# Advances in Decentralized State Estimation for Power Systems

Xiao Li and Anna Scaglione
Dept. of Electrical and Computer Engineering, University of California, Davis, email: {eceli, ascaglione}@ucdavis.edu

***Abstract*—Distributed learning via network diffusion is a popular trend in signal processing, which addresses the need of obtaining scalable analytics from networked sensor systems. This paper describes relevant advances in distributed power system state estimation (PSSE) via diffusion. Considering a hybrid sensor measurements system, we show that the Gauss-Newton approach, typically favored in PSSE, can be used as a primitive to derive a gossip-based algorithm that outperforms first order diffusion methods proposed in the literature. We also study analytically and numerically the dependency between measurement placement, grid topology and physical parameters, communication network and the performance of the decentralized PSSE.**

## I. Introduction

Conforming to many classic sensor fusion problems, the original formulation of power system state estimation (PSSE) was cast as a non-linear least squares (NLLS) problem [1] after gathering power measurements through the Supervisory Control and Data Acquisition (SCADA) system at a control center. The Gauss-Newton (GN) procedure (or its variants) have been widely adopted to solve for the state numerically. While the system grew in size and connectivity, together with the insertion of Phasor Measurements Units (PMUs) that provide phasor measurements at a fast rate, decentralized solutions have been proposed to alleviate the communication and computation burden at the control center [2]–[4]. In these solutions, different areas estimate their local states using redundant measurements that guarantee local observability, and then refine them hierarchically by tuning the estimates on neighboring buses. These methods typically rely on *aggregation trees* that limit their flexibility.

Recently, the authors in [5], [6] proposed distributed algorithms that relax the requirement of local observability. In particular, [5] used the so called *alternating direction method of multipliers* (AD-MoM) to distribute PSSE algorithms using PMU measurements. However, the communications are constrained by the grid topology and the effects of having hybrid measurements are not studied. A step further towards relaxing network and measurements constraints is the approach proposed in [6], solving for the global state through *gossiping*. The approach is inspired by [7], [8] that combine a local descent step with a diffusion step. Compared to other decentralized methods, gossip-based learning has minimal network requirements. But the convergence of these algorithms depends on the convexity of the cost and a small (or diminishing) step-size which considerably slows down the algorithm. Because PSSE is in general non-convex when SCADA measurements are in the pool (i.e. power flow and injection), convergence is not guaranteed and the initialization of the algorithm is crucial.

Given that the PMUs provide direct state measurements, placing them strategically and exploiting such measurements for initialization can aid significantly the convergence. There is vast literature on PMU placement problem by minimizing the deployment cost under observability constraints (see e.g. [9], [10]) or by optimizing the achievable estimation accuracy [11], [12]. However, having numerically stable decentralized PSSE and being able to remove the influence of bad measurements are two crucial aspects that were overlooked in the studies of PMU placement, until recently. To overcome these issues, we proposed in [13] a *Decentralized Adaptive Re-weighted State Estimation* (DARSE) scheme and furthermore, derived a new metric in [14] called Convergence-Observability-Performance (COP) metric to account for the observability, accuracy, and convergence of PSSE via Gauss-Newton, and used it successfully to ensure the numerical stability of the DARSE algorithm.

Our objective in this paper is to present these results in a combined framework and perform numerical comparisons that showcase how the judicious placement of sensors as well as the use of second order methods contribute to improve significantly the attainable performance in gossip-based PSSE solutions. Hence, future database systems for the power grid can be designed to have a peer to peer randomized network structure, which is more scalable, resilient and self-healing compared to the current aggregation architectures.

## II. State Estimation via Network Gossiping

Optimization via network diffusion is possible if the learning problem can be formulated as a regression over the sum of functions [7], [8]. In this section we first show how the Maximum Likelihood (ML) PSSE can be cast as an instance of such type of regression problem and then introduce a methodology that gradually reduces the influence of outliers, by learning the sensor measurement variances online.

### A. System Model

Considering a power grid with the set $\mathcal{N} \triangleq \{1, \cdots, N\}$ of $N$ *buses* and the set $\mathcal{E} \triangleq \{(n,m)\}$ of size $|\mathcal{E}| = L$ of transmission lines between all bus pairs $(n,m)$, the grid state is the vector of voltage phasors. Instead of the conventional polar coordinates, we resort to the Cartesian coordinates $\mathbf{v} = [\Re\{V_1\}, \cdots, \Re\{V_N\}, \Im\{V_1\}, \cdots, \Im\{V_N\}]^T$ where $V_n$ is the voltage phasor at bus $n$. Measurements for estimation include the voltage phasor $\mathbf{z}_{\mathcal{V}}[t] \in \mathbb{R}^{2N}$ and current phasor $\mathbf{z}_{\mathcal{C}}[t]$ acquired via PMUs, and power injections $\mathbf{z}_{\mathcal{I}}[t] \in \mathbb{R}^{2N}$ at each bus and power flows $\mathbf{z}_{\mathcal{F}}[t]$ on each line acquired by traditional SCADA systems that lack synchronization. The subscripts mean voltage, current, injection and flow. The model for measurement of type $\mathcal{K} \in \mathcal{A} \triangleq \{\mathcal{V}, \mathcal{C}, \mathcal{I}, \mathcal{F}\}$ is

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

where $\mathbf{f}_{\mathcal{K}}(\mathbf{v})$ is the power flow equation vector for type $\mathcal{K}$, $\bar{\mathbf{v}}[t]$ is the true state, and $\mathbf{r}_{\mathcal{K}}[t]$ is the Gaussian noise with an unknown covariance $\mathbf{R}_{\mathcal{K}}$. We model outliers (due to attacks or malfunction) as having large variances. The entries that have large variances are what we call *bad data*.

The data collection architecture consists of $I$ interconnected *areas*, where each *area* records a subset of $\mathbf{z}_{\mathcal{K}}$ in (1). We use binary selection matrices $\mathbf{T}_{\mathcal{K},i} \in \{0,1\}^{M_{\mathcal{K},i} \times 2N}$ for

978-1-4673-3146-3/13/$31.00 ©2013IEEE

nodal measurements $\mathcal{K} \in \{\mathcal{V}, \mathcal{I}\}$ and $\mathbf{T}_{\mathcal{K},i} \in \{0,1\}^{M_{\mathcal{K},i} \times 4L}$ for branch measurements $\mathcal{K} \in \{\mathcal{C}, \mathcal{F}\}$ to model the specific measurements taken in that area's collection system. Then, the vector of measurements for each type in the $i$-th area is

$$\mathbf{T}_{\mathcal{K},i}\mathbf{z}_{\mathcal{K}}[t] = \mathbf{T}_{\mathcal{K},i}\mathbf{f}_{\mathcal{K}}(\bar{\mathbf{v}}[t]) + \mathbf{T}_{\mathcal{K},i}\mathbf{r}_{\mathcal{K}}[t]. \quad (2)$$

Thus, the ML state and covariances estimates are

$$\min_{\mathbf{v}\in\mathbb{V},\mathbf{R}_{\mathcal{K},i}} \sum_{i=1}^{I}\sum_{\mathcal{K}\in\mathcal{A}} \|\mathbf{T}_{\mathcal{K},i}\left(\mathbf{z}_{\mathcal{K}}[t] - \mathbf{f}_{\mathcal{K}}(\mathbf{v})\right)\|^2_{\mathbf{R}^{-1}_{\mathcal{K},i}} + \log\det(\mathbf{R}_{\mathcal{K},i}),$$

where $\mathbf{R}_{\mathcal{K},i} = \mathbf{T}_{\mathcal{K},i}\mathbf{R}_{\mathcal{K}}\mathbf{T}^T_{\mathcal{K},i}$ is the unknown covariance in the $i$-th area for type $\mathcal{K}$. For simplicity, the covariance is assumed to be diagonal with $\mathbf{R}_{\mathcal{K},i} = \mathrm{diag}[\boldsymbol{\gamma}_{\mathcal{K},i}]$. Setting its derivatives to zero, the ML estimates satisfy the following equations

$$\widehat{\mathbf{v}} = \arg\min_{\mathbf{v}\in\mathbb{V}} \sum_{i=1}^{I}\sum_{\mathcal{K}\in\mathcal{A}} \|\mathbf{T}_{\mathcal{K},i}\left(\mathbf{z}_{\mathcal{K}}[t] - \mathbf{f}_{\mathcal{K}}(\mathbf{v})\right)\|^2_{\widehat{\mathbf{R}}^{-1}_{\mathcal{K},i}} \quad (3)$$

$$\widehat{\boldsymbol{\gamma}}_{\mathcal{K},i}[t] = \mathbf{T}_{\mathcal{K},i}\left[\mathbf{z}_{\mathcal{K}}[t] - \mathbf{f}_{\mathcal{K}}(\widehat{\mathbf{v}})\right] \odot \mathbf{T}_{\mathcal{K},i}\left[\mathbf{z}_{\mathcal{K}}[t] - \mathbf{f}_{\mathcal{K}}(\widehat{\mathbf{v}})\right] \quad (4)$$

which requires substituting $\widehat{\boldsymbol{\gamma}}_{\mathcal{K},i}[t]$ back into (3) to solve for $\widehat{\mathbf{v}}$. However, this requires complex computations. Hence, the best option is to alternate between estimating the state and the variances as measurements stream in. Since, the ML estimation has the typical modular structure of optimization problems that can be solved by network diffusion, next we show how the *Decentralized Adaptive Re-weighted State Estimation* (DARSE) scheme proposed in [13] takes advantage of this formulation.

### *B. Decentralized Adaptive Re-weighted State Estimation*

The objective of the DARSE diffusion algorithm is to lead each area $i$ to be in consensus with *the global state* estimate. Instead of only accruing information on the state in area $i$ each participating site effectively becomes a *mirror* for information about the state of the entire grid, that can be shared with client applications in full or partially. If the noise covariance is known, the state (3) can be obtained from conventional PSSE [1]. In the DARSE scheme, we propose to use the previous covariance estimate as a substitute of $\widehat{\mathbf{R}}_{\mathcal{K},i}[t]$ to re-weight the measurements and, thus, reduce the impact of highly noisy data on the overall cost[1]. This strategy is an alternative to bad data elimination techniques, used in centralized PSSE [1]. An area involved in the DARSE scheme follows the steps below:

1) Predict $\boldsymbol{\Gamma}_{\mathcal{K},i} = \widehat{\mathbf{R}}_{\mathcal{K},i}[t-1]$, $i = 1, \cdots, I$.
2) Update state estimates collaboratively
$$\widehat{\mathbf{v}}[t] = \arg\min_{\mathbf{v}\in\mathbb{V}} \sum_{i=1}^{I} \|\mathbf{T}_{\mathcal{K},i}\left[\mathbf{z}_{\mathcal{K}}[t] - \mathbf{f}_{\mathcal{K}}(\mathbf{v})\right]\|^2_{\boldsymbol{\Gamma}^{-1}_{\mathcal{K},i}} \quad (5)$$
3) Adjust $\widehat{\mathbf{R}}_{\mathcal{K},i}[t] = \mathrm{diag}[\widehat{\boldsymbol{\gamma}}_{\mathcal{K},i}[t]]$ according to (4).

Since *step (1)* and *(3)* are decoupled, their decentralized implementations are straightforward. Now we omit the index $t$ and focus on solving *step (2)*. The global estimate $\widehat{\mathbf{v}}$ is traditionally obtained by the Gauss-Newton (GN) algorithm

$$\mathbf{v}_i^{k+1} = P_{\mathbb{V}}\left[\mathbf{v}_i^k + \mathbf{d}_i^k\right], \quad \mathbf{d}_i^k = \mathbf{Q}^{-1}(\mathbf{v}_i^k)\mathbf{q}(\mathbf{v}_i^k), \quad (6)$$

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

$$\mathbf{q}(\mathbf{v}_i^k) = \frac{1}{I}\sum_{p=1}^{I}\sum_{\mathcal{K}\in\mathcal{A}} \mathbf{F}^T_{\mathcal{K}}(\mathbf{v}_i^k)\mathbf{T}^T_{\mathcal{K},p}\boldsymbol{\Gamma}^{-1}_{\mathcal{K},p}\mathbf{T}_{\mathcal{K},p}\left[\mathbf{z}_{\mathcal{K}} - \mathbf{f}_{\mathcal{K}}(\mathbf{v}_i^k)\right]$$

$$\mathbf{Q}(\mathbf{v}_i^k) = \frac{1}{I}\sum_{p=1}^{I}\sum_{\mathcal{K}\in\mathcal{A}} \mathbf{F}^T_{\mathcal{K}}(\mathbf{v}_i^k)\mathbf{T}^T_{\mathcal{K},p}\boldsymbol{\Gamma}^{-1}_{\mathcal{K},p}\mathbf{T}_{\mathcal{K},p}\mathbf{F}_{\mathcal{K}}(\mathbf{v}_i^k), \quad (7)$$

with the Jacobians given by $\mathbf{F}_{\mathcal{V}}(\mathbf{v}) = \mathbf{I}_{2N}$, $\mathbf{F}_{\mathcal{C}}(\mathbf{v}) = \mathbf{H}_{\mathcal{C}}$, $\mathbf{F}_{\mathcal{I}}(\mathbf{v}) = (\mathbf{I}_{2N}\otimes\mathbf{v}^T)\mathbf{H}_{\mathcal{I}}$ and $\mathbf{F}_{\mathcal{F}}(\mathbf{v}) = (\mathbf{I}_{4L}\otimes\mathbf{v}^T)\mathbf{H}_{\mathcal{F}}$ defined in [13]. However, each area knows only its own estimate $\mathbf{v}_i^k$ and part of the measurements/functions in (7), which makes it impossible to run *step (2)* in a decentralized setting. DARSE solves *step (2)* via the Gossip-based Gauss-Newton (GGN) algorithm we developed in [15], which emulates the exact GN update in (6) by averaging via near-neighbor communications. Specifically, the GGN algorithm alternates between the GN *update* denoted by "$k$", and the gossip *exchange* denoted by "$\ell$". All areas have a clock that determines the time $t = \tau_k$ for the $k$-th update across the network. After the $k$-th update, the areas exchange information via gossiping at $\tau_{k,\ell} \in [\tau_k, \tau_{k+1})$ for $\ell_k$ times to compute the "network average" $\bar{\mathbf{h}}_k$ and $\bar{\mathbf{H}}_k$

$$\bar{\mathbf{h}}_k = \frac{1}{I}\sum_{i=1}^{I}\sum_{\mathcal{K}\in\mathcal{A}} \mathbf{F}^T_{\mathcal{K}}(\mathbf{v}_i^k)\mathbf{T}^T_{\mathcal{K},i}\boldsymbol{\Gamma}^{-1}_{\mathcal{K},i}\mathbf{T}_{\mathcal{K},i}\left[\mathbf{z}_{\mathcal{K}} - \mathbf{f}_{\mathcal{K}}(\mathbf{v}_i^k)\right] \quad (8)$$

$$\bar{\mathbf{H}}_k = \frac{1}{I}\sum_{i=1}^{I}\sum_{\mathcal{K}\in\mathcal{A}} \mathbf{F}^T_{\mathcal{K}}(\mathbf{v}_i^k)\mathbf{T}^T_{\mathcal{K},i}\boldsymbol{\Gamma}^{-1}_{\mathcal{K},i}\mathbf{T}_{\mathcal{K},i}\mathbf{F}_{\mathcal{K}}(\mathbf{v}_i^k) \quad (9)$$

in order to approximate $\mathbf{q}(\mathbf{v}_i^k)$ and $\mathbf{Q}(\mathbf{v}_i^k)$. For this computation, each area combines the information from its neighbors with a weight matrix $\mathbf{W}_k(\ell) \triangleq [W^k_{ij}(\ell)]_{I\times I}$ during $[\tau_{k,\ell}, \tau_{k,\ell+1})$, where $W^k_{ij}(\ell)$ is the weight associated to the edge $\{i,j\}$, which is non-zero if and only if the pair of areas $\{i,j\}$ communicate with each other. Define the local vector at the $i$-th area for the $\ell$-th gossip

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

with $\mathbf{h}_{k,i}(0) \triangleq \sum_{\mathcal{K}\in\mathcal{A}} \mathbf{F}^T_{\mathcal{K}}(\mathbf{v}_i^k)\mathbf{T}^T_{\mathcal{K},i}\boldsymbol{\Gamma}^{-1}_{\mathcal{K},i}\mathbf{T}_{\mathcal{K},i}\left[\mathbf{z}_{\mathcal{K}} - \mathbf{f}_{\mathcal{K}}(\mathbf{v}_i^k)\right]$ and $\mathbf{H}_{k,i}(0) \triangleq \sum_{\mathcal{K}\in\mathcal{A}} \mathbf{F}^T_{\mathcal{K}}(\mathbf{v}_i^k)\mathbf{T}^T_{\mathcal{K},i}\boldsymbol{\Gamma}^{-1}_{\mathcal{K},i}\mathbf{T}_{\mathcal{K},i}\mathbf{F}_{\mathcal{K}}(\mathbf{v}_i^k)$. The $i$-th area mixes the local information with its neighbors as

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

for all $i = 1, \cdots, I$. There are many ways to choose the weights under different protocols (see [16]). The DARSE scheme exploits the *Uncoordinated Random Exchange* (URE) protocol specified in [13]. After $\ell_k$ exchanges, the GGN descent for the $(k+1)$-th update at the $i$-th area is computed as $\mathbf{d}_i^k(\ell_k) = \mathbf{H}^{-1}_{k,i}(\ell_k)\mathbf{h}_{k,i}(\ell_k)$, which leads to

$$\mathbf{v}_i^{k+1} = P_{\mathbb{V}}\left[\mathbf{v}_i^k + \mathbf{d}_i^k(\ell_k)\right]. \quad (12)$$

## III. Placement Design for DARSE

As we said hybrid PSSE is a non convex problem and initialization is of paramount importance. We explored this problem in [14] and in [13] suggested to initialize DARSE using the PMUs as follows:

$$\mathbf{v}_0 = \mathbf{v}_{\text{nominal}} + (\mathbf{z}_{\mathcal{V}} - \mathbf{v}_{\text{nominal}})\sum_{i=1}^{I}\mathbf{T}^T_{\mathcal{K},i}\mathbf{T}_{\mathcal{K},i}$$

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

where $\mathbf{v}_{\text{nominal}}$ is a nominal profile obtained from previous estimates ( see [13] for more details) with a placement $\mathbf{T}_{\mathcal{K},i}$ for the PMUs that is designed for convergence based on the solution in [14][2]. In [14] the installed PMUs capture the voltage and all incident current measurements on that bus (see [10], [12]), so that $\{\mathbf{T}_{\mathcal{C},i}\}_{i=1}^{I}$ depend entirely on $\{\mathbf{T}_{\mathcal{V},i}\}_{i=1}^{I}$. Therefore, we define the PMU placement vector

$$\boldsymbol{\mathcal{V}} \triangleq [\mathcal{V}_1, \cdots, \mathcal{V}_N]^T, \quad \mathcal{V}_n \in \{0,1\}, \tag{13}$$

indicating if the $n$-th bus has a PMU.

Assuming the state is observable, we were able to find an upper bound of the ML mean square error (MSE)

$$\mathbb{E}\left\|\widehat{\mathbf{v}} - \bar{\mathbf{v}}\right\|^2 = \mathrm{Tr}\left[\left[\mathcal{P}(\boldsymbol{\mathcal{V}}) + \mathbf{S}(\bar{\mathbf{v}}\bar{\mathbf{v}}^T)\right]^{-1}\right] \leq \frac{2N}{\beta(\boldsymbol{\mathcal{V}})}. \tag{14}$$

where, with $\lambda_{\min}[\cdot]$ denoting the minimum eigenvalue, $\beta(\boldsymbol{\mathcal{V}})$ is the following function of the PMU placement

$$\beta(\boldsymbol{\mathcal{V}}) = \inf_{\mathbf{v}\in\mathbb{V}} \lambda_{\min}\left[\mathcal{P}(\boldsymbol{\mathcal{V}}) + \mathbf{S}(\mathbf{v}\mathbf{v}^T)\right]. \tag{15}$$

In (15), $\mathcal{P}(\boldsymbol{\mathcal{V}}) = \sum_{n=1}^{N} \mathcal{V}_n \left(\mathbf{H}_{I,n}^T \boldsymbol{\Gamma}_{I,n}^{-1} \mathbf{H}_{I,n} + \mathbf{H}_{J,n}^T \boldsymbol{\Gamma}_{J,n}^{-1} \mathbf{H}_{J,n}\right) + \sum_{n=1}^{N} \mathcal{V}_n \left(\boldsymbol{\Gamma}_{\mathcal{V},n}^{-1} \otimes \mathbf{e}_n \mathbf{e}_n^T\right)$ and

$$\mathbf{S}(\mathbf{v}\mathbf{v}^T) = \mathbf{H}_{\mathcal{I}}^T \left(\sum_{i=1}^{I} \mathbf{T}_{\mathcal{I},i}^T \boldsymbol{\Gamma}_{\mathcal{I},i}^{-1} \mathbf{T}_{\mathcal{I},i} \otimes \mathbf{v}\mathbf{v}^T\right) \mathbf{H}_{\mathcal{I}} \tag{16}$$

$$+ \mathbf{H}_{\mathcal{F}}^T \left(\sum_{i=1}^{I} \mathbf{T}_{\mathcal{F},i}^T \boldsymbol{\Gamma}_{\mathcal{F},i}^{-1} \mathbf{T}_{\mathcal{F},i} \otimes \mathbf{v}\mathbf{v}^T\right) \mathbf{H}_{\mathcal{F}}, \tag{17}$$

where $\boldsymbol{\Gamma}_{I,n}, \boldsymbol{\Gamma}_{J,n} \in \mathbb{R}^{2L_n \times 2L_n}$ correspond[3] to the variances of the real and imaginary parts of current measurements associated with line $(n,m)$ at bus $n$ (i.e., selected from $\boldsymbol{\Gamma}_{\mathcal{C},i}$ if a PMU is installed at bus $n$ in area $i$), and $\boldsymbol{\Gamma}_{\mathcal{V},n} \in \mathbb{R}^{2\times 2}$ contains the variances of the real and imaginary parts of voltage measurements in $\{\boldsymbol{\Gamma}_{\mathcal{V},i}\}_{i=1}^{I}$ if bus $n$ is in area $i$, and arbitrary if otherwise since $\mathcal{V}_n = 0$. The matrices $\mathcal{P}(\boldsymbol{\mathcal{V}})$ and $\mathbf{S}(\mathbf{v}\mathbf{v}^T)$ are the two terms present in the expression of the Hessian, due to PMUs and SCADA measurements respectively:

$$\mathbf{Q}(\mathbf{v}) = [\mathcal{P}(\boldsymbol{\mathcal{V}}) + \mathbf{S}(\mathbf{v}\mathbf{v}^T)]/I, \tag{18}$$

Therefore, we argue that PMU should be placed to maximize $\beta(\boldsymbol{\mathcal{V}})$. It is shown next that maximizing $\beta(\boldsymbol{\mathcal{V}})$ also indirectly improves the convergence.

As analyzed in [15, Theorem 2], the error made by the decentralized update with respect to the exact update can be bounded by an arbitrarily small constant

$$\left\|\mathbf{d}_i^k(\ell_k) - \mathbf{d}_i^k\right\| \leq C\lambda_\alpha^{\ell_{\min}}, \tag{19}$$

where $\ell_{\min} \triangleq \min\{\ell_k\}$ is the minimum number of gossip exchanges in the network for each update $k$, $C$ is a constant determined by the grid parameters and measurements and $\lambda_\alpha = (1-\alpha^{IT})^{IT}$ is a measure of the network connectivity with $\alpha$ being the minimal non-zero weight in matrix $\{\mathbf{W}_k(\ell)\}_{\ell=1,\cdots,\ell_k}^{k=1,\cdots,\infty}$. It is shown in [15, Theorem 1] that as long as the minimum cost $\sum_{i=1}^{I} \left\|\mathbf{T}_{\mathcal{K},i}\left[\mathbf{z}_{\mathcal{K}}[t] - \mathbf{f}_{\mathcal{K}}(\widehat{\mathbf{v}})\right]\right\|_{\boldsymbol{\Gamma}_{\mathcal{K},i}^{-1}}^2$ is sufficiently small and $\ell_{\min}$ is sufficiently large, the iterative error is bounded as $\limsup_{k\to\infty} \left\|\mathbf{v}_i^k - \widehat{\mathbf{v}}\right\| \leq C\lambda_\alpha^{\ell_{\min}}$ if

$$\left\|\mathbf{v}_i^0 - \widehat{\mathbf{v}}\right\| < 2\sqrt{\frac{\beta(\boldsymbol{\mathcal{V}})}{\phi(\boldsymbol{\mathcal{V}})}} - C\lambda_\alpha^{\ell_{\min}}, \tag{20}$$

where $\phi(\boldsymbol{\mathcal{V}})$ is a bound on the Lipschitz constant of the power flow equations, approximated in [14] as

$$\phi(\boldsymbol{\mathcal{V}}) \approx \lambda_{\max}[(\mathbf{I}_{2N} - \mathbf{J}_{\mathcal{V}})^T \mathbf{S}(\mathbf{I}_{2N}) (\mathbf{I}_{2N} - \mathbf{J}_{\mathcal{V}})], \tag{21}$$

where $\mathbf{S}(\mathbf{I}_{2N})$ is obtained by letting $\mathbf{V} = \mathbf{I}_{2N}$ in (16). Thus, the greater the COP metric $\rho(\boldsymbol{\mathcal{V}}) = \beta(\boldsymbol{\mathcal{V}})/\phi(\boldsymbol{\mathcal{V}})$, the less sensitive PSSE is to initialization and the faster the algorithm converges. Therefore, the PMU can play a dual role of both providing a good initialization and lowering the error. Thus, the *optimal* PMU placement design aims at maximizing

$$\begin{aligned} \max_{\boldsymbol{\mathcal{V}}} \quad & \frac{\lambda_{\min}\left[\mathcal{P}(\boldsymbol{\mathcal{V}}) + \mathbf{S}(\mathbf{v}_{\text{nom}}\mathbf{v}_{\text{nom}}^T)\right]}{\lambda_{\max}\left[(\mathbf{I}_{2N} - \mathbf{J}_{\mathcal{V}})^T \mathbf{S}(\mathbf{I}_{2N}) (\mathbf{I}_{2N} - \mathbf{J}_{\mathcal{V}})\right]} \\ \text{s.t.} \quad & \mathbf{J}_{\mathcal{V}} = \mathbf{I}_2 \otimes \mathrm{diag}(\boldsymbol{\mathcal{V}}), \quad \mathbf{1}_N^T \boldsymbol{\mathcal{V}} \leq N_{\text{PMU}}, \quad \mathcal{V}_n \in \{0,1\}, \end{aligned} \tag{22}$$

where $\mathbf{v}_{\text{nom}} = [\mathbf{1}_N^T, \mathbf{0}_N^T]^T$ is a nominal profile used to avoid the search in (15) and $N_{\text{PMU}}$ is the budget of PMU deployment. Solving this problem efficiently requires formulating the problem as a semi-definite program under the relaxation of $\mathcal{V}_n \in \{0,1\}$ to $[0,1]$ and the Charnes-Cooper transformation in [14], but due to space contraints we do not repeat the lengthy reformulation here and refer the readers to the details in [14].

## IV. Study of Communication and Sensing Graphs

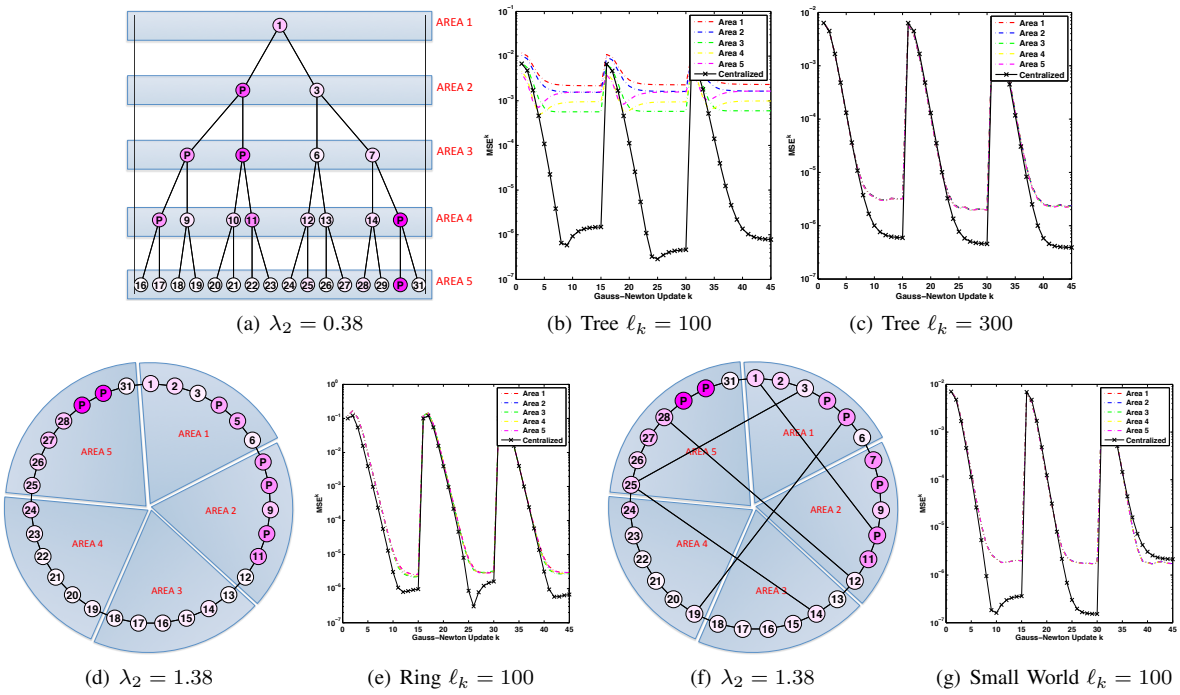
In this section, we test the DARSE algorithm with optimal PMU placements based on (22) with $N_{\text{PMU}} = 6$ for different grid configurations and communication graphs. They all consist of $I = 5$ areas: Figs. 1(a) is a hierarchical graph from Area 1 to Area 5 while Figs. 1(d) and 1(f) form a typical ring network between areas. These examples are chosen considering the typical topologies found in the distribution and transmission grids respectively, and assuming that the communication lines are laid over these infrastructure, resembling the underlying graphs. The figures also show the algebraic connectivities of these three topologies (second smallest eigenvalue $\lambda_2$ of the Laplacian), which shows that the tree network has much lower connectivity since there is no link between Area 1 and Area 5. For each grid topology we have $N = 31$ buses[4] (PMU deployment ratio of $18.75\%$). PMU locations are highlighted as $\mathbf{P}$. Let $\mathbf{U}$ and $\boldsymbol{\Sigma}$ be the eigenvectors and eigenvalues of the matrix $\mathbf{S}(\mathbf{I}_{2N})$ determined by SCADA measurements and grid topologies. The intensity of the color at each node is given by $\|\mathbf{U}(n,:)\boldsymbol{\Sigma}\| / \max_n \|\mathbf{U}(n,:)\boldsymbol{\Sigma}\|$ as a measure of its significance in the grid eigen-structure. A first observation is that the optimal PMU placements align very well with these highly influential nodes.

All the scenarios are tested using the same $80\%$ of the SCADA measurements randomly chosen from the ensemble over 3 snapshots with 1 bad data in each area with 10 times larger variances than the normal variance $\sigma^2 = 10^{-6}$, where each snapshot is generated by perturbing the demand by $10\%$. The state estimation and tracking performances for the DARSE

[2]We assume that SCADA measurements are given.

[3]The parameter $L_n$ is the number of incident lines at bus $n$

[4]The generators are located at bus 1 to 5 and the line parameters are adapted from the IEEE-30 test case.




(a) $\lambda_2 = 0.38$ (b) Tree $\ell_k = 100$ (c) Tree $\ell_k = 300$

(d) $\lambda_2 = 1.38$ (e) Ring $\ell_k = 100$ (f) $\lambda_2 = 1.38$ (g) Small World $\ell_k = 100$

Fig. 1. Performances of individual areas for different grids using different communication networks.

scheme in different scenarios are illustrated in Figs. 1(b), 1(c), 1(e) and 1(g), compared with the centralized solution with no bad data. We restrict the number of random pair-wise exchanges across the entire network to be $\ell_k = 100$ for all $K$ updates and average over 50 experiments.

For the *tree network*, $\ell_k = 100$ is insufficient to diffuse the local information on the state, which is, in turn, not accurately estimated and tracked. When the exchanges increase to $\ell_k = 300$ the scheme achieves much better performance. Both the lower algebraic connectivity ($\lambda_2 = 0.38$ vs. $\lambda_2 = 1.38$) and the asymmetry in buses and sensor information gathered across areas render the DARSE over the tree network slower compared to the ring and small world grid cases. The asymmetry explains also why different areas saturate at different error floors in the tree network with insufficient gossip exchanges. In contrast, for the *ring network* and *small world network*, it is observed that they achieve consensus and saturate at the same error floor even with $\ell_k = 100$ exchanges because of the more balanced information carried by each area and the higher communication connectivity $\lambda_2 = 1.38$.

This brief analysis suggests that transmission networks potentially more so than distribution networks can benefit from gossip algorithm that capitalize on the relatively short nodal distance attained by overlaying communication lines over the *small world* nature of transmission lines.

## 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. 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.

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

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

[6] 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.

[7] 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.

[8] 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.

[9] B. Gou, "Optimal Placement of PMUs by Integer Linear Programming," *Power Systems, IEEE Transactions on*, vol. 23, no. 3, pp. 1525–1526, 2008.

[10] T. Baldwin, L. Mili, M. Boisen Jr, and R. Adapa, "Power System Observability with Minimal Phasor Measurement Placement," *Power Systems, IEEE Transactions on*, vol. 8, no. 2, pp. 707–715, 1993.

[11] Q. Li, R. Negi, and M. Ilic, "Phasor Measurement Units Placement for Power System State Estimation: a Greedy Approach," in *Power and Energy Society General Meeting, 2011 IEEE*. IEEE, 2011, pp. 1–8.

[12] V. Kekatos and G. Giannakis, "A Convex Relaxation Approach to Optimal Placement of Phasor Measurement Units," in *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2011 4th IEEE International Workshop on*. IEEE, 2011, pp. 145–148.

[13] X. Li and A. Scaglione, "Robust decentralized state estimation and tracking for power systems via network gossiping," *Selected Areas in Communications, IEEE Journal on*, vol. 31, no. 7, pp. 1184–1194, 2013.

[14] X. Li, A. Scaglione, and T.-H. Chang, "Optimal Sensor Placement for Hybrid State Estimation in Smart Grid," *Acoustics, Speech and Signal Processing, 2013. ICASSP 2013 Proceedings. 2013 IEEE International Conference on*.

[15] X. Li and A. Scaglione, "Convergence and Applications of a Gossip-based Gauss-Newton Algorithm," *accepted by IEEE Trans. on Signal Processing, arXiv preprint arXiv:1210.0056*, 2013.

[16] 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.