

Tractable Learning in Underexcited Power Grids

Deepjyoti Deka , Senior Member, IEEE, Harish Doddi , Sidhant Misra ,
and Murti V. Salapaka , Fellow, IEEE

Abstract—Estimating the structure of physical flow networks, such as power grids, is critical to secure delivery of energy. This article discusses statistical structure estimation in power grids in the “underexcited” regime, where a subset of internal nodes has zero injection fluctuations. Prior estimation algorithms based on nodal voltages fail for such grids as the voltage covariance matrix is not invertible. We propose a novel topology learning algorithm for learning underexcited general networks. Our algorithm uses physics-informed conservation laws to first identify the zero-injection buses and their neighbors, and then estimates the remaining edges in the grid. We prove the asymptotic correctness of our algorithm for grids with non-adjacent internal zero-injection nodes. More important, we theoretically analyze our algorithm’s efficacy under noisy measurements, and determine bounds on maximum noise under which asymptotically correct recovery is guaranteed. Our approach is validated through simulations with voltage samples generated on test distribution grids with real injection data and nonlinear power flow models.

Index Terms—Distribution grid, flow conservation, inverse covariance graph, inverse graph Laplacian, Matpower, noisy regression.

I. INTRODUCTION

TOPOLOGY estimation is a crucial part of situational awareness in power grids, which is the backbone for energy needs in the modern economy [1]. Operationally, the power grid can be distinguished between high-voltage transmission grids and low-voltage distribution grids [2]. Structurally, the transmission grids are meshed/loopy (have cycles), while distribution grids are primarily radial (tree-like). In either setting, the grid topology is established by the status of breakers/switches on an underlying set of permissible lines/edges as shown in

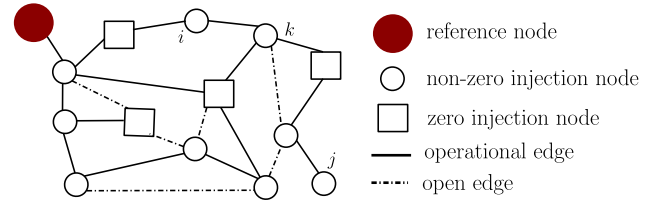


Fig. 1. Loopy power grid with internal (i, k) and terminal (j) nodes.

Fig. 1. In the modern grid, the presence of stochastic resources, such as solar energy, wind farms, and aggregated controllable loads, has made topology and state estimation with guarantees of paramount importance for secure operation. This problem is exacerbated in distribution grids due to the absence of sufficient real-time line measurements [3]. In recent years, researchers [4] have proposed the use of nodal voltage measurements collected from high-fidelity nodal measurement devices, such as phasor measurement units (PMUs) [5], micro-PMUs [6], and advanced metering infrastructure (AMI) to aid in topology estimation in distribution grids. Similarly, voltage-based analytics for topology change detection and fault localization have been proposed as use-cases for PMUs in transmission grids, to aid traditional relay-based methods [7]. Among algorithms with theoretical guarantees, passive greedy schemes [1], [8], [9] and active probing methods [10], [11] that employ voltage measurements are proposed for learning radial distribution grids. However, they do not generalize to meshed/loopy networks in transmission grids and urban distribution grids [12], [13]. Similarly, algorithms for joint estimation of topology and line parameters using both voltage, and active and reactive injection measurements have been designed using least-squares fitting [14] and low-ranked decomposition [15]. Unlike these related works, in this article, we focus on the problem of topology recovery using only nodal voltage measurements, both magnitude and/or phase, collected from PMUs and AMI, without any additional measurement or statistics of nodal injections. Note that such devices, currently available in large transmission grids, are being extensively deployed in distribution grids as well. In the line of work on voltage-based topology estimation, consistent algorithms based on inverse covariance or mutual independence (MI) of nodal voltages are used for learning grid topology in both radial [16]–[18] and loopy settings [19], [20]. Its theoretical generalizability notwithstanding, the validity of inverse voltage covariance or MI-based methods relies on a key assumption: all nodes/buses are excited, i.e., have nonzero injection fluctuations. This is necessary to ensure that the voltage covariance matrix

Manuscript received March 13, 2021; revised June 16, 2021; accepted August 29, 2021. Date of publication November 2, 2021; date of current version June 20, 2022. This work was supported in part by the U.S. Department of Energy (DOE) Advanced Grid Modeling (AGM) Program, in part by an internal LDRD DR project “The Optimization of Machine Learning,” and in part by the U.S. DOE’s ARPA-E program under Grant DE-AR0001016. Recommended by Associate Editor Ioannis Lestas. (Corresponding author: Deepjyoti Deka.)

Deepjyoti Deka and Sidhant Misra are with Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87544 USA (e-mail: deepjyoti@lanl.gov; sidhant@lanl.gov).

Harish Doddi is with the Department of Mechanical Engineering, University of Minnesota, Minneapolis, MN 55414 USA (e-mail: doddi003@umn.edu).

Murti V. Salapaka is with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: murtis@umn.edu).

Digital Object Identifier 10.1109/TCNS.2021.3124299

is full-ranked and, hence, invertible. While terminal nodes in grids generally have loads or generation, an unidentified set of internal nodes with zero injection may be present in realistic transmission and distribution grids for routing power flows [2]. In such settings, voltage measurement-based learning algorithms proposed in prior work do not extend theoretically. The overarching goal of this work is to overcome this drawback and develop a provably correct topology learning algorithm using only *nodal voltages* from a *general (loopy)* power grid, where an *unidentified* subset of internal nodes may be *unexcited*. As linearized power flow models are examples of structural equation models (SEM) [21] with external excitation, our work enables tractable learning for SEMs satisfying conservation laws, in the underexcited regime. Extension of our work to loopy grids with partial voltage availability, but supplemented with additional injection measurements, is the objective of our future work.

A. Contribution

Our learning algorithm is based on the physics of linearized power flows and uses nodal voltage measurements. For the dc power model, this refers to measurements of voltage phase angles, while for the ac power model, measurements imply voltage magnitude and phase angles. As the voltage covariance matrix is rank-deficient in the underexcited grid, we propose a two-step learning approach. In the *first* step, we develop a regression-based test that identifies the internal nodes of zero injection, and subsequently learns their true neighbors. In the *second* step, we extract the voltage covariance pertaining only to nodes with nonzero injection, and estimate the rest of the network. We prove that under noiseless measurements, our algorithm correctly estimates the underlying topology of radial grids, provided internal nodes with zero-injection are nonadjacent. For loopy grids of minimum loop-size four, our algorithm ensures that correct learning provided loops of size four do not include zero-injection buses. Note that the minimum loop size of four is nonrestrictive as real grids have large girth [12], and is necessary for correct estimation even in the fully excited setting [20]. We show the necessity of our assumptions for unique and consistent recovery, through multiple counter examples. Under noisy data, we determine the maximum noise margin for consistency of each step of our learning algorithm, and express this margin as an intuitive function of system parameters. We validate the developed algorithms through experiments with linear and nonlinear power-flow voltage samples from test grids with real injection data.

The next section discusses power flows in grids and related voltage statistics. Section III discusses properties of nodal voltages which are used to extract zero-injection buses and their neighbors. Section IV determines edges between nonzero injection buses and completes the topology estimation. The analysis under noise for each step is provided in its corresponding section. Section V includes simulations results on IEEE test networks, using nonlinear samples generated using real injection data. Finally, Section VI concludes this article.

II. PRELIMINARIES

A. Notation

We denote the power grid topology by an undirected graph $\mathcal{G} \equiv (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{i, j, \dots\}$ is the set of $N + 1$ nodes/nodes, and $\mathcal{E} = \{(ij), (jk), \dots\}$ is the set of operational lines/edges (see Fig. 1). Two nodes that share an edge are called “neighbors.” The degree of a node is the total number of edges that involve the node. A node with degree 1 is called a “terminal node,” and an “internal” node otherwise. A “loop” of size $l \geq 3$ refers to a set of distinct edges $\{(ij_1), (j_1j_2), \dots, (j_{l-1}i)\}$. A radial (loop-less) graph is defined to have loops of minimum size ∞ . Each edge (ij) in grid $\mathcal{G}$ is associated with susceptance $\beta_{ij} > 0$, and conductance $g_{ij} > 0$. Each node i has a corresponding voltage magnitude v_i , phase angle θ_i , active power p_i , and reactive power q_i . We use v, θ, p , and q to denote the corresponding vectors with entries for each node.

B. Power Grid Basics

The Kirchoff’s law for ac power flow (ac-PF) is given by

$$\forall i \in \mathcal{V}, p_i + \hat{q}_i = [lr] \sum_{j:(ij) \in \mathcal{E}} (v_i^2 - v_i v_j e^{i\theta_i - i\theta_j}) (g_{ij} + i\beta_{ij}). \quad (1)$$

Considering small deviations in v from a reference bus voltage and small differences in θ between neighboring nodes, the nonlinear ac-PF can be linearized in v, θ to obtain the *linearized* ac power flow (LC-PF) [1], [22], [23] where

$$p + \hat{q} = H_{g+\hat{i}\beta}(v - \hat{i}\theta) \quad \text{with} \quad \begin{bmatrix} p \\ q \end{bmatrix} = \begin{bmatrix} H_g & H_\beta \\ H_\beta & -H_g \end{bmatrix} \begin{bmatrix} v \\ \theta \end{bmatrix}. \quad (2)$$

The susceptance weighted symmetric Laplacian matrix H_β is given by

$$H_\beta(i, j) = \begin{cases} \sum_{k:(ik) \in \mathcal{E}} \beta_{ik} & \text{if } i = j \\ -\beta_{ij} & \text{if } (ij) \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases}. \quad (3)$$

H_g is defined similarly. If voltage magnitude deviations are negligible, or lines are primarily susceptive, the linear *dc power flow* (dc-PF) is derived [22]

$$p_i = [lr] \sum_{j:(ij) \in \mathcal{E}} \beta_{ij} (\theta_i - \theta_j) \Rightarrow p = H_\beta \theta. \quad (4)$$

Note that both LC-PF and dc-PF models are lossless. One node’s voltage can be taken as reference to measure all other voltages, while its injection is given by the negative sum of all other injections. The PF models can then be reduced to the N nonreference nodes by removing entries in p, q, v, θ , and rows and columns in H_β, H_g that correspond to the reference node. From this point onward, we use p, q, v, θ, H_β , and H_g to refer to their reduced versions. As reduced H_β, H_g have full rank, LC-PF and dc-PF models are invertible. In the rest of the article, we focus on dc-PF for our theoretical results. Our analysis naturally extends to the

LC-PF model, as noted in subsequent sections. Furthermore, our simulation results demonstrate the performance of our learning algorithm on nonlinear ac-PF voltage measurements.

C. Structure Recovery Problem

The structure recovery problem pertains to estimating the set of edges $\mathcal{E}$ of $\mathcal{G}$ using the measurements $\{\theta_i^t, i \in \mathcal{V}, t = 1, \dots, T\}$ of the phase angles (and voltages for the LC-PF model). For the theoretical analysis, we consider the asymptotic setting $T \rightarrow \infty$, that ensures a sufficient number of collected voltage observations for correct estimation of the second-order statistics of the nodal voltages. Our simulation results demonstrate the performance in the low-sample regime. In an underexcited grid $\mathcal{G}$, we denote the set of internal nodes with zero-injection by $\mathcal{U}$, and the remaining nodes in $\mathcal{V} - \mathcal{U}$ by $\mathcal{W}$, respectively. Without a loss of generality, we consider the nodes numbered 1 to $|\mathcal{U}|$ ($|\mathcal{U}|$ denotes the cardinality of set $\mathcal{U}$) as zero-injection and rest as nonzero injection nodes. We partition $p = \begin{bmatrix} \mathbf{0} \\ p^{\mathcal{W}} \end{bmatrix}$, and $\theta = \begin{bmatrix} \theta^{\mathcal{U}} \\ \theta^{\mathcal{W}} \end{bmatrix}$. We assume the following regarding covariance nodal injections.

Assumption 1: Fluctuations in $p^{\mathcal{W}}$ are uncorrelated

$$\Sigma_p = \begin{bmatrix} 0 & 0 \\ 0 & \Sigma_{p^{\mathcal{W}}} \end{bmatrix}$$

and $\Sigma_{p^{\mathcal{W}}} = \mathbb{E}[p^{\mathcal{W}} p^{\mathcal{W}T}] \succ 0$ is diagonal.

Assumption 1 is used in prior work [16]–[18] for tractable grid learning. Further, it also holds for real loads and renewables [19], [24], when small linear trends are empirically detrended. While Assumption 1 is used for proving the correctness of a part of our algorithm (Section IV), we consider real correlated injection data in our numerical simulations to demonstrate accurate learning, in its absence.

The relevant statistics is the covariance matrix of the observables Σ_θ , which under the dc-PF (4) is given by

$$\Sigma_\theta = \mathbb{E}[\theta\theta^T] = J_\beta \begin{bmatrix} 0 & 0 \\ 0 & \Sigma_{p^{\mathcal{W}}} \end{bmatrix} J_\beta^T, \text{ where } J_\beta = H_\beta^{-1}. \quad (5)$$

A similar formula for the voltage covariance $\Sigma_{(v,\theta)} = \begin{bmatrix} \Sigma_v & \Sigma_{v\theta} \\ \Sigma_{\theta v} & \Sigma_\theta \end{bmatrix}$ can be derived for the LC-PF (2).

D. Fully Excited Case

If $\mathcal{U}$ is empty, Σ_p and Σ_θ are invertible, and

$$\Sigma_\theta^{-1} = H_\beta^T \Sigma_p^{-1} H_\beta, \text{ where using (3), we verify that}$$

$$\Sigma_\theta^{-1}(i, j) \begin{cases} < 0 \text{ if } (ij) \in \mathcal{E}, \text{ and } \forall k, (ik), (jk) \notin \mathcal{E} \\ > 0 \text{ if } (ij) \notin \mathcal{E}, \text{ and } \exists k, (ik), (jk) \in \mathcal{E} \\ = 0 \text{ if } (ij) \notin \mathcal{E}, \text{ and } \forall k, (ik), (jk) \notin \mathcal{E}. \end{cases} \quad (6)$$

Note that if i, j are not part of a loop of size three, then they must exist in one of the configurations in (6). The following thus holds.

Lemma 1: [16], [20] Let nodes i, j in fully excited $\mathcal{G}$ not be part of a three-node loop. Under Assumption 1, edge (ij) exists if and only if $\Sigma_\theta^{-1}(ij) < 0$.

Note: In the LC-PF model [20], under faithfulness assumptions, it can be shown that if i, j are not in a three-node loop, (ij) exists iff $\Sigma_{(v,\theta)}^{-1}(i, j) + \Sigma_{(v,\theta)}^{-1}(i + N, j + N) < 0$.

If the minimum loop size for grid $\mathcal{G}$ is four, no node pair exists in a three-node loop. Thus, Lemma 1 enables *correct topology estimation in the fully excited setting* [16], [20], under both dc-PF and LC-PF. Notably, Lemma 1 fails in the underexcited grid as Σ_θ or $\Sigma_{(v,\theta)}$ are not invertible. The rest of the article discusses tractable learning in this underexcited setting. To learn grids with zero-injection node set $\mathcal{U}$, we propose a two-step topology learning algorithm. First, we identify zero injection nodes and their neighbors. Next, we estimate edges between node pairs with nonzero injection. The next section presents a regression-based framework for the first step.

III. IDENTIFYING ZERO-INJECTION NODES AND NEIGHBORS

We first start with stating the required assumption for structure recovery in an underexcited grid.

A. Structural Assumptions

Terminal nodes in power grids have loads or generators and typically have nonzero injection. For theoretical completion, consider a zero-injection terminal node i with unique neighbor j in $\mathcal{G}$. From (4), $\theta_i = \theta_j$. Thus, any voltage-based learning algorithm can estimate $\mathcal{G}$ only up to a permutation between i, j , as i, j can be interchanged without affecting the algorithm's output. However, the pair i, j can be checked using $(\theta_i - \theta_j)^2 = 0$. We thus identify node groups with overlapping phase angles. To enable unique estimation, we retain one node per group as neighbor and remove the rest labeled as terminal zero-injection nodes, before estimating the rest of the network. This, along with another structural assumption required for our regression approach, is stated below.

Assumption 2: In grid $\mathcal{G}$, all zero-injection nodes are internal (degree > 1) and nonadjacent.

The nonadjacency assumption prevents cases where a large fraction/majority of internal nodes has no injection. We subsequently show through counterexamples, that in the absence of Assumption 2, both zero-injection buses and their neighbors may be incorrectly identified. We first consider the setting with noiseless voltage measurements.

B. Noiseless Setting

To begin our analysis, we partition H_β (J_β) into rows (columns) corresponding to $\mathcal{U}$ and $\mathcal{W}$: $H_\beta = \begin{bmatrix} H_\beta^{\mathcal{U}} \\ H_\beta^{\mathcal{W}} \end{bmatrix}$, $J_\beta = \begin{bmatrix} J_\beta^{\mathcal{U}} \\ J_\beta^{\mathcal{W}} \end{bmatrix}$. As $H_\beta J_\beta = \mathbb{I}$, $H_\beta^{\mathcal{U}} J_\beta^{\mathcal{W}} = \mathbf{0}$. Thus

$$y^T J_\beta^{\mathcal{W}} = 0 \iff y^T = c^T H_\beta^{\mathcal{U}} \text{ for some } c. \quad (7)$$

Equation (7) follows from the fact that both rows-space of $H_\beta^{\mathcal{U}}$ and left null-space of $J_\beta^{\mathcal{W}}$ have rank $|\mathcal{U}|$. Following (3) and Assumption 2, the matrix $H_\beta^{\mathcal{U}}$ satisfies the following properties:

$$H_\beta^{\mathcal{U}}(i, i) > 0 \quad \forall i \in \mathcal{U} \quad (8a)$$

$$H_\beta^{\mathcal{U}}(i, j) = 0 \quad \forall i \in \mathcal{U} \quad \forall j \in \mathcal{U} - \{i\} \quad (8b)$$

$$H_\beta(i, k) \leq 0 \quad \forall i \in \mathcal{U} \quad \forall k \in \mathcal{W} \quad (8c)$$

$$\forall i \in \mathcal{U}, \exists \{k_1^i, k_2^i\} \in \mathcal{W} \text{ such that}$$

$$H_\beta(i, k) = -\beta_{ik} < 0, \quad k \in \{k_1^i, k_2^i\}. \quad (8d)$$

Let θ_{-i} represent the vector of phase angles at all nodes but i . Consider the following regression problem, solved for each $i \in \mathcal{V}$.

Nodal regression

$$a^{i*} = \arg[lr] \min_{x \in \mathbb{R}^{N-1}} \mathbb{E}[(\theta_i - \theta_{-i}^T x)^2] \text{ s.t. } x \geq \mathbf{0}, \mathbf{1}^T x \leq 1. \quad (9)$$

The next result shows that (9) can identify nodes in $\mathcal{U}$.

Theorem 1: In grid $\mathcal{G}$, under Assumption 2, i is a zero-injection bus in $\mathcal{U}$ if and only if the minimum value for Problem (9) is zero.

Once the nodes in $\mathcal{U}$ have been identified, we use the following regression to identify the *neighbors* of each node $i \in \mathcal{U}$.

Constrained nodal regression:

$$b^{i*} = \arg[lr] \min_{x \in \mathbb{R}^W} \mathbb{E}[(\theta_i - \theta^W x)^2] \text{ s.t. } x \geq \mathbf{0}, \mathbf{1}^T x \leq 1. \quad (10)$$

Theorem 2: For $i \in \mathcal{U}$, consider Problem (10). Under Assumption 2, b^{i*} has cost 0, and neighbors of i are given by $\{j : j \in \mathcal{W}, b_{j-|U|}^{i*} \neq 0\}$.

The regression problem (10) differs from (9) in that the regression vector x is now constrained to be zero for nodes in $\mathcal{U}$ that have been identified using Theorem 1. This constraint leverages the structural property in Assumption 2. We now prove both of these results.

Proof of Theorem 1: Consider $y \in \mathbb{R}^N$ with $y_i = 1, y_{-i} = -x$. The constraints and cost in Problem (9) are reformulated as

$$(\text{Constraints}) y_i = 1, y_{-i} \leq 0, \mathbf{1}^T y \geq 0 \quad (11)$$

$$(\text{Cost}) \mathbb{E}(\theta y)^2 = y^T \Sigma_\theta y = y^T J_\beta^W \Sigma_{p^W} J_\beta^W y.$$

Note that, [using (7)]

$$\mathbb{E}(\theta y)^2 = 0 \iff y^T = c^T H_\beta^U \text{ for some } c. \quad (12)$$

For $i \in \mathcal{U}$, let $y^T = \frac{H_\beta^U(i, :)}{H_\beta(i, i)}$. From (12), $\mathbb{E}(\theta y)^2 = 0$. Using (3), $y_i = 1$, while $y_j \leq 0 \quad \forall j \neq i$ and $\mathbf{1}^T y = 0$. Thus, a feasible y with cost 0 exists for $i \in \mathcal{U}$.

For $i \in \mathcal{W}$, suppose $y^T = c^T H_\beta^U$ with $y_i = 1, y_{-i} \leq 0$. Note that for all $j \in \mathcal{U}, j \neq i$ and using (8), $y_j = c_j H_\beta(j, j)$. $y_j \leq 0$ thus enforces $c_j \leq 0$ for all $j \in \mathcal{U}$. Similarly, $y_i = 1$ implies $\exists j_1 \in \mathcal{U}$ such that $c_{j_1} \neq 0$ and is strictly negative. Under Assumption 2, $\exists k \in \mathcal{W} - \{i\}$ such that edge $(j_1 k) \in \mathcal{E}$. Then, $y_k = -c_{j_1} \beta_{j_1 k} + \sum_{j \in \mathcal{U} - \{j_1\}} c_j H_\beta(j, k)$. As $H_\beta(j, k) \leq 0$, we have $y_k > 0$, which violates $y_{-i} \leq 0$. Thus, no $y = c^T H_\beta^U$ is feasible with (11) and the cost in Problem (9) is nonzero for $i \in \mathcal{W}$. ■

Proof of Theorem 2: For $i \in \mathcal{U}$, all of its neighbors belong to $\mathcal{W}$, under Assumption 2. Consider the following reformulation

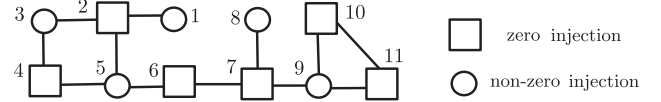


Fig. 2. Grid that violates Assumption 2 due to nodes 6, 7 and 10, 11.

of the cost and constraints of Problem (10):

$$(\text{Cost:}) \mathbb{E}(\theta y)^2 = y^T J_\beta^W \Sigma_p^W J_\beta^W y \text{ with constraints} \quad (13)$$

$$y_i = 1, y_j = 0 \quad \forall j \in \mathcal{U} - \{i\}, y_{-i} \leq 0, \mathbf{1}^T y \geq 0. \quad (14)$$

Suppose y is a solution with cost zero. Using (12), there exists c such that $y^T = c^T H_\beta^U$. As nodes in $\mathcal{U}$ are nonadjacent, $y_j = c_j H_\beta(j, j)$ for $j \in \mathcal{U}$. To ensure $y_i = 1$ and $y_j = 0$, for all $j \in \mathcal{U} - \{i\}$, we have $c_i = 1/H_\beta(j, j)$ and $c_j = 0$ for all $j \neq i$. Thus, $y^T = \frac{H_\beta^U(i, :)}{H_\beta(i, i)}$ is the unique solution. From (8), it follows that y satisfies constraints in (14). Note that $j \neq i$, $y_j = H_\beta(i, j)/H_\beta(i, i)$ and, thus, is nonzero if and only if edge $(ij) \in \mathcal{E}$. Thus, it identifies the neighbors of i . ■

Note:

a) The constraints in Problem (9) stem from the power-flow conservation law. Using a similar analysis for (v, θ) in the LC-PF (2), it can be shown that nodes in $\mathcal{U}$ are identified by zero-cost solutions of the following complex-valued regression problem:

$$\begin{aligned} & \arg[lr] \min_{x \in \mathbb{C}^{N-1}} \mathbb{E}[|(v_i - \hat{i}\theta_i - (v_{-i} - \hat{i}\theta_{-i})^T x)|^2] \\ & \text{s.t. } \text{Real}(x) \geq \mathbf{0}, \mathbf{1}^T \text{Real}(x) \leq 1 \\ & \quad -1 \leq \text{Imag}(x) \leq 1, \mathbf{1}^T \text{Imag}(x) = 0. \end{aligned} \quad (15)$$

b) Following (15), the LC-PF version of Theorem 2 can be constructed as a complex-valued regression problem over (v, θ) , with regression coefficients restricted to $\mathcal{W}$.

Note that Assumption 1 is not used in either identification of nodes in $\mathcal{U}$ (Theorem 1), or estimation of their neighbors (Theorem 2). We next show that nonadjacency of nodes in $\mathcal{U}$ (Assumption 2) is necessary for both steps.

Requirement of Assumption 2: Consider the grid graph in Fig. 2, with adjacent node pairs 6, 7 and 10, 11.

Consider Problem (9) for node $9 \in \mathcal{W}$. By combining rows $H_\beta^T(10, :), H_\beta^T(11, :)$ of adjacent zero-injection nodes, we can create a feasible zero-cost solution y^T for $i = 9$, which will incorrectly identify 9 as a zero-injection node. For example, if $\min(\beta_{9,10}, \beta_{9,11}) > \beta_{9,11}$, then $y^T = -\frac{H_\beta^T(10, :)}{2\beta_{9,10}} -$

$\frac{H_\beta^T(11, :)}{2\beta_{9,11}}$ can be verified as a valid zero-cost solution for $i = 9$.

Now consider node $6 \in \mathcal{U}$. It can be verified that zero-cost solutions in Problem (9) for $i = 6$ are given by

$$y^T = \frac{H_\beta(7, 7)H_\beta^T(6, :) + \epsilon\beta_{67}H_\beta^T(7, :)}{H_\beta(6, 6)H_\beta(7, 7) - \epsilon\beta_{67}^2} \text{ with } 0 \leq \epsilon \leq 1.$$

Taking $\epsilon = 1$ ensures $y_7^T = 0$, and produces a zero-cost solution for Problem (10), that incorrectly estimates 5,8,9 as neighbors of 6. Assumption 2, is thus, necessary for correctness of Theorems 1 and 2. Next, we analyze the asymptotic correctness of Theorems 1 and 2 in the practical setting, where measurements are noisy.

C. Noisy Setting

Let $\tilde{\theta}$ be the phase measurement corrupted by independent noise n of mean zero and covariance $\Sigma_n \succ 0$, i.e.,

$$\tilde{\theta} = \theta + n \Rightarrow \Sigma_{\tilde{\theta}} = \Sigma_{\theta} + \Sigma_n. \quad (16)$$

First, we define the following parameters:

$$\begin{aligned} \beta_{\min} &= [lr] \min_{(ij) \in \mathcal{E}} \beta_{ij}, S_{id} = [lr] \max_{i \in \mathcal{V}} \frac{H_{\beta}(i, i)}{\beta_{\min}} \\ \text{SNR} &= \frac{\sigma_{\min}(\Sigma_{\theta^W})}{\sigma_{\max}(\Sigma_n)}. \end{aligned} \quad (17)$$

Here, $\sigma_{\min}(X)$ and $\sigma_{\max}(X)$ denote the minimum and maximum singular values of matrix X , respectively. Σ_{θ^W} is the covariance of θ^W .

Let the minimum for Problem (9) for node i under noise be achieved at a_n^{i*} . Let $y_n \in \mathbb{R}^N$, with $y_{ni} = 1, y_{n-i} = -a_n^{i*}$. Similarly, denote $y \in \mathbb{R}^N$, with $y_i = 1, y_{-i} = -a^{i*}$, where a^{i*} is the noiseless optima. For $i \in \mathcal{U}$, optimality of y_n under noise, gives

$$\begin{aligned} y_n^T (\Sigma_{\theta} + \Sigma_n) y_n &\leq y^T (\Sigma_{\theta} + \Sigma_n) y \stackrel{(a)}{=} y^T \Sigma_n y \\ &\leq (1 + \|a^{i*}\|_2^2) \sigma_{\max}(\Sigma_n) \stackrel{(b)}{\leq} 2 \sigma_{\max}(\Sigma_n). \end{aligned} \quad (18)$$

Here, (a) follows from Theorem 1 for $i \in \mathcal{U}$. (b) follows from $\|a^{i*}\|_2 \leq \|a^{i*}\|_1 \leq 1$. In contrast, when $i \in \mathcal{W}$, the following result holds.

Lemma 2: For $i \in \mathcal{W}$, consider $y_n \in \mathbb{R}^N$ where $y_{ni} = 1$, and $y_{n-i} = -a_n^{i*}$ is the optimal solution for Problem (9) under noise. Then, $y_n^T (\Sigma_{\theta} + \Sigma_n) y_n > \frac{\sigma_{\min}(\Sigma_{\theta^W})}{1 + S_{id}^4 + S_{id}^2}$, where S_{id} is given in (17).

The proof is provided in Appendix 1. The following result combines Lemma 2 with (18) to ensure correct identification of $\mathcal{U}$ under noise.

Theorem 3: If $\text{SNR} \geq 2(1 + S_{id}^4 + S_{id}^2)$, then only nodes in $\mathcal{U}$ have cost lower than $\frac{\sigma_{\min}(\Sigma_{\theta^W})}{1 + S_{id}^4 + S_{id}^2}$ in Problem (9), where S_{id} , SNR are defined in (17).

We now analyze Theorem 2 for noisy estimation of the neighbors of zero-injection nodes. For node $i \in \mathcal{U}$, let the optimal for Problem (10) be achieved at b_n^{i*} under noise, and at b^{i*} in the noiseless setting. Let $\theta^M = \begin{bmatrix} \theta^i \\ \theta^W \end{bmatrix}$, $n^M = \begin{bmatrix} n^i \\ n^W \end{bmatrix}$, where $\mathcal{M} = \{i\} \cup \mathcal{W}$. By optimality of b_n^{i*} in the noisy setting, we have

$$\begin{bmatrix} 1 \\ -b_n^{i*} \end{bmatrix}^T (\Sigma_{\theta^M} + \Sigma_{n^M}) \begin{bmatrix} 1 \\ -b_n^{i*} \end{bmatrix} \leq \begin{bmatrix} 1 \\ -b^{i*} \end{bmatrix}^T (\Sigma_{\theta^M} + \Sigma_{n^M}) \begin{bmatrix} 1 \\ -b^{i*} \end{bmatrix}$$

$$\begin{aligned} &\Rightarrow \begin{bmatrix} 0 \\ b_n^{i*} - b^{i*} \end{bmatrix}^T (\Sigma_{\theta^M} + \Sigma_{n^M}) \begin{bmatrix} 0 \\ b_n^{i*} - b^{i*} \end{bmatrix} \\ &\leq 2 \begin{bmatrix} 0 \\ b_n^{i*} - b^{i*} \end{bmatrix}^T (\Sigma_{\theta^M} + \Sigma_{n^M}) \begin{bmatrix} 1 \\ -b^{i*} \end{bmatrix} \\ &\stackrel{(a)}{=} 2 \begin{bmatrix} 0 \\ b_n^{i*} - b^{i*} \end{bmatrix}^T \Sigma_{n^M} \begin{bmatrix} 1 \\ -b^{i*} \end{bmatrix} \stackrel{(b)}{\leq} 4\sqrt{2} \sigma_{\max}(\Sigma_n). \end{aligned} \quad (19)$$

Here, (a) follows from optimal cost of 0 for b^{i*} in the noiseless setting (see Theorem 2). (b) follows from $\|b_n^{i*} - b^{i*}\|_2 \leq \|b_n^{i*} - b^{i*}\|_1 \leq \|b_n^{i*}\|_1 + \|b^{i*}\|_1 \leq 2$, and $\sigma_{\max}(\Sigma_{n^M}) \leq \sigma_{\max}(\Sigma_n)$. Further, as $\Sigma_n \succ 0$, we have

$$\begin{aligned} &\begin{bmatrix} 0 \\ b_n^{i*} - b^{i*} \end{bmatrix}^T (\Sigma_{\theta^M} + \Sigma_{n^M}) \begin{bmatrix} 0 \\ b_n^{i*} - b^{i*} \end{bmatrix} > \\ &\quad (b_n^{i*} - b^{i*})^T \Sigma_{\theta^W} (b_n^{i*} - b^{i*}) \\ &\Rightarrow 4\sqrt{2} \sigma_{\max}(\Sigma_n) > \sigma_{\min}(\Sigma_{\theta^W}) \|b_n^{i*} - b^{i*}\|_{\infty}^2. \end{aligned} \quad (20)$$

The last inequality follows from (19) and the full-rank of Σ_{θ^W} . Using this, the next result ensures consistent estimation of neighbors of $\mathcal{U}$.

Theorem 4: For node $i \in \mathcal{U}$, let b_n^{i*} be the optimal solution in Problem (10) under noise. If $\text{SNR} \geq 16\sqrt{2} S_{id}^2$, neighbors of i are given by $\{j : j \in \mathcal{W}, b_n^{i*}{}_{j-|U|} \geq .5/S_{id}\}$, where S_{id} , SNR are defined in (17).

Proof: Consider $j \in \mathcal{W}$. By Theorem 2, noiseless solution b^{i*} has $b_{j-|U|}^{i*} = \beta_{ij}/H_{\beta}(i, i) \geq 1/S_{id}$ if $(ij) \in \mathcal{E}$, and 0 otherwise. If $\text{SNR} \geq 16\sqrt{2} S_{id}^2$, using (20), $\|b_n^{i*} - b^{i*}\|_{\infty} < .5/S_{id}$. Thus, entries in b_n^{i*} for neighbors of i are greater than $.5/S_{id}$, and less than $.5/S_{id}$ otherwise. ■

It is worth mentioning that, estimation of nodes in $\mathcal{U}$ and identification of their neighbors are done separately in Theorems 1 and 2. To complete this section, we discuss why this is necessary and whether both these steps can be combined.

D. Joint Identification and Neighborhood Estimation for $\mathcal{U}$

Theorem 1 identifies zero-injection nodes in $\mathcal{U}$ by zero-cost solution for Problem (9). However, even when Assumption 2 holds, the solution to (9) need not be unique. For example, consider node $2 \in \mathcal{U}$ in the grid in Fig. 2. Combining $H_{\beta}^T(2, :)$, $H_{\beta}^T(4, :)$, a family of zero-cost solutions $y = \frac{H_{\beta}^T(2, :)}{H_{\beta}(2, 2)} - \epsilon \frac{H_{\beta}^T(4, :)}{H_{\beta}(4, 4)}$ is constructed for $i = 2$, where $0 \leq \epsilon < \frac{H_{\beta}(4, 4)}{H_{\beta}(2, 2)} \min\left(\frac{\beta_{23}}{\beta_{43}}, \frac{\beta_{25}}{\beta_{45}}\right)$ ensures $y_{-2} \leq 0$. Hence, the regression coefficients of Theorem 1 may reflect incorrectly on the neighbor set. Observe that the correct neighbors of 2 are identified only if $\epsilon = 0$. Indeed, $\epsilon = 0$ is indirectly enforced in Problem (10) [unlike Problem (9)] by restricting the regression coefficients to nodes in $\mathcal{W}$.

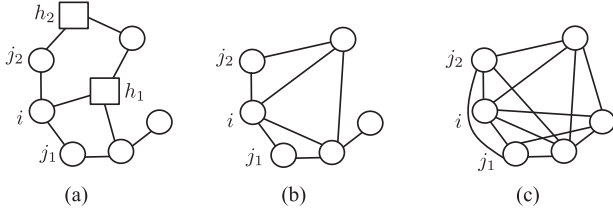


Fig. 3. (a) $\mathcal{G}$ with nodes in $\mathcal{U}$ marked square, h_1, h_2 violate Assumption 3. (b) $\hat{\mathcal{G}}$ after Kron-reducing $\mathcal{U}$. (c) Nonzero entries in inverse voltage covariance in $\mathcal{W}$.

If additional assumptions are allowed, joint identification of zero-injection nodes and their neighborhood estimation in Problem (9) is possible. The next result demonstrates this under a sufficient topological assumption, that holds trivially for radial grids.

Theorem 5: In grid $\mathcal{G}$ with Assumption 2, $\forall i \in \mathcal{U}$, assume that no $j \in \mathcal{U} - \{i\}$ exists such that all neighbors of j are also neighbors of i . Then, true neighbors of $i \in \mathcal{U}$ are given by the nonzero entries in the solution of Problem (9).

The proof is given in Appendix 2. Note that nodes 2,4 in Fig. 2 violate the condition in Theorem 5, hence estimation of 2s neighbors needs Problem (10). The consistency of Theorem 5 under noisy measurements can be shown using a similar analysis as Theorem 4. In the next section, we complete the learning process by estimating the remaining grid edges between nonzero injection nodes.

IV. LEARNING EDGES BETWEEN NONZERO INJECTION NODES

In this section, we construct a method to reconstruct the remaining graph, i.e., identifying the edges between nonzero injection nodes.

A. Structural Assumptions

First, we partition H_β, J_β into blocks for zero and nonzero nodes, $H_\beta = \begin{bmatrix} H_\beta^{\mathcal{U}\mathcal{U}} & H_\beta^{\mathcal{U}\mathcal{W}} \\ H_\beta^{\mathcal{W}\mathcal{U}} & H_\beta^{\mathcal{W}\mathcal{W}} \end{bmatrix}$, $J_\beta = \begin{bmatrix} J_\beta^{\mathcal{U}\mathcal{U}} & J_\beta^{\mathcal{U}\mathcal{W}} \\ J_\beta^{\mathcal{W}\mathcal{U}} & J_\beta^{\mathcal{W}\mathcal{W}} \end{bmatrix}$. Then

$$\Sigma_{\theta^{\mathcal{W}}} = J_\beta^{\mathcal{W}\mathcal{W}} \Sigma_{p^{\mathcal{W}}} J_\beta^{\mathcal{W}\mathcal{W}} \quad [\text{by (5), and } p^{\mathcal{W}} = \mathbf{0}] \quad (21)$$

$$\text{where } J_\beta^{\mathcal{W}\mathcal{W}} = (H_\beta^{\mathcal{W}\mathcal{W}} - H_\beta^{\mathcal{W}\mathcal{U}} H_\beta^{\mathcal{U}\mathcal{U}^{-1}} H_\beta^{\mathcal{U}\mathcal{W}})^{-1}. \quad (22)$$

Note that $H_\beta^{\mathcal{U}\mathcal{U}}$ and its inverse are diagonal matrices as nodes in $\mathcal{U}$ are not adjacent. Here, $J_\beta^{\mathcal{W}\mathcal{W}}$ represents the inverse weighted Laplacian for a Kron-reduced $\hat{\mathcal{G}}$ obtained by removing nodes in $\mathcal{U}$ from $\mathcal{G}$ [25]. See Fig. 3(b) for an illustrative example. To identify edges between nodes in $\mathcal{W}$, we impose the following topological restriction for zero-injection nodes.

Assumption 3: The minimum size of a loop in $\mathcal{G}$ is 4. Further, any loop of size 4 does not include nodes in $\mathcal{U}$, while any loop of size 5 includes at most one node in $\mathcal{U}$.

Note that Assumption 3 holds trivially for radial grids, that include a majority of distribution grids. The minimum loop size of 4 is nonrestrictive in real grids with large girth [12], [13]. In fact, it is necessary for consistent estimation even in the

fully excited setting [20] (see Lemma 1). Further, as nodes in $\mathcal{U}$ have degree at least two, neighbors of any $j \in \mathcal{U}$ cannot all be neighbors of another node $i \in \mathcal{U}$, as that would create a loop of size 4. Assumption 3, thus, implies Theorem 5 and enables joint identification of nodes in $\mathcal{U}$ and estimation of their neighbors.

B. Noiseless Setting

Consider nodes $i, j \in \mathcal{W}$. If $\exists k \in \mathcal{U}$ with edges $(ik), (jk)$ (estimated by Theorem 2), then $(ij) \notin \mathcal{E}$ as that will create a loop of size 3. To complete the topology learning, we, thus, need to estimate only edges between $i, j \in \mathcal{W}$ that do not have a common neighbor in $\mathcal{U}$. The following result identifies such edges, in the noiseless setting.

Theorem 6: Consider nodes $i, j \in \mathcal{W}$ without any common neighbor in $\mathcal{U}$. Under Assumption 1,3, $(ij) \in \mathcal{E}$ if and only if $\Sigma_{\theta^{\mathcal{W}}}^{-1}(i - |\mathcal{U}|, j - |\mathcal{U}|) < 0$.

Before proving the theorem, we state the following lemma.

Lemma 3: Let $i, j \in \mathcal{W}$ without a common neighbor in $\mathcal{U}$. Then $J_\beta^{\mathcal{W}\mathcal{W}^{-1}}(i - |\mathcal{U}|, j - |\mathcal{U}|) = H_\beta(i, j)$.

Proof: Consider $i, j \in \mathcal{W}$ without common neighbor in $\mathcal{U}$. The $(i - |\mathcal{U}|, j - |\mathcal{U}|)^{\text{th}}$ entry in $J_\beta^{\mathcal{W}\mathcal{W}^{-1}}$ corresponds to (i, j) in Kron-reduced $\hat{\mathcal{G}}$. As $H_\beta^{\mathcal{U}\mathcal{U}}$ is diagonal under Assumption 2 and that i and j do not have a common neighbor in $\mathcal{U}$, it follows from (22) that $J_\beta^{\mathcal{W}\mathcal{W}^{-1}}(i - |\mathcal{U}|, j - |\mathcal{U}|) = H_\beta^{\mathcal{W}\mathcal{W}}(i - |\mathcal{U}|, j - |\mathcal{U}|) = H_\beta(i, j)$. ■

Proof of Theorem 6: Using Lemma 3, (ij) is an edge in $\hat{\mathcal{G}}$, if and only if $(i, j) \in \mathcal{E}$. We now prove that i, j are not part of a loop of size 3 in $\hat{\mathcal{G}}$. The result then holds by applying Lemma 1 in $\hat{\mathcal{G}}$.

If $(i, j) \notin \mathcal{E}$, then edge (ij) does not exist in $\hat{\mathcal{G}}$ and i, j are not part of a three node loop. Next let $(i, j) \in \mathcal{E}$. As minimum loop size in $\mathcal{G}$ is four, a three node loop with i, j and some k exists in $\hat{\mathcal{G}}$ if (A) $\{(ik), (jh), (kh)\} \subseteq \mathcal{E}$ or $\{(ih), (kh), (jk)\} \subseteq \mathcal{E}$, $\exists h \in \mathcal{U}$, or (B) $\{(ih_i), (kh_i), (jh_j), (kh_j)\} \subseteq \mathcal{E}$, $\exists h_i, h_j \in \mathcal{U}$ [see Fig. 3(a)]. (A) implies h is in a four node loop, while (B) implies h_i, h_j are in a five node loop. (A), (B), thus, contradict Assumption 3. Hence, i, j are not part of a three node loop in $\hat{\mathcal{G}}$. Using Lemma 1 on $\hat{\mathcal{G}}$, $\Sigma_{\theta^{\mathcal{W}}}^{-1}(i - |\mathcal{U}|, j - |\mathcal{U}|) < 0$ for edge (ij) , and ≥ 0 otherwise. ■

Note: (a) For the LC-PF (2), as per the remark following Lemma 1, an edge between $(i + |\mathcal{U}|), (j + |\mathcal{U}|) \in \mathcal{W}$ without common neighbor in $\mathcal{U}$ is present iff $\Sigma_{(v^{\mathcal{W}}, \theta^{\mathcal{W}})}^{-1}(i + |\mathcal{W}|, j + |\mathcal{W}|) < 0$.

(b) In loopy $\mathcal{G}$, violation of Assumption 3 can create three node loops in the Kron-reduced $\hat{\mathcal{G}}$ [see Fig. 3(b)]. Lemma 1 may fail to distinguish true edges in that setting. We now discuss Theorem 6 in the presence of noisy measurements.

C. Noisy Setting

Under the noise model in (16), consider $\tilde{\theta}^{\mathcal{W}} = \theta^{\mathcal{W}} + n^{\mathcal{W}}$ with noisy inverse covariance matrix $\Sigma_{\tilde{\theta}^{\mathcal{W}}}^{-1}$. Using Woodbury formula [20], [26], we have

$$\Delta \Sigma_{\tilde{\theta}^{\mathcal{W}}}^{-1} = \Sigma_{\theta^{\mathcal{W}}}^{-1} - \Sigma_{\tilde{\theta}^{\mathcal{W}}}^{-1} = \Sigma_{\theta^{\mathcal{W}}}^{-1} (\Sigma_{n^{\mathcal{W}}}^{-1} + \Sigma_{\theta^{\mathcal{W}}}^{-1})^{-1} \Sigma_{\theta^{\mathcal{W}}}^{-1}. \quad (23)$$

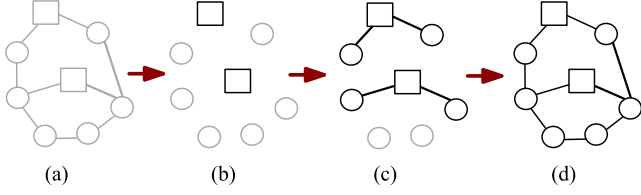


Fig. 4. Steps in Algorithm 1 for a test grid.

Using positive-definiteness of $\Delta\Sigma_{\theta\mathcal{W}}^{-1}$ in (23), we have

$$\begin{aligned} \max_{i,j} |\Delta\Sigma_{\theta\mathcal{W}}^{-1}(i,j)| &\leq \max_i \Delta\Sigma_{\theta\mathcal{W}}^{-1}(i,i) \leq \sigma_{\max}(\Delta\Sigma_{\theta\mathcal{W}}^{-1}) \\ &\leq \frac{\sigma_{\max}^2(\Sigma_{\theta\mathcal{W}}^{-1})}{\sigma_{\min}(\Sigma_{\theta\mathcal{W}}^{-1} + \Sigma_{\theta\mathcal{W}}^{-1})} < \sigma_{\max}^2(\Sigma_{\theta\mathcal{W}}^{-1})\sigma_{\max}(\Sigma_{\theta\mathcal{W}}^{-1}) \\ &\leq \sigma_{\max}(\Sigma_n)/\sigma_{\min}^2(\Sigma_{\theta\mathcal{W}}). \end{aligned} \quad (24)$$

We use this to modify Theorem 6 under noise.

Theorem 7: Consider $i, j \in \mathcal{W}$ with no common neighbor in $\mathcal{U}$. If $\text{SNR} \geq \frac{\max_i \Sigma_p(i,i)}{\beta_{\min}^2 \sigma_{\min}(\Sigma_{\theta\mathcal{W}})}$, edge (ij) exists if and only if $\Sigma_{\theta\mathcal{W}}^{-1}(i - |\mathcal{U}|, j - |\mathcal{U}|) \leq -\beta_{\min}^2 / \max_i \Sigma_p(i,i)$, where $\beta_{\min}$, SNR are given in (17).

Proof: Consider $i, j \in \mathcal{W}$ without a common neighbor in $\mathcal{U}$. For $(ij) \notin \mathcal{E}$, by Theorem 6, $\Sigma_{\theta\mathcal{W}}^{-1}(i - |\mathcal{U}|, j - |\mathcal{U}|) \geq 0$. If SNR satisfies the stated bound, using (24), $\Sigma_{\theta\mathcal{W}}^{-1}(i - |\mathcal{U}|, j - |\mathcal{U}|) > -\frac{1}{\text{SNR}\sigma_{\min}(\Sigma_{\theta\mathcal{W}})} \geq -\beta_{\min}^2 / \max_i \Sigma_p(i,i)$.

For $(ij) \in \mathcal{E}$, using Lemma 3 and the positive-definiteness of $J_{\beta}^{\mathcal{W}\mathcal{W}^{-1}}$, we get $J_{\beta}^{\mathcal{W}\mathcal{W}^{-1}}(i - |\mathcal{U}|, i - |\mathcal{U}|) \geq \beta_{ij}$ and $J_{\beta}^{\mathcal{W}\mathcal{W}^{-1}}(j - |\mathcal{U}|, j - |\mathcal{U}|) \geq \beta_{ij}$. Using (22), we get

$$\begin{aligned} \Sigma_{\theta\mathcal{W}}^{-1}(i - |\mathcal{U}|, j - |\mathcal{U}|) &\leq -\beta_{ij}^2 (\Sigma_p^{-1}(i,i) + \Sigma_p^{-1}(j,j)) \\ &\leq -2\beta_{\min}^2 / \max_i \Sigma_p(i,i). \end{aligned}$$

Using (24) and the stated SNR bound, $\Sigma_{\theta\mathcal{W}}^{-1}(i - |\mathcal{U}|, j - |\mathcal{U}|) \leq -\beta_{\min}^2 / \max_i \Sigma_p(i,i)$. ■

Note: Using (21), $\sigma_{\max}(\Sigma_{\theta\mathcal{W}}^{-1}) \leq \frac{\sigma_{\max}(\Sigma_{\mathcal{P}\mathcal{W}}^{-1})}{\sigma_{\min}^2(J_{\beta}^{\mathcal{W}\mathcal{W}})} \leq \frac{\sigma_{\max}(\Sigma_{\mathcal{P}\mathcal{W}}^{-1})}{\sigma_{\min}^2(J_{\beta})} = \frac{\sigma_{\max}^2(H_{\beta})}{\min_{i \in \mathcal{W}} \Sigma_p(i,i)}$. Using this, Theorem 7 holds when $\text{SNR} \geq \frac{\max_i \Sigma_p(i,i)}{\min_{i \in \mathcal{W}} \Sigma_p(i,i)} \left(\frac{\sigma_{\max}(H_{\beta})}{\beta_{\min}} \right)^2$.

Learning algorithm: The overall steps of our voltage-based learning algorithm are listed in Algorithm 1. Theorem 1 is used in Steps 2–7 to identify zero-injection nodes. Then their neighborhood nodes are identified in Steps 8–12 using Theorem 2. Finally, edges between nonzero injection nodes are determined in Steps 14–18 through Theorem 6. An example of the learning steps for a test grid is given in Fig. 4.

Computational complexity: Solving each of the constrained regression problems (9) and (10) for all nodes, needs $O(N^4)$ operations. Inverse covariance estimation takes $O(N^3)$ operations. Edge estimation and creation of $\mathcal{N}_2$ in Steps 10–12 for all nodes in $\mathcal{U}$ is bounded by $O(N^3)$ operations. Adding edges between nodes in $\mathcal{W}$ needs $O(N^3)$ operations. The overall complexity is thus $O(N^4)$. Note that the complexity analysis is for the general

Algorithm 1: Topology Learning for Underexcited Grids.

Input: θ samples for nodes in $\mathcal{V}$, positive thresholds

τ_1, τ_2, τ_3

Output: Edge set $\mathcal{E}$ in grid $\mathcal{G}$

```

1:  $\mathcal{U} \leftarrow \{\}$ .  $\mathcal{N}_2 \leftarrow \{\}$ 
2: for all node  $i$  in  $\mathcal{V}$ 
3:   Solve Problem (9) to get cost  $c_i$ .
4:   if  $c_i \leq \tau_1$  then
5:      $\mathcal{U} \leftarrow \mathcal{U} \cup \{i\}$ 
6:   end if
7: end for
8: node  $i$  in  $\mathcal{U}$ 
9:   Solve Problem (10) to get solution  $b^{i*}$ .
10:   $\mathcal{E} \leftarrow \mathcal{E} \cup \{(ij) : b_{j-|\mathcal{U}|}^{i*} \geq \tau_2\}$ .
11:   $\mathcal{N}_2 \leftarrow \mathcal{N}_2 \cup \{(jk) : b_{j-|\mathcal{U}|}^{i*}, b_{k-|\mathcal{U}|}^{i*} \geq \tau_2\}$ .
12: end for
13: Compute  $\Sigma_{\theta\mathcal{W}}^{-1}$  using  $\theta$  samples for nodes in
     $\mathcal{W} = \mathcal{V} - \mathcal{U}$ .
14: for all  $i \neq j \in \mathcal{W}$ 
15:   if  $\Sigma_{\theta\mathcal{W}}^{-1}(i,j) < -\tau_3$  and  $(ij) \notin \mathcal{N}_2$  then
16:      $\mathcal{E} \leftarrow \mathcal{E} \cup \{(ij)\}$ 
17:   end if
18: end for
```

case without any prior information on existing or permissible edges. In real world implementation, topology estimation and verification may be conducted as and when desired by the estimator. However, once a true topology has been learnt, the estimator may analyze changes in the topology from the existing one and focus on change detection. Such change detection is outside the focus of this article. In a modified setting with three-phase unbalanced systems, the number of variables will be $3N$, hence the computational complexity as a function of the size of the network will stay the same.

Availability of prior information: In realistic grids, system operators have prior information of a set of candidate transmission lines. Algorithm 1 is easily modified in the presence of such information. In particular, while identifying zero-injection nodes and their neighbors, the set of neighbors in Problems 9 and 10 is restricted to only nodes where candidate edges can exist. Similarly, the check for edges between nonzero injection buses in Theorem 6 is restricted to node-pairs with potential edges.

Thresholds: We include three thresholds: τ_1 in Step 4 (identify $\mathcal{U}$ nodes), τ_2 in Step 10 (identify neighbors of $\mathcal{U}$ nodes), and τ_3 in Step 15 (identify edges between $\mathcal{W}$ nodes). This is done to mitigate the effect of finite voltage samples and measurement noise. In the large sample limit, we combine results from Theorems 3, 4, and 7 to determine the following thresholds, for consistent estimation in the presence of noise.

Theorem 8: If $\text{SNR} \geq \max(\frac{\max_i \Sigma_p(i,i)}{\beta_{\min}^2 \sigma_{\min}(\Sigma_{\theta\mathcal{W}})}, 16\sqrt{2}s_{id}^2, 2(1 + s_{id}^4 + s_{id}^2))$, then taking $\tau_1 = \frac{\sigma_{\min}(\Sigma_{\theta})}{1 + s_{id}^4 + s_{id}^2}$, $\tau_2 = \frac{1}{2s_i}$, $\tau_3 = \frac{\beta_{\min}^2}{\max_i \Sigma_p(i,i)}$ in Algorithm 1 gives asymptotically

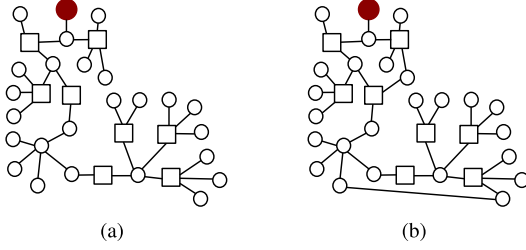


Fig. 5. The 33-bus test networks [27] for simulations. (a) Radial. (b) Loopy. Nonzero injection nodes are marked square.

correct topology recovery, in the presence of noise, where $\beta_{\min}$, SNR , S_{id} are defined in (17).

In the next section, we present simulation results on the performance of our algorithm, in particular on noisy voltage data generated using real-world injection data through a nonlinear power flow model.

V. NUMERICAL SIMULATIONS

We test Algorithm 1 on different test systems, with varying injections (real/simulated) and noise profiles, and power flow models (dc/ac, linear/nonlinear). First, we consider two power grids, constructed from the IEEE 33-bus system [27]: (a) radial [see Fig. 5(a)] with 9 zero-injection nodes, and (b) meshed/loopy [see Fig. 5(b)] with 8 zero-injection nodes, that respects Assumptions 2 and 3. It is worth noting that Algorithm 1 only takes voltage measurements as input. No additional information regarding line parameters, values of injection statistics, or set of feasible lines are considered. In such a setting, the number of possible edges in our grid of 32 nonreference nodes is 496, with 32^{30} possible connected tree realizations (using Cayley's formula) and even more nonradial realizations. It is, however, noted that the performance of the algorithm can be improved if information regarding existence/nonexistence of certain lines is known. For both grids, we generate voltages using linear and nonlinear models of (i) dc power flow and, (ii) ac power flow. While the linear models are given by (2) and (4), the nonlinear PF models are evaluated using Matpower [28]. For generating voltages in all cases, we consider uncorrelated nodal injection fluctuations around base loads, that are sampled using zero-mean Gaussian random variables of standard deviation 10^{-1} . Eventually, we also test on voltage samples generated using real injection data from [29], that may be correlated between nodes. To demonstrate the performance under noise, we corrupt our generated voltage samples in all cases with zero-mean Gaussian noise of variance measured as a fraction of the variance of voltage measurements.

We solve Problems (9) and (10) in Algorithm 1 using CVX [30]. We calibrate the 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 , τ_2 , and τ_3 in Algorithm 1 are tuned to give minimal estimation error for a large sample size of 10^4 . These thresholds are then fixed, and the estimation

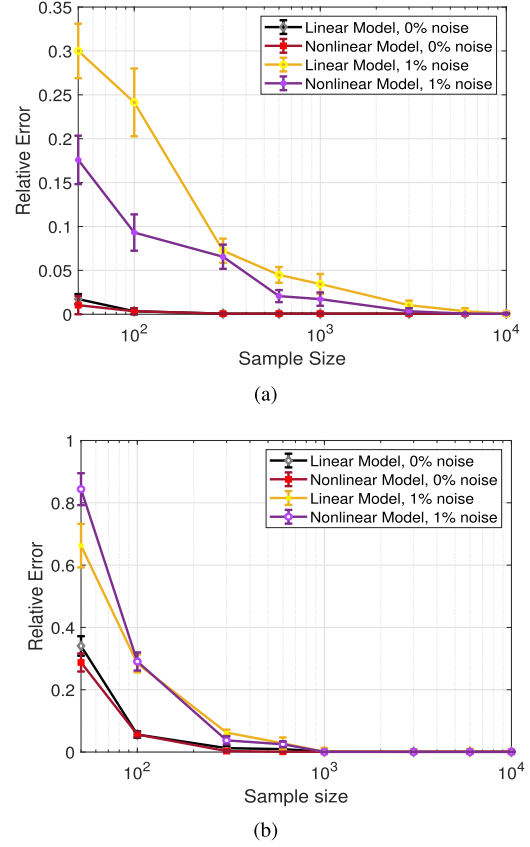
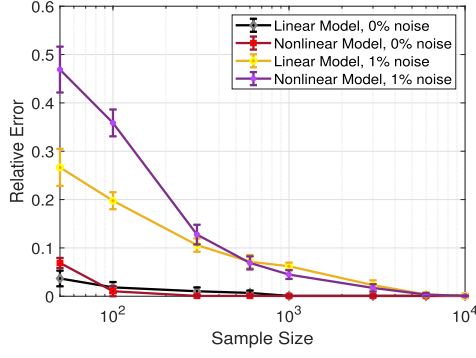


Fig. 6. Performance with linear and nonlinear dc-PF voltages for (a) radial grid in Fig. 5(a), (b) loopy grid in Fig. 5(b).

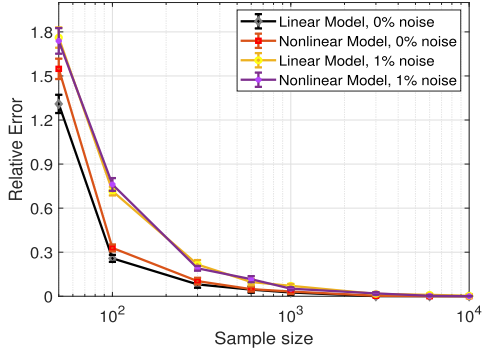
errors at different sample sizes are computed by averaging over 15 independent runs.

Fig. 6(a) shows the performance for the radial grid in Fig. 5(a) for linear and nonlinear dc-PF voltages generated using simulated injection data. With noiseless data samples, the number of samples required for exact topology learning is approximately 300 in either case. Perfect reconstruction under voltages corrupted with relative noise variance of 1% is observed at approximately 10^4 samples. However, at 600 samples, the average error is below 5% for both linear and nonlinear models. While the performance improves with high sample sizes, note that the thresholds used in Algorithm 1 are not updated at each sample size, but fixed to the ones decided for asymptotically perfect recovery. In a realistic setting, the learning algorithm should use thresholds based on the predetermined number of samples to improve the performance. Further, the performance with voltage samples obtained from linear and nonlinear power flows is almost identical at sample sizes above 600. This follows from the acceptable accuracy of dc-PF in modeling its nonlinear counterpart. Fig. 6(b) shows the performance of the algorithm for dc-PF for the loopy grid in Fig. 5(b). The performance, as expected, improves as the number of samples is increased. Beyond 300 samples, the performance is comparable for linear and nonlinear models, under both noisy and noiseless regimes.

Next, we demonstrate the performance of Algorithm 1 for ac-PF voltage samples (v, θ) , generated using the linear LC-PF



(a)



(b)

Fig. 7. Performance with linear and nonlinear ac-PF voltages for (a) radial grid in Fig. 5(a), (b) loopy grid in Fig. 5(b).

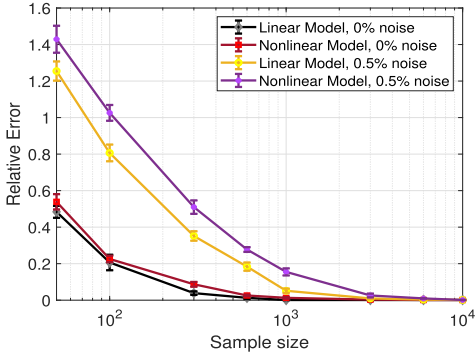


Fig. 8. Performance on loopy grid in Fig. 5(b), with linear and nonlinear dc-PF voltages generated from real injection data.

(2), and the nonlinear Matpower solver. The performance for the radial and loopy networks is given in Fig. 7(a) and (b), respectively. In the noiseless setting, average errors go down to zero for the radial case at 600 samples, and for the loopy case at 1000 samples, under both linear and nonlinear models. With noise, the average errors are below 5% for the radial case at 1000 samples, and below 5% for the loopy case at 1900 samples. Perfect recovery in the noisy setting requires approximately 6000 samples in the radial case and 3000 samples in the loopy grid.

We now demonstrate performance in the absence of Assumption 1 by using real power/active load data, sampled at 15 min intervals from real households [29]. Fig. 8 demonstrates the

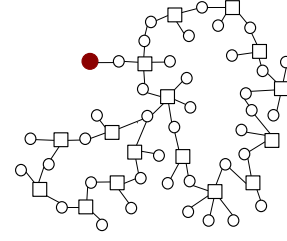
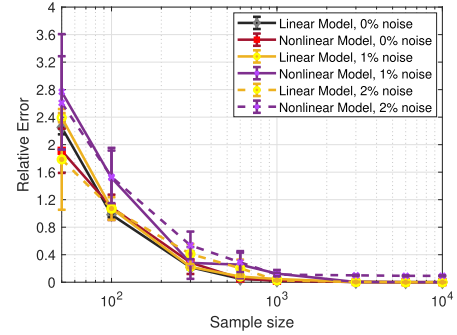
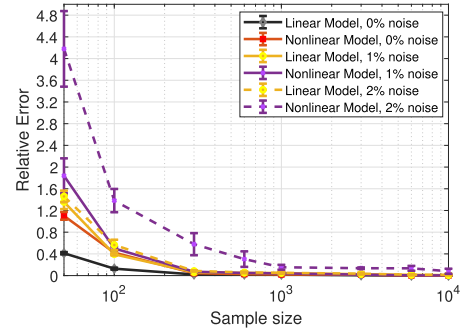


Fig. 9. Loopy 56 bus test network derived from [16]. Nonzero injection nodes are marked square.



(a)



(b)

Fig. 10. Performance for the loopy grid in Fig. 9, with (a) linear and nonlinear dc-PF voltages, and (b) linear and nonlinear ac-PF voltages.

performance of our learning algorithm for the loopy grid in Fig. 5(b), with linear and nonlinear dc-PF samples. Note that the errors, in the noiseless setting, are below 2.67% at 600 samples for either case, despite no prior information about the topology. On the other hand, in the noisy setting with 5% noise variance, the average errors are below 5% at 3000 samples.

Finally, we consider a larger 56-bus loopy test network with 16 zero-injection buses as shown in Fig. 9, that is derived from the radial case in [16]. First, we consider performance for linear and nonlinear dc-PF voltages in Fig. 10(a). For either power flow model, the errors go down to zero at 3000 samples and are below 5% at 2300 samples. Similarly, for the ac-PF linear and nonlinear voltage profiles, both with and without noise, the errors reduce to zero for 6000 samples as shown in Fig. 10(b). At lower samples, however, it is clear that the errors increase with an increase in noise variance.

Note that all the experiments have asymptotically correct recovery (0 errors) at large sample sizes. They validate the theoretical consistency of our learning algorithm proven in the previous section.

Time requirements: We implemented the algorithm for all the networks in MATLAB R2020b on a workstation with Intel Xeon@ 3.7 GHz 32 GB RAM. On average, running Algorithm 1 for the 33-bus system takes 13.61 s. For the 56-bus system, it takes 26.89 s. Out of this, the first step involving estimating connections to zero-injection buses takes 26.88 s, while the second step of estimating the connections between nonzero injection buses takes 0.01 s. Note that the time-requirement for the first step can be reduced further by solving the regression problems, one for each bus, in parallel instead of sequentially as reported in this article.

VI. CONCLUSION

This article discusses topology learning using nodal voltages in general power grids under a realistic regime: one where a subset of the internal nodes in the grid have zero injection. This regime makes the voltage covariance matrix degenerate, and standard inverse covariance or mutual information-based learning methods do not apply. We provide a novel learning algorithm to overcome the presence of unknown zero-injection nodes, by first identifying them and their neighboring nodes, and then identifying the remaining edges in a Kron-reduced network without the zero-injection nodes. While the article considers zero-injection buses, the analysis based on voltage covariances extends naturally to grids with buses of constant injection and zero-injection variance. Notably, aside from voltage measurements, our algorithm does not use any prior information of structure, line parameters, values of nodal injections, or their statistics. The theoretical contributions are validated by simulation results on IEEE test networks, with both linear and nonlinear dc and ac voltage samples, corrupted with noise. Further, voltages generated using real injection data are used to calibrate the performance of our algorithm.

While our work focuses on power grids, it naturally extends to learning underexcited structured equation models that follow mass and flow conservation laws. Extensions to three-phase networks are possible using voltage statistics proposed for the fully excited setting [18]. While this article focuses on voltage availability at all nodes, we hope to extend our analysis to underexcited grids with partial voltage observability but with additional injection measurements. Similarly, extending our work to learning general dynamical systems [31] in the underexcited regime, is another direction of future work.

APPENDIX 1 PROOF OF LEMMA 2

We decompose any $a \in \mathbb{R}$ as $a = a^+ - a^-$, where $a^+ = \max(0, a) \geq 0$, and $a^- = -\min(0, a) \geq 0$. We list two inequalities that we use in the proof

$$\forall a, b \in \mathbb{R}, \text{ if } b \geq 0, \text{ then } (a + b)^+ \geq a^+ \quad (25)$$

$$\text{if } a > b, \text{ then } a^+ \geq b^+. \quad (26)$$

Consider y with $y_i = 1, y_{-i} = -a^{i*}$, where $a^{i*} > 0$ is the optimal noiseless solution in Problem (9) for $i \in \mathcal{W}$. Let $\sigma_{\min^+}(\Sigma_\theta)$ denote the smallest nonzero singular value of Σ_θ . We have

$$\begin{aligned} y_n^T (\Sigma_\theta + \Sigma_n) y_n &\geq y_n^T \Sigma_\theta y_n \stackrel{(a)}{\geq} y^T \Sigma_\theta y \geq \sigma_{\min^+}(\Sigma_\theta) \|y^R\|_2^2 \\ &\stackrel{(b)}{\geq} \sigma_{\min}(\Sigma_{\theta^w}) \|y^R\|_2^2 \end{aligned} \quad (27)$$

where (a) follows from optimality of y in the noiseless setting. y^R denotes the projection of y in the range of Σ_θ , and (b) follows from Cauchy's Interlace Theorem as Σ_{θ^w} is a principal submatrix of Σ_θ . By (12), $y^{RT} = y^T - c^T H_\beta^U$ for some $c \in \mathbb{R}^{|\mathcal{U}|}$. Then

$$\begin{aligned} \exists j_1, j_2 \in \mathcal{U}, \text{ s.t. } c_{j_1}^- &= \max_{j \in \mathcal{U}} c_j^-, c_{j_2}^+ = \max_{j \in \mathcal{U}} c_j^+, \text{ and} \\ \exists k_1 \in \mathcal{W} - \{i\}, \text{ s.t. } (j_1 k_1) &\in \mathcal{E} \text{ (using Assumption 2).} \end{aligned} \quad (28)$$

Using the structure of y and H_β^U , we have, for $i \in \mathcal{W}$

$$\|y_R\|_2^2 = \|y_R^T\|_2^2 = A + B + C \text{ where} \quad (29)$$

$$\begin{aligned} A &\left(1 - \sum_{j \in \mathcal{U}} c_j H_\beta(i, j)\right)^2 \\ &= \left(1 + \sum_{j \in \mathcal{U}} c_j^- H_\beta(i, j) - \sum_{j \in \mathcal{U}} c_j^+ H_\beta(i, j)\right)^2 \\ &\stackrel{(a_1)}{\geq} \left(1 + \sum_{j \in \mathcal{U}} c_j^- H_\beta(i, j)\right)^{+2} \stackrel{(a_2)}{\geq} (1 - c_{j_1}^- H_\beta(i, i))^{+2} \end{aligned} \quad (30)$$

$$\begin{aligned} B &\sum_{j \in \mathcal{U}} (-a_j - c_j H_\beta(j, j))^2 \\ &= \sum_{j \in \mathcal{U}} (c_j^+ H_\beta(j, j) + a_j - c_j^- H_\beta(j, j))^2 \\ &\stackrel{(b_1)}{\geq} (c_{j_2}^+ H_\beta(j_2, j_2))^2 \end{aligned} \quad (31)$$

$$\begin{aligned} C &\sum_{k \in \mathcal{W} - \{i\}} \left(-a_k - \sum_{j \in \mathcal{U}} c_j H_\beta(k, j)\right)^2 \\ &\geq \left(a_{k_1} + \sum_{j \in \mathcal{U}} c_j H_\beta(k_1, j)\right)^2 \\ &\stackrel{(c_1)}{\geq} \left(\sum_{j \in \mathcal{U}} c_j^+ H_\beta(k_1, j) - \sum_{j \in \mathcal{U}} c_j^- H_\beta(k_1, j)\right)^{+2} \\ &\stackrel{(c_2)}{\geq} \left(-c_{j_2}^+ H_\beta(k_1, k_1) - \sum_{j \in \mathcal{U}} c_j^- H_\beta(k_1, j)\right)^{+2} \\ &\geq (c_{j_1}^- H_\beta(k_1, j_1) - c_{j_2}^+ H_\beta(k_1, k_1))^{+2} \end{aligned}$$

$$\stackrel{(c_3)}{\geq} (c_{j_1}^- H_\beta(i, i)/s_{id} - c_{j_2}^+ H_\beta(j_2, j_2)s_{id})^2. \quad (32)$$

Here, $(a_1), (b_1), (c_1)$ follow from (25), and $(a_2), (c_2)$ follow from (26) and (28). (c_3) follows from the definition of s_{id} in (17) and (26). Using (30)–(32) in (29), we have

$$\|y_R\|_2^2 \geq (1 - x_1)^2 + x_2^2 + (x_1/s_{id} - x_2 s_{id})^2 \quad (33)$$

where $x_1 = c_{j_1}^- H_\beta(i, i) \geq 0$ and $x_2 = c_{j_2}^+ H_\beta(j_2, j_2) \geq 0$. Using the minimum value of quadratic functions, consider the cases for x_1, x_2 and their effect on $\|y_R\|_2^2$

$$(a) \ x_2 \geq 1/s_{id}^2 :$$

$$\|y_R\|_2^2 \geq 1/s_{id}^4 \quad (34)$$

$$(b) \ x_2 < 1/s_{id}^2, x_1 \geq 1 :$$

$$\|y_R\|_2^2 \geq \min_{x_2} x_2^2 + (1/s_{id} - x_2 s_{id})^2 = \frac{1}{s_{id}^4 + s_{id}^2} \quad (35)$$

$$(c) \ x_2 < 1/s_{id}^2, x_2 s_{id}^2 < x_1 < 1 :$$

$$\begin{aligned} \|y_R\|_2^2 &\geq \min_{x_1, x_2} (1 - x_1)^2 + x_2^2 + (x_1/s_{id} - x_2 s_{id})^2 \\ &= 1/(1 + s_{id}^4 + s_{id}^2) \end{aligned} \quad (36)$$

$$(d) \ x_2 < 1/s_{id}^2, x_1 \leq x_2 s_{id}^2 :$$

$$\|y_R\|_2^2 \geq \min_{x_2} (1 - x_2 s_{id}^2)^2 + x_2^2 = 1/(1 + s_{id}^4). \quad (37)$$

Using (34)–(37), $\|y_R\|_2^2 \geq 1/(1 + s_{id}^4 + s_{id}^2)$. Using this in (27) completes the proof.

APPENDIX 2 PROOF OF THEOREM 5

Consider $i \in \mathcal{U}$. From the proof of Theorem 1, note that zero-cost solutions in Problem (9) are given by $y^T = c^T H_\beta^U$, where $y_i = 1, y_{-i} < 0$.

$y_j \leq 0 \ \forall j \in \mathcal{U} - \{i\}$, necessitates $c_j \leq 0 \ \forall j \neq i$. Consider $j_1 \in \mathcal{U} - \{i\}$. Under the assumption in the statement, $\exists k_1 \in \mathcal{W}$ such that $(k_1 j_1) \in \mathcal{E}$, but $(k_1 i) \notin \mathcal{E}$. Thus, $y_{k_1} = -c_{j_1} \beta_{j_1 k_1} + \sum_{j \in \mathcal{U} - \{i, j_1\}} c_j H_\beta(j, k_1)$, where $H_\beta(j, k_1) \leq 0$ follows from (8). As c_j is nonpositive for $j \neq i, y_{k_1} \leq 0$ holds only if $c_{j_1} = 0$. Following the same reasoning $\forall j \in \mathcal{U} - \{i\}$, $c_j = 0$. Finally, $c_i = 1/H_\beta(i, i)$ is needed for $y_i = 1$. Thus, $y = \frac{H_\beta^T(i, :)}{H_\beta(i, i)}$ is the unique zero-cost solution for $i \in \mathcal{U}$. It identifies the true neighbors of i , due to the structure of $H_\beta^T(i, :)$.

REFERENCES

- [1] D. Deka, S. Backhaus, and M. Chertkov, "Structure learning in power distribution networks," *IEEE Control Netw. Syst.*, vol. 5, no. 3, pp. 1061–1074, Sep. 2018.
- [2] W. H. Kersting, "Radial distribution test feeders," in *Proc. Power Eng. Soc. Winter Meeting*, 2001, vol. 2, pp. 908–912.
- [3] R. Hoffman, "Practical state estimation for electric distribution networks," in *Proc. Power Syst. Conf. Expo.*, 2006, pp. 510–517.
- [4] R. Arghandeh and Y. Zhou, *Big Data Application in Power Systems*. Amsterdam, The Netherlands: Elsevier, 2017.
- [5] A. Phadke, "Synchronized phasor measurements in power systems," *IEEE Comput. Appl. Power*, vol. 6, no. 2, pp. 10–15, Apr. 1993.
- [6] A. Von Meier, D. Culler, A. McEachern, and R. Arghandeh, "Micro-synchrophasors for distribution systems," in *Proc. Innov. Smart Grid Technol. Conf.*, 2014, pp. 1–5.
- [7] M. U. Usman and M. O. Faruque, "Applications of synchrophasor technologies in power systems," *J. Modern Power Syst. Clean Energy*, vol. 7, no. 2, pp. 211–226, 2019.
- [8] J. Peppanen, J. Grimaldo, M. J. Reno, S. Grijalva, and R. G. Harley, "Increasing distribution system model accuracy with extensive deployment of smart meters," in *Proc. IEEE Power Energy Soc. Gen. Meeting Conf. Expo.*, 2014, pp. 1–5.
- [9] S. Park, D. Deka, and M. Chertkov, "Exact topology and parameter estimation in distribution grids with minimal observability," in *Proc. Power Syst. Comput. Conf.*, 2018, pp. 1–6.
- [10] G. Cavararo 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.
- [11] G. Cavararo and V. Kekatos, "Inverter probing for power distribution network topology processing," *IEEE Control Netw. Syst.*, vol. 6, no. 3, pp. 980–992, Sep. 2019.
- [12] C. Rudin *et al.*, "Machine learning for the new york city power grid," *IEEE Trans. Mach. Intell.*, vol. 34, no. 2, pp. 328–345, Feb. 2012.
- [13] D. Wolter, M. Zdrallek, M. Stötzl, C. Schacherer, I. Mladenovic, and M. Biller, "Impact of meshed grid topologies on distribution grid planning and operation," *CIREN-Open Access Proc. J.*, vol. 2107, pp. 2338–2341, 2017.
- [14] J. Yu, Y. Weng, and R. Rajagopal, "PaToPaEM: A data-driven parameter and topology joint estimation framework for time-varying system in distribution grids," *IEEE Trans. Power Syst.*, vol. 34, no. 3, pp. 1682–1692, May 2018.
- [15] Y. Yuan, O. Ardakanian, S. Low, and C. Tomlin, "On the inverse power flow problem," 2016, *arXiv:1610.06631*.
- [16] 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*, 2013, pp. 1659–1664.
- [17] M. He and J. Zhang, "A dependency graph approach for fault detection and localization towards secure smart grid," *IEEE Trans. Smart Grid*, vol. 2, no. 2, pp. 342–351, Jun. 2011.
- [18] D. Deka, M. Chertkov, and S. Backhaus, "Topology estimation using graphical models in multi-phase power distribution grids," *IEEE Trans. Power Syst.*, vol. 35, no. 3, pp. 1663–1673, May 2020.
- [19] 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.
- [20] D. Deka, S. Talukdar, M. Chertkov, and M. V. Salapaka, "Graphical models in meshed distribution grids: Topology estimation, change detection and limitations," *IEEE Trans. Smart Grid*, vol. 11, no. 5, pp. 4299–4310, 2020.
- [21] K. A. Bollen and M. D. Noble, "Structural equation models and the quantification of behavior," in *Proc. Nat. Acad. Sci.*, vol. 108, Sep. 13, 2011.
- [22] M. Baran and F. Wu, "Network reconfiguration in distribution systems for loss reduction," *IEEE Trans. Power Del.*, vol. 4, no. 2, pp. 1401–1407, Apr. 1989.
- [23] S. Bolognani and S. Zampieri, "On the existence and linear approximation of the power flow solution in power distribution networks," *IEEE Trans. Power Syst.*, vol. 31, no. 1, pp. 163–172, Jan. 2016.
- [24] Y. Dvorkin, M. Lubin, S. Backhaus, and M. Chertkov, "Uncertainty sets for wind power generation," *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 3326–3327, Jul. 2016.
- [25] F. Dorfler and F. Bullo, "Kron reduction of graphs with applications to electrical networks," *IEEE Trans. Circuits Syst.*, vol. 60, no. 1, pp. 150–163, Jan. 2013.
- [26] W. W. Hager, "Updating the inverse of a matrix," *SIAM Rev.*, vol. 31, no. 2, pp. 221–239, 1989.
- [27] "IEEE 1547 Standard for interconnecting distributed resources with electric power systems," 2014. [Online]. Available: <http://grouper.ieee.org/>
- [28] R. D. Zimmerman and C. E. Murillo-Sanchez, *Matpower 4.1 User's Manual*, Power Systems Engineering Research Center (PSERC), vol. 20, 2011.
- [29] R. Pedersen, C. Sloth, G. B. Andresen, and R. Wisniewski, "Disc: A simulation framework for distribution system voltage control," in *Proc. Eur. Control Conf.*, 2015, pp. 1056–1063.
- [30] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," 2014.
- [31] S. Talukdar, D. Deka, H. Doddi, D. Materassi, M. Chertkov, and M. V. Salapaka, "Physics informed topology learning in networks of linear dynamical systems," *Automatica*, vol. 112, 2020, Art. no. 108705.



Deepjyoti Deka (Senior Member, IEEE) received the undergraduate degree in electrical engineering from IIT Guwahati, Guwahati, India, and the M.S. and Ph.D. degrees in electrical engineering from the University of Texas, Austin, TX, USA, in 2011 and 2015, respectively.

He is a Staff Scientist with the Theoretical Division, Los Alamos National Laboratory (LANL), Los Alamos, NM, USA, where he was previously a Postdoctoral Research Associate with the Center for Nonlinear Studies. His research interests include data-analysis of power grid structure and security, and optimization in social and physical networks. At LANL, he serves as a Co-Principal Investigator for DOE projects on machine learning in power systems and in cyber-physical security. Dr. Deka was the recipient of IIT Guwahati's silver medal as the best outgoing student of the department in 2009.



Sidhant Misra received the bachelor's degree in electrical engineering from the Indian Institute of Technology, Kanpur, India, in 2008 and the M.S. and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2011 and 2014, respectively.

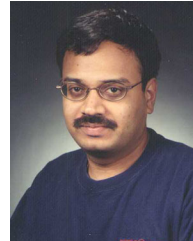
He is a Staff Scientist with the Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, USA, and part of the Advanced Network Sciences Initiative (ANSI). His research interests include intersection of machine learning and optimization and applications to planning and operations of energy infrastructure systems.



Harish Doddi received the B.Tech. degree in mechanical engineering from the Indian Institute of Technology Madras, Chennai, India, in 2014. He is currently working toward the Ph.D. degree in mechanical engineering at the University of Minnesota, Minneapolis, MN, USA.

He worked on thermal analysis of commercial buildings with the TCS Innovation Lab, Chennai, India till 2016. His research includes information thermodynamics, energy limits in computing and structure learning of graphical

models.



Murti V. Salapaka (Fellow, IEEE) received the B.Tech. degree in mechanical engineering from the Indian Institute of Technology, Madras, Chennai, India, in 1991 and the M.S. and Ph.D. degrees in mechanical engineering from the University of California at Santa Barbara, Santa Barbara, CA, USA, in 1993 and 1997, respectively.

He was a Faculty Member with the Electrical and Computer Engineering Department, Iowa State University, Ames, IA, USA, from 1997 to 2007. Currently, he is the Director of Graduate Studies and the Vincentine Hermes Luh Chair Professor with the Electrical and Computer Engineering Department, University of Minnesota, Minneapolis, MN, USA. His research interests include control and network science, nanoscience, and single molecule physics.

Dr. Salapaka was the recipient of the 1997 National Science Foundation CAREER Award.