

Voltage Analytics for Power Distribution Network Topology Verification

Guido Cavraro, Vassilis Kekatos, *Senior Member, IEEE*, and Sriharsha Veeramachaneni

Abstract—Distribution grids constitute complex networks of lines oftentimes reconfigured to minimize losses, balance loads, alleviate faults, or for maintenance purposes. Topology monitoring becomes a critical task for optimal grid scheduling. While synchrophasor installations are limited in low-voltage grids, utilities have an abundance of smart meter data at their disposal. In this context, a statistical learning framework is put forth for verifying single-phase grid structures using non-synchronized voltage data. The related maximum likelihood task boils down to minimizing a non-convex function over a non-convex set. The function involves the sample voltage covariance matrix and the feasible set is relaxed to its convex hull. Asymptotically in the number of data, the actual topology yields the global minimizer of the original and the relaxed problems. Under a simplified data model, the function turns out to be convex, thus offering optimality guarantees. Prior information on line statuses is also incorporated via a maximum a-posteriori approach. The formulated tasks are tackled using solvers having complementary strengths. Numerical tests using real data on benchmark feeders demonstrate that reliable topology estimates can be acquired even with a few smart meter data, while the non-convex schemes exhibit superior line verification performance at the expense of additional computational time.

Index Terms—Maximum likelihood; inverse covariance matrix estimation; linearized distribution flow model.

I. INTRODUCTION

Low commercial interest together with the sheer coverage of residential low-voltage grids have resulted in their limited instrumentation. Traditionally, utility operators collect voltage, current, and power readings only from a few grid points. With the growing interest in integrating solar generation along with demand-response and electric vehicle charging programs, utilities need more refined models of their assets to accomplish grid scheduling tasks. To this end and given the currently prohibitive cost of installing synchrophasors on a wide scale, data from advanced metering infrastructure (AMI) and power inverters can provide useful grid information.

One such critical piece of information is the operational structure of a grid. Power networks are built with line redundancy for efficiency, reliability, and maintenance purposes. Although operators know the available lines along with their characteristics, the energized topology under which the grid

operates at any given time may not be precisely known. At the transmission level, the grid topology is typically acquired using historical data, measurements, and the generalized state estimator [1]. Detecting sudden single- and double-line outages via efficient enumerations has been suggested in [2], while outages of multiple lines are unveiled via the sparse overcomplete representation of [3]. Given power injections across the network over multiple times, grid connectivity is recovered via a blind matrix factorization in [4]. A Gaussian Markov random field has been postulated over bus voltage angles to localize transmission grid faults [5]. Instead of using electrical quantities, power system topologies are tracked using publicly available electricity prices in [6].

Focusing on distribution grids, reference [7] exploits the time delays of power line communication signals to reveal the grid structure. The topology recovery scheme of [8] relies on the properties of the inverse covariance matrix of bus voltage magnitudes. After developing a linearized distribution grid model, reference [9] generalizes the previous schemes to voltage data from multiple feeders, correlated power injections, and grids with variable resistance-to-reactance ratios. The graph recovery algorithms of [9] have been extended to incorporate covariances of power injections from terminal nodes [10]. Grid topology recovery has been tackled using graphical models by fitting a spanning tree based on the mutual information of voltage data in [11]. The aforementioned approaches rely on the *ensemble* rather than *sample* moments of meter data. In a grid of N buses, the sample covariance of voltage data becomes invertible after collecting at least N data. If meters report every few minutes, the sample covariance can be inverted only after some hours. Even then, its inverse would deviate substantially from its ensemble counterpart.

Topology inference methods relying on synchrophasor data have also been suggested for distribution grids. The scheme of [12] selects the topology attaining the best least-squares fit in an exhaustive fashion whose complexity grows exponentially in the number of configurations. A data-driven algorithm for detecting switching events based on topology signatures has been reported [13]. A sparse linear model capturing the voltage dependence between every node and all other nodes is sought via ℓ_1 -penalized regression in [14]; yet the per-node models may not agree. In [15], the bus admittance matrix is found via linear regression and its observability is characterized presuming all non-metered buses do not inject power. Albeit synchrophasors for distribution grids are underway, their current cost inhibits wide adoption.

The task of verifying grid topologies using non-synchronized voltage data is considered here. Our contribution is three-fold. First, topology verification in single-phase grids

Manuscript received Apr 1, 2017; revised July 18, 2017 and September 1, 2017; accepted September 28, 2017. Date of publication DATE; date of current version DATE. Paper no. TSG.00453.2017.

G. Cavraro and V. Kekatos are with the Bradley Dept. of Electrical and Computer Engr., Virginia Tech, Blacksburg, VA 24061, USA. S. Veeramachaneni is with Windlogics Inc., Saint Paul, MN 55108, USA. Emails: {cavraro,kekatos}@vt.edu, Sriharsha.Veeramachaneni@windlogics.com.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier XXXXXX

is posed as a maximum likelihood (ML) problem involving the *sample* covariance matrix of voltage data. After reviewing the grid model in Section II, the grid topology is captured by a binary vector (Section III-A). The associated non-convex set is relaxed to its convex hull and a stationary point of the non-convex likelihood function is found via gradient projection. Asymptotically in the number of data, the true topology constitutes the global minimizer for both problems. Second, the novel learning schemes simplify if lines are assumed to exhibit identical resistance-to-reactance ratios, see also [8]. By further ignoring noise, the likelihood function becomes convex, and hence, numerical bounds on the suboptimality of the relaxation are obtained (Section III-B). Lastly, possible prior information on the status of individual lines is incorporated in terms of a maximum *a posteriori* (MAP) estimator (Section IV-C). The numerical tests of Section V using actual data on benchmark feeders corroborate our findings, and conclusions are drawn in Section VI.

Regarding *notation*, lower- (upper-) case boldface letters denote column vectors (matrices), with the exception of the line complex power flow vector \mathbf{S} . Calligraphic symbols are reserved for sets. Symbol $^\top$ stands for transposition and \bar{z} is the complex conjugate of z . Vectors $\mathbf{0}$ and $\mathbf{1}$ are the all-zero and all-one vectors. Symbol $\|\mathbf{x}\|_2$ denotes the ℓ_2 -norm of \mathbf{x} and $\text{dg}(\mathbf{x})$ defines a diagonal matrix having \mathbf{x} on its diagonal. A symmetric positive (semi)definite matrix is denoted as $\mathbf{X} \succ \mathbf{0}$ ($\mathbf{X} \succeq \mathbf{0}$).

II. MODELING PRELIMINARIES

The proposed topology verification schemes build on the approximate linearized model for distribution grids that are briefly reviewed next. A radial single-phase grid having $N+1$ buses can be represented by a tree graph $\mathcal{T} = (\mathcal{N}_o, \mathcal{L})$, whose nodes $\mathcal{N}_o := \{0, \dots, N\}$ correspond to buses and its edges \mathcal{L} to distribution lines. The tree is rooted at the substation bus indexed by $n = 0$, while every non-root bus $n \in \mathcal{N} := \{1, \dots, N\}$ is connected to its unique parent bus π_n via line n . The grid is modeled by the branch flow equations [16]

$$s_n = \sum_{k \in \mathcal{C}_n} S_k - S_n + I_n z_n \quad (1a)$$

$$v_n = v_{\pi_n} - 2 \text{Re}[\bar{z}_n S_n] + I_n |z_n|^2 \quad (1b)$$

$$|S_n|^2 = v_{\pi_n} I_n \quad (1c)$$

where $z_n = r_n + jx_n$ is the impedance of line n ; I_n is the squared current magnitude on line n ; S_n is the complex power flow sent on line n from the parent node π_n ; and \mathcal{C}_n is the set of children buses for n . Further, v_n and $s_n = p_n + jq_n$ are the squared voltage magnitude and power injection at bus n . Equation (1a) follows from conservation of power; equation (1b) describes the drop in squared voltages along line n [16]; and (1c) stems from the definition of complex power flow.

Since the nonlinearity in (1c) complicates power flow calculations, distribution grids are sometimes studied using the *linear distribution flow* (LDF) model [16]. The latter follows from (1) upon dropping the last summands in the right-hand sides (RHS) of (1a)–(1b). It can also be derived assuming voltage magnitudes to be close to unity and voltage angle

differences between neighboring buses to be small [9]; or alternatively, via a first-order Taylor series approximation of power injections as functions of voltages [17].

The topology of a grid with $N+1$ buses and L distribution lines is captured by the branch-bus incidence matrix $\hat{\mathbf{A}} \in \{0, \pm 1\}^{L \times (N+1)}$ that can be partitioned into its first and the rest of its columns as $\hat{\mathbf{A}} = [\mathbf{a}_0 \ \mathbf{A}]$. For a radial grid ($L = N$), the *reduced incidence matrix* \mathbf{A} is square and invertible [18]. After ignoring losses, equations (1a)–(1b) can be compactly expressed in matrix-vector form as [19]

$$\mathbf{s} = \mathbf{A}^\top \mathbf{S} \quad (2a)$$

$$\mathbf{A} \mathbf{v} = 2 \text{Re}[\text{dg}(\bar{\mathbf{z}}) \mathbf{S}] - \mathbf{a}_0 v_0 \quad (2b)$$

where $\mathbf{s} = \mathbf{p} + j\mathbf{q}$ is the vector of complex nodal injections; \mathbf{S} is the vector of complex power flows; $\mathbf{z} = \mathbf{r} + j\mathbf{x}$ is the vector of line impedances; and v_0 is the squared voltage magnitude at the substation. Substituting $\mathbf{S} = \mathbf{A}^{-\top} \mathbf{s}$ from (2a) in (2b), and exploiting the fact that $\mathbf{a}_0 + \mathbf{A} \mathbf{1} = \mathbf{0}$ (due to $\hat{\mathbf{A}} \mathbf{1} = \mathbf{0}$), the vector \mathbf{v} collecting the squared voltage magnitudes at all buses in \mathcal{N} can be approximated as [20], [9], [17]

$$\mathbf{v} \simeq \mathbf{R} \mathbf{p} + \mathbf{X} \mathbf{q} + v_0 \mathbf{1}_N \quad (3)$$

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

$$\mathbf{R} := 2(\mathbf{A}^\top \text{dg}^{-1}(\mathbf{r}) \mathbf{A})^{-1} \quad (4a)$$

$$\mathbf{X} := 2(\mathbf{A}^\top \text{dg}^{-1}(\mathbf{x}) \mathbf{A})^{-1}. \quad (4b)$$

The RHS of (3) has been shown to underestimate the actual squared voltage magnitudes with the bias depending on I_n 's; see [21]. Numerical tests report voltage magnitude approximation errors less than 0.005 pu; see [19, Fig. 6], [22].

Squared voltage magnitudes are collected at all nodes with a sampling period of T_s . Samples are indexed by the subscript $t = 0, \dots, T$. Upon applying (3) over two successive samples, the fluctuations in squared voltage magnitudes $\tilde{\mathbf{v}}_t := \mathbf{v}_t - \mathbf{v}_{t-1}$ caused by perturbations in power injections $\tilde{\mathbf{p}}_t := \mathbf{p}_t - \mathbf{p}_{t-1}$ and $\tilde{\mathbf{q}}_t := \mathbf{q}_t - \mathbf{q}_{t-1}$ follow the model

$$\tilde{\mathbf{v}}_t = \mathbf{R} \tilde{\mathbf{p}}_t + \mathbf{X} \tilde{\mathbf{q}}_t + \mathbf{n}_t \quad (5)$$

where the vector \mathbf{n}_t captures measurement noise, the approximation error introduced by LDF, and modeling inaccuracies.

III. TOPOLOGY VERIFICATION

Power grids are built with redundancy in line infrastructure. This redundancy improves system reliability against failures or during scheduled maintenance, while grids are regularly reconfigured for loss minimization [16]. Therefore, the set of energized lines \mathcal{L} is a subset of the existing lines denoted by the set \mathcal{L}_e with $|\mathcal{L}_e| = L_e$, and $\mathcal{L} \subseteq \mathcal{L}_e$. For example, line 742-744 in Fig. 1 belongs to \mathcal{L}_e but not to \mathcal{L} .

To verify grid topologies using smart meter data, let us express (\mathbf{R}, \mathbf{X}) defined in (4) in terms of the active lines as

$$\mathbf{R} = \left(\sum_{\ell \in \mathcal{L}} \frac{1}{2r_\ell} \mathbf{a}_\ell \mathbf{a}_\ell^\top \right)^{-1} \quad \text{and} \quad \mathbf{X} = \left(\sum_{\ell \in \mathcal{L}} \frac{1}{2x_\ell} \mathbf{a}_\ell \mathbf{a}_\ell^\top \right)^{-1} \quad (6)$$

where \mathbf{a}_ℓ^\top is the ℓ -th row of \mathbf{A} . By slightly abusing notation, matrix \mathbf{A} here has been augmented to include both active and

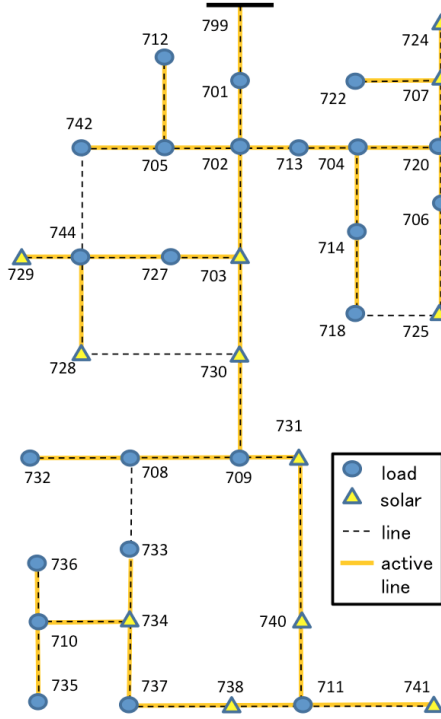


Fig. 1. Line infrastructure for the IEEE 37-bus feeder benchmark.

non-energized lines. Let us introduce the variable b_ℓ for line ℓ taking the values $b_\ell = 1$ if $\ell \in \mathcal{L}_e$; and $b_\ell = 0$ otherwise. Collecting b_ℓ 's in the binary L_e -length vector \mathbf{b} , matrices (\mathbf{R}, \mathbf{X}) can be written as

$$\mathbf{R}(\mathbf{b}) = \left(\sum_{\ell \in \mathcal{L}_e} \frac{b_\ell}{2r_\ell} \mathbf{a}_\ell \mathbf{a}_\ell^\top \right)^{-1}, \quad \mathbf{X}(\mathbf{b}) = \left(\sum_{\ell \in \mathcal{L}_e} \frac{b_\ell}{2x_\ell} \mathbf{a}_\ell \mathbf{a}_\ell^\top \right)^{-1}$$

where now the summations extend over all lines in \mathcal{L}_e . Under this representation, verifying the grid topology entails finding \mathbf{b} from meter data. Meter readings are collected by the utility over T time slots. Given the current and forthcoming communication capabilities of the smart metering infrastructure, samples are collected every $T_s = 5$ or 15 minutes. Within one hour, at most $T = 4$ to 12 voltage vectors can be recorded.

Smart meter data comprise voltage magnitudes, active powers, and power factors among other quantities. 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 to 10 residential costumers. Hence, power readings from meters may not be readily used. Instead, we statistically characterize power injections and rely on the second-order statistics of voltage readings to verify the grid topology.

A. Verification under a detailed data model

To arrive at a tractable statistical formulation, differential injection data $(\tilde{\mathbf{p}}_t, \tilde{\mathbf{q}}_t)$ are assumed to be zero-mean random vectors with covariance matrices $\Sigma_p := \mathbb{E}[\tilde{\mathbf{p}}_t \tilde{\mathbf{p}}_t^\top]$, $\Sigma_q := \mathbb{E}[\tilde{\mathbf{q}}_t \tilde{\mathbf{q}}_t^\top]$, and $\Sigma_{pq} := \mathbb{E}[\tilde{\mathbf{p}}_t \tilde{\mathbf{q}}_t^\top]$. The noise vector \mathbf{n}_t in (5) is modeled as zero-mean independent Gaussian with variance

σ_n^2 , that is $\mathbf{n}_t \sim \mathcal{N}(\mathbf{0}, \sigma_n^2 \mathbf{I})$. Under these assumptions and from (5), the differential squared voltage magnitudes $\tilde{\mathbf{v}}_t$ (hereafter termed *voltage data* for brevity) are also zero-mean with ensemble covariance matrix explicitly parameterized by the line indicator vector \mathbf{b} as

$$\begin{aligned} \Sigma(\mathbf{b}) = & \mathbf{R}(\mathbf{b}) \Sigma_p \mathbf{R}(\mathbf{b}) + \mathbf{X}(\mathbf{b}) \Sigma_q \mathbf{X}(\mathbf{b}) \\ & + \mathbf{R}(\mathbf{b}) \Sigma_{pq} \mathbf{X}(\mathbf{b}) + \mathbf{X}(\mathbf{b}) \Sigma_{pq}^\top \mathbf{R}(\mathbf{b}) + \sigma_n^2 \mathbf{I}_N. \end{aligned} \quad (7)$$

Based on the preceding modeling, the topology verification problem is stated next.

Definition 1. *The task of topology verification amounts to finding the subset \mathcal{L} of active distribution lines given:*

- (i) the distribution line infrastructure $\{\mathbf{a}_\ell\}_{\ell \in \mathcal{L}_e}$;
- (ii) the line parameters $\{r_\ell, x_\ell\}_{\ell \in \mathcal{L}_e}$;
- (iii) the covariance matrices $\Sigma_p, \Sigma_q, \Sigma_{pq}$; and
- (iv) voltage data $\{\tilde{\mathbf{v}}_t\}_{t=1}^T$ over T time instances.

Different from the task of topology identification where both the existing lines and their impedances are unknown [17], [9], in topology verification the utility operator knows the existing line infrastructure. Here, the operator collects voltage readings, but power injections are described only through their first- and second-order moments. This is to cater the scenario where smart meter data from one house can be used to trace back the voltage magnitude at the related pole transformer. The net injection on the pole transformer serving 5-10 additional houses however cannot be inferred. The covariances can be estimated from historical data. Postulating structure on covariances (e.g., diagonal matrices) reduces the number of parameters.

To build a statistical model of voltage data, we invoke contemporary variants of the central limit theorem (CLT) and 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.

Remark 1. *Note that $\tilde{\mathbf{v}}_t = \sum_{n=1}^N (\tilde{p}_{t,n} \mathbf{r}_n + \tilde{q}_{t,n} \mathbf{x}_n)$, where $(\tilde{p}_{t,n}, \tilde{q}_{t,n})$ are the n -th entries of $(\tilde{\mathbf{p}}_t, \tilde{\mathbf{q}}_t)$ and $(\mathbf{r}_n, \mathbf{x}_n)$ are the n -th columns of (\mathbf{R}, \mathbf{X}) , accordingly. The N -dimensional vector $\tilde{\mathbf{v}}_t$ is therefore expressed as the sum of N independent random vectors. Different from the classical multivariate CLT, the dimension of $\tilde{\mathbf{v}}_t$ grows with the number of its summands. Nevertheless, its pdf can still be approximated by a Gaussian pdf for bounded power injections [23, Sec. 2].*

The statistical tests based on actual data in Sec. V validate this assumption. Thus, the pdf of $\tilde{\mathbf{v}}_t$ can be approximated as

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

To fully characterize the collected data $\{\tilde{\mathbf{v}}_t\}_{t=1}^T$, their joint pdf should be provided. To that end, assuming independence across time is convenient and computationally tractable. In fact, the normalized autocorrelation functions depicted in Fig. 3 demonstrate that voltage data are relatively uncorrelated across time, especially for sampling periods larger than $T_s = 5$ min. Due to Gaussianity, uncorrelatedness implies

independence. Therefore, the joint pdf for the entire voltage dataset becomes $p(\{\tilde{\mathbf{v}}_t\}_{t=1}^T; \mathbf{b}) = \prod_{t=1}^T p(\tilde{\mathbf{v}}_t; \mathbf{b})$, or

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

Upon observing the data $\{\tilde{\mathbf{v}}_t\}_{t=1}^T$, function $p(\{\tilde{\mathbf{v}}_t\}_{t=1}^T; \mathbf{b})$ becomes the likelihood function of the unknown line indicator vector \mathbf{b} . Adopting a maximum likelihood (ML) approach, the sought \mathbf{b} can be found as the maximizer of $p(\{\tilde{\mathbf{v}}_t\}_{t=1}^T; \mathbf{b})$, or equivalently, as the minimizer of $-\log p(\{\tilde{\mathbf{v}}_t\}_{t=1}^T; \mathbf{b})$. After scaling by $2/T$ and ignoring constant terms, the negative log-likelihood function is

$$f(\mathbf{b}) := \log |\Sigma(\mathbf{b})| + \frac{1}{T} \sum_{t=1}^T \tilde{\mathbf{v}}_t^\top \Sigma^{-1}(\mathbf{b}) \tilde{\mathbf{v}}_t.$$

Exploiting the linearity of the trace operator $\text{Tr}(\cdot)$, the function $f(\mathbf{b})$ can be equivalently expressed as

$$f(\mathbf{b}) = \log |\Sigma(\mathbf{b})| + \text{Tr}(\Sigma^{-1}(\mathbf{b}) \hat{\Sigma}) \quad (8)$$

where $|\cdot|$ denotes the matrix determinant and $\hat{\Sigma}$ is the sample covariance matrix of the voltage data

$$\hat{\Sigma} := \frac{1}{T} \sum_{t=1}^T \tilde{\mathbf{v}}_t \tilde{\mathbf{v}}_t^\top. \quad (9)$$

Function $f(\mathbf{b})$ consists of two terms. The term $\text{Tr}(\Sigma^{-1}(\mathbf{b}) \hat{\Sigma})$ aggregates all the information from voltage data and can be understood as a data-fitting term. The term $\log |\Sigma(\mathbf{b})|$ on the other hand acts as a regularizer guarding $\Sigma(\mathbf{b})$ within the positive definite matrix cone and away from singularity [6].

Vector \mathbf{b} can now be found by minimizing

$$f^* := \min_{\mathbf{b} \in \mathcal{B}, \mathbf{1}^\top \mathbf{b} = L} f(\mathbf{b}) \quad (10)$$

where $\mathcal{B} := \{0, 1\}^{L_e}$ and the constraint $\mathbf{1}^\top \mathbf{b} = L$ fixes the number of active lines to L . In a grid with $N + 1$ nodes, requiring $L = N$ in (10) enforces a radial structure. To see this, note that the graph induced by any N edges not forming a tree is not a connected graph. The corresponding reduced incidence matrix $\mathbf{A}(\mathbf{b})$ is not full column-rank. Consequently, matrices $(\mathbf{R}(\mathbf{b}), \mathbf{X}(\mathbf{b}))$ become singular and $f(\mathbf{b})$ goes to infinity.

Solving (10) is non-trivial due to the binary constraints. The set \mathcal{B} can be surrogated by its convex hull $\mathcal{B}_r := [0, 1]^{L_e}$ and a lower bound on f^* can be found by solving

$$f_r^* := \min_{\mathbf{b} \in \mathcal{B}_r, \mathbf{1}^\top \mathbf{b} = L} f(\mathbf{b}). \quad (11)$$

The function $f(\mathbf{b})$ can be shown to be non-convex in general. A projected gradient descent (PGD) scheme with a sufficiently small step size $\mu > 0$ is guaranteed to converge to a point $\tilde{\mathbf{b}}$ satisfying the necessary optimality condition for (11), that is [24, Prop. 6.1.3]

$$(\nabla f(\tilde{\mathbf{b}}))^\top (\mathbf{b} - \tilde{\mathbf{b}}) \geq 0 \text{ for all } \mathbf{b} \in \mathcal{B}_r \text{ and } \mathbf{1}^\top \mathbf{b} = L. \quad (12)$$

After initializing \mathbf{b} at some \mathbf{b}^0 , the k -th PGD iteration reads

$$\mathbf{b}^{k+1} := \arg \min_{\mathbf{b} \in \mathcal{B}_r, \mathbf{1}^\top \mathbf{b} = L} \|\mathbf{b} - \mathbf{b}^k + \mu \nabla f(\mathbf{b}^k)\|_2^2. \quad (13)$$

The projection in (13) can be handled either by a generic linearly-constrained convex quadratic program; by the lambda iteration method [1, Sec. 5.2.4]; or by dual ascent upon dualizing the constraint $\mathbf{1}^\top \mathbf{b} = L$. The gradient $\nabla f(\mathbf{b})$ is calculated using standard matrix differential rules as shown in the Appendix.

Lemma 1. *The ℓ -th entry of the gradient of $f(\mathbf{b})$ is*

$$\begin{aligned} \frac{\partial f(\mathbf{b})}{\partial b_\ell} = & -\frac{1}{r_\ell} \mathbf{a}_\ell^\top \mathbf{R}(\mathbf{b}) \Sigma_p \mathbf{R}(\mathbf{b}) \mathbf{F}(\mathbf{b}) \mathbf{R}(\mathbf{b}) \mathbf{a}_\ell \\ & -\frac{1}{x_\ell} \mathbf{a}_\ell^\top \mathbf{X}(\mathbf{b}) \Sigma_q \mathbf{X}(\mathbf{b}) \mathbf{F}(\mathbf{b}) \mathbf{X}(\mathbf{b}) \mathbf{a}_\ell \\ & -\frac{1}{r_\ell} \mathbf{a}_\ell^\top \mathbf{R}(\mathbf{b}) \Sigma_{pq} \mathbf{X}(\mathbf{b}) \mathbf{F}(\mathbf{b}) \mathbf{R}(\mathbf{b}) \mathbf{a}_\ell \\ & -\frac{1}{x_\ell} \mathbf{a}_\ell^\top \mathbf{X}(\mathbf{b}) \mathbf{F}(\mathbf{b}) \mathbf{R}(\mathbf{b}) \Sigma_{pq} \mathbf{X}(\mathbf{b}) \mathbf{a}_\ell \end{aligned} \quad (14)$$

where $\mathbf{F}(\mathbf{b}) := \Sigma^{-1}(\mathbf{b}) - \Sigma^{-1}(\mathbf{b}) \hat{\Sigma} \Sigma^{-1}(\mathbf{b})$.

Consistent with the properties of the maximum likelihood estimator (see [25]), the true line indicator vector minimizes (10) and (11), when the number of data T grows to infinity; see the Appendix for a proof.

Lemma 2. *Let \mathbf{b}_o be the true line indicator vector. If the sample covariance $\hat{\Sigma}$ has converged to the ensemble covariance $\Sigma(\mathbf{b}_o)$, then \mathbf{b}_o is a stationary point of $f(\mathbf{b})$ and global minimizer for (10) and (11).*

The PGD estimate $\tilde{\mathbf{b}}$ may not lie in the original non-convex set \mathcal{B} . Then, a feasible vector $\hat{\mathbf{b}}$ can be heuristically obtained by selecting the lines corresponding to the L largest entries of $\tilde{\mathbf{b}}$; by finding the minimum spanning tree on a graph having $\tilde{\mathbf{b}}$ as edge weights; or by choosing $\hat{\mathbf{b}}$ as the mean of independent Bernoulli distributions used to draw a number of line configurations from which the feasible one attaining the smallest $f(\mathbf{b})$ is returned as $\hat{\mathbf{b}}$. Obviously, it holds that $f^* \leq f(\hat{\mathbf{b}})$ and $f(\hat{\mathbf{b}}) \leq f(\tilde{\mathbf{b}})$. Although $\hat{\mathbf{b}}$ yields reasonable verification performance (see Sec. V), vector $\tilde{\mathbf{b}}$ may not be a global minimizer of (11) since $f(\mathbf{b})$ is non-convex.

B. Verification under a simplified data model

To alleviate the issues with the non-convexity of $f(\mathbf{b})$, we alternatively pursue a simplified model for voltage data by resorting to two assumptions [17]:

(a1) The resistance-to-reactance ratios $\alpha_\ell := r_\ell/x_\ell$ are identical for all lines, i.e., $\alpha_\ell = \alpha$ for all $\ell \in \mathcal{L}_e$.

(a2) The noise term \mathbf{n}_t is negligible, or $\sigma_n^2 = 0$.

Regarding (a1), the ratios α_ℓ 's practically tend to lie within a limited range $[\alpha_{\min}, \alpha_{\max}]$ as listed on Table I. The parameter α can be picked as the average or the median of the known actual ratios α_ℓ 's. Ignoring the noise term simplifies considerably the approach.

Under (a1)–(a2), the voltage data model in (5) simplifies as

$$\tilde{\mathbf{v}}_t = \alpha \mathbf{X} \bar{\mathbf{p}}_t + \mathbf{X} \tilde{\mathbf{q}}_t. \quad (15)$$

Defining $\Sigma_\alpha := \alpha^2 \Sigma_p + \Sigma_q + \alpha(\Sigma_{pq} + \Sigma_{pq}^\top)$, the ensemble covariance of the voltage data becomes

$$\tilde{\Sigma}(\mathbf{b}) = \mathbf{X}(\mathbf{b}) \Sigma_\alpha \mathbf{X}(\mathbf{b}). \quad (16)$$

TABLE I
DISTRIBUTION LINE RESISTANCE-TO-REACTANCE RATIOS

Feeder	α_{\min}	α_{\max}	mean	std	median
IEEE 34-bus	1.00	1.88	1.41	0.29	1.37
IEEE 37-bus	1.48	2.70	2.72	0.45	1.93
IEEE 123-bus	0.42	2.02	0.74	0.38	0.97

Using properties of the determinant and trace, the original negative log-likelihood in (8) is surrogated by the function

$$\tilde{f}(\mathbf{b}) := -2 \log |\mathbf{X}^{-1}(\mathbf{b})| + \text{Tr}[\mathbf{X}^{-1}(\mathbf{b})\Sigma_{\alpha}^{-1}\mathbf{X}^{-1}(\mathbf{b})\hat{\Sigma}] \quad (17)$$

which enjoys convexity as shown in the Appendix.

Lemma 3. *The function $\tilde{f}(\mathbf{b})$ is strictly convex over $\mathbf{b} \in \mathbb{R}_{+}^L$.*

Based on the simplified data model of (15), the line indicator vector can be recovered as the minimizer of

$$\tilde{\mathbf{b}}^* := \min_{\mathbf{b} \in \mathcal{B}, \mathbf{1}^{\top} \mathbf{b} = L} \tilde{f}(\mathbf{b}). \quad (18)$$

Again, to deal with the non-convexity of the feasible set, function $\tilde{f}(\mathbf{b})$ is now minimized over its convex hull instead

$$\tilde{\mathbf{b}} := \arg \min_{\mathbf{b} \in \mathcal{B}_r, \mathbf{1}^{\top} \mathbf{b} = L} \tilde{f}(\mathbf{b}). \quad (19)$$

Different from (11), the optimization in (19) is convex. Upon replacing $\Sigma(\mathbf{b})$, $\mathbf{R}(\mathbf{b})$, and r_{ℓ} , respectively by $\tilde{\Sigma}(\mathbf{b})$, $\alpha \mathbf{X}(\mathbf{b})$, and αx_{ℓ} onto (14), the ℓ -th entry of $\nabla \tilde{f}(\mathbf{b})$ becomes

$$\frac{\partial \tilde{f}(\mathbf{b})}{\partial b_{\ell}} = \frac{1}{x_{\ell}} \mathbf{a}_{\ell}^{\top} [\Sigma_{\alpha}^{-1} \mathbf{X}^{-1}(\mathbf{b}) \hat{\Sigma} - \mathbf{X}(\mathbf{b})] \mathbf{a}_{\ell}, \quad \forall \ell \in \mathcal{L}_e.$$

It is easy to verify that under the model of (15), Lemma 2 carries over to function \tilde{f} . Precisely, if \mathbf{b}_o is the true line indicator vector and $\hat{\Sigma}$ has converged to the ensemble covariance $\tilde{\Sigma}(\mathbf{b}_o) = \mathbf{X}(\mathbf{b}_o) \Sigma_{\alpha} \mathbf{X}(\mathbf{b}_o)$ under the approximate model, then $\nabla \tilde{f}(\mathbf{b}_o) = \mathbf{0}$. The latter together with the strong convexity of \tilde{f} imply that \mathbf{b}_o is the unique minimizer of (18) and (19).

The minimizer of (19) can be found by the PGD iteration

$$\mathbf{b}_r^{k+1} = \arg \min_{\mathbf{b} \in \mathcal{B}_r, \mathbf{1}^{\top} \mathbf{b} = L} \|\mathbf{b} - \mathbf{b}^k + \mu \nabla \tilde{f}(\mathbf{b}^k)\|_2^2. \quad (20)$$

The Frank-Wolfe (FW) or conditional gradient scheme is a competing alternative for tackling (19); see also [26], [27]. The scheme is suitable for minimizing a differentiable convex function over a compact set. Tailoring the FW iterates to the problem at hand provides

$$\tilde{\mathbf{b}}_r^{k+1} \in \arg \min_{\mathbf{b} \in \mathcal{B}_r, \mathbf{1}^{\top} \mathbf{b} = L} \mathbf{b}^{\top} \nabla \tilde{f}(\mathbf{b}_r^k) \quad (21a)$$

$$\mathbf{b}_r^{k+1} = \mathbf{b}_r^k + \frac{2}{k+2} (\tilde{\mathbf{b}}_r^{k+1} - \mathbf{b}_r^k). \quad (21b)$$

The step in (21a) involves a linear program (LP) whose minimizer can be found by simply setting the entries of $\tilde{\mathbf{b}}_r^{k+1}$ corresponding to the L smallest entries of $\nabla \tilde{f}(\mathbf{b}_r^k)$ to one, and setting the remaining entries to zero. The FW scheme converges sublinearly in $\mathcal{O}(1/k)$ iterations, while the PGD scheme exhibits linear convergence [24]. Nonetheless, the LP step of FW in (21a) is computationally cheaper than the costly projection in (20). Practically, the FW scheme is preferred over PGD when a low-accuracy solution suffices [27].

If the minimizer $\tilde{\mathbf{b}}$ of (19) belongs to \mathcal{B} , it is a minimizer of (18) too. Otherwise, a suboptimal \mathbf{b}' can be obtained from $\tilde{\mathbf{b}} \in \mathcal{B}_r$ using the heuristics discussed earlier. In this case, thanks to the convexity in (19), the suboptimality gap can be bounded as $f(\tilde{\mathbf{b}}) \leq \tilde{f}^* \leq f(\mathbf{b}')$; see also [28, Sec. 7.5].

Comparing the two approaches, the one in Sec. III-A relies on the detailed model of (5) and fully exploits the line data $\{r_{\ell}, x_{\ell}\}$. However, it entails a non-convex negative log-likelihood $f(\mathbf{b})$ that can be minimized only up to a point satisfying (12). The approach of this section builds on the simplified model of (15) requiring only $\{x_{\ell}\}$ and the approximate common ratio α . Because the associated function $\tilde{f}(\mathbf{b})$ is convex, the relaxation in (19) can be solved to optimality and with quantifiable gap from the non-convex task of (18).

IV. EXTENSIONS

Our learning toolbox is extended next to cope with meshed topologies, disjoint feeders, and prior information on lines.

A. Meshed topologies

Up to this point, the underlying grid topology was presumed to be radial. However, urban power distribution grids often-times exhibit operationally meshed structures. For distribution grids with $L \geq N$, the model in (5) generalizes to [9], [17]

$$\tilde{\mathbf{v}}_t = 2(\mathbf{G} + \mathbf{B}\mathbf{G}^{-1}\mathbf{B})^{-1} \tilde{\mathbf{p}}_t + 2(\mathbf{B} + \mathbf{G}\mathbf{B}^{-1}\mathbf{G})^{-1} \tilde{\mathbf{q}}_t + \mathbf{n}_t \quad (22)$$

for the bus conductance and susceptance matrices

$$\mathbf{G} := \mathbf{A}^{\top} \text{dg} \left(\left\{ \frac{r_{\ell}}{r_{\ell}^2 + x_{\ell}^2} \right\}_{\ell} \right) \mathbf{A} \quad (23a)$$

$$\mathbf{B} := \mathbf{A}^{\top} \text{dg} \left(\left\{ \frac{x_{\ell}}{r_{\ell}^2 + x_{\ell}^2} \right\}_{\ell} \right) \mathbf{A}. \quad (23b)$$

It is not hard to show that for radial networks where $L = N$, matrix \mathbf{A} is square and the model in (22) simplifies to (5).

The scheme of Sec. III-A can be extended to meshed networks based on (22). The gradient of $f(\mathbf{b})$ can be derived by adopting the proof of Lemma 1 as outlined next. The formulas of (28)–(29) carry over as long as \mathbf{R} is substituted by $\tilde{\mathbf{R}} := 2(\mathbf{G} + \mathbf{B}\mathbf{G}^{-1}\mathbf{B})^{-1}$, and \mathbf{X} is substituted by $\tilde{\mathbf{X}} := 2(\mathbf{B} + \mathbf{G}\mathbf{B}^{-1}\mathbf{G})^{-1}$. In this case however, the partial derivatives in (30) are replaced by

$$\begin{aligned} \frac{\partial \tilde{\mathbf{R}}(\mathbf{b})}{\partial b_{\ell}} &= -\frac{2r_{\ell}}{r_{\ell}^2 + x_{\ell}^2} \tilde{\mathbf{R}}(\mathbf{b}) \mathbf{a}_{\ell} \mathbf{a}_{\ell}^{\top} \tilde{\mathbf{R}}(\mathbf{b}) \\ &+ \frac{2r_{\ell}}{r_{\ell}^2 + x_{\ell}^2} \mathbf{B}(\mathbf{b}) \mathbf{G}^{-1}(\mathbf{b}) \mathbf{a}_{\ell} \mathbf{a}_{\ell}^{\top} \mathbf{G}^{-1}(\mathbf{b}) \mathbf{B}(\mathbf{b}) \\ &+ \frac{2x_{\ell}}{r_{\ell}^2 + x_{\ell}^2} \tilde{\mathbf{R}}(\mathbf{b}) \mathbf{B}(\mathbf{b}) \mathbf{G}^{-1}(\mathbf{b}) \mathbf{a}_{\ell} \mathbf{a}_{\ell}^{\top} \tilde{\mathbf{R}}(\mathbf{b}) \\ &+ \frac{2x_{\ell}}{r_{\ell}^2 + x_{\ell}^2} \tilde{\mathbf{R}}(\mathbf{b}) \mathbf{a}_{\ell} \mathbf{a}_{\ell}^{\top} \mathbf{G}^{-1}(\mathbf{b}) \mathbf{B}(\mathbf{b}) \tilde{\mathbf{R}}(\mathbf{b}) \end{aligned}$$

for all ℓ . An analogous expression holds for $\frac{\partial \tilde{\mathbf{X}}(\mathbf{b})}{\partial b_{\ell}}$.

On the other hand, under assumptions (a1)–(a2), the matrices in (23) simplify as $\mathbf{G} = \frac{2\alpha}{\alpha^2 + 1} \mathbf{X}^{-1}$ and $\mathbf{B} = \frac{2}{\alpha^2 + 1} \mathbf{X}^{-1}$. Simple algebraic manipulations show that (15)–(17) and the schemes of Sec. III-B apply *unaltered* to meshed grids. This point further justifies the benefit of adopting the approximate model under (a1)–(a2).

B. Disjoint feeders and multiple substation buses

Voltage readings may be collected from multiple radial feeders or a meshed grid rooted at multiple substation buses [9]. These scenarios can be handled in a uniform fashion under a simple modification. It is known that the nullity of the Laplacian matrix $\mathbf{L} := \mathbf{\tilde{A}}^\top \text{dg}^{-1}(\mathbf{r})\mathbf{\tilde{A}}$, i.e., the number of its zero eigenvalues, is equal to the number of the connected components in the underlying graph [18]. Moreover, every principal minor of \mathbf{L} is strictly positive definite if the graph is connected. By definition, the nullity of \mathbf{L} coincides with the nullspace dimension of $\mathbf{\tilde{A}}$.

When the grid is a single connected feeder, then $\dim(\text{null}(\mathbf{\tilde{A}})) = 1$. Dropping any column of $\mathbf{\tilde{A}}$ provides a full column-rank matrix. Recall that \mathbf{A} is derived from $\mathbf{\tilde{A}}$ by removing the column corresponding to the substation bus.

The idea extends naturally. If data are collected from different feeders, the corresponding Laplacian can be transformed under column- and row-permutations to a block diagonal matrix whose diagonal blocks constitute the Laplacian matrices of the connected grid. To guarantee the invertibility of the individual Laplacians, the matrix \mathbf{A} used in our schemes is obtained from $\mathbf{\tilde{A}}$ by dropping the columns related to all substation buses.

C. Incorporating prior information on lines

In meshed grids, the utility may not know the exact number of active lines L . On the other hand, prior information on individual lines being active could be known through a generalized state estimator; current magnitude readings on transformers; or commonly used grid configurations (e.g., seasonally). To cope with these setups, a maximum *a-posteriori* (MAP) approach is 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 conveniently expressed as $\Pr(b_\ell) = \pi_\ell^{b_\ell} (1 - \pi_\ell)^{1-b_\ell}$. To ease modeling, it is postulated that lines are active independently. Then, the joint pmf for \mathbf{b} is $\Pr(\mathbf{b}) = \prod_{\ell \in \mathcal{L}_e} \Pr(b_\ell)$ up to normalization and its negative logarithm is

$$\begin{aligned} -\log \Pr(\mathbf{b}) &= -\sum_{\ell \in \mathcal{L}_e} b_\ell \log \pi_\ell + (1 - b_\ell) \log(1 - \pi_\ell) \\ &= \sum_{\ell \in \mathcal{L}_e} b_\ell \beta_\ell - \log(1 - \pi_\ell) \end{aligned} \quad (24)$$

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 vector maximizing the posterior pdf $p(\mathbf{b}|\{\tilde{\mathbf{v}}_t\}_{t=1}^T)$. By Bayes' rule, the latter is proportional to the product $p(\{\tilde{\mathbf{v}}_t\}_{t=1}^T; \mathbf{b}) \Pr(\mathbf{b})$ scaled by the inconsequential term $p(\{\tilde{\mathbf{v}}_t\}_{t=1}^T)$. The MAP estimate can be equivalently found by minimizing

$$-\log p(\mathbf{b}|\{\tilde{\mathbf{v}}_t\}_{t=1}^T) = -\log p(\{\tilde{\mathbf{v}}_t\}_{t=1}^T; \mathbf{b}) - \log \Pr(\mathbf{b}).$$

By collecting the given parameters β_ℓ 's in vector $\boldsymbol{\beta}$ and ignoring constants, the latter leads to the problem

$$\mathbf{b}_{\text{MAP}} := \arg \min_{\mathbf{b} \in \mathcal{B}} \frac{T}{2} f(\mathbf{b}) + \boldsymbol{\beta}^\top \mathbf{b}. \quad (25)$$

If the simplified model of (15) is used in lieu of (5), the function $f(\mathbf{b})$ in (25) is replaced by the convex $\tilde{f}(\mathbf{b})$. Contrasted to (18), the task in (25) leverages prior information on lines

- If line ℓ is likely to be active, then $\pi_\ell > 1/2$ and $\beta_\ell < 0$.
- If line ℓ is known to be active, then $\pi_\ell = 1$ and $\beta_\ell = -\infty$, thus forcing the ℓ -th entry of \mathbf{b}_{MAP} to one.
- No prior on line ℓ entails $\pi_\ell = 1/2$ and $\beta_\ell = 0$.
- If all lines statuses exhibit the same mean π_c , that is $\mathbb{E}[b_\ell] = \pi_\ell = \pi_c$ for $\ell \in \mathcal{L}_e$, the linear term in the cost of (25) becomes $\pi_c \mathbf{1}^\top \mathbf{b}$. The latter can be interpreted as the Lagrangian counterpart of (18).

The feasible set \mathcal{B} can be again relaxed to \mathcal{B}_r . Lacking the coupling constraint $\mathbf{1}^\top \mathbf{b} = L$, the PGD update

$$\mathbf{b}^{k+1} = \arg \min_{\mathbf{b} \in \mathcal{B}_r} \|\mathbf{b} - \mathbf{b}^k + \mu \nabla f(\mathbf{b}^k) + \mu \boldsymbol{\beta}\|_2^2 \quad (26)$$

decouples across lines and the ℓ -th entry of \mathbf{b}^{k+1} is found as

$$b_\ell^{k+1} := \left[b_\ell^k - \mu \frac{\partial f(\mathbf{b}^k)}{\partial b_\ell} - \mu \beta_\ell \right]_0^1 \quad (27)$$

where $[x]_0^1$ projects x onto the interval $[0, 1]$.

V. NUMERICAL TESTS

The developed schemes were validated using the IEEE 37-bus and 123-bus feeders depicted in Figures 1 and 2 [29]. Buses $\{150, 195, 251, 451\}$ served as possible substations for the IEEE 123-bus feeder. The feeders were modified to single-phase using the process described in [21]. Zero-injection buses were assigned loads uniformly randomly chosen between the minimum and maximum bus loads provided by the benchmark.

Each bus hosted a uniformly random number of four to ten houses. Data collected from the Pecan Street project over January 1, 2013 were used as power profiles for houses with a random assignment [30]. These net power profiles correspond to residences with and without solar generation. Since the dataset does not provide reactive loads, power factors were drawn uniformly within 0.95 leading to 0.95 lagging for each bus at every time interval. The load aggregates at each bus were normalized so that their maxima matched the benchmark nominal values. The PV inverters implemented the following reactive control rule: if the generated active power exceeded half the PV capacity, the inverter would absorb more reactive power so that the inverter power factor decreases linearly with active power [31]. This localized rule was selected due to its simplicity and because it abides by the German codes [31].

Voltages were obtained through a power flow solver. The measurement noise was modeled as zero-mean Gaussian with a 3-sigma deviation matching 0.5% of the actual value. This is consistent with a study where 99% of the tested meters attained relative accuracy of less than 0.5% [32]. The schemes were run using MATLAB on a 2.7 GHz Intel Core i5 (12GB RAM).

First, the Gaussianity of the voltage data was tested through the Kolmogorov-Smirnov test for various sampling periods using one-day data from the IEEE 123-bus feeder with significance level 0.1. The results are listed in Table II. For the buses not passing the test, the test can still output the maximum distance between the sample cumulative distribution function

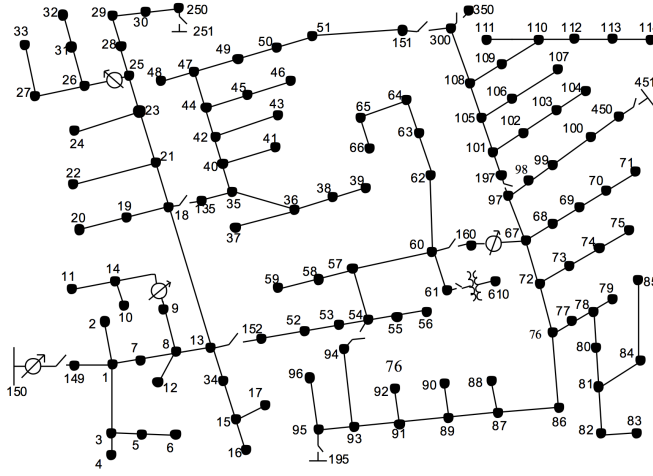


Fig. 2. The IEEE 123-bus distribution feeder with nine switches.

TABLE II
KOLMOGOROV-SMIRNOV TEST RESULTS

T_s [min]	5	15	30	60
# of samples T	288	96	48	24
# of buses failing the test	47	3	1	0
d_{ave}	0.1132	0.1582	0.2068	

(cdf) and the Gaussian cdf. The third row of Table II reports the average of those distances. Although Table II shows that the hypothesis on Gaussianity is less accurate for smaller T_s , the numerical tests shown later demonstrate that the developed schemes exhibit reasonable performance even for $T_s = 5$ min. Having Gaussianity in place, independence is ensured if data are uncorrelated over time. Figure 3 plots the normalized autocorrelation function for voltage data at buses 52 and 114 over T_s , which decays fast.

The ML verification schemes of Sec. III were then evaluated using 50 Monte Carlo runs and a sampling period of $T_s = 5$ min. The actual topologies for both grids were randomly chosen allowing for both radial and meshed configurations. The covariances Σ_p , Σ_q , and Σ_{pq} needed for evaluating gradients were approximated as diagonal matrices whose diagonal entries were estimated from historical data. In the first setup, line impedances were known and statuses were verified under the detailed model by solving (11). In the second setup, line resistance-to-reactance ratios $\{\alpha_\ell\}_{\ell \in \mathcal{L}_e}$ were approximated by their average, and line verification was tackled via (19). Because (11) is non-convex, the solution found by the PGD iterates of (13) depends on the initialization \mathbf{b}^0 . The latter was selected as the minimizer of the convex problem in (19). The solutions to (11) and (19) were projected onto \mathcal{B} by setting them as the mean of independent Bernoulli distributions used to draw a number of line configurations from which the feasible one associated with the smallest cost function is chosen as the selected configuration.

The convex problem in (19) was solved using both the PGD iterates of (13), and the FW iterates of (21). The non-convex problem of (11) was solved solely by the PGD scheme, since the FW iterates are guaranteed to converge only under a

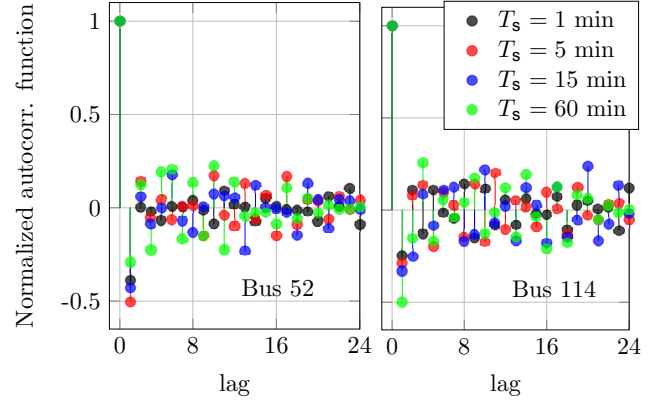


Fig. 3. Normalized autocorrelation function for voltage data on buses 52 (left) and 114 (right) computed using one-day data on the IEEE 123-bus feeder.

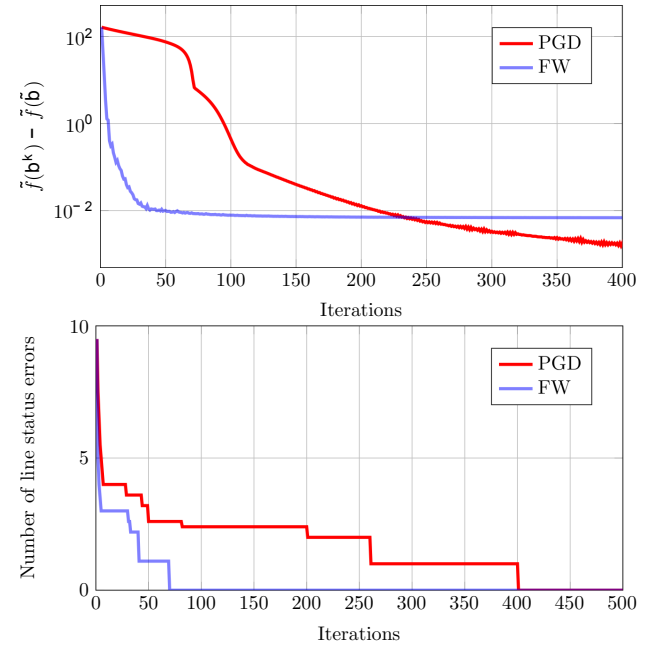


Fig. 4. (top) Convergence of the solvers of (19) for the IEEE 37-bus grid; (bottom) Detection performance after projecting the intermediate iterates.

non-trivial step size selection [33]. The top panel of Fig. 4 shows the convergence of the PGD and FW iterates for a particular instance of (19) with $T = 100$. Although solving the LP is computationally cheaper than the PGD projection, the execution time in both schemes is dominated by calculating $\nabla \tilde{f}(\mathbf{b})$. Numerical tests verify that the per-iteration time is similar for PGD and FW. Although PGD converges faster overall, the FW iterates exhibit faster initial convergence. The bottom panel of Fig. 4 shows the number of line status errors, if the intermediate estimates \mathbf{b}_r^k are used as the mean vector of a multivariate Bernoulli distribution from which 2,000 samples are drawn and evaluated. The latter indicates a potential advantage of the FW iterates.

The average running times for solving (11) and (19) reported in Table III indicate that the ML verification scheme using the simplified model is roughly seven to ten times faster. All algorithms here were run till convergence and without the early

TABLE III
AVERAGE RUNNING TIMES AND STEP SIZES FOR $T = 50$

ML and MAP solvers	IEEE 37-bus		IEEE 123-bus	
	Time [sec]	μ	Time [sec]	μ
PGD solver for (11)	60	0.007	336	0.0005
FW solver for (19)	5		45	
PGD solver for (25)	204	0.08	1332	0.001

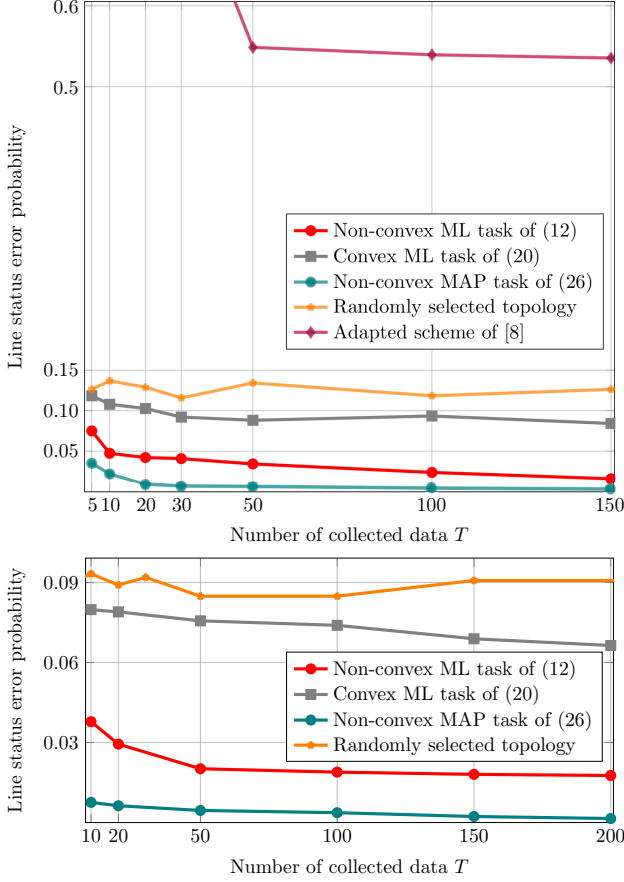


Fig. 5. Probability of line status errors for the IEEE 37-bus (top) and the IEEE 123-bus (bottom) feeders with $T_s = 5$ min.

stop procedure described earlier.

Our algorithms were compared with a random selection of an admissible topology, and an adapted version of the scheme of [8]. The latter computes a pseudo-inverse of the sample voltage covariance. The grid topology is identified by passing the signs of the entries in the latter matrix to a spanning tree algorithm. Since this is an identification scheme and to make the comparison more fair, we checked only existing lines (energized or inactive). The scheme was tested only for the 37-bus feeder, because it cannot not handle disjoint feeders.

The average number of line errors for the two feeders is depicted in Fig. 5. It is worth emphasizing that reliable line verification can be obtained even for $T < N$, when the sample covariance $\hat{\Sigma}$ is *singular*. The performance of the scheme from [8] is worse than that of a randomly selected topology, since the covariance matrix is away from its ensemble value and line impedances have not been exploited. Figure 6 tests the effect of the data sampling period on the line status error probability

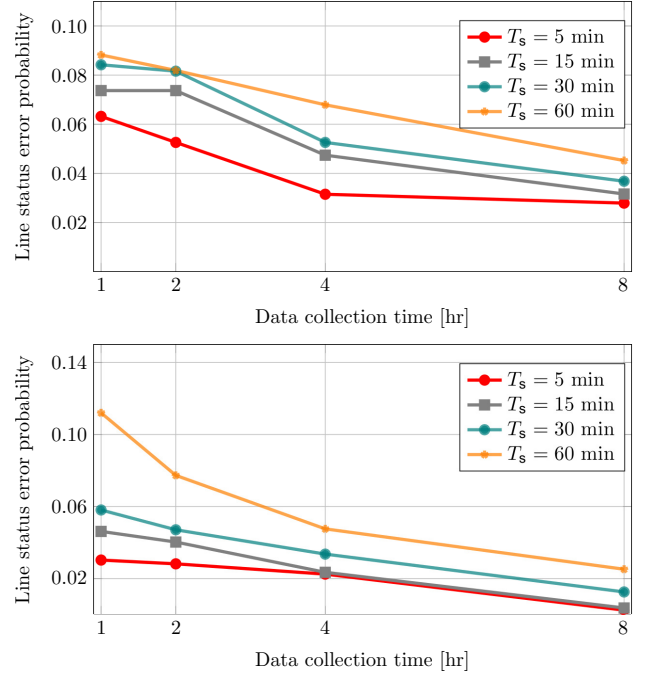


Fig. 6. Probability of line status errors for the IEEE 37-bus (top) and the IEEE 123-bus (bottom) feeders for varying sampling periods T_s .

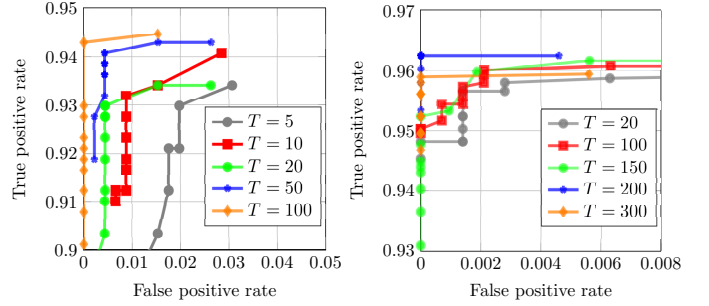


Fig. 7. True- versus false-positive rates using varying truncation thresholds for the IEEE 37-bus (left) and 123-bus (right) grids with $T_s = 5$ min.

achieved by the non-convex ML task of (11). The performance improves as T_s and the total collection time increase.

The MAP approach of Sec. IV-C was finally tested via 50 Monte Carlo runs. The prior probabilities π_ℓ were set to 0.5 for switches and 0.9 for lines. A stationary point of (25) was found through the PGD iterates of (26)–(27) after random initialization within $[0, 1]$. The entries of the PGD solution were truncated to binary upon thresholding. Varying the threshold from 0.05 to 0.60 in increments of 0.05 resulted in detection curves of Fig. 7. The threshold yielding the point closer to the top-left corner of these plots was used in the MAP curves of Fig. 5 and the related running times and step sizes are listed in Table III for $T = 50$. A direct comparison between the ML and MAP estimates may not be fair since the latter uses prior information on lines. The previous tests included both radial and meshed grid configurations. Table IV reports the line status error probabilities obtained separately for the two configuration types, and demonstrates there is no significant difference in the performance of the developed approaches.

TABLE IV
LINE STATUS ERROR PROBABILITY FOR THE IEEE 37-BUS GRID WITH
 $T_s = 5$ MIN AND $T = 100$.

Configuration	Radial	Meshed
Non-convex ML task of (11)	0.0237	0.0234
Convex ML task of (19)	0.0884	0.0895
Non-convex ML task of (25)	0.0039	0.0047

VI. CONCLUSIONS

Distribution grid topology verification has been posed as a statistical inference problem. By judiciously expressing the observed voltage magnitudes as functions of the underlying grid topology, maximum likelihood and maximum *a posteriori* probability detection schemes have been proposed. The novel learning tasks minimize (non)-convex objectives depending on the accuracy of the adopted grid data model. Either way, the optimization is over the non-convex feasible set of active line configurations that is relaxed to its convex hull. Solvers of complementary strengths have been devised. Numerical tests using real data on benchmark feeders demonstrate that our schemes perform well even when the number of data is smaller than the network size. Depending on prior information and load activity, collecting smart meter data over 0.5–3 hours can verify topologies over a hundred of nodes.

Extending the approach to multiphase grids and dynamically selecting meters constitute interesting research directions. Setups where voltage data are collected over a subset of buses and/or in the presence of zero-injection buses are practically relevant and need to be addressed. Coping with the task of topology identification where line parameters are unknown too is practically relevant too. To improve scalability, second-order (Newton) and/or accelerated optimization techniques could be pursued. Since data arrive sequentially, the iterates could be initialized to their most recent value.

APPENDIX

Proof of Lemma 1: The partial derivatives $\frac{\partial f(\mathbf{b})}{\partial b_\ell}$ for $\ell \in \mathcal{L}_e$ will be derived using the matrix differentiation rules [34]:

- (r1) $\partial(\mathbf{X}\mathbf{Y}) = (\partial\mathbf{X})\mathbf{Y} + \mathbf{X}(\partial\mathbf{Y})$;
- (r2) $\partial(\log |\mathbf{X}|) = \text{Tr}(\mathbf{X}^{-1}\partial\mathbf{X})$; and
- (r3) $\partial(\mathbf{X}^{-1}) = -\mathbf{X}^{-1}(\partial\mathbf{X})\mathbf{X}^{-1}$.

Rules (r2) and (r3) together with the commutativity property of the trace provide that

$$\frac{\partial f(\mathbf{b})}{\partial b_\ell} = \text{Tr} \left(\mathbf{F}(\mathbf{b}) \frac{\partial \Sigma(\mathbf{b})}{\partial b_\ell} \right) \quad (28)$$

where $\mathbf{F}(\mathbf{b}) := \Sigma^{-1}(\mathbf{b}) - \Sigma^{-1}(\mathbf{b})\hat{\Sigma}\Sigma^{-1}(\mathbf{b})$.

From the definition of $\Sigma(\mathbf{b})$ in (7) and (r1), it follows

$$\begin{aligned} \frac{\partial \Sigma(\mathbf{b})}{\partial b_\ell} &= \frac{\partial \mathbf{R}(\mathbf{b})}{\partial b_\ell} \Sigma_p \mathbf{R}(\mathbf{b}) + \mathbf{R}(\mathbf{b}) \Sigma_p \frac{\partial \mathbf{R}(\mathbf{b})}{\partial b_\ell} \\ &\quad + \frac{\partial \mathbf{X}(\mathbf{b})}{\partial b_\ell} \Sigma_q \mathbf{X}(\mathbf{b}) + \mathbf{X}(\mathbf{b}) \Sigma_q \frac{\partial \mathbf{X}(\mathbf{b})}{\partial b_\ell} \\ &\quad + \frac{\partial \mathbf{R}(\mathbf{b})}{\partial b_\ell} \Sigma_{pq} \mathbf{X}(\mathbf{b}) + \mathbf{R}(\mathbf{b}) \Sigma_{pq} \frac{\partial \mathbf{X}(\mathbf{b})}{\partial b_\ell} \\ &\quad + \frac{\partial \mathbf{X}(\mathbf{b})}{\partial b_\ell} \Sigma_{pq}^\top \mathbf{R}(\mathbf{b}) + \mathbf{X}(\mathbf{b}) \Sigma_{pq}^\top \frac{\partial \mathbf{R}(\mathbf{b})}{\partial b_\ell}. \end{aligned} \quad (29)$$

Using again rule (r3) for the matrix inverse differential yields

$$\frac{\partial \mathbf{R}(\mathbf{b})}{\partial b_\ell} = -\frac{1}{2r_\ell} \mathbf{R}(\mathbf{b}) \mathbf{a}_\ell \mathbf{a}_\ell^\top \mathbf{R}(\mathbf{b}). \quad (30)$$

A similar expression holds for $\frac{\partial \mathbf{X}(\mathbf{b})}{\partial b_\ell}$. Plugging (29)–(30) into (28) and exploiting the symmetry of the matrices and the commutativity of the trace yield the result. ■

Proof of Lemma 2: It can be easily verified that $\mathbf{F}(\mathbf{b}_o) = \mathbf{0}$ when $\hat{\Sigma} = \Sigma(\mathbf{b}_o)$, and therefore, $\nabla f(\mathbf{b}_o) = \mathbf{0}$ from (14). Hence, the vector \mathbf{b}_o is a stationary point of $f(\mathbf{b}_o)$.

It is next proved that \mathbf{b}_o is also a global minimizer for problems (10) and (11). It can be shown that the Kullback-Leibler divergence between two Gaussian probability density functions $p_i(\mathbf{x}) = \mathcal{N}(\mathbf{0}, \Sigma_i)$ for $i = 1, 2$, with $\mathbf{x} \in \mathbb{R}^N$ is [35, Ex. 2.13]

$$\begin{aligned} D(p_1||p_2) &:= \int p_1(\mathbf{x}) \log \left[\frac{p_1(\mathbf{x})}{p_2(\mathbf{x})} \right] d\mathbf{x} \\ &= \frac{1}{2} \left[\log \frac{|\Sigma_2|}{|\Sigma_1|} - N + \text{Tr}(\Sigma_2^{-1}\Sigma_1) \right]. \end{aligned}$$

Recall also that $D(p_1||p_2) \geq 0$. Since $\Sigma(\mathbf{b}) \succ \mathbf{0}$ for all $\mathbf{b} \in \mathcal{B}_r$ attaining a finite value, matrix $\Sigma(\mathbf{b})$ qualifies as a covariance matrix. Selecting Σ_2 to take the parametric form $\Sigma(\mathbf{b})$ and fixing Σ_1 to the given matrix $\Sigma(\mathbf{b}_o) \succ \mathbf{0}$, it follows that the objective in (10) is an affine transformation of $D(p_1||p_2)$ as $f(\mathbf{b}) = 2D(p_1||p_2) + \log |\Sigma(\mathbf{b}_o)| + N$. Since $D(p_1||p_2)$ attains zero only if $\Sigma_1 = \Sigma_2$, the minimum value of $f(\mathbf{b})$ over all $\mathbf{b} \in \mathcal{B}_r$ is hence achieved at $\mathbf{b} = \mathbf{b}_o$. ■

Proof of Lemma 3: By definition of $\mathbf{X}(\mathbf{b})$, the matrix $\mathbf{X}^{-1}(\mathbf{b}) = \sum_{\ell \in \mathcal{L}_e} \frac{b_\ell}{x_\ell} \mathbf{a}_\ell \mathbf{a}_\ell^\top$ is symmetric positive semi-definite and depends linearly on $\mathbf{b} \in \mathbb{R}_+^L$. Although there exist \mathbf{b} for which the matrix $\sum_{\ell \in \mathcal{L}_e} \frac{b_\ell}{x_\ell} \mathbf{a}_\ell \mathbf{a}_\ell^\top$ becomes singular, these points are practically avoided while minimizing $\tilde{f}(\mathbf{b})$ since the term $-2 \log |\mathbf{X}^{-1}(\mathbf{b})|$ tends to infinity.

The negative log-determinant of a positive definite matrix is known to be a strictly convex function [28]. The convexity of the second term of $\tilde{f}(\mathbf{b})$ stems from the equivalence:

$$\begin{aligned} \text{Tr}(\mathbf{Y}\Sigma_\alpha^{-1}\mathbf{Y}\hat{\Sigma}) &= \min_{\mathbf{Z}} \text{Tr}(\mathbf{Z}) \\ \text{s.t. } &\begin{bmatrix} \Sigma_\alpha^{1/2}\mathbf{Z}\Sigma_\alpha^{1/2} & \mathbf{Y}\hat{\Sigma}^{1/2} \\ \hat{\Sigma}^{1/2}\mathbf{Y} & \mathbf{I} \end{bmatrix} \succeq \mathbf{0} \end{aligned} \quad (31)$$

where $\hat{\Sigma}^{1/2}$ is the square root of $\hat{\Sigma}$ and likewise for $\Sigma_\alpha^{1/2}$. Recall that although Σ_α is invertible, the sample covariance $\hat{\Sigma}$ may be singular, for example when $T < N$.

To prove the equivalence in (31), heed that by Schur's complement the constraint yields $\Sigma_\alpha^{1/2}\mathbf{Z}\Sigma_\alpha^{1/2} \succeq \mathbf{Y}\hat{\Sigma}\mathbf{Y}$, or equivalently, $\mathbf{Z} \succeq \Sigma_\alpha^{-1/2}\mathbf{Y}\hat{\Sigma}\mathbf{Y}\Sigma_\alpha^{-1/2}$. Any feasible \mathbf{Z} can be expressed as $\mathbf{Z} = \Sigma_\alpha^{-1/2}\mathbf{Y}\hat{\Sigma}\mathbf{Y}\Sigma_\alpha^{-1/2} + \mathbf{W}$ for some \mathbf{W} . Because $\mathbf{Z} \succeq \Sigma_\alpha^{-1/2}\mathbf{Y}\hat{\Sigma}\mathbf{Y}\Sigma_\alpha^{-1/2}$, it holds that $\mathbf{W} \succeq \mathbf{0}$. Then, the trace of \mathbf{Z} is strictly larger than $\text{Tr}(\mathbf{Y}\Sigma_\alpha^{-1}\mathbf{Y}\hat{\Sigma})$ unless $\mathbf{W} = \mathbf{0}$. The equivalence follows since the minimizer of (31) is $\mathbf{Z}^* = \Sigma_\alpha^{-1/2}\mathbf{Y}\hat{\Sigma}\mathbf{Y}\Sigma_\alpha^{-1/2}$.

The constraint in (31) is a linear matrix inequality with respect to (\mathbf{Y}, \mathbf{Z}) , and therefore, convex. Function $\text{Tr}(\mathbf{Y}\Sigma_\alpha^{-1}\mathbf{Y}\hat{\Sigma})$ is convex in \mathbf{Y} as the partial minimization of a jointly convex problem over (\mathbf{Y}, \mathbf{Z}) . ■

REFERENCES

- [1] A. Gómez-Expósito, A. J. Conejo, and C. Canizares, Eds., *Electric Energy Systems, Analysis and Operation*. Boca Raton, FL: CRC Press, 2009.
- [2] J. E. Tate and T. J. Overbye, "Double line outage detection using phasor angle measurements," in *Proc. of IEEE Power & Energy Society General Meeting*, Calgary, Alberta, Canada, Jul. 2009.
- [3] H. Zhu and G. B. Giannakis, "Sparse overcomplete representations for efficient identification of power line outages," *IEEE Trans. Power Syst.*, vol. 27, no. 4, pp. 2215–2224, Nov. 2012.
- [4] X. Li, V. Poor, and A. Scaglione, "Blind topology identification for power systems," in *Proc. IEEE Smart Grid Communications Conf.*, Vancouver, BC, Canada, Oct. 2013.
- [5] 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, Jul. 2011.
- [6] V. Kekatos, G. Giannakis, and R. Baldick, "Online energy price matrix factorization for power grid topology tracking," *IEEE Trans. Smart Grid*, vol. 7, no. 7, pp. 1239–1248, May 2016.
- [7] T. Erseghe, S. Tomasin, and A. Vigato, "Topology estimation for smart micro grids via powerline communications," *IEEE Trans. Signal Process.*, vol. 61, no. 13, pp. 3368–3377, Jul. 2013.
- [8] 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. on Decision and Control*, Florence, Italy, Dec. 2013, pp. 1659–1664.
- [9] D. Deka, M. Chertkov, and S. Backhaus, "Structure learning in power distribution networks," *IEEE Trans. Control Netw. Syst.*, vol. PP, no. 99, pp. 1–1, 2017.
- [10] D. Deka, S. Backhaus, and M. Chertkov, "Tractable structure learning in radial physical flow networks," 2016, (submitted). [Online]. Available: <http://arxiv.org/abs/1608.05064>
- [11] Y. Weng, Y. Liao, and R. Rajagopal, "Distributed energy resources topology identification via graphical modeling," *IEEE Trans. Power Syst.*, 2017, early access.
- [12] R. Arghandeh, M. Gahr, A. von Meier, G. Cavraro, M. Ruh, and G. Andersson, "Topology detection in microgrids with micro-synchrophasors," in *Proc. IEEE Power & Energy Society General Meeting*, Denver, CO, Jul. 2015.
- [13] G. Cavraro, R. Arghandeh, K. Poolla, and A. von Meier, "Data-driven approach for distribution network topology detection," in *Proc. IEEE Power & Energy Society General Meeting*, Denver, CO, Jul. 2015.
- [14] Y. Liao, Y. Weng, and R. Rajagopal, "Urban distribution grid topology reconstruction via Lasso," in *Proc. IEEE Power & Energy Society General Meeting*, Boston, MA, Jul. 2016.
- [15] Y. Yuan, O. Ardakanian, S. Low, and C. Tomlin, "On the inverse power flow problem," Oct. 2016, preprint. [Online]. Available: <https://arxiv.org/abs/1610.06631>
- [16] M. Baran and F. Wu, "Network reconfiguration in distribution systems for loss reduction and load balancing," *IEEE Trans. Power Del.*, vol. 4, no. 2, pp. 1401–1407, Apr. 1989.
- [17] S. Bolognani and F. Dörfler, "Fast power system analysis via implicit linearization of the power flow manifold," in *Proc. Allerton Conf.*, Allerton, IL, Sep. 2015, pp. 402–409.
- [18] C. Godsil and G. Royle, *Algebraic Graph Theory*. New York, NY: Springer, 2001.
- [19] V. Kekatos, L. Zhang, G. B. Giannakis, and R. Baldick, "Voltage regulation algorithms for multiphase power distribution grids," *IEEE Trans. Power Syst.*, vol. 31, no. 5, pp. 3913–3923, Sep. 2016.
- [20] M. Farivar, L. Chen, and S. Low, "Equilibrium and dynamics of local voltage control in distribution systems," in *Proc. IEEE Conf. on Decision and Control*, Florence, Italy, Dec. 2013, pp. 4329–4334.
- [21] L. Gan, N. Li, U. Topcu, and S. Low, "On the exactness of convex relaxation for optimal power flow in tree networks," in *Proc. IEEE Conf. on Decision and Control*, Maui, HI, Dec. 2012, pp. 465–471.
- [22] 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.
- [23] V. Chernozhukov, D. Chetverikov, and K. Kato, "Central limit theorems and bootstrap in high dimensions," 2016, (under review). [Online]. Available: <https://arxiv.org/abs/1412.3661>
- [24] D. P. Bertsekas, *Convex Optimization Algorithms*. Belmont, MA: Athena Scientific, 2015.
- [25] S. M. Kay, *Fundamentals of Statistical Signal Processing, Vol. I: Estimation Theory*. Prentice Hall, 1993.
- [26] L. Zhang, V. Kekatos, and G. B. Giannakis, "A generalized Frank-Wolfe approach to decentralized electric vehicle charging," in *Proc. IEEE Conf. on Decision and Control*, Las Vegas, NV, Dec. 2016.
- [27] M. Jaggi, "Revisiting Frank-Wolfe: Projection-free sparse convex optimization," in *Proc. Intl. Conf. on Machine Learning*, Atlanta, GA, Jun. 2013, pp. 427–435.
- [28] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY: Cambridge University Press, 2004.
- [29] W. H. Kersting, *Distribution System Modeling and Analysis*. New York, NY: CRC Press, 2001.
- [30] (2013) Pecan Street Inc. Dataport. [Online]. Available: <https://dataport.pecanstreet.org/>
- [31] A. Samadi, R. Eriksson, L. Soder, B. G. Rawn, and J. C. Boemer, "Coordinated active power-dependent voltage regulation in distribution grids with PV systems," *IEEE Trans. Power Del.*, vol. 29, no. 3, pp. 1454–1464, Jun. 2014.
- [32] EEI-AEIC-UTC. (2011) Smart meters and smart meter systems: A metering industry perspective. [Online]. Available: <http://www.eei.org/issuesandpolicy/grid-enhancements/Documents/smartmeters.pdf>
- [33] S. Lacoste-Julien, "Convergence rate of Frank-Wolfe for non-convex objectives," 2016, preprint. [Online]. Available: <https://arxiv.org/abs/1607.00345>
- [34] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge University Press, 1991.
- [35] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY: Springer, 2006.



Guido Cavraro received the B.S. degree in Information Engr., the M.S. degree in Automation Engineering, and the Ph.D. degree in Information Engr. from the Univ. of Padova, Italy, in 2008, 2011, and 2015, respectively. In 2015 and 2016, he was a postdoctoral associate at Information Engr. Dept. of the Univ. of Padova. He is currently a postdoctoral associate at the Bradley Dept. of ECE at Virginia Tech. His current research interests include identification, control and optimization applied to power systems and smart grids.



Vassilis Kekatos (SM'16) is an Assistant Professor of ECE at the Bradley Dept. of ECE at Virginia Tech. He obtained his Diploma, M.Sc., and Ph.D. in computer science and engr. from the Univ. of Patras, Greece, in 2001, 2003, and 2007, respectively. He was a recipient of the Marie Curie Fellowship during 2009–2012, and a research associate with the ECE Dept. at the Univ. of Minnesota, where he received the postdoctoral career development award (honorable mention). During 2014, he stayed with the Univ. of Texas at Austin and the Ohio State Univ. as a visiting researcher. His research focus is on optimization and learning for future energy systems. He is currently serving in the editorial board of the IEEE Trans. on Smart Grid.



Sriharsha Veeramachaneni is a researcher at Windlogics Inc., a subsidiary of NextEra Energy Inc., specializing in statistics, machine learning and optimization. He holds a Ph.D. in computer engineering from the Rensselaer Polytechnic Institute.