a stationary point of the LAV objective in (7).

The computational burden of the ADMM based deterministic solver is dominated by the projection step of (19), which incurs per-iteration computational complexity on the order of $\mathcal{O}(MN^2)$. This complexity can be afforded in small- or medium-size PSSE tasks, but may not be efficient enough for nowadays increasingly interconnected power networks. This motivates our stochastic alternative of the ensuing section that relies on very inexpensive iterations.

IV. Stochastic Prox-linear LAV Solver

Finding the minimizer of (11) exactly per iteration of the deterministic LAV solver may be computationally expensive, and can be intractable when the network size grows very large. Considering the wide applicability of LAV estimation as well

as the increasing interconnection of microgrids, scalable online and stochastic approaches become of substantial interest. In this section, a stochastic linear proximal algorithm of [24] is adapted to our PSSE task, and enables the prox-linear method to efficiently solve the LAV estimation problem at scale. Advantages of the stochastic approaches over their deterministic counterparts include oftentimes simple closed-form updates as well as faster convergence to yield an (approximately) optimal solution. Leveraging the sparsity structure of the measurement matrices and judiciously grouping measurements into small mini-batches can considerably speedup implementation of the stochastic solver.

A. Stochastic LAV solver

Instead of dealing with the quadratic subproblems in (11), each iteration of the stochastic LAV solver samples a datum $m_t \in \{1, 2, \ldots, M\}$ uniformly at random from the total M of measurements, and substitutes functions (h, c) by (h_{m_t}, c_{m_t}) associated with the sampled datum in the local linearization of (9), hence also in (11), yielding

$$\mathbf{v}_{t+1} := \arg\min_{\mathbf{v}\in\mathbb{C}^N} \left\{ \left|\mathbf{a}_{m_t}^{\mathcal{H}}(\mathbf{v}-\mathbf{v}_t) - c_{m_t}\right| + \frac{1}{2\mu_t}\|\mathbf{v}-\mathbf{v}_t\|_2^2 \right\} \tag{21}$$

where the coefficients are given by

$$\mathbf{a}_{m_t} := 2\mathbf{H}_{m_t}\mathbf{v}_t \tag{22a}$$

$$c_{m_t} := z_{m_t} - \mathbf{v}_t^{\mathcal{H}}\mathbf{H}_{m_t}\mathbf{v}_t. \tag{22b}$$

Different from iteratively seeking solutions of (11) based on ADMM iterations, the minimization of (21) admits a simple closed-form minimizer presented in the next result, which is proved in the Appendix.

Proposition 3. *Given $\mathbf{a} \in \mathbb{C}^N$ and $c \in \mathbb{R}$, the solution of*

$$\underset{\mathbf{w}\in\mathbb{C}^N}{\text{minimize}} \quad \left|\mathbf{a}^{\mathcal{H}}\mathbf{w} - c\right| + \frac{1}{2\tau}\|\mathbf{w}\|_2^2 \tag{23}$$

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

Based on Proposition 3, the solution of (21) is given by

$$\mathbf{v}_{t+1} := \mathbf{v}_t + \text{proj}_{\mu_t}(c_{m_t}/\|\mathbf{a}_{m_t}\|_2^2)\cdot\mathbf{a}_{m_t}. \tag{24}$$

Intuitively, measurements with a relatively small absolute residual, namely $|c_{m_t}| \leq \|\mathbf{a}_{m_t}\|_2^2$, are deemed 'nominal,' and $\mathbf{v}_t$ is updated with a step of $c_{m_t}/\|\mathbf{a}_{m_t}\|_2^2$ along the current direction of $\mathbf{a}_{m_t}$. On the other hand, the measurements of larger absolute residuals are likely to be outliers, so $\mathbf{v}_t$ is updated along its direction $\mathbf{a}_{m_t}$ by only a step of τ_t as opposed to $c_{m_t}/\|\mathbf{a}_{m_t}\|_2^2$.

The proposed stochastic prox-linear LAV solver is listed in Table II. For convergence, a diminishing stepsize sequence $\{\mu_t\}$ is required. Specifically, we consider stepsizes that are square summable but not summable; that is,

$$\sum_{t=0}^{\infty}\mu_t = \infty, \quad \text{and} \quad \sum_{t=0}^{\infty}\mu_t^2 < \infty. \tag{25}$$

TABLE II: Stochastic LAV Solver

1: Input data $\{(z_m, \mathbf{H}_m)\}_{m=1}^M$, stepsize $\{\mu_t > 0\}$, and initialization $\mathbf{v}_0 = \mathbf{1}$.
2: **For** $t = 0, 1, \ldots$ **do**
3: Draw m_t uniformly at random from $\{1, 2, \ldots, M\}$, or, by cycling through $\{1, 2, \ldots, M\}$.
4: Evaluate $\mathbf{a}_{m_t}$ and c_{m_t} as in (22).
5: Update $\mathbf{v}_{t+1}$ using (24).
6: **End for**

For instance, one can choose $\mu_t = \alpha t^{-\beta}$ with appropriately selected constants $\alpha > 0$ and $\beta \in (0.5, 1]$. Then the sequence $\{\mathbf{v}_t\}$ converges to a stationary point of the LAV objective in (7) almost surely [24, Thm. 1].

In terms of computational complexity, it can be verified that each $\mathbf{H}_m$ matrix corresponding to a square voltage magnitude or (re)active power flow measurement [cf. (1) and (5)] has exactly one or three nonzero entries, respectively. As such, if the available measurements include only these two types of measurements, evaluating $(\mathbf{a}_{m_t}, c_{m_t})$ as well as updating $\mathbf{v}_t$ per stochastic iteration requires just a small number (≤ 10) of scalar multiplications and additions, therefore incurring per-iteration complexity of $\mathcal{O}(1)$, which is *independent of the network size* N. This complexity evidently scales favorably to very large interconnected power networks. It is also worth highlighting that only one or two entries of $\mathbf{v}_t$ are updated depending on whether a voltage or power flow measurement is processed at each iteration. On the other hand, even if the power injections are measured too, the number of scalar operations per iteration increases to the order of the number of neighboring buses, which still remains much smaller than N in most real-world networks.

Remark 1. PMU measurements can be easily accounted for in (7). The developed prox-linear LAV schemes apply without any algorithmic modification.

Remark 2. An auxiliary sparse variable $\mathbf{o} \in \mathbb{R}^M$ has been introduced to model explicitly the outliers in [31]. These robust schemes with WLS estimation regularized by $\|\mathbf{o}\|_1$ can also benefit from our efficient and scalable algorithms.

B. Accelerated implementation using mini-batches

Although it involves only simple closed-form updates, the stochastic solver in Table II may require a large number of iterations to converge for high-dimensional power networks. Stochastic approaches based on mini-batches of measurements have been recently popular in large-scale machine learning tasks, because they offer a means of accelerating the stochastic algorithms. However, the naive way of designing mini-batches by grouping measurements randomly would yield a sequence of quadratic programs as in (11) of the deterministic LAV solver, which does not have closed-form solutions due to the ℓ_1 term. The novelty here is to fully exploit the sparsity of $\mathbf{H}_m$ matrices to group measurements into mini-batches in a way that closed-form solutions of the resulting quadratic programs are possible.

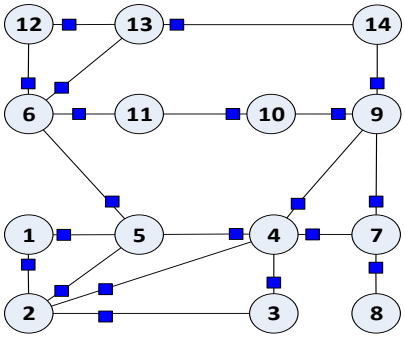

Fig. 1: The IEEE 14-bus benchmark configuration.

Groups	Power flow measurements (Bus index)
1	$(1,2)$, $(3,4)$, $(5,6)$, $(7,9)$
2	$(1,5)$, $(2,4)$, $(6,11)$, $(12,13)$
3	$(2,3)$, $(4,5)$, $(6,12)$, $(9,14)$
4	$(2,5)$, $(4,7)$, $(6,13)$, $(10,11)$
5	$(4,9)$, $(7,8)$, $(9,10)$, $(13,14)$

TABLE III: Mini-batching power flow measurements.

Suppose that active and reactive power flows over all lines and the square voltage magnitude at all buses are measured. Since $\mathbf{H}_n^V$ in (1) has exactly one nonzero entry at the (n,n)-th position, the corresponding updating vector $\mathbf{a}_n$ in (22a) has (at most) one nonzero entry at the n-th position. Updating $\mathbf{v}_t$ using (24) modifies the n-th position only. Hence, refining $\mathbf{v}$ using the N voltage measurements sequentially in N stochastic iterations is equivalent to updating $\mathbf{v}$ using all N measurements simultaneously in a single iteration. For power flows, every $\mathbf{H}_m$ has three nonzero entries in two rows. For instance, $\mathbf{H}^P_{nn'}$ has nonzero entries indexed by (n,n'), (n,n), and (n',n); and so does $\mathbf{H}^Q_{nn'}$. Processing the active or reactive power flow measurement over line (n,n') amounts to updating the n-th and n'-th entries of $\mathbf{v}_t$. Hence, so long as each of a mini-batch of measurements does not share common indices with the remaining ones, processing a mini-batch of such measurements one after the other boils down to processing all measurements simultaneously in one iteration.

For illustration, consider the IEEE 14-bus test system depicted in Fig. 1 [32]. Consider a total of 54 measurements, which include 14 square voltage magnitudes, as well as 20 'from-end' active and reactive power flows each. All voltage magnitudes can be grouped as a single mini-batch, or into several mini-batches by any means. One way of mini-batching each type of power flow measurements is suggested in Table III, where 20 active (reactive) power flows yield 5 mini-batches of equal size. It can be easily verified that any two measurements within a group (mini-batch) are measured over two lines of non-overlapping indices.

Let the entire measurements be divided into B mini-batches denoted by $\{\mathcal{B}_b\}_{b=1}^B \subseteq \{1,2,\ldots,M\}$. If a mini-batch $\mathcal{B}_{b_t}$ of measurements is drawn uniformly at random from $\{\mathcal{B}_b\}_{b=1}^B$ at iteration t, the accelerated implementation by means of mini-batching, updates the state estimate according to [cf. (24)]

$$\mathbf{v}_{t+1} := \mathbf{v}_t + \sum_{m\in\mathcal{B}_{b_t}} \mathrm{proj}_{\mu_t}(c_m/\|\mathbf{a}_m\|_2^2)\cdot\mathbf{a}_m \qquad (26)$$

which is in sharp contrast to that of using ADMM iterations to deal with the quadratic subproblems (10) in the deterministic LAV solver.

V. Numerical Tests

The proposed linear proximal LAV solvers were numerically tested on an Intel CPU @ 3.4 GHz (32 GB RAM) computer using MATLAB R2016a. Three power network benchmarks including the IEEE 14-, 118-bus, and the PEGASE 9,241-bus systems were simulated, following the MATLAB-based toolbox MATPOWER [32], [33]. The 'workhorse' Gauss-Newton iterations for the WLS-based PSSE were adopted as a baseline, and were implemented using the SE function 'doSE.m' in MATPOWER. Moreover, when all squared voltage magnitudes are measured, the initial point is taken to be the voltage magnitude vector, unless otherwise specified. In order to fix the phase ambiguity, the phase generated at the reference bus was set to 0 in all tests. For numerical stability, and to eliminate the effect of leverage measurements [19], the developed solvers were implemented using the normalized data, namely $\{(\frac{z_m}{\|\mathbf{H}_m\|_2}, \frac{\mathbf{H}_m}{\|\mathbf{H}_m\|_2})\}_{m=1}^M$.

A. Noiseless case

The first experiment simulates the noiseless data to evaluate the convergence and runtime of the novel algorithms relative to the WLS-based Gauss-Newton iterations on the IEEE 14-bus test system. The default voltage profile was employed. Measurements including all active and reactive power flows, as well as all squared voltage magnitudes were obtained from MATPOWER [33]. The deterministic solver in Table I was implemented with 200 ADMM iterations and stepsizes $\mu = \rho = 0.35$. The stochastic algorithm in Table II used the diminishing stepsize $1/t^{0.8}$. The accelerated scheme in (26) was implemented with stepsize 0.8 using a total of 11 mini-batches: 1 for all voltage magnitudes, and 5 of equal size for active and reactive power flows, each grouped as in Table III. The WLS objective $(1/M)\sum_{m=1}^M(z_m - \mathbf{v}^{\mathcal{H}}\mathbf{H}_m\mathbf{v})^2$ was evaluated at every Gauss-Newton iteration, while the LAV objective in (7) was computed per iteration of the deterministic, and every M iterations of the stochastic and accelerated schemes. Figure 2 plots the objective values for the Gauss-Newton, deterministic, stochastic, and accelerated schemes, whose corresponding runtime is 0.034s, 0.148s, 0.053s, and 0.042s. The Gauss-Newton stopped after a few iterations, but at a near-optimal point of relative root mean-square error (RMSE) $\|\hat{\mathbf{v}} - \mathbf{v}\|_2/\|\mathbf{v}\|_2 = 1.6\times 10^{-3}$, where $\mathbf{v}$ denotes the true voltage profile, and $\hat{\mathbf{v}}$ the estimate. The accelerated implementation was faster than the deterministic and stochastic LAV solvers, and yielded an accurate solution

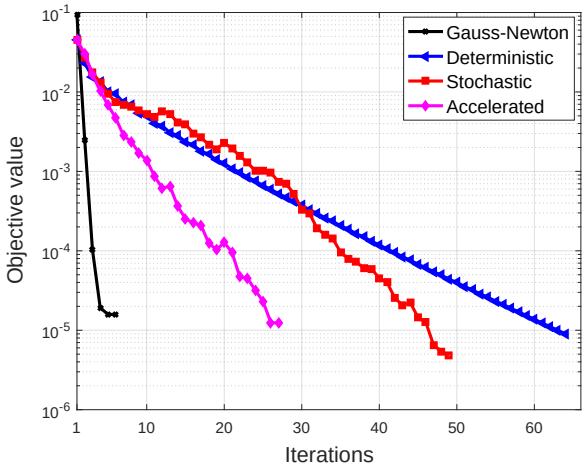

Fig. 2: Convergence performance for the IEEE 14-bus system.

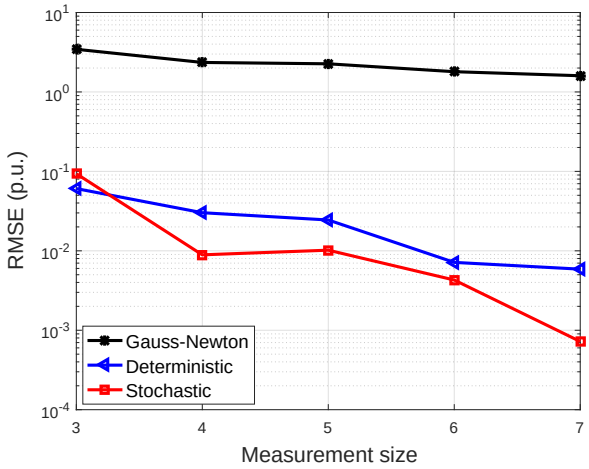

Fig. 3: Robustness to outliers for the IEEE 118-bus system.

with RMSE $= 4.88 \times 10^{-5}$ in time comparable to the Gauss-Newton iterations.

B. Presence of outliers

The second experiment was set to assess the robustness of the novel deterministic and stochastic solvers to measurements with outliers using the IEEE 118-bus benchmark network [32]. The WLS-based Gauss-Newton iterations were simulated as a baseline. The actual voltage magnitude (in p.u.) and angle (in radians) of each bus were uniformly distributed over $[0.95, 1.05]$, and over $[-0.05\pi, 0.05\pi]$. To assess the PSSE performance versus the measurement size, an additional type of measurements was included in a deterministic manner, as described next. All seven types of SCADA measurements were first enumerated as $\{|V_n|^2, P^f_{nk}, Q^f_{nk}, P_n, Q_n, P^t_{nk}, Q^t_{nk}\}$. Each x-axis value in Fig. 3 signifies the number of ordered types of measurements used in the experiment to yield the corresponding RMSEs, which are obtained by averaging over 100 independent realizations. For example, 5 implies that the first 5 types of data (i.e., $|V_n|^2, P^f_{nk}, Q^f_{nk}, P_n, Q_n$ over all buses and lines) were measured. Additive noise was independently generated from Gaussian distributions having zero-mean and standard deviation 0.004, 0.008, and 0.01 p.u. for the voltage magnitude, line flow, and power injection measurements, respectively [17]. Five percent of the measured data were corrupted according to model M1, which were chosen uniformly at random from line flows and bus injections. The outlying data $\{\xi_m\}$ were drawn independently from a Gaussian distribution with zero-mean and variance 5. The subproblems (10) with $\mu = 0.1$ of the deterministic scheme were solved with CVX [28], while the stochastic one was run with a diminishing stepsize $\mu_t = 10/t^{0.8}$. It is evident from Fig. 3 that our novel prox-linear LAV schemes are resilient to outlying measurements under corruption model M1.

The last experiment tests the scalability and efficacy of the stochastic iterations on a larger power network of $9,241$ buses available in MATPOWER [33]. The true voltage magnitude of each bus was uniformly distributed over $[0.9, 1.1]$, and its angle over $[-0.1\pi, 0.1\pi]$. All seven types of SCADA data were measured with additive noise described in the last experiment, 5% of which were compromised under model M2. The corrupted data $\xi_m := \tilde{\mathbf{v}}^{\mathcal{H}} \mathbf{H}_m \tilde{\mathbf{v}}$ relying on the individual $\mathbf{H}_m$ were generated using $\tilde{\mathbf{v}} \in \mathbb{R}^N$ from the standardized multivariate Gaussian distribution. In words, a number of $18,481$ variables were estimated, a total of $91,919$ measurements were obtained, $4,595$ of which were purposefully manipulated by adversaries. Initialized with the flat voltage profile point, the WLS-based Gauss-Newton iterations yielded an estimate of RMSE 0.9656, while the stochastic scheme with stepsize $\mu_t = 100/t^{0.8}$ attained an RMSE of 0.0163. Moreover, the novel stochastic LAV scheme is at least three times faster than the Gauss-Newton iterations.

VI. Conclusions

Robust power system state estimation was pursued using contemporary tools of composite optimization. Building upon recent algorithmic advances, two novel solvers were put forward to efficiently handle the LAV-based PSSE. Specifically, a deterministic scheme was developed based on a linear proximal method, yielding a sequence of convex quadratic subproblems that can be efficiently solved using off-the-shelf solvers, or, through ADMM iterations. Inspired by the sparse connectivity inherent to power networks, a highly scalable stochastic scheme that can afford simple closed-form updates was also devised. When only line flows and voltage magnitudes are measured, each stochastic iteration performs merely a few (less than 10) scalar operations, incurring per-iteration computational runtime of $\mathcal{O}(1)$, regardless of the number of buses in the network. If the power injections are included as well, this complexity goes down to the order of the number of neighboring buses, which still remains much smaller than the network size in general. A mini-batching technique was also suggested to further accelerate the stochastic iterations by judiciously leveraging the sparsity structure of the measurement matrices. Numerical tests on a variety of benchmark networks showcase the robustness and computational efficiency of the novel approaches.

Devising decentralized and parallel implementations of the novel approaches constitutes interesting future research di-

rections. Coping with the Y- and Δ-connection, as well as extending the approaches to multiphase unbalanced distribution systems are practically relevant future research topics too. Another possibility includes leveraging recent advances in solving systems of quadratic equations to tackle PSSE [34].

APPENDIX

Proof of Prop. 2: Letting $\boldsymbol{\chi}$ denote the dual variable associated with the constraint $\mathbf{u} = \mathbf{A}\mathbf{w}$, the KKT optimality conditions are given by [29]

$$\begin{aligned} \mathbf{w}^* - \mathbf{b} + \mathbf{A}^{\mathcal{H}}\boldsymbol{\chi}^* &= \mathbf{0} \\ \mathbf{u}^* - \mathbf{d} - \boldsymbol{\chi}^* &= \mathbf{0} \\ \mathbf{A}\mathbf{w}^* - \mathbf{u}^* &= \mathbf{0} \end{aligned}$$

or in the following compact form

$$\begin{bmatrix} \mathbf{I}_N & \mathbf{0} & \mathbf{A}^{\mathcal{H}} \\ \mathbf{0} & \mathbf{I}_M & -\mathbf{I}_M \\ \mathbf{A} & -\mathbf{I}_M & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{w}^* \\ \mathbf{u}^* \\ \boldsymbol{\chi}^* \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ \mathbf{d} \\ \mathbf{0} \end{bmatrix}.$$

Eliminating the dual variable via $\boldsymbol{\chi}^* = \mathbf{u}^* - \mathbf{d}$ from the KKT system, yields

$$\begin{bmatrix} \mathbf{I}_N & \mathbf{A}^{\mathcal{H}} \\ \mathbf{A} & -\mathbf{I}_M \end{bmatrix} \begin{bmatrix} \mathbf{w}^* \\ \mathbf{u}^* \end{bmatrix} = \begin{bmatrix} \mathbf{b} + \mathbf{A}^{\mathcal{H}}\mathbf{d} \\ \mathbf{0} \end{bmatrix}. \tag{27}$$

By further eliminating $\mathbf{d}^*$ and solving for $\mathbf{b}^*$, the solution to (27) and also to the minimization in (18) can be found in two steps as

$$\begin{aligned} \mathbf{w}^* &:= \left(\mathbf{I} + \mathbf{A}^{\mathcal{H}}\mathbf{A}\right)^{-1}\left(\mathbf{b} + \mathbf{A}^{\mathcal{H}}\mathbf{d}\right) \\ \mathbf{u}^* &:= \mathbf{A}\mathbf{w}^* \end{aligned}$$

which completes the proof of the claim. ■

Proof of Prop. 3: For any constant $c \in \mathbb{R}$, it is easy to verify that $\mathbf{a}^{\mathcal{H}}\mathbf{w}$ at the optimum $\mathbf{w}^*$ takes real values by contradiction. The optimality condition for (23) is

$$\mathbf{0} \in \partial\big(\big|\mathbf{a}^{\mathcal{H}}\mathbf{w} - c\big|\big) + \frac{1}{\tau}\mathbf{w} \Longleftrightarrow \mathbf{0} \in \partial\big|\mathbf{a}^{\mathcal{H}}\mathbf{w} - c\big| \cdot \mathbf{a} + \frac{1}{\tau}\mathbf{w}$$

or equivalently, $\mathbf{0} \in \partial\big|\mathbf{a}^{\mathcal{H}}\mathbf{w} - c\big| \cdot (\tau\mathbf{a}) + \mathbf{w}$, where ∂ denotes the subdifferential. Let us first examine the case where $\mathbf{a}^{\mathcal{H}}\mathbf{w} - c \neq 0$. We thus have $\partial|\mathbf{a}^{\mathcal{H}}\mathbf{w} - c| = \mathrm{sign}(\mathbf{a}^{\mathcal{H}}\mathbf{w} - c)$, which yields the optimum

$$\mathbf{w}^* = -\tau\mathrm{sign}(\mathbf{a}^{\mathcal{H}}\mathbf{w} - c) \cdot \mathbf{a}.$$

Note that if $c/\|\mathbf{a}\|_2^2 \geq \tau$, or $\mathbf{a}^H\mathbf{w}^* - c = -\tau\|\mathbf{a}\|_2^2\,\mathrm{sign}(\mathbf{a}^{\mathcal{H}}\mathbf{w}^* - c) - c < 0$, then $\mathbf{w}^* = \tau\mathbf{a}$. Equivalently, if $c/\|\mathbf{a}\|_2^2 \leq -\tau$, or $\mathbf{a}^{\mathcal{H}}\mathbf{w}^* - c = -\tau\|\mathbf{a}\|_2^2\,\mathrm{sign}(\mathbf{a}^{\mathcal{H}}\mathbf{w}^* - c) - c > 0$, then $\mathbf{w}^* = -\tau\mathbf{a}$.

If $\mathbf{a}^{\mathcal{H}}\mathbf{w} - c = 0$, the subdifferential of the absolute operator is the interval $[-1, 1]$; hence, the optimality condition becomes

$$\mathbf{0} \in -[-1, 1] \cdot (\tau\mathbf{a}) + \mathbf{w} \Longleftrightarrow \mathbf{w} \in [-\tau, \tau]\mathbf{a}.$$

Upon letting $\mathrm{proj}_\tau(x)$ denote the projection of a real number x onto the interval $[-\tau, \tau]$, one can combine the aforementioned three cases, and write collectively the optimum as

$$\mathbf{w}^* := \mathrm{proj}_\tau(c/\|\mathbf{a}\|_2^2) \cdot \mathbf{a}$$

which concludes the proof of the proposition.

■

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

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

[5] P. Fairley, "Cybersecurity at US utilities due for an upgrade: Tech to detect intrusions into industrial control systems will be mandatory," *IEEE Spectr.*, vol. 53, no. 5, pp. 11–13, May 2016.

[6] D. U. Case, "Analysis of the cyber attack on the Ukrainian power grid," 2016.

[7] S. Zonouz, K. M. Rogers, R. Berthier, R. B. Bobba, W. H. Sanders, and T. J. Overbye, "SCPSE: Security-oriented cyber-physical state estimation for power grid critical infrastructures," *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 1790–1799, Dec. 2012.

[8] E. Caro and A. Conejo, "State estimation via mathematical programming: A comparison of different estimation algorithms," *IEE Proc.-Gener. Transm. Distrib.*, vol. 6, no. 6, pp. 545–553, Jun. 2012.

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

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

[11] 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.

[12] 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.

[13] R. Madani, M. Ashraphijuo, J. Lavaei, and R. Baldick, "Power system state estimation with a limited number of measurements," Preprint, 2016.

[14] 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).

[15] L. Mili, M. G. Cheniae, and P. J. Rousseeuw, "Robust state estimation of electric power systems," *IEEE Trans. Circuits Syst. I*, vol. 41, no. 5, pp. 349–358, May 1994.

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

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

[18] R. Jabr and B. Pal, "Iteratively re-weighted least-absolute-value method for state estimation," *IEE Proc.-Gener. Transm. Distrib.*, vol. 150, no. 4, pp. 385–391, July 2003.

[19] M. Gol and A. Abur, "LAV based robust state estimation for systems measured by PMUs," *IEEE Trans. Smart Grid*, vol. 5, no. 4, pp. 1808–1814, Jul. 2014.

[20] C. Xu and A. Abur, "A fast and robust linear state estimator for very large scale interconnected power grids," *IEEE Trans. Smart Grid*, vol. PP, no. 99, pp. 1–1, 2017.

[21] J. V. Burke, "Descent methods for composite nondifferentiable optimization problems," *Math. Program.*, vol. 33, no. 3, pp. 260–279, Dec. 1985.

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

[23] D. Drusvyatskiy and A. S. Lewis, "Error bounds, quadratic growth, and linear convergence of proximal methods," *arXiv:1602.06661*, 2016.

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

[25] ——, "Solving (most) of a set of quadratic equalities: Composite optimization for robust phase retrieval," *arXiv:1705.02356*, 2017.

[26] P. J. Huber, "Robust Statistics," in *International Encyclopedia of Statistical Science*. Springer, 2011, pp. 1248–1251.

[27] G. Wang, A. S. Zamzam, G. B. Giannakis, and N. D. Sidiropoulos, "Power system state estimation via feasible point pursuit: Algorithms and Crmér-Rao bound," *arXiv:1705.04031*, 2017.

[28] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming," 2008. [Online]. Available: http://cvxr.com/cvx.

[29] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, pp. 1–122, 2010.

[30] A. Maleki, L. Anitori, Z. Yang, and R. G. Baraniuk, "Asymptotic analysis of complex LASSO via complex approximate message passing (CAMP)," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4290–4308, July 2013.

[31] V. Kekatos and G. B. Giannakis, "Distributed robust power system state estimation," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1617–1626, May 2013.

[32] Power systems test case archive. Univ. of Washington. [Online]. Available: http://www.ee.washington.edu/research/pstca.

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

[34] 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.