ROBUST COLLABORATIVE STATE ESTIMATION FOR SMART GRID MONITORING

Xiao Li and Anna Scaglione

Dept. of Electrical and Computer Engineering, University of California, Davis, CA, USA

ABSTRACT

This paper proposes a decentralized state estimation scheme via network gossiping with applications in smart grid wide-area monitoring. The proposed scheme allows distributed control areas to solve for an accurate global state estimate collaboratively using the proposed Gossip-based Gauss-Newton (GGN) algorithm. Furthermore, the proposed scheme mitigates the influence of bad data by adaptively updating the noise variances and re-weighting the contributions of the most recent measurements for state estimation. Compared with other distributed techniques, our scheme via gossiping is more flexible and resilient in case of network reconfigurations and failures. We further prove that the power flow equations satisfy the sufficient condition for the GGN algorithm to converge to the desired solution. Simulations of the IEEE-118 system show that the proposed scheme estimates and tracks the global state robustly, and degrades gracefully when there are random failures and bad data.

Index Terms— hybrid state estimation, convergence, gossiping

1. INTRODUCTION

Traditionally, power system state estimation (PSSE) has been solved via the Gauss-Newton (GN) procedure (or its variants) by fitting the measurements from Supervisory Control and Data Acquisition (SCADA) systems [1] to the power flow equations. Recently, measurements from Phasor Measurement Units (PMU), deployed in the Wide-Area Measurement Systems (WAMS), are gaining increasing attention since the state estimator becomes linear [2]. Due to the limited deployment of PMUs, researchers have proposed hybrid estimation schemes [3, 4] by integrating WAMS and SCADA data.

Many works in literature [2, 5–11] have further considered distributing the state estimation proccedures across distributed areas. Each area obtains estimates of the local state relying on redundant measurements that guarantee local observability, and then refines them hierarchically by tuning the estimates on buses shared with neighboring areas. Although these methods alleviate the burdens at control centers, they depend on *aggregation trees* that limit their flexibility in network reconfiguration during failures or attacks.

Recently, [12, 13] proposed distributed schemes that do not require local observability. Specifically, [12] follows a similar formulation in [7–9] and uses the *alternating direction method of multipliers* (AD-MoM) to distribute the algorithm. However, its information exchange model is constrained by the grid topology, which is also inflexible for network reconfiguration. Furthermore, the simulations in [12] are based on PMU data and the algorithm convergence in the presence of SCADA measurements is not discussed. While the merits of hybrid decentralized schemes are evident from [3, 4], these papers do not provide an analytical proof of convergence. Inspired by network diffusion algorithms, the approach in [13] solves for the global state in a fully distributed manner without relying on hierarchical aggregation. These algorithms have been developed for linear filtering problems [14], convex optimizations [15] and adaptive estimation [16], which combine a local descent step with a diffusion step via *gossiping*. The convergence of these algorithms depends on the convexity of the cost function, and more importantly a small (or diminishing) step-size which considerably slows down the algorithm. Moreover, it is not clear how they perform in practice because PSSE using hybrid measurements is non-convex.

Compared to [13, 17], another major issue addressed in this paper is bad data processing. There has been extensive work on χ^2-test or largest normalized residual (LNR) test [1, 18–21] for bad data detection. These methods repeatedly remove or compensate the identified bad data after iterative re-estimation [9, 22], where a distributed scheme is later developed in [23]. A more comprehensive review on bad data processing techniques can be found in [24, 25].

In this paper, we develop a *Decentralized Adaptive Re-weighted State Estimation* (DARSE) scheme using the Gossip-based Gauss-Newton (GGN) algorithm we proposed in [26], and prove that the power flow equations satisfy the sufficient convergence condition for the GGN algorithm. For bad data processing, the DARSE scheme follows a similar approach in [27, 28] by adjusting the weights for measurements based on its quality to reduce their impacts on the state estimates. We show that the DARSE scheme is consistent with the maximum likelihood (ML) framework, where the noise variances are unknown and adaptively updated. Finally, simulations of the IEEE-118 bus system corroborate our claims on the estimation and tracking performance of the DARSE scheme, as well as its robustness to bad data injection.

2. SYSTEM MODEL

We consider a power grid with N *buses* (i.e., substations) that represent interconnections, generators or loads. Denote the bus set $\mathcal{N} \triangleq \{1, \cdots, N\}$ and the edge set $\mathcal{E} \triangleq \{(n, m)\}$, where $\{(n, m)\}$ indicates the transmission line between n and m. Let the cardinality of the edge set be $|\mathcal{E}| = L$. We define $\mathcal{N}(n) \triangleq \{m : (n, m) \in \mathcal{E}\}$ as the neighbor of bus n and let $L_n = |\mathcal{N}(n)|$. The Energy Management Systems (EMS) at control centers collect measurements on certain buses and transmission lines to estimate the state of the power system, i.e., the voltage phasor $V_n = |V_n|e^{\mathrm{i}\theta_n} \in \mathbb{C}$ at each bus $n \in \mathcal{I}$ with $\theta_n = \angle V_n$. Instead of (V_n, θ_n), we consider the Cartesian coordinates $\mathbf{v} = [\Re\{V_1\}, \cdots, \Re\{V_N\}, \Im\{V_1\}, \cdots, \Im\{V_N\}]^T$.

2.1. Measurement Model

Since there are 2 complex injection measurements at each bus and 4 complex flow measurements on each line, this amounts to twice as many real variables. Thus the measurement ensemble has $M = 4N+8L$ entries in an aggregate vector partitioned into four sections[1] $\mathbf{z}[t] = [\mathbf{z}_{\mathcal{V}}^T[t], \mathbf{z}_{\mathcal{C}}^T[t], \mathbf{z}_{\mathcal{I}}^T[t], \mathbf{z}_{\mathcal{F}}^T[t]]^T$, containing the length-$2N$ voltage phasor $\mathbf{z}_{\mathcal{V}}[t]$ and power injection vector $\mathbf{z}_{\mathcal{I}}[t]$ at bus $n \in \mathcal{N}$, the length-$4L$ current phasor $\mathbf{z}_{\mathcal{C}}[t]$ and power flow vector $\mathbf{z}_{\mathcal{F}}[t]$ on line $(n, m) \in \mathcal{E}$ at bus n. Defining the power flow equations $\mathbf{f}_{(\cdot)}(\mathbf{v})$ in Appendix A and letting $\bar{\mathbf{v}}[t]$ be the true state at time t, the individual vector $\mathbf{z}_{(\cdot)}[t] = \mathbf{f}_{(\cdot)}(\bar{\mathbf{v}}[t]) + \mathbf{r}_{(\cdot)}[t]$ contains observations corrupted

[1] Subscripts $\{\mathcal{V}, \mathcal{C}, \mathcal{I}, \mathcal{F}\}$ mean voltage, current, injection and flow.

by measurement noise $\mathbf{r}_{(\cdot)}[t]$ that arises from instrumentation imprecision and random outliers whose variances are potentially much larger due to attacks or equipment malfunction. The entries that have large variances are what we call *bad data*. Then we have

$$\mathbf{z}[t] = \mathbf{f}(\bar{\mathbf{v}}[t]) + \mathbf{r}[t], \tag{1}$$

where $\mathbf{r}[t] = [\mathbf{r}_{\mathcal{V}}^T[t], \mathbf{r}_{\mathcal{C}}^T[t], \mathbf{r}_{\mathcal{I}}^T[t], \mathbf{r}_{\mathcal{F}}^T[t]]^T$ is the aggregate noise vector and $\mathbf{f}(\mathbf{v}) = [\mathbf{f}_{\mathcal{V}}^T(\mathbf{v}), \mathbf{f}_{\mathcal{C}}^T(\mathbf{v}), \mathbf{f}_{\mathcal{I}}^T(\mathbf{v}), \mathbf{f}_{\mathcal{F}}^T(\mathbf{v})]^T$.

A practical data collection architecture in power systems (compatible with WAMS and SCADA) consists of I interconnected *areas*, where each *area* records a subset of $\mathbf{z}$ in (1). We apply a binary matrix $\mathbf{T}_{i,(\cdot)} \in \{0,1\}^{M_{i,(\cdot)} \times M}$ on $\mathbf{z}$ to obtain a length-$M_{i,(\cdot)}$ vector to pick the corresponding element in each category from $\{\mathcal{V}, \mathcal{C}, \mathcal{I}, \mathcal{F}\}$. Letting $\mathbf{T}_i \triangleq [\mathbf{T}_{i,\mathcal{V}}^T, \mathbf{T}_{i,\mathcal{C}}^T, \mathbf{T}_{i,\mathcal{I}}^T, \mathbf{T}_{i,\mathcal{F}}^T]^T$ be the selection matrix in the i-th area, then there are a total of $M_i = M_{i,\mathcal{V}} + M_{i,\mathcal{C}} + M_{i,\mathcal{I}} + M_{i,\mathcal{F}}$ measurements selected as

$$\mathbf{c}_i[t] = \mathbf{f}_i(\bar{\mathbf{v}}[t]) + \mathbf{r}_i[t], \tag{2}$$

where $\mathbf{c}_i[t] \triangleq \mathbf{T}_i\mathbf{z}[t] = [\mathbf{c}_{i,\mathcal{V}}^T[t], \mathbf{c}_{i,\mathcal{C}}^T[t], \mathbf{c}_{i,\mathcal{I}}^T[t], \mathbf{c}_{i,\mathcal{F}}^T[t]]^T$ and similarly $\mathbf{f}_i(\cdot) = \mathbf{T}_i\mathbf{f}(\cdot)$, $\mathbf{r}_i[t] = \mathbf{T}_i\mathbf{r}[t]$. We assume that $\mathbf{r}_i[t]$'s are Gaussian and uncorrelated between different areas, which has an unknown covariance denoted by $\mathbf{R}_i[t] \triangleq \mathrm{diag}[\epsilon_{i,1}[t], \cdots, \epsilon_{i,M_i}[t]]$.

2.2. Maximum Likelihood (ML) Estimation Framework

The ML estimate is obtained by maximizing the likelihood function with unknown noise covariances $\{\mathbf{R}_i[t]\}_{i=1}^I$. Using (2), it is equivalent to minimizing the following objective over the state space $\mathbb{V}$

$$\left\{\widehat{\mathbf{v}}[t], \widehat{\mathbf{R}}_i[t]\right\} = \arg\min_{\mathbf{v}\in\mathbb{V}, \mathbf{R}_i} \sum_{i=1}^I \|\mathbf{c}_i[t] - \mathbf{f}_i(\mathbf{v})\|^2_{\mathbf{R}_i^{-1}} + \sum_{i=1}^I \log|\mathbf{R}_i|.$$

Setting its derivatives with respect to $\mathbf{v}$ and $\{\epsilon_{i,m}\}_{m=1}^{M_i}$ for $i = 1, \cdots, I$ to zero, we have the estimates $\widehat{\mathbf{v}}[t]$ and $\widehat{\epsilon}_{i,m}[t]$ as follows

$$\widehat{\mathbf{v}}[t] = \arg\min_{\mathbf{v}\in\mathbb{V}} \sum_{i=1}^I [\mathbf{c}_i[t] - \mathbf{f}_i(\mathbf{v})]^T \widehat{\mathbf{R}}_i^{-1}[t] [\mathbf{c}_i[t] - \mathbf{f}_i(\mathbf{v})] \tag{3}$$

$$\widehat{\mathbf{R}}_i[t] = \mathrm{diag}[\cdots, \widehat{\epsilon}_{i,m}[t], \cdots],\ \widehat{\epsilon}_{i,m}[t] = |c_{i,m}[t] - f_{i,m}(\widehat{\mathbf{v}}[t])|^2,$$

which would require substituting $\widehat{\mathbf{R}}_i[t]$ (with the unknown $\mathbf{v}$) back to (3) to jointly solve for the variances $\widehat{\epsilon}_{i,m}[t]$ and the state $\widehat{\mathbf{v}}[t]$. However, this approach is highly non-linear and requires a considerable amount of computations. Thus, in the following, we take advantage of the fact that measurements are streaming, to switch adaptively between the estimate of the state and the estimate of the variances.

3. DECENTRALIZED STATE ESTIMATION

Our objective is to harness the computation capabilities in each area to perform state estimation in Section 2.2 online in a decentralized fashion. Note that each area estimates *the global state*, rather than the portion that pertains to its local facilities. If the noise covariance is known, the state (3) can be obtained directly from conventional PSSE [1]. Therefore, we propose to use the previous covariance estimate as a substitute[2] of $\widehat{\mathbf{R}}_i[t]$ to re-weight the measurements in the current snapshot, and propose the *Decentralized Adaptive Reweighted State Estimation* (DARSE) scheme in Algorithm[3] 1.

[2]In general, a better substitute can be predicted using the temporal statistics of the random process $\mathbf{r}[t]$, but here we simply use the previous estimate.

[3]If desired, one can iterate once again the state estimation after the outlier covariance has been updated to give a better state.

Algorithm 1 DARSE Scheme

1: Predict outlier covariance $\mathbf{\Gamma}_i = \widehat{\mathbf{R}}_i[t-1]$, $i = 1, \cdots, I$

2: Update state estimates collaboratively

$$\widehat{\mathbf{v}}[t] = \arg\min_{\mathbf{v}\in\mathbb{V}} \sum_{i=1}^I [\mathbf{c}_i[t] - \mathbf{f}_i(\mathbf{v})]^T \mathbf{\Gamma}_i^{-1} [\mathbf{c}_i[t] - \mathbf{f}_i(\mathbf{v})]$$

3: Adjust outlier covariance $\widehat{\mathbf{R}}_i[t] = \mathrm{diag}[\widehat{\epsilon}_{i,1}[t], \cdots, \widehat{\epsilon}_{i,M_i}[t]]$

$$\widehat{\epsilon}_{i,m}[t] = |c_{i,m}[t] - f_{i,m}(\widehat{\mathbf{v}}[t])|^2. \tag{9}$$

Since *step (1)* and *step (3)* in Algorithm 1 are decoupled between different areas, their decentralized implementations are straightforward. Now we omit the time index t and focus on solving *step (2)*

$$\widehat{\mathbf{v}} = \arg\min_{\mathbf{v}\in\mathbb{V}} \sum_{i=1}^I \|\tilde{\mathbf{c}}_i - \tilde{\mathbf{f}}_i(\mathbf{v})\|^2, \tag{4}$$

where $\tilde{\mathbf{f}}_i(\mathbf{v}) = \mathbf{\Gamma}_i^{-\frac{1}{2}}\mathbf{f}_i(\mathbf{v})$ and $\tilde{\mathbf{c}}_i = \mathbf{\Gamma}_i^{-\frac{1}{2}}\mathbf{c}_i$. Note that we propose to solve (4) in a decentralized setting, where each area has a local estimate $\mathbf{v}_i^k$ that is in consensus with other agents $i' \neq i$ and asymptotically converges to the global estimate $\widehat{\mathbf{v}}$. The global estimate $\widehat{\mathbf{v}}$ is traditionally obtained by the Gauss-Newton (GN) algorithm

$$\mathbf{v}_i^{k+1} = P_{\mathbb{V}}\left[\mathbf{v}_i^k + \mathbf{d}_i^k\right], \quad \mathbf{d}_i^k = \mathbf{Q}^{-1}(\mathbf{v}_i^k)\mathbf{q}(\mathbf{v}_i^k), \tag{5}$$

where $P_{\mathbb{V}}(\cdot)$ is a projection on the space $\mathbb{V}$, $\mathbf{q}(\mathbf{v}_i^k)$ and $\mathbf{Q}(\mathbf{v}_i^k)$ are scaled gradients and GN Hessian of the cost function

$$\mathbf{q}(\mathbf{v}_i^k) = \frac{1}{I}\sum_{p=1}^I \tilde{\mathbf{F}}_p^T(\mathbf{v}_i^k)(\tilde{\mathbf{c}}_p - \tilde{\mathbf{f}}_p(\mathbf{v}_i^k)) \tag{6}$$

$$\mathbf{Q}(\mathbf{v}_i^k) = \frac{1}{I}\sum_{p=1}^I \tilde{\mathbf{F}}_p^T(\mathbf{v}_i^k)\tilde{\mathbf{F}}_p(\mathbf{v}_i^k),$$

with $\tilde{\mathbf{F}}_i(\mathbf{v}) = \mathbf{\Gamma}_i^{-\frac{1}{2}}\mathbf{T}_i\mathrm{d}\mathbf{f}(\mathbf{v})/\mathrm{d}\mathbf{v}^T = \mathbf{\Gamma}_i^{-\frac{1}{2}}\mathbf{T}_i\mathbf{F}(\mathbf{v})$ computed from $\mathbf{F}(\mathbf{v})$ in Appendix A. However, each area knows only its own estimate $\mathbf{v}_i^k$ and part of the measurements/functions in (6), which makes it impossible to implement *step (2)* in a decentralized setting. Therefore, we use the "network average" of different areas as surrogates of $\mathbf{q}(\mathbf{v}_i^k)$ and $\mathbf{Q}(\mathbf{v}_i^k)$, which can be obtained via gossiping

$$\bar{\mathbf{h}}_k = \frac{1}{I}\sum_{i=1}^I \tilde{\mathbf{F}}_i^T(\mathbf{v}_i^k)\left(\tilde{\mathbf{c}}_i - \tilde{\mathbf{f}}_i(\mathbf{v}_i^k)\right) \tag{7}$$

$$\bar{\mathbf{H}}_k = \frac{1}{I}\sum_{i=1}^I \tilde{\mathbf{F}}_i^T(\mathbf{v}_i^k)\tilde{\mathbf{F}}_i(\mathbf{v}_i^k). \tag{8}$$

The enabling technique we propose here for solving *step (2)* is the Gossip-based Gauss-Newton (GGN) algorithm we developed in [26], which emulates the exact GN update in (5) by averaging via near-neighbor communications. Next we describe the GGN algorithm to make the paper self-contained.

3.1. Gossip-based Gauss-Newton (GGN) Algorithm

We interchangeably use *area* and *agent* as the entity for communications and computations. The GGN algorithm runs on two time scales. One is the GN *update* denoted by "k", and the other is the gossip *exchange* denoted by "ℓ" between every two GN updates. All

agents have a common clock that determines the time $t = \tau_k$ for the k-th update across the network. Then the agents exchange information via gossiping at $\tau_{k,\ell} \in [\tau_k, \tau_{k+1})$ for $\ell = 1, \cdots, \ell_k$.

After the k-th update by each agent at time τ_k, the network enters gossip exchange stage $[\tau_k, \tau_{k+1})$ to compute the surrogates $\bar{\mathbf{h}}_k$ and $\bar{\mathbf{H}}_k$ in (7). Each agent combines the information from its neighbors with a weight matrix $\mathbf{W}_k(\ell) \triangleq [W_{ij}^k(\ell)]_{I\times I}$ during $[\tau_{k,\ell}, \tau_{k,\ell+1})$, where $W_{ij}^k(\ell)$ is the weight associated to the edge $\{i,j\}$, which is non-zero if and only if $\{i,j\} \in \mathcal{M}_{k,\ell}$. As an example, we exploit the Uncoordinated Random Exchange (URE) protocol often discussed in literature [29–32], where an agent i chooses a neighbor agent j to communicate during $[\tau_k, \tau_{k+1})$. The exchanges are pairwise and local [29]. Suppose agent $I_{k,\ell}$ wakes up at $\tau_{k,\ell} \in [\tau_k, \tau_{k+1})$ and picks $J_{k,\ell}$ to communicate. Given $0 < \beta < 1$, the weight matrix is $\mathbf{W}_k(\ell) = \mathbf{I} - \beta(\mathbf{e}_{I_{k,\ell}} + \mathbf{e}_{J_{k,\ell}})(\mathbf{e}_{I_{k,\ell}} + \mathbf{e}_{J_{k,\ell}})^T$, where $\mathbf{e}_n$ is the canonical basis with 1 on the n-th entry and 0 otherwise.

Define local information vector at the i-th agent for the ℓ-th gossip

$$\boldsymbol{\mathcal{H}}_{k,i}(\ell) = \begin{bmatrix} \mathbf{h}_{k,i}(\ell) \\ \mathrm{vec}\left[\mathbf{H}_{k,i}(\ell)\right] \end{bmatrix}, \tag{10}$$

evolving from initial conditions $\mathbf{h}_{k,i}(0) \triangleq \tilde{\mathbf{F}}_i^T(\mathbf{v}_i^k)(\tilde{\mathbf{c}}_i - \tilde{\mathbf{f}}_i(\mathbf{v}_i^k))$ and $\mathbf{H}_{k,i}(0) \triangleq \tilde{\mathbf{F}}_i^T(\mathbf{v}_i^k)\tilde{\mathbf{F}}_i(\mathbf{v}_i^k)$. The i-th agent mixes the local information with its neighbors as

$$\boldsymbol{\mathcal{H}}_{k,i}(\ell+1) = W_{ii}^k(\ell)\boldsymbol{\mathcal{H}}_{k,i}(\ell) + \sum_{j\neq i} W_{ij}^k(\ell)\boldsymbol{\mathcal{H}}_{k,j}(\ell) \tag{11}$$

for all $i = 1, \cdots, I$. After ℓ_k exchanges, the local GGN descent for the $(k+1)$-th update at the i-th agent is performed as

$$\mathbf{v}_i^{k+1} = P_{\mathbb{V}}\left[\mathbf{v}_i^k + \mathbf{d}_i^k(\ell_k)\right], \;\; \mathbf{d}_i^k(\ell_k) = \mathbf{H}_{k,i}^{-1}(\ell_k)\mathbf{h}_{k,i}(\ell_k). \tag{12}$$

Remark: It is simulated in [2] that the overall delay from substations in a IEEE-14 bus system to the control center is around 2ms with bandwidth 100-1000 Mbits/s. Thus we bound the worst case hop delay by discounting it with the network diameter $2/7 \approx 0.6$ms. We assume that the state estimation here is performed every 10 seconds rather than today's periodicity (minutes) [2]. If information is stored with 64-bits per entry, the data packets sent by each agent per exchange has $64(2N + 4N^2)$-bits. For a power system with $N = 118$ buses with a communication bandwidth 100 Mbits/s, the maximum number of exchange that can be accomodated in 10 seconds is $10\times10^8/(0.6\times10^{-3}\times10^8+64(2\times118+4\times118^2)) \approx 300$, which is far more than 10 exchanges as examined in simulations.

3.2. Convergence Analysis

Condition 1. *First, we assume the following conditions:*

(1) The state space $\mathbb{V}$ is closed and convex.

(2) The cost $\sum_{i=1}^I \|\tilde{\mathbf{c}}_i - \tilde{\mathbf{f}}_i(\mathbf{v})\|$ is bounded for all $\mathbf{v} \in \mathbb{V}$.

(3) Denote by $\lambda_{\min}(\cdot)$ the minimum eigenvalues and let

$$\sigma_F = \min_{\mathbf{v}\in\mathbb{V}} \sqrt{\lambda_{\min}\left(\sum_{i=1}^I \tilde{\mathbf{F}}_i^T(\mathbf{v})\tilde{\mathbf{F}}_i(\mathbf{v})\right)} > 0.$$

(4) There exist finite ℓ_k's such that $\|\mathbf{d}_i^k(\ell_k) - \mathbf{d}_i^k\| \leq \kappa$ for $\forall i, k$.

Conditions 1-(2) and (3) are satisfied if the system is observable [33] and the noise is finite. As shown in [26, Prop. 1], condition 2-(4) holds as long as ℓ_k's are chosen properly, and all agents are initialized with consensus estimates $\mathbf{v}_1^0 = \cdots = \mathbf{v}_I^0$. Furthermore, we prove that the Jacobians $\{\tilde{\mathbf{F}}_i(\mathbf{v})\}_{i=1}^I$ to satisfy the Lipschitz condition, which is important for the convergence analysis.

Lemma 1. *The Jacobian matrix $\tilde{\mathbf{F}}_i(\mathbf{v})$ satisfies the Lipschitz condition for all i and arbitrary $\mathbf{v}, \mathbf{v}' \in \mathbb{V}$*

$$\left\|\tilde{\mathbf{F}}_i(\mathbf{v}) - \tilde{\mathbf{F}}_i(\mathbf{v}')\right\| \leq \omega \left\|\mathbf{v} - \mathbf{v}'\right\|, \quad \forall i = 1, \cdots, I \tag{13}$$

where ω is the Lipschitz constant given in (27).

Proof. See Appendix B. □

The GGN algorithm is initialized with $\mathbf{v}_i^0$ at each agent and continues until a stopping criterion is met. Since (4) is a non-linear least squares (NLLS) problem that is non-convex, the iterate $\mathbf{v}_i^k$ may stop at any fixed point $\mathbf{v}^\star$ of (5) satisfying the first order condition

$$\sum_{i=1}^I \tilde{\mathbf{F}}_i^T(\mathbf{v}^\star)\left(\tilde{\mathbf{c}}_i - \tilde{\mathbf{f}}_i(\mathbf{v}^\star)\right) = \mathbf{0}. \tag{14}$$

Clearly, the ML estimate $\widehat{\mathbf{v}}$ in (4) is one of the fixed points and it is desirable to have the algorithm converge to this point. The analysis presented in our previous paper [26] is particularly useful here after we have proven Lemma 1 specifically for power systems. The resultant theorem from [26] can be re-stated as follows.

Theorem 1. *[26, Lemma 1, 2, 3 & Theorem 1] Given Lemma 1 under Condition 1, the error between $\mathbf{v}_i^k$ in* (12) *and $\widehat{\mathbf{v}}$ in* (4) *satisfies*

$$\left\|\mathbf{v}_i^{k+1} - \widehat{\mathbf{v}}\right\| \leq T_1 \left\|\mathbf{v}_i^k - \widehat{\mathbf{v}}\right\|^2 + T_2 \left\|\mathbf{v}_i^k - \widehat{\mathbf{v}}\right\| + \kappa, \tag{15}$$

where $T_1 \triangleq \omega/2\sigma_F$, $T_2 \triangleq \sqrt{2}\omega\tilde{\epsilon}/\sigma_F^2$, $\tilde{\epsilon} = \sum_{i=1}^I \|\tilde{\mathbf{c}}_i - \tilde{\mathbf{f}}_i(\widehat{\mathbf{v}})\|$. If $T_2 \ll 1$ and $\kappa \leq (1-T_2)^2/4T_1$ with $\mathbf{v}_1^0 = \cdots = \mathbf{v}_I^0$, then for any $\mathbf{v}_i^0$ satisfying $\|\mathbf{v}_i^0 - \widehat{\mathbf{v}}\| < 2\sigma_F/\omega - \kappa$, the iterative error is asymptotically bounded by $\limsup_{k\to\infty} \left\|\mathbf{v}_i^{k+1} - \widehat{\mathbf{v}}\right\| \leq \kappa$.

Theorem 1 implies that all the agents will converge to an arbitrarily small neighborhood of the ML estimate $\widehat{\mathbf{v}}$ if each area is initialized sufficiently close. In fact, an effective initializer is the re-scaled average of the voltage phasor measurements $\mathbf{c}_{i,\mathcal{V}}$ of all areas because it povides direct state measurements accurately. We do not elaborate on the details here because it is straightforward.

4. NUMERICAL RESULTS

We illustrate the Mean Square Error (MSE) performance of the DARSE scheme, where the MSE is the averaged over all agents $\mathrm{MSE}_V = \sum_{i=1}^I \mathbb{E}\left\|\mathbf{v}_i^k - \bar{\mathbf{v}}\right\|^2/I$ for each iteration k. We used MATPOWER 4.0 to simulate the IEEE-118 system with 3 snapshots and perform DARSE over 100 runs. We divide the system into $I = 10$ areas where 9 areas have 12 buses in each and 1 area has the remaining 10 buses, all randomly drawn from 1 to 118 without repetition. For each area, we randomly choose 50% of the SCADA measurements and exploit the 36 PMU data in Area 1, 2 and 3. In the 3 snapshots, the measurements are corrupted by Gaussian errors $r_{i,m}[t] \sim \mathcal{N}(0, \sigma^2)$ with $\sigma = 10^{-2}$, among which 10 data entries are chosen at random to inject outliers with variances $30\sigma^2$.

We compare the DARSE scheme with the centralized GN procedure with and without bad data in Fig.1 in terms of MSE against the iteration $k = 1, \cdots, 20$ for each snapshot (i.e., 60 updates for 3 snapshots). Between every two updates, each agent proceeds under the URE protocol and talks to each agent 10 times on average such that the network traffic is comparable to the centralized scheme as if

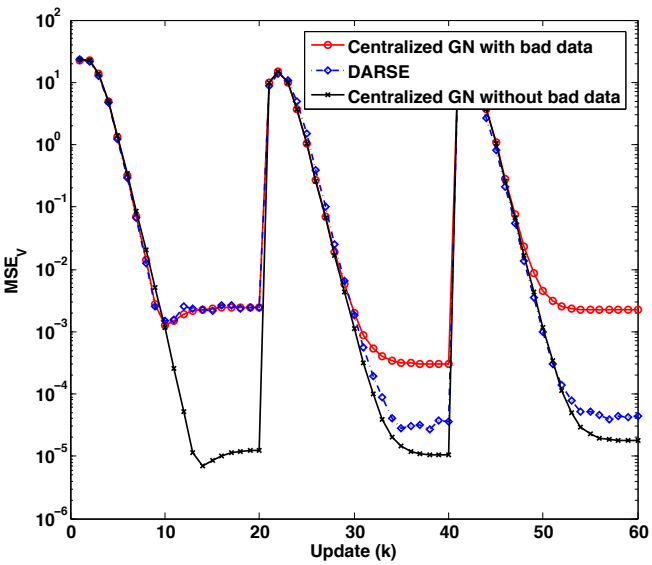

Fig. 1. Performance of the DARSE scheme (blue) against the centralized GN approach with (red) and without (black) bad data

the local measurements were routed through the network. We also impose link failures between agents in the DARSE scheme, where any established link $\{i, j\}$ fails with probability 0.1 independently.

The centralized GN algorithm which processes perfect data (black dotted line) serves as the ultimate performance benchmark. It is seen that DARSE scheme only has a small performance loss compared with the centralized GN with perfect data. When there is no link failures, the DARSE scheme approximately touches the benchmark, but we did not present it due to lack of space. When bad data are present, thanks to the re-weighted bad data suppression, the DARSE scheme (blue dotted line) outperforms significantly the centralized GN approach without re-weighting procedure (red dotted line).

5. CONCLUSIONS

We proposed a DARSE scheme for hybrid power system state estimation integrating WAMS and SCADA system, which adaptively estimates the global state vector along with an updated noise covariance. The simulation shows that the DARSE scheme is able to deliver accurate estimates of the global state at each distributed area, even in the presence of bad data and random link failures.

A. POWER FLOW EQUATIONS AND JACOBIAN MATRIX

The matrix $\mathbf{Y} = [-Y_{nm}]$ includes line admittances $Y_{nm} = G_{nm} + \mathrm{i}B_{nm}$, $(n, m) \in \mathcal{E}$, shunt admittances $\bar{Y}_{nm} = \bar{G}_{nm} + \mathrm{i}\bar{B}_{nm}$ of the line $(n, m) \in \mathcal{E}$, and self-admittance $Y_{nn} = -\sum_{m \neq n}(\bar{Y}_{nm} + Y_{nm})$. Using $\mathbf{e}_n = [0, \cdots, 1, \cdots, 0]^T$, we define the following

$$\mathbf{Y}_n \triangleq \mathbf{e}_n\mathbf{e}_n^T\mathbf{Y}, \ \mathbf{Y}_{nm} \triangleq (Y_{nm} + \bar{Y}_{nm})\mathbf{e}_n\mathbf{e}_n^T - Y_{nm}\mathbf{e}_n\mathbf{e}_l^T. \quad (16)$$

Letting $\mathbf{G}_n = \Re\{\mathbf{Y}_n\}$, $\mathbf{B}_n = \Im\{\mathbf{Y}_n\}$, $\mathbf{G}_{nm} = \Re\{\mathbf{Y}_{nm}\}$ and $\mathbf{B}_{nm} = \Im\{\mathbf{Y}_{nm}\}$, we further define the following matrices

$$\mathbf{N}_{P,n} \triangleq \begin{bmatrix} \mathbf{G}_n & -\mathbf{B}_n \\ \mathbf{B}_n & \mathbf{G}_n \end{bmatrix} \qquad \mathbf{N}_{Q,n} \triangleq -\begin{bmatrix} \mathbf{B}_n & \mathbf{G}_n \\ -\mathbf{G}_n & \mathbf{B}_n \end{bmatrix}$$

$$\mathbf{E}_{P,nm} \triangleq \begin{bmatrix} \mathbf{G}_{nm} & -\mathbf{B}_{nm} \\ \mathbf{B}_{nm} & \mathbf{G}_{nm} \end{bmatrix} \ \mathbf{E}_{Q,nm} \triangleq -\begin{bmatrix} \mathbf{B}_{nm} & \mathbf{G}_{nm} \\ -\mathbf{G}_{nm} & \mathbf{B}_{nm} \end{bmatrix}$$

$$\mathbf{C}_{I,nm} \triangleq \begin{bmatrix} \mathbf{G}_{nm} & \mathbf{0} \\ \mathbf{0} & -\mathbf{B}_{nm} \end{bmatrix} \ \mathbf{C}_{J,nm} \triangleq \begin{bmatrix} \mathbf{B}_{nm} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{nm} \end{bmatrix}.$$

The SCADA systems collect the active/reactive power injection (P_n, Q_n) at bus n and flow (P_{nm}, Q_{nm}) at bus n from line (n, m)

$$P_n = \mathbf{v}^T\mathbf{N}_{P,n}\mathbf{v}, \quad Q_n = \mathbf{v}^T\mathbf{N}_{Q,n}\mathbf{v}, \quad (17)$$

$$P_{nm} = \mathbf{v}^T\mathbf{E}_{P,nm}\mathbf{v}, \quad Q_{nm} = \mathbf{v}^T\mathbf{E}_{Q,nm}\mathbf{v}. \quad (18)$$

We stack these functions in the power flow equation vectors

$$\mathbf{f}_{\mathcal{I}}(\mathbf{v}) = [\cdots, P_n, \cdots, \cdots, Q_n, \cdots]^T \quad (19)$$

$$\mathbf{f}_{\mathcal{F}}(\mathbf{v}) = [\cdots, P_{nm}, \cdots, \cdots, Q_{nm}, \cdots]^T. \quad (20)$$

The WAMS collects the voltage phasor $(\Re\{V_n\}, \Im\{V_n\})$ at bus n and the current phasor (I_{nm}, J_{nm}) at bus n on line (n, m)

$$I_{nm} = (\mathbf{1}_2 \otimes \mathbf{e}_n)^T \mathbf{C}_{I,nm}\mathbf{v}, \ J_{nm} = (\mathbf{1}_2 \otimes \mathbf{e}_n)^T \mathbf{C}_{J,nm}\mathbf{v}, \quad (21)$$

where $\otimes$ is the Kronecker product, and stacks them as

$$\mathbf{f}_{\mathcal{V}}(\mathbf{v}) = \mathbf{v}, \quad \mathbf{f}_{\mathcal{C}}(\mathbf{v}) = [\cdots, I_{nm}, \cdots, \cdots, J_{nm}, \cdots]^T. \quad (22)$$

The Jacobian $\mathbf{F}(\mathbf{v})$ can then be derived from (21), (17) and (18) as

$$\mathbf{F}(\mathbf{v}) = \begin{bmatrix} \mathbf{I}_{2N} & \mathbf{H}_{\mathcal{C}}^T & \mathbf{H}_{\mathcal{I}}^T(\mathbf{I}_{2N} \otimes \mathbf{v}) & \mathbf{H}_{\mathcal{F}}^T(\mathbf{I}_{4L} \otimes \mathbf{v}) \end{bmatrix}^T, \quad (23)$$

where $\mathbf{H}_{\mathcal{I}} \triangleq \left[\cdots, \mathbf{N}_{P,n} + \mathbf{N}_{P,n}^T, \cdots, \mathbf{N}_{Q,n} + \mathbf{N}_{Q,n}^T, \cdots\right]^T$, and $\mathbf{H}_{\mathcal{F}} \triangleq \left[\cdots, \mathbf{E}_{P,nm} + \mathbf{E}_{P,nm}^T, \cdots, \mathbf{E}_{Q,nm} + \mathbf{E}_{Q,nm}^T, \cdots\right]^T$ and $\mathbf{H}_{\mathcal{C}} \triangleq \left[\cdots, \mathbf{H}_{I,n}^T, \cdots, \cdots, \mathbf{H}_{J,n}^T, \cdots\right]^T$, $\mathbf{H}_{I,n} \triangleq \mathbf{S}_n\mathbf{C}_{I,n}$ with $\mathbf{C}_{I,n} \triangleq \left[\cdots, \mathbf{C}_{I,nm}^T, \cdots\right]^T$ and $\mathbf{H}_{J,n} \triangleq \mathbf{S}_n\mathbf{C}_{J,n}$ with $\mathbf{C}_{J,n} \triangleq \left[\cdots, \mathbf{C}_{J,nm}^T, \cdots\right]^T$ using $\mathbf{S}_n \triangleq \mathbf{I}_{L_n} \otimes (\mathbf{1}_2 \otimes \mathbf{e}_n)^T$.

B. PROOF OF LEMMA 1

The F-norm inequality $\|\cdot\| \leq \|\cdot\|_F$ gives us $\|\tilde{\mathbf{F}}_i(\mathbf{v}) - \tilde{\mathbf{F}}_i(\mathbf{v}')\|^2 \leq \|\tilde{\mathbf{F}}_i(\mathbf{v}) - \tilde{\mathbf{F}}_i(\mathbf{v}')\|_F^2$ for all i. Since $\tilde{\mathbf{F}}_i(\mathbf{v}) = \mathbf{\Gamma}_i^{-1/2}\mathbf{T}_i\mathbf{F}(\mathbf{v})$, we further use the multiplicative norm inequality $\|\mathbf{AB}\|_F \leq \|\mathbf{A}\|_F \|\mathbf{B}\|_F$

$$\left\|\tilde{\mathbf{F}}_i(\mathbf{v}) - \tilde{\mathbf{F}}_i(\mathbf{v}')\right\|_F^2 = \left\|\mathbf{\Gamma}_i^{-1/2}\mathbf{T}_i\left[\mathbf{F}(\mathbf{v}) - \mathbf{F}(\mathbf{v}')\right]\right\|_F^2 \quad (24)$$

$$\leq \lambda_{\min}^{-1}(\mathbf{\Gamma}_i)\left\|\mathbf{F}(\mathbf{v}) - \mathbf{F}(\mathbf{v}')\right\|_F^2. \quad (25)$$

From (23), we have

$$\mathbf{F}(\mathbf{v}) - \mathbf{F}(\mathbf{v}') = \begin{bmatrix} \mathbf{0}_{2N \times 2N} \\ \mathbf{0}_{4L \times 2N} \\ \left[\mathbf{I}_{2N} \otimes (\mathbf{v} - \mathbf{v}')^T\right]\mathbf{H}_{\mathcal{I}} \\ \left[\mathbf{I}_{4L} \otimes (\mathbf{v} - \mathbf{v}')^T\right]\mathbf{H}_{\mathcal{F}} \end{bmatrix}. \quad (26)$$

According to the F-norm definition $\|\mathbf{A}\|_F^2 = \mathrm{Tr}\left(\mathbf{A}^T\mathbf{A}\right)$ and the properties of the trace operator, we have for any $\mathbf{v}, \mathbf{v}'$

$$\left\|\mathbf{F}(\mathbf{v}) - \mathbf{F}(\mathbf{v}')\right\|_F^2 = \mathrm{Tr}\left[\left(\mathbf{I}_{2N} \otimes (\mathbf{v} - \mathbf{v}')(\mathbf{v} - \mathbf{v}')^T\right)\mathbf{H}_{\mathcal{I}}\mathbf{H}_{\mathcal{I}}^T\right] + \mathrm{Tr}\left[\left(\mathbf{I}_{4L} \otimes (\mathbf{v} - \mathbf{v}')(\mathbf{v} - \mathbf{v}')^T\right)\mathbf{H}_{\mathcal{F}}\mathbf{H}_{\mathcal{F}}^T\right].$$

Expanding $\mathbf{H}_{\mathcal{I}}$ and $\mathbf{H}_{\mathcal{F}}$ in (23) and using their symmetric properties, we have $\|\mathbf{F}(\mathbf{v}) - \mathbf{F}(\mathbf{v}')\|_F^2 = (\mathbf{v} - \mathbf{v}')^T\mathbf{M}(\mathbf{v} - \mathbf{v}')$, where $\mathbf{M} = \mathbf{H}_{\mathcal{I}}^T\mathbf{H}_{\mathcal{I}} + \mathbf{H}_{\mathcal{F}}^T\mathbf{H}_{\mathcal{F}}$. It is well-known that any quadratic form of a symmetric matrix is determined by its 2-norm

$$(\mathbf{v} - \mathbf{v}')^T\mathbf{M}(\mathbf{v} - \mathbf{v}') \leq \|\mathbf{M}\|\left\|\mathbf{v} - \mathbf{v}'\right\|^2. \quad (27)$$

The result follows by setting $\omega = \max_i \sqrt{\|\mathbf{M}\|\lambda_{\min}^{-1}(\mathbf{\Gamma}_i)}$.

C. REFERENCES

[1] F. Schweppe and E. Handschin, "Static State Estimation in Electric Power Systems," *Proceedings of the IEEE*, vol. 62, no. 7, pp. 972–982, 1974.

[2] T. Yang, H. Sun, and A. Bose, "Transition to a Two-Level Linear State Estimator : Part I & II," *IEEE Trans. Power Syst.*, no. 99, pp. 1–1, 2011.

[3] A. Phadke, J. Thorp, and K. Karimi, "State Estimation with Phasor Measurements," *Power Engineering Review, IEEE*, no. 2, pp. 48–48, 1986.

[4] R. Zivanovic and C. Cairns, "Implementation of PMU Technology in State Estimation: an Overview," in *AFRICON, 1996., IEEE AFRICON 4th*, vol. 2. IEEE, 1996, pp. 1006–1011.

[5] C. Brice and R. Cavin, "Multiprocessor Static State Estimation," *IEEE Trans. Power App. Syst.*, no. 2, pp. 302–308, 1982.

[6] A. Gómez-Expósito, A. Abur, A. de la Villa Jaén, and C. Gómez-Quiles, "A Multilevel State Estimation Paradigm for Smart Grids," *Proceedings of the IEEE*, no. 99, pp. 1–25, 2011.

[7] D. Falcao, F. Wu, and L. Murphy, "Parallel and Distributed State Estimation," *IEEE Trans. Power Syst.*, vol. 10, no. 2, pp. 724–730, 1995.

[8] S. Lin, "A Distributed State Estimator for Electric Power Systems," *IEEE Trans. Power Syst.*, vol. 7, no. 2, pp. 551–557, 1992.

[9] R. Ebrahimian and R. Baldick, "State Estimation Distributed Processing," *IEEE Trans. Power Syst.*, vol. 15, no. 4, pp. 1240–1246, 2000.

[10] L. Zhao and A. Abur, "Multi-area State Estimation using Synchronized Phasor Measurements," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 611–617, 2005.

[11] W. Jiang, V. Vittal, and G. Heydt, "A Distributed State Estimator Utilizing Synchronized Phasor Measurements," *IEEE Trans. Power Syst.*, vol. 22, no. 2, pp. 563–571, 2007.

[12] V. Kekatos and G. Giannakis, "Distributed Robust Power System State Estimation," *Arxiv preprint arXiv:1204.0991*, 2012.

[13] L. Xie, D. Choi, S. Kar, and H. Poor, "Fully Distributed State Estimation for Wide-Area Monitoring Systems," *Smart Grid, IEEE Transactions on*, vol. 3, no. 3, pp. 1154–1169, 2012.

[14] C. Lopes and A. Sayed, "Diffusion Least-Mean Squares over Adaptive Networks: Formulation and Performance Analysis," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3122–3136, 2008.

[15] A. Nedic and A. Ozdaglar, "Distributed Subgradient Methods for Multi-agent Optimization," *Automatic Control, IEEE Transactions on*, vol. 54, no. 1, pp. 48–61, 2009.

[16] S. Kar, J. Moura, and K. Ramanan, "Distributed Parameter Estimation in Sensor Networks: Nonlinear Observation Models and Imperfect Communication," *Information Theory, IEEE Transactions on*, vol. 58, no. 6, pp. 3575 –3605, june 2012.

[17] X. Li, Z. Wang, and A. Scaglione, "Decentralized Data Processing and Management in Smart Grid via Gossiping," in *Sensor Array and Multichannel Signal Processing Workshop (SAM), 2012 IEEE 7th*. IEEE, 2012, pp. 1–4.

[18] T. Van Cutsem, M. Ribbens-Pavella, and L. Mili, "Bad Data Identification Methods in Power System State Estimation-A Comparative Study," *Power Apparatus and Systems, IEEE Transactions on*, no. 11, pp. 3037–3049, 1985.

[19] A. Garcia, A. Monticelli, and P. Abreu, "Fast Decoupled State Estimation and Bad Data Processing," *Power Apparatus and Systems, IEEE Transactions on*, no. 5, pp. 1645–1652, 1979.

[20] R. Bobba, K. Rogers, Q. Wang, H. Khurana, K. Nahrstedt, and T. Overbye, "Detecting False Data Injection Attacks on DC State Estimation," in *Preprints of the First Workshop on Secure Control Systems, CPSWEEK 2010*, 2010.

[21] Y. Liu, P. Ning, and M. 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.

[22] A. Monticelli, "State Estimation in Electric Power Systems: A Generalized Approach, 1999."

[23] D. Choi and L. Xie, "Fully distributed bad data processing for wide area state estimation," in *Smart Grid Communications (SmartGridComm), 2011 IEEE International Conference on*. IEEE, 2011, pp. 546–551.

[24] K. Lo, P. Ong, R. McColl, A. Moffatt, and J. Sulley, "Development of a Static State Estimator Part i: Estimation and Bad Data Suppression," *Power Apparatus and Systems, IEEE Transactions on*, no. 8, pp. 2486–2491, 1983.

[25] E. Handschin, F. Schweppe, J. Kohlas, and A. Fiechter, "Bad Data Analysis for Power System State Estimation," *Power Apparatus and Systems, IEEE Transactions on*, vol. 94, no. 2, pp. 329–337, 1975.

[26] X. Li and A. Scaglione, "Convergence and Applications of a Gossip-based Gauss-Newton Algorithm," *submitted to IEEE Trans. on Signal Processing, arXiv preprint arXiv:1210.0056*, 2012.

[27] H. Merrill and F. Schweppe, "Bad Data Suppression in Power System Static State Estimation," *Power Apparatus and Systems, IEEE Transactions on*, no. 6, pp. 2718–2725, 1971.

[28] D. Falcão, P. Cooke, and A. Brameller, "Power System Tracking State Estimation and Bad Data Processing," *Power Apparatus and Systems, IEEE Transactions on*, no. 2, pp. 325–333, 1982.

[29] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized Gossip Algorithms," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2508–2530, 2006.

[30] A. Dimakis, S. Kar, J. Moura, M. Rabbat, and A. Scaglione, "Gossip Algorithms for Distributed Signal Processing," *Proc. IEEE*, vol. 98, no. 11, pp. 1847–1864, 2010.

[31] B. Johansson, T. Keviczky, M. Johansson, and K. Johansson, "Subgradient Methods and Consensus Algorithms for Solving Convex Optimization Problems," in *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*. IEEE, 2008, pp. 4185–4190.

[32] J. Tsitsiklis, "Problems in Decentralized Decision Making and Computation." DTIC Document, Tech. Rep., 1984.

[33] K. Clements, "Observability Methods and Optimal Meter Placement," *International Journal of Electrical Power & Energy Systems*, vol. 12, no. 2, pp. 88–93, 1990.