Fast LAV Estimation via Composite Optimization

Gang Wang and Gerogios B. Giannakis
Department of ECE and Digital Technology Center
University of Minnesota
Minneapolis, MN 55455, USA
E-mails: {gangwang, georgios}@umn.edu

Jie Chen
Key Lab of Intell. Control and Decision of Complex Systems
School of Automation, Beijing Institute of Technology
Beijing 100081, China
E-mail: chenjie@bit.edu.cn

***Abstract*—Accurate and robust power system state estimation (PSSE) is an essential prerequisite for reliable operation of smart power grids. In contrast to the commonly employed weighted least squares (WLS) one, the least-absolute-value (LAV) estimator is well documented for its robustness. Due to the non-convexity and non-smoothness however, existing LAV implementations are typically slow, thus inadequate for real-time system monitoring. In this context, this paper puts forward a novel LAV estimator leveraging recent algorithmic advances in composite optimization. Concretely, the estimator is based on a proximal linear procedure that deals with a sequence of convex quadratic problems, each efficiently solvable by means of either standard convex optimization methods, or the alternating direction method of multipliers. Simulated tests using two IEEE benchmark networks showcase its improved robustness and computational efficiency relative to several competing alternatives.**

I. Introduction

Accurately monitoring the power grid's operating condition is central to a number of control and optimization tasks, including optimal power flow, reliability and contingency analysis, and network expansion planning [1]. Current power grids are primarily monitored by supervisory control and data acquisition (SCADA) systems. Parameter uncertainty, instrument mis-calibration, delays, and unreported topology switches can yield grossly corrupted SCADA measurements (a.k.a. bad data) [2]. Designed for functionality and efficiency yet with little attention paid to security, today's SCADA systems are vulnerable to cyberattacks [3]. Bad data can 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 [4]. Any of these events can occur which will cause a given data collection to be much more inaccurate than is assumed by our mathematical models. In the smart grid context, robust PSSE methods against cyber threats are thus well motivated.

Commonly adopted PSSE criteria consist of the weighted least-squares (WLS) and the least-absolute value (LAV) [1]. However, WLS estimators are sensitive to bad data [2], which may yield poor estimates in the presence of outliers. This issue was to some extent mitigated by incorporating the largest normalized residual (LNR) test for bad data removal [2], or, by reformulating WLS into a semidefinite program via convex relaxation [5], [6]. The least-median-squares and the least-trimmed-squares estimators have provably improved performance under certain conditions [7]. Unfortunately, their computational and storage requirements scale unfavorably with the number of buses in the network [8].

On the other hand, LAV estimators simultaneously identify and reject bad data while acquiring an accurate estimate of the state [9]. Recent efforts have focused on coping with the non-convexity and non-smoothness in LAV estimation. Upon linearizing the nonlinear measurement functions at the current iterate, a series of linear programs was solved in [9]. Suggestions for improving the linear programming by exploiting the system's structure [10], or by iterative reweighting [11] have also been reported. Despite these efforts, LAV estimators have not been widely employed yet in today's power networks due mostly to their computational inefficiency.

LAV-based PSSE is revisited in this paper from the viewpoint of composite optimization [12], 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. A novel LAV estimator is developed based on a proximal linear (prox-linear) procedure. Concretely, it involves minimizing a sequence of convex quadratic subproblems, constructed from a linearized approximation to f and a quadratic regularization. Each subproblem can be efficiently solved via either off-the-shelf solvers, or, the alternating direction method of multipliers (ADMM) as we will elaborate on.

Notation. Matrices (column vectors) are denoted by upper- (lower-) case boldface letters; e.g., $\mathbf{I}$ ($\mathbf{1}$) are identity matrices (all-one vectors) of suitable dimensions. Symbol $(\cdot)^{\mathcal{T}}$ ($(\cdot)^{\mathcal{H}}$) represents (Hermitian) transpose, while $\Re(\cdot)$ ($\Im(\cdot)$) takes the real (imaginary) part of a complex number.

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 $\mathcal{E} := \{(n, n')\} \subseteq \mathcal{N} \times \mathcal{N}$ correspond 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 written 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

The work of G. Wang and G. B. Giannakis was supported by NSF grants 1508993, 1509040, and 1711471. The work of J. Chen was supported by the National Natural Science Foundation of China grants U1509215, 61621063, and the Program for Changjiang Scholars and Innovative Research Team in University (IRT1208).

bus admittance matrix $\mathbf{Y} = \mathbf{G} + j\mathbf{B} \in \mathbb{C}^N$ [1]. In rectangular coordinates, the active and reactive power injections p_n and q_n per bus n can be compactly given by

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

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

Recognize that the line current from bus n to n' at the sending end 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 sending end. Ohm's and Kirchhoff's laws assert that the sending end power flow over line (n, n') can be expressed as $\overline{S}_{nn'}^f = P_{nn'}^f - jQ_{nn'}^f = \overline{v_n}i_{nn'}^f = (\mathbf{v}^{\mathcal{H}}\mathbf{e}_n)(\mathbf{e}_{nn'}^{\mathcal{T}}\mathbf{i}_f) = \mathbf{v}^{\mathcal{H}}\mathbf{e}_n\mathbf{e}_{nn'}^{\mathcal{T}}\mathbf{Y}_f\mathbf{v}$, yielding

$$P_{nn'}^f = \mathbf{v}^{\mathcal{H}}\mathbf{H}_{nn'}^P\mathbf{v}, \text{ with } \mathbf{H}_{nn'}^P := \frac{\mathbf{Y}_f^{\mathcal{H}}\mathbf{e}_{nn'}\mathbf{e}_n^{\mathcal{T}} + \mathbf{e}_n\mathbf{e}_{nn'}^{\mathcal{T}}\mathbf{Y}_f}{2} \tag{3a}$$

$$Q_{nn'}^f = \mathbf{v}^{\mathcal{H}}\mathbf{H}_{nn'}^Q\mathbf{v}, \text{ with } \mathbf{H}_{nn'}^Q := \frac{\mathbf{Y}_f^{\mathcal{H}}\mathbf{e}_{nn'}\mathbf{e}_n^{\mathcal{T}} - \mathbf{e}_n\mathbf{e}_{nn'}^{\mathcal{T}}\mathbf{Y}_f}{2j}. \tag{3b}$$

The active and reactive power flows measured at the receiving end $P_{nn'}^t$ 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 data including squared voltage magnitudes as well as active and reactive power injections and flows can be expressed as quadratic functions of $\mathbf{v} \in \mathbb{C}^N$. 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 meter locations of the corresponding type, we have available the following (possibly noisy or even corrupted) measurements: $\{\check{V}_n\}_{n\in\mathcal{S}_V}$, $\{\check{p}_n\}_{n\in\mathcal{S}_p}$, $\{\check{q}_n\}_{n\in\mathcal{S}_q}$, $\{\check{P}_{nn'}^f\}_{(n,n')\in\mathcal{S}_P^f}$, $\{\check{Q}_{nn'}^f\}_{(n,n')\in\mathcal{S}_Q^f}$, $\{\check{P}_{nn'}^t\}_{(n,n')\in\mathcal{S}_P^t}$, and $\{\check{Q}_{nn'}^t\}_{(n,n')\in\mathcal{S}_Q^t}$, henceforth concatenated in the vector $\mathbf{z} \in \mathbb{R}^M$, where M is the total number of measurements.

In this paper, the following data corruption model is considered [13]: If $\{\xi_i\} \subseteq \mathbb{R}$ models an arbitrary attack sequence, given the measurement matrices $\{\mathbf{H}_m\}_{m=1}^M$ in (1)-(3a), we observe for $1 \le m \le M$ the samples

$$z_m \approx \begin{cases} \mathbf{v}^{\mathcal{H}}\mathbf{H}_m\mathbf{v} & \text{if } m \in \mathcal{I}^{nom} \\ \xi_m & \text{if } m \in \mathcal{I}^{out} \end{cases} \tag{4}$$

where additive measurement noise can be included if $\approx$ is replaced by equality, and sets $\mathcal{I}^{nom}, \mathcal{I}^{out} \subseteq \{1, 2, \ldots, M\}$ 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. 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, we assume for the attacks that

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

It is worth highlighting that the model M1) requires full independence between the corruption and measurements. That is, the attacker may only corrupt ξ_m without knowing $\mathbf{H}_m$.

Having outlined the system and corruption models, PSSE can be stated as follows: Given $\mathbf{Y}$, $\mathbf{Y}_f$, and measurements collected in $\mathbf{z} \in \mathbb{R}^M$, recover $\mathbf{v} \in \mathbb{C}^N$. The first attempt may be seeking the WLS estimate, or the ML one when assuming independent additive Gaussian noise [14]. It is known however that the WLS criterion is sensitive to outliers, and may yield very poor estimates even if there are very few grossly corrupted measurements [2]. As is well documented in statistics and optimization, the ℓ_1-based losses yield median-based estimators [15], and they 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 [9]

$$\underset{\mathbf{v}\in\mathbb{C}^N}{\text{minimize}} \quad f(\mathbf{v}) := \frac{1}{M}\sum_{m=1}^{M}\left|\mathbf{v}^{\mathcal{H}}\mathbf{H}_m\mathbf{v} - z_m\right|. \tag{5}$$

Because of $\{\mathbf{v}^{\mathcal{H}}\mathbf{H}_m\mathbf{v}\}_{m=1}^M$ and the absolute-value operation, the LAV objective is non-smooth, non-convex, and not even locally convex near the optima $\pm\mathbf{v}^*$. This is clear from the real-valued scalar case $f(v) = |v^*v - 1|$, where $v \in \mathbb{R}$. A local analysis based on convexity and smoothness is impossible, and $f(\mathbf{v})$ is in general difficult to minimize. For this reason, Gauss-Newton is not applicable to minimize (5). Nevertheless, the criterion $f(\mathbf{v})$ possesses several unique structural properties, which we will exploit next to develop efficient algorithms.

Remark 1. For an N-bus power system, most existing PSSE approaches have relied on optimizing over $(2N - 1)$ real variables, which consist of either the polar or the rectangular components of the complex voltage phasors after excluding the angle or the imaginary part of the reference bus that is often set to 0. Nevertheless, when iterative algorithms are used, working directly with the N-dimensional complex voltage vector has in general lower complexity and computational burden than in the real case. This is due to the compact quadratic representations of all SCADA quantities in complex voltage phasors, namely the natural sparsity of quadratic measurement matrices in the unknown complex voltage phasor vector.

III. Prox-Linear LAV Estimator

In this section, we will develop an efficient solver for (5). Toward this objective, we start rewriting the function in (5) as

$$\underset{\mathbf{v}\in\mathbb{C}^N}{\text{minimize}} \quad f(\mathbf{v}) := h(c(\mathbf{v})) \tag{6}$$

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 [12]. Evidently, this general form subsumes (5) as a special case, for which we can take $h(\mathbf{u}) = (1/M)\|\mathbf{u}\|_1$ and $c(\mathbf{v}) = \left[\mathbf{v}^{\mathcal{H}}\mathbf{H}_m\mathbf{v} - z_m\right]_{1\le m\le M}$. This unique compositional structure lends itself well to a proximal linear algorithm, which can be viewed a variant of the Gauss-Newton iterations [12]. Specifically, define close to a given $\mathbf{v}$ the local "linearization" of f as

$$f_{\mathbf{v}}(\mathbf{w}) := h\big(c(\mathbf{v}) + \Re(\nabla^{\mathcal{H}}c(\mathbf{v})(\mathbf{w} - \mathbf{v}))\big) \tag{7}$$

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 [16, Appendix]. In contrast to the nonconvex

$f(\mathbf{v})$, function $f_{\mathbf{v}}(\mathbf{w})$ in (7) becomes convex, which is the key behind the prox-linear method. Starting with some $\mathbf{v}_0$, e.g. the flat-voltage profile point $\mathbf{1}$, construct the iteration

$$\mathbf{v}_{t+1} := \arg\min_{\mathbf{v}\in\mathbb{C}^N} \left\{ f_{\mathbf{v}_t}(\mathbf{v}) + \frac{1}{2\mu_t}\|\mathbf{v}-\mathbf{v}_t\|_2^2 \right\} \tag{8}$$

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

Evidently, the subproblem (8) to be solved at every iteration of the prox-linear algorithm is convex, and can be handled by standard convex programming methods. 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 [17].

When specifying f to be the LAV objective of (5), the minimization in (8) becomes

$$\mathbf{v}_{t+1} = \arg\min_{\mathbf{v}\in\mathbb{C}^N} \|\Re(\mathbf{A}_t(\mathbf{v}-\mathbf{v}_t)) - \mathbf{c}_t\|_1 + \frac{1}{2}\|\mathbf{v}-\mathbf{v}_t\|_2^2 \tag{9}$$

with coefficients given by

$$\mathbf{A}_t := \left[(2\mu_t/M)\mathbf{v}_t^{\mathcal{H}}\mathbf{H}_m\right]_{1\le m\le M} \tag{10a}$$

$$\mathbf{c}_t := \left[(\mu_t/M)(z_m - \mathbf{v}_t^{\mathcal{H}}\mathbf{H}_m\mathbf{v}_t)\right]_{1\le m\le M}. \tag{10b}$$

For brevity, let $\mathbf{w} := \mathbf{v}-\mathbf{v}_t$, and rewrite (9) equivalently as a constrained optimization problem

$$\underset{\mathbf{u}\in\mathbb{C}^M,\ \mathbf{w}\in\mathbb{C}^N}{\text{minimize}} \quad \|\Re(\mathbf{u}) - \mathbf{c}_t\|_1 + \frac{1}{2}\|\mathbf{w}\|_2^2 \tag{11a}$$

$$\text{subject to} \quad \mathbf{A}_t\mathbf{w} = \mathbf{u}. \tag{11b}$$

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 (11) into

$$\underset{\tilde{\mathbf{u}},\,\tilde{\mathbf{w}},\,\mathbf{u},\,\mathbf{w}}{\text{minimize}} \quad \|\Re(\tilde{\mathbf{u}}) - \mathbf{c}_t\|_1 + \frac{1}{2}\|\tilde{\mathbf{w}}\|_2^2 \tag{12a}$$

$$\text{subject to} \quad \tilde{\mathbf{u}} = \mathbf{u},\ \tilde{\mathbf{w}} = \mathbf{w},\ \mathbf{A}_t\mathbf{w} = \mathbf{u}. \tag{12b}$$

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 (12b) can be expressed as

$$\mathcal{L}(\tilde{\mathbf{w}},\tilde{\mathbf{u}},\mathbf{w},\mathbf{u};\boldsymbol{\lambda},\boldsymbol{\nu}) := \|\Re(\tilde{\mathbf{u}}) - \mathbf{c}_t\|_1 + \frac{1}{2}\|\tilde{\mathbf{w}}\|_2^2 + \frac{\rho}{2}\|\tilde{\mathbf{w}}-\mathbf{w}\|_2^2$$
$$+ \Re(\boldsymbol{\lambda}^{\mathcal{H}}(\tilde{\mathbf{w}}-\mathbf{w})) + \Re(\boldsymbol{\nu}^{\mathcal{H}}(\tilde{\mathbf{u}}-\mathbf{u})) + \frac{\rho}{2}\|\tilde{\mathbf{u}}-\mathbf{u}\|_2^2 \tag{13}$$

where $\rho > 0$ is a predefined step size. With $k \in \mathbb{N}$ denoting the iteration index for solving (11), or equivalently (8), ADMM cycles through the following recursions

$$\tilde{\mathbf{w}}^{k+1} := \arg\min_{\tilde{\mathbf{w}}} \left\{ \frac{1}{2}\|\tilde{\mathbf{w}}\|_2^2 + \frac{\rho}{2}\left\|\tilde{\mathbf{w}} - (\mathbf{w}^k - \boldsymbol{\lambda}^k)\right\|_2^2 \right\} \tag{14a}$$

$$\tilde{\mathbf{u}}^{k+1} := \arg\min_{\tilde{\mathbf{u}}} \left\{ \frac{1}{2}\|\Re(\tilde{\mathbf{u}}) - \mathbf{c}_t\|_1 + \frac{\rho}{2}\left\|\tilde{\mathbf{u}} - (\mathbf{u}^k - \boldsymbol{\nu}^k)\right\|_2^2 \right\} \tag{14b}$$

$$\left\{\mathbf{w}^{k+1}, \mathbf{u}^{k+1}\right\} :=$$
$$\arg\min_{\mathbf{w},\,\mathbf{u}} \ \left\|\mathbf{w} - (\tilde{\mathbf{w}}^{k+1} + \boldsymbol{\lambda}^k)\right\|_2^2 + \left\|\mathbf{u} - (\tilde{\mathbf{u}}^{k+1} + \boldsymbol{\nu}^k)\right\|_2^2$$
$$\text{subject to } \mathbf{A}_t\mathbf{w} = \mathbf{u} \tag{14c}$$

$$\begin{bmatrix} \boldsymbol{\lambda}^{k+1} \\ \boldsymbol{\nu}^{k+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\lambda}^k + (\tilde{\mathbf{w}}^{k+1} - \mathbf{w}^{k+1}) \\ \boldsymbol{\nu}^k + (\tilde{\mathbf{u}}^{k+1} - \mathbf{u}^{k+1}) \end{bmatrix} \tag{14d}$$

where all the dual variables have been scaled by $\rho > 0$ [17].

Interestingly enough, the solutions of (14a)-(14c) can be provided in closed form, as we elaborate in the following two propositions, whose proofs will be provided in our full paper due to space limitation.

Proposition 1. *The solutions of* (14a)-(14b) *are respectively*

$$\tilde{\mathbf{w}}^{k+1} := \frac{\rho}{1+\rho}\left(\mathbf{w}^k - \boldsymbol{\lambda}^k\right) \tag{15a}$$

$$\tilde{\mathbf{u}}^{k+1} := \mathbf{c}_t + \mathcal{S}_{1/2\rho}\left(\Re(\mathbf{u}^k - \boldsymbol{\nu}^k) - \mathbf{c}_t\right) + i\Im(\mathbf{u}^k - \boldsymbol{\nu}^k) \tag{15b}$$

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

The constrained minimization of (14c) 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.

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

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

is given as

$$\mathbf{w}^* := \left(\mathbf{I} + \mathbf{A}^{\mathcal{H}}\mathbf{A}\right)^{-1}\left(\mathbf{b} + \mathbf{A}^{\mathcal{H}}\mathbf{d}\right) \tag{16a}$$

$$\mathbf{u}^* := \mathbf{A}\mathbf{w}^*. \tag{16b}$$

Using Proposition 2, the minimizer of (14c) is found as

$$\mathbf{w}^{k+1} := (\mathbf{I} + \mathbf{A}_t^{\mathcal{H}}\mathbf{A}_t)^{-1}\left[\tilde{\mathbf{w}}^{k+1} + \boldsymbol{\lambda}^k + \mathbf{A}^{\mathcal{H}}(\tilde{\mathbf{u}}^{k+1} + \boldsymbol{\nu}^k)\right] \tag{17a}$$

$$\mathbf{u}^{k+1} := \mathbf{A}\mathbf{w}^{k+1}. \tag{17b}$$

The four updates in (14) are computationally simple except for the matrix inversion of (17a), 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 (9) can be obtained as

$$\mathbf{v}_{t+1} := \mathbf{v}_t + \mathbf{w}^* \tag{18}$$

where $\mathbf{w}^*$ is the converged $\mathbf{w}$-iterate of the ADMM iterations in (15), (17), and (14d).

For implementation, our ADMM-based LAV estimator is summarized in Table I, in which the inner loop (i.e., Steps 3-8) can be replaced with standard convex programming methods to solve (9) for $\mathbf{v}_{t+1}$. As far as convergence is concerned, if h is L-Lipschitz and ∇c is β-Lipschitz, taking a constant stepsize $\mu \le 1/(L\beta)$ in (8) guarantees that [18, Sec. 5]:

i) the solver in Table I is a descent method for (9), and;

ii) the iterate sequence $\{\mathbf{v}_t\}$ converges to a stationary point of the LAV objective in (5).

TABLE I: Efficient 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 (10).
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 (15), (17), and (14d), respectively.
7: **End for**
8: Update $\mathbf{v}_{t+1}$ via (18).
9: **End for**

It is also worth remarking that local quadratic convergence can be attained when certain error bound conditions are satisfied [18, Thm. 5.5]. This is indeed observed in our experiments when noise and outliers are absent. The computational burden of the ADMM based deterministic solver is dominated by the projection step of (17), which incurs a per-iteration computational complexity of $\mathcal{O}(MN^2)$. This complexity can be afforded in small- or medium-size power networks.

IV. Numerical Tests

The proposed linear proximal LAV solver was numerically examined on an Intel CPU @ 3.4 GHz (32 GB RAM) computer using MATLAB R2016a. Two power network benchmarks, namely the IEEE 14-, and 118-bus systems were simulated, following the MATLAB-based toolbox MATPOWER [19].

The linear programming (LP) and the iteratively reweighted least-squares (IRLS) based LAV estimators [10], [20], along with the "workhorse" Gauss-Newton iterations for the WLS-based PSSE [1] were adopted as baselines. It is worth pointing out that the LP-based implementation can be regarded as a special case of the prox-linear algorithm with a constant step size of ∞. To see this, per iteration, the LP-based scheme [10] solves the minimization problem in (8) but *without* the regularization term $\frac{1}{2\mu_t}\|\boldsymbol{v}-\boldsymbol{v}_t\|_2^2$, or equivalently with $\mu_t = \infty$. Each linear program was formulated over $2N-1$ real variables consisting of the real and imaginary parts of the unknown voltage phasor vector, after excluding the imaginary part of the reference bus which was set 0. Per iteration, the resultant linear program was solved with SeDuMi [21]. Given that there is no parameter in the LP-based LAV estimator, although the time performance may vary if different toolboxes are used for solving the resultant linear programs, its convergence behavior in terms of the number of iterations is independent of the toolbox used. Furthermore, the Gauss-Newton iterations were implemented by calling for the embedded state estimation function 'doSE.m' in MATPOWER.

Regarding the initialization, when all squared voltage magnitudes are measured, the initial point is taken to be the voltage magnitude vector, unless otherwise specified. Each simulated scheme stops either when a maximum number 100 of iterations are reached, or when the normalized distance between two consecutive estimates becomes smaller than 10^{-10}, namely $\|\boldsymbol{v}_t - \boldsymbol{v}_{t-1}\|_2/\sqrt{N} \leq 10^{-10}$. In order to fix the phase ambiguity, the phase generated at the reference bus was set to 0 in all tests.

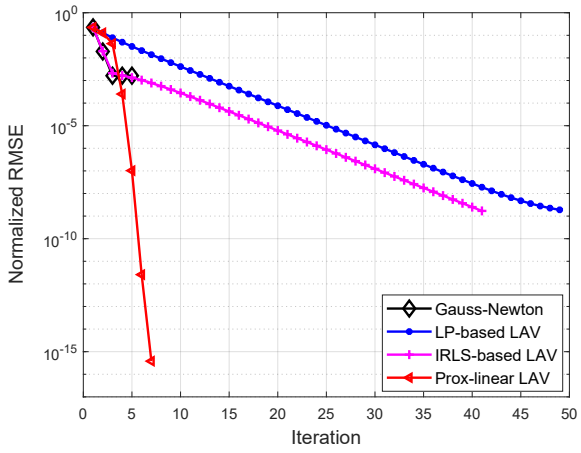

Fig. 1: Convergence performance for the IEEE 14-bus system.

A. Noiseless measurements

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, as well as the LP- and RLS-based LAV estimators 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 [19]. The ADMM-based prox-linear LAV estimator in Table I was implemented with $\mu = 200$, in which each quadratic subproblem was solved using a maximum of 150 ADMM iterations with $\rho = 100$. Our prox-linear estimator can also be implemented using standard convex programming methods to directly solve (8). It typically converges in very few (less than 10) iterations yet at a higher computational complexity. The normalized root mean-square error (RMSE) $\|\mathbf{v}_t - \mathbf{v}\|_2/\|\mathbf{v}\|_2$ was evaluated, where $\mathbf{v}$ is the true voltage profile, and $\mathbf{v}_t$ denotes the estimate obtained at the t-th iteration.

Figure 1 compares the normalized RMSE for the LP-based, IRLS-based, and prox-linear LAV estimators with that of Gauss-Newton iterations. Evidently, the novel scheme is the fastest in terms of the number of iterations, and converges to a point of machine accuracy (i.e., 10^{-16}) in 8 iterations. The IRLS is also fast, but it requires inverting a matrix per iteration. Even though the time of solving an LP may vary across toolboxes, convergence of the LP-based estimator in terms of the number of iterations will be the same. Evidently, solving an LP of $2M$ constraints and $2N-1$ real variables is indeed computationally more cumbersome. The Gauss-Newton method terminated after six iterations, but at a sub-optimal point of normalized RMSE 10^{-3} or so. The proposed prox-linear solver for LAV estimation requires much less iterations than the LP, and IRLS-based implementations.

B. Presence of outlying data

The second experiment was performed to assess the robustness of the novel LAV estimator to measurements with

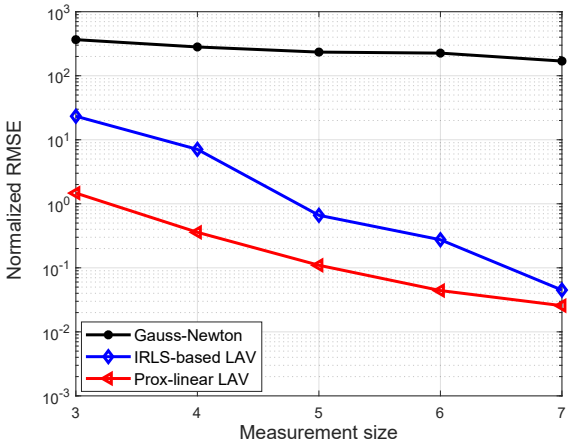

Fig. 2: Robustness to outliers for the IEEE 118-bus system.

outliers using the IEEE 118-bus benchmark network, while the IRLS-based LAV implementation and the WLS-based Gauss-Newton iterations were simulated as baselines. The actual voltage magnitude (in p.u.) and angle (in radians) of each bus were uniformly distributed over $[0.9,\ 1.1]$, and over $[-0.1\pi,\ 0.1\pi]$. 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. 2 signifies the number of ordered types of measurements used in the experiment to yield the corresponding normalized RMSEs, obtained by averaging over 100 independent runs. For instance, 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 used. Additive noise was independently generated from Gaussian distributions having zero-mean and standard deviations 0.004, 0.008, and 0.01 p.u. for the voltage magnitude, line flow, and power injection measurements, respectively [10]. Ten percent of the measured data were corrupted according to model M1, chosen randomly from line flows and bus injections. The outlying data $\{\xi_m\}$ were drawn independently from a Laplacian distribution with zero-mean and standard deviation 30. The subproblems (8) with $\mu = 100$ were solved using a maximum of 200 ADMM iterations with $\rho = 100$. It is evident from Fig. 2 that our prox-linear LAV estimator is resilient to outlying measurements under corruption model M1, yielding improved performance relative to the IRLS and WLS estimators. Furthermore, IRLS works well when the number of measurements grows large.

V. Conclusions

Robust power system state estimation was pursued using contemporary tools of composite optimization. Building on recent algorithmic advances, a novel proximal linear algorithm was put forward to efficiently deal with the non-convexity and non-smoothness of LAV-based PSSE. The algorithm relies on solving a sequence of convex quadratic subproblems, constructed from a linearized approximation to the original objective and a quadratic regularization. Numerical tests using two IEEE benchmark networks showcase the robustness and computational efficiency of the developed approach relative to competing alternatives for LAV and WLS estimation.

References

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

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

[3] 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.

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

[5] 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.

[6] 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.

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

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

[9] 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.

[10] 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.

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

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

[13] J. C. Duchi and F. Ruan, "Solving (most) of a set of quadratic equalities: Composite optimization for robust phase retrieval," *Information and Inference: A Journal of the IMA (to appear)*, 2018.

[14] 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.

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

[16] G. Wang, A. S. Zamzam, G. B. Giannakis, and N. D. Sidiropoulos, "Power system state estimation via feasible point pursuit: Algorithms and Cramér-Rao bound," *IEEE Trans. Signal Process.*, vol. 66, no. 6, pp. 1649–1658, Mar. 2018.

[17] 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.

[18] D. Drusvyatskiy and A. S. Lewis, "Error bounds, quadratic growth, and linear convergence of proximal methods," *Math. Oper. Res.*, vol. 43, no. 3, pp. 919–948, Aug. 2018.

[19] 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.

[20] R. Jabr and B. Pal, "Iteratively reweighted Least-Squares implementation of the WLAV state-estimation method," *IET Gener. Transm. Dis.*, vol. 151, no. 1, pp. 103–108, Feb. 2004.

[21] 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.