

# A Privacy Preserving Distributed Model Identification Algorithm for Power Distribution Systems

Chin-Yao Chang

**Abstract**—Distributed control/optimization is a promising approach for network systems due to its advantages over centralized schemes, such as robustness, cost-effectiveness, and improved privacy. However, distributed methods can have drawbacks, such as slower convergence rates due to limited knowledge of the overall network model. Additionally, ensuring privacy in the communication of sensitive information can pose implementation challenges. To address this issue, we propose a distributed model identification algorithm that enables each agent to identify the sub-model that characterizes the relationship between its local control and the overall system outputs. The proposed algorithm maintains the privacy of local agents by only communicating through dummy variables. We demonstrate the efficacy of our algorithm in the context of power distribution systems by applying it to the voltage regulation of a modified IEEE distribution system. The proposed algorithm is well-suited to the needs of power distribution controls and offers an effective solution to the challenges of distributed model identification in network systems.

## I. INTRODUCTION

In recent years, there has been a growing trend towards emphasizing the privacy of users. This has led to the adoption of stricter regulations and policies to protect user privacy, such as the General Data Protection Regulation (GDPR) in Europe [1] and the California Consumer Privacy Act (CCPA) [2] in the United States. As the trend towards emphasizing privacy is likely to continue, how to extract values from data without sacrificing privacy becomes very relevant in many applications such as healthcare, insurance, and FinTech. The technology behind is known as federated learning, which has found success in language model, image recognition, etc [3], [4]. Some recent works [5], [6] also found that federated learning can be useful in developing energy management strategies for future power grids, which could involve millions of controllable devices [7]. However, those are restricted to gaining knowledge of the pattern of certain classes of power consumption or generation, while can not help identify a system-level model (e.g. power flow model), which is very much needed for sophisticated control of future power distribution systems.

C.-Y. Chang is with the National Renewable Energy Laboratory, Golden, CO 80401, USA (Emails: {chinyao.chang}@nrel.gov).

This work was authored in part by NREL, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by DOE Office of Electricity, Advanced Grid Modeling Program, through agreement NO. 33652. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

Distribution systems are mostly unobservable largely because of the cost-ineffectiveness of installing all the sensors. Many existing works are then on identifying some valuable information for grid controls, such as topology of the distribution system [8], [9] and state estimation [10]–[12]. Though distribution system state estimation (DSSE) is about estimating the voltage and currents for buses without sensors, many DSSE still requires some knowledge of the admittance matrix or similar information [11], [12]. The knowledge of the admittances is also valuable for various controls of distribution systems, e.g., [13], [14], but such information is not necessarily easy to obtain, especially considering the data collection hurdles when privacy comes into the equation.

The aforementioned works assume a central entity collects the data from all the sensors in the distribution system for the estimations, which is also the case for estimating the admittance matrix [15]. Collecting all the available data, especially the power consumption patterns of the local users or distributed energy resources can be challenging. Generally speaking, there are some distributed algorithms that can estimate the system model, say the admittance matrix, through distributed communication [16], [17]. However, distributed algorithms mechanically have a consensus element that are usually about the consensus of the estimated models or even the local state variables of the agents, which leaves the valuable information floating on the communication network that could be susceptible for cyber attacks. Another downside of many distributed algorithms is that they are essentially gradient descent that makes them robust over package drops or delays, while the convergence rate is compromised. Adam algorithm (Adaptive Moment Estimation) [18] provides sophisticated step-size (learning rate) selection that can improve the convergence rate of the distributed algorithms. Adam algorithm is also found very effective for large-scale optimizations [19], [20], thus it grows as one of the mainstream algorithms in the machine learning field [21]. Bringing in Adam algorithm elements in distributed algorithms can potentially improve the convergence rate.

*Contributions:* The contribution of the paper is mainly on developing a distributed model identification algorithm such that each agent in the network system identifies the sub-model for the correlation between its local controls and the overall system outputs, which can be understood as a sub-matrix of the LinDistFlow model (or its equivalence) for the distribution system if the inputs are the power injections and outputs are the voltage magnitudes. The proposed algorithm takes elements of Adam algorithm to improve the convergence rate. On top of it, it has an appealing property that the packets exchanged between the agents are dummy

variables so that local data and the identified model are kept private. We demonstrate the effectiveness of the proposed distributed model identification algorithm with a modified IEEE test system.

## II. PRELIMINARY

### A. Notations

For  $x \in \mathbb{R}^n$ , we denote its  $l^2$ -norm and quadratic norm in terms of matrix  $A \succ 0$  by  $\|x\|_2$  and  $\|x\|_A$ , respectively. For matrices  $A_1, A_2, \dots, A_N$  with  $A_i \in \mathbb{R}^{m \times n}$  for all  $i = 1, 2, \dots, N$ , we denote  $\text{blkdiag}(\{A_i\}_{i=1}^N)$  as the block diagonal matrix of all the  $A_i$ ;  $[A_1; A_2; \dots; A_n] \in \mathbb{R}^{Nm \times n}$  and  $[A_1, A_2, \dots, A_n] \in \mathbb{R}^{m \times Nn}$  respectively as the vertical and horizontal concatenations. The Hadamard product and division of  $A_i$  and  $A_j$  are denoted as  $A_i \odot A_j$  and  $A_i./A_j$ , respectively. The Kronecker product between  $A_i$  and  $A_j$  is written as  $A_i \otimes A_j$ . For a matrix  $A \in \mathbb{R}^{m \times n}$ ,  $\text{vec}(A) \in \mathbb{R}^{mn}$  is a column vector that vectorizes  $A$  by concatenating the column vectors of  $A$  from left to right. The kernel of matrix  $A$  is denoted as  $\ker(A)$ . The scalar element of the  $i^{\text{th}}$  row and  $k^{\text{th}}$  column of  $A$  is denoted as  $A(i, k)$ ;  $A(:, k)$  is the  $k^{\text{th}}$  column vector of  $A$ . An identity matrix and all-ones vector with dimension  $n$  are denoted as  $I_n$  and  $\mathbf{1}_n$ , respectively.

### B. System modeling with input-output data

In the following, we briefly overview the data-driven modeling framework which will be used for the remainder of this paper. Let  $u(k) \in \mathbb{R}^m$  and  $y(k) \in \mathbb{R}^n$  be the input and output data at time instant  $k$  for an unknown system. Given the input-output data collected for time  $k = 0, \dots, T-1$ , we define the data matrices shown in the following:

$$U = [\phi_u(u(0)), \dots, \phi_u(u(T-1))], \quad (1a)$$

$$Y = [\phi_y(y(0)), \dots, \phi_y(y(T-1))], \quad (1b)$$

where  $\phi_u : \mathbb{R}^m \rightarrow \mathbb{R}^m$  and  $\phi_y : \mathbb{R}^n \rightarrow \mathbb{R}^n$  are the mappings that capture the known (can be nonlinear) physics of the system to be identified. Assuming that the only unknown part of the targeted system is linear and characterized by  $A \in \mathbb{R}^{n \times m}$ , the following equation holds:

$$Y = AU. \quad (2)$$

The model identification problem with given data ( $U$  and  $Y$ ) can be understood as solving the linear equation (2). If  $U$  has full row rank, then the system model  $A$  is uniquely defined. The straight data representation of the system model with data (2) is useful for controller design purposes, more details are available in [22], [23].

## III. DISTRIBUTED MODEL IDENTIFICATION ALGORITHM

In this section, we will first show how (2) is formulated in a network system, and then lay out how to reformulate it as a distributed optimization problem. We next develop a distributed model identification algorithm leveraging Adam adaptive step-size.

### A. System modeling of network systems

We consider a network system that is partitioned into  $N$  number of regions, and each region has an associated agent that collect all the actuator and sensor data in the region. We assume such a network partitioning setup effectively makes each agent  $i \in \mathcal{N} := \{1, 2, \dots, N\}$  only be able to get information on certain rows of  $U$  and  $Y$  for the purpose of formulating model identification problem (2). Define  $\mathcal{D}_{u,i}$  and  $\mathcal{D}_{y,i}$  as the sets of rows of  $U$  and  $Y$  that are known for agent  $i$ , respectively. We assume the full observability of the system in the sense that  $\cup_{i \in \mathcal{N}} \mathcal{D}_{u,i} = \{1, 2, \dots, m\}$  and  $\cup_{i \in \mathcal{N}} \mathcal{D}_{y,i} = \{1, 2, \dots, n\}$  with  $\mathcal{D}_{u,i} \cap \mathcal{D}_{u,j} = \emptyset$  and  $\mathcal{D}_{y,i} \cap \mathcal{D}_{y,j} = \emptyset$  for all  $i \neq j$  for simplicity. In this setup, the agent  $i$  can capture how its regional controls affect the output  $Y$  by knowing just some columns of  $A$  instead of the full matrix, for which we define the sub-matrix by  $A_{\mathcal{D}_{u,i}} \in \mathbb{R}^{n \times |\mathcal{D}_{u,i}|}$  and without loss of generality,  $A = [A_{\mathcal{D}_{u,1}}, A_{\mathcal{D}_{u,2}}, \dots, A_{\mathcal{D}_{u,N}}]$ . Note that even only  $A_{\mathcal{D}_{u,i}}$  is needed for agent  $i$ , it is achievable to identify and reach consensus on the full model  $A$  for all the agents; however, that can lead to massive data exchanges between the agents that overwhelm the distributed communication network. Therefore, the goal for each agent  $i$  is identifying  $A_{\mathcal{D}_{u,i}}$  by distributed communication and locally available sub-matrices of  $U$  and  $Y$ .

### B. Distributed reformulation of the system modeling

In this section, we go through a series of reformulations of (2) for convenience of distributed algorithm design. Define  $\mathbf{x}_i = \text{vec}(A_{\mathcal{D}_{u,i}})$  and  $\mathbf{x} = [\mathbf{x}_1; \mathbf{x}_2; \dots; \mathbf{x}_n] \in \mathbb{R}^{nm}$ . In this rearrangement, we consider the following formulation of (2):

$$\min_{\mathbf{x}} \frac{1}{2} \sum_{k=1}^T \left\| \left( \sum_{i=1}^N U_{k, \mathcal{D}_{u,i}} \cdot \mathbf{x}_i \right) - Y(:, k) \right\|_2^2, \quad (3)$$

where  $U_{k, \mathcal{D}_{u,i}} \in \mathbb{R}^{n \times n \cdot |\mathcal{D}_{u,i}|}$  is the horizontal concatenation of  $U(i, k) \cdot I_n$  for all  $i \in \mathcal{D}_{u,i}$ . One can verify that the optimal solution of (3) is a solution of (2) by direct algebra. The reason for the optimization formulation is that practically, it is unlikely to find  $\mathbf{x}$  such that (2) holds due to the noisy data, communication disturbances or other disturbances. We next rewrites (3) in a more compact way. Defining  $U_{\mathcal{D}_{u,i}} = [U_{1, \mathcal{D}_{u,i}}; U_{2, \mathcal{D}_{u,i}}; \dots; U_{T, \mathcal{D}_{u,i}}]$ ,  $Y_{\mathcal{D}_{y,i}} = [Y_{1, \mathcal{D}_{y,i}}; Y_{2, \mathcal{D}_{y,i}}; \dots; Y_{T, \mathcal{D}_{y,i}}]$  with  $Y_{k, \mathcal{D}_{y,i}} \in \mathbb{R}^n$  such that

$$Y_{k, \mathcal{D}_{y,i}}(j) = \begin{cases} Y(i, k) & \text{if } j \in \mathcal{D}_{y,i}, \\ 0 & \text{otherwise,} \end{cases}$$

we rewrite (3) as

$$\min_{\mathbf{x}} f(\mathbf{x}), \quad f(\mathbf{x}) := \frac{1}{2} \left\| \sum_{i=1}^N (U_{\mathcal{D}_{u,i}} \mathbf{x}_i - Y_{\mathcal{D}_{y,i}}) \right\|_2^2. \quad (4)$$

The objective function  $f(\mathbf{x})$  couples the variables and data for all the agents, which is not yet solvable with distributed communications. In the following, we first leverage the results in [17] for an algorithm that can solve (4) distributively, followed with some modifications with Adam step-size and illustrating the privacy preserving properties. Define

the graph associated with the distributed communication network as  $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ , where  $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$  is the set of edges (communication links). We next consider the following optimization:

$$\min_{\mathbf{x}, \mathbf{w}} \hat{f}(\mathbf{x}, \mathbf{w}), \quad \hat{f}(\mathbf{x}, \mathbf{w}) := \frac{1}{2} \left\| \hat{U} \mathbf{x} - \hat{Y} - \mathbf{L}^{\frac{1}{2}} \mathbf{w} \right\|_2^2, \quad (5)$$

where  $\mathbf{w} \in \mathbb{R}^{TN}$  is a newly introduced slack variable,  $\hat{U} = \text{blkdiag}(\{\hat{U}_{\mathcal{D}_{u,i}}\}_{i=1}^N)$ ,  $\hat{Y} = [Y_{\mathcal{D}_{y,1}}; Y_{\mathcal{D}_{y,2}}; \dots; Y_{\mathcal{D}_{y,N}}]$ ,  $\mathbf{L} = (L \otimes I_{nT})$ ,  $L$  is the Laplacian matrix associated with  $\mathcal{G}$ , and  $\mathbf{L}^{\frac{1}{2}}$  is the square root of  $\mathbf{L}$ . Lemma III.1 (a compact version of [17, Lemma 3.1]) shows that the optimal solutions of (5) are also the ones for (4) through KKT optimality condition arguments.

**Lemma III.1. (Optimal solutions of (4) and (5)).** *If  $\mathcal{G}$  is connected and  $(\mathbf{x}^*, \mathbf{w}^*)$  is an optimal solution of (5), then  $\mathbf{x}^*$  is an optimal solution of (4).*

*Proof:* By KKT conditions [24],  $(\mathbf{x}^*, \mathbf{w}^*)$  is an optimal solution of (5) if and only if

$$\hat{U}^\top (\hat{U} \mathbf{x}^* - \hat{Y} - \mathbf{L}^{\frac{1}{2}} \mathbf{w}^*) = 0, \quad (6a)$$

$$\mathbf{L}^{\frac{1}{2}\top} (\hat{U} \mathbf{x}^* - \hat{Y} - \mathbf{L}^{\frac{1}{2}} \mathbf{w}^*) = 0. \quad (6b)$$

By (6b) and the property of  $\ker(\mathbf{L}^{\frac{1}{2}}) = \ker(\mathbf{L})$ , we can define  $\mathbf{z}^* = \mathbf{1}_N \otimes \mathbf{z}^* = \hat{U} \mathbf{x}^* - \hat{Y} - \mathbf{L}^{\frac{1}{2}} \mathbf{w}^*$ . We derive the following equation by left multiplying  $\mathbf{z}^*$  by  $\mathbf{1}_N^\top \otimes I_{nT}$ :

$$Nz^* = \sum_{i=1}^N (\hat{U}_{\mathcal{D}_{u,i}} \mathbf{x}_i^* - \hat{Y}_{\mathcal{D}_{y,i}}), \quad (7)$$

where  $\ker(\mathbf{L}^{\frac{1}{2}}) = \ker(\mathbf{L})$  is used again in deriving (7). Substituting (7) to (6a) gives

$$\hat{U}_{\mathcal{D}_{u,j}}^\top \left( \sum_{i=1}^N (\hat{U}_{\mathcal{D}_{u,i}} \mathbf{x}_i^* - \hat{Y}_{\mathcal{D}_{y,i}}) \right) = 0, \quad \forall j \in \mathcal{N}. \quad (8)$$

Because (8) is the KKT condition for the optimality of optimization (4), we conclude that  $\mathbf{x}^*$  is also an optimal solution of (4) and complete the proof.  $\square$

With Lemma III.1, we focus on solving (5) distributively for the original model identification problem.

### C. Distributed algorithm for model identification

The first step of the distributed algorithm for (5) is looking into the gradient descent of (5):

$$\dot{\mathbf{x}} = -\hat{U}^\top (\hat{U} \mathbf{x} - \hat{Y} - \mathbf{L}^{\frac{1}{2}} \mathbf{w}), \quad (9a)$$

$$\dot{\mathbf{w}} = \mathbf{L}^{\frac{1}{2}} (\hat{U} \mathbf{x} - \hat{Y} - \mathbf{L}^{\frac{1}{2}} \mathbf{w}). \quad (9b)$$

Because  $\mathbf{L}^{\frac{1}{2}}$  is not necessarily sparse, (9) can not be directly implemented in a distributed way. Even if we formulate (5) by replacing  $\mathbf{L}^{\frac{1}{2}}$  with  $\mathbf{L}$  (the results of Lemma III.1 still hold), the associated gradient descent with  $\mathbf{L}^{\frac{1}{2}}$  replaced by  $\mathbf{L}$  in (9) still requires packet exchanges between two-hop neighbors, and the local input-output data should be shared between the agents, which is not desirable. By introducing a change of variable

$$\mathbf{z} = \hat{U} \mathbf{x} - \hat{Y} - \mathbf{L}^{\frac{1}{2}} \mathbf{w}, \quad (10)$$

we instead consider the following alternative gradient method:

$$\dot{\mathbf{x}} = -\hat{U}^\top \mathbf{z}, \quad (11a)$$

$$\dot{\mathbf{z}} = -\mathbf{L} \mathbf{z} + \hat{U} \dot{\mathbf{x}} = -(\mathbf{L} + \hat{U} \hat{U}^\top) \mathbf{z}. \quad (11b)$$

Because  $\hat{U}$  is block diagonal with the off-diagonal elements being zeros, the implementation of (11) only requires distributed communication of  $\mathbf{z}$  with  $\mathbf{z}$  being partitioned properly in a way that  $\mathbf{z} = [\mathbf{z}_1; \mathbf{z}_2; \dots; \mathbf{z}_N]$ ,  $\mathbf{z}_i \in \mathbb{R}^{nT}$ . [17] has shown that (11) and its Euler discrete formulation (12) converge to the optimal solution of (5).

$$\mathbf{x}(k+1) = \mathbf{x}(k) - \alpha \hat{U}^\top \mathbf{z}(k), \quad (12a)$$

$$\mathbf{z}(k+1) = \mathbf{z}(k) - \alpha (\mathbf{L} + \hat{U} \hat{U}^\top) \mathbf{z}(k). \quad (12b)$$

To improve the convergence rate of (12), we propose to introduce Adam adaptive step-size to (12). Defining  $D = \mathbf{L} + \hat{U} \hat{U}^\top$  and  $N_I$  as the number of iterations for the model identification, we propose the distributed model identification algorithm with Adam adaptive step-size in Algorithm 1.

---

**Algorithm 1** Distributed Model Identification Algorithm with Adam adaptive step-size

---

**Require:**  $\beta_1, \beta_2 \in [0, 1)$ ,  $\epsilon > 0$ ,  $k = 0$

**Require:** Initialize  $s_1(0) = s_2(0) = \hat{s}_1(0) = \hat{s}_2(0) = 0$ , and  $\mathbf{x}(0)$ ,  $\mathbf{w}(0)$ ,  $\mathbf{z}(0)$  that satisfies (10),

1: **while**  $k \leq N_I - 1$  **do**

2:    $k \leftarrow k + 1$

3:    $g(k) \leftarrow \hat{U}^\top \mathbf{z}(k-1)$

4:    $s_1(k) \leftarrow \beta_1 s_1(k-1) + (1 - \beta_1) g(k)$

5:    $s_2(k) \leftarrow \beta_2 s_2(k-1) + (1 - \beta_2) (g(k) \odot g(k))$

6:    $\hat{s}_1(k) \leftarrow s_1(k) / (1 - \beta_1^k)$

7:    $\hat{s}_2(k) \leftarrow s_2(k) / (1 - \beta_2^k)$

8:    $\mathbf{x}(k) \leftarrow \mathbf{x}(k-1) - \alpha (\hat{s}_1(k) / (\sqrt{\hat{s}_2(k)} + \epsilon))$

9:    $\mathbf{z}(k) \leftarrow \mathbf{z}(k-1) - \alpha D \mathbf{z}(k-1)$

10: **end while**

---

Note that the adaptive step-size is only used on the  $\mathbf{x}$  dynamics in line 8 but not on  $\mathbf{z}$  in line 9 because such a adaptive step-size on  $\mathbf{z}$  can break the consensus established through  $\mathbf{L}$ . Although we do not conclude a faster convergence rate in the proof of convergence of Algorithm 1 stated in Theorem III.2, empirical results suggest faster convergence rate of Adam algorithm over gradient descent method generally [21], and we observe similar results in our simulation studies.

**Theorem III.2. (Convergence of Algorithm 1).** *Algorithm 1 has  $\mathbf{x}(k) \rightarrow \mathbf{x}^*$  as  $k \rightarrow \infty$  if  $\|\hat{U}\|_2 < \infty$ ,  $\|\hat{Y}\|_2 < \infty$ ,  $\beta_1, \beta_2 \in [0, 1)$  and  $\alpha$  is chosen such that*

$$(\alpha^2 + \mu) D^\top P D - \alpha (D^\top P + P D) \preceq 0 \quad (13)$$

for some  $P \succ 0$  and  $\mu > 0$ .

*Proof:* With line 9 of Algorithm 1, (13) implies

$$\begin{aligned} \mathbf{z}^\top(k) P \mathbf{z}(k) - \mathbf{z}^\top(k-1) P \mathbf{z}(k-1) \\ \leq -\mu \mathbf{z}^\top(k-1) D^\top P D \mathbf{z}(k-1), \end{aligned}$$

$$\Rightarrow \|\mathbf{z}(k)\|_P - \|\mathbf{z}(k-1)\|_P \leq -\mu \|D\mathbf{z}(k-1)\|_P. \quad (14)$$

We next sum up (14) for  $k = 1, \dots, N_I$ , leading to

$$\|\mathbf{z}(N_I)\|_P \leq \|\mathbf{z}(0)\|_P - \sum_{k=0}^{N_I-1} \mu \|D\mathbf{z}(k)\|_P. \quad (15)$$

Inequality (15) implies that  $\|D\mathbf{z}(k)\|_P \rightarrow 0$  as  $k \rightarrow \infty$ . Because  $D = \mathbf{L} + \hat{U}\hat{U}^\top$ , and  $\mathbf{L}$  and  $\hat{U}\hat{U}^\top$  are positive semidefinite, for any  $\mathbf{z}(k) \neq 0$ ,  $\|D\mathbf{z}(k)\|_P = 0$  if and only if  $\mathbf{L}\mathbf{z}(k) = \hat{U}\hat{U}^\top\mathbf{z}(k) = 0$ . Recall that  $\mathbf{z}^* := \hat{U}\mathbf{x}^* - \hat{Y} - \mathbf{L}^{\frac{1}{2}}\mathbf{w}^*$  is the optimal solution if and only if (6) holds, so we can conclude  $\mathbf{z}(k) \rightarrow \mathbf{z}^*$  by  $\|D\mathbf{z}(k)\|_P \rightarrow 0$ . Because  $\|\mathbf{z}^*\|_2 < \infty$  and  $\mathbf{L}^{\frac{1}{2}}\mathbf{z}^* = 0$ , for any  $\mathbf{x}(k)$  with  $\|\mathbf{x}(k)\|_2 < \infty$ , there exists a  $\mathbf{w}(k)$  that solves

$$\mathbf{z}^* = \hat{U}\mathbf{x}(k) - \hat{Y} - \mathbf{L}^{\frac{1}{2}}\mathbf{w}(k).$$

The  $\mathbf{x}(k)$  and  $\mathbf{w}(k)$  that solve the equation above are the optimal solutions of (5). In other words, we can conclude  $\mathbf{x}(k) \rightarrow \mathbf{x}^*$  given  $\mathbf{z}(k) \rightarrow \mathbf{z}^*$  as  $k \rightarrow \infty$  by showing that  $\|\mathbf{x}(k)\|_2 < \infty$  for all  $k \in \mathbb{N}$ , or

$$\|\mathbf{x}(k_1) - \mathbf{x}(k_2)\|_2 < \infty, \quad \forall k_1, k_2 \in \mathbb{N} \quad (16)$$

To show (16) holds, we first state that there exists a  $C_g < \infty$  such that

$$\sum_{k=1}^{\infty} \|\hat{U}^\top \mathbf{z}(k)\|_2 = \sum_{k=1}^{\infty} \|g(k)\|_2 \leq C_g. \quad (17)$$

Otherwise, (15) does not hold when  $N_I \rightarrow \infty$ . We next analyze the series of  $\hat{s}_1(k) / (\sqrt{\hat{s}_2(k)} + \epsilon)$  that appears in the updates of  $\mathbf{x}(k)$  in Algorithm 1:

$$\begin{aligned} & \left\| \sum_{k=1}^{\infty} \left( \hat{s}_1(k) / (\sqrt{\hat{s}_2(k)} + \epsilon) \right) \right\|_2 \leq \frac{1}{\epsilon} \left\| \sum_{k=1}^{\infty} \hat{s}_1(k) \right\|_2 \\ &= \frac{1}{\epsilon} \left\| \sum_{k=1}^{\infty} \frac{\beta_1^k s_1(0) + (1 - \beta_1) \sum_{i=1}^k \beta_1^{k-i} g(i)}{1 - \beta_1^k} \right\|_2, \quad (18) \end{aligned}$$

where line 4 of Algorithm 1 is used in deriving (18). Because  $s_1(0) = 0$ , we can simplify (18) to:

$$\begin{aligned} & \left\| \sum_{k=1}^{\infty} \left( \hat{s}_1(k) / (\sqrt{\hat{s}_2(k)} + \epsilon) \right) \right\|_2 \\ & \leq \frac{1 - \beta_1}{\epsilon} \left\| \sum_{k=1}^{\infty} \frac{1}{1 - \beta_1^k} \sum_{i=1}^k \beta_1^{k-i} g(i) \right\|_2 \\ &= \frac{1 - \beta_1}{\epsilon} \left\| \sum_{j=1}^{\infty} g(j) \cdot \left( \sum_{h=j}^{\infty} \frac{\beta_1^{h-j}}{1 - \beta_1^h} \right) \right\|_2 \\ & \leq \frac{1 - \beta_1}{\epsilon} \cdot \frac{1}{(1 - \beta_1)^2} \sum_{j=1}^{\infty} \|g(j)\|_2 \leq \frac{C_g}{\epsilon(1 - \beta_1)} < \infty \quad (19) \end{aligned}$$

The inequality (19) implies (16) holds, which completes the proof.  $\square$

**Remark III.3. (Comparison of Algorithm 1 with Adam algorithm).** By dropping line 9 of Algorithm 1 and changing the meaning of  $g(k)$  from  $\hat{U}^\top \mathbf{z}(k-1)$  to the gradient of

a certain objective function of  $\mathbf{x}$ , Algorithm 1 is actually Adam algorithm. Such a change leads in very different routes to prove the convergence to the optimal  $\mathbf{x}^*$ . However, the fundamental assumptions are similar. The convergence statement of Adam algorithm [18, Theorem 10.5] straight assumes (16) and (17) hold, and for Algorithm 1, such an assumption is indirectly embedded in (13), which actually leads to (16) and (17) as shown in the proof. Last but not least, the convergence statement of Adam algorithm [18, Theorem 10.5] has additional assumptions on  $\beta_1$  and  $\beta_2$ , which are used for deriving the bounds of the regret (or characterizing the convergence rate). Because Algorithm 1 does not use  $\beta_1$  and  $\beta_2$  for all the state variables (it only uses them in updating  $\mathbf{x}$ ), it is unclear on how to make additional assumptions on  $\beta_1$  and  $\beta_2$  to help characterize the convergence rate of Algorithm 1. Practically, choosing  $\beta_1$  and  $\beta_2$  for Algorithm 1 based on numerical experiences of Adam algorithm in the literature works reasonably well.

We conclude this section by noting that Algorithm 1 possesses some desirable properties: (1) only the dummy variables  $\mathbf{z}$  are exchanged through the communication network; (2) retrieving the sub-model,  $\mathbf{x}_i$  (or  $A_{\mathcal{D}_{u,i}}$ ), that is useful for agent  $i$  from  $\mathbf{z}$  requires local data  $U_{\mathcal{D}_{u,i}}$  and  $Y_{\mathcal{D}_{u,i}}$ , which simultaneously preserves the privacy and enhances the security. We view those as the unique strengths of Algorithm 1 compared to other model identification algorithms.

#### IV. NUMERICAL STUDIES

We validate the proposed distributed model identification algorithm with two test cases. We first demonstrate that by adding Adam algorithm adaptive step-size, a faster convergence rate is observed for Algorithm 1 compared to (12). We next apply the proposed algorithm to identify the LinDistFlow model [25] (or its equivalence) for a modified IEEE 37 buses system, and show a satisfactory control performance by leveraging the identified model.

##### A. A small-scale example

We randomly generate a linear networked system given as

$$\mathbf{y} = [y_1; y_2; \dots; y_N] = \sum_{i=1}^N A_{\mathcal{D}_{u,i}} u_i, \quad (20)$$

where  $u_i \in \mathbb{R}^{|\mathcal{D}_{u,i}|}$  and  $y_i \in \mathbb{R}^{|\mathcal{D}_{y,i}|}$  are the input and output of agent  $i$ ,  $|\mathcal{D}_{u,i}| \in \{4, 5\}$ ,  $|\mathcal{D}_{y,i}| \in \{3, 4\}$ , and  $N = 5$ . All the entries of matrix  $A = [A_{\mathcal{D}_{u,1}}, A_{\mathcal{D}_{u,2}}, \dots, A_{\mathcal{D}_{u,N}}]$  and  $u_i$  are randomly generated with the standard normal distribution with the standard deviation of 1. We assume that the model identification algorithm (12) and Algorithm 1 start when the data matrices  $U$  and  $Y$  are constructed with  $T$  steps,  $T = 2 \sum_{i=1}^N |\mathcal{D}_{u,i}|$ . The algorithmic parameters are set as  $\alpha = 10^{-3}$ ,  $\beta_1 = 0.9$ ,  $\beta_2 = 0.95$ , and  $\epsilon = 10^{-8}$ . As shown in Figures 1 and 2, both (12) and Algorithm 1 have the estimated model converges to the actual ones. Algorithm 1 has a noticeably faster convergence rate, especially at the first hundreds iterations. We conjecture the dominating factor for the convergence after few hundred steps is on the consensus of  $\mathbf{z}$ , so there is no much difference between (12) and Algorithm 1 afterward.

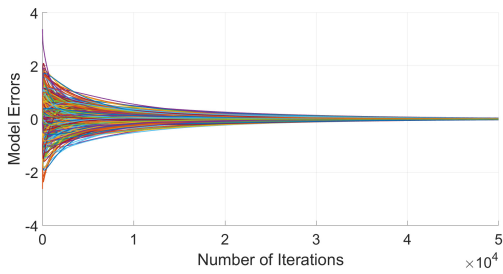


Fig. 1. The absolute value difference for each element of the estimated  $A$  and  $A^*$  by implementing (12).

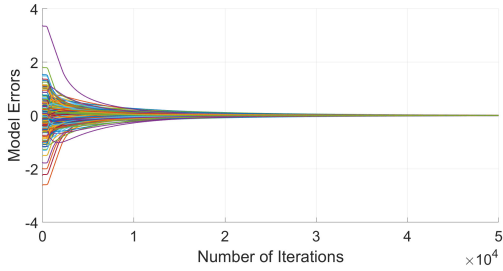


Fig. 2. The absolute value difference for each element of the estimated  $A$  and  $A^*$  for Algorithm 1.

### B. A modified IEEE 37-bus test system

In this section, we apply Algorithm 1 to identify the LinDistFlow model  $A$  for the voltage magnitude regulation purpose. We modify the IEEE 37-bus test system with penetration of PVs as illustrated in Figure 3. We assume that every PV bus acts as an agent and knows its local active/reactive power injections and the voltage magnitude of the bus, and it can adjust its power injections to help regulate the voltage magnitudes. The power injections of PV bus  $i$  is collected as the control variable  $u_i \in \mathbb{R}^2$  and the voltage magnitude measurements are collected by  $y_i \in \mathbb{R}_+$ . We assume that a connected distributed communication network  $\mathcal{G}$  is established for the purposes of running Algorithm 1 and the distributed feedback-based algorithm developed in [14].

Implementing the distributed feedback-based control algorithm developed in [14] requires a LinDistFlow model  $A$  or

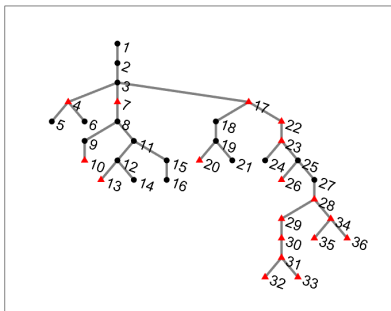


Fig. 3. Illustration of the modified IEEE 37-bus system with the buses highlighted in red triangles are PV buses

its equivalence. Although the model  $A$  does not need to be very accurate as explained in [26], a reasonably accurate  $A$  is still needed. We simulate the following scenarios: (i) no control on the voltage magnitudes; (ii) distributed feedback-based control with  $A$  derived by the given knowledge of line impedances and network topology; (iii) distributed feedback-based control with  $A$  identified by Algorithm 1 offline; (iv) distributed feedback-based control with  $A$  identified by Algorithm 1 online whenever the new pair of  $u$  and  $y$  is available to replace the oldest pair. The algorithmic parameters  $\alpha$ ,  $\beta_1$ ,  $\beta_2$  and  $\epsilon$  are set the same as the first numerical example for Algorithm 1. The number of time instances of constructing input and output data matrices,  $U$  and  $Y$ , is set by  $T = 140$ .

Figures 4-7 illustrate the voltage magnitudes for the four scenarios. Unsurprisingly, the control performance with the known  $A$  is among the bests. An interesting observation is that for scenario (iii), the  $A$  concatenated by the identified  $A_{\mathcal{D}_{u,i}}$  for all the agents is very different from the  $A$  derived from the given impedance and topology, regardless how we initialize  $A_{\mathcal{D}_{u,i}}$  (or  $x_i$ ). However, the concatenated  $A$  still matches the data well with the per unit voltage estimation errors being at the order of  $10^{-3}$  ( $10^{-4}$  for the LinDistFlow  $A$ ), and a similar voltage regulation result to scenario (ii) is achieved. Our explanation is that another form of linearized power flow is identified through Algorithm 1 because there can be many valid linearized models for the underlying nonlinear power flow physics. Scenario (iv) is expected to perform best because it adapts the model  $A$  in accordance with the operating points online (could be understood as adjusting the linearization point). However, as shown in Figure 7, we observe a bit more fluctuation of the regulated voltage magnitudes compared to scenarios (ii) and (iii). Some fine-tuning of the number of steps before updating  $U$  and  $Y$  may be needed for a better performance. Overall, both scenarios (iii)-(iv) achieve satisfactory voltage regulations while keeping the power injections and voltage magnitude measurements local, and the packet exchanges are limited to dual variables associated with voltage constraints and dummy variables for the distributed algorithms.

### V. CONCLUSION

In this paper, we proposed a distributed model identification algorithm such that each agent identifies a sub-model that describes its local controls with the overall system outputs. The algorithm is designed such that the agents do not need to share their local data and the convergence rate is practically better than the gradient descent consensus algorithm because of adding the adaptive step-size of the Adam algorithm. We focus on the power distribution system applications in this paper, but we envision the proposed algorithm is potentially useful for decentralized or distributed controls in many other networked systems such as robotics, economics, telecommunications. Our future works will be on refining the online distributed model identification for improved linearized models for nonlinear or time-varying linear systems. More in-depth testing of the proposed algorithm for some potential applications is also among our future works.

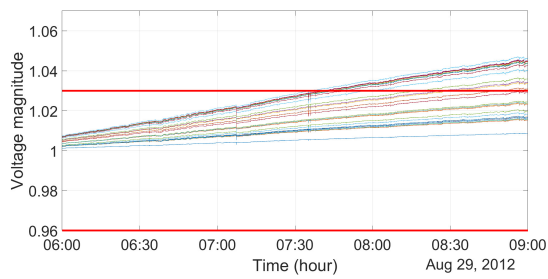


Fig. 4. The voltage magnitudes (p.u.) over time without control.

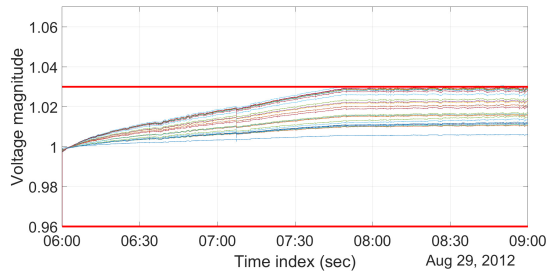


Fig. 5. The voltage magnitudes (p.u.) over time with distributed feedback-based control and known LinDistFlow model.

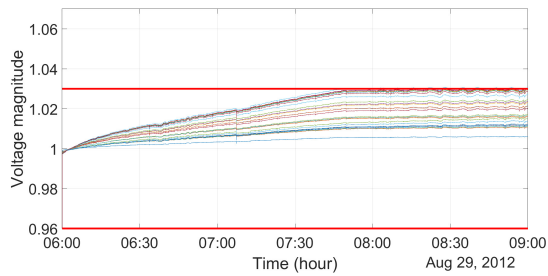


Fig. 6. The voltage magnitudes (p.u.) over time with distributed feedback-based control and the identified LinDistFlow model.

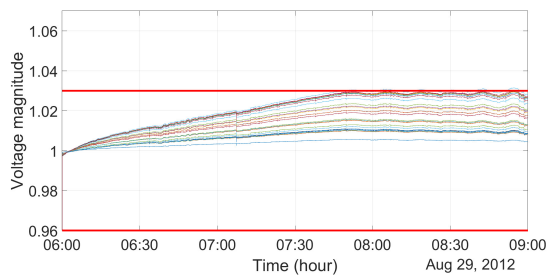


Fig. 7. The voltage magnitudes (p.u.) over time with distributed feedback-based control and the identified LinDistFlow model updated online.

## REFERENCES

[1] P. Voigt and A. Von dem Bussche, "The EU general data protection regulation (GDPR)," *A Practical Guide, 1st Ed.*, Cham: Springer International Publishing, vol. 10, no. 3152676, pp. 10–5555, 2017.

[2] P. Bukaty, *The California Consumer Privacy Act (CCPA): An Implementation Guide*. IT Governance Ltd, 2019.

[3] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," *arXiv preprint arXiv:1610.05492*, 2016.

[4] L. Li, Y. Fan, M. Tse, and K.-Y. Lin, "A review of applications in federated learning," *Computers & Industrial Engineering*, vol. 149, p. 106854, 2020.

[5] Y. Wang, I. L. Bennani, X. Liu, M. Sun, and Y. Zhou, "Electricity consumer characteristics identification: A federated learning approach," *IEEE Transactions on Smart Grid*, vol. 12, no. 4, pp. 3637–3647, 2021.

[6] J. Lin, J. Ma, and J. Zhu, "A privacy-preserving federated learning method for probabilistic community-level behind-the-meter solar generation disaggregation," *IEEE Transactions on Smart Grid*, vol. 13, no. 1, pp. 268–279, 2021.

[7] B. Kroposki, A. Bernstein, J. King, D. Vaidhyanathan, X. Zhou, C.-Y. Chang, and E. Dall'Anese, "Autonomous energy grids: Controlling the future grid with large amounts of distributed energy resources," *IEEE Power and Energy Magazine*, vol. 18, no. 6, pp. 37–46, 2020.

[8] S. Bolognani, N. Bof, D. Michelotti, R. Muraro, and L. Schenato, "Identification of power distribution network topology via voltage correlation analysis," in *52nd IEEE Conference on Decision and Control*, 2013, pp. 1659–1664.

[9] O. Ardakanian, V. W. Wong, R. Dobbe, S. H. Low, A. von Meier, C. J. Tomlin, and Y. Yuan, "On identification of distribution grids," *IEEE Transactions on Control of Network Systems*, vol. 6, no. 3, pp. 950–960, 2019.

[10] A. Primadianto and C.-N. Lu, "A review on distribution system state estimation," *IEEE Transactions on Power Systems*, vol. 32, no. 5, pp. 3875–3883, 2016.

[11] M. C. de Almeida and L. F. Ochoa, "An improved three-phase amb distribution system state estimator," *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 1463–1473, 2016.

[12] Y. Guo, X. Zhou, C. Zhao, Y. Chen, T. Summers, and L. Chen, "Solving optimal power flow for distribution networks with state estimation feedback," in *2020 American Control Conference (ACC)*, 2020, pp. 3148–3155.

[13] R. H. Salim and R. A. Ramos, "A model-based approach for small-signal stability assessment of unbalanced power systems," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 2006–2014, 2012.

[14] C.-Y. Chang, M. Colombino, J. Cortés, and E. Dall'Anese, "Saddle-flow dynamics for distributed feedback-based optimization," *IEEE Control Systems Letters*, 2019.

[15] J. Zhang, P. Wang, and N. Zhang, "Distribution network admittance matrix estimation with linear regression," *IEEE Transactions on Power Systems*, vol. 36, no. 5, pp. 4896–4899, 2021.

[16] P. Stegagno and C. Yuan, "Distributed cooperative adaptive state estimation and system identification for multi-agent systems," *IET Control Theory & Applications*, vol. 13, no. 6, pp. 815–822, 2019.

[17] Y. Huang, Z. Meng, and J. Sun, "Scalable distributed least square algorithms for large-scale linear equations via an optimization approach," *Automatica*, vol. 146, p. 110572, 2022.

[18] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[19] T. Chilimbi, Y. Suzue, J. Apacible, and K. Kalyanaraman, "Project adam: Building an efficient and scalable deep learning training system," in *11th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 14)*, 2014, pp. 571–582.

[20] C. Jose and F. Fleuret, "Scalable metric learning via weighted approximate rank component analysis," in *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part V 14*. Springer, 2016, pp. 875–890.

[21] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv preprint arXiv:1609.04747*, 2016.

[22] C. De Persis and P. Tesi, "Formulas for data-driven control: Stabilization, optimality, and robustness," *IEEE Transactions on Automatic Control*, vol. 65, no. 3, pp. 909–924, 2019.

[23] C.-Y. Chang and A. Bernstein, "Robust data-driven control for systems with noisy data," *arXiv preprint arXiv:2207.09587*, 2022.

[24] A. Ben-Tal and A. Nemirovski, *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. SIAM, 2001.

[25] M. E. Baran and F. F. Wu, "Network reconfiguration in distribution systems for loss reduction and load balancing," *IEEE Transactions on Power delivery*, vol. 4, no. 2, pp. 1401–1407, 1989.

[26] M. Colombino, J. W. Simpson-Porco, and A. Bernstein, "Towards robustness guarantees for feedback-based optimization," *arXiv preprint arXiv:1905.07363*, 2019.