

Learning partially observed meshed distribution grids

Harish Doddi[‡], Deepjyoti Deka[†], Murti Salapaka[‡]

([‡]) Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, USA

([†]) Theoretical Division, Los Alamos National Laboratory, New Mexico, USA

Abstract—This article analyzes statistical learning methods to identify the topology of meshed power distribution grids under partial observability. The learning algorithms use properties of the probability distribution of nodal voltages collected at the observed nodes. Unlike prior work on learning under partial observability, this work does not presume radial structure of the grid, and furthermore does not use injection measurements at any node. To the best of our knowledge, this is the first work for topology recovery in partially observed distribution grids, that uses voltage measurements alone. The developed learning algorithms are validated with non-linear power flow samples generated by Matpower in test grids.

Keywords—distribution grids, topology, statistical learning, low rank, sparsity.

I. INTRODUCTION

The power grid's topology is determined by the status of operational breakers/switches that have been turned on (see Fig. 1). Learning the operational structure of power grids is an important problem for situational awareness, network constrained control operations, and verification of operator actions [1]. The learning problem is complicated in low voltage power distribution grids that serve end-users due to the traditionally limited presence of real-time meters, in particular of line and breaker statuses. In recent years, due to modernization efforts at the edge of the grid, such problems have received significant attention. To overcome issues with limited line-based measurements, researchers have proposed the use of nodal measurements such as bus voltage and injection, for topology estimation. These measurements have become available in the smart grid era, due to the installation of micro-PMUs [2], sensors on smart devices, and other advanced meters [3]. Among others, statistical properties of the time-series of grid bus voltages were harvested for topology estimation in prior work [4], [5], [6]. However, a majority of these learning efforts, in particular the ones with consistency results, require voltage data at *all nodes*. This may not be practical, as real-time metering, though expanding, is still not ubiquitous [3]. In scenarios with low nodal observability, voltage measurements supplemented with injection data have been used for topology estimation in radial grids. Examples of such efforts include passive statistical methods [7], [8], [9] and active probing analysis [10], [11]. However the problem of topology identification using *only* partially observed voltage data remains largely unresolved. While current distribution grids are radial [7], urban grids [12], [13] are being proposed to be meshed/loopy in the near future for reliability considerations. It is thus imperative for learning methods to be extensible to non-radial grids. Few existing work, aside from [6], [14] that require complete voltage observability, analyze learning in non-radial/meshed grids. The goal of this work is to develop algorithms for estimating distribution grids that overcome shortcomings in

prior literature, by considering a difficult regime under the following conditions: (a) partial observability of voltages, (b) absence of any load measurements, and (c) meshed structure.

A. Contribution

Our learning algorithm for topology recovery is based on statistics of nodal voltages obtained from observed nodes in the grid. To model the statistics, we investigate the *empirical inverse covariance matrix* of available voltages in the partially observed meshed distribution grid. Unlike prior work on complete observability [4], the partial inverse covariance is dense and cannot directly be used for topology inference. Using the ac power flow equations, we show that the partial inverse can be decomposed into low rank and sparse submatrices, and analyze their algebraic properties. We use these properties to design a novel learning algorithm that estimates the true topology in meshed grids with *cycle length greater than four*, provided any two unobserved nodes are *greater than two hops away*. We further show that the separation between unobserved nodes can be reduced to *two hops* for radial grids. To the best of our knowledge, this is the first work to discuss topology learning using only voltage measurements for partially observed grids (both radial or meshed). Simulations with linear and non-linear ac power flow samples demonstrate the performance of the learning algorithm on IEEE test cases.

The rest of the paper is organized as follows. The next section presents power flow relations and the corresponding statistical measures used. Section III discusses properties of the partially observed voltage covariance matrix and its inverse. Section IV includes the development of our topology learning algorithm for partially observed systems. Section V includes detailed simulations results of our work on test cases. Conclusions and future work are included in Section VI.

II. POWER GRID BASICS

Notation: We represent a power grid by graph, $\mathcal{G} \equiv (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{i, j, \dots\}$ is the set of $N + 1$ buses/nodes, and $\mathcal{E} = \{(ij), (jk), \dots\}$ is the set of operational undirected lines/edges that need to be discovered (see Fig. 1). A node's neighbors are the nodes it shares an edge with. We define path $\mathcal{P}_i^j \equiv \{(ik_1), (k_1k_2), \dots, (k_{n-1}j)\}$ to be a set of n distinct undirected edges that connect node i and node j . If $i = j$ and path $\mathcal{P}_i^j$ has more than two edges, we term it a 'cycle'. Non-neighbor nodes that are separated by paths of length two are termed 'two-hop neighbors'. For the remaining of the paper, we include the following non-restrictive assumption on the topology. The minimum size of a cycle in the grid is called its 'minimum cycle length'.

Assumption 1: The minimum cycle length in grid $\mathcal{G}$ is *five*.

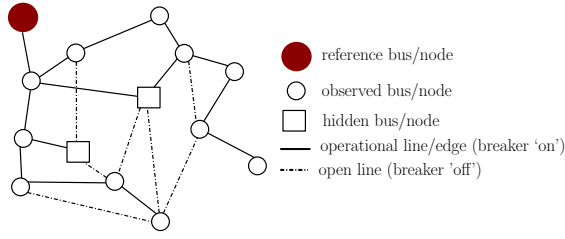


Fig. 1. Meshed distribution grid with a substation, and open and operational lines. Our learning algorithm does *not* require knowledge of the set of permissible lines (open or closed).

Assumption 1 is true for almost all practical and test networks that are either radial (infinite cycle length) or have large girth with few cycles [12], [13]. It is worth mentioning that a minimum cycle length of four is necessary for unique estimation even under complete voltage observability [15].

For line (ij) , let r_{ij} , x_{ij} , and $z_{ij} = r_{ij} + \hat{i}x_{ij}$ denote its resistance, reactance and impedance, respectively. The non-linear power flow equation at each node $i \in \mathcal{V}$ is given by:

$$P_i = p_i + \hat{i}q_i = \sum_{j:(ij) \in \mathcal{E}} \frac{v_i^2 - v_i v_j \exp(\hat{i}\theta_i - \hat{i}\theta_j)}{z_{ij}^*}. \quad (1)$$

Here real valued v_i , θ_i , p_i and q_i denote the voltage magnitude, voltage phase, active and reactive power injections, respectively, at node i . During normal/ambient operation, the second order behavior in the power flow equations can be ignored to obtain the *Linearized AC power flow* (LC-PF) equation [7], [16], [17]:

$$P_i = p_i + \hat{i}q_i = \sum_{j:(ij) \in \mathcal{E}} \frac{(v_i - v_j) - \hat{j}(\theta_i - \theta_j)}{z_{ij}^*}. \quad (2)$$

Mathematically, Eq. 2 is derived by taking small deviations in phase difference of neighboring nodes ($|\theta_i - \theta_j| \ll 1$ for edge (ij)), and voltage magnitude deviations from the reference voltage ($|v_i - 1| \ll 1$ for node i). It is lossless as injection at the substation is equal to the negative sum of the other buses. Without a loss of generality, the reference bus is ignored and the voltages at all other buses are measured relative to it. The power flow equations are hence restricted to the N non-reference buses in the grid, and represent the following invertible map between voltage vectors v, θ and injection vectors p, q :

$$\begin{bmatrix} v \\ \theta \end{bmatrix} = H_{(g,\beta)}^{-1} \begin{bmatrix} p \\ q \end{bmatrix}, \text{ where } H_{(g,\beta)} = \begin{bmatrix} H_g & H_\beta \\ H_\beta & -H_g \end{bmatrix} \quad (3)$$

Here H_β (H_g) is the reduced weighted Laplacian matrix of $\mathcal{G}$ with edge weights given as susceptances β (conductances g). By definition, $g_{ij} + \hat{i}\beta_{ij}$ equals $\frac{1}{r_{ij} - \hat{i}x_{ij}}$. We have

$$H_g(i, j) = \begin{cases} \sum_{k:(ik) \in \mathcal{E}} g_{ik} & \text{if } i = j \\ -g_{ij} & \text{if } (ij) \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases}. \quad (4)$$

The reduction is derived by removing the row and column corresponding to the reference bus from H_β (H_g). It is worth mentioning that ignoring voltage magnitude deviations and reactive power converts LC-PF to the standard *DC power flow* (DC-PF) model [18]: $H_\beta \theta_{DC} = p$.

All our analysis naturally extends to DC-PF as well, as mentioned in comments after important results.

Voltage Statistics in the Grid: As the LC-PF Eq. 3 and its DC reduction represent an invertible map between voltages and injections at the non-substation buses, we can relate centered covariance matrix of bus voltages $\Sigma_{(v,\theta)}$ for LC-PF with covariance of injection fluctuations $\Sigma_{(p,q)}$ through

$$\Sigma_{(v,\theta)} = \mathbb{E} \begin{bmatrix} v \\ \theta \end{bmatrix} \begin{bmatrix} v \\ \theta \end{bmatrix}^T = \begin{bmatrix} \Sigma_{vv} & \Sigma_{v\theta} \\ \Sigma_{\theta v} & \Sigma_{\theta\theta} \end{bmatrix} = H_{(g,\beta)}^{-1} \Sigma_{(p,q)} H_{(g,\beta)}^{-1}. \quad (5)$$

In the remaining of the paper, we assume the following condition on injection covariance matrix $\Sigma_{(p,q)}$.

Assumption 2: Fluctuations in nodal injections in the grid are uncorrelated, i.e., $\Sigma_{(p,q)} = \begin{bmatrix} \Sigma_{pp} & \Sigma_{pq} \\ \Sigma_{qp} & \Sigma_{qq} \end{bmatrix}$ where each of the block matrices is diagonal.

The assumption of uncorrelated injection fluctuations has been verified for loads and renewables, and used for theoretical analysis in prior literature [19], [6], [7]. Intuitively, the uncorrelation arises from independent load fluctuations in distribution grids, once small linear trends are empirically detrended, as observed in prior work [20], [6]. Under the block diagonal structure of injection covariance in Assumption 2, the voltage covariance matrix $\Sigma_{(v,\theta)}$ has the following inverse [15],

$$\Sigma_{(v,\theta)}^{-1} = \begin{bmatrix} J_{vv} & J_{v\theta} \\ J_{v\theta}^T & J_{\theta\theta} \end{bmatrix} \text{ where} \quad (6)$$

$$\begin{aligned} J_{vv} &= H_g D^{-1} (\Sigma_{qq} H_g - \Sigma_{pq} H_\beta) - H_\beta D^{-1} (\Sigma_{pq} H_g - \Sigma_{pp} H_\beta), \\ J_{v\theta} &= H_g D^{-1} (\Sigma_{qq} H_\beta + \Sigma_{pq} H_g) - H_\beta D^{-1} (\Sigma_{pq} H_\beta + \Sigma_{pp} H_g), \\ J_{\theta\theta} &= H_\beta D^{-1} (\Sigma_{qq} H_\beta + \Sigma_{pq} H_g) + H_g D^{-1} (\Sigma_{pq} H_\beta + \Sigma_{pp} H_g), \\ D(i, i) &= \Sigma_{pp}(i, i) \Sigma_{qq}(i, i) - \Sigma_{pq}^2(i, i) \text{ for diagonal } D. \end{aligned}$$

Note: For DC-PF, we have the following relation regarding the voltage covariance $\Sigma_{\theta_{DC}}$:

$$\Sigma_{\theta_{DC}} = H_\beta^{-1} \Sigma_{pp} H_\beta^{-1} \Rightarrow J_{\theta_{DC}} = \Sigma_{\theta_{DC}}^{-1} = H_\beta \Sigma_{pp}^{-1} H_\beta. \quad (7)$$

The inverse covariance matrix $\Sigma_{(v,\theta)}^{-1}$ has two specific algebraic properties, that follow from Eq. 6.

Theorem 1 (Theorems 3, 6 in [15]). *Consider inverse covariance matrix of voltages $\Sigma_{(v,\theta)}^{-1} = \begin{bmatrix} J_{vv} & J_{v\theta} \\ J_{v\theta}^T & J_{\theta\theta} \end{bmatrix}$ in grid $\mathcal{G}$ under*

Assumptions 1 and 2. Then,

- (a) $J_{vv}(i, j) \neq 0$, $J_{v\theta}(i, j) \neq 0$, $J_{\theta\theta}(i, j) \neq 0$ if and only if i, j are neighbors or two-hop neighbors.
- (b) $J_{vv}(i, j) + J_{\theta\theta}(i, j) < 0$ if and only if (ij) is an edge.

The proof of both statements follows directly by algebraic expansion of $\Sigma_{(v,\theta)}^{-1}$ using Eq. 6. The first statement implies that non-zero terms in J_{vv} or $J_{\theta\theta}$ create an inverse covariance graph that includes true edges in $\mathcal{G}$ and ‘spurious’ edges between two hop neighbors (see Fig. 2(b)). More importantly, the second statement provides an easy mechanism to *construct the true graph if the full inverse covariance matrix is accurately estimated*.

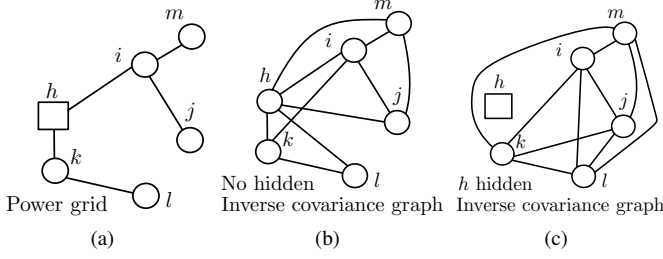


Fig. 2. (a) Grid graph $\mathcal{G}$ with hidden node h . (b) Sparsity of inverse voltage covariance $\mathcal{J}$ if all nodes are observed. (c) Sparsity of observed inverse voltage covariance $\mathcal{J}^{OO}$ if node h is hidden. Nodes l, m separated by 4 hops in $\mathcal{G}$ get an edge due to hidden node h .

Note: For DC-PF Eq. 7, it can be verified [15] under Assumptions 1, 2 that

$$J_{\theta DC}(i, j) := \begin{cases} < 0 & \text{if } (ij) \text{ is an edge} \\ > 0 & \text{if } i, j \text{ are two hop neighbors.} \\ = 0 & \text{otherwise} \end{cases} \quad (8)$$

Thus, edges can be identified by negative values in the DC-voltage inverse covariance matrix.

In the next section, we discuss the inverse voltage covariance statistics under partial observability, that form the basis for the algorithm development in this work.

III. VOLTAGE STATISTICS FOR PARTIAL OBSERVABILITY

We consider the practical setting where voltage measurements (both magnitude and phase) are limited to a set of observed nodes $\mathcal{O} \subset \mathcal{V}$, while nodes in set $\mathcal{H} = \mathcal{V} - \mathcal{O}$ are hidden. In the remaining article, we consider hidden nodes to satisfy the following mild assumption.

Assumption 3: Hidden nodes in the set $\mathcal{H}$ have a degree greater than 1 and any two elements of $\mathcal{H}$ are three or more hops away.

Assumption 3 implies that hidden nodes are not leaves. This is necessary for consistency/uniqueness [8] as one can always add the injection of a hidden leaf to the injection of its single neighbor and generate the same set of observed voltages (ignoring the unobserved leaf). Similarly, adjacent internal hidden nodes (see Fig. 3) can be merged while maintaining the same measurements at end-points thereby leading to non-unique estimation. Similarly, we discuss later (see Fig. 4) that separation of two hops between hidden nodes can lead to non-unique estimation for meshed grids (though not for radial grids).

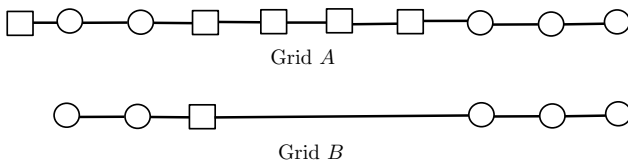


Fig. 3. Grid A has hidden nodes (squares) that are leaves or are adjacent. Case B with one hidden node is consistent with the same measurements as A, which can lead to non-unique estimation.

Consider the following permutation of the voltage vector

into observed and hidden parts: $\begin{bmatrix} v^{\mathcal{O}} \\ \theta^{\mathcal{O}} \\ v^{\mathcal{H}} \\ \theta^{\mathcal{H}} \end{bmatrix} = M \begin{bmatrix} v \\ \theta \end{bmatrix}$, where M is

a binary (0 – 1) permutation matrix [21]. Thus, the covariance of the permuted voltage vector is given by

$$\tilde{\Sigma}_{(v,\theta)} = M \Sigma_{(v,\theta)} M^T = \begin{bmatrix} \Sigma^{\mathcal{OO}} & \Sigma^{\mathcal{OH}} \\ \Sigma^{\mathcal{HO}} & \Sigma^{\mathcal{HH}} \end{bmatrix}, \quad (9)$$

where $\Sigma^{\mathcal{OO}} = \begin{bmatrix} \Sigma_{vv}^{\mathcal{OO}} & \Sigma_{v\theta}^{\mathcal{OO}} \\ \Sigma_{\theta v}^{\mathcal{OO}} & \Sigma_{\theta\theta}^{\mathcal{OO}} \end{bmatrix}$ is the covariance of the observed voltages. As permutation matrices satisfy $M^T = M^{-1}$ [21], using Eq. 6, we have

$$\tilde{\Sigma}_{(v,\theta)}^{-1} = M \Sigma_{(v,\theta)}^{-1} M^T =: \begin{bmatrix} J^{\mathcal{OO}} & J^{\mathcal{OH}} \\ J^{\mathcal{HO}} & J^{\mathcal{HH}} \end{bmatrix} \quad (10)$$

$$\text{where } \Sigma^{\mathcal{OO}-1} = J^{\mathcal{OO}} - J^{\mathcal{OH}} J^{\mathcal{HH}-1} J^{\mathcal{HO}}. \quad (11)$$

Eq. 11 follows from Eq. 9 through Schur's formula [21] for block inversion. For the inverse of the observed covariance $\Sigma^{\mathcal{OO}}$, we look at each term on the right side of Eq. 11. The following result provides the inverse of $J^{\mathcal{HH}}$.

Lemma 1. Consider $J^{\mathcal{HH}} = \begin{bmatrix} J_{vv}^{\mathcal{HH}} & J_{v\theta}^{\mathcal{HH}} \\ J_{\theta v}^{\mathcal{HH}} & J_{\theta\theta}^{\mathcal{HH}} \end{bmatrix}$, in Eqs. 10, 11. Under Assumption 3,

$$J^{\mathcal{HH}-1} = \begin{bmatrix} D^{\mathcal{H}} J_{\theta\theta}^{\mathcal{HH}} & -D^{\mathcal{H}} J_{\theta v}^{\mathcal{HH}} \\ -D^{\mathcal{H}} J_{v\theta}^{\mathcal{HH}} & D^{\mathcal{H}} J_{vv}^{\mathcal{HH}} \end{bmatrix}, \text{ where } D^{\mathcal{H}} \text{ is diagonal,}$$

$$D^{\mathcal{H}}(i, i) = (J_{vv}^{\mathcal{HH}}(i, i) J_{\theta\theta}^{\mathcal{HH}}(i, i) - J_{v\theta}^{\mathcal{HH}}(i, i) J_{\theta v}^{\mathcal{HH}}(i, i))^{-1}.$$

Proof: As hidden nodes in $\mathcal{H}$ are three or more hops away (Assumption 3), it follows from Theorem 1(a) that each block in $J^{\mathcal{HH}}$ is a diagonal matrix. Following inverse of diagonal block matrices, it can be verified that $J^{\mathcal{HH}-1}$ has a similar structure with diagonals on each block. ■

It is clear from Eq. 11 that the non-zero terms in observed $\Sigma^{\mathcal{OO}-1}$ may differ from the corresponding elements in $\Sigma_{(v,\theta)}^{-1}$. The following result formalizes this observation.

Theorem 2. In grid $\mathcal{G}$ with observed buses $\mathcal{O}$, let $\Sigma^{\mathcal{OO}-1}$ be the inverse voltage covariance of observed buses. For observed nodes i, j ,

- (a) If i, j are one or two hops away, $\Sigma^{\mathcal{OO}-1}(i, j) \neq 0$.
- (b) If i, j are separated by 5 or more hops, the terms in $\Sigma^{\mathcal{OO}-1}(i, j) = 0$.
- (c) If i, j are 3 or 4 hops away, $\Sigma^{\mathcal{OO}-1}(i, j)$ is non-zero, if and only if there is at least one common hidden node r within two hops of i and j .

Proof: Consider $\Sigma^{\mathcal{OO}-1}$ given by Eq. 11. As $J^{\mathcal{OO}}$, $J^{\mathcal{OH}}$ and $J^{\mathcal{HO}}$ are sub-matrices of $\Sigma_{(v,\theta)}^{-1}$, using Theorem 1(a), their non-zero terms are limited to nodes that are two or less hops away. This proves statement (a).

For i, j that are greater than two hops away, $J^{\mathcal{OO}}(i, j) = 0$. Using Lemma 1, non-zero terms in $J^{\mathcal{HH}-1}$ correspond to voltages at the same hidden node. For nodes i, j that are greater than 4 hops away, there exist no node r (hidden

or observed) that is within two hops of both i, j . Hence, $J^{\mathcal{O}\mathcal{H}} J^{\mathcal{H}\mathcal{H}^{-1}} J^{\mathcal{H}\mathcal{O}}(i, j) = 0$. This proves statement (b).

Using a similar argument, if i, j are 3 or 4 hops away but have no common hidden node r within two hops, $\Sigma^{\mathcal{O}\mathcal{O}^{-1}}(i, j) = 0$. If such hidden node r exists, then $J^{\mathcal{O}\mathcal{H}}(i, r), J^{\mathcal{H}\mathcal{H}^{-1}}(r, r), J^{\mathcal{H}\mathcal{O}}(r, j)$ are non-zero and $\Sigma^{\mathcal{O}\mathcal{O}^{-1}}(i, j)$ is non-zero. This proves statement (c). ■

Note: For DC-PF, we permute θ_{DC} into observed $\theta_{DC}^{\mathcal{O}}$ and hidden $\theta_{DC}^{\mathcal{H}}$, and divide voltage covariance $\Sigma_{\theta_{DC}}$ and its inverse $J_{\theta_{DC}}$ in Eq. 7 into observed and hidden blocks. Using Schur's complement, the inverse covariance of the observed voltage $\theta_{DC}^{\mathcal{O}}$ satisfies:

$$\Sigma_{\theta_{DC}}^{\mathcal{O}\mathcal{O}^{-1}} = J_{\theta_{DC}}^{\mathcal{O}\mathcal{O}} - J_{\theta_{DC}}^{\mathcal{O}\mathcal{H}} J_{\theta_{DC}}^{\mathcal{H}\mathcal{H}^{-1}} J_{\theta_{DC}}^{\mathcal{H}\mathcal{O}}, \quad (12)$$

As hidden nodes are separated by 3 or more hops, $J_{\theta_{DC}}^{\mathcal{H}\mathcal{H}}$ and its inverse are diagonal. Using a similar analysis as Theorem 2, it can be shown that $\Sigma_{\theta_{DC}}^{\mathcal{O}\mathcal{O}^{-1}}(i, j) = 0$ if observed nodes i, j are 5 or more hops away.

Using Theorem 2 for LC-PF and its extension to DC-PF, it is clear that the inverse covariance graph constructed using non-zero entries in the inverse covariance of observed voltages can have additional edges linking buses separated by 4 or less hops (see Fig. 2(c)). In the next section, we determine a learning algorithm that identifies the true edge in $\mathcal{G}$ using properties of $\Sigma^{\mathcal{O}\mathcal{O}^{-1}}$.

IV. LEARNING ALGORITHM USING PARTIALLY OBSERVED VOLTAGES

In observed inverse covariance $\Sigma^{\mathcal{O}\mathcal{O}^{-1}}$, given in Eq. 11, let

$$J^{\mathcal{O}\mathcal{O}} = \begin{bmatrix} J_{vv}^{\mathcal{O}\mathcal{O}} & J_{v\theta}^{\mathcal{O}\mathcal{O}} \\ J_{\theta v}^{\mathcal{O}\mathcal{O}} & J_{\theta\theta}^{\mathcal{O}\mathcal{O}} \end{bmatrix}. \quad (13)$$

Under Assumption 3, two hidden nodes are not adjacent. Thus, to identify all edges $\mathcal{E}$ in grid $\mathcal{G}$, we need to determine the edges between observed nodes in $\mathcal{O}$, and edges between observed and hidden nodes in $\mathcal{H}$. The following result allows exact recovery in this setting using properties of $J^{\mathcal{O}\mathcal{O}}$ in Eq. 13.

Theorem 3. In grid $\mathcal{G}$ with cycle length at least 5,

- (a) For observed $i, j \in \mathcal{O}$, $(ij) \in \mathcal{E}$ if and only if $J_{vv}^{\mathcal{O}\mathcal{O}}(i, j) + J_{\theta\theta}^{\mathcal{O}\mathcal{O}}(i, j) < 0$.
- (b) For $i, j \in \mathcal{O}$ with $(ij) \notin \mathcal{E}$, there exists one hidden h with edges $(ih), (jh) \in \mathcal{E}$ if and only if $J_{vv}^{\mathcal{O}\mathcal{O}}(i, j) \neq 0$, but $\nexists r \in \mathcal{O}$ with $(ir), (jr) \in \mathcal{E}$.

Proof: $J^{\mathcal{O}\mathcal{O}}$ (and consequently $J_{vv}^{\mathcal{O}\mathcal{O}}$) is a sub-matrix of the full inverse $\Sigma_{(v,\theta)}^{-1}$ in Eq. 10-11. Thus, statement (a) follows immediately from Theorem 1(b).

For statement (b), consider $i, j \in \mathcal{O}$ with $(ij) \notin \mathcal{E}$. Using Theorem 1(a), $J_{vv}^{\mathcal{O}\mathcal{O}}(i, j) \neq 0$ implies i, j are two-hops away. However as $\nexists r \in \mathcal{O}$ with $(ir), (jr) \in \mathcal{E}$, there should be a hidden node h that is neighbor of both i, j . The hidden neighbor is unique because otherwise the two hidden neighbors and i, j would form a cycle of size 4 that violates Assumption 1. For the converse, if a hidden node h exists with $(ih), (jh) \in \mathcal{E}$, there cannot exist another common neighbor r (hidden or

observed) of i, j as the minimum cycle length is 5. Hence the statement holds true. ■

Note: For DC-PF, the inverse covariance of observed voltages $\Sigma_{\theta_{DC}}^{\mathcal{O}\mathcal{O}^{-1}}$ is given in Eq. 12. As $J_{\theta_{DC}}^{\mathcal{O}\mathcal{O}}$ is a sub-matrix of the full inverse $J_{\theta_{DC}}$, it has the same sparsity properties mentioned in Eq. 8. Following the reasoning in Theorem 3, it can be shown that for observed buses $i, j \in \mathcal{O}$:

- (a) $J_{\theta_{DC}}^{\mathcal{O}\mathcal{O}}(i, j) < 0$ if and only if $(ij) \in \mathcal{E}$, and
- (b) There exists one hidden h with edges $(ih), (jh) \in \mathcal{E}$ if and only if $J_{\theta_{DC}}^{\mathcal{O}\mathcal{O}}(i, j) > 0$, but $\nexists r \in \mathcal{O}$ with $(ir), (jr) \in \mathcal{E}$.

If $J^{\mathcal{O}\mathcal{O}}$ (or its DC counterpart $J_{\theta_{DC}}^{\mathcal{O}\mathcal{O}}$) is known/estimated, Theorem 3(a) enables the identification of edges between all observed nodes. Next, Theorem 3(b) identifies observed nodes that have a hidden node as common neighbor. Consider a set of observed nodes $\mathcal{S} \subset \mathcal{O}$ such that $\forall i, j \in \mathcal{S}$, i, j satisfy Theorem 3(b). Hidden nodes can exist in multiple configurations to satisfy Theorem 3(b) as shown in Fig. 4. However, under Assumption 3, each observed node can have *at most* one hidden node as neighbor. Hence only the configuration (Fig. 4A) where *one* hidden node is the common neighbor of *all* nodes in $\mathcal{S}$ is consistent. To identify edges with hidden nodes, we iteratively find a disjoint observed set $\mathcal{S}$ of two hop neighbors, and connect one hidden node to all nodes in $\mathcal{S}$. A similar setup follows for the DC-PF case.

Note: For radial grids, configuration (Fig. 4B) with loops is not permissible even if hidden nodes are separated by two hops. Unique estimation in radial grids is thus possible if Assumption 3 is relaxed to *hidden nodes not being adjacent*.

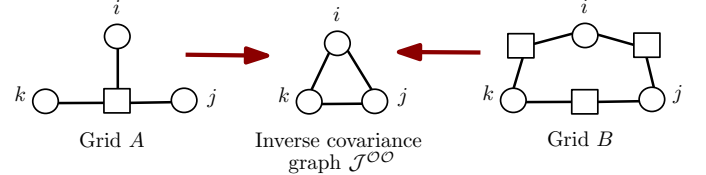


Fig. 4. Grid A and B with different hidden nodes configurations will lead to the same inverse covariance graph due to signed entries in $J^{\mathcal{O}\mathcal{O}}$. Only grid A satisfies Assumption 3.

Correct topology estimation is thus possible if $J_{vv}^{\mathcal{O}\mathcal{O}}$ (for LC-PF) or $J_{\theta_{DC}}^{\mathcal{O}\mathcal{O}}$ (for DC-PF) is estimated from the observed voltage measurements. Next we discuss its estimation based on its algebraic properties discussed previously.

A. Decomposition of partial inverse voltage covariance

Consider n samples of the observed LC-PF voltage vector $y_k = \begin{bmatrix} v_k^{\mathcal{O}} \\ \theta_k^{\mathcal{O}} \end{bmatrix}$ for $1 \leq k \leq n$ in the grid. The empirical observed covariance is given by

$$\Sigma^{\mathcal{O}\mathcal{O}} = \frac{1}{n-1} \sum_{k=1}^n (y_k - \sum_{k=1}^n \frac{y_k}{n}) (y_k - \sum_{k=1}^n \frac{y_k}{n})^T. \quad (14)$$

Note that $\Sigma^{\mathcal{O}\mathcal{O}^{-1}}$ can be estimated using direct inversion or Graphical Lasso [22], [14], which is the maximum likelihood estimator for multi-variate Gaussian random variables. However, edge estimation requires us to extract $J^{\mathcal{O}\mathcal{O}}$. From Eq. 11, $\Sigma^{\mathcal{O}\mathcal{O}^{-1}}$ is the sum of $J^{\mathcal{O}\mathcal{O}}$ and $J_2 = -J^{\mathcal{O}\mathcal{H}} J^{\mathcal{H}\mathcal{H}^{-1}} J^{\mathcal{H}\mathcal{O}}$.

Under Assumption 3, the number of observed nodes is greater (at least twice) the number of hidden nodes. Thus the rank of J_2 is at most $|\mathcal{H}| < |\mathcal{O}|$, and it is *rank deficient*. As non-zero terms in $J^{\mathcal{O}\mathcal{O}}$ are limited to observed nodes within two-hop distance, $J^{\mathcal{O}\mathcal{O}}$ is *sparse* for realistic distribution grids that are either radial or have large cycle width. We thus use a low-rank + sparse decomposition [23] for the estimated $\Sigma^{\mathcal{O}\mathcal{O}-1}$ to separate $J^{\mathcal{O}\mathcal{O}}$. This is done via the following semi-definite program (SDP):

$$\hat{J}^{\mathcal{O}\mathcal{O}} = \arg \min_{S \succ 0, S=S^T} \gamma \|S\|_1 + \|\Sigma^{\mathcal{O}\mathcal{O}-1} - S\|_* \quad (15)$$

Note that as matrix rank and sparsity are non-convex functions, we use their convex envelopes, nuclear norm ($\|\cdot\|_*$) and l_1 norm ($\|\cdot\|_1$) respectively, in Eq. 15. The hyperparameter γ is tuned using historical data. Once $J^{\mathcal{O}\mathcal{O}}$ is estimated, the determination of true edges, and identification of the location of hidden nodes follow from Theorem 3 and following discussion. The overall steps of the learning algorithm are listed in Algorithm IV-A.

Algorithm 1 Topology Learning with hidden nodes

Input: n samples of voltage vector $y_k = \begin{bmatrix} v_k^{\mathcal{O}} \\ \theta_k^{\mathcal{O}} \end{bmatrix}$ for $1 \leq k \leq n$
at observed nodes $\mathcal{O}$, threshold $\tau_1 > 0, \tau_2 > 0, \gamma > 0$
Output: Edge set $\mathcal{E}$ in grid $\mathcal{G}$.

- 1: Compute $\Sigma^{\mathcal{O}\mathcal{O}}$ and its inverse using Eq. 14.
 - 2: Estimate $J^{\mathcal{O}\mathcal{O}} = \begin{bmatrix} J_{vv}^{\mathcal{O}\mathcal{O}} & J_{v\theta}^{\mathcal{O}\mathcal{O}} \\ J_{\theta v}^{\mathcal{O}\mathcal{O}} & J_{\theta\theta}^{\mathcal{O}\mathcal{O}} \end{bmatrix}$ by solving Problem 15
 - 3: Construct $\mathcal{E} = \{\}, \mathcal{E}_2 = \{(ij) : |J_{vv}^{\mathcal{O}\mathcal{O}}(i, j)| > \tau_1\}$.
 - 4: **for all** $(ij) \in \mathcal{E}_2$ **do**
 - 5: **if** $J_{vv}^{\mathcal{O}\mathcal{O}}(i, j) + J_{\theta\theta}^{\mathcal{O}\mathcal{O}}(i, j) < -\tau_2$ **then**
 - 6: $\mathcal{E} \leftarrow \mathcal{E} \cup \{(ij)\}, \mathcal{E}_2 \leftarrow \mathcal{E}_2 - \{(ij)\}$
 - 7: **end if**
 - 8: **end for**
 - 9: **for all** $(ij) \in \mathcal{E}_2$ **do**
 - 10: **if** $\exists k.s.t.(ki) \text{ and } (kj) \in \mathcal{E}$ **then**
 - 11: $\mathcal{E}_2 \leftarrow \mathcal{E}_2 - \{(ij)\}$
 - 12: **end if**
 - 13: **end for**
 - 14: **for all edge** $(ij) \in \mathcal{E}_2$ **do**
 - 15: $S \leftarrow \{k : (ik), (jk) \in \mathcal{E}_2\}$. Insert hidden node h in $\mathcal{G}$
 - 16: $\mathcal{E} \leftarrow \mathcal{E} \cup \{(hk) : k \in S \cup \{i, j\}\}$
 - 17: $\mathcal{E}_2 \leftarrow \mathcal{E}_2 - \{(k_1 k_2) : k_1 \neq k_2 \in S \cup \{i, j\}\}$
 - 18: **end for**
-

Thresholds in Algorithm IV-A: Aside from γ , we include thresholds $\tau_1, \tau_2 > 0$ as empirical estimates of inverse voltage covariance may differ from the true value. τ_1 is used to identify observed nodes within two-hops through Theorem 3(a), while τ_2 is used to identify true edges through Theorem 3(b). These thresholds can be trained using historical data for optimal performance.

Computational Complexity: Matrix inversion and SDP-based decomposition for estimating $J^{\mathcal{O}\mathcal{O}}$ (Step 2) has a complexity of $O(N^3)$. Determining true edges between observed nodes and removing common two-hop neighbors (Step 4-13) has a complexity of $O(N^3)$. After Step 13, $\mathcal{E}_2$ includes only two-hop neighbors without common observed neighbor. Determining groups S of two-hop neighbors connected to a

common hidden node and adding edges (Step 14-18) has a complexity of $O(N^3)$. Thus the overall complexity of the learning algorithm is $O(N^3)$.

V. NUMERICAL SIMULATIONS

We consider two distribution grids with 4 hidden nodes, constructed from the radial IEEE 33 bus system [24]: radial (Fig. 5(a)) and loopy (Fig. 6(a)). The fluctuations of nodal injections around their mean loads are sampled using zero-mean Gaussian random variables of standard deviation 10^{-1} . Then voltage samples are constructed for each grid using (a) linear LC-PF Eq. 3, and (b) non-linear PF equations using Matpower [25]. We solve Problem (15) in Algorithm IV-A using CVX [26]. We calibrate algorithm's accuracy by relative estimation errors, computed as the sum of false edges and missed edges (false positives and true negatives) relative to the number of true edges in $\mathcal{E}$. The thresholds τ_1 and τ_2 in Algorithm IV-A are tuned to give minimal estimation error for a large sample size of 10^5 . These thresholds are fixed, and the estimation errors at different sample sizes are computed by averaging over 10 runs with independent samples. Fig. 5(b) shows the performance for the radial grid in Fig. 5(a) where we consider 3 and 4 hidden nodes separated by 2 hops. Note that the errors reduce to around 0 for 10^4 samples for both cases. Further the performance with linear and non-linear voltage samples are almost identical. This is true as the LC-PF model approximates the non-linear voltages correctly. Fig. 6(b) shows the performance for the loopy grid in Fig. 6(a) where hidden nodes are separated by 3 hops, following Assumption 3. The performance improves as the number of samples is increased and is comparable for linear and non-linear samples. As expected, the performance in both radial and loopy grids worsens with higher number of hidden nodes, although exact recovery is demonstrated in the high sample regime of 10^5 .

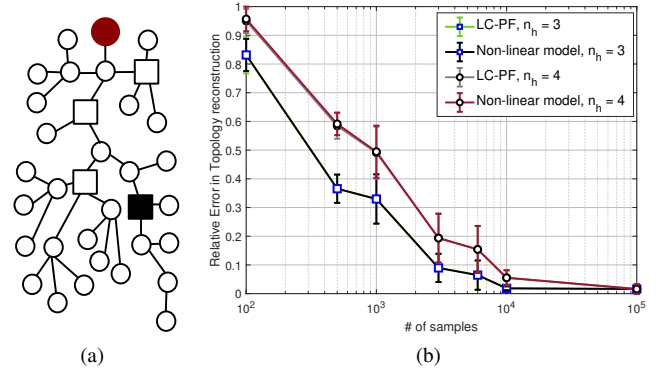


Fig. 5. Learning for radial network with 33 nodes, and hidden nodes separated by at least 2 hops, over 10 iterations. Number of hidden nodes n_h is taken as 3 and 4 (additional hidden node colored black).

VI. CONCLUSION

In this paper, we discuss topology learning in partially observed meshed distribution grids using only nodal voltage samples. Using algebraic properties combined with a low-rank + sparse decomposition of the inverse covariance matrix of observed voltages, we design a novel learning algorithm that is able to recover the topology of meshed grids when

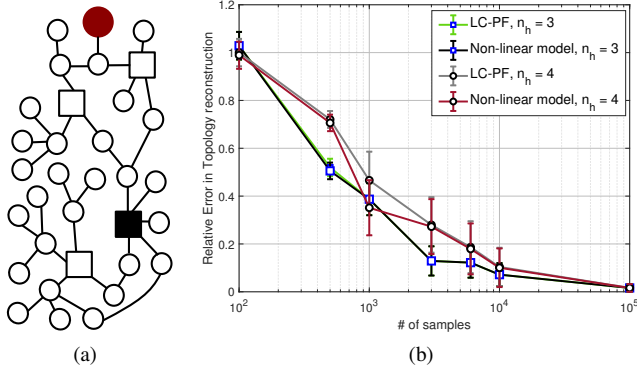


Fig. 6. Learning for loopy network with 33 nodes, and hidden nodes separated by atleast 3 hops, over 10 iterations. Number of hidden nodes n_h is taken as 3 and 4 (additional hidden node colored black).

hidden nodes are separated by three or more hops. For radial grids, our algorithm is able to recover the true topology for non-adjacent hidden nodes. Unlike prior work, our learning algorithm has low information requirement, as it does not require any injection samples or statistics, or knowledge of line impedances. In future, we plan to extend this work to three phase power flow models [4], as well as to learning under power flow dynamics [20]. Furthermore, we propose to study learning under partial observability in practical systems with bus-splitting and side-information regarding nodal injections.

REFERENCES

- [1] R. Hoffman, "Practical state estimation for electric distribution networks," in *IEEE PES Power Systems Conference and Exposition*. IEEE, 2006, pp. 510–517.
- [2] A. von Meier and et al., "Micro-synchrophasors for distribution systems," in *IEEE Innovative Smart Grid Technologies Conference*, 2014.
- [3] US-DoE, "Advanced metering infrastructure & customer system," 2018.
- [4] D. Deka, M. Chertkov, and S. Backhaus, "Topology estimation using graphical models in multi-phase power distribution grids," *IEEE Transactions on Power Systems*, 2019.
- [5] S. Bolognani and et al., "Identification of power distribution network topology via voltage correlation analysis," in *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*. IEEE, 2013.
- [6] Y. Liao, Y. Weng, G. Liu, and R. Rajagopal, "Urban mv and lv distribution grid topology estimation via group lasso," *IEEE Transactions on Power Systems*, 2018.
- [7] D. Deka, S. Backhaus, and M. Chertkov, "Structure learning in power distribution networks," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 3, pp. 1061–1074, 2018.
- [8] S. Park, D. Deka, and M. Chertkov, "Exact topology and parameter estimation in distribution grids with minimal observability," in *Power Systems Computation Conference (PSCC), 2018*. IEEE, 2018.
- [9] D. Deka, S. Backhaus, and M. Chertkov, "Learning topology of distribution grids using only terminal node measurements," in *IEEE Smartgridcomm*, 2016.
- [10] G. Cavraro and V. Kekatos, "Inverter probing for power distribution network topology processing," *IEEE Transactions on Control of Network Systems*, 2019.
- [11] G. Cavraro and V. Kekatos, "Graph algorithms for topology identification using power grid probing," *IEEE control systems letters*, vol. 2, no. 4, pp. 689–694, 2018.
- [12] C. Rudin and et al., "Machine learning for the new york city power grid," *IEEE Trans. on pattern analysis and machine intelligence*, 2011.
- [13] D. Wolter and et al., "Impact of meshed grid topologies on distribution grid planning and operation," *CIGRE-Open Access Proceedings Journal*, 2017.
- [14] D. Deka, S. Talukdar, M. Chertkov, and M. Salapaka, "Graphical models in meshed distribution grids: Topology estimation, change detection limitations," *IEEE Transactions on Smart Grid*, 2020.
- [15] D. Deka, M. Chertkov, S. Talukdar, and M. V. Salapaka, "Topology estimation in bulk power grids: Theoretical guarantees and limits," in *Bulk Power Systems Dynamics and Control Symposium-IREP*, 2017.
- [16] M. Baran and F. Wu, "Network reconfiguration in distribution systems for loss reduction and load balancing," *Power Delivery, IEEE Transactions on*, vol. 4, no. 2, pp. 1401–1407, Apr 1989.
- [17] S. Bolognani and S. Zampieri, "On the existence and linear approximation of the power flow solution in power distribution networks," *Power Systems, IEEE Transactions on*, vol. 31, no. 1, pp. 163–172, 2016.
- [18] A. Abur and A. G. Exposito, *Power system state estimation: theory and implementation*. CRC Press, 2004.
- [19] Y. Dvorkin, M. Lubin, S. Backhaus, and M. Chertkov, "Uncertainty sets for wind power generation," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 3326–3327, 2016.
- [20] S. Talukdar and et al., "Learning exact topology of a loopy power grid from ambient dynamics," in *Proceedings of the Eighth International Conference on Future Energy Systems*, 2017.
- [21] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 2012.
- [22] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, no. 3, pp. 432–441, 2008.
- [23] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky, "Sparse and low-rank matrix decompositions," *IFAC Proceedings Volumes*, vol. 42, no. 10, pp. 1493–1498, 2009.
- [24] "IEEE 1547 Standard for Interconnecting Distributed Resources with Electric Power Systems." [Online]. Available: http://grouper.ieee.org/groups/scc21/1547/1547_index.html
- [25] R. D. Zimmerman and C. E. Murillo-Sánchez, "Matpower 4.1 user's manual."
- [26] M. Grant and S. Boyd, "Cvx: Matlab software for disciplined convex programming, version 2.1," 2014.