Grid Topology Identification With Hidden Nodes via Structured Norm Minimization

Rajasekhar Anguluri, *Member, IEEE*, Gautam Dasarathy, *Member, IEEE*, Oliver Kosut, *Member, IEEE*, and Lalitha Sankar, *Senior Member, IEEE*

***Abstract*—This letter studies a topology identification problem for an electric distribution grid using sign patterns of the inverse covariance matrix of bus voltage magnitudes and angles, while accounting for hidden buses. Assuming the grid topology is sparse and the number of hidden buses are fewer than those of the observed buses, we express the observed voltages inverse covariance matrix as the sum of three structured matrices: *sparse matrix, low-rank matrix with sparse factors*, and *low-rank matrix*. Using the sign patterns of the first two of these matrices, we develop an algorithm to identify the topology of a distribution grid with a minimum cycle length greater than three. To estimate the structured matrices from the empirical inverse covariance matrix, we formulate a novel convex optimization problem with appropriate sparsity and structured norm constraints and solve it using an alternating minimization method. We validate the proposed algorithm's performance on a modified IEEE 33 bus system.**

***Index Terms*—Power systems, smart grid, estimation, atomic norm, and alternating minimization.**

I. Introduction

THE KNOWLEDGE of electric distribution grid topology[1] is crucial to many power system applications, including state estimation, control of energy resources, and cybersecurity [1], [2]. However, operators have limited or no access to the grid's topology in real-time, and they need to identify it from measurements [1]. Identifying topology from measurements is a challenging problem because of the nonlinear relationship between the measured quantities and the topology, noise in the measurements, and the compromised or missing data.

Several data-driven methods have appeared in the literature for the topology identification problem. In [3], a linear regression framework, with unknown predictors encoding the topology, is used to identify the topology. Along these lines, [1], [4] reconstructed topology by invoking the sparse nature of the distribution grid via group and adaptive LASSO formulations. In [5], a decision theoretic framework is used to reconstruct tree structured grids; instead, [6] uses graphical models to estimate tree and meshed grid topologies. In contrast to the preceding offline methods, online methods for joint estimation of topology and line parameters are considered in [7], [8]. Finally, topology identification in the presence of hidden buses (unmeasured buses) is considered in [9].

Lately, [10] showed that the topology could be identified using sparsity pattern of the inverse covariance matrix of voltage magnitude. Assuming the grid's minimum cycle length (MCL) is greater than three, [11] provided a simple topology identification algorithm using sign patterns of the inverse covariance matrix of voltage magnitudes and angles measurements recorded from all the buses. By definition, radial grids satisfy the MCL constraint; however, for meshed grids, the preceding size constraint on MCL is necessary to uniquely identify the topology [11]. For a recent summary on topology identification from voltage correlations, see [12].

In this letter, for the grid topology identification problem using voltage measurements, we relax (i) the full observability assumption of [11] to the case where only a subset of buses are measured (called the *observed buses*); and (ii) the MCL assumption of greater than four in [13] to that of the theoretically possible limit of greater than three (see Theorem 2). Our MCL assumption is the theoretical limit below which topology identification is not possible (see Section IV). Further, via numerical simulations, we show that the sparse plus low rank decomposition method, used in [13, Algorithm 1], fails for specific grid topologies (see Remark 2). The main contributions of our paper are as follows.

1) Assuming the underlying grid topology is sparse, and no two hidden buses are adjacent, we decompose the inverse covariance matrix of observed voltage magnitudes and angles into structured matrices: sparse, low-rank with sparse factors, and low-rank matrices.
2) We present a recursive method (Algorithm 1) to estimate the grid's topology with MCL greater than three using sign patterns of the sparse and low-rank with sparse factors matrices (see Theorem 2 for details).
3) We formulate a convex optimization problem with a structured norm-based regularizer to estimate the structured matrices from the empirical covariance matrix.

Manuscript received March 4, 2021; revised May 17, 2021; accepted June 9, 2021. Date of publication June 17, 2021; date of current version July 1, 2021. This work was supported in part by the National Science Foundation (NSF) under Award OAC-1934766 and Award CCF-2029044, and in part by the National Institutes of Health (NIH) under Award 1R01GM140468-01. Recommended by Senior Editor R. S. Smith. *(Corresponding author: Rajasekhar Anguluri.)*

The authors are with the School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85281 USA (e-mail: rangulur@asu.edu; gautamd@asu.edu; okosut@asu.edu; lalithasankar@asu.edu).

Digital Object Identifier 10.1109/LCSYS.2021.3089993

[1]By topology, we mean physical connectivity (active lines or transformers) among the buses in a distribution grid.

Finally, to solve the optimization problem, we propose an alternating minimization algorithm that combines ADMM and Frank-Wolfe methods.

II. Problem Setup and Preliminary Notions

We represent a distribution grid of $N+1$ buses with the graph $\mathcal{G} := (\mathcal{V}, \mathcal{E})$, where the nodes $\mathcal{V} = \{1, \ldots, N+1\}$ and the edges $(ij) \in \mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ denote the buses and the lines, respectively. A set of distinct undirected edges $\mathcal{P}_i^k = \{(il_1), (l_1 l_2), \ldots, (l_{t-1}, k)\}$ that connect node i and j is referred to as *path* of length of t. A *cycle* is a path $\mathcal{P}_i^k$ with $i = k$ and length greater than two, i.e., $t > 2$. The *minimum cycle length* is the length of the smallest cycle. The neighbors of a node i are the set of nodes j such that $(ij) \in \mathcal{E}$. Nodes i and j are referred to as *k-hop* neighbors if the shortest path between them equals k.

A. Power Distribution Grid Model

Let $p_n + jq_n$ and $v_n + j\theta_n$ (where $j = \sqrt{-1}$) denote the complex valued power injection and voltage at the bus $n \in \mathcal{V}$. We refer the quantities v_n, θ_n, p_n, and q_n as *voltage magnitude, voltage phase, active power*, and *reactive power*, respectively. Let $y_{nm} = g_{nm} + j\beta_{nm}$ be the admittance of the line (nm), where g_{nm} and β_{nm} are the conductance and susceptance, resp. Consider the nonlinear power flow model:

$$p_n + jq_n = \sum_{m:(nm)\in\mathcal{E}} y_{nm}[v_n^2 - v_n v_m \exp(j(\theta_m - \theta_n))],$$

where $n \in \mathcal{V}$. Following [11], [13], we assume that $|\theta_m - \theta_n| << 1$ for $(mn) \in \mathcal{E}$ and $|v_n - 1| << 1$. Thus, we have the following linear coupled power flow (LC-PF) model:[2]

$$p_n + jq_n = \sum_{m:(nm)\in\mathcal{E}} y_{nm}[(v_n - v_m) - j(\theta_n - \theta_m)]. \tag{1}$$

Without loss of generality, we ignore the reference bus, $r \in \mathcal{V}$, and consider the power flow model (1) for the remaining N non-reference buses.

Let $\mathcal{O} \subseteq \mathcal{V}$ be the set of observed (measured) buses of size o, and $\mathcal{Z} := \mathcal{V} \setminus \mathcal{O}$ be the set of hidden (unmeasured) buses of size z. Let $\mathbf{v}_\mathcal{O} = [v_{n_1}, \ldots, v_{n_o}]^\mathsf{T}$, $\boldsymbol{\theta}_\mathcal{O} = [\theta_{n_1}, \ldots, \theta_{n_o}]^\mathsf{T}$, $\mathbf{p}_\mathcal{O} = [p_{n_1}, \ldots, p_{n_o}]^\mathsf{T}$, and $\mathbf{q}_\mathcal{O} = [q_{n_1}, \ldots, q_{n_o}]^\mathsf{T}$, where $n_j \in \mathcal{O}$. Let $\mathbf{y}_\mathcal{O}^\mathsf{T} = [\mathbf{v}_\mathcal{O}^\mathsf{T} \;\; \boldsymbol{\theta}_\mathcal{O}^\mathsf{T}]$ and $\mathbf{x}_\mathcal{O}^\mathsf{T} = [\mathbf{p}_\mathcal{O}^\mathsf{T} \;\; \boldsymbol{q}_\mathcal{O}^\mathsf{T}]$, respectively, be the vector of voltage magnitudes and angles and the vector of real and reactive power injections of buses in $\mathcal{O}$. Analogously, define the quantities $\mathbf{v}_\mathcal{Z}$, $\boldsymbol{\theta}_\mathcal{Z}$, $\mathbf{p}_\mathcal{Z}$, $\mathbf{q}_\mathcal{Z}$, $\mathbf{y}_\mathcal{Z}$ and $\mathbf{x}_\mathcal{Z}$ for the buses in $\mathcal{Z}$.

For all $n \in \mathcal{V}$, by doubling the complex valued equation (1) into two real valued equations and rearranging terms, we have the following equivalent representation for (1)

$$\begin{bmatrix} \mathbf{y}_\mathcal{O} \\ \mathbf{y}_\mathcal{Z} \end{bmatrix} = \underbrace{\begin{bmatrix} H_{\mathcal{O}\mathcal{O}} & H_{\mathcal{O}\mathcal{Z}} \\ H_{\mathcal{O}\mathcal{Z}}^\mathsf{T} & H_{\mathcal{Z}\mathcal{Z}} \end{bmatrix}^{-1}}_{\triangleq H^{-1}} \begin{bmatrix} \mathbf{x}_\mathcal{O} \\ \mathbf{x}_\mathcal{Z} \end{bmatrix}, \tag{2}$$

where $H \in \mathbb{R}^{2N\times 2N}$. Define the edge sets $\mathcal{E}_{\mathcal{O}\mathcal{O}} \subset \mathcal{O} \times \mathcal{O}$, $\mathcal{E}_{\mathcal{O}\mathcal{Z}} \subset \mathcal{O} \times \mathcal{Z}$, and $\mathcal{E}_{\mathcal{Z}\mathcal{Z}} \subset \mathcal{Z} \times \mathcal{Z}$. Then, for any index

[2]In this letter, we use the linearized power flow model; however, due to mathematical equivalence, our identification algorithm (see Section III-A) can be used on the traditional linear circuit model as well.

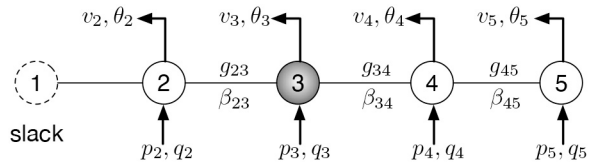

Fig. 1. Line grid with 1 as the reference bus. The hidden bus is 3, and the rest are the observed buses.

$\zeta \in \{\mathcal{O}\mathcal{O}, \mathcal{O}\mathcal{Z}, \mathcal{Z}\mathcal{Z}\}$, it follows that

$$H_\zeta = \begin{bmatrix} H_g^{(\zeta)} & H_\beta^{(\zeta)} \\ H_\beta^{(\zeta)} & -H_g^{(\zeta)} \end{bmatrix} \tag{3}$$

with the (i,j)-th entry of $H_g^{(\zeta)}$ (similarly for $H_\beta^{(\zeta)}$) given by

$$H_g^{(\zeta)}(i,j) = \begin{cases} \sum_{l:(il)\in\mathcal{E}_\zeta} g_{il}, & \text{for } i = j \\ -g_{ij}, & \text{for } (ij) \in \mathcal{E}_\zeta \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

From (4), notice that $H_g^{(\mathcal{O}\mathcal{Z})}(i,j) = 0$ when there is no edge between buses $i \in \mathcal{O}$ and $j \in \mathcal{Z}$. Similar conclusion holds for $H_g^{(\mathcal{O}\mathcal{O})}$ and $H_g^{(\mathcal{Z}\mathcal{Z})}$. Thus, blocks $H_{\mathcal{O}\mathcal{O}}$, $H_{\mathcal{O}\mathcal{Z}}$, and $H_{\mathcal{Z}\mathcal{Z}}$ encode the connectivity among the observed buses, between an observed and a hidden bus, and among the hidden buses, respectively; see Example 1. With a slight abuse of notation we refer H in (2) to as the Laplacian matrix of $\mathcal{G}$.

Example 1: Consider the line grid shown in Fig. 1.

For the grid in Fig. 1, $\mathbf{y}_\mathcal{O}^\mathsf{T} = [v_2 \;\; v_4 \;\; v_5 \;\; \theta_2 \;\; \theta_4 \;\; \theta_5]$ and $\mathbf{y}_\mathcal{Z}^\mathsf{T} = [v_3 \;\; \theta_3]$, and from (3) and (4), it follows that

$$H_{\mathcal{O}\mathcal{O}} = \left[\begin{array}{ccc|ccc} g_{23} & 0 & 0 & \beta_{23} & 0 & 0 \\ 0 & (g_{34}+g_{45}) & -g_{45} & 0 & (\beta_{34}+\beta_{45}) & -\beta_{45} \\ 0 & -g_{45} & g_{45} & 0 & -\beta_{45} & \beta_{45} \\ \hline \beta_{23} & 0 & 0 & -g_{23} & 0 & 0 \\ 0 & (\beta_{34}+\beta_{45}) & -\beta_{45} & 0 & -(g_{34}+g_{45}) & g_{45} \\ 0 & -\beta_{45} & \beta_{45} & 0 & g_{45} & -g_{45} \end{array}\right];$$

$$H_{\mathcal{O}\mathcal{Z}}^\mathsf{T} = \left[\begin{array}{ccc|ccc} -g_{23} & -g_{34} & 0 & -\beta_{23} & -\beta_{34} & 0 \\ \hline -\beta_{23} & -\beta_{34} & 0 & g_{23} & g_{34} & 0 \end{array}\right]; \; H_\mathcal{Z}^\mathsf{T} = \begin{bmatrix} \widetilde{g}_{24} & \widetilde{\beta}_{24} \\ \widetilde{\beta}_{24} & -\widetilde{g}_{24} \end{bmatrix};$$

where $\widetilde{g}_{24} = g_{23} + g_{34}$ and $\widetilde{\beta}_{24} = \beta_{23} + \beta_{34}$.

B. Stochastic Model for Power Injections and Voltages

We establish a relation between the Laplacian H (defined in (2)) and the covariance matrix of $(\mathbf{y}_\mathcal{O}, \mathbf{y}_\mathcal{Z})$. This relation plays a key role in our topology identification algorithm, which we discuss later. We begin with the following assumption.

Assumption 1 (Power injections): For $n \in \mathcal{G}$, the power injection $\mathbf{x}_n = [p_n, \; q_n]^\mathsf{T}$ is a zero-mean Gaussian random vector with non-degenerate covariance matrix. Also, $\mathbf{x}_m$ and $\mathbf{x}_n$ (for $m \neq n$) are uncorrelated; that is, $\mathbb{E}[\mathbf{x}_m \mathbf{x}_n^\mathsf{T}] = \mathbf{0}$.

The validity of Gaussian distribution for power injections is discussed in [11], [12]. Under Assumption 1, $\mathbb{E}[\mathbf{x}_n \mathbf{x}_n^\mathsf{T}]$ need not be diagonal; that is, we allow for correlations among the real and reactive power injections at any bus. Since $(\mathbf{x}_\mathcal{O}, \mathbf{x}_\mathcal{Z})$ is obtained by stacking $\mathbf{x}_n$, it follows that $\mathbb{E}[\mathbf{x}_\mathcal{O} \mathbf{x}_\mathcal{Z}^\mathsf{T}] = \mathbf{0}$, and the joint covariance matrix of $(\mathbf{x}_\mathcal{O}, \mathbf{x}_\mathcal{Z})$ is given by

$$\Sigma_{(\mathbf{x}_\mathcal{O}, \mathbf{x}_\mathcal{Z})} = \mathbb{E}\begin{bmatrix} \mathbf{x}_\mathcal{O} \\ \mathbf{x}_\mathcal{Z} \end{bmatrix} [\mathbf{x}_\mathcal{O}^\mathsf{T} \quad \mathbf{x}_\mathcal{Z}^\mathsf{T}] = \begin{bmatrix} \Sigma_{\mathbf{x}_\mathcal{O}, \mathbf{x}_\mathcal{O}} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & \Sigma_{\mathbf{x}_\mathcal{Z}, \mathbf{x}_\mathcal{Z}} \end{bmatrix}, \tag{5}$$

where $\Sigma_{\mathbf{x}_\mathcal{O},\mathbf{x}_\mathcal{O}} = \mathbb{E}[\mathbf{x}_\mathcal{O}\mathbf{x}_\mathcal{O}^\mathsf{T}]$ and $\Sigma_{\mathbf{x}_\mathcal{Z},\mathbf{x}_\mathcal{Z}} = \mathbb{E}[\mathbf{x}_\mathcal{Z}\mathbf{x}_\mathcal{Z}^\mathsf{T}]$ are the cross covariance matrices. Note the difference in the subscript notations of the joint- and cross-covariance matrices.

Let $\Sigma_{(\mathbf{y}_\mathcal{O},\mathbf{y}_\mathcal{Z})}$ be the joint covariance matrix of the full voltage vector $(\mathbf{y}_\mathcal{O}, \mathbf{y}_\mathcal{Z})$, and consider the parmeterization

$$\Sigma^{-1}_{(\mathbf{y}_\mathcal{O},\mathbf{y}_\mathcal{Z})} = \begin{bmatrix} \Sigma_{\mathbf{y}_\mathcal{O},\mathbf{y}_\mathcal{O}} & \Sigma_{\mathbf{y}_\mathcal{O},\mathbf{y}_\mathcal{Z}} \\ \Sigma^\mathsf{T}_{\mathbf{y}_\mathcal{O},\mathbf{y}_\mathcal{Z}} & \Sigma_{\mathbf{y}_\mathcal{Z},\mathbf{y}_\mathcal{Z}} \end{bmatrix}^{-1} = \begin{bmatrix} K_{\mathcal{O}\mathcal{O}} & K_{\mathcal{O}\mathcal{Z}} \\ K^\mathsf{T}_{\mathcal{O}\mathcal{Z}} & K_{\mathcal{Z}\mathcal{Z}} \end{bmatrix}. \quad (6)$$

To obtain expressions for $K_{\mathcal{O}\mathcal{O}}$, $K_{\mathcal{O}\mathcal{Z}}$, and $K_{\mathcal{Z}\mathcal{Z}}$ in terms of the block matrices of H (2) proceed as follows. From (2) and (5), first express $\Sigma_{(\mathbf{y}_\mathcal{O},\mathbf{y}_\mathcal{Z})}$ in terms of $\Sigma_{\mathbf{x}_\mathcal{O},\mathbf{x}_\mathcal{O}}$ and $\Sigma_{\mathbf{x}_\mathcal{Z},\mathbf{x}_\mathcal{Z}}$. Then, take the inverse of $\Sigma^{-1}_{(\mathbf{y}_\mathcal{O},\mathbf{y}_\mathcal{Z})}$ to note the following:

$$\begin{aligned} K_{\mathcal{O}\mathcal{O}} &= \underbrace{H_{\mathcal{O}\mathcal{O}}\Sigma^{-1}_{\mathbf{x}_\mathcal{O},\mathbf{x}_\mathcal{O}}H_{\mathcal{O}\mathcal{O}}}_{\triangleq S} + \underbrace{H_{\mathcal{O}\mathcal{Z}}\Sigma^{-1}_{\mathbf{x}_\mathcal{Z},\mathbf{x}_\mathcal{Z}}H^\mathsf{T}_{\mathcal{O}\mathcal{Z}}}_{\triangleq L} \\ K_{\mathcal{O}\mathcal{Z}} &= H_{\mathcal{O}\mathcal{O}}\Sigma^{-1}_{\mathbf{x}_\mathcal{O},\mathbf{x}_\mathcal{O}}H_{\mathcal{O}\mathcal{Z}} + H_{\mathcal{O}\mathcal{Z}}\Sigma^{-1}_{\mathbf{x}_\mathcal{Z},\mathbf{x}_\mathcal{Z}}H_{\mathcal{Z}\mathcal{Z}} \\ K_{\mathcal{Z}\mathcal{Z}} &= H_{\mathcal{Z}\mathcal{Z}}\Sigma^{-1}_{\mathbf{x}_\mathcal{Z},\mathbf{x}_\mathcal{Z}}H_{\mathcal{Z}\mathcal{Z}} + H^\mathsf{T}_{\mathcal{O}\mathcal{Z}}\Sigma^{-1}_{\mathbf{x}_\mathcal{O},\mathbf{x}_\mathcal{O}}H_{\mathcal{O}\mathcal{Z}}. \end{aligned} \quad (7)$$

Using the Schur's formula, we have the following formula to evaluate the observed inverse covariance matrix:

$$\Sigma^{-1}_{\mathbf{y}_\mathcal{O},\mathbf{y}_\mathcal{O}} = (\mathbb{E}[\mathbf{y}_\mathcal{O}\mathbf{y}_\mathcal{O}^\mathsf{T}])^{-1} = K_{\mathcal{O}\mathcal{O}} - \underbrace{K_{\mathcal{O}\mathcal{Z}}K^{-1}_{\mathcal{Z}\mathcal{Z}}K^\mathsf{T}_{\mathcal{O}\mathcal{Z}}}_{\triangleq M}. \quad (8)$$

Thus, the inverse covariance matrix $\Sigma^{-1}_{\mathbf{y}_\mathcal{O},\mathbf{y}_\mathcal{O}}$, if it exists, can always be decomposed as $S+L-M$. The conditions that ensure the sparse nature of S and low-rank factorization of L are discussed in Section IV. Note that the triple $(K_{\mathcal{O}\mathcal{O}}, K_{\mathcal{O}\mathcal{Z}}, K_{\mathcal{Z}\mathcal{Z}})$ encodes the complete grid topology via $(H_{\mathcal{O}\mathcal{O}}, H_{\mathcal{O}\mathcal{Z}}, H_{\mathcal{Z}\mathcal{Z}})$. Thus, the complete grid topology is also encoded in both the full and the observed inverse covariance matrices, $\Sigma^{-1}_{(\mathbf{y}_\mathcal{O},\mathbf{y}_\mathcal{Z})}$ and $\Sigma^{-1}_{\mathbf{y}_\mathcal{O},\mathbf{y}_\mathcal{O}}$, respectively.

Topology Identification Problem: Assuming the knowledge of $\Sigma^{-1}_{\mathbf{y}_\mathcal{O},\mathbf{y}_\mathcal{O}}$ (or an estimate of it) and the decomposition (8), infer the non-zero entries (topology) of H in (2).

To address the above problem, we define a graph whose edges are given by the non-zero entries of the full inverse covariance $\Sigma^{-1}_{(\mathbf{y}_\mathcal{O},\mathbf{y}_\mathcal{Z})}$. Formally, for the graph $(\mathcal{V}_{GM}, \mathcal{E}_{GM})$, associate a random variable $z_n \in (\mathbf{y}_\mathcal{O}, \mathbf{y}_\mathcal{Z})$ to n-th node in $\mathcal{V}_{GM} = \{1, \ldots, 2N\}$. The size of $\mathcal{V}_{GM}$ is $2N$ because for every n-th bus, we have two scalars $(v_n, \theta_n) \in (\mathbf{y}_\mathcal{O}, \mathbf{y}_\mathcal{Z})$. The edge $(n, m) \notin \mathcal{E}_{GM}$ iff $\Sigma^{-1}_{m,n} = 0$. In what follows, we characterize the induced graph structures of $\Sigma^{-1}_{\mathbf{y}_\mathcal{O},\mathbf{y}_\mathcal{O}}$ and $K_{\mathcal{O}\mathcal{O}}$ based on $(\mathcal{V}_{GM}, \mathcal{E}_{GM})$ and the grid graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$.

If $(\mathbf{y}_\mathcal{O}, \mathbf{y}_\mathcal{Z})$ follows a zero-mean Gaussian distribution, the graph $(\mathcal{V}_{GM}, \mathcal{E}_{GM})$ is equivalent to the underlying un-directed graphical model. Recall that a graphical model encodes conditional dependencies between pairs of random variables [14].

III. Grid Topology Identification

This section provides an algorithm to identify the complete grid topology of $\mathcal{G}$ using the sign pattern of $K_{\mathcal{O}\mathcal{O}}$.

Let $\Sigma^{(\mathcal{O}\mathcal{O})}_{pp} = \mathbb{E}[\mathbf{p}_\mathcal{O}\mathbf{p}_\mathcal{O}^\mathsf{T}]$; $\Sigma^{(\mathcal{O}\mathcal{O})}_{qq} = \mathbb{E}[\mathbf{q}_\mathcal{O}\mathbf{q}_\mathcal{O}^\mathsf{T}]$; and $\Sigma^{(\mathcal{O}\mathcal{O})}_{pq} = \mathbb{E}[\mathbf{p}_\mathcal{O}\mathbf{q}_\mathcal{O}^\mathsf{T}]$. Similarly, let $\Sigma^{(\mathcal{O}\mathcal{Z})}_{pp} = \mathbb{E}[\mathbf{p}_\mathcal{O}\mathbf{p}_\mathcal{Z}^\mathsf{T}]$, $\Sigma^{(\mathcal{O}\mathcal{Z})}_{qq} = \mathbb{E}[\mathbf{q}_\mathcal{O}\mathbf{q}_\mathcal{Z}^\mathsf{T}]$, and $\Sigma^{(\mathcal{O}\mathcal{Z})}_{pq} = \mathbb{E}[\mathbf{p}_\mathcal{O}\mathbf{q}_\mathcal{Z}^\mathsf{T}]$.

Lemma 1: Let $K_{\mathcal{O}\mathcal{O}} = S + L$ be as in (7). Then,

$$S = \begin{bmatrix} J^{(\mathcal{O}\mathcal{O})}_{vv} & J^{(\mathcal{O}\mathcal{O})}_{v\theta} \\ J^{(\mathcal{O}\mathcal{O})}_{\theta v} & J^{(\mathcal{O}\mathcal{O})}_{\theta\theta} \end{bmatrix} \text{ and } L = \begin{bmatrix} J^{(\mathcal{O}\mathcal{Z})}_{vv} & J^{(\mathcal{O}\mathcal{Z})}_{v\theta} \\ J^{(\mathcal{O}\mathcal{Z})}_{\theta v} & J^{(\mathcal{O}\mathcal{Z})}_{\theta\theta} \end{bmatrix}, \quad (9)$$

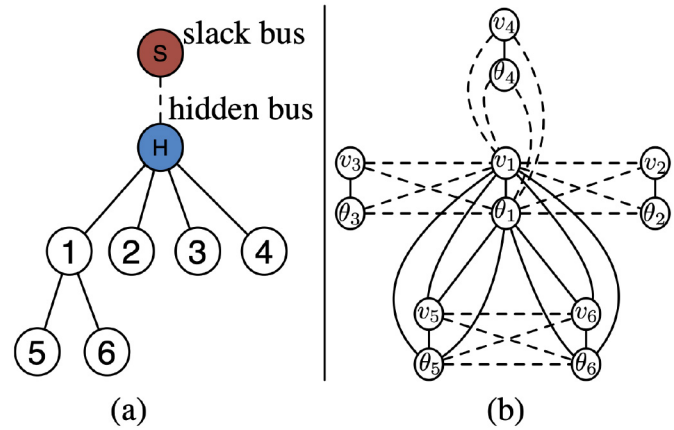

Fig. 2. (a) Radial grid with one hidden bus. (b) Two-hop network for the observed voltage magnitude and angles. Each bus of the radial grid contributes to two nodes in the two-hop network. In (b), two nodal quantities are connected by a solid (dashed) line if their underlying buses in the radial grid are one-hop (two-hop) neighbors.

and for $l \in \{\mathcal{O}\mathcal{O}, \mathcal{O}\mathcal{Z}\}$, we have

$$\begin{aligned} J^{(l)}_{vv} &= H^{(l)}_g[\widetilde{\Sigma}^{(l)}_{qq}H^{(l)}_g - \widetilde{\Sigma}^{(l)}_{pq}H^{(l)}_\beta] - H^{(l)}_\beta[\widetilde{\Sigma}^{(l)}_{pq}H^{(l)}_g - \widetilde{\Sigma}^{(l)}_{pp}H^{(l)}_\beta]; \\ J^{(l)}_{v\theta} &= H^{(l)}_g[\widetilde{\Sigma}^{(l)}_{qq}H^{(l)}_\beta + \widetilde{\Sigma}^{(l)}_{pq}H^{(l)}_g] - H^{(l)}_\beta[\widetilde{\Sigma}^{(l)}_{pq}H^{(l)}_\beta + \widetilde{\Sigma}^{(l)}_{pp}H^{(l)}_g]; \\ J^{(l)}_{\theta v} &= H^{(l)}_\beta[\widetilde{\Sigma}^{(l)}_{qq}H^{(l)}_g - \widetilde{\Sigma}^{(l)}_{pq}H^{(l)}_\beta] - H^{(l)}_g[\widetilde{\Sigma}^{(l)}_{pq}H^{(l)}_g - \widetilde{\Sigma}^{(l)}_{pp}H^{(l)}_\beta]; \\ J^{(l)}_{\theta\theta} &= H^{(l)}_\beta[\widetilde{\Sigma}^{(l)}_{qq}H^{(l)}_\beta + \widetilde{\Sigma}^{(l)}_{pq}H^{(l)}_g] + H^{(l)}_g[\widetilde{\Sigma}^{(l)}_{pq}H^{(l)}_\beta + \widetilde{\Sigma}^{(l)}_{pp}H^{(l)}_g]; \end{aligned}$$

where $\widetilde{\Sigma}^{(l)}_{qq} = (D^{(l)})^{-1}\Sigma^{(l)}_{qq}$, and similarly for $\widetilde{\Sigma}^{(l)}_{pp}$ and $\widetilde{\Sigma}^{(l)}_{pq}$, and $D^{(l)}(i,i) = \Sigma^{(l)}_{pp}(i,i)\Sigma^{(l)}_{qq}(i,i) - \Sigma^{(l)}_{pq}(i,i)\Sigma^{(l)}_{pq}(i,i)$ for diagonal $D^{(l)}$.

Proof: Invoke Assumption 1 and apply block matrix inversion formula on $\Sigma_{\mathbf{x}_\mathcal{O},\mathbf{x}_\mathcal{O}}$ and $\Sigma_{\mathbf{x}_\mathcal{Z},\mathbf{x}_\mathcal{Z}}$ to see that the inverses are block matrices, with each block being a diagonal. The expressions of $J^{(l)}_{vv}, J^{(l)}_{v\theta}$, and $J^{(l)}_{\theta\theta}$ can now be obtained by explicit matrix multiplication. ■

Lemma 1 allows us to compute each entry of S and L (and hence $K_{\mathcal{O}\mathcal{O}}$) using the entries of Laplacian H. The following result characterizes the graph of $K_{\mathcal{O}\mathcal{O}}$.

Theorem 1 (Graphical Model of $K_{\mathcal{O}\mathcal{O}}$: Two-Hop Network): For graph $\mathcal{G}$, the graphical model of $K_{\mathcal{O}\mathcal{O}}$ includes edges between voltage magnitudes and phases that are at the same observed bus, one-hop observed neighbor buses, and two-hop observed neighbor buses.

Proof: We use the proof technique in [11]. We prove the theorem by showing that there is no edge between voltage magnitudes (v) and phases (θ) at the buses $k \in \mathcal{O}$ and $l \in \mathcal{O}$ three or more hops away. This is equivalent to showing that $K_{\mathcal{O}\mathcal{O}}(k,l) = K_{\mathcal{O}\mathcal{O}}(k,2l) = K_{\mathcal{O}\mathcal{O}}(2k,l) = K_{\mathcal{O}\mathcal{O}}(2k,2l) = 0$. We show that $K_{\mathcal{O}\mathcal{O}}(k,l) = 0$, and omit the details for the remaining cases. Recall that $K_{\mathcal{O}\mathcal{O}} = S + L$ and consider $S(k,l) = [J^{(\mathcal{O}\mathcal{O})}_{vv}]_{k,l}$. From Lemma 1, note that $\widetilde{\Sigma}^{(\mathcal{O}\mathcal{O})}_{qq}$, $\widetilde{\Sigma}^{(\mathcal{O}\mathcal{O})}_{qp}$, and $\widetilde{\Sigma}^{(\mathcal{O}\mathcal{O})}_{pp}$ are all diagonals. Instead, the non-zero entries in $H^{(\mathcal{O}\mathcal{O})}_g$ (and $H^{(\mathcal{O}\mathcal{O})}_\beta$) are the diagonal terms and the (i,j)-th entries for which $i \in \mathcal{O}$ and $j \in \mathcal{O}$ are neighbors. By invoking this observation and simplifying S, we see that $S(k,l) = 0$ if k and l are neither neighbors nor have an observed neighbor in common. Similarly, $L(k,l) = 0$ if k and l do not have an hidden neighbor in common. ■

Note that the *two-hop neighbors* in a graph are the nodes connected by a shortest path of length two. Instead, *two-hop network* has at most two-hop neighbors. Theorem 1 says that the graph of K_{OO} includes (spurious) edges between observed nodal quantities even if their corresponding buses do not have an edge in the grid; see Fig. 2. Note that unlike $K_{\mathcal{O}\mathcal{O}}$, the observed inverse covariance matrix $\Sigma^{-1}_{\mathbf{y}_{\mathcal{O}},\mathbf{y}_{\mathcal{O}}}$ need not have a two-hop network structure [11].

A. Topology Identification

We present an algorithm that identifies the complete grid topology—that is, the edge connections among the observed buses, between observed and hidden buses, and among the hidden buses—using the components S and L of $K_{\mathcal{O}}$ (9). We begin with an assumption that is necessary to uniquely identifying the true topology [11], [13].

Assumption 2: In grid $\mathcal{G}$, hidden buses are neither the leaf nodes nor adjacent to each other.

Theorem 2 (Sign Based Rules for Topology Identification): For grid $\mathcal{G}$ with a minimum cycle length (MCL) greater than three, consider S and L in (9). Then,

1) $[J^{(\mathcal{O}\mathcal{O})}_{vv} + J^{(\mathcal{O}\mathcal{O})}_{\theta\theta}]_{i,j} < 0$ **iff** (ij) is a true edge in $\mathcal{G}$.
2) $[J^{(\mathcal{O}\mathcal{O})}_{vv} + J^{(\mathcal{O}\mathcal{O})}_{\theta\theta}]_{i,j} > 0$ **iff** there is a path $i{-}k{-}j$ linking the observed buses i and j, and k is an observed bus.
3) $[J^{(\mathcal{O}\mathcal{Z})}_{vv} + J^{(\mathcal{O}\mathcal{Z})}_{\theta\theta}]_{i,j} > 0$ **iff** there is a path $i{-}k{-}j$ linking the observed buses i and j, where k is a hidden bus.

Proof: For part (i), we proceed as follows. If part: let i and j be the observed neighbor buses and $(ij) \in \mathcal{G}$. As the minimum cycle length is greater than three, there are no common neighbors of i and j. Thus, from Lemma 1, we have $[J^{(\mathcal{O}\mathcal{O})}_{vv} + J^{(\mathcal{O}\mathcal{O})}_{\theta\theta}]_{i,j} < 0$. Here, we used the fact that $\widetilde{\Sigma}^{(l)}_{qq}$, $\widetilde{\Sigma}^{(l)}_{pp}$, and $\widetilde{\Sigma}^{(l)}_{pq}$ are diagonals with positive entries. Only if: we prove the contrapositive. Suppose i and j are not neighbors. If i and j are more than two-hops away, from Theorem 1, we have $[J^{(\mathcal{O}\mathcal{O})}_{vv}]_{i,j} = [J^{(\mathcal{O}\mathcal{O})}_{\theta\theta}]_{i,j} = 0$. If i and j are two-hop neighbors with a common observed neighbor, then from Lemma 1 it follows that $[J^{(\mathcal{O}\mathcal{O})}_{vv} + J^{(\mathcal{O}\mathcal{O})}_{\theta\theta}]_{i,j} > 0$. Similarly, we can prove (ii) and (iii), and the details are omitted. ■

If $\mathcal{Z} = \emptyset$, [11, Th. 4] follows a corollary to Theorem 2. Further, Theorem 2 (i) allows us to uniquely identify the connectivity among any pair of observed buses. Instead, Theorem 2 (ii)-(iii) can be used to identify connectivity between an observed and hidden bus. Finally, from Assumption 2, notice that no two hidden buses share an edge. Thus, we have the complete grid topology. We summarize these steps in Algorithm 1, which requires components of S, L (9) as an input. However, recall that we have access only to $\Sigma^{-1}_{\mathbf{y}_{\mathcal{O}},\mathbf{y}_{\mathcal{O}}}$, but not to its decomposition (8). The thresholds τ_1 and τ_2 help us counteract the bias introduced by the sample estimate of $\Sigma^{-1}_{\mathbf{y}_{\mathcal{O}},\mathbf{y}_{\mathcal{O}}}$. We discuss these issues in the following sections.

Theorem 2 vs. [13, Th. 3]: Note that Theorem 2 and [13, Th. 3] primarily differs on the assumptions of size of MCL. In particular, we require MCL to be at least four and [13, Th. 3] requires MCL to be at least five. We highlight that MCL of size four is the theoretically possible limit on uniquely identifying topology using measurements alone. In other words, there exist topologies with MCL size at most three for which topology identification is not possible. Hence, our Algorithm 1 is applicable to several classes of distribution grids than [13, Algorithm 1].

Algorithm 1: Grid Topology Identification

Input: Matrices $J^{(\mathcal{O}\mathcal{O})}_{vv}$, $J^{(\mathcal{O}\mathcal{O})}_{\theta\theta}$, $J^{(\mathcal{O}\mathcal{Z})}_{vv}$, and $J^{(\mathcal{O}\mathcal{Z})}_{\theta\theta}$; o:= dimension of $\Sigma^{-1}_{\mathbf{y}_{\mathcal{O}},\mathbf{y}_{\mathcal{O}}}$; z := number of hidden nodes; and thresholds $\tau_1, \tau_2 > 0$.
Output: Reconstructed graph $\widehat{G} = (\widehat{\mathcal{V}}, \widehat{\mathcal{E}})$

```
initialization: V̂_o = {1, ..., o} and Ê_o = {}
for i ∈ V̂_o, i < j do
    if [J_vv^(OO) + J_θθ^(OO)]_{i,j} < −τ_1 then Ê_o ← {(i, j)}
end
initialization: V̂ ← V̂_o and Ê ← Ê_o
counter k = 1
for i ∈ V̂_o, i < j do
    if [J_vv^(OZ) + J_θθ^(OZ)]_{i,j} > τ_2 then
        V̂ ← V̂ ∪ {o + k}; Ê ← Ê ∪ {(i, o + k)}
        k = k + 1
    end
    if k == z then break
end
```

Remark 1: The statement of Theorem 2 may fail to hold for a specific values of H (4) that lie in a zero measure set. However, for all practical purposes, these specific values should not hinder the applicability of our results.

IV. A Convex Framework to Decompose Observed Inverse Covariance Matrix of Voltages

This section presents an optimization framework to extract matrices—S, L, and the number of hidden nodes—which are inputs to Algorithm 1, from the sample covariance matrix

$$\widehat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}\Big(\mathbf{y}^{(t)}_{\mathcal{O}} - \overline{\mathbf{y}}_{\mathcal{O}}\Big)\Big(\mathbf{y}^{(t)}_{\mathcal{O}} - \overline{\mathbf{y}}_{\mathcal{O}}\Big)^{\mathsf{T}}, \tag{10}$$

where $y^{(t)}$ are the i.i.d samples, and $\overline{\mathbf{y}}_{\mathcal{O}} = T^{-1}\sum_{t=1}^{T} y^{(t)}$ is the sample mean. We discuss a few properties of the triple (S, L, M) (8) that helps us decompose $\widehat{\Sigma}^{-1}$ into $\widehat{S} + \widehat{L} - \widehat{M}$, where the hatted terms are the sample estimates.

Recall that distribution grids have sparse connections [1]. Thus, the Laplaican H is sparse, and hence S (7) is sparse. Instead, ranks of L and M are much smaller than that of S. In fact, from (7) and (8), we have $\text{Rank}(L), \text{Rank}(M) \leq 2z$; where as, from Assumption 2, $\text{Rank}(S) = 2o >> 2z$.

Further, M can be dense (non-sparse), provided there are sufficient number of connections between *observed* neighbors and non-neighbors of a hidden bus; see Fig. 3; Instead, L, although non-sparse, admits a low-rank factorization with sparse factors. To see this, let supp(M) be the support of M: a $(0, 1)$-matrix with (i, j)-th entry equal to 1 if $[M]_{i,j} \neq 0$, and equal to zero, otherwise.

Proposition 1: (a) There is a unique B with $\Sigma^{-1}_{\mathbf{x}_{\mathcal{Z}},\mathbf{x}_{\mathcal{Z}}} = B^2$ such that $L = (H_{\mathcal{O}\mathcal{Z}}B)(H_{\mathcal{O}\mathcal{Z}}B)^{\mathsf{T}}$ and $\text{supp}(H_{\mathcal{O}\mathcal{Z}}) = \text{supp}(H_{\mathcal{O}\mathcal{Z}}B)$, and (b) the i- and $2i$-th columns of $H_{\mathcal{O}\mathcal{Z}}$ are each (at most) $2s_i$-sparse, where s_i is the number of *observed* neighbors of the i-th hidden bus.

Proof (Part (a)): From Assumption 1, note that $\Sigma_{\mathbf{x}_{\mathcal{Z}},\mathbf{x}_{\mathcal{Z}}}$ is 2×2 block matrix with blocks being diagonal (call this

property $\mathcal{P}$), and so is $\Sigma^{-1}_{\mathbf{x}_Z,\mathbf{x}_Z}$ (follows by direct evaluation). Thus, there is a permutation matrix Π and a block diagonal matrix M such that $\Sigma^{-1}_{\mathbf{x}_Z,\mathbf{x}_Z} = \Pi M \Pi^\mathsf{T}$. Let $M = \widetilde{B}^2$, for a block diagonal $\widetilde{B}$, and define $B = (\Pi \widetilde{B} \Pi^\mathsf{T})$. Then, $\Sigma^{-1}_{\mathbf{x}_Z,\mathbf{x}_Z} = B^2$, where we use the identity $\Pi\Pi^\mathsf{T} = I$. Note that B not only satisfies the property $\mathcal{P}$, but is unique because $\Sigma^{-1}_{\mathbf{x}_Z,\mathbf{x}_Z}$ is a positive definite matrix. Using these facts, we have $\text{supp}(H_{\mathcal{O}\mathcal{Z}}) = \text{supp}(H_{\mathcal{O}\mathcal{Z}}B)$. **Part (b)**: For a hidden bus $i \in \mathcal{Z}$, from (3) and (4), the i-th (or $2i$-th) column of $H_{\mathcal{O}\mathcal{Z}}$ is the concatenation of i-th column of $H_g^{(\mathcal{O}\mathcal{Z})}$ and that of $H_\beta^{(\mathcal{O}\mathcal{Z})}$, each column being s_i-sparse. Thus, the i- and $2i$-th columns of $H_{\mathcal{O}\mathcal{Z}}$ are each $2s_i$-sparse. ■

From Proposition 1 we note that L is a low-rank matrix with sparse factors, provided s_i is not large; we illustrate this property in Fig. 3. Here $L = \mathbf{u}\mathbf{u}^\mathsf{T}$, for $\mathbf{u}^\mathsf{T} = [\mathbf{u}_1, \mathbf{u}_2]$, and $\mathbf{u}_j = [\star\ \star\ \star\ \star 0\ 0\ \star\ \star\ \star\ \star\ 0\ 0]^\mathsf{T}$ is 8-sparse, where $\star$ means a non-zero entry.

Before we present our optimization framework that allows us recover $(\widehat{S}, \widehat{L}, \widehat{M})$ from (10), we recall the notion of an *atomic norm*—when used as a regularizer in (12) promotes low-rank factorization property of L. To that end, let $X \in \mathbb{R}^{m\times m}$ satisfy $X = \sum_{i=1} c_i a_i$, where $c_i \geq 0$; $a_i \in \mathcal{A} \subseteq \mathbb{R}^{m\times m}$ are the *atoms*; and $\mathcal{A} \subset \mathbb{R}^{m\times m}$ is the *atomic* set. Then, the *atomic norm* [15] of X is given by

$$\|X\|_{\mathcal{A}} = \inf\left\{\sum_i c_i : X = \sum_{i=1} c_i a_i, c_i \geq 0, a_i \in \mathcal{A}\right\}.$$

The *atomic norm* returns the minimal sum of weights c_i over all decompositions of X with respect to the set $\mathcal{A}$. For example, the atomic set of trace norm (sum of singular values of A) is the set of rank-one matrices with unit norm: $\mathcal{A} = \{\mathbf{u}\mathbf{v}^\mathsf{T}, \|\mathbf{u}\|_2 = \|\mathbf{v}\|_2 = 1\}$. Thus, smaller the trace-norm, smaller the Rank(X). However, we require L to be a low-rank matrix with sparse factors. To capture this structure, let $\mathcal{A} = \cup_k \mathcal{A}_k$, where $\mathcal{A}_k = \{\mathbf{u}\mathbf{u}^\mathsf{T} : \|\mathbf{u}\|_2 = 1, \|\mathbf{u}\|_0 \leq k\}$, and consider the resulting *structured atomic norm* [16], [17]

$$\Omega(X) = \inf\left\{\sum_i \sum_{k=1}^m \omega_k c_{i,k} : X = \sum_i \sum_{k=1}^m c_{i,k} a_{i,k}, \right. \\ \left. \times\ c_{i,k} \geq 0, \text{ and } \mathbf{u}_{i,k} \in \mathcal{A}_k\right\}, \quad (11)$$

where $\|\cdot\|_2$ is the Euclidean norm, and the norm $\|\mathbf{u}\|_0$ counts the number of non-zero entries in $\mathbf{u}$. The weights $w_k \geq 0$ can be used to give importance to the sparsity level k. Higher the weight, sparser the columns of optimal X.

A. Convex Optimization Framework

Consider the following convex optimization problem that recovers S, L, and E with aforementioned properties

$$(\widehat{S}, \widehat{L}, \widehat{M}) = \arg\min_{S,L,M} \ell(S + L - M; \widehat{\Sigma}) + \mathcal{R}(S, L, M) \\ \text{s.t.} \quad S + L - M \succeq 0, L \succeq 0, M \succeq 0, \quad (12)$$

where $\ell(\cdot)$ is a loss function, which we shall discuss below, and the structure promoting regularizer $\mathcal{R}(\cdot)$ is given by

$$\mathcal{R}(S, L, M) = \lambda_1 \|S\|_1 + \lambda_2 \Omega(L) + \lambda_3 \text{tr}(M), \quad (13)$$

Algorithm 2: Alternating Minimization

Input: Sample covariance matrix $\widehat{\Sigma} \in \mathbb{R}^{2o\times 2o}$; maximum iterations T_{in} and T_{out}.

Intialization: $\widehat{S}_0 = \mathbf{0}; \widehat{L}_0 = \mathbf{0}$; and $\widehat{E}_0 = \mathbf{0}$.

1 **for** $t = 1{:}T_{out}$ **do**

2 Fix $\widehat{L}_{t-1}$, and apply ADMM on (12) for T_{in} iterations to compute S_t and E_t.

3 Fix $\widehat{S}_{t-1}$ and $\widehat{E}_{t-1}$, and apply FCG on (12) to compute $\widehat{L}_t$.

4 **end**

Return: The recovered matrices $(\widehat{S}_t, \widehat{L}_t, \widehat{E}_t)$.

and $\lambda_i \geq 0$. The $\|S\|_1 := \sum_{i,j} |s_{ij}|$ norm enforces sparsity among the observed buses, the tr(M)-norm enforces the low-rank property, and the atomic norm $\Omega(L)$ enforces L to be a low-rank matrix with sparse factors.

In Gaussian graphical model selection problems, for $K \succ 0$, one lets $\ell(K; \hat{\Sigma}) = -\log\det(K) + \text{tr}(K\Sigma)$, and solve (12) using proximal methods, such as the alternating direction multiplier method (ADMM) [17]. However, evaluating the proximal operator of $\Omega(L)$ is computationally demanding. To overcome this problem, we solve (12) using an alternating minimization method (see below) that combines ADMM and the fast column generation (FCG) method—a variant of Frank-Wolfe method [15]. We use the quadratic loss function [18]

$$\ell(K; \widehat{\Sigma}) = \frac{1}{2}\text{tr}(K\widehat{\Sigma}K) - \text{tr}(K). \quad (14)$$

Here $\ell(K; \widehat{\Sigma})$ is convex in K, and the unique minimizer of $\ell(K; \widehat{\Sigma})$ occurs at $\widehat{\Sigma}^{-1}$; see [18] for more details.

We provide only a brief description of alternating minimization method in Algorithm 2. For the line 2 in Algorithm 2, we use the standard ADMM (the detailed steps are given in [19]). For Line 3 in Algorithm 2, we use the FCG method (the detailed steps are given in [15, Algorithm 1]).

Remark 2 (Ineffectiveness of Sparse and (One) Low Rank Decomposition): In [13], the authors use a sparse plus trace norm regularizer to decompose $\Sigma^{-1}_{\mathbf{y}_\mathcal{O},\mathbf{y}_\mathcal{O}}$ into matrices $(S + L)$ and M. For this naïve regularizer to work, matrix $S + L$ needs to be sufficiently sparse. However, for sparsely connected distribution networks, $S + L$ maybe non-sparse. In fact, in Fig. 3, for a sparse radial network, we demonstrate that $S+L$ is non-sparse (the first column is full). Instead, in the right-bottom of Fig. 3, we show that S is sparse and L is low-rank with sparse factors (also see text below Proposition 4.1). From this discussion, we conclude that one should consider sophisticated regularizers, such as (13), to accurately decompose the inverse covariance matrices that arise from structured models, such as the power flow model (2).

V. Numerical Simulations

We evaluate and compare the performance of Algorithm 1 with that of [13, Algorithm 1] on a modified IEEE 33 bus distribution grid with three hidden nodes; see Fig. 4. We add extra edges to ensure that the grid's minimum cycle length is greater than three and the hidden nodes are not adjacent to each other. The power injection $(\mathbf{x}_\mathcal{O}, \mathbf{x}_\mathcal{Z})$ is modeled as an i.i.d. zero-mean Gaussian random vector with covariance matrix $\sigma^2 I_{2N\times 2N}$ (where $N = 32$) and $\sigma^2 = 0.1$. The observed

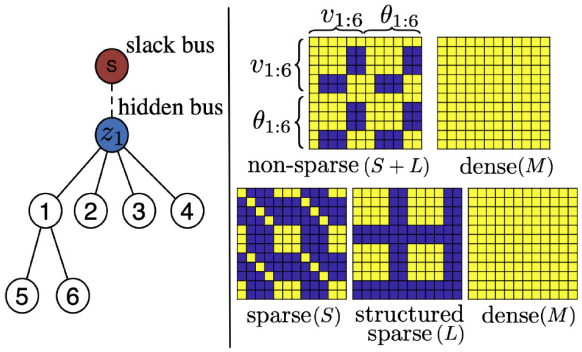

Fig. 3. (Left) radial grid with one hidden bus and six observed buses. (Right) visualization of components of $\Sigma^{-1}_{\mathbf{y}_{\mathcal{O}},\mathbf{y}_{\mathcal{O}}}$: blue and yellow boxes indicate zero and non-zero entries. (Right-top) decomposition of $\Sigma^{-1}_{\mathbf{y}_{\mathcal{O}},\mathbf{y}_{\mathcal{O}}}$ into non-sparse and dense low-rank matrices. (Right-bottom) decomposition of $\Sigma^{-1}_{\mathbf{y}_{\mathcal{O}},\mathbf{y}_{\mathcal{O}}}$ into a sparse matrix, a low-rank matrix with sparse factors, and a dense low-rank matrix.

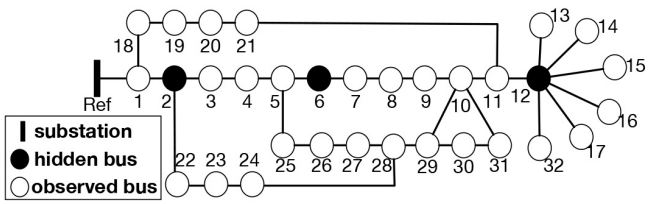

Fig. 4. Modified IEEE 33 bus system with a minimum cycle length of size four. Bus 1 is the reference bus. The three hidden buses, labeled {3, 7, 13}, are more than two-hops away.

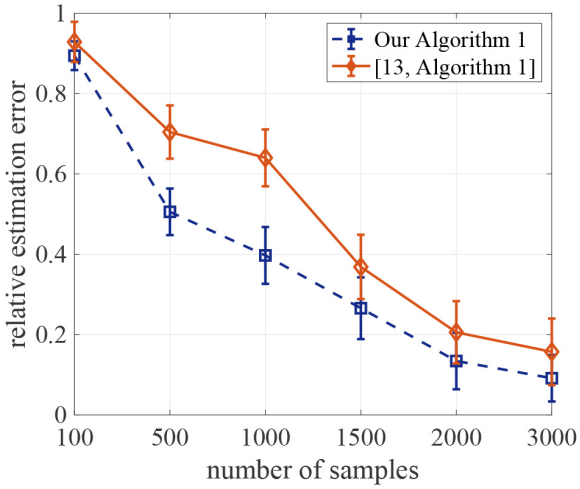

Fig. 5. Relative estimation error in Algorithm 1 and Algorithm used in [13] for the reconstructed topology over 10 independent trails.

voltage samples $\mathbf{y}_{\mathcal{O}}$ are generated from the linear model (1). Using these samples, we obtain matrices S and L, given as an input to Algorithm 1, by solving optimization problem (12). Instead, as suggested in [13], we employ a low-rank plus sparse decomposition method to obtain matrix $S+L$, given as an input [13, Algorithm 1]. Finally, we set $\tau_1 = \tau_2 = 0.1$. We assess the accuracy of both the algorithms by relative estimation error (REE): the ratio of the sum of false and missed edges to that of the number of true edges in $\mathcal{E}$. Fig. 5 shows that the REE for both the algorithms decrease with an increase in the sample size. Furthermore, Algorithm 1 has a minimum REE compared with that of [13, Algorithm 1].

VI. Conclusion

This letter provides an algorithm to identify the unknown topology of a distribution grid from bus voltage magnitude and angle measurements, while accounting for hidden buses. Our identification algorithm relies on the sign patterns of the sparse and low-rank (with sparse factors) components of the inverse covariance matrix of observed voltages. By carefully studying the nuanced structure of the inverse covariance matrix, we relaxed the existing minimum cycle length conditions [13]. We validated our algorithm's performance on a modified IEEE-33 bus test system. Future works include obtaining theoretical conditions (e.g., subspace incoherency condition) under which the inverse covariance matrix can be decomposed into the required structural components.

References

[1] O. Ardakanian *et al.*, "On identification of distribution grids," *IEEE Control Netw. Syst.*, vol. 6, no. 3, pp. 950–960, Sep. 2019.

[2] J. Zhang and L. Sankar, "Physical system consequences of unobservable state-and-topology cyber-physical attacks," *IEEE Trans. Smart Grid*, vol. 7, no. 4, pp. 2016–2025, Jul. 2016.

[3] K. Moffat, M. Bariya, and A. Von Meier, "Unsupervised impedance and topology estimation of distribution networks—Limitations and tools," *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 846–856, Jan. 2020.

[4] Y. Liao, Y. Weng, G. Liu, and R. Rajagopal, "Urban MV and LV distribution grid topology estimation via group lasso," *IEEE Trans. Power Syst.*, vol. 34, no. 1, pp. 12–27, Jan. 2019.

[5] S. Bolognani, "Chapter 13—Grid topology identification via distributed statistical hypothesis testing," in *Big Data Application in Power Systems*. Amsterdam, The Netherlands: Elsevier, 2018, pp. 281–301.

[6] Y. Weng, Y. Liao, and R. Rajagopal, "Distributed energy resources topology identification via graphical modeling," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 2682–2694, Jul. 2017.

[7] E. Fabbiani, P. Nahata, G. De Nicolao, and G. Ferrari-Trecate, "Identification of AC networks via online learning," 2020. [Online]. Available: arXiv:2003.06210.

[8] G. Cavraro and V. Kekatos, "Graph algorithms for topology identification using power grid probing," *IEEE Control Syst. Lett.*, vol. 2, no. 4, pp. 689–694, Oct. 2018.

[9] G. Cavraro, A. Bernstein, V. Kekatos, and Y. Zhang, "Real-time identifiability of power distribution network topologies with limited monitoring," *IEEE Control Syst. Lett.*, vol. 4, no. 2, pp. 325–330, Apr. 2020.

[10] S. Bolognani, N. Bof, D. Michelotti, R. Muraro, and L. Schenato, "Identification of power distribution network topology via voltage correlation analysis," in *Proc. IEEE Conf. Decis. Control*, Firenze, Italy, 2013, pp. 1659–1664.

[11] D. Deka, S. Talukdar, M. Chertkov, and M. V. Salapaka, "Graphical models in meshed distribution grids: Topology estimation, change detection & limitations," *IEEE Trans. Smart Grid*, vol. 11, no. 5, pp. 4299–4310, Sep. 2020.

[12] G. Cavraro, V. Kekatos, L. Zhang, and G. B. Giannakis, "Learning power grid topologies," in *Advanced Data Analytics for Power Systems*, Cambridge, U.K.: Cambridge Univ. Press, 2021, pp. 3–27.

[13] H. Doddi, D. Deka, and M. Salapaka, "Learning partially observed meshed distribution grids," in *Proc. Int. Conf. Probab. Methods Appl. Power Syst. (PMAPS)*, Liege, Belgium, 2020, pp. 1–6.

[14] S. Lauritzen, *Graphical Models*. New York, NY, USA: Clarendon, 1996.

[15] M. Vinyes and G. Obozinski, "Fast column generation for atomic norm regularization," in *Proc. 20th Int. Conf. Artif. Intell. Stat.*, vol. 54, Apr. 2017. pp. 547–556.

[16] M. Vinyes and G. Obozinski, "Learning the effect of latent variables in Gaussian graphical models with unobserved variables," 2018. [Online]. Available: arXiv:1807.07754.

[17] D. A. Tarzanagh and G. Michailidis, "Estimation of graphical models through structured norm minimization," *J. Mach. Learn. Res.*, vol. 18, no. 209, pp. 1–48, 2018.

[18] T. Zhang and H. Zou, "Sparse precision matrix estimation via lasso penalized D-trace loss," *Biometrika*, vol. 101, no. 1, pp. 103–120, 2014.

[19] C. Wu, H. Zhao, H. Fang, and M. Deng, "Graphical model selection with latent variables," *Electron. J. Stat.*, vol. 11, no. 2, pp. 3485–3521, 2017.