DECENTRALIZED POWER SYSTEM STATE ESTIMATION VIA NON-CONVEX MULTI-AGENT OPTIMIZATION

Aritra Konar, Ahmed S. Zamzam, and Nicholas D. Sidiropoulos

Dept. of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN

ABSTRACT

Power system state estimation (PSSE) is a critical task for ensuring reliable, efficient system operation in power grids. As a result of the quadratic relationship between the voltage state variables and the observable quantities, the problem is non-convex and NP–Hard in the worst case. In this paper, we present a decentralized approach for PSSE by casting it in the framework of multi-agent optimization. Two synchronous, gradient-consensus schemes are brought to bear upon the problem for the purpose of computing approximate solutions by performing gradient descent updates at each bus followed by local averaging over single-hop neighbors. In contrast to the prevailing state-of-art in decentralized PSSE, these algorithms feature guaranteed convergence to stationary solutions of the non-convex PSSE problem under very mild assumptions, at the expense of each bus having to maintain a local copy of the global state variable. Numerical experiments are used to demonstrate the effectiveness of the methods.

1. INTRODUCTION

Power system state estimation (PSSE) is a fundamental problem in power systems engineering where the objective is to estimate the complex voltages across the constituent buses of an electrical power grid from a set of measurements pertaining to the observable quantities [1]. Performing accurate and efficient state estimation is of paramount importance for enhancing our ability to monitor the operating conditions of the power grid in real-time. A major impediment towards realizing this goal is the fact that PSSE is a non-convex optimization problem which is NP–Hard in the worst case [2].

PSSE has been traditionally performed at a single control center using supervisory control and data acquisition (SCADA) measurements. However, such centralized approaches are increasingly being rendered impractical by the scale of network interconnections, complexity issues incurred at a single control center, privacy concerns amongst regional operators and susceptibility to failure/cyberattacks. This underscores the need for developing *decentralized* algorithms which are capable of computing high quality sub-optimal solutions for this class of hard optimization problems in polynomial-time.

Recognizing this requirement, recent years have witnessed a flurry of research activity in the development of such decentralized algorithms, ranging from gossip-based schemes [3, 4], dual decomposition / operator-splitting methods [6–8], to distributed implementations of semidefinite programming [9, 10] and the Gauss-Newton method [11]. A salient feature of *all* the aforementioned approaches is that they lack provable convergence guarantees; i.e., given any network operating at an arbitrary choice of a valid voltage profile, one cannot theoretically characterize the solution obtained via these approaches.

Contact: (konar006,zamza002,nikos)@umn.edu. Supported in part by NSF CIF-1525194 and a Doctoral Dissertation Fellowship from the University of Minnesota.

In this paper, we consider the PSSE problem in the context of multi-agent optimization [12–14], where each bus corresponds to an agent and only has access to SCADA measurements corresponding to quantities it can directly observe and can communicate intermittently with its single-hop neighbors. We evaluate the performance of two decentralized, synchronous, gradient-consensus schemes [15, 16] on the resulting non-convex network optimization problem. These methods require each bus to maintain a local copy of the global state variable, which is updated via inexpensive gradient descent steps performed on the individual cost function of each bus followed by a weighted averaging step computed using the local updates of the single-hop neighbors for the ultimate purpose of attaining consensus over the network. In contrast to the previously mentioned approaches for decentralized PSSE, these methods are guaranteed to (asymptotically) converge to the set of stationary points of the PSSE problem under very mild assumptions, which are always satisfied in practice. Simulation results are used to demonstrate the performance of these methods and to illustrate the claims.

Finally, we point out that a convergence result for decentralized PSSE in a similar multi-agent scenario was also derived in [17]. However, our results hold under considerably more relaxed and realistic conditions: i) the convergence of our methods requires exchanges amongst buses over the network defined by the grid topology whereas in [17], the communication model is decoupled from the grid topology altogether; ii) the communication overhead is substantially larger in [17] since exchanges amongst buses involve sharing matrices in addition to vectors while multiple rounds of exchanges must be executed to guarantee convergence. In contrast, in our algorithms, at each round, each bus only needs to exchange a vector with its neighbors *once* while still preserving convergence; and iii) the result in [17] only pertains to the Gauss-Newton method whereas one of the methods we use [16] can be used a generate an entire family of convergent algorithms for the problem at hand. [1]

2. SYSTEM MODEL

Given an electrical power transmission network, we model it as an undirected graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N} := \{1, \cdots, N\}$ denotes the set of buses (nodes) and $\mathcal{E}$ represents the set of transmission lines (edges), with each transmission line $(m, n) \in \mathcal{E}$ corresponding to an unordered pair of distinct buses. At each bus $n \in \mathcal{N}$, we define the following complex nodal quantities: voltage $V_n := |V_n|e^{j\theta_n}$, current injection $I_n := |I_n|e^{j\phi_n}$, and power injection $S_n := P_n + jQ_n$ (here P_n and Q_n denote the active and reactive power injections respectively). Associated with each transmission line $(m, n) \in \mathcal{E}$ are the following line quantities: I_{mn} is the complex current flowing from bus m to n, while $S_{mn} := P_{mn} + jQ_{mn}$ is the apparent power flow from bus m to n (here P_{mn} and Q_{mn} denote the active and reactive power flow respectively), as seen at the send-

[1] Owing to space limitations, we only present one algorithm from this family.

ing end. For notational simplicity, we represent the nodal quantities $\{V_n\}_{n\in\mathcal{N}}, \{I_n\}_{n\in\mathcal{N}}, \{P_n\}_{n\in\mathcal{N}}$ and $\{Q_n\}_{n\in\mathcal{N}}$ in the form of vectors $\mathbf{v} := [V_1^*, \cdots, V_N^*]^H \in \mathbb{C}^N$, $\mathbf{i} := [I_1^*, \cdots, I_N^*]^H \in \mathbb{C}^N$, $\mathbf{p} := [p_1, \cdots, p_N]^T \in \mathbb{R}^N$, and $\mathbf{q} := [q_1, \cdots, q_N]^T \in \mathbb{R}^N$ respectively. Bus $n \in \mathcal{N}$ has access to a set $\mathcal{L}_n$ of (possibly noisy) SCADA measurements $\{z_l\}_{l\in\mathcal{L}_n}$ corresponding to the voltage magnitude $|V_n|$, active and reactive power injections P_n and Q_n, and possibly the active and reactive power flows P_{nm} and Q_{nm}, $\forall\, m \in \{m|(n,m) \in \mathcal{E}\}$ (i.e., if bus n corresponds to the sending end).

In an AC power flow model, current and voltage laws mandate that the state variables $\mathbf{v}$ are quadratically related to the SCADA measurements $\{z_l\}_{l\in\mathcal{L}_n}, \forall\, n \in \mathcal{N}$. This holds trivially for the square of the voltage magnitude measurements, since $|V_n|^2 = V_n V_n^*, \forall\, n \in \mathcal{N}$. In order to see that such a relationship exists between $\mathbf{v}$ and the other power measurements, let $\mathbf{Y} \in \mathbb{C}^{N\times N}$ denote the bus-admittance matrix whose entries are given by

$$Y_{mn} := \begin{cases} -y_{mn}, & (m,n) \in \mathcal{E} \\ \bar{y}_{nn} + \sum_{k\in\mathcal{N}_n} y_{nk}, & m = n \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where y_{mn} is the admittance of line $(m,n) \in \mathcal{E}$, $\bar{y}_{nn}$ is the admittance to ground at bus $n \in \mathcal{N}$, and $\mathcal{N}_n := \{k|(n,k) \in \mathcal{E}\}$ denotes the immediate neighborhood of bus n. We point out that $\mathbf{Y}$ is symmetric but non-Hermitian, and is also sparse. Combining Kirchoff's current law and the multivariate form of Ohm's law, the relationship between the nodal voltages and currents can be expressed as

$$\mathbf{i} = \mathbf{Y}\mathbf{v} \tag{2}$$

For the power injections, under the AC power flow model, it holds that $P_n + jQ_n = V_n I_n^*, \forall\, n \in \mathcal{N}$. Utilizing (2), we obtain the following matrix-vector relationship

$$\mathbf{p} + j\mathbf{q} = \text{diag}(\mathbf{v})\mathbf{i}^* = \text{diag}(\mathbf{v})\mathbf{Y}^*\mathbf{v}^* \tag{3}$$

Meanwhile, by appealing to Ohm's and Kirchoff's laws, the line current I_{mn} can be expressed as

$$I_{mn} = \bar{y}_{mn} V_m + y_{mn}(V_m - V_n), \forall\, (m,n) \in \mathcal{E} \tag{4}$$

where $\bar{y}_{mn}$ denotes the shunt admittance at bus m corresponding to line (m,n). The reverse direction current I_{nm} can be obtained similarly by symmetry. Note that $I_{mn} \neq -I_{nm}$ as $\bar{y}_{mn} \neq 0$. The sending-end active and reactive power flow from bus m to n can now be expressed as

$$\begin{aligned} P_{mn} + jQ_{mn} &= V_m I_{mn}^* \\ &= (\bar{y}_{mn}^* + y_{mn}^*)|V_m|^2 - y_{mn}^* V_m V_n^*, \forall\, (m,n) \in \mathcal{E} \end{aligned} \tag{5}$$

where in the second step we have made use of (4). From (3) and (5), it can be observed that the power measurements are quadratically related to $\mathbf{v}$.

We now make explicit the relationship between the measurements $\{z_l\}_{l\in\mathcal{L}_n}$ and $\mathbf{v}$. At each bus $n \in \mathcal{N}$, the available measurements can be expressed in the quadratic form

$$z_l = \mathbf{v}^H \mathbf{H}_l \mathbf{v} + n_l, \forall\, l \in \mathcal{L}_n \tag{6}$$

where $\mathbf{H}_l \in \mathbb{C}^{N\times N}$ (to be specified shortly) and $n_l \sim \mathcal{N}(0, \sigma_l^2)$ represents additive noise (assumed independent across all meters). Considering the measurement of the voltage magnitude squared at bus n, we have $|V_n|^2 = V_n V_n^* = \mathbf{v}^H \mathbf{e}_n \mathbf{e}_n^T \mathbf{v}$ (where $\mathbf{e}_n \in \mathbb{R}^N$ represents the n^{th} canonical basis vector of $\mathbb{R}^N$). Hence, in this case, we have $\mathbf{H}_{V_n} = \mathbf{e}_n \mathbf{e}_n^T$. In order to define $\{\mathbf{H}_l\}_{l\in\mathcal{L}_n\setminus V_n}$ for the active and reactive power injection and flow measurements, we first define the admittance-related matrices

$$\mathbf{Y}_n := \mathbf{e}_n \mathbf{e}_n^T \mathbf{Y}, \forall\, n \in \mathcal{N} \tag{7a}$$

$$\mathbf{Y}_{mn} := (\bar{y}_{mn} + y_{mn})\mathbf{e}_m \mathbf{e}_m^T - y_{mn}\mathbf{e}_m \mathbf{e}_n^T, \forall\, (m,n) \in \mathcal{E} \tag{7b}$$

By equating the real and imaginary parts of (3) and (5), we obtain the matrices

$$\begin{aligned} &\mathbf{H}_{P_n} := \frac{1}{2}(\mathbf{Y}_n + \mathbf{Y}_n^H), \qquad \mathbf{H}_{Q_n} := \frac{j}{2}(\mathbf{Y}_n - \mathbf{Y}_n^H) \\ &\mathbf{H}_{P_{mn}} := \frac{1}{2}(\mathbf{Y}_{mn} + \mathbf{Y}_{mn}^H),\ \mathbf{H}_{Q_{mn}} := \frac{j}{2}(\mathbf{Y}_{mn} - \mathbf{Y}_{mn}^H) \end{aligned} \tag{8}$$

We point out that for each bus $n \in \mathcal{N}$, the measurement matrices $\{\mathbf{H}_l\}_{l\in\mathcal{L}_n\setminus V_n}$ are rank-2, indefinite Hermitian matrices, while $\mathbf{H}_{V_n}$ is rank-1, positive semidefinite.

3. PROBLEM FORMULATION

Assuming the availability of the noisy measurements $\{z_l\}_{l\in\mathcal{L}_n}$ across all buses $n \in \mathcal{N}$, adopting a maximum-likelihood estimation approach results in the following weighted least squares optimization problem

$$\min_{\mathbf{v}\in\mathcal{K}} \left\{ F(\mathbf{v}) := \sum_{n=1}^N f_n(\mathbf{v}) := \sum_{n=1}^N \sum_{l\in\mathcal{L}_n} \frac{(\mathbf{v}^H \mathbf{H}_l \mathbf{v} - z_l)^2}{\sigma_l^2} \right\} \tag{9}$$

where $\mathcal{K} \subset \mathbb{C}^N$ is a simple[2], convex, compact set which represents constraints on $\mathbf{v}$ known *apriori*. In this paper, we will work with the rectangular coordinate representation of the state variables $\mathbf{v}$. For this purpose, we define $\bar{\mathbf{v}} := [\mathbf{v}_r^T, \mathbf{v}_i^T]^T \in \mathbb{R}^{2N}$, where $\mathbf{v}_r = \text{Re}(\mathbf{v})$ and $\mathbf{v}_i = \text{Im}(\mathbf{v})$. We also define a set of symmetric measurement matrices $\bar{\mathbf{H}}_l \in \mathbb{R}^{2N\times 2N}$ as

$$\bar{\mathbf{H}}_l := \begin{bmatrix} \text{Re}\{\mathbf{H}_l\} & -\text{Im}\{\mathbf{H}_l\} \\ \text{Im}\{\mathbf{H}_l\} & \text{Re}\{\mathbf{H}_l\} \end{bmatrix}, \forall\, l \in \mathcal{L}_n, \forall\, n \in \mathcal{N} \tag{10}$$

Note that we have

$$\mathbf{v}^H \mathbf{H}_l \mathbf{v} = \bar{\mathbf{v}}^T \bar{\mathbf{H}}_l \bar{\mathbf{v}},\ \forall\, l \in \mathcal{L}_n, \forall\, n \in \mathcal{N} \tag{11}$$

Hence, in terms of rectangular coordinates, (9) can be equivalently represented as

$$\min_{\bar{\mathbf{v}}\in\bar{\mathcal{K}}} \left\{ F(\bar{\mathbf{v}}) := \sum_{n=1}^N f_n(\bar{\mathbf{v}}) := \sum_{n=1}^N \sum_{l\in\mathcal{L}_n} \frac{(\bar{\mathbf{v}}^T \bar{\mathbf{H}}_l \bar{\mathbf{v}} - z_l)^2}{\sigma_l^2} \right\} \tag{12}$$

where $\bar{\mathcal{K}} \subset \mathbb{R}^{2N}$ corresponds to the representation of $\mathcal{K}$ in rectangular coordinates. Note that (12) is a non-convex optimization problem which is NP–Hard in the worst case. Here, we aim to compute high-quality approximate solutions for (12) in polynomial-time via decentralized algorithms. Towards this end, we propose to cast (12) in the framework of multi-agent optimization, where each bus corresponds to an agent. From (12), it is evident that the global cost function $F(.)$ is decomposable across buses (agents). We make the following standard assumptions in multi-agent scenarios:

- (A1) The true voltage profile $\bar{\mathbf{v}}$ is fixed.
- (A2) The graph of the network $\mathcal{G}$ is connected (always satisfied).
- (A3) Each bus $n \in \mathcal{N}$ only has knowledge of its own cost function $f_n(.)$ and the feasible set $\bar{\mathcal{K}}$.
- (A4) Each bus is only aware of its interconnections to its immediate neighbors instead of knowing the full topology of the network.

Given these assumptions, some degree of local coordination is required amongst neighboring buses in order to minimize the global cost function $F(.)$. Furthermore, note that (12) exhibits the following properties:

- (B1) The set of minima of $F(.)$ over $\bar{\mathcal{K}}$ is non-empty and compact.
- (B2) Each $f_n(.)$ is continuously differentiable on an open set containing $\bar{\mathcal{K}}$.

[2]By *simple*, we mean that the Euclidean projection operation onto the set $\mathcal{K}$ can be computed in closed form.

(B3) Each $\nabla f_n(.)$ is locally Lipschitz continuous on $\bar{\mathcal{K}}$ with constant $L_n > 0$.

(B4) $F(.)$ is locally Lipschitz continuous on $\bar{\mathcal{K}}$; i.e., $\|\nabla F(\bar{\mathbf{v}})\|_2 \leq L_F, \forall\, \bar{\mathbf{v}} \in \bar{\mathcal{K}}$ for some $L_F > 0$.

Our goal is to compose simple, decentralized algorithms that converge to stationary points of (12) in the setting described by A1-A4 while only exploiting the (admittedly mild) properties B1-B4.

Definition 1. (Stationary point) [18, Proposition 3.1.1] If $\bar{\mathbf{v}}^*$ is a stationary point of (12) over $\bar{\mathcal{K}}$, then

$$\nabla F(\bar{\mathbf{v}}^*)^T(\bar{\mathbf{v}} - \bar{\mathbf{v}}^*) \geq 0, \forall\, \bar{\mathbf{v}} \in \bar{\mathcal{K}} \tag{13}$$

We denote the set of such stationary points by $\mathcal{S}$.

4. DECENTRALIZED GRADIENT CONSENSUS METHODS

In this section, we propose to use two synchronous, consensus driven, decentralized first-order methods (FOMs) for computing stationary points of (12) under the assumptions stated in the previous section. At a high level, these algorithms can be summarized as follows: first, each bus performs local gradient decent iterations with its own local estimate of $\bar{\mathbf{v}}$, followed by a second round of exchanges amongst neighboring buses with the goal of ultimately attaining consensus over the network with regard to the value of the minimizer of $F(\bar{\mathbf{v}})$. Each of the communicating buses then updates its own estimate by taking a weighted average of its current estimate and the estimates received from neighboring buses. Note that the extreme sparsity of power networks (the average degree of each bus has been observed to be around 2 [19]) facilitates the implementation of such synchronous algorithms since each bus only has to share its updates and aggregate estimates in coordination with few other buses. A more detailed exposition regarding these algorithms now follows.

4.1. Decentralized Gradient Descent (DGD)

In this scheme [15], given the current local estimates $\{\bar{\mathbf{v}}_n[k]\}_{n=1}^N$ at iteration $k \in \mathbb{N}$, the updates of DGD can be expressed as

$$\bar{\mathbf{z}}_n[k] = \mathbf{\Pi}_{\bar{\mathcal{K}}}(\bar{\mathbf{v}}_n[k] - \alpha[k]\nabla f_n(\bar{\mathbf{v}}_n[k])) \tag{14a}$$

$$\bar{\mathbf{v}}_n[k+1] = \sum_{m \in \overline{\mathcal{N}}_n} W_{nm}\bar{\mathbf{z}}_m[k],\ \forall\, n \in \mathcal{N}, \forall\, k \in \mathbb{N} \tag{14b}$$

where $\mathbf{\Pi}_{\bar{\mathcal{K}}}(.)$ denotes the Euclidean projection operator onto $\bar{\mathcal{K}}$, $\{\alpha[k]\}_{k\in\mathbb{N}}$ is the step-size sequence, $\overline{\mathcal{N}}_n := \mathcal{N}_n \cup \{n\}$ is defined as the in-neighborhood of bus n, and $\{W_{nm}\}_{n,m=1}^N$ are the averaging weights. The first step (14a) denotes the local GD step (performed by all buses simultaneously) while (14b) denotes the consensus step. The weights $\{W_{nm}\}_{n,m=1}^N$ can be represented in the form of a matrix $\mathbf{W} \in \mathbb{R}^{N\times N}$, which is typically chosen to be doubly stochastic for ensuring consensus.[3] Choices of weights which satisfy this condition include the uniform weights [20], Laplacian weights [21], as well as the maximum degree weight, Metropolis Hastings and the least-mean square consensus weight rules [22]. Let $\{\bar{\mathbf{v}}_{\text{avg}}[k]\}_{k\in\mathbb{N}} := \{(1/N)\sum_{n=1}^N \bar{\mathbf{v}}_n[k]\}_{k\in\mathbb{N}}$ be the sequence formed by averaging the local variables generated at each step by DGD. If $\mathbf{W}$ is doubly stochastic and $\{\alpha[k]\}_{k\in\mathbb{N}}$ is chosen such that

$$\alpha[k] \geq 0, \sum_{k\in\mathbb{N}} \alpha[k] = \infty,\ \sum_{k\in\mathbb{N}} \alpha^2[k] < \infty \tag{15}$$

then, under (A1)-(A4) and (B1)-(B4), [15, Theorem 1] asserts that: i) every limit point of $\{\bar{\mathbf{v}}_{\text{avg}}[k]\}_{k\in\mathbb{N}}$ is a stationary point of (12); and ii) all sequences $\{\bar{\mathbf{v}}_n[k]\}_{k\in\mathbb{N}}$ asymptotically attain consensus; i.e., $\|\bar{\mathbf{v}}_n[k] - \bar{\mathbf{v}}_{\text{avg}}[k]\| \underset{k\to\infty}{\longrightarrow} 0, \forall\ n \in \mathcal{N}$. If we define the sequence $\{\bar{\mathbf{v}}[k]\}_{k\in\mathbb{N}} := \{[\bar{\mathbf{v}}_1[k]^T, \cdots, \bar{\mathbf{v}}_N[k]^T]^T\}_{k\in\mathbb{N}}$, then the above result implies that $\{\bar{\mathbf{v}}[k]\}_{k\in\mathbb{N}}$ converges to the set $\{\mathbf{1}\otimes\bar{\mathbf{v}} | \bar{\mathbf{v}} \in \mathcal{S}\}$.

[3] Since we model the graph as being undirected, we have $\mathbf{W} = \mathbf{W}^T$ in this case.

4.2. In-Network Non-convex Optimization (NEXT)

NEXT [16] is a sophisticated algorithm where each agent approximates the global cost function $F(.)$ via a local convex surrogate at every step followed by utilizing dynamic consensus as a means of enforcing agreement amongst the agents. Here, we utilize a simple version which, while being similar to DGD in being gradient based, utilizes first-order information in a different way. The updates of NEXT are given by

$$\tilde{\mathbf{v}}_n[k] = \mathbf{\Pi}_{\bar{\mathcal{K}}}\left(\bar{\mathbf{v}}_n[k] - \frac{1}{\tau}(\nabla f_n(\bar{\mathbf{v}}_n[k]) + \tilde{\boldsymbol{\pi}}_n[k])\right) \tag{16a}$$

$$\bar{\mathbf{z}}_n[k] = \bar{\mathbf{v}}_n[k] + \alpha[k](\tilde{\mathbf{v}}_n[k] - \bar{\mathbf{v}}_n[k]) \tag{16b}$$

$$\bar{\mathbf{v}}_n[k+1] = \sum_{m\in\overline{\mathcal{N}}_n} W_{nm}\bar{\mathbf{z}}_m[k] \tag{16c}$$

$$\bar{\mathbf{y}}_n[k+1] = \sum_{m\in\overline{\mathcal{N}}_n} W_{nm}\bar{\mathbf{y}}_m[k] + (\nabla f_n(\bar{\mathbf{v}}_n[k+1]) - \nabla f_n(\bar{\mathbf{v}}_n[k]) \tag{16d}$$

$$\tilde{\boldsymbol{\pi}}_n[k+1] = N\bar{\mathbf{y}}_n[k+1] - \nabla f_n(\bar{\mathbf{v}}_n[k+1]),\ \forall\, n \in \mathcal{N}, \forall\, k \in \mathbb{N} \tag{16e}$$

where $\tau > 0$ is a constant, $\{\alpha[k]\}_{k\in\mathbb{N}}$ is a step-size sequence, and $\tilde{\boldsymbol{\pi}}_n[k]$ is a local variable that aims to dynamically mimic the term $\boldsymbol{\pi}_n[k] := \sum_{m\neq n}\nabla f_m(\bar{\mathbf{v}}_n[k])$ (i.e., the gradients of the unobserved cost functions at the other buses). This implies that in (16a), the local variables are updated with information (although inexact) of the *full* gradient $\nabla F(.)$ relative to DGD, where each bus only utilizes information of the local gradient $\nabla f_n(.)$. As shown later, the information in $\tilde{\boldsymbol{\pi}}_n[k]$ becomes more accurate with iterations. A convex combination of the result and the previous local variable is then performed in (16b). The consensus update of $\bar{\mathbf{v}}_n[k]$ in (16c) is similar to (14b). In addition, NEXT uses a local auxiliary variable $\bar{\mathbf{y}}_n[k]$ for asymptotically tracking the quantity $\overline{\nabla F}(\bar{\mathbf{v}}_n[k]) := (1/N)\sum_{n=1}^N \nabla g_n(\bar{\mathbf{v}}_n[k])$ via an additional dynamic consensus step (16d). This update is then used in (16e) for updating the estimate $\tilde{\boldsymbol{\pi}}_n([k])$. Hence, while each step of NEXT benefits from the global information provided about $\nabla F(.)$ (which is not exploited by DGD), this comes at the expense of twice the communication cost. According to [16, Theorem 3], if $\mathbf{W}$ is doubly stochastic and the step-size sequence $\{\alpha[k]\}_{k\in\mathbb{N}}$ is chosen such that and $\alpha[k] \in (0,1], \forall\, k \in \mathbb{N}$ and

$$\sum_{k\in\mathbb{N}} \alpha[k] = \infty,\ \sum_{k\in\mathbb{N}} \alpha^2[k] < \infty \tag{17}$$

then NEXT converges to the set of stationary points of (12) and asymptotically attains consensus in the same sense as DGD. Note that the second result implies that $\|\overline{\nabla F}(\bar{\mathbf{v}}_n[k]) - \bar{\mathbf{y}}_n[k+1]\| \underset{k\to\infty}{\longrightarrow} 0, \forall\ n \in \mathcal{N}$, and thus $\|\tilde{\boldsymbol{\pi}}_n[k] - \boldsymbol{\pi}_n[k]\| \underset{k\to\infty}{\longrightarrow} 0, \forall\ n \in \mathcal{N}$, as we claimed previously.

Remark 1. Both algorithms feature guaranteed convergence to stationary points of (12) for *any* choice of initialization $\bar{\mathbf{v}} \in \bar{\mathcal{K}}$ while only utilizing first-order information in a distributed fashion. Note that these convergence results do not require explicit knowledge of the function parameters $L_F, \{L_n\}_{n\in\mathcal{N}}$; i.e., we merely need to guarantee their existence. Furthermore, the sparsity of the matrices in the quadratic forms can be exploited to form the gradient at each step at each bus n in time proportional to the number of non-zeros in $\{\bar{\mathbf{H}}_l\}_{l\in\mathcal{L}_n}$. Combined with the fact that the operator $\mathbf{\Pi}_{\bar{\mathcal{K}}}(.)$ can

be computed efficiently, this overall translates into low per-iteration complexity at each bus. Finally, these methods require each bus to maintain and share a copy of the global state variable with its single-hop neighbors. While this does result in increased memory and communication overhead, we point out that NEXT exploits this additional cost judiciously to extract information regarding the global cost function $F(.)$ at each bus which results in improved coordination amongst the buses. This singular feature is not shared by DGD or the methods [6–8, 11], which only maintain copies of the local state at each bus.

5. SIMULATION RESULTS

In this section, we present simulation results to demonstrate the performance of the algorithms on IEEE benchmark networks. For a given network, the voltage magnitude at each bus was generated from a uniform distribution over the interval $[0.9, 1.1]$, while the voltage angle was uniformly distributed over $[-0.1\pi, 0.1\pi]$. We fixed the phase of the reference bus to 0 in order to resolve the phase ambiguity. The quadratic matrices for generating the SCADA measurements were obtained using MATPOWER [23]. We define the feasible set $\bar{\mathcal{K}}$ to impose upper bounds on the maximum value of the voltage magnitude squared; i.e.,

$$\bar{\mathcal{K}} = \{\bar{\mathbf{v}} \in \mathbb{R}^{2N} | \, \bar{v}_n^2 + \bar{v}_{n+N}^2 \leq (1.1)^2, \forall \, n \in \mathcal{N}\} \tag{18}$$

Note that the projection operator $\mathbf{\Pi}_{\bar{\mathcal{K}}}(.)$ in this case can be computed in closed form. Meanwhile, the entries of $\mathbf{W}$ are chosen according to the Metropolis-Hastings rule, which is given by

$$W_{nm} = \begin{cases} 1/(\max\{d_n, d_m\} + 1), & (n,m) \in \mathcal{E} \\ 1 - \sum_{m \in \mathcal{N}_n} 1/(\max\{d_n, d_m\} + 1), & n = m \\ 0, & \text{otherwise} \end{cases} \tag{19}$$

where $d_n = |\mathcal{N}_n|$ is the degree of bus $n \in \mathcal{N}$. Meanwhile, we used the following step-size rule for both DGD and NEXT.

$$\alpha[k+1] = \alpha[k](1 - \gamma\alpha[k]), \forall \, k \in \mathbb{N} \tag{20}$$

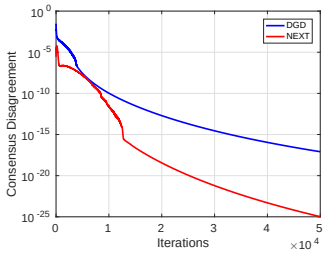
with $\alpha[0] \in (0, 1]$ and $\gamma \in (0, 1)$. Both algorithms were initialized from the flat voltage profile and are run for 5×10^4 iterations. In order to compare the two algorithms, we use the following performance metrics.

$$C[k] := \frac{1}{N}\sum_{n=1}^{N} \|\bar{\mathbf{v}}_n[k] - \bar{\mathbf{v}}_{\text{avg}}[k]\|^2, E[k] := \frac{\|\bar{\mathbf{v}} - \bar{\mathbf{v}}_{\text{avg}}[k]\|_2}{\|\bar{\mathbf{v}}\|_2},$$
$$S[k] := \|\bar{\mathbf{v}}_{\text{avg}}[k] - \mathbf{\Pi}_{\bar{\mathcal{K}}}(\bar{\mathbf{v}}_{\text{avg}}[k] - \nabla F(\bar{\mathbf{v}}_{\text{avg}}[k]))\|_\infty, \forall \, k \in \mathbb{N} \tag{21}$$

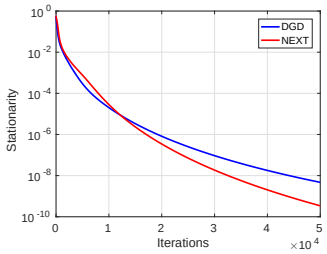
The metric $C[k]$ measures the disagreement amongst the local variables of the agents while $E[k]$ represents the normalized mean square error (NMSE) (here $\bar{\mathbf{v}}$ is the true voltage profile). The last metric $S[k]$ measures the sub-optimality of $\bar{\mathbf{v}}_{\text{avg}}[k]$; we have $S[k] = 0$ if and only if $\bar{\mathbf{v}}_{\text{avg}}[k]$ is a stationary solution of (12).

First, we consider the case of the IEEE-14 bus network in the noiseless regime. Here, we set $\alpha[0] = 10^{-3}$ for DGD and $\alpha[0] = 10^{-3}, \tau = 10$ for NEXT, with $\gamma = 0.1$ for both algorithms. The results are depicted in Figure 1, where it can be observed that both methods attain consensus and converge to stationary solutions. In fact, as evidenced by the figures, globally optimality is attained in this case using only gradient information; with NEXT demonstrating better performance across all metrics. We also carried out additional experiments on larger networks in the noisy case (with independent, zero-mean Gaussian noise with power of 0 and 3 dBm added to the voltage magnitudes and power flow/injections respectively), the results of which are summarized in Tables 1 and 2. Owing to space limitations, we do not explicitly detail the step-sizes, but they were chosen to be slightly less aggressive compared to the noiseless case.

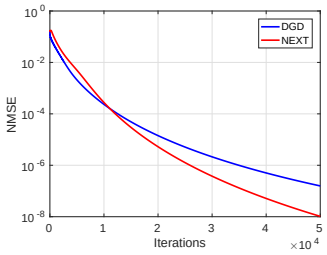
It can be observed that NEXT outperforms DGD in terms of NMSE, while attaining slightly worse solutions with respect to stationarity.



(a) Consensus Disagreement



(b) Quality of Solution



(c) NMSE

Fig. 1: Results for IEEE-14 bus network

6. CONCLUSIONS

We considered the decentralized PSSE problem from the perspective of multi-agent optimization. Two synchronous, gradient-consensus algorithms with low per-iteration complexity (at each agent) were used for obtaining approximate solutions for the resulting non-convex network optimization problem, with guaranteed convergence to stationary solutions. Numerical results demonstrated the superior performance of NEXT over DGD on IEEE benchmark networks. In future, we will use more sophisticated variants of NEXT which utilize Hessian information with the goal of improving convergence speed.

	$C[k]$	$E[k]$	$S[k]$
DGD	8.71×10^{-7}	0.0346	0.0116
NEXT	6.62×10^{-8}	0.0020	0.0131

Table 1: Results for IEEE-57 bus network

	$C[k]$	$E[k]$	$S[k]$
DGD	4.86×10^{-6}	0.1247	0.0339
NEXT	9.16×10^{-6}	0.0352	0.3194

Table 2: Results for IEEE-118 bus network

7. REFERENCES

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

[2] K. Lehmann, A. Grastien, and P. Van Hentenryck, "AC-feasibility on tree networks in NP–Hard", *IEEE Trans. Power Syst.*, vol. 31, no. 1, pp. 798-801, Jan. 2016.

[3] Y. Hu, A. Kuh, T. Yang, and A. Kavcic, "A belief propagation based power distribution system state estimator," *IEEE Computat. Intell. Mag.*, vol. 6, no. 3, pp. 36-46, Aug. 2011.

[4] L. Xie, D.-H. Choi, S. Kar, and H. V. Poor, "Fully distributed state estimation for wide-area monitoring systems," *IEEE Trans. Smart Grid*, vol. 3, no. 3, pp. 1154-1169, May 2012.

[5] X. Li and A. Scaglione, "Robust decentralized state estimation and tracking for power systems via network gossiping," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1184-1194, June 2013.

[6] A. Conejo, S. de la Torre, and M. Canas, "An optimization approach to multiarea state estimation," *IEEE Trans. Power Syst.*, vol. 22, no. 1, pp. 213-221, Feb. 2007.

[7] E. Caro, A. Conejo, and R. Minguez, "Decentralized state estimation and bad measurement identification: An efficient Lagrangian relaxation approach," *IEEE Trans. Power Syst.*, vol. 26, no. 4, pp. 2500-2508, Nov. 2011.

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

[9] Y. Weng, Q. Li, R. Negi, and M. Ilic, "Distributed algorithm for SDP state estimation," in *Proc. of Innovative Smart Grid Tech.*, Washington, D.C., Feb. 2013, pp. 1-6.

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

[11] A. Minot, Y. M. Lu, and N. Li, "A distributed Gauss-Newton method for power system state estimation", *IEEE Trans. Power Syst.*, vol. 31, no. 5, pp. 3804-3815, Sept. 2016.

[12] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. Automat. Control*, vol. 31, no. 9, pp. 803-812, Sep. 1986.

[13] A. Nedic, and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization", *IEEE Trans. Automat. Control,* vol. 54, no. 1, pp. 48–61, Jan. 2009.

[14] A. Nedic, A. Ozdaglar, and P. Parrilo "Constrained consensus and optimization in multi-agent networks", *IEEE Trans. Automat. Control,* vol. 55, no. 4, pp. 922-938, Apr. 2010.

[15] P. Bianchi, and J. Jakubowicz, "Convergence of a multi-agent projected stochastic gradient algorithm for non-convex optimization," *IEEE Trans. Automat. Control,* vol. 58, no. 2, pp. 391-405, Feb. 2013.

[16] P. Di Lorenzo, and G. Scutari, "NEXT: In-network nonconvex optimization," *IEEE Trans. Signal and Info. Process. Networks* , vol. 2, no. 2, pp. 120–136, June 2016.

[17] X. Li and A. Scaglione, "Convergence and applications of a gossip-based Gauss-Newton algorithm," *IEEE Trans. Signal Process.*, vol. 61, no. 21, pp. 5231-5246, Nov. 2013.

[18] D. P. Bertsekas, *Nonlinear Programming*, 3rd ed., Belmont, MA, USA: Athena Scientific, 2016.

[19] D. J. Watts, and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[20] V. D. Blondel, J. M. Hendrickx, A. Olshevsky, and J. N. Tsitsiklis, "Convergence in multiagent coordination, consensus, and flocking," in *Proc. 44th IEEE Conf. Decision Control Eur. Control Conf. (CDC-ECC)*, Seville, Spain, Dec. 12-15, 2005, pp. 2996-3000.

[21] D. S. Scherber, and H. C. Papadopoulos, "Locally constructed algorithms for distributed computations in ad hoc networks," in *Proc. 3rd Int. Symp. Inf. Process. Sensor Netw. (IPSN)*, Berkeley, CA, USA, Apr. 26-27, 2004, pp. 11-19.

[22] L. Xiao, S. Boyd, and S.-J. Kim, "Distributed average consensus with least-mean-square deviation," *J. Parallel Distrib. Comput.*, vol. 67, no. 1, pp. 33-46, Jan. 2007.

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