# Fast Nonconvex SDP Solvers for Large-scale Power System State Estimation

Yu Lan, *Student Member, IEEE*, Hao Zhu, *Senior Member, IEEE*, and Xiaohong Guan, *Fellow, IEEE*

***Abstract*—Fast power system state estimation (SE) solution is of paramount importance for achieving real-time decision making in power grid operations. Semidefinite programming (SDP) reformulation has been shown effective to obtain the global optimum for the nonlinear SE problem, while suffering from high computational complexity. Thus, we leverage the recent advances in nonconvex SDP approach that allows for the simple first-order gradient-descent (GD) updates. Using the power system model, we can verify that the SE objective function enjoys nice properties (strongly convex, smoothness) which in turn guarantee a linear convergence rate of the proposed GD-based SE method. To further accelerate the convergence speed, we consider the accelerated gradient descent (AGD) extension, as well as their robust versions under outlier data and a hybrid GD-based SE approach with additional synchrophasor measurements. Numerical tests on the IEEE 118-bus, 300-bus and the synthetic ACTIVSg2000-bus systems have demonstrated that FGD-SE and AGD-SE, can approach the near-optimal performance of the SDP-SE solution at significantly improved computational efficiency, especially so for AGD-SE.**

***Index Terms*—Power system state estimation, semidefinite programming, gradient descent, accelerated gradient descent, robust estimation.**

## I. Introduction

POWER system state estimation (SE) aims to obtain the operating condition of the grid, namely nodal complex voltages, from noisy measurements at buses and branches. The SE problem is of paramount importance for reliable control and economic operation of power systems; see e.g., [1], [2].

Due to a nonlinear measurement model, SE is traditionally formulated as a nonlinear least-squares (LS) problem and solved by Gauss-Newton (GN) iterations [3]. The GN method iteratively updates the variables by minimizing an approximate objective through linearization. Albeit computationally efficient per iteration, the convergence of GN iterations is not guaranteed. Recent work [4]–[6] and references therein have provided insightful understandings on GN divergence, such as heavy loading conditions or bad data including topology errors. To tackle this issue, a semidefinite programming (SDP) reformulation of the SE problem has been proposed in [7] through rank relaxation. The resultant convex problem achieves near-optimal guarantees to solve the original nonconvex SE. To promote lower-rank solutions, [8] suggested a nuclear norm based penalization. General penalization terms are designed in [9], [10] for recovery guarantees and quantifiable estimation error of the SDP-SE formulation. To account for bad data, robust SDP-SE has been developed by modeling outliers using a sparse vector [5], [11]. Although global optima of SDP-SE can be obtained by generic convex solvers such as the interior-point method, the high-order polynomial complexity therein could challenge real-time implementation in large-scale systems [11]. Recent work has focused on using conic relaxation or composite optimization techniques for the SDP-SE problem [9], [12]. A parallelizable SDP-SE solution was also developed in [7], [13] using graph-specific decomposition. However, it remains open on how to develop fast SDP-SE solvers with simple implementation steps.

Recently, a nonconvex approach to solving SDPs by representing the solution matrix using the Burer-Monteiro factorization [14], [15] has become popular. The factorization form can easily eliminate the positive semi-definite (PSD) constraint of SDP problems. This way, the first-order gradient descent (GD) updates are readily applicable to the resultant nonconvex objective, leading to the so-termed factored GD (FGD) algorithm [16]. Linear convergence guarantee for FGD is available if the original SDP objective is strongly convex and smooth with sufficiently close initialization. Albeit at linear convergence rate, FGD may suffer from slow convergence speed depending on the conditioning of objective function. To tackle this, the popular extension of accelerated GD (AGD) method [17] was considered for the nonconvex SDP reformulation in [18].

The goal of the present work is to accelerate large-scale SDP-based SE method by leveraging latest advances in fast SDP solution techniques. Our contribution is two-fold. First, we develop a class of fast GD-based solvers for the SDP-SE formulation that can achieve guaranteed linear convergence rate to a closely approximate SDP-SE solution. Specifically, we have successfully adopted both the FGD and AGD methods for the SDP-SE problem with excellent performance under regular SE conditions. Second, we further enhance the FGD-/AGD-SE solvers to account for practical concerns such as the presence of outliers or synchrophasor data. Our proposed robust GD-SE solutions incur a simple hard thresholding step to tackle outliers, while a hybrid GD-SE method is developed by extending the gradient direction to include synchrophasor measurements. All the GD-SE methods and extensions enjoy extremely efficient update per iteration, with linear convergence rate verified by our analysis and numerical tests.

Y. Lan is with Ministry of Education Key Lab for Intelligent Networks and Network Security, Xi'an Jiaotong University, Xi'an, Shaanxi, 710049, China. Y. Lan was a visiting student the Department of ECE, The University of Texas at Austin, Austin, TX 78705, USA. E-mail: ylan@sei.xjtu.edu.cn.

H. Zhu is with the Department of ECE, The University of Texas at Austin, Austin, TX 78705, USA. Email: haozhu@utexas.edu. H. Zhu's work was supported in part by the National Science Foundation under Award ECCS-1802319.

X. Guan is with Ministry of Education Key Lab for Intelligent Networks and Network Security, Xi'an Jiaotong University, Xi'an, Shaanxi, 710049, China. E-mail: xhguan@xjtu.edu.cn. Y. Lan and X. Guan's work was supported in part by Key R&D Project of China 2016YFB0901900.

Numerical results also confirm the excellent performance of the GD-SE methods in estimating the voltage phasors and identifying outliers.

The rest of the paper is organized as follows. The SDP-SE problem and the nonconvex reformulation are introduced in Section II. Section III presents the FGD-/AGD-SE methods along with the convergence analysis specifically for the power system models. Section IV develops the robust GD-SE methods and the PMU-aided extensions. Several numerical tests presented in Section V corroborate the faster computation time of FGD and AGD relative to the SDP-SE solver, and improved estimation performance over GN iterations. The paper is wrapped up in Section VI.

*Notation:* Upper (lower) boldface symbols stand for matrices (vectors); $|\cdot|$ stands for the magnitude; $(\cdot)^{\mathcal{T}}$ denotes transposition; $(\cdot)^{\mathcal{H}}$ complex-conjugate transposition; $\Re(\cdot)/\Im(\cdot)$ the real/imaginary part; $\mathrm{Tr}(\cdot)$ the matrix trace; $\mathrm{rank}(\cdot)$ the matrix rank; $\|\cdot\|_F$ the Frobenius norm; $\|\cdot\|_2$ the spectral norm, and $\mathbf{V}\succeq\mathbf{0}$ denotes $\mathbf{V}$ is a positive semi-definite (PSD) matrix.

## II. Problem Formulation

Consider a transmission network modeled as a graph $\mathcal{G}=(\mathcal{N},\mathcal{E})$, with the set of buses (nodes) in $\mathcal{N}:=\{1,...,N\}$ and the set of lines (edges) in $\mathcal{E}:=\{(n,n')\}$. The complex voltage phasor $V_n$ per bus $n\in\mathcal{N}$ can be expressed in the rectangular coordinate as $V_n=\Re(V_n)+\mathrm{j}\Im(V_n)$. All nodal voltages form the full system state vector $\mathbf{v}:=[V_1,...,V_N]^{\mathcal{T}}\in\mathbb{C}^N$. To estimate nodal voltages in $\mathbf{v}$, a subset of the following system variables are measured:

- $P_n(Q_n)$: the active (reactive) power injection at bus $n$;
- $P_{nn'}(Q_{nn'})$: the active (reactive) power line flow from bus $n$ to bus $n'$;
- $|V_n|$: the voltage magnitude at bus $n$.

The ac power flow model [1, Ch. 4] asserts power variables are nonlinearly (quadratically) related to the state $\mathbf{v}$.

Collecting the noisy measurements in the vector $\mathbf{z}:=\left[\{\check{P}_n\},\{\check{Q}_n\},\{\check{P}_{nn'}\},\{\check{Q}_{nn'}\},\{|\check{V}_n|^2\}\right]^T\in\mathbb{R}^L$, where $L$ is the total number of measurements, one can write the $\ell$-th measurement in $\mathbf{z}$ as

$$z_\ell=h_\ell(\mathbf{v})+\epsilon_\ell,\ \forall\ell=1,\ldots,L \tag{1}$$

where $h_\ell(\cdot)$ stands for the nonlinear transformation from $\mathbf{v}$, and $\epsilon_\ell$ accounts for the additive measurement error. Given this measurement model, the SE problem can be cast as a (weighted) LS minimization one over $\mathbf{v}$. Without loss of generality (Wlog), we assume the weight coefficients are included by the model (1). Thus, it suffices to consider the *unweighted* LS-SE objective throughout the paper. Due to its nonlinear objective, Gauss-Newton (GN) has been the workhorse solution for LS-SE; see e.g., [19]. GN iteratively approximates the objective by linearizing (1) at the latest solution. This iterative linearization procedure, though computationally efficient if convergent, can be potentially divergent under heavy loading or bad data conditions; see e.g., [4], [6]. The lack of statistical robustness was identified in [4] to cause divergence of GN in addition to numerical instability concerns.

One approach to tackle the nonlinearity is to introduce the outer-product matrix $\mathbf{V}:=\mathbf{v}\mathbf{v}^{\mathcal{H}}\in\mathbb{C}^{N\times N}$, consisting of all quadratic terms involving $\mathbf{v}$. This way, each measurement in (1) is linearly related to $\mathbf{V}$, as given by

$$z_\ell=\mathbf{v}^{\mathcal{H}}\mathbf{H}_\ell\mathbf{v}+\epsilon_\ell=\mathrm{Tr}(\mathbf{H}_\ell\mathbf{V})+\epsilon_\ell,\ \forall\ell=1,\ldots,L. \tag{2}$$

where $\mathbf{H}_\ell\in\mathbb{C}^{N\times N}$ is Hermitian matrix depending on the network topology and line parameters (see e.g., [7], [20] for the definitions). Reformulating the LS objective about $\mathbf{v}$ using (2) leads to the following semidefinite program (SDP) for $\mathbf{V}$:

$$\hat{\mathbf{V}}=\arg\min_{\mathbf{V}}\ f(\mathbf{V}):=\sum_{\ell=1}^{L}\frac{1}{2}\left[z_\ell-\mathrm{Tr}(\mathbf{H}_\ell\mathbf{V})\right]^2 \tag{3a}$$

$$\text{s.t. }\mathbf{V}\succeq\mathbf{0} \tag{3b}$$

where the PSD constraint in (3b) together with $\mathrm{rank}(\mathbf{V})=1$ can guarantee the existence of $\mathbf{v}$ that satisfies $\hat{\mathbf{V}}=\mathbf{v}\mathbf{v}^{\mathcal{H}}$. Due to the nonconvexity of rank constraint, it is dropped through a well-appreciated semidefinite relaxation (SDR) procedure that leads to the convex SDP-SE formulation (3). Solving the latter can achieve a near-optimal performance for the SE problem as the solution tends to have very low-rank; see [10], [11]. General convex solvers such as the popular interior-point method based solver SeDuMi [21] can obtain the optimal $\hat{\mathbf{V}}$ in polynomial time. Nonetheless, these solution methods can scale unfavorably as the number of buses or measurements increase, with worst-case complexity at $\mathcal{O}(N^{4.5})$ [11]. Thus, it is necessary to develop accelerated algorithms for solving large-scale SDP-SE in real-time.

Motivated by recent work on nonconvex SDP solvers [16], we consider an equivalent formulation of (3). The idea is to use matrix factorization to represent the PSD matrix as product of two factor matrices, as first proposed by [14], [15]. Accordingly, it can tackle the main computational challenge caused by the PSD conic constraint (3b). In general, expressing $\mathbf{V}=\mathbf{U}\mathbf{U}^{\mathcal{H}}$ using a rank$-r$ matrix $\mathbf{U}\in\mathbb{C}^{N\times r}$, one can reformulate (3) as an unconstrained one for the complex $\mathbf{U}$:

$$\hat{\mathbf{U}}=\arg\min_{\mathbf{U}\in\mathbb{C}^{N\times r}}\ g(\mathbf{U}):=f(\mathbf{U}\mathbf{U}^{\mathcal{H}}). \tag{4}$$

The rest of paper will consider the case of $r=1$; i.e., $\mathbf{U}$ becomes a vector $\mathbf{u}$. The advantage of using a rank-one $\mathbf{u}$ here is two-fold: (i) the relaxed SDP problem (3) is likely to attain a nearly rank-one solution; and (ii) searching for rank-one solutions has the lowest computational complexity. Note that the function $g(\mathbf{u})$ in (4) is equivalently the traditional (W)LS-SE objective, which is again nonconvex. Interestingly, based on the SDP formulation (3), one can show that the gradient descent (GD) updates for the reformulation (4) can achieve guaranteed convergence to a closely approximate solution to (3). Thus, we develop a class of GD-SE methods that enjoy both low computational time and guaranteed convergence.

## III. Gradient-descent Based SE Solvers

Thanks to the unconstrained structure and convenient gradient computation, we can use first-order methods such as gradient descent (GD) to solve (4). This simple approach, termed as factored GD (FGD) in [16], has been shown to

converge to the global optimum of general SDP problems, at a linear rate similar to that of classical GD method for strongly convex and smooth functions [22, Ch. 9]. Inspired by these results, we first use power flow analysis to prove the convergence rate of FGD for our LS-SE problem. Furthermore, we will propose an accelerated scheme of FGD for improved convergence rate.

### A. Factored Gradient Descent Method

To develop the FGD updates for (4), it suffices to compute the derivative $\nabla g(\mathbf{u})$ for any complex $\mathbf{u}$. Using the chain rule, one can write

$$\nabla g(\mathbf{u}) = 2\nabla f(\mathbf{u}\mathbf{u}^{\mathcal{H}}) \times \mathbf{u} = \sum_{\ell=1}^{L} 2\big[\mathbf{u}^{\mathcal{H}}\mathbf{H}_\ell\mathbf{u} - z_\ell\big]\mathbf{H}_\ell\mathbf{u}. \quad (5)$$

The rigorous derivation for a complex derivative is slightly more complicated, with the details in Appendix A. Given step-size $\eta$, the main FGD-SE steps are tabulated in Algorithm 1 for general rank-$r$ solutions. To select the $\eta$ value and establish the convergence results, one needs to investigate the characteristics of the original SDP function $f(\mathbf{V})$.

Specifically, consider a compact form $f(\mathbf{V}) = \frac{1}{2}\|\mathbf{z} - \mathcal{H}(\mathbf{V})\|_2^2$, where $\mathcal{H}(\cdot)$ stands for the linear mapping in the SDP objective (3). In general, strong convexity is needed for the loss function $f$. Nonetheless, it is challenging to show that over any PSD $\mathbf{V}$ of an arbitrary rank $r$. Instead, we can restrict it to a subset of $\mathbf{V}$'s [23]. Specifically, we consider the subset

$$\mathcal{V} := \{\mathbf{V} | \text{rank}(\mathbf{V}) = 1 \text{ and } \underline{V}^2 \le \mathbf{V}_{nn} \le \bar{V}^2 \ \forall n\} \quad (6)$$

where $\underline{V}/\bar{V}$ are lower/upper bounds on bus voltage magnitude. These bounds (e.g., $\pm 5\%$ from unity) are easily available thanks to well-designed power system voltage control mechanism. Using the subset $\mathcal{V}$, we can show the following bounds as given in Appendix B.

**Proposition 1.** *For any $\mathbf{V} \in \mathcal{V}$ as given by* (6)*, its mapped output $\mathcal{H}(\mathbf{V})$ for the SDP-SE objective $f$ in* (3) *satisfies*

$$m \cdot \|\mathbf{V}\|_F^2 \le \|\mathcal{H}(\mathbf{V})\|_2^2 \le M \cdot \|\mathbf{V}\|_F^2 \quad (7)$$

*where $m$ and $M$ are positive coefficients. Hence, the objective $f$ is $m$-strongly convex and $M$-smooth over restricted set $\mathcal{V}$ with the condition number $\kappa = M/m$.*

Proposition 1 is shown by leveraging the power flow model embedded in the mapping $\mathcal{H}$. It is a key result to ensure the convergence of FGD for the objective $g$, even though the latter is not (strongly) convex itself. For example, heavy loading conditions would lead to larger angular separation across the transmission line. Our numerical experience suggests that this increased angular separation has resulted in decreased value of $m$, and thereby higher condition number $\kappa = M/m$. This observation coincides with that for GN-SE method, as heavy loading conditions make it more likely to diverge [4]. Sec. III-B will introduce a GD scheme that can better tackle ill-conditioned problems due to e.g., heavy loading.

As developed in [16], using the initial $\mathbf{V}_0 = \mathbf{u}_0\mathbf{u}_0^{\mathcal{H}}$ one can set the step-size according to

$$\eta = \frac{1}{c(M\|\mathbf{V}_0\|_2 + \|\nabla f(\mathbf{V}_0)\|_2)} \quad (8)$$

**Algorithm 1** FGD-SE

**Input:** Function $f$, rank $r$, maximum iteration number $K$.
**Output:** $\mathbf{u}$ and $\mathbf{V} = \mathbf{u}\mathbf{u}^{\mathcal{H}}$
1: Initialize $\mathbf{u}_0 \in \mathbb{C}^{n\times r}$ and set $\mathbf{V}_0 = \mathbf{u}_0\mathbf{u}_0^{\mathcal{H}}$.
2: Set the step-size $\eta$ as in (8).
3: **for** $k = 0$ to $K$ **do**
4: $\quad \mathbf{u}_{k+1} = \mathbf{u}_k - \eta\nabla g(\mathbf{u}_k)$
5: **end for**
6: **return** $\mathbf{u} = \mathbf{u}_K$ and $\mathbf{V} = \mathbf{u}_K\mathbf{u}_K^{\mathcal{H}}$

where the constant ratio $1/c$ was picked as $1/16$ in [16] but can be better tuned up according to specific problems. In practice, the smoothness parameter $M$ can be approximated by the ratio $\|\nabla f(\mathbf{V}_0) - \nabla f(\mathbf{V})\|_F / \|\mathbf{V}_0 - \mathbf{V}\|_F$ for any $\mathbf{V} \in \mathcal{V}$.

To establish the local convergence of Algorithm 1 for the SDP-SE problem, we need the following two assumptions regarding the approximation to the global optimal $\hat{\mathbf{u}}$ and $\hat{\mathbf{V}}^1 := \hat{\mathbf{u}}\hat{\mathbf{u}}^{\mathcal{H}}$, using small constant numbers $\rho_u$ and $\rho_v$.

**(as1)** The initial $\mathbf{u}_0$ satisfies $\text{Dist}(\mathbf{u}_0, \hat{\mathbf{u}}) \le \frac{\rho_u}{\kappa}\|\hat{\mathbf{u}}\|_2$, where Dist denotes the minimum Euclidean distance between the two complex vectors up to any rotational change.

**(as2)** The optimum $\hat{\mathbf{V}}$ satisfies $\|\hat{\mathbf{V}} - \hat{\mathbf{V}}^1\|_F \le \frac{\rho_v}{\kappa^{1.5}}\|\hat{\mathbf{u}}\|_2^2$.

Assumption (as1) requires the initial guess to be sufficiently close to the optimal $\hat{\mathbf{u}}$. As the upper bound scales with $\|\hat{\mathbf{u}}\|_2$, it could be reasonable for good initialization such as flat start or the SE solution from dc power flow model. Assumption (as2) relates to the rank-one approximation to $\hat{\mathbf{V}}$. Since the SDP-SE solution $\hat{\mathbf{V}}$ is nearly rank-1 [10], [11], the upper bound therein could also be satisfied in practice.

**Proposition 2.** *Suppose (as1) and (as2) hold and use the step-size in* (8)*. Under Proposition 1 with the SE objective $f$ satisfying* (7)*, the FGD-SE updates in Algorithm 1 converge linearly to a neighborhood of $\hat{\mathbf{u}}$. Specifically, it can be shown that the update at iteration $k$ satisfies*

$$\text{Dist}(\mathbf{u}_{k+1}, \hat{\mathbf{u}})^2 \le \alpha \cdot \text{Dist}(\mathbf{u}_k, \hat{\mathbf{u}})^2 + \beta \cdot \|\hat{\mathbf{V}} - \hat{\mathbf{V}}^1\|_F^2 \quad (9)$$

*where $0 < \alpha < 1$ is the contraction rate and $\beta$ is a constant number, both depending on the optimal $\hat{\mathbf{u}}$.*

A sketch of the proof is provided here following from [16, Lem. 14-16]. Due to the rotational ambiguity, define $\text{Dist}(\mathbf{u}_k, \hat{\mathbf{u}}) := \|\boldsymbol{\delta}_k\|_2 := \|\mathbf{u}_k - \hat{\mathbf{u}}\mathbf{R}_k\|_2$, where the rotation matrix $\mathbf{R}_k$ yields the smallest Euclidean distance between the two. Since $\mathbf{u}_{k+1} - \hat{\mathbf{u}}\mathbf{R}_k = (\mathbf{u}_{k+1} - \mathbf{u}_k) + (\mathbf{u}_k - \hat{\mathbf{u}}\mathbf{R}_k)$, one can derive the following bound:

$$\|\boldsymbol{\delta}_{k+1}\|_2^2 \le \|\mathbf{u}_{k+1} - \mathbf{u}_k\|_2^2 + \|\boldsymbol{\delta}_k\|_2 - 2\langle \mathbf{u}_{k+1} - \mathbf{u}_k, \ -\boldsymbol{\delta}_k\rangle.$$

The FGD update dictates $\mathbf{u}_{k+1} - \mathbf{u}_k = \eta\nabla f(\mathbf{V}_k)\mathbf{u}_k$. It is important to characterize the third term in relation to the first two. By considering $\hat{\mathbf{V}}^1$, the third term is re-written as $2\langle \mathbf{u}_{k+1} - \mathbf{u}_k, \ -\boldsymbol{\delta}_k\rangle = \eta\langle \nabla f(\mathbf{V}_k), \ \mathbf{V}_k - \hat{\mathbf{V}}^1\rangle + \eta\langle \nabla f(\mathbf{V}_k), \ \boldsymbol{\delta}_k\boldsymbol{\delta}_k^{\mathcal{H}}\rangle$ where $\eta$ is the fixed stepsize in (8). Using Proposition 1 and (as1), one can show $\langle \nabla f(\mathbf{V}_k), \ \mathbf{V}_k - \hat{\mathbf{V}}^1\rangle \ge \theta_1\|\nabla f(\mathbf{V}_k)\mathbf{u}_k\|^2 + \alpha_1\|\boldsymbol{\delta}_k\|^2 - \beta_1\|\hat{\mathbf{V}} - \hat{\mathbf{V}}^1\|_F^2$, with all positive constants $\theta_1$, $\alpha_1$ and $\beta_1$. Additionally, using (as2) we have $\langle \nabla f(\mathbf{V}_k), \ \boldsymbol{\delta}_k\boldsymbol{\delta}_k^{\mathcal{H}}\rangle \ge -\theta_2\|\nabla f(\mathbf{V}_k)\mathbf{u}_k\|^2 - \alpha_2\|\boldsymbol{\delta}_k\|^2$ with

positive $\theta_2$ and $\alpha_2$ again. Adding these two lower bounds together, one can bound $\text{Dist}(\mathbf{u}_{k+1}, \hat{\mathbf{u}})^2 := \|\boldsymbol{\delta}_{k+1}\|_2^2$ as

$$\|\boldsymbol{\delta}_{k+1}\|_2^2 \leq -\theta_3 \|\nabla f(\mathbf{V}_k)\mathbf{u}_k\|_F^2 + \alpha\|\boldsymbol{\delta}_k\|_2^2 + \beta\|\hat{\mathbf{V}} - \hat{\mathbf{V}}^1\|_F^2.$$

Since $-\theta_3$ is negative, the first term can be ignored, which completes the proof for (9).

Note that (9) shows that the accuracy of the convergent solution depends on the term $\beta\|\hat{\mathbf{V}} - \hat{\mathbf{V}}^1\|_F^2$. If the optimum $\hat{\mathbf{V}}$ is rank-one, the FGD-SE updates will accordingly converge to the globally optimal $\hat{\mathbf{V}}$. This condition relates to the approximation performance of SDP-SE (3) to the original unrelaxed problem, which has been corroborated by both analytical and numerical results [7], [10], [11]. Hence, we will not focus on this aspect but instead discuss more on the convergence speed of FGD-SE, as related to the $\alpha$ value. Faster convergence speed requires a smaller $\alpha$, which critically depends on the initial $\mathbf{u}_0$ and the condition number $\kappa$. Various initialization schemes are available for SE, including flat start and dc power flow based solution. Our numerical tests suggest that these options work well for the FGD-SE updates, even though they do not strictly satisfy (as1). Our empirical experience has identified the main issue to be that the SE objective $f$ tends to have a very large condition number $\kappa$. As a result, the convergence speed for FGD would gradually decrease as the number of iterations $k$ increases. This motivates us to develop an accelerated scheme for FGD-SE as follows.

### B. Accelerated Gradient Descent Method

Slower convergence speed as a result of ill-conditioned objective function is a common issue for first-order methods. One popular improvement is to use the Nesterov's acceleration scheme first proposed in [17]. This improved first-order method has been shown to achieve *superlinear* convergence rate for convex objectives. Loosely speaking, it can reduce the condition number to $\sqrt{\kappa}$. Here, we develop the accelerated GD (AGD) based SE method for the nonconvex problem (4).

Different from FGD that uses only the instantaneous gradient, AGD takes the information from the past two iterations to compute the update. Per iteration $k$, a time-varying interpolation is first performed to obtain

$$\mathbf{u}^+ = \mathbf{u}_k + \left(\frac{k-1}{k+2}\right)(\mathbf{u}_k - \mathbf{u}_{k-1}), \forall k = 1, 2, \ldots \quad (10)$$

which is used to update $\mathbf{u}_{k+1}$ as the gradient descent on $\mathbf{u}^+$. The ratio $\mu = \frac{k-1}{k+2}$ is termed as the momentum parameter for AGD, which increases to 1 as $k$ approaches infinity. Note that the paper [17] proposed a momentum parameter $\mu = \frac{t_k - 1}{t_{k+1}}$ by generating a positive sequence $\{t_k\}$. It was shown in [24] that $t_k \geq \frac{k+1}{2}$ and thus one can approximately set a simpler parameter $\mu = \frac{t_k - 1}{t_{k+1}}$ with $t_k = \frac{k+1}{2}$. For $\mathbf{u}_0$ at $k = 0$, the update $\mathbf{u}_1$ is simply computed using the FGD rule. The AGD method is tabulated in Algorithm 2.

**Remark 1.** *(*AGD convergence for nonconvex $g$.) Recent results in [18] show that AGD can achieve linear convergence rate for approximating the SDP problem as well. Compared to (as1), the initial condition can be updated to $\text{Dist}(\mathbf{u}_0, \hat{\mathbf{u}}) \leq \frac{\rho_a}{\sqrt{\kappa}}\|\hat{\mathbf{u}}\|_2$ for AGD where $\rho_a$ is again a small constant. Clearly, this condition is more relaxed than (as1) as it increases the upper bound by a factor of $\sqrt{\kappa}$. In addition to relaxed condition, our numerical tests have shown that the AGD-SE method will achieve faster convergence speed than FGD-SE under exactly the same settings of initialization and step-size.

**Algorithm 2** AGD-SE

**Input:** Function $f$, rank $r$, maximum iteration number $K$.
**Output:** $\mathbf{u}$ and $\mathbf{V} = \mathbf{u}\mathbf{u}^{\mathcal{H}}$

1: Initialize $\mathbf{u}_0 \in \mathbb{C}^{n\times r}$ and set $\mathbf{V}_0 = \mathbf{u}_0\mathbf{u}_0^{\mathcal{H}}$.
2: Set the step-size $\eta$ as in (8).
3: Compute $\mathbf{u}_1 = \mathbf{u}_0 - \eta\nabla g(\mathbf{u}_0)$.
4: **for** $k = 1$ to $K$ **do**
5: Update $\mathbf{u}^+$ as (10) and $\mathbf{u}_{k+1} = \mathbf{u}^+ - \eta\nabla g(\mathbf{u}^+)$
6: **end for**
7: **return** $\mathbf{u}_K$ and $\mathbf{V} = \mathbf{u}_K\mathbf{u}_K^{\mathcal{H}}$

## IV. Practical Extensions for GD-SE Solvers

In addition to convergence guarantees and speed, it is truly important to consider practical issues in SE, such as i) robustness to bad data (outliers), ii) incorporation of additional meter types, and iii) multi-area implementation. Traditionally, outliers arise in SE due to data contamination, meter failure and synchronization issues [3], [4], [20]. More recently, malicious cyber attacks [25] and topology errors [8] can also contribute to SE outliers. Meanwhile, recent development of phasor measurement units (PMUs) motivates us to expand the SE solvers to incorporate synchrophasor data as well. There is also increasing trend to consider multi-area SE among different control centers [7], [13], [26]. All these practical concerns have been shown to potentially worsen the convergence issue of GN-SE method; see e.g., [4], [7], [8]. Due to page limit, we focus on discussing the first two extensions for GD-SE. Multi-area (Distributed) GD-SE can be developed using the framework of distributed linear SE method such as [26].

### A. Robust GD-SE against Outliers

One popular approach to tackle the presence of bad data is to introduce a sparse vector $\boldsymbol{\tau} \in \mathbb{R}^L$ with $\tau_\ell \neq 0$ indicating an outlier entry. This way, the robust SDP-SE (RSDP-SE) problem can be formulated as [11]

$$\min_{\mathbf{V},\boldsymbol{\tau}} \; f'(\mathbf{V}, \boldsymbol{\tau}) = \sum_{\ell=1}^{L} \frac{1}{2}\big\{[z_\ell - \text{Tr}(\mathbf{H}_\ell\mathbf{V})] - \tau_\ell\big\}^2 \quad (11a)$$

$$\text{s.t.} \;\; \mathbf{V} \succeq \mathbf{0}, \; \|\boldsymbol{\tau}\|_0 \leq \rho L \quad (11b)$$

with the pseudo-norm $\|\boldsymbol{\tau}\|_0 = \sum_\ell \mathbb{1}\{\tau_\ell \neq 0\}$ as the number of nonzero entries and $\rho$ is the given fraction of outliers. Hence, the constraint (11b) ensures that $\boldsymbol{\tau}$ is sparse with at most $\rho L$ non-zero entries.

To generalize the GD-SE methods to the RSDP-SE problem (11), we consider the nonconvex counterpart as $g'(\mathbf{u}, \boldsymbol{\tau}) := f'(\mathbf{u}\mathbf{u}^{\mathcal{H}}, \boldsymbol{\tau})$. Note that the GD updates now have to satisfy the additional sparsity constraint (11b). To this end, we adopt the idea of hard thresholding [27], [28] to obtain the *truncated*

**Algorithm 3** RFGD-SE and RAGD-SE

**Input:** Function $f$, rank $r$, maximum iteration number $K$, and outlier fraction $\rho$.
**Output:** $\mathbf{u}$ and $\mathbf{V}=\mathbf{u}\mathbf{u}^{\mathcal{H}}$

Initialize $\mathbf{u}_0\in\mathbb{C}^{n\times r}$ and set $\mathbf{V}_0=\mathbf{u}_0\mathbf{u}_0^{\mathcal{H}}$.
Compute the current error $\boldsymbol{\chi}_0$ and threshold it to $\boldsymbol{\tau}_0$.
Update $\mathbf{u}_1=\mathbf{u}_0-\eta\nabla_{\mathbf{u}}\, g'(\mathbf{u}_0,\boldsymbol{\tau}_0)$.
Set the step-size $\eta$ as in (8).
**for** $k=1$ to $K$ **do**
  **if** RAGD updates **then**
    Compute $\mathbf{u}^+=\mathbf{u}_k+(\frac{k-1}{k+2})(\mathbf{u}_k-\mathbf{u}_{k-1})$
  **else if** RFGD updates **then**
    Set $\mathbf{u}^+=\mathbf{u}_k$
  **end if**
  Compute the current error $\boldsymbol{\chi}_k=\mathbf{z}-\mathcal{H}(\mathbf{u}^+(\mathbf{u}^+)^{\mathcal{H}})$.
  Threshold the error to $\boldsymbol{\tau}_k=\mathcal{D}_{\rho L}(\boldsymbol{\chi}_k)$.
  Update $\mathbf{u}_{k+1}=\mathbf{u}^+-\eta\nabla_{\mathbf{u}}\, g'(\mathbf{u}^+,\boldsymbol{\tau}_k)$.
**end for**
**return** $\mathbf{u}_K$ and $\mathbf{V}=\mathbf{u}_K\mathbf{u}_K^{\mathcal{H}}$

gradient by removing the data samples with significantly high mismatch error. For a given integer $\gamma\leq L$, define the hard thresholding operator $\mathcal{D}_\gamma$ as

$$\big[\mathcal{D}_\gamma(\boldsymbol{\chi})\big]_\ell:=\begin{cases}\chi_\ell, & \text{if } |\chi_\ell|\geq|\boldsymbol{\chi}^{(\gamma)}|,\\ 0, & \text{otherwise,}\end{cases}\tag{12}$$

where $|\boldsymbol{\chi}^{(\gamma)}|$ denotes the entry in $\boldsymbol{\chi}\in\mathbb{R}^L$ with the $\gamma$-largest absolute value. Equivalently, $\mathcal{D}$ throws away the $\gamma$-largest values of $\boldsymbol{\chi}$. For any given $\mathbf{u}$, one can use the thresholding operator on the instantaneous error mismatch, namely $\chi_\ell=z_\ell-\mathbf{u}^{\mathcal{H}}\mathbf{H}_\ell\mathbf{u}$. Hence, by setting the outlier indicator $\boldsymbol{\tau}=\mathcal{D}_{\rho L}(\boldsymbol{\chi})$, the former always satisfies the sparsity constraint, and the gradient of $g'$ at $\mathbf{u}$ becomes

$$\begin{aligned}\nabla_{\mathbf{u}}\, g'(\mathbf{u},\boldsymbol{\tau})&=2\nabla_{\mathbf{u}}f'(\mathbf{u}\mathbf{u}^{\mathcal{H}},\boldsymbol{\tau})\times\mathbf{u}\\&=\sum_{\ell=1}^{L}2\big(\mathbf{u}^{\mathcal{H}}\mathbf{H}_\ell\mathbf{u}-z_\ell+\tau_\ell\big)\mathbf{H}_\ell\mathbf{u}.\end{aligned}\tag{13}$$

Since $\boldsymbol{\tau}$ is the hard thresholded output of $\boldsymbol{\chi}$, the summation in (13) only takes $(L-\rho L)$ measurements of smaller mismatch errors and rules out the rest of higher mismatch errors.

Using the truncated gradient in (13), we can develop the robust (R)FGD and (R)AGD methods for SE, as tabulated in Algorithm 3. The two algorithms are jointly presented as they are only different in the gradient updates, on either $\mathbf{u}^+$ or $\mathbf{u}_k$. Other settings such as initialization and step-size follow from the original FGD/AGD method.

**Remark 2.** *(*Performance of Robust GD-SE) The GD based updates using truncated gradient in (13) has been proposed for solving nonconvex optimization problems under outliers, such as the robust phase retrieval problem [27], [29], and more recently the robust matrix factorization problem [28]. Results therein suggest the truncated GD can achieve linear convergence rate with sufficiently good initialization, given that the original function is strongly convex and smooth. However, the convergence guarantee and rate of the truncated AGD updates are still an open question for nonconvex optimization. Our numerical tests have shown that the RAGD-SE method could achieve accelerated convergence performance than RFGD-SE, while both are shown to converge numerically.

### *B. Augmented GD-SE with PMU Data*

Compare to legacy quadratic measurement model of $\mathbf{v}$, PMUs provide synchronous phasor data that are linearly related to the state $\mathbf{v}$. If bus $n$ is equipped with a PMU, then its voltage phasor $V_n$ and the incident line current phasors $\{I_{nn'}\}$ are available with high accuracy. When there are sufficient PMU measurements making the system observable, SE will be non-iterative and fast thanks to the the linearity between the PMU measurements and the system states. Currently and in near future, power systems still have limited PMUs due to the high installation and networking costs. Hence, SE must be performed using both the legacy meters and PMU data.

Let $\{\boldsymbol{\zeta}_n\}_{n\in\mathcal{P}}$ denote all the PMU measurements, where $\mathcal{P}\subseteq\mathcal{N}$ denotes the PMU-instrumented buses. The noisy PMU data at bus $n$ can be modeled as $\boldsymbol{\zeta}_n=\boldsymbol{\Phi}_n\mathbf{v}+\boldsymbol{\varepsilon}_n$, where $\boldsymbol{\Phi}_n$ denotes the measurement matrix constructed in accordance to the bus index $n$ and line parameters [30] and $\boldsymbol{\varepsilon}_n$ the measurement noise assumed to be complex zero-mean Gaussian. Given both $\mathbf{z}$ and $\{\boldsymbol{\zeta}_n\}_{n\in\mathcal{P}}$, the joint SE problem aims to minimize the augmented error objective as

$$\min_{\mathbf{u}\in\mathbb{C}^N} g^a(\mathbf{u}):=g(\mathbf{u})+\frac{1}{2}\sum_{n\in\mathcal{P}}\|\boldsymbol{\zeta}_n-\boldsymbol{\Phi}_n\mathbf{u}\|_2^2.\tag{14}$$

The PMU-augmented SE problem (14) is still nonconvex due to $g(\mathbf{u})$. Nonetheless, its gradient function can be formed by

$$\nabla g^a(\mathbf{u})=\nabla g(\mathbf{u})+\sum_{n\in\mathcal{P}}\boldsymbol{\Phi}_n^{\mathcal{H}}(\boldsymbol{\Phi}_n\mathbf{u}-\boldsymbol{\zeta}_n).\tag{15}$$

Using this gradient function, both FGD-SE and AGD-SE methods can be extended to include PMU data. The new LS term from PMU data can be thought of as an additional regularization term on the original objective, which can improve the effectiveness of gradient update thanks to the accuracy of PMU data. Accordingly, the augmented FGD-SE or AGD-SE methods could accelerate their respective counterparts, as verified by our numerical results.

## V. Numerical Results

The proposed FGD/AGD-SE methods and their extensions have been tested on a laptop with Intel® CPU @ 2.2GHz (8GB RAM) in the MATLAB® R2017a simulator. They are compared with the SDP-SE solutions using the MATLAB-based optimization modeling package `CVX` [31] together with SeDuMi [21] solver. Three power transmission system test cases, namely, the IEEE 118-bus, 300-bus, and the synthetic ACTIVSg2000-bus [32] systems are used with the pertinent power flow solver and GN-SE iterations implemented by the MATLAB-based toolbox MATPOWER [33]. To generate the measurements, random Gaussian noise (in p.u.) is added to the power flow output, with $\sigma_\ell=0.02$ at line flow meters, 0.04 at power injection meters, 0.004 at voltage meters, and 0.0004 at PMUs, respectively. Empirical root mean-square estimation

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2019.2950392, IEEE Transactions on Power Systems



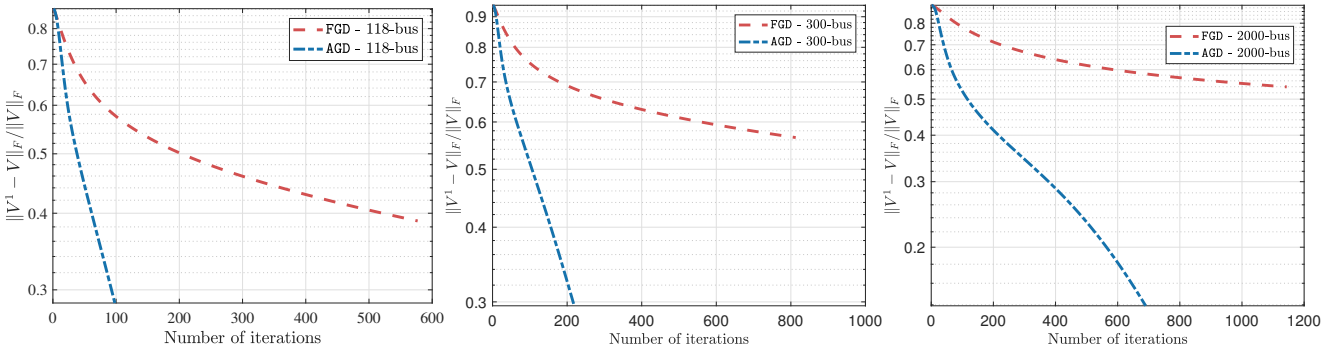


Fig. 1: Average iterative error from the ground-truth $\mathbf{V}$ for the (left) 118-bus, (middle) 300-bus and (right) 2000-bus systems.

TABLE I: RMSE and GN Convergence Rate (GNCR)

| SE Error | 118-bus | 300-bus | 2000-bus |
|---|---|---|---|
| GN | 0.022 (96%) | 0.103 (70%) | 0.256 (30%) |
| SDP-GN | 0.003 (100%) | 0.017 (100%) | N/A (N/A) |
| FGD-GN | 0.003 (100%) | 0.017 (100%) | 0.004 (100%) |
| AGD-GN | 0.003 (100%) | 0.017 (100%) | 0.004 (100%) |

TABLE II: Average Runtime of LS-SE Solutions

| Time | 118-bus | 300-bus | 2000-bus |
|---|---|---|---|
| GN | 0.059s | 2.039s | 151.801s |
| SDP | 8.650s | 98.557s | N/A |
| FGD | 0.068s | 0.749s | 73.101s |
| AGD | 0.039s | 0.277s | 35.733s |

error (RMSE) $||\hat{\mathbf{v}} - \mathbf{v}||_2/\|\mathbf{v}\|_2$ is computed by averaging over 100 Monte-Carlo realizations. For each realization, the actual bus voltage magnitude (in p.u.) and phase angle (in rad) of each bus are uniformly distributed over $[0.95, 1.05]$ and $[-0.35\pi, 0.35\pi]$, respectively. To construct the normalized systems, we follow the earlier approaches [12], [34] scale with measurement as $\{\frac{z_\ell}{\|\mathbf{H}_\ell\|_F}; \frac{\mathbf{H}_\ell}{\|\mathbf{H}_\ell\|_F}\}_{\ell=1}^{L}$.

To initialize all iterative updates, we set the bus voltage magnitude to be its measured value, or 1 p.u.. The phase angles are initialized as the linear SE solutions based on dc power flow model. For all GD-based SE approaches throughout the tests, the stepsize is chosen using (8) by fixing $1/c = 1/4$. Generally, the SDP-SE solution is not perfectly rank-one. Thus, the popular eigen-decomposition approach is used to retrieve the best rank-one approximation; see e.g., [7]. All estimated voltage vectors by GD-SE or SDP-SE will be further improved using the GN updates to reduce the optimality gap. Empirically, 3-4 GN iterations are sufficient for convergence, and thus this additional computational time is neglected for evaluating the runtime of SDP-/GD-SE methods.

*1) Test Case 1 on LS-SE Objective:* We first test all methods on the LS objective function (4). The measurements include all bus voltage magnitudes, and all active and reactive line flows at the 'from' end. The stopping criteria for GD-SE iterations are based on consecutive change of the iterate and that of the objective value, while the stop criterion for GN is based on the first-order optimal condition.

To demonstrate the performance of FGD and AGD in estimating the ground-truth $\mathbf{V}$, the average iterative error $\|\mathbf{V}^1 - \mathbf{V}\|_F/\|\mathbf{V}\|_F$, where matrix $\mathbf{V}^1$ stands for the instantaneous matrix solution per iteration, is plotted in Fig. 1. Both GD methods use the same step-size, showing nearly linear convergence rate. Meanwhile, the AGD updates always outperform FGD in terms of the convergence speed, thanks to the improved condition number of AGD as in Remark 1.

We also compare the SE performance achieved by GD-SE methods, to that by GN-SE and SDP-SE, as listed in Table I, along with the percentage of convergence for the respective GN updates. As mentioned earlier, the SDP-SE and GD-SE solutions are used to initialize the GN updates (indicated by the -GN) for reduced optimality gap. Table I shows that these solutions achieve 100% GN convergence rate (GNCR), verifying their near-optimal performance. As a comparison, the GN-SE experiences divergence issues, even more seriously as the system sizes increases. In addition, both FGD-GN and AGD-GN achieve the same SE performance as the benchmark SDP-GN. For the largest 2000-bus case, SDP-SE cannot be solved by the generic CVX solver in reasonable time, and thus its error performance is not reported. We can notice that FGD-/AGD-GN still demonstrate superior error performance for the 2000-bus case.

To better investigate the computational time improvement, we have listed the average runtime of all solution techniques over 100 realizations in Table II. Compared to the SDP solver, both FGD and AGD scale nicely with the system size, especially for AGD thanks to its improved initial condition. Since the SDP solution is not found for the 2000-bus case, its runtime is not reported. In contrast, AGD takes less than one minute to converge. Compared to the two GD-based solutions, the average runtime of GN from the original initialization is much longer, mainly due to the divergence issue. Note that the GN updates are set to stop at 200 iterations if not convergent. This comparison again verifies the improvement of gradient-based iterations, in approaching the globally optimal SE solutions.

*2) Test Case 2 on Robust SE:* To generate outliers, five measurement meters are randomly picked in each Monte-Carlo run, with the corresponding data purposely manipulated to 5 times of actual value. To find the benchmark performance for the proposed RFGD/RAGD method, we consider a re-formulated robust SDP-SE model (RSDP) by using the $l_1$-norm instead of $l_0$-norm term in (11) [35]. Specifically, we solve its Lagrangian form augmenting the objective function

TABLE III: RMSE and Outlier Identification Rate (OIR)

| SE | 118-bus | 300-bus | 2000-bus |
|---|---|---|---|
| LNR | 0.177(48%) | 0.529(4%) | 1.363 (3%) |
| LAV | 0.187(30%) | 0.402(8%) | 1.099 (2%) |
| SPL | 0.218(4%) | 0.358(3%) | 0.451 (1%) |
| RSDP | 0.028(60%) | 0.078(45%) | N/A (N/A) |
| RFGD | 0.027(67%) | 0.065(54%) | 0.036 (55%) |
| RAGD | 0.021(73%) | 0.054(60%) | 0.026 (72%) |

TABLE IV: Average Runtime of Robust SE Solutions

| Time | 118-bus | 300-bus | 2000-bus |
|---|---|---|---|
| LNR | 2.875s | 24.626s | 3091.6s |
| LAV | 17.185s | 42.623s | 182.423s |
| SPL | 1.424s | 4.131s | 90.342s |
| RSDP | 8.345s | 120.768s | N/A |
| RFGD | 0.419s | 1.776s | 79.165s |
| RAGD | 0.226s | 1.034s | 42.188s |

TABLE V: RMSE and OIR under Interacting Outliers (IO)

| SE | 118-bus | 300-bus | 2000-bus |
|---|---|---|---|
| LNR | 29.924(38%) | 0.422(10%) | 1.205 (1%) |
| LAV | 0.246(26%) | 0.298(20%) | 0.882 (0%) |
| SPL | 0.197(9%) | 0.321(2%) | 0.379 (0%) |
| RSDP | 0.056(47%) | 0.060(41%) | N/A (N/A) |
| RFGD | 0.043(52%) | 0.044(49%) | 0.029 (54%) |
| RAGD | 0.036(60%) | 0.037(55%) | 0.027 (58%) |

TABLE VI: Runtime of Robust SE Solutions under IO

| Time | 118-bus | 300-bus | 2000-bus |
|---|---|---|---|
| LNR | 3.625s | 29.075s | 2583.1s |
| LAV | 32.659s | 45.445s | 180.909s |
| SPL | 1.551s | 4.397s | 92.465s |
| RSDP | 9.359s | 132.801s | N/A |
| RFGD | 0.490s | 2.139s | 90.286s |
| RAGD | 0.240s | 1.605s | 43.793s |

as $\hat{\mathbf{V}} = \arg\min_{\mathbf{V},\boldsymbol{\tau}} f'(\mathbf{V},\boldsymbol{\tau}) + \lambda\|\boldsymbol{\tau}\|_1$, where $\lambda$ is a positive parameter. The GN-based largest normalized residual (LNR) test as developed in [19, Ch. 5] is included as well. LNR repeats a procedure consisting of bad data detection, identification, and deletion of single measurement of largest error, until no bad data are to be found. We also implement the classical robust least-absolute value (LAV) estimator [3], which iteratively solves the linearized problem using the latest iterates. Recently, the robust LAV objective has been solved using the so-termed stochastic prox-linear (SPL) method [20], which can accelerate the convergence speed by using a special linearization on the complex phasor representation.

We test all robust solvers, namely LNR, LAV, SPL, RSDP, RFGD/RAGD, on the three test cases. For simplicity, the initialization is set to be flat start. Again, the $l_1$-norm based RSDP is solved by CVX, where $\lambda$ is set based on the level of 3 times noise standard deviation. For all methods, the total number of outliers is set to be 10. This choice is very conservative but shown reasonable performance under good measurement redundancy. More advanced methods for determining the number of outliers can be found in [36, Ch. 6]. Upon the convergence of all methods, 10 measurements with the largest deviation errors will be identified as outliers and compared to the ground truth to compute the identification success rate. Afterwards, the other measurements together with the solution of $\mathbf{u}$ will be used to evoke the GN iterations for evaluating the SE error performance.

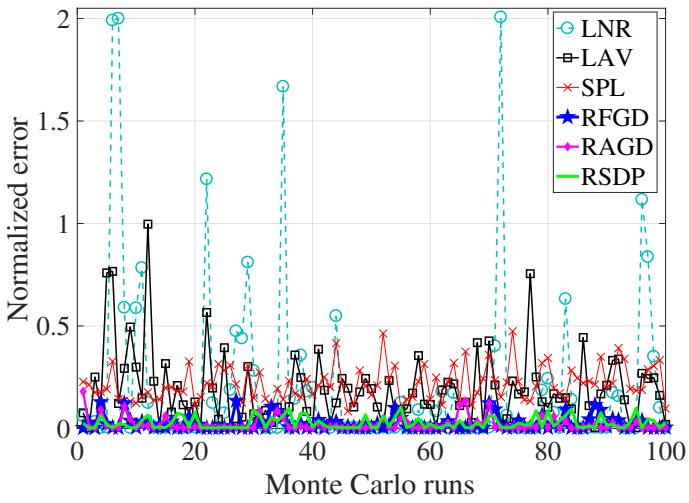


Fig. 2: Estimation performance for the IEEE 118-bus system with 5 actual outliers.

Table III lists the RMSE and success rate of identifying outliers for all robust SE methods, while Table IV provides the average runtime. Similar to Test Case 1, RFGD and RAGD consistently outperform the other robust solutions, in both the SE error and outlier identification rate (OIR), with similar computational time. Compared to Table I, all the methods experience slightly larger RMSE because of the existence of outlier data. However, the performance of RFGD/RAGD is still elegant as they both can identify the majority of outliers. Interestingly for the robust case, RAGD at fastest runtime slightly outperforms RFGD in terms of error performance. Also, both of them can even improve the RSDP solution. This is perhaps because the robust GD methods perform outlier thresholding at every iteration and thus exclude outlier data more effectively. Notably, Table III indicates that the OIR for LNR, LAV, or SPL decreases as the system size increases from 118-bus to 2000-bus. This points out the importance of having a well-designed robust SE algorithm for practical systems. The normalized error for each Monte-Carlo run using the 118-bus system as plotted in Fig. 2 also confirms that GD-based robust SE solutions consistently outperform other counterparts.

We further test the proposed robust SE methods in the presence of multiple interacting outliers (IO). Following the steps in [19, Ch. 5], in every realization we pick five bad data to be *interacting*, such that residuals of these measurements are strongly correlated based on the measurement residual sensitivity matrix. The RMSE and the OIR are listed in Table V, and Table VI shows the average runtime. Compared to the results for random outliers in Table III, the presence of IO can lead to increased RMSE values, especially for the LNR and LAV estimators using the 118-bus case. For the larger 300-bus and 2000-bus cases, the impact of IO is very minimal. By and large, the proposed RFGD and RAGD methods still outperform other robust estimators in both the RMSE and OIR, even for the IO case.

*3) Test Case 3 on Low-Redundancy (LR) System:* To better address numerical stability concerns, we test a modified 118-bus system with LR in measurements. This LR 118-bus system

TABLE VII: LS-SE Results for the LR 118-bus System

| SE | GN | SDP-GN | FGD-GN | AGD-GN |
|---|---|---|---|---|
| RMSE | 0.075 | 0.007 | 0.007 | 0.007 |
| GNCR | 94% | 100% | 100% | 100% |
| Runtime | 0.052s | 7.854s | 0.056s | 0.042s |

TABLE VIII: Robust-SE Results for the LR 118-bus System

| SE | LNR | LAV | SPL | RSDP | RFGD | RAGD |
|---|---|---|---|---|---|---|
| RMSE | 0.843 | 0.628 | 0.387 | 0.091 | 0.079 | 0.072 |
| OIR | 12% | 13% | 5% | 55% | 64% | 69% |
| Time | 3.466s | 31.835s | 1.501s | 10.932s | 0.563s | 0.382s |

has a total of 418 measurements, including all the bus voltage magnitudes and 300 power measurements; see e.g., the setting in [4]. All other settings follow from Test Case 1. The SE results are shown in Table VII, including all three metrics. Compared with the original system results in Tables I, the LR system shows similar comparisons among the four SE methods. Notably, the convergence rate of GN updates slightly decreases and RMSE error increases, as a result of less measurements. Thus, the proposed GD-SE methods are more robustness to low measurement redundancy.

We also test all robust solvers using the LR 118-bus system with 5 randomly selected outliers. All other settings follow from Test Case 2. The RMSE, OIR and average runtime are listed in Table VIII. The results for both LNR and LAV using the LR system show high level of performance degradation compared to those for the original system in Table III. Specifically, the OIR of LNR decreases from 48% to 12%, while that of LAV decreases from 30% to 13%. On the contrary, our proposed RFGD/RAGD methods can still maintain the same performance ($> 60\%$ of OIR) for the LR system. Hence, the LR system tests further demonstrate the convergence issue of traditional SE methods, while confirming that our proposed ones are more effective in dealing with ill-conditioned problems or bad data.

*4) Test Case 4 on PMU-aided SE:* We further evaluate the SE performance with the additional PMU data on the three test cases. All methods in Test Case 1 are considered, with their PMU-aided counterparts denoted by the subscript $_p$ here. For $\text{GN}_p$, the sequential approach of including PMU data in [37] is adopted. It entails two steps: i) the LS-based SE is performed first to process the legacy measurements, followed by ii) a post-processing step which only involves a linear models together with PMU data. Four buses are selected to equip with PMUs, namely $\{10, 12, 27, 15\}$, based on the PMU placement work in [26]. For these buses, the legacy meters are no longer included.

Similarly, Table IX lists the SE error and average runtime for all PMU-augmented solutions. Compare to Table I, the SE error has been improved for all scenarios, thanks to the accurate PMU data. Both $\text{FGD}_p$ and $\text{AGD}_p$ still achieve the same estimation performance as $\text{SDP}_p$. Compared to the runtime in Table II, the two GD-based solutions are even faster with PMU data, thanks to the additional regularization that the latter provides to the objective function.

TABLE IX: RMSE and Average Runtime of PMU-aided SE

| SE Error (Time) | 118-bus | 300-bus | 2000-bus |
|---|---|---|---|
| $\text{GN}_p$ | 0.016 (0.289s) | 0.091 (7.837s) | 0.182 (151.504s) |
| $\text{SDP}_p$ | 0.002 (8.415s) | 0.015 (99.070s) | N/A (N/A) |
| $\text{FGD}_p$ | 0.002 (0.058s) | 0.015 (0.469s) | 0.003 (72.848s) |
| $\text{AGD}_p$ | 0.002 (0.019s) | 0.015 (0.114s) | 0.003 (30.626s) |

## VI. Conclusions

This paper presented a gradient descent (GD) based framework for solving the nonconvex SE formulation, in order to accelerate the convex SDP-based SE for power system monitoring. The SDP formulation can offer near-optimality performance and improve the divergence issue of the iterative GN method. To tackle the high computational complexity of SDP-SE, we propose to adopt the factored (F)GD-update for the nonconvex objective on voltage phasor vector. Furthermore, the accelerated (A)GD-update is developed to improve the condition of the objective function. For FGD-SE, a linear convergence rate is guaranteed based on the analysis of the strong convexity and smoothness of the LS objective, whereas simiar result also holds for AGD-SE. Both proposed FGD-/AGD-SE methods can be extended to include practical scenarios of measurement outliers and PMU data. Extensive numerical comparisons have demonstrated the near-optimal error performance of the proposed approaches, while greatly reducing the computation time.

Interesting future research directions open up, including the convergence guarantee and rate of robust AGD-SE method. Moreover, we are interested to pursue the constrained SDP extension for SE problem under voltage limits or even optimal power flow problem.

## Appendix

### A. Calculation of Complex Gradient of a Real Function

Consider a real function $g$ with the complex $\mathbf{u}$ input, the complex gradient direction is given by [38, Ch. 4]

$$\nabla g(\mathbf{u}) = \frac{\partial g(\mathbf{u})}{\partial \Re(\mathbf{u})} + \mathrm{j}\frac{\partial g(\mathbf{u})}{\partial \Im(\mathbf{u})}. \tag{16}$$

For the nonconvex SE objective $g$ in (4), we can use the chain-rule to find

$$\nabla g(\mathbf{u}) = \sum\nolimits_{\ell=1}^{L} (\mathbf{u}^{\mathcal{H}}\mathbf{H}_\ell\mathbf{u} - z_\ell)\nabla(\mathbf{u}^{\mathcal{H}}\mathbf{H}_\ell\mathbf{u}). \tag{17}$$

To find the second-term in (17), let $\mathbf{u} = \mathbf{r} + \mathrm{j}\mathbf{x}$ and $\mathbf{H}_\ell = \mathbf{R}_\ell + \mathrm{j}\mathbf{X}_\ell$, and thus the real function $(\mathbf{u}^{\mathcal{H}}\mathbf{H}_\ell\mathbf{u}) = \mathrm{Tr}[\mathbf{R}_\ell(\mathbf{r}\mathbf{r}^{\mathcal{T}} + \mathbf{x}\mathbf{x}^{\mathcal{T}}) - \mathbf{X}_\ell(\mathbf{x}\mathbf{r}^{\mathcal{T}} - \mathbf{r}\mathbf{x}^{\mathcal{T}})]$. Thus, we have

$$\frac{\partial(\mathbf{u}^{\mathcal{H}}\mathbf{H}_\ell\mathbf{u})}{\partial \mathbf{r}} \overset{(i)}{=} 2\mathbf{R}_\ell\mathbf{r} - \mathbf{X}_\ell\mathbf{x} + \mathbf{X}_\ell^{\mathcal{T}}\mathbf{x} \overset{(ii)}{=} 2\mathbf{R}_\ell\mathbf{r} - 2\mathbf{X}_\ell\mathbf{x}$$

where $(i)$ is due to the derivative of trace, and $(ii)$ follows from $\mathbf{X}_\ell^{\mathcal{T}} = -\mathbf{X}_\ell$ as $\mathbf{H}_\ell$ is Hermitian. Simiarly, we can find the partial derivative with respect to $\mathbf{x}$ and thus the complex gradient $\nabla(\mathbf{u}^{\mathcal{H}}\mathbf{H}_\ell\mathbf{u}) = (2\mathbf{R}_\ell\mathbf{r} - 2\mathbf{X}_\ell\mathbf{x}) + \mathrm{j}(2\mathbf{R}_\ell\mathbf{x} + 2\mathbf{X}_\ell\mathbf{r})$, which is exactly equal to the complex product of $2\mathbf{H}_\ell\mathbf{u}$. This completes the proof of (5).

### *B. Analysis of the SDP-SE Objective Function $f$*

Consider any $\mathbf{V} \in \mathcal{V}$, $\mathcal{H}(\mathbf{V})$ is the vector of error-free measurements. Thus, we have

$$\|\mathcal{H}(\mathbf{V})\|_2^2 = \sum_{n\in\mathcal{N}_V} |V_n|^4 + \sum_{(n,n')\in\mathcal{E}_S} P_{nn'}^2 + \sum_{(n,n')\in\mathcal{E}_S} Q_{nn'}^2 + \sum_{n\in\mathcal{N}_S} P_n^2 + \sum_{n\in\mathcal{N}_S} Q_n^2, \tag{18}$$

where the sets denote the corresponding meter locations. Using the power flow model, we can derive upper/lower bounds for each of the summand terms in (18). First, using the voltage limits in (6) we have

$$\underline{V}^4 \le |V_n|^4 \le \bar{V}^4. \tag{19}$$

As for the line flows, a trivial lower bound is 0, with the upper bound given by

$$\begin{aligned} &P_{nn'}^2 + Q_{nn'}^2 = |S_{nn'}|^2 = |V_n(y_{nn'}(V_n - V_{n'}))^{\mathcal{H}}|^2 \\ \le&|V_n|^2|y_{nn'}|^2(|V_n| + |V_{n'}|)^2 \le 4|y_{nn'}|^2\bar{V}^4, \end{aligned} \tag{20}$$

by applying the triangle inequality. Similarly, the power injection is upper bounded by

$$\begin{aligned} &P_n^2 + Q_n^2 = |S_n|^2 = |V_n i_n^{\mathcal{H}}|^2 = |V_n(\sum_{\nu\in\mathcal{N}_n} y_{n\nu}V_\nu)^{\mathcal{H}}|^2 \\ \le&|V_n|^2\Big(\sum_{(n,\nu)\in\mathcal{E}} |y_{n\nu}|\bar{V}\Big)^2 \le \Big(\sum_{(n,\nu)\in\mathcal{E}} |y_{n\nu}|\Big)^2\bar{V}^4. \end{aligned} \tag{21}$$

This way, we can find the upper/lower bounds for $\|\mathcal{H}(\mathbf{V})\|_2^2$ to quantify the two positive parameters $m$ and $M$. Note that we use a trivial lower bound of 0 for power measurements. This bound can be further improved by assuming a minimal angular separation between the two end buses of any line or using the system's total power demand.

## REFERENCES


[1] A. J. Wood and B. F. Wollenberg, *Power generation, operation, and control*. John Wiley & Sons, 2012.

[2] G. B. Giannakis, V. Kekatos, N. Gatsis, S.-J. Kim, H. Zhu, and B. F. Wollenberg, "Monitoring and optimization for power grids: A signal processing perspective," *IEEE Signal Processing Magazine*, vol. 30, no. 5, pp. 107–128, 2013.

[3] A. Monticelli, "Electric power system state estimation," *Proceedings of the IEEE*, vol. 88, no. 2, pp. 262–282, 2000.

[4] J. Zhao, L. Mili, and R. C. Pires, "Statistical and numerical robust state estimator for heavily loaded power systems," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6904–6914, 2018.

[5] R. Madani, J. Lavaei, R. Baldick, and A. Atamtürk, "Power system state estimation and bad data detection by means of conic relaxation," in *Proc. 50th Hawaii Intl. Conf. on System Sciences*, 2017.

[6] R. Zhang, J. Lavaei, and R. Baldick, "Spurious critical points in power system state estimation," in *Proceedings of the 51st Hawaii International Conference on System Sciences*, 2018.

[7] H. Zhu and G. B. Giannakis, "Power system nonlinear state estimation using distributed semidefinite programming," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 6, pp. 1039–1050, 2014.

[8] Y. Weng, M. D. Ilić, Q. Li, and R. Negi, "Convexification of bad data and topology error detection and identification problems in ac electric power systems," *IET Generation, Transmission & Distribution*, vol. 9, no. 16, pp. 2760–2767, 2015.

[9] Y. Zhang, R. Madani, and J. Lavaei, "Conic relaxations for power system state estimation with line measurements," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 3, pp. 1193–1205, 2017.

[10] R. Madani, J. Lavaei, and R. Baldick, "Convexification of power flow equations for power systems in presence of noisy measurements," *IEEE Transactions on Automatic Control*, 2019.

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

[12] G. Wang, G. B. Giannakis, and J. Chen, "Robust and scalable power system state estimation via composite optimization," *IEEE Transactions on Smart Grid*, 2019.

[13] R. Madani, A. Kalbat, and J. Lavaei, "A low-complexity parallelizable numerical algorithm for sparse semidefinite programming," *IEEE Transactions on Control of Network Systems*, 2017.

[14] S. Burer and R. D. Monteiro, "A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization," *Mathematical Programming*, vol. 95, no. 2, pp. 329–357, 2003.

[15] ——, "Local minima and convergence in low-rank semidefinite programming," *Mathematical Programming*, vol. 103, no. 3, pp. 427–444, 2005.

[16] S. Bhojanapalli, A. Kyrillidis, and S. Sanghavi, "Dropping convexity for faster semi-definite optimization," in *Conference on Learning Theory*, 2016, pp. 530–582.

[17] Y. Nesterov, "A method of solving a convex programming problem with convergence rate o (1/k2)," in *Soviet Mathematics Doklady*, vol. 27, no. 2, 1983, pp. 372–376.

[18] A. Kyrillidis, S. Ubaru, G. Kollias, and K. Bouchard, "Run procrustes, run! on the convergence of accelerated procrustes flow," *arXiv preprint arXiv:1806.00534*, 2018.

[19] A. Abur and A. G. Exposito, *Power system state estimation: theory and implementation*. CRC press, 2004.

[20] G. Wang, H. Zhu, G. B. Giannakis, and J. Sun, "Robust power system state estimation from rank-one measurements," *IEEE Transactions on Control of Network Systems*, 2019.

[21] J. F. Sturm, "Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones," *Optimization methods and software*, vol. 11, no. 1-4, pp. 625–653, 1999.

[22] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[23] S. Negahban and M. J. Wainwright, "Restricted strong convexity and weighted matrix completion: Optimal bounds with noise," *Journal of Machine Learning Research*, vol. 13, no. May, pp. 1665–1697, 2012.

[24] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[25] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Transactions on Information and System Security (TISSEC)*, vol. 14, no. 1, p. 13, 2011.

[26] V. Kekatos and G. B. Giannakis, "Distributed robust power system state estimation," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1617–1626, 2013.

[27] J. Chen, L. Wang, X. Zhang, and Q. Gu, "Robust wirtinger flow for phase retrieval with arbitrary corruption," *arXiv: 1704.06256*, 2017.

[28] Y. Li, Y. Chi, H. Zhang, and Y. Liang, "Non-convex low-rank matrix recovery with arbitrary outliers via median-truncated gradient descent," *Information and Inference: A Journal of the IMA*, 05 2019. [Online]. Available: https://doi.org/10.1093/imaiai/iaz009

[29] H. Zhang, Y. Chi, and Y. Liang, "Median-truncated nonconvex approach for phase retrieval with outliers," *IEEE Transactions on Information Theory*, vol. 64, no. 11, pp. 7287–7310, 2018.

[30] V. Kekatos, G. B. Giannakis, and B. Wollenberg, "Optimal placement of phasor measurement units via convex relaxation," *IEEE Transactions on power systems*, vol. 27, no. 3, pp. 1521–1530, 2012.

[31] M. Grant, S. Boyd, and Y. Ye, "Cvx: Matlab software for disciplined convex programming," 2008.

[32] A. B. Birchfield, T. Xu, K. M. Gegner, K. S. Shetye, and T. J. Overbye, "Grid structural characteristics as validation criteria for synthetic networks," *IEEE Transactions on power systems*, vol. 32, no. 4, pp. 3258–3265, 2017.

[33] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "Matpower: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Transactions on power systems*, vol. 26, no. 1, pp. 12–19, 2011.

[34] 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, 2014.

[35] W. Zheng, W. Wu, A. Gomez-Exposito, B. Zhang, and Y. Guo, "Distributed robust bilinear state estimation for power systems with nonlinear measurements," *IEEE Transactions on Power Systems*, vol. 32, no. 1, pp. 499–509, 2017.

[36] P. J. Rousseeuw and A. M. Leroy, *Robust regression and outlier detection*. John wiley & sons, 2005, vol. 589.

[37] M. Zhou, V. A. Centeno, J. S. Thorp, and A. G. Phadke, "An alternative for including phasor measurements in state estimators," *IEEE transactions on power systems*, vol. 21, no. 4, pp. 1930–1937, 2006.
[38] K. B. Petersen, M. S. Pedersen *et al.*, "The matrix cookbook," *Technical University of Denmark*, vol. 7, no. 15, p. 510, 2008.

**Yu Lan** (S'14) received the B.E. degree in automation from Xidian University, Xian, China, in 2012. Currently, she is working towards the Ph.D. degree at the System Engineering Institute, Xian Jiaotong University, Xian, China. She was a visiting student in University of Texas at Austin from 2017 to 2018. Her research interests include control of smart power grid, distributed optimization of large-scale electric power systems, large-scale power system state estimation.

**Hao Zhu** (M'12, SM'19) received the B.E. degree from Tsinghua University in 2006, and the M.Sc. and Ph.D. degrees from the University of Minnesota in 2009 and 2012, all in electrical engineering. She has been an Assistant Professor of Electrical and Computer Engineering with The University of Texas at Austin since 2017. Before that, she was a Post-Doctoral Research Associate and then an Assistant Professor at the University of Illinois at Urbana-Champaign. Her current research interests include power grid monitoring, power system operations and control, and energy data analytics. She received the NSF CAREER Award and the Siebel Energy Institute Seed Grant Award in 2017. Dr. Zhu supervised and co- and the 2nd Best Paper Award at the 2016 North American Power Symposium. She is currently a member of the Big Data Task Force of the IEEE Power & Energy Society (IEEE-PES), the Signal Processing Theory and Methods (SPTM) Technical Committee (TC) and Big Data Special Interest Group (SIG) of the IEEE Signal Processing Society (IEEE-SPS), and the Smart Grid SIG of the IEEE Communications Society (COMSOC).

**Xiaohong Guan** (M'93-SM'94-F'07) received the B.S. and M.S. degrees in control engineering from Tsinghua University, Beijing, China in 1982 and 1985 respectively, and the Ph.D. degree in electrical engineering from University of Connecticut in 1993. From 1993 to 1995, he was a Senior Consulting Engineer at Pacific Gas and Electric Company (PG&E), San Francisco, CA, USA. He was a Visiting Scholar with the Division of Engineering and Applied Science, Harvard University, from 1999 to 2000. Since 1995, he has been with the Systems Engineering Institute, Xi'an Jiaotong University, Xi'an, China, and has been the Cheung Kong Professor of Systems Engineering since 1999, and the Dean of the Faculty of Electronic and Information Engineering since 2008. He has served as the Head of the Department of Automation from 2003 to 2008 at Tsinghua University. His research interests include the optimization of power and energy systems, electric power markets, allocation and scheduling of complex networked resources, network security, and sensor networks