Learning Power Grid Topologies

Guido Cavraro, Vassilis Kekatos, Liang Zhang, and
Georgios B. Giannakis [1]

0.1   Introduction

To perform any meaningful grid optimization task, distribution utilities need to know the topology and line impedances of their grid assets. One may distinguish two major topology learning tasks: *i)* In *topology detection*, the utility knows the existing line infrastructure and impedances, and wants to find which lines are energized; and *ii)* In *topology identification*, the utility aims at identifying both the connectivity and line impedances; hence it is a harder task.

Grid topology learning oftentimes relies on second-order statistics from smart meter data [1, 2, 3]. Nonetheless, sample statistics converge to their ensemble values only after a large number of grid data has been collected. Detecting which lines are energized can be posed as a maximum likelihood detection task [4]; sparse linear regression [5, 6]; or as a spanning tree recovery task using the notion of graph cycles [7]. Line impedances are estimated via a total least-squares fit in [8]. In [9], deep neural networks are trained to detect which lines are energized; nevertheless, the dataset feeding the classifiers may not be available in distribution grids. A Kron-reduced admittance matrix is recovered via a low rank-plus-sparse decomposition in [10], though the deployment of micro-phasor measurement units presumed there occurs at a slower pace in distribution grids.

This chapter presents a gamut of statistical tools for learning the topology of distribution grids. Towards this end, utilities could rely on smart meter data polled from customers' sites. In addition to passively collecting data, this chapter puts forth an active data acquisition paradigm that we term *grid probing using smart inverters*. The rest of the chapter is organized as follows. Section 0.2 reviews an approximate grid model. Assuming smart meter data, Section 0.3 poses topology detection as a statistical learning task. The method of partial correlations is extended to the nonlinear setting to detect meshed grid topologies in Section 0.4. Section 0.5 puts forth the novel data acquisition paradigm of grid probing through smart inverters. Section 0.6 provides conditions on inverter placement to ensure topology identifiability. Once specific blocks of the inverse Laplacian matrix have been recovered from probing data, a radial grid topology can be identified using the graph algorithms of Section 0.7. Because these algorithms may become impractical under low signal-to-noise ratios, the grid topology can be identified using the convex relaxation approach of Section 0.8. The different algorithmic alternatives for topology identifiability using probing data are tested in Section 0.9. The chapter is concluded in Section 0.10.

Regarding notation, lower- (upper-) case boldface letters denote column vectors (matrices), and calligraphic letters are reserved for sets. Vectors $\mathbf{0}$, $\mathbf{1}$, and $\mathbf{e}_n$ denote respectively the all-zero, all-one, and the n-th canonical vectors. The superscript ^{T} stands for transposition and $\text{Tr}(\mathbf{X})$ is the trace of $\mathbf{X}$. The operator $\text{dg}(\mathbf{x})$ returns a diagonal matrix with $\mathbf{x}$ on its main diagonal. For $\mathbf{W} \succ \mathbf{0}$, define the norm $\|\mathbf{X}\|_{\mathbf{W}}^2 := \text{Tr}(\mathbf{X}^\top \mathbf{W} \mathbf{X}) = \|\mathbf{W}^{1/2}\mathbf{X}\|_F^2$, where $\|\mathbf{X}\|_F$ is the Frobenius norm of $\mathbf{X}$. The pseudo-norm $\|\mathbf{X}\|_{0,\text{off}}$ counts the number of non-zero off-diagonal entries of $\mathbf{X}$.

0.2 Grid Modeling

We build upon an approximate distribution grid model briefly reviewed next. A radial single-phase grid having $N + 1$ buses can be represented by a tree graph $\mathcal{G} = (\mathcal{N}_o, \mathcal{L})$, whose nodes $\mathcal{N}_o := \{0, \dots, N\}$ correspond to buses and its edges $\mathcal{L}$ to lines. The tree is rooted at the substation indexed by $n = 0$. Let $p_n + jq_n$ be the complex power injection and $v_n e^{j\phi_n}$ the complex voltage phasor at bus $n \in \mathcal{N}$. Collect the voltage magnitudes and phases and power injections of buses in $\mathcal{N}$ in the vectors $\mathbf{v}$, $\boldsymbol{\phi}$, $\mathbf{p}$ and $\mathbf{q}$, respectively. The impedance of line $\ell : (m, n) \in \mathcal{L}$ is denoted by $r_\ell + jx_\ell$ or $r_{mn} + jx_{mn}$, depending on the context. The grid connectivity is captured by the branch-bus incidence matrix $\tilde{\mathbf{A}} \in \{0, \pm 1\}^{L \times (N+1)}$; which can be partitioned into the first and the rest of its columns as $\tilde{\mathbf{A}} = [\mathbf{a}_0 \ \mathbf{A}]$. For a radial grid $(L = N)$, the *reduced incidence matrix* $\mathbf{A}$ is square and invertible [11].

The complex power injections are nonlinear functions of voltages. Nonetheless, the power flow equations are oftentimes linearized at the flat voltage profile of $v_n = 1$ and $\phi_n = 0$ for all n, to yield the linearized grid model [12, 13, 14]

$$p_n = \sum_{(n,m)\in\mathcal{L}} g_{nm}(v_n - v_m) + b_{nm}(\phi_n - \phi_m) \qquad (0.1a)$$

$$q_n = \sum_{(n,m)\in\mathcal{L}} b_{nm}(v_n - v_m) - g_{nm}(\phi_n - \phi_m). \qquad (0.1b)$$

where $g_{nm} + jb_{nm} := 1/(r_{nm} + jx_{nm})$ is the admittance of line (n, m). The model in (0.1) constitutes a system of $2N$ linear equations of $2N$ unknowns, namely the voltage magnitudes and phases. The system can be inverted to obtain

$$\begin{bmatrix} \mathbf{v} \\ \boldsymbol{\phi} \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{X} \\ \mathbf{X} & -\mathbf{R} \end{bmatrix} \begin{bmatrix} \mathbf{p} \\ \mathbf{q} \end{bmatrix} + \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix}. \qquad (0.2)$$

where the $N \times N$ matrices $\mathbf{R}$ and $\mathbf{X}$ are defined as

$$\mathbf{R} := (\mathbf{A}^\top \mathrm{dg}^{-1}(\mathbf{r})\mathbf{A})^{-1} \quad \text{and} \quad \mathbf{X} := (\mathbf{A}^\top \mathrm{dg}^{-1}(\mathbf{x})\mathbf{A})^{-1}. \qquad (0.3)$$

It is worth mentioning that the widely used *linearized distribution flow (LDF)* model originally proposed by [12], involves the squared voltage magnitudes rather than the voltage magnitudes in (0.2). For this reason, the matrices $\mathbf{R}$ and $\mathbf{X}$ appearing in LDF take the values of (0.3) divided by a factor of 2; see also [13]. Note finally that different from (0.2), the LDF model does not provide a linearized model to approximate voltage angles by complex power injections.

The topology learning schemes presented later in this chapter do not consider voltage phases. This is because, despite the recent developments of phasor measurement unity (PMU) for distribution grids, smart meters report to utilities only voltage magnitudes, henceforth simply referred to as voltages.

Voltage samples are collected at a sampling period of T_s and indexed by $t = 0, \dots, T$. Upon applying (0.2) over two consecutive voltage samples, the changes

in voltages caused by changes in power injections $\tilde{\mathbf{p}}_t := \mathbf{p}_t - \mathbf{p}_{t-1}$ and $\tilde{\mathbf{q}}_t := \mathbf{q}_t - \mathbf{q}_{t-1}$ can be modeled as

$$\tilde{\mathbf{v}}_t := \mathbf{v}_t - \mathbf{v}_{t-1} = \mathbf{R}\tilde{\mathbf{p}}_t + \mathbf{X}\tilde{\mathbf{q}}_t + \mathbf{n}_t \tag{0.4}$$

where $\mathbf{n}_t$ captures measurement noise; the approximation error introduced by the linearized grid model; and unmodeled load dynamics. The advantages of using the differential model of (0.4) over (0.2) are two: When dealing smart meter data, the operator may be observing some or none of the entries of $(\mathbf{p}_t, \mathbf{q}_t)$, so then one can only exploit the second-order moments of $(\tilde{\mathbf{p}}_t, \tilde{\mathbf{q}}_t)$ via blind signal processing techniques (see Section 0.3). With active data collection, power injections can be intentionally perturbed over short intervals, so that $(\tilde{\mathbf{p}}_t, \tilde{\mathbf{q}}_t)$ take zero entries for non-actuated buses and known non-zero entries for actuated buses. Then, topology learning can be cast as a system identification task; see Section 0.5.

0.3 Topology Detection using Smart Meter Data

Feeders are built with redundancy in line infrastructure. This redundancy improves system reliability against failures or during scheduled maintenance, while grids are reconfigured for loss minimization [12]: At each time, not all existing lines in $\mathcal{L}$ are energized. We will be considering a period of operation where the subset of energized lines $\mathcal{E} \subset \mathcal{L}$ with $|\mathcal{E}| = L_e$ remains constant, yet unknown.

0.3.1 A maximum likelihood (ML) approach

To capture the status of each line, introduce variable b_ℓ for line ℓ taking values $b_\ell = 1$ if line is energized ($\ell \in \mathcal{L}_e$); and $b_\ell = 0$ otherwise. Collect the b_ℓ's in the binary L_e–length vector $\mathbf{b}$. To verify grid topologies using smart meter data, parameterize $(\mathbf{R}, \mathbf{X})$ from (0.3) as

$$\mathbf{R}(\mathbf{b}) = \left(\sum_{\ell \in \mathcal{L}} \frac{b_\ell}{r_\ell} \mathbf{a}_\ell \mathbf{a}_\ell^\top\right)^{-1} \quad \text{and} \quad \mathbf{X}(\mathbf{b}) = \left(\sum_{\ell \in \mathcal{L}} \frac{b_\ell}{x_\ell} \mathbf{a}_\ell \mathbf{a}_\ell^\top\right)^{-1} \tag{0.5}$$

where $\mathbf{a}_\ell^\top$ is the ℓ-th row of $\mathbf{A}$. By slightly abusing notation, matrix $\mathbf{A}$ here has been augmented to include both energized and non-energized lines. Under this representation, verifying the grid topology entails finding $\mathbf{b}$ from grid data.

The utility collects voltage readings, but power injections are described only through their first- and second-order moments. Smart meter data comprise voltages and (re)active powers. Nonetheless, a smart meter monitors a single household, which may not necessarily correspond to a bus. The nodes of low-voltage grids are typically mapped to pole transformers, each one serving $5 - 10$ residential costumers. Hence, power readings from meters may not be useful. Interpreting powers and voltages as system inputs and outputs, respectively, topology detection has to be posed as a blind system identification problem where power

injections (system inputs) are characterized only statistically. To this end, we postulate the ensuing statistical model; see also [14, 15, 2].

ASSUMPTION 1   *Differential injection data* $(\tilde{\mathbf{p}}_t, \tilde{\mathbf{q}}_t)$ *are zero-mean random vectors with covariance matrices* $\boldsymbol{\Sigma}_p := \mathbb{E}[\tilde{\mathbf{p}}_t\tilde{\mathbf{p}}_t^\top]$; $\boldsymbol{\Sigma}_q := \mathbb{E}[\tilde{\mathbf{q}}_t\tilde{\mathbf{q}}_t^\top]$; *and* $\boldsymbol{\Sigma}_{pq} := \mathbb{E}[\tilde{\mathbf{p}}_t\tilde{\mathbf{q}}_t^\top]$. *Noise* $\mathbf{n}_t$ *is a zero-mean independent identically distributed (iid) Gaussian random vector with covariance matrix* $\sigma_n^2\mathbf{I}_N$.

Under Assumption 1 and from (0.4)–(0.5), the differential voltages $\{\tilde{\mathbf{v}}_t\}$ (termed *voltage data* for brevity) are zero-mean with covariance matrix parameterized as

$$\boldsymbol{\Sigma}(\mathbf{b}) = \mathbf{R}(\mathbf{b})\boldsymbol{\Sigma}_p\mathbf{R}(\mathbf{b}) + \mathbf{X}(\mathbf{b})\boldsymbol{\Sigma}_q\mathbf{X}(\mathbf{b}) + \mathbf{R}(\mathbf{b})\boldsymbol{\Sigma}_{pq}\mathbf{X}(\mathbf{b}) + \mathbf{X}(\mathbf{b})\boldsymbol{\Sigma}_{pq}^\top\mathbf{R}(\mathbf{b}) + \sigma_n^2\mathbf{I}_N.$$

We postulate that the probability density function (pdf) of each random vector $\tilde{\mathbf{v}}_t$ converges asymptotically in N to a multivariate Gaussian pdf, even if $(\tilde{\mathbf{p}}_t, \tilde{\mathbf{q}}_t)$ are not Gaussian; this stems from contemporary variants of the central limit theorem as detailed in [16]. Statistical tests on actual data validate this assumption [16]. Thus, the pdf of $\tilde{\mathbf{v}}_t$ for each t can be approximated as

$$p(\tilde{\mathbf{v}}_t; \mathbf{b}) = \frac{|\boldsymbol{\Sigma}(\mathbf{b})|^{-1/2}}{(2\pi)^{N/2}} \exp\left(-\frac{1}{2}\tilde{\mathbf{v}}_t^\top\boldsymbol{\Sigma}^{-1}(\mathbf{b})\tilde{\mathbf{v}}_t\right).$$

To fully characterize the collected voltage data $\{\tilde{\mathbf{v}}_t\}_{t=1}^T$, their joint pdf should be provided. It has been demonstrated that voltage data are relatively uncorrelated across time, especially for sampling periods larger than $T_s = 5$ min, see [16, Fig. 3]. Due to Gaussianity, uncorrelatedness implies independence. Therefore, the joint pdf for the entire voltage dataset becomes

$$p(\{\tilde{\mathbf{v}}_t\}_{t=1}^T; \mathbf{b}) = \prod_{t=1}^T p(\tilde{\mathbf{v}}_t; \mathbf{b}) = \frac{|\boldsymbol{\Sigma}(\mathbf{b})|^{-T/2}}{(2\pi)^{NT/2}} \exp\left(-\frac{1}{2}\sum_{t=1}^T \tilde{\mathbf{v}}_t^\top\boldsymbol{\Sigma}^{-1}(\mathbf{b})\tilde{\mathbf{v}}_t\right). \quad (0.6)$$

From the preceding modeling, *topology detection* amounts to finding the subset $\mathcal{E}$ given: the line infrastructure, that is $\{r_\ell, x_\ell, \mathbf{a}_\ell\}_{\ell\in\mathcal{L}}$; the covariance matrices $\boldsymbol{\Sigma}_p, \boldsymbol{\Sigma}_q, \boldsymbol{\Sigma}_{pq}$; and voltage data $\{\tilde{\mathbf{v}}_t\}_{t=1}^T$. Upon observing $\{\tilde{\mathbf{v}}_t\}_{t=1}^T$, function (0.6) becomes the likelihood function of the line indicator vector $\mathbf{b}$. After ignoring constant terms and adopting a maximum likelihood (ML) approach, vector $\mathbf{b}$ can be found as the minimizer of the negative log-likelihood function

$$\hat{\mathbf{b}} := \arg\min_{\mathbf{b}} \left\{f(\mathbf{b}) : \mathbf{b} \in \{0, 1\}^L, \mathbf{1}^\top\mathbf{b} = N\right\} \quad (0.7)$$

where $f(\mathbf{b}) := \log|\boldsymbol{\Sigma}(\mathbf{b})| + \text{Tr}(\boldsymbol{\Sigma}^{-1}(\mathbf{b})\hat{\boldsymbol{\Sigma}})$; the operator $|\cdot|$ is the matrix determinant; and $\hat{\boldsymbol{\Sigma}} := \frac{1}{T}\sum_{t=1}^T \tilde{\mathbf{v}}_t\tilde{\mathbf{v}}_t^\top$ is the sample covariance of voltage data. The second summand in $f(\mathbf{b})$ aggregates the information from data, while the first one acts as a regularizer guarding $\boldsymbol{\Sigma}(\mathbf{b})$ within the positive definite matrix cone and away from singularity [17]. The constraint $\mathbf{1}^\top\mathbf{b} = N$ ensures a tree structure.

Solving (0.7) is non-trivial due to the binary variable and the nonlinear objective. A lower bound can be obtained by solving the box relaxation of (0.7)

$$\check{\mathbf{b}} := \arg\min_{\mathbf{b}} \left\{f(\mathbf{b}) : \mathbf{b} \in [0, 1]^L, \mathbf{1}^\top\mathbf{b} = N\right\}. \quad (0.8)$$

Since $f(\mathbf{b})$ is non-convex in general, one may only be able to find a local minimum of (0.8). Consistent with the properties of the MLE, the true indicator vector minimizes (0.7) and (0.8), when the number of data T grows to infinity.

PROPOSITION 1 ([16])   *Let $\mathbf{b}_o$ be the true line indicator vector. If the sample covariance $\hat{\boldsymbol{\Sigma}}$ has converged to the ensemble covariance $\boldsymbol{\Sigma}(\mathbf{b}_o)$, then $\mathbf{b}_o$ is a stationary point of $f(\mathbf{b})$ and global minimizer for (0.7) and (0.8).*

To obtain a feasible $\mathbf{b}$, one may apply a heuristic on $\check{\mathbf{b}}$, such as selecting the lines corresponding to the L largest entries of $\check{\mathbf{b}}$; or finding the minimum spanning tree on a graph having $\check{\mathbf{b}}$ as edge weights. Obviously $f^\star \leq f(\hat{\mathbf{b}})$ and $f(\check{\mathbf{b}}) \leq f(\hat{\mathbf{b}})$. Even though $\hat{\mathbf{b}}$ yields reasonable detection performance (see Sec. 0.9), vector $\check{\mathbf{b}}$ may not be a global minimizer of (0.8).

The issues with the non-convexity of $f(b)$ can be alleviated by resorting to two simplifying assumptions: *a1)* the resistance-to-reactance ratios are identical or $r_\ell/x_\ell = \alpha$ for all lines $\ell \in \mathcal{L}$; and *a2)* the noise term n_t is negligible. In this case, the voltage data model of (0.4) simplifies as $\tilde{\mathbf{v}}_t = \alpha\mathbf{X}\tilde{\mathbf{p}}_t + \mathbf{X}\tilde{\mathbf{q}}_t$, and the ensemble covariance of voltage data becomes

$$\boldsymbol{\Sigma}(\mathbf{b}) := \mathbf{X}(\mathbf{b})(\alpha^2\boldsymbol{\Sigma}_p + \boldsymbol{\Sigma}_q + \alpha(\boldsymbol{\Sigma}_{pq} + \boldsymbol{\Sigma}_{pq}^\top)\mathbf{X}(\mathbf{b}).$$

Under *a1)-a2)*, the original negative log-likelihood is surrogated by

$$\tilde{f}(\mathbf{b}) := -2\log|\mathbf{X}^{-1}(\mathbf{b})| + \mathrm{Tr}\left(\mathbf{X}^{-1}(\mathbf{b})\boldsymbol{\Sigma}_\alpha^{-1}\mathbf{X}^{-1}(\mathbf{b})\hat{\boldsymbol{\Sigma}}^{-1}\right).$$

Interestingly, function $\tilde{f}(\mathbf{b})$ is convex and so (0.8) becomes a convex program [16].

0.3.2   A maximum a-posteriori probability (MAP) approach

In meshed grids, the utility may not know the exact number of energized lines L. However, prior information on individual lines being energized could be known through the generalized state estimator or current readings on transformers. To cope with such scenarios, a maximum *a-posteriori* probability (MAP) approach can be adopted. The indicator b_ℓ for line ℓ is modeled as a Bernoulli random variable with given mean $\mathbb{E}[b_\ell] = \pi_\ell$. The prior probability mass function (pmf) for b_ℓ can be expressed as $\mathrm{Pr}(b_\ell) = \pi_\ell^{b_\ell}(1-\pi_\ell)^{1-b_\ell}$. To derive a tractable statistical model, lines are assumed to be energized independently. Then, the joint pmf for $\mathbf{b}$ is $\mathrm{Pr}(\mathbf{b}) = \prod_{\ell\in\mathcal{L}_e}\mathrm{Pr}(b_\ell)$ up to a normalization constant and then

$$-\log\mathrm{Pr}(\mathbf{b}) = \sum_{\ell\in\mathcal{L}_e} b_\ell\beta_\ell - \log(1-\pi_\ell) \tag{0.9}$$

where $\beta_\ell := \log\left(\frac{1-\pi_\ell}{\pi_\ell}\right)$ for all $\ell \in \mathcal{L}_e$. The MAP estimate for $\mathbf{b}$ is defined as the maximizer of the posterior $p(\mathbf{b}|\{\tilde{\mathbf{v}}_t\}_{t=1}^T)$. From Bayes' rule, the latter is proportional to the product $p(\{\tilde{\mathbf{v}}_t\}_{t=1}^T;\mathbf{b})\mathrm{Pr}(\mathbf{b})$, so that the MAP estimate can be found by minimizing

$$-\log p(\mathbf{b}|\{\tilde{\mathbf{v}}_t\}_{t=1}^T) = -\log p(\{\tilde{\mathbf{v}}_t\}_{t=1}^T;\mathbf{b}) - \log\mathrm{Pr}(\mathbf{b}).$$
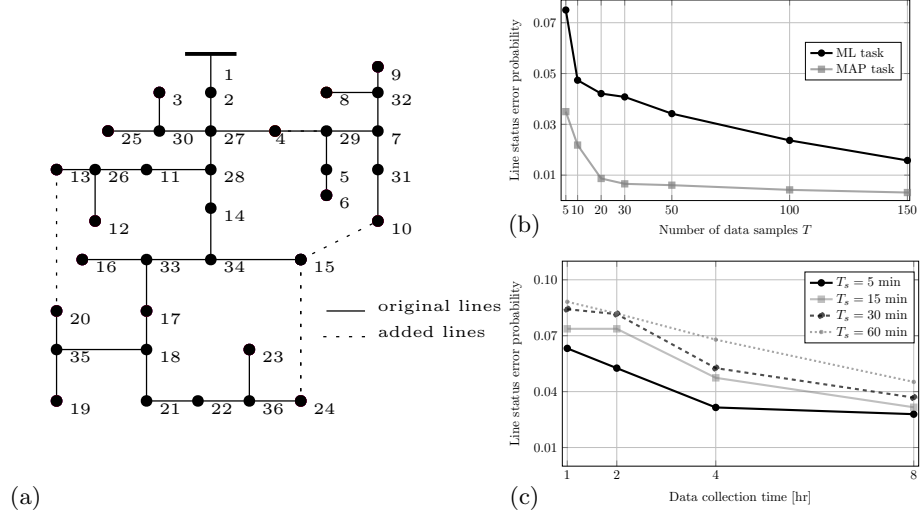
Figure 0.1 a) Line structure for the IEEE 37-bus feeder. b) Line error probability for the ML and the MAP task. c) Line error probability for the ML task (0.10). c) Effect of T_s on the line status error probability for the ML task.

Collecting β_ℓ's in $\boldsymbol{\beta}$ and ignoring constants, the latter leads to the problem

$$\mathbf{b}_{\mathrm{MAP}} := \arg \min_{\mathbf{b} \in \{0,1\}^L} \; \frac{T}{2} f(\mathbf{b}) + \boldsymbol{\beta}^\top \mathbf{b} \tag{0.10}$$

Contrasted to (0.7), problem (0.10) leverages prior information: If line ℓ is likely to be energized, then $\pi_\ell > 1/2$ and $\beta_\ell < 0$. If line ℓ is known to be energized, then $\pi_\ell = 1$ and $\beta_\ell = -\infty$, thus forcing the ℓ-th entry of $\mathbf{b}_{\mathrm{MAP}}$ to one. No prior information on line ℓ means $\pi_\ell = 1/2$ and $\beta_\ell = 0$. If $\pi_\ell = \pi_o$ for $\ell \in \mathcal{L}_e$, then $\boldsymbol{\beta}^\top \mathbf{b} = \pi_o \mathbf{1}^\top \mathbf{b}$, and (0.10) takes a Lagrangian form of (0.7). Similar to (0.7), a box relaxation of (0.10) can be pursued.

0.3.3 Numerical tests on topology detection using smart meter data

The schemes were validated using the modified version of the IEEE 37-bus system depicted in Figures 0.1 a), in which additional lines were added. Data collected from the Pecan Street project were used as power profiles [18]. Voltages were obtained through a power flow solver and the measurement noise was modeled as zero-mean Gaussian with a 3-sigma deviation matching 0.5% of the actual value [16]. The ML task (0.8) was solved by a projected gradient (PGD) scheme over 50 Monte Carlo runs and a sampling period of $T_s = 5$ min. The actual topology was randomly chosen. The solutions $\check{\mathbf{b}}$ were projected onto the feasible set via randomization by treating $\check{\mathbf{b}}$ as the mean of a multivariate Bernoulli distribution [16]. The MAP approach (0.10) was tested via 50 Monte Carlo runs and solved by a PGD scheme too. The prior probabilities π_ℓ were set to 0.5 for

switches and 0.9 for lines. The entries of the solution were truncated to binary upon thresholding. The line error probabilities for the two problems are depicted in Fig. 0.1 b). It is worth emphasizing that reliable line detection can be obtained even for $T < N$, when matrix $\hat{\boldsymbol{\Sigma}}$ is *singular*. Figure 0.1 c) tests the effect of T_s on the line status error probability achieved by the non-convex ML task of (0.8). The performance improves as T_s and the total collection time increase.

0.4   Topology Detection via Partial Correlations

Collect the data associated with bus n for $t =, 1 \ldots, T$, into vector $\mathbf{x}_n :=$ $[x_n[1] \ \ldots \ x_n[T]]^\top$. For power network topology inference, vector $\mathbf{x}_n$ may collect voltage readings across time [2]. Given $\{\mathbf{x}_n\}_{n \in \mathcal{N}}$, we would like to recover the edge set $\mathcal{L}$. To this end, we adopt the method of Partial Correlations (PC), which is effective in capturing unmediated linear influence between nodes [2, 19, 20]. Consider two buses $\{n, m\}$ and define the error vector $\boldsymbol{\epsilon}_{n|m} = \mathbf{x}_n - \hat{\mathbf{x}}_{n|m}$, where $\hat{\mathbf{x}}_{n|m}$ is the estimate of $\mathbf{x}_n$ based on $\{\mathbf{x}_i\}$ for $i \notin \{n, m\}$. The *empirical PC coefficient* between $\mathbf{x}_n$ and $\mathbf{x}_m$ is given by [20]

$$\rho_{nm} := \frac{\left(\boldsymbol{\epsilon}_{n|m} - \bar{\boldsymbol{\epsilon}}_{n|m}\right)^\top \left(\boldsymbol{\epsilon}_{m|n} - \bar{\boldsymbol{\epsilon}}_{m|n}\right)}{\|\boldsymbol{\epsilon}_{n|m} - \bar{\boldsymbol{\epsilon}}_{n|m}\|_2 \cdot \|\boldsymbol{\epsilon}_{m|n} - \bar{\boldsymbol{\epsilon}}_{m|n}\|_2} \tag{0.11}$$

where $\bar{\boldsymbol{\epsilon}}_{n|m} := \frac{1}{T}\boldsymbol{\epsilon}_{n|m}^\top \mathbf{1}_T \mathbf{1}_T$. Having computed the PC coefficients ρ_{nm}'s for all pairs $(n, m) \in \mathcal{N} \times \mathcal{N}$, determining whether node m is connected with n entails a hypothesis test: An edge between m and n is declared present if $|\rho_{nm}| \geq \tau > 0$, where τ trades off the relative true positive for the false positive decisions.

0.4.1   Non-linear partial correlations

The PC method assumes that $\hat{\mathbf{x}}_{n|m}$ is a linear function of $\{\mathbf{x}_i\}_{i \notin \{n,m\}}$. It is worth mentioning here that partial correlation coefficients are preferred over ordinary correlation coefficients because they can reveal direct (non-mediated) links rather than mediated ones e.g., [20]. Nonetheless, the dependence of $\mathbf{x}_n$ on $\{\mathbf{x}_i\}_{i \notin \{n,m\}}$ may be nonlinear. To accommodate that, an ℓ_2-norm based multi-kernel partial correlation-based approach has been studied in [19], which works as follows. Let vector $\boldsymbol{\chi}_{nm}[t]$ collect the data associated with time t and all buses except for $\{n, m\}$. By replacing $\boldsymbol{\chi}_{nm}[t]$ with its lifted image via a feature map $\boldsymbol{\phi}_{nm}[t]$, a nonlinear data generation model can be postulated as

$$x_n[t] = \langle \boldsymbol{\phi}_{nm}[t], \boldsymbol{\beta}_{nm} \rangle + e_n[t] \tag{0.12}$$

where $\boldsymbol{\beta}_{nm}$ is a parameter vector to learn and $e_n[t]$ captures modeling inaccuracies. Along the lines of ridge regression, $\boldsymbol{\beta}_{nm}$ can be estimated as

$$\hat{\boldsymbol{\beta}}_{nm} := \arg\min_{\boldsymbol{\beta}} \quad \frac{C}{N}\|\boldsymbol{\xi}_n\|_2^2 + \frac{1}{2}\|\boldsymbol{\beta}\|_2^2 \tag{0.13a}$$

$$\text{s.to} \quad \boldsymbol{\xi}_n = \mathbf{x}_n - \boldsymbol{\Phi}_{nm}^\top \boldsymbol{\beta} \tag{0.13b}$$

where matrix $\boldsymbol{\Phi}_{nm}$ has $\boldsymbol{\phi}_{nm}[t]$ for $t = 1, \ldots, T$ as its columns, and $C \geq 0$ is a given constant. Since $\boldsymbol{\phi}_{nm}[t]$ has high (potentially infinite) dimension, the dual of (0.13) which has only T variables will be used henceforth. Specifically, the dual of (0.13) can be succinctly written as [21]

$$\max_{\boldsymbol{\alpha}} -\mu \boldsymbol{\alpha}^\top \boldsymbol{\alpha} + 2\boldsymbol{\alpha}^\top \mathbf{x}_n - \boldsymbol{\alpha}^\top \mathbf{K}_{nm} \boldsymbol{\alpha} \tag{0.14}$$

where $\boldsymbol{\alpha} \in \mathbb{R}^T$ denotes the Lagrange multiplier associated with (0.13b); constant $\mu := N/(2C)$; and $\mathbf{K}_{nm} := \boldsymbol{\Phi}_{nm}^\top \boldsymbol{\Phi}_{nm}$. The maximizer of (0.14) can be found in closed form as $\hat{\boldsymbol{\alpha}}_n = (\mathbf{K}_{nm} + \mu \mathbf{I}_N)^{-1} \mathbf{x}_n$, and so $\hat{\mathbf{x}}_{n|m}$ is obtained as

$$\hat{\mathbf{x}}_{n|m} = \boldsymbol{\Phi}_{nm}^\top \hat{\boldsymbol{\beta}}_{nm} = \boldsymbol{\Phi}_{nm}^\top \boldsymbol{\Phi}_{nm} \hat{\boldsymbol{\alpha}}_n = \mathbf{K}_{nm} (\mathbf{K}_{nm} + \mu \mathbf{I}_N)^{-1} \mathbf{x}_n. \tag{0.15}$$

The latter entails computing inner products between high-dimensional feature vectors of the form $\langle \boldsymbol{\phi}_{nm}[t], \boldsymbol{\phi}_{nm}[t'] \rangle$. Fortunately, such a costly computation can be reduced by invoking the so-termed *kernel trick* [22], which allows computing the wanted inner products in (0.15) by evaluating a kernel function $\kappa(\boldsymbol{\phi}_{nm}[t], \boldsymbol{\phi}_{nm}[t'])$ for all pairs (t, t').

The accuracy of the estimates in (0.15) depends on the selected kernel function $\kappa(\cdot, \cdot)$ [22]. To choose a suitable kernel, multi-kernel ridge regression (MKRR) is invoked here [23], which seeks $\kappa(\cdot, \cdot)$ as a conic combination of user-defined kernel functions; that is, $\kappa(\cdot, \cdot) := \sum_{i=1}^M \theta_i \kappa_i(\cdot, \cdot)$. The coefficients $\{\theta_i \geq 0\}_{i=1}^M$ can be deciphered from data by solving [23]

$$\boldsymbol{\theta}^* := \arg \min_{\boldsymbol{\theta} \in \Theta_p} \max_{\boldsymbol{\alpha} \in \mathbb{R}^N} -\mu \boldsymbol{\alpha}^\top \boldsymbol{\alpha} + 2\boldsymbol{\alpha}^\top \mathbf{x}_n - \sum_{i=1}^M \theta_i \boldsymbol{\alpha}^\top \mathbf{K}_{nm}^i \boldsymbol{\alpha} \tag{0.16}$$

where the (t, t')-th entry of the kernel matrix is $\kappa_i(\boldsymbol{\chi}_{nm}[t], \boldsymbol{\chi}_{nm}[t'])$, and the constraint set Θ_p is defined as $\Theta_p := \{\boldsymbol{\theta} \in \mathbb{R}^M \mid \boldsymbol{\theta} \geq \mathbf{0}, \|\boldsymbol{\theta}\|_p \leq \Lambda\}$ with $p \geq 1$ and $\Lambda > 0$ is a pre-selected constant. Define the optimal value of the inner optimization (maximization over $\boldsymbol{\alpha}$) as $F(\boldsymbol{\theta})$. Then, the problem in (0.16) can be compactly expressed as

$$\boldsymbol{\theta}^* := \arg \min_{\boldsymbol{\theta} \in \Theta_p} F(\boldsymbol{\theta}) \tag{0.17}$$

where $F(\boldsymbol{\theta}) := \mathbf{x}_n^\top \left(\mu \mathbf{I} + \sum_{m=1}^M \theta_m \mathbf{K}_{nm}^i \right)^{-1} \mathbf{x}_n$. Upon obtaining $\boldsymbol{\theta}^*$ and $\mathbf{K}_{nm} = \sum_{m=1}^M \theta_m^* \mathbf{K}_{nm}^i$, we get from (0.15)

$$\hat{\mathbf{x}}_{nm} = \mathbf{K}_{nm} (\mu \mathbf{I} + \mathbf{K}_{nm})^{-1} \mathbf{x}_n. \tag{0.18}$$

This formulation generalizes the nonlinear estimator in [19] beyond $p = 2$. This generalization is well motivated since $\|\boldsymbol{\epsilon}_{n|m}\|_2$ can be reduced for $p \neq 2$; see [24].

Returning to the grid topology inference task at hand, let vector $\mathbf{x}_n$ collect the voltage angles $\phi_n[t]$ for $t = 1, \ldots, T$. Giver data $\{\mathbf{x}_n\}_{n \in \mathcal{N}}$, the goal is to infer the connectivity between buses. For each bus pair (n, m), we follow the

ensuing procedure. Having selected candidate kernel functions $\{\kappa_i\}_{i=1}^M$, one first forms the kernel matrices $\{\mathbf{K}_{nm}^i\}_{i=1}^M$, and learns the best kernel combination $\mathbf{K}_{nm} = \sum_{i=1}^M \theta_i^* \mathbf{K}_{nm}^i$ by solving (0.17). The next step is to obtain the nonlinear nonlinear estimators $\hat{\mathbf{x}}_{n|m} = \mathbf{K}_{nm}(\mu\mathbf{I} + \mathbf{K}_{nm})^{-1}\mathbf{x}_n$, and likewise for $\hat{\mathbf{x}}_{m|n}$. The PC coefficient ρ_{nm} is found from (0.11), and an edge between buses n and m is claimed to be present if $|\rho_{nm}| > \tau$. This pairwise hypotheses test is repeated for all pairs (n, m). We next present an algorithm for solving (0.17).

0.4.2 A Frank-Wolfe-based solver

Though the ℓ_p-norm based MKRR (0.17) can lead to improved estimation accuracy, solving it may not be easy unless $p = 1$ or 2; see [24], [25]. For this reason, we put forward an efficient solver by leveraging the Frank-Wolfe or conditional gradient method [26]. The latter algorithm targets the convex problem

$$\mathbf{y}^* \in \arg\min_{\mathbf{y}\in\mathcal{Y}} \; F(\mathbf{y}) \tag{0.19}$$

where F is differentiable and $\mathcal{Y}$ is compact. The Frank-Wolfe solver starts with an arbitrary point $\mathbf{y}^0$, and iterates between the updates [27]

$$\mathbf{s}^k \in \arg\min_{\mathbf{s}\in\mathcal{Y}} \; \mathbf{s}^\top\nabla F(\mathbf{y}^k) \tag{0.20a}$$

$$\mathbf{y}^{k+1} := \mathbf{y}^k + \eta_k(\mathbf{s}^k - \mathbf{y}^k) \tag{0.20b}$$

where $\eta_k = 2/(k+2)$. The iterates $\{\mathbf{y}^k\}$ remain feasible for all k because $\eta^0 = 1$, $\mathbf{y}^1 = \mathbf{s}^0 \in \mathcal{Y}$, and $\mathbf{s}^k \in \mathcal{Y}$; see [26].

Since the cost in (0.17) is convex and differentiable, and set Θ_p is convex and compact, problem (0.17) complies with the form in (0.19). The negative gradient of $F(\boldsymbol{\theta})$ can be computed as

$$-\nabla F(\boldsymbol{\theta}) = [\hat{\boldsymbol{\alpha}}^\top\mathbf{K}_{nm}^1\hat{\boldsymbol{\alpha}} \; \cdots \; \hat{\boldsymbol{\alpha}}^\top\mathbf{K}_{nm}^M\hat{\boldsymbol{\alpha}}]^\top \tag{0.21}$$

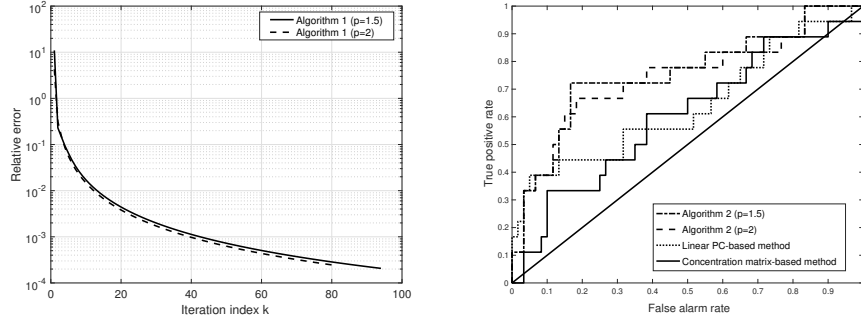where $\hat{\boldsymbol{\alpha}}$ depends on $\boldsymbol{\theta}$ through

$$\hat{\boldsymbol{\alpha}} = \left(\mu\mathbf{I} + \sum_{i=1}^M \theta_i\mathbf{K}_{mn}^i\right)^{-1}\mathbf{x}_n. \tag{0.22}$$

Since $\mathbf{K}_{nm}^i \succeq \mathbf{0}$ for all i, it follows that $\mathbf{g}^k := -\nabla F(\boldsymbol{\theta}^k) \geq \mathbf{0}$ for all $\boldsymbol{\theta}^k \in \Theta_p$. Applying (0.20a) to (0.17) yields

$$\mathbf{s}^k \in \arg\max_{\mathbf{s}\in\Theta_p} \; \mathbf{s}^\top\mathbf{g}^k. \tag{0.23}$$

By Hölder's inequality and for all $\mathbf{s} \in \Theta_p$, it holds that $\mathbf{s}^\top\mathbf{g}^k \leq \|\mathbf{s}\|_p\|\mathbf{g}^k\|_q \leq \Lambda\|\mathbf{g}^k\|_q$ for any (p, q) with $1/p + 1/q = 1$. Since $\mathbf{g}^k = -\nabla F(\boldsymbol{\theta}^k) \geq \mathbf{0}$, it can be deduced that the solution to (0.23) satisfies the previous inequalities with equalities, and so

$$s_i^k = \Lambda\frac{(g_i^k)^{q-1}}{\|\mathbf{g}^k\|_q^{q-1}}, \quad i = 1,\ldots,M. \tag{0.24}$$

Figure 0.2 Topology recovery for the IEEE 14-bus system. *Left*: Convergence speed for the devised Frank-Wolfe iterations; *Right*: ROC curves for line detection.

The Frank-Wolfe solver of (0.17) is summarized next: The algorithm is initialized at $\boldsymbol{\theta}^0 = \mathbf{0}$. At iteration k, $\hat{\boldsymbol{\alpha}}^k$ is found from (0.22); the negative gradient $\mathbf{g}^k$ is updated from (0.21); the direction $\mathbf{s}^k$ is found from (0.24); and the sought vector is updated as $\boldsymbol{\theta}^{k+1} = \boldsymbol{\theta}^k + \eta_k(\mathbf{s}^k - \boldsymbol{\theta}^k)$. The algorithm converges to $\boldsymbol{\theta}^*$ at sublinear rate, that is $F(\boldsymbol{\theta}^k) - F(\boldsymbol{\theta}^*) \leq \mathcal{O}(1/k)$; see [26].

0.4.3 Numerical tests on topology detection using partial correlations

The performance of the proposed method was evaluated based on voltage angles from the IEEE 14-bus benchmark system, using real load data from [28]. Specifically, the first ten-day loads of zones $1-14$ were normalized to match the scale of active demands in the benchmark, and then corrupted by noise. Voltage angle measurements $\phi_n[t]$ across $T = 240$ times were found by solving the AC power flow equations. We employed a total of $M = 20$ kernels to form the dictionary, which consists of 10 polynomial kernels of orders varying by 1 from 1 to 10, as well as 10 Gaussian kernels with variances distributed uniformly from 0.5 to 5. The regularization coefficients in (0.16) were set as $\mu = 1$ and $\Lambda = 3$.

The first test assesses the convergence and computational performance of the Frank-Wolfe solver. To serve as a benchmark, problem (0.17) is first reformulated as a semidefinite program (SDP) and solved by SeDuMi [29]; see [25]. The left panel of Fig 0.2 depicts the evolution of the relative error $(F(\boldsymbol{\theta}^k) - F(\boldsymbol{\theta}^*))/F(\boldsymbol{\theta}^*)$ of (0.17) with $p = 1.5$ and $p = 2$, where $\{\mathbf{x}_n\}_{n=1}^2$ and $\{\mathbf{x}_n\}_{n=5}^{14}$ were used to predict $\mathbf{x}_3$. The solver converged roughly at rate $\mathcal{O}(1/k)$.

We next tested the topology recovery performance of the proposed scheme for $p = 1.5$ and $p = 2$, against the linear PC- and concentration matrix-based methods [30]. The right panel of Fig. 0.2 depicts the obtained empirical receiver operating characteristics (ROC). For our scheme and its linear PC counterpart, the ROC curves were obtained using $|\rho_{nm}|$'s as test statistics. For the concentration matrix-based method, entries of the negative concentration matrix were used as test statistics. The area under the curve for our scheme with $p = 1.5$

and $p = 2$, the linear PC-, and the concentration matrix-based methods were 0.755, 0.743, 0.646, and 0.604, accordingly. The results demonstrate the improved recovery performance of the novel scheme, and the advantage of selecting $p \neq 2$.

0.5   Grid Probing for Topology Learning

Existing topology processing schemes rely on passively collected smart meter data. We next put forth an active data acquisition approach to topology learning. The idea is to leverage the actuation and sensing capabilities of smart inverters. An inverter can be commanded to shed solar generation, (dis)-charge a battery, or change its power factor within milliseconds. The distribution feeder as an electric circuit responds within a second and reaches a different steady-state voltage profile. This enables a new data collection paradigm, where the operator purposefully probes a grid by changing inverter injections and measuring the circuit response to identify the grid topology. Rather than processing smart meter data on a 15-min basis, probing actively senses voltages on a per-second basis.

The buses hosting controllable inverters are collected in $\mathcal{P} \subseteq \mathcal{N}$ with $P := |\mathcal{P}|$. Consider the probing action at time t. Each bus $m \in \mathcal{P}$ perturbs its active injection by $\delta_m(t)$. Vector $\boldsymbol{\delta}(t)$ collects all inverter perturbations at time t. The incurred perturbation in voltage magnitudes is expressed from (0.4) as

$$\tilde{\mathbf{v}}(t) = \mathbf{R}\mathbf{I}_{\mathcal{P}}\boldsymbol{\delta}(t) + \boldsymbol{\epsilon}(t). \tag{0.25}$$

where the $N \times P$ matrix $\mathbf{I}_{\mathcal{P}}$ collects the canonical vectors associated with $\mathcal{P}$ and basically selects the rows of $\mathbf{R}$ related to inverters. Vector $\boldsymbol{\epsilon}(t)$ captures measurement noise, modeling errors, and voltage deviations attributed to possible load variations during probing.

The grid is perturbed over T probing periods. Stacking the probing actions $\{\boldsymbol{\delta}(t)\}_{t=1}^T$, the measured voltage deviations $\{\tilde{\mathbf{v}}(t)\}_{t=1}^T$, and the error terms $\{\boldsymbol{\epsilon}(t)\}_{t=1}^T$ as columns of matrices $\boldsymbol{\Delta}$, $\tilde{\mathbf{V}}$, and $\mathbf{E}$ accordingly, yields

$$\tilde{\mathbf{V}} = \mathbf{R}\mathbf{I}_{\mathcal{P}}\boldsymbol{\Delta} + \mathbf{E}. \tag{0.26}$$

Define the *weighted Laplacian* matrix $\tilde{\boldsymbol{\Theta}} = \tilde{\mathbf{A}}^\top \mathrm{dg}^{-1}(\mathbf{r})\tilde{\mathbf{A}}$ and the *reduced weighted Laplacian* $\boldsymbol{\Theta}$ as

$$\boldsymbol{\Theta} := \mathbf{R}^{-1} = (\mathbf{A}^\top \mathrm{dg}(\mathbf{r})\mathbf{A})^{-1}. \tag{0.27}$$

The grid topology is equivalently captured by the reduced weighted Laplacian and the non-reduced Laplacian matrix $\tilde{\boldsymbol{\Theta}} = \tilde{\mathbf{A}}^\top \mathrm{dg}^{-1}(\mathbf{r})\tilde{\mathbf{A}}$. To see this, note that the two matrices are related through the linear mapping $\boldsymbol{\Phi} : \mathbb{S}^N \to \mathbb{S}^{N+1}$ as

$$\tilde{\boldsymbol{\Theta}} = \boldsymbol{\Phi}(\boldsymbol{\Theta}) := \begin{bmatrix} \mathbf{1}^\top \boldsymbol{\Theta}\mathbf{1} & -\mathbf{1}^\top \boldsymbol{\Theta} \\ -\boldsymbol{\Theta}\mathbf{1} & \boldsymbol{\Theta} \end{bmatrix}. \tag{0.28}$$

Topology identification can be now posed as the system identification task of finding $\boldsymbol{\Theta}$ given $(\tilde{\mathbf{V}}, \boldsymbol{\Delta})$ from (0.26). This is a major advantage over the blind

schemes of Section 0.3 and [2, 1]. Albeit we have so far considered perturbing active power injections, the developments carry over to reactive ones too.

0.6   Identifiability Analysis of Grid Probing

This section studies whether the actual topology can be uniquely recovered by probing the buses in $\mathcal{P}$ [31, 32]. We first review some graph theory background. Let $\mathcal{G} = (\mathcal{N}, \mathcal{L})$ be an undirected tree graph, where $\mathcal{N}$ is the set of nodes and $\mathcal{L}$ the set of edges $\mathcal{L} := \{(m, n) : m, n \in \mathcal{N}\}$. A tree is termed rooted if one of its nodes, henceforth indexed by 0, is designated as the root.

In a tree graph, a *path* is the unique sequence of edges connecting two nodes. The set of nodes adjacent to the edges forming the path between nodes n and m will be denoted by $\mathcal{P}_{n,m}$. The nodes belonging to $\mathcal{A}_m := \mathcal{P}_{0,m}$ are termed the *ancestors* of node m. If $n \in \mathcal{A}_m$, then m is a *descendant* of node n. Reversely, if $n \in \mathcal{A}_m$, m is a *descendant* of node n. The descendants of node m comprise the set $\mathcal{D}_m$. By convention, $m \in \mathcal{A}_m$ and $m \in \mathcal{D}_m$. If $n \in \mathcal{A}_m$ and $(m, n) \in \mathcal{E}$, node n is the *parent* of m. A node without descendants is called a *leaf* or *terminal* node. Leaf nodes are collected in the set $\mathcal{F}$, while non-leaf nodes will be termed *internal* nodes; see Fig. 0.3. The *depth* d_m of node m is defined as the number of its ancestors, i.e., $d_m := |\mathcal{A}_m|$. The depth of the entire tree is $d_{\mathcal{G}} := \max_{m \in \mathcal{N}} d_m$. If $n \in \mathcal{A}_m$ and $d_n = k$, node n is the unique k-depth ancestor of node m and will be denoted by α_m^k for $k = 0, \ldots, d_m$. Let also $\mathcal{T}_m^k$ denote the subset of the nodes belonging to the subtree of $\mathcal{G}$ rooted at the k-depth node m and containing all the descendants of m. Finally, the *k-th level set*[1] of node m is defined as [31, 32]

$$\mathcal{N}_m^k := \begin{cases} \mathcal{D}_{\alpha_m^k} \setminus \mathcal{D}_{\alpha_m^{k+1}} & , \ k = 0, \ldots, d_m - 1 \\ \mathcal{D}_m & , \ k = d_m. \end{cases} \qquad (0.29)$$

The concept of the level set is illustrated in Figure 0.3. In essence, the level set $\mathcal{N}_m^k$ consists of node α_m^k and all the subtrees rooted at α_m^k excluding the one containing node m.

The entries of the reduced Laplacian inverses $(\mathbf{R}, \mathbf{X})$ can also be defined exploiting graphical quantities. If $R_{m,n}$ is the (m, n)-th entry of $\mathbf{R}$, then [14]

$$R_{mn} = \sum_{\substack{\ell = (c,d) \in \mathcal{L} \\ c,d \in \mathcal{A}_m \cap \mathcal{A}_n}} r_\ell. \qquad (0.30)$$

The entry $R_{m,n}$ can be equivalently interpreted as the voltage drop between the substation and bus m when a unitary active power is injected as bus n and the remaining buses are unloaded. Through this interpretation, the entries of $\mathbf{R}$ relate to the levels sets in $\mathcal{G}$ as follows.

LEMMA 1 ([31])   *Let m, n, s be nodes in a radial grid.*

---

[1] The notion of level sets has been used in [33] to derive a meter placement strategy for detecting which switches are energized.
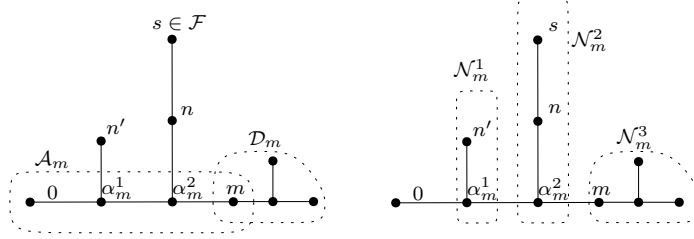
Figure 0.3 Node s is a leaf node, while m is an internal node of the left graph. The ancestor ($\mathcal{A}_m$) and descendant ($\mathcal{D}_m$) sets for node m are also shown. The level set $\mathcal{N}_m^1$ consists of α_m^1 and the subtrees rooted at α_m^1 excluding the subtree containing m.

(i) if $m \in \mathcal{F}$, then $R_{mm} > R_{nm}$ for all $n \neq m$;

(ii) $n, s \in \mathcal{N}_m^k$ for a k if and only if $R_{nm} = R_{sm}$; and

(iii) if $n' \in \mathcal{N}_m^{k-1}$, $s \in \mathcal{N}_m^k$, then $R_{sm} = R_{mn'} + r_{\alpha_m^{k-1}, \alpha_m^k}$.

We next provide conditions on inverter placement under which the topology of a radial grid can be identified. We consider the case where voltages are collected at all buses (complete data), or only at probed buses (partial data).

0.6.1   Topology identifiability with complete voltage data

As customary in identifiability analysis, probing data are considered noiseless ($\mathbf{E} = \mathbf{0}$), so that (0.26) yields

$$\tilde{\mathbf{V}} = \mathbf{R}_{\mathcal{P}} \boldsymbol{\Delta} \tag{0.31}$$

where $\mathbf{R}_{\mathcal{P}} := \mathbf{R}\mathbf{I}_{\mathcal{P}}$. It is convenient to design $\boldsymbol{\Delta} \in \mathbb{R}^{C \times T}$ to be full row-rank. For example, one can set $T = C$ and $\boldsymbol{\Delta}_1 := \mathrm{dg}(\{\delta_m\})$ with $\delta_m \neq 0$ for all $m \in \mathcal{C}$. A diagonal $\boldsymbol{\Delta}$ requires simple synchronization, since each $m \in \mathcal{C}$ probes the grid at different time slots. Another practical choice is $\boldsymbol{\Delta}_2 := \mathrm{dg}(\{\delta_m\}) \otimes [+1 \ -1]$, where $\otimes$ is the Kronecker product. In this case, $T = 2C$ and each inverter induces two probing actions: It first drops its generation to zero, yielding a perturbation of δ_m, and then resumes generation at its nominal value, thus incurring a perturbation of $-\delta_m$. The process is repeated over all C inverters.

The task of topology identification can be split into three stages:

s1) Finding $\mathbf{R}_{\mathcal{P}}$ from (0.31). If $\boldsymbol{\Delta}$ is full row-rank, then matrix $\mathbf{R}_{\mathcal{P}}$ can be uniquely recovered as $\tilde{\mathbf{V}}\boldsymbol{\Delta}^+$, where $\boldsymbol{\Delta}^+$ is the pseudo-inverse of $\boldsymbol{\Delta}$. Under this setup, probing for $T = P$ times suffices to find $\mathbf{R}_{\mathcal{P}}$.

s2) Recovering the level sets for all buses in $\mathcal{P}$. Let $\mathbf{r}_m$ be the m-th column of $\mathbf{R}$. Using Lemma 1, we can recover the level sets for each bus $m \in \mathcal{P}$ by grouping together indices associated with the same entries of $\mathbf{r}_m$; see Lemma 1-(ii). The depth of each level set can be found by ranking the unique values of $\mathbf{r}_m$ in increasing order; see Lemma 1–(iii).

s3) Recover the grid topology and resistances given the level sets for all $m \in \mathcal{P}$.
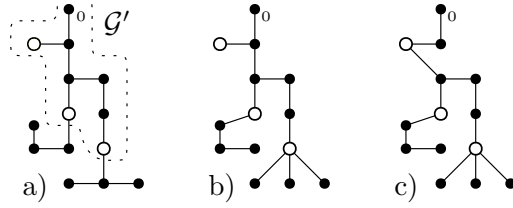
Figure 0.4 The white (black) nodes represent probing (non-probing) nodes. The dashed portion of the network represent graph $\mathcal{G}'$. Panel a) represents the actual network. Panel b) reports one of the networks that can be identified when the three white nodes are chosen as probing buses. Note that the topology is correctly identified down to the probing buses. Panel c) shows one of the erroneous topologies recovered if not all leaf nodes of $\mathcal{G}'$ are probed.

The true $\boldsymbol{\Theta}$ may not be identifiable from (0.31). A sufficient condition guaranteeing identifiability is provided next.

THEOREM 1   *Given probing data* $(\tilde{\mathbf{V}}, \boldsymbol{\Delta})$ *where rank*$(\boldsymbol{\Delta}) = P$, *the resistive network topology is identifiable if the grid is probed at all leaf nodes, that is* $\mathcal{F} \subseteq \mathcal{P}$, *and voltage data are collected at all nodes.*

Theorem 1 establishes that the topology is identifiable if the grid is probed at all leaf nodes and voltages are collected at all nodes. Under this setup, one needs at least $T = |\mathcal{F}|$, which can be significantly smaller than N. When not all leaf nodes are probed, a portion of the network may still be identifiable. Let $\mathcal{P}_F$ be the set collecting the probing buses who have only regular nodes as descendants, i.e., $n \in \mathcal{P}_F$ if $\mathcal{D}_N \cap \mathcal{P} = \emptyset$. Consider graph $\mathcal{G}'$ obtained from $\mathcal{G}$ by removing the descendants of buses in $\mathcal{P}_F$. Theorem 1 ensures that the topology of $\mathcal{G}'$ can be recovered if $\mathcal{G}'$ is a tree whose leaf nodes are exactly nodes in $\mathcal{P}_F$. That is, $\mathcal{G}$ can be reconstructed up to the descendants of nodes $\mathcal{P}_F$, see Fig. 0.4.

0.6.2   Topology identifiability with partial voltage data

Section 0.6.1 assumed that voltages are collected at all buses. This may be unrealistic in grids where the operator can only access the probed buses of $\mathcal{P}$. This section considers probing under the next more realistic setup.
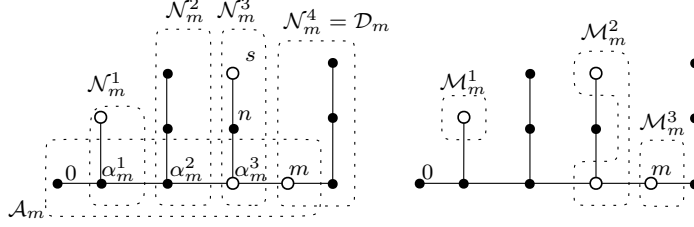
ASSUMPTION 2   *Voltage differences are metered only in* $\mathcal{P}$.

Under Assumption 2, the probing model (0.26) becomes

$$\tilde{\mathbf{V}} = \mathbf{R}_{\mathcal{P}\mathcal{P}}\boldsymbol{\Delta} \tag{0.32}$$

where now $\tilde{\mathbf{V}}$ is of dimension $P \times T$ and $\mathbf{R}_{\mathcal{P}\mathcal{P}} := \mathbf{I}_{\mathcal{P}}^{\top}\mathbf{R}\mathbf{I}_{\mathcal{P}}$ is obtained from $\mathbf{R}$ upon maintaining only the rows and columns in $\mathcal{P}$. Similar to (0.31), $\mathbf{R}_{\mathcal{P}\mathcal{P}}$ is identifiable if $\boldsymbol{\Delta}$ is full row-rank. This is the equivalent of stage *s1)* in Section 0.5 under the partial data setup.

Figure 0.5 White nodes represent metered nodes. The level sets and the metered level sets of node m are reported in the left panel and in the right panel, respectively.

Towards the equivalent of stage *s2)*, since column $\mathbf{r}_m$ is partially observed, only the *metered level sets* of node $m \in \mathcal{P}$ defined as [32]

$$\mathcal{M}_m^k(\mathbf{w}) = \mathcal{N}_m^h(\mathbf{w}) \cap \mathcal{O}, \tag{0.33}$$

can be found, where $\mathcal{N}_m^h(\mathbf{w})$ is the k-th level set having at least one observed node, see Fig. 0.5. The metered level sets for node m can be obtained by grouping the indices associated with the same values in the observed subvector of $\mathbf{r}_m$. Albeit the topology cannot be fully recovered based on $\mathcal{M}_m^k$'s, one can recover a *reduced grid* relying on the concept of internally identifiable nodes; see Fig. 0.6. The set $\mathcal{I} \subset \mathcal{N}$ of *internally identifiable* nodes consists of all buses in $\mathcal{G}$ having at least two children with each of one of them being the ancestor of a probing bus.

The reduced grid induced by $\mathcal{P}$ is defined as the graph $\mathcal{G}^r := (\mathcal{N}^r, \mathcal{L}^r)$ with

- node set $\mathcal{N}^r := \mathcal{P} \cup \mathcal{I}$;
- $\ell = (m, n) \in \mathcal{L}^r$ if $m, n \in \mathcal{N}^r$ and all other nodes on the path from m to n in $\mathcal{G}$ do not belong to $\mathcal{N}^r$; and
- line $\ell = (m, n) \in \mathcal{L}^r$ resistance equals the effective resistance between m and n in $\mathcal{G}$, i.e., the sum of resistances across the $m - n$ path [34].

Let $\mathbf{R}^r$ be the inverse reduced Laplacian associated with $\mathcal{G}^r$. From the properties of effective resistances, it holds [34]

$$\mathbf{R}_{\mathcal{PP}}^r = \mathbf{R}_{\mathcal{PP}}. \tag{0.34}$$

Heed the grid $\mathcal{G}^r$ is not the only electric grid sharing $\mathbf{R}_{\mathcal{PP}}$ as the top-left block of its $\mathbf{R}$ matrix with $\mathcal{G}$: the (meshed) Kron reduction of $\mathcal{G}$ given $\mathcal{P}$; and even grids having extra nodes relative to $\mathcal{N}$ can yield the same $\mathbf{R}_{\mathcal{PP}}$; see Fig. 0.6. However, $\mathcal{G}^r$ features desirable properties: i) it is radial; ii) it satisfies (0.34) with the minimal number of nodes; and iii) its resistances correspond to the effective resistances of $\mathcal{G}$. Furthermore, $\mathcal{G}^r$ conveys all the information needed to solve an optimal power flow task [35]. The next result provides a sufficient condition for identifying $\mathcal{G}^r$.

THEOREM 2 *Let the tree $\mathcal{G} = (\mathcal{N}, \mathcal{L})$ represent a distribution grid, and let Assumption 2 hold. Given probing data $(\tilde{\mathbf{V}}, \boldsymbol{\Delta})$, where $rank(\boldsymbol{\Delta}) = P$, the resistive*
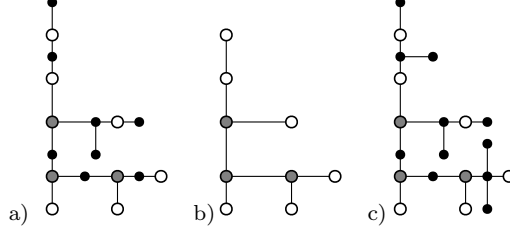
Figure 0.6 *a)* the original feeder; *b)* its reduced equivalent; and *c)* another feeder with the same $\mathbf{R}_{\mathcal{PP}}$. White nodes are probed; black and gray are not. Gray nodes are internally identifiable nodes comprising $\mathcal{I}$.

network topology of its reduced graph $\mathcal{G}^r = (\mathcal{N}^r, \mathcal{L}^r)$ *is identifiable if the grid is probed at all leaf nodes, that is* $\mathcal{F} \subseteq \mathcal{P}$.

0.7   Graph Algorithms for Topology Identification using Probing

This section presents graph algorithms for identifying grid topologies using probing data under the following assumption.

ASSUMPTION 3   *All leaf nodes are probed, that is* $\mathcal{F} \subseteq \mathcal{P}$.

Note that although Assumption 3 ensures topology identifiability, it does not provide a solution for *s3)*, which will be devised next.

0.7.1   Topology identification with complete voltage data

A recursive algorithm for topology identification is presented next [32]. The input to the recursion is a depth k and a maximal subset of probing nodes $\mathcal{P}_n^k$ having the *same* $(k-1)$-depth and k-depth ancestors and the level sets $\mathcal{N}_m^k$ for all $m \in \mathcal{P}_n^k$. The $(k-1)$-depth ancestor α_n^{k-1} is known. The k-depth ancestor n is known to exist, yet it is unknown for now. The recursion proceeds in three steps.

The *first step* finds n as the unique intersection of the sets $\mathcal{N}_m^k$ for all $m \in \mathcal{P}_n^k$.

At the *second step*, node n is connected to node α_n^{k-1}. Since $n = \alpha_m^k \in \mathcal{N}_m^k$ and $\alpha_n^{k-1} = \alpha_m^{k-1} \in \mathcal{N}_m^{k-1}$, the resistance of line (n, α_n^{k-1}) is given by

$$r_{(\alpha_n^{k-1}, n)} = r_{(\alpha_m^{k-1}, \alpha_m^k)} = R_{\alpha_m^k m} - R_{\alpha_m^{k-1} m} \tag{0.35}$$

for any $m \in \mathcal{P}_n^k$; see Lemma 1-(iii).

The *third step* partitions $\mathcal{P}_n^k \setminus \{n\}$ into subsets of buses sharing the same $(k+1)$-depth ancestor. The buses forming one of these partitions $\mathcal{P}_s^{k+1}$ have the same k-depth and $(k+1)$-depth ancestors. Node n was found to be the k-depth ancestor. The $(k+1)$-depth ancestor is known to exist and is assigned the symbol s. The value of s is found by invoking the recursion with new inputs the depth

$(k+1)$, the set of buses $\mathcal{P}_s^{k+1}$ along with their $(k+1)$-depth level sets, and their common k-depth ancestor (node n).

Algorithm 1 Topology Identification with Complete Data

Require: $\mathcal{N}$, $\{\mathcal{N}_m^k\}_{k=0}^{d_m}$ for all $m \in \mathcal{P}$.
 1: Run `Root&Branch`$(\mathcal{P}, \emptyset, 0)$.
Ensure: Radial grid and line resistances over $\mathcal{N}$.
Function `Root&Branch`$(\mathcal{P}_n^k, \alpha_n^{k-1}, k)$
 1: Identify the node n as the common k-depth ancestor for all buses in $\mathcal{P}_n^k$.
 2: **if** $k > 0$, **then**
 3:     Connect node n to α_n^{k-1} with the resistance of (0.35).
 4: **end if**
 5: **if** $\mathcal{P}_n^k \setminus \{n\} \neq \emptyset$, **then**
 6:     Partition $\mathcal{P}_n^k \setminus \{n\}$ into groups of buses $\{\mathcal{P}_s^{k+1}\}$ having identical k-depth level sets.
 7:     Run `Root&Branch`$(\mathcal{P}_s^{k+1}, n, k+1)$ for all s.
 8: **end if**

To *initialize* the recursion, set $\mathcal{P}_n^0 = \mathcal{P}$ since every probing bus has the substation as 0-depth ancestor. At $k = 0$, the second step is skipped as the substation does not have any ancestors to connect. The recursion *terminates* when $\mathcal{P}_n^k$ is a singleton $\{m\}$. In this case, the first step identifies m as node n; the second step links m to its known ancestor α_m^{k-1}; and the third step has no partition to accomplish. The recursion is tabulated as Alg. 1.

0.7.2   Topology identification with partial voltage data

A three-step recursion operating on metered rather than ordinary level sets and aiming at reconstructing the reduced grid is detailed next [32]. Suppose we are given the set of probing nodes $\mathcal{P}_n^k$ having the same $(k-1)$-depth and k-depth ancestors (known and unknown, respectively), along with their k-depth metered level sets.

At the *first step*, if there exists a node $m \in \mathcal{P}_n^k$ such that $\mathcal{M}_m^k = \mathcal{P}_n^k$, then the k-depth ancestor n is set as m. Otherwise, a non-probing node is added and assigned to be the k-depth ancestor.

At the *second step*, node $n = \alpha_m^k$ is connected to node $\alpha_n^{k-1} = \alpha_m^{k-1}$. The line resistance can be found through the modified version of (0.35)

$$r_{(\alpha_m^{k-1}, \alpha_m^k)} = R_{\alpha_m^k m} - R_{\alpha_m^{k-1} m} = R_{s'm} - R_{sm}. \tag{0.36}$$

At the *third step*, the set $\mathcal{P}_n^k \setminus \{n\}$ is partitioned into subsets of buses having the same $(k+1)$-depth ancestor, by comparing their k-depth metered level sets.

The recursion is tabulated as Alg. 2. It is initialized at $k = 1$, since the substation is not probed and $\mathcal{M}_m^0$ does not exist; and is terminated as in Section 0.6.

Algorithm 2 Topology Recovery with Partial Data

Require: $\mathcal{M}$, $\{\mathcal{M}_m^k\}_{k=1}^{d_m}$ for all $m \in \mathcal{P}$.
 1: Run `Root&Branch-P`$(\mathcal{P}, \emptyset, 1)$.
Ensure: Reduced grid $\mathcal{G}^r$ and resistances over $\mathcal{L}^r$.

Function `Root&Branch-P`$(\mathcal{P}_n^k, \alpha_n^{k-1}, k)$
 1: **if** $\exists$ node n such that $\mathcal{M}_n^k = \mathcal{P}_n^k$, **then**
 2:     Set n as the parent node of subtree $\mathcal{T}_n^k$.
 3: **else**
 4:     Add node $n \in \mathcal{I}$ and set it as the root of $\mathcal{T}_n^k$.
 5: **end if**
 6: **if** $k > 1$, **then**
 7:     Connect n to α_n^{k-1} via a line with resistance (0.36).
 8: **end if**
 9: **if** $\mathcal{P}_n^k \setminus \{n\} \neq \emptyset$, **then**
10:     Partition $\mathcal{P}_n^k \setminus \{n\}$ into groups of buses $\{\mathcal{P}_s^{k+1}\}$ having identical k-depth metered level sets.
11:     Run `Root&Branch-P`$(\mathcal{P}_s^{k+1}, n, k+1)$ for all s.
12: **end if**

0.7.3   Graph algorithms operating under noisy data

When measurements are corrupted by noise and loads are time varying, a least-squares estimate of $\mathbf{R}_{\mathcal{P}}$ can be found as

$$\hat{\mathbf{R}}_{\mathcal{P}} := \arg\min_{\boldsymbol{\Theta}} \ \|\tilde{\mathbf{V}} - \boldsymbol{\Theta}\boldsymbol{\Delta}\|_F^2 = \tilde{\mathbf{V}}\boldsymbol{\Delta}^+. \tag{0.37}$$

To facilitate its statistical characterization and implementation, assume the probing protocol $\boldsymbol{\Delta}_2$. The m-th column of $\mathbf{R}_{\mathcal{P}}$ can be found as the scaled sample mean of voltage differences obtained when node m was probed

$$\hat{\mathbf{r}}_m = \sum_{t \in \mathcal{T}_m} \frac{1}{(-1)^{t-1}\delta_m T_m}, \quad \mathcal{T}_m := \left\{\sum_{\tau=1}^{m-1} T_\tau + 1, \dots, \sum_{\tau=1}^{m} T_\tau\right\} \tilde{\mathbf{v}}(t). \tag{0.38}$$

Let Assumption 1 hold and, for simplicity, let $\{\tilde{\mathbf{p}}(t), \tilde{\mathbf{q}}(t), \mathbf{n}(t)\}$ have diagonal covariances $\sigma_p^2 \mathbf{I}$, $\sigma_q^2 \mathbf{I}$, and $\sigma_n^2 \mathbf{I}$. Hence, the error vector $\boldsymbol{\epsilon}(t)$ is zero-mean with covariance $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} := \sigma_p^2 \mathbf{R}^2 + \sigma_q^2 \mathbf{X}^2 + \sigma_n^2 \mathbf{I}$ and the estimate $\hat{\mathbf{r}}_m$ can be approximated as zero-mean Gaussian with covariance $\frac{1}{\delta_m^2 T_m} \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$. By increasing T_m and/or δ_m, $\hat{\mathbf{r}}_m$ goes arbitrarily close to $\mathbf{r}_m$ and their distance can be bounded probabilistically using $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$. Note however, that $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ depends on the unknown $(\mathbf{R}, \mathbf{X})$. To resolve this issue, suppose the spectral radii $\rho(\mathbf{R})$ and $\rho(\mathbf{X})$, and the variances $(\sigma_p^2, \sigma_q^2, \sigma_w^2)$ are known; see [35] for upper bounds. Then, it holds that $\rho(\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}) \leq \sigma^2$, where $\sigma^2 := \sigma_p^2 \rho^2(\mathbf{R}) + \sigma_q^2 \rho^2(\mathbf{X}) + \sigma_n^2$. The standard Gaussian concentration inequality bounds the deviation of the n-th entry of $\hat{\mathbf{r}}_m$ from its actual value as $\Pr\left(|\hat{R}_{nm} - R_{nm}| \geq \frac{4\sigma}{\delta_m \sqrt{T_m}}\right) \leq \pi_0 := 6 \cdot 10^{-5}$.

For *s2)*, no two noisy entries of $\hat{\mathbf{r}}_m$ will be identical almost surely. The entries will be concentrated around their actual values. To identify groups of similar values, sort the entries of $\hat{\mathbf{r}}_m$ in increasing order, and take the differences of the successive sorted entries. Lemma 1-(iii) guarantees that the minimum difference between the entries of $\mathbf{r}_m$ is larger or equal to the smallest line resistance $r_{\min}$. Hence, if all estimates were confined within $|\hat{R}_{nm} - R_{nm}| \leq r_{\min}/4$, a difference of sorted $\hat{R}_{nm}$'s larger than $r_{\min}/2$ would pinpoint the boundary between two bus groups. In practice, if the operator knows $r_{\min}$ and selects (T_m, δ_m) so that

$$\delta_m \sqrt{T_m} \geq 16 \,\sigma/r_{\min} \tag{0.39}$$

then $|\hat{R}_{nm} - R_{nm}| \leq r_{\min}/4$ will be satisfied with probability larger than 99.95%. Taking the union bound, the probability of recovering all level sets is larger than $1 - N^2 \pi_0$. The argument carries over to $\mathbf{R}_{\mathcal{PP}}$ under the partial data setup.

0.8 Topology Identification using Probing via Convex Relaxation

Section 0.7 first estimated a part of $\mathbf{R}$, and then run a graph algorithm to identify the grid topology. Because the latter algorithm may become impractical for low signal-to-noise ratios, this section estimates $\boldsymbol{\Theta}$ directly from probing data. From (0.3), it follows that $\boldsymbol{\Theta} \succ \mathbf{0}$, since $\mathbf{A}$ is non-singular [36]. The off-diagonal terms of $\boldsymbol{\Theta}$ are non-positive, while the diagonal ones are positive. Also, since the first column of $\tilde{\boldsymbol{\Theta}}$ has non-positive off-diagonal entries and $\tilde{\boldsymbol{\Theta}}\mathbf{1} = \mathbf{0}$, it also follows that $\boldsymbol{\Theta}\mathbf{1} \geq \mathbf{0}$. Finally, the grid operator may know that two specific buses are definitely connected, e.g., through flow sensors or line status indicators. To model known line statuses, introduce $\tilde{\boldsymbol{\Gamma}} \in \mathbb{S}^{N+1}$ with $\tilde{\Gamma}_{mn} = 0$ if line (m,n) is known to be non-energized; and $\tilde{\Gamma}_{mn} = 1$ if there is no prior information for line (m,n). If there is no information for any line, then $\tilde{\boldsymbol{\Gamma}} = \mathbf{1}\mathbf{1}^\top$. Based on $\tilde{\boldsymbol{\Gamma}}$, define the set

$$\mathcal{S}(\tilde{\boldsymbol{\Gamma}}) := \left\{ \boldsymbol{\Theta} : \begin{array}{l} \Theta_{mn} \leq 0, \text{ if } \tilde{\Gamma}_{mn} = 1 \\ \Theta_{mn} = 0, \text{ if } \tilde{\Gamma}_{mn} = 0 \end{array} m, n \in \mathcal{N}, m \neq n \right\}.$$

The set $\mathcal{S}(\tilde{\boldsymbol{\Gamma}})$ ignores possible prior information on lines fed directly by the substation. This information is encoded on the zero-th column of $\tilde{\boldsymbol{\Gamma}}$. In particular, if $\tilde{\Gamma}_{0n} = 1$, then $\tilde{\Theta}_{0n} \leq 0$ and $\sum_{m=1}^{N} \Theta_{mn} \geq 0$. Otherwise, it holds that $\tilde{\Theta}_{0n} = \sum_{m=1}^{N} \Theta_{mn} = 0$. The two properties pertaining to lines directly connected to the substation are captured by the set

$$\mathcal{S}_0(\tilde{\boldsymbol{\Gamma}}) := \left\{ \boldsymbol{\Theta} : \begin{array}{l} \mathbf{e}_n^\top \boldsymbol{\Theta}\mathbf{1} \geq 0, \text{ if } \tilde{\Gamma}_{0n} = 1 \\ \mathbf{e}_n^\top \boldsymbol{\Theta}\mathbf{1} = 0, \text{ if } \tilde{\Gamma}_{0n} = 0 \end{array}, n \in \mathcal{N} \right\}.$$

Summarizing, the set of admissible reduced Laplacian matrices for arbitrary graphs with prior edge information $\tilde{\boldsymbol{\Gamma}}$ is

$$\mathcal{C} := \left\{ \boldsymbol{\Theta} : \boldsymbol{\Theta} \in \mathcal{S}(\tilde{\boldsymbol{\Gamma}}) \cap \mathcal{S}_0(\tilde{\boldsymbol{\Gamma}}), \boldsymbol{\Theta} = \boldsymbol{\Theta}^\top \right\}. \tag{0.40}$$

By invoking the Gershgorin's disc theorem, it can be shown that $\boldsymbol{\Theta} \succeq \mathbf{0}$ for all

$\boldsymbol{\Theta} \in \mathcal{C}$, that is $\mathcal{C} \subseteq \mathbb{S}^+$. The reduced Laplacian matrices included in $\mathcal{C}$ correspond to possibly meshed and/or disconnected graphs. Enforcing two additional properties can render $\boldsymbol{\Theta}$ a proper reduced Laplacian of a tree: First, matrix $\boldsymbol{\Phi}(\boldsymbol{\Theta})$ should have exactly $2N$ non-zero off-diagonal entries. Second, matrix $\boldsymbol{\Theta}$ should be strictly positive definite since the graph is connected. Then, define the set

$$\mathcal{T} := \{ \boldsymbol{\Theta} : \|\boldsymbol{\Phi}(\boldsymbol{\Theta})\|_{0,\text{off}} = 2N, \boldsymbol{\Theta} \succ \mathbf{0} \}. \tag{0.41}$$

Back to topology identification using (0.26), matrix $\boldsymbol{\Theta}$ can be estimated via a (weighted) least-squares (LS) fit of the probing data under Laplacian constraints

$$\min_{\boldsymbol{\Theta} \in \mathcal{C} \cap \mathcal{T}} f(\boldsymbol{\Theta}) := \frac{1}{2} \|\boldsymbol{\Theta} \tilde{\mathbf{V}} - \mathbf{I}_{\mathcal{C}} \boldsymbol{\Delta}\|_{\mathbf{W}}^2. \tag{0.42}$$

Albeit its objective and set $\mathcal{C}$ are convex, the optimization in (0.42) is challenging because $\mathcal{T}$ is non-convex and open. To arrive at a practical solution, we surrogate $\mathcal{T}$ by adding two penalties in the objective of (0.42). Heed that the property $\boldsymbol{\Theta} \succ \mathbf{0}$ is equivalent to enforcing a finite lower bound on $\log |\boldsymbol{\Theta}|$. On the other hand, the non-convex pseudo-norm $\|\boldsymbol{\Phi}(\boldsymbol{\Theta})\|_{0,\text{off}}$ can be relaxed by its convex envelope $\|\boldsymbol{\Phi}(\boldsymbol{\Theta})\|_{1,\text{off}} := \sum_{m,n \neq m} |\Phi(\boldsymbol{\Theta})_{mn}|$; see also [37, 38, 39] for related approaches aiming to recover sparse inverse covariance or Laplacian matrices. By defining $\boldsymbol{\Pi} := \mathbf{I} + \mathbf{1}\mathbf{1}^\top$, $\|\boldsymbol{\Phi}(\boldsymbol{\Theta})\|_{1,\text{off}}$ can be rewritten as

$$\|\boldsymbol{\Phi}(\boldsymbol{\Theta})\|_{1,\text{off}} = \text{Tr}(\boldsymbol{\Theta}\boldsymbol{\Pi})$$

Upon dualizing the constraints comprising $\mathcal{T}$, (0.42) can be convexified as

$$\hat{\boldsymbol{\Theta}} := \arg \min_{\boldsymbol{\Theta} \in \mathcal{C}} \ \frac{1}{2} \|\boldsymbol{\Theta} \tilde{\mathbf{V}} - \mathbf{I}_{\mathcal{C}} \boldsymbol{\Delta}\|_{\mathbf{W}}^2 + \lambda \, \text{Tr}(\boldsymbol{\Theta}\boldsymbol{\Pi}) - \mu \log |\boldsymbol{\Theta}| \tag{0.43}$$

where $\lambda, \mu > 0$ are tunable parameters. Since the minimizer of (0.43) does not necessarily belong to $\mathcal{T}$, one may invoke heuristics to convert it to the reduced Laplacian of a tree graph. As suggested in [2], a Laplacian $\tilde{\boldsymbol{\Theta}}$ belonging to $\mathcal{T}$ can be found by running a minimum spanning tree algorithm for the graph defined by $\boldsymbol{\Phi}(\hat{\boldsymbol{\Theta}})$. Problem (0.43) can be solved using the alternating direction method of multipliers (ADMM).

Before closing this section, two comments are in order: *i)* beyond identification, probing can be used for topology detection as well [31]; and *ii)* Instead of the convex relaxation proposed here, topology identification using probing can be solved *exactly* via the mixed-integer linear program (MILP) formulation of [40].

0.9   Numerical Tests on Topology Identification using Probing

Our probing algorithms were validated on the IEEE 37-bus feeder, see also Section 0.3.3. Figures 0.7 a) and b) show the actual and reduced topologies. Probing buses were equipped with inverters having the same rating as the related load. Loads were generated by adding a zero-mean Gaussian variation to the benchmark data, with standard deviation 0.067 times the average of nominal loads.
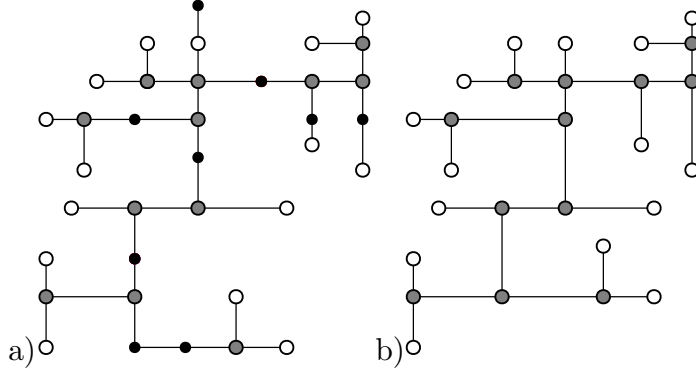
Figure 0.7 Line infrastructure for the IEEE 37-bus feeder benchmark.

Table 0.1 Numerical tests under complete and partial noisy data

T_m	1	10	40	90		1	5	20	39
EP [%]	98.5	55.3	3.1	0.2		97.2	45.8	18.9	0.1
MPE [%]	35.1	32.5	30.9	28.5		97.2	45.8	18.9	0.1
		Alg. 1					Alg. 2		

Voltages were obtained via a power flow solver, and then corrupted by zero-mean Gaussian noise with 3σ deviation of 0.01% per unit (pu). Probing actions were performed using the probing protocol $\mathbf{\Delta}_2$. The algorithms were tested through 200 Monte Carlo tests. At every run, the actual topology was randomly drawn.

Firstly, the graph algorithms of Section 0.7 were tested. For the 37-bus feeder, $r_{\min} = 0.0014$ pu. From the rated δ_m's; the $r_{\min}$; and (0.39), the number of probing actions was set as $T_m = 90$. In the partial data case, the smallest effective resistance was 0.0021 pu, yielding $T_m = 39$. Level sets obtained using the procedure described in Sec. 0.7.3 were given as inputs to Alg. 1 and 2. Table 0.1 demonstrates that the error probability (EP) in topology recovery and the mean percentage error (MPE) of line resistances decay as T_m increases.

Secondly, the identification problem in (0.42) was solved for $\lambda = 5 \cdot 10^{-3}$, $\mu = 1$, and $\mathbf{W} = \mathbf{I}$. No prior information on line statuses was assumed ($\tilde{\mathbf{\Gamma}} = \mathbf{11}^\top$). Kruskal's algorithm was used to obtain a Laplacian matrix $\tilde{\mathbf{\Theta}}$ corresponding to a radial grid. The connectivity of $\tilde{\mathbf{\Theta}}$ was compared against the actual one. The average number of line status errors, including both the energized lines not detected and the non-energized lines detected, are reported in Table 0.2.

Table 0.2 Probability of detecting a wrong topology via (0.42)

T_m	1	2	5	10
Topology error prob. [%]	5.1	3.9	3.7	2.6

0.10   Discussion

This chapter has put forth a grid topology learning toolbox. If the operator collects data passively, grid topology can be identified using ML, MAP, or partial correlation-based schemes. To accelerate learning, the operator may perturb inverter injections and collect voltage responses to infer the grid topology. We have provided conditions under which the grid topology can be correctly identified by probing. Computationally, under high SNR setups, the topology can be revealed using a recursive graph algorithm. For lower SNR setups and complete data, a convex relaxation approach has been suggested instead. The presented works motivate several interesting questions such as extending these results to multiphase setups; using the tap changes of voltage regulators and capacitor banks as a means for active topology learning; and combining actively and passively collected grid data. Additional directions include active semi-blind approaches along the lines of passive ones dealt with for a general graph in [41].

Acknowledgments

Bibliography

[1] Deka D, Chertkov M, Backhaus S. Structure Learning in Power Distribution Networks. IEEE Trans Control Netw Syst. 2018 Sep;5(3):1061–1074.

[2] Bolognani S, Bof N, Michelotti D, Muraro R, Schenato L. Identification of power distribution network topology via voltage correlation analysis. In: Proc. IEEE Conf. on Decision and Control. Florence, Italy; 2013. p. 1659–1664.

[3] Park S, Deka D, Chertkov M. Exact Topology and Parameter Estimation in Distribution Grids with Minimal Observability. In: Proc. Power Syst. Comput. Conf. Dublin, Ireland; 2018. .

[4] Sharon Y, Annaswamy AM, Motto AL, Chakraborty A. Topology identification in distribution network with limited measurements. In: Proc. IEEE Conf. on Innovative Smart Grid Technologies. Washington, DC; 2012. .

[5] Kekatos V, Giannakis GB. Joint Power System State Estimation and Breaker Status Identification. In: Proc. North American Power Symposium. Urbana-Champaign, IL; 2012. .

[6] Zhu H, Giannakis GB. Sparse Overcomplete Representations for Efficient Identification of Power Line Outages. IEEE Trans Power Syst. 2012 Nov;27(4):2215–2224.

[7] Sevlian RA, Zhao Y, Rajagopal R, Goldsmith A, Poor HV. Outage Detection Using Load and Line Flow Measurements in Power Distribution Systems. IEEE Trans Power Syst. 2018 Mar;33(2):2053–2069.

[8] Yu J, Weng Y, Rajagopal R. PaToPa: A Data-Driven Parameter and Topology Joint Estimation Framework in Distribution Grids. IEEE Trans Power Syst. 2017 Jul;33(4):4335–4347.

[9] Zhao Y, Chen J, Poor HV. A Learning-to-Infer Method for Real-Time Power Grid Topology Identification; 2017. Available from: `https://arxiv.org/pdf/1710.07818.pdf`.

[10] Ardakanian O, Yuan Y, Wong V, Dobbe R, Low S, von Meier A, et al. On Identification of Distribution Grids. IEEE Trans Control Netw Syst. 2019;PP(99).

[11] Godsil C, Royle G. Algebraic Graph Theory. New York, NY: Springer; 2001.

[12] Baran ME, Wu FF. Network reconfiguration in distribution systems for loss reduction and load balancing. IEEE Trans Power Del. 1989 Apr;4(2):1401–1407.

[13] Bolognani S, Dorfler F. Fast power system analysis via implicit linearization of the power flow manifold. In: Proc. Allerton Conf. Allerton, IL; 2015. p. 402–409.

[14] Deka D, Backhaus S, Chertkov M. Learning topology of distribution grids using only terminal node measurements. In: Proc. IEEE Intl. Conf. on Smart Grid Commun. Syndey, Australia; 2016. .

[15] Cavraro G, Arghandeh R. Power Distribution Network Topology Detection With Time-Series Signature Verification Method. IEEE Trans Power Syst. 2018 Jul;33(4):3500–3509.

[16] Cavraro G, Kekatos V, Veeramachaneni S. Voltage Analytics for Power Distribution Network Topology Verification. IEEE Trans Smart Grid. 2019 Jan;10(1):1058–1067.

[17] Kekatos V, Giannakis GB, Baldick R. Online Energy Price Matrix Factorization for Power Grid Topology Tracking. IEEE Trans Smart Grid. 2016 May;7(7):1239–1248.

[18] Pecan Street Inc. Dataport; 2013. Available from: https://dataport.pecanstreet.org/.

[19] Karanikolas GV, Giannakis GB, Slavakis K, Leahy RM. Multi-Kernel Based Nonlinear Models for Connectivity Identification of Brain Networks. In: Proc. IEEE Intl. Conf. on Acoustics, Speech, and Signal Process. Shanghai, China; 2016. p. 6315 – 6319.

[20] Giannakis GB, Shen Y, Karanikolas GV. Topology identification and learning over graphs: Accounting for nonlinearities and dynamics. Proc IEEE. 2018 May;106(5):787–807.

[21] Saunders C, Gammerman A, Vovk V. Ridge regression learning algorithm in dual variables. In: Proc. Intl. Conf. Machine Learning. Madison, WI; 1998. .

[22] Shawe-Taylor J, Cristianini N. Kernel Methods for Pattern Analysis. New York, NY: Cambridge University Press; 2004.

[23] Lanckriet GRG, Cristianini N, Bartlett P, Ghaoui LE, Jordan MI. Learning the kernel matrix with semidefinite programming. J Mach Lear Res. 2004 Jan;5:27–72.

[24] Kloft M, Brefeld U, Laskov P, Müller KR, Zien A, Sonnenburg S. Efficient and accurate ℓ_p-norm multiple kernel learning. In: Neural Inf. Process. Syst. Vancouver, Canada; 2009. .

[25] Zhang L, Romero D, Giannakis GB. Fast convergent algorithms for multi-kernel regression. In: IEEE Wkshp. on Statistical Signal Process. Mallorca, Spain; 2016. .

[26] Jaggi M. Revisiting Frank-Wolfe: Projection-free sparse convex optimization. In: Intl. Conf. on Machine Learning. vol. 28. Atlanta, GA; 2013. p. 427–435.

[27] Zhang L, Wang G, Romero D, Giannakis GB. Randomized Block Frank-Wolfe for Convergent Large-Scale Learning. IEEE Trans Signal Process. 2017 Sep;65(24):6448 – 6461.

[28] Kaggle. Global Energy Forecasting Competition; 2012. Available from: https://www.kaggle.com/c/global-energy-forecasting-competition-2012-load-forecasting/data.

[29] Sturm JF. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. Optim Method Softw. 1999 Jan;11(1-4):625–653.

[30] Deka D, Talukdar S, Chertkov M, Salapaka M. Topology Estimation in Bulk Power Grids: Guarantees on Exact Recovery; 2017. [Online]. Available: https://arxiv.org/abs/1707.01596.

[31] Cavraro G, Kekatos V. Inverter Probing for Power Distribution Network Topology Processing. IEEE Trans Control Netw Syst. 2019;PP(99).

[32] Cavraro G, Kekatos V. Graph Algorithms for Topology Identification using Power Grid Probing. IEEE Control System Letters. 2018 Oct;2(4):689–694.

[33] Cavraro G, Bernstein A, Kekatos V, Zhang Y. Real-Time Identifiability of Power Distribution Network Topologies with Limited Monitoring; 2019. (submitted for publication). Available from: `https://www.faculty.ece.vt.edu/kekatos/papers/L-CSS-2019.pdf`.

[34] Dorfler F, Bullo F. Kron Reduction of Graphs With Applications to Electrical Networks. IEEE Trans Circuits Syst I. 2013 Jan;60(1):150–163.

[35] Bolognani S, Cavraro G, Zampieri S. A Distributed Feedback Control Approach to the Optimal Reactive Power Flow Problem. Heidelberg: Springer Intl. Publishing; 2013. p. 259–277.

[36] Kekatos V, Zhang L, Giannakis GB, Baldick R. Voltage Regulation Algorithms for Multiphase Power Distribution Grids. IEEE Trans Power Syst. 2016 Sep;31(5):3913–3923.

[37] Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. Biostatistics. 2008 Dec;9(3):432–441.

[38] Kekatos V, Giannakis GB, Baldick R. Grid Topology Identification using Electricity Prices. In: Proc. IEEE Power & Energy Society General Meeting. Washington, DC; 2014. p. 1–5.

[39] Egilmez HE, Pavez E, Ortega A. Graph Learning From Data Under Laplacian and Structural Constraints. IEEE J Sel Topics Signal Process. 2017 Sep;11(6):825–841.

[40] Taheri S, Kekatos V, Cavraro G. An MILP Approach for Distribution Grid Topology Identification using Inverter Probing. In: IEEE PowerTech. Milan, Italy; 2019. p. 1–5.

[41] Ioannidis VN, Shen Y, Giannakis GB. Semi-Blind Inference of Topologies and Dynamical Processes over Graphs. IEEE Trans Signal Process. 2019 Apr;67(9):2263–2274.

Index