2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Leveraging the Physics of AC Power Flow in Support Vector Regression to Identify Power System Topology

Elizabeth Glista and Somayeh Sojoudi
University of California, Berkeley

Abstract— **Understanding an electric power system's topology, including both its nodal connectivity and physical parameters, is critically important to the reliable operation and control of the power grid. In cases where this power system topology may be unavailable, due to data collection deficiencies, real-time line switching, or intentional cyberattacks, it is important to be able to estimate the real power system topology with high accuracy. In this paper, we propose a new data-driven constrained support vector regression (SVR) method that aims to map voltage data collected from phasor measurement units (PMUs) to data collected by Supervisory Data Acquisition and Control (SCADA) systems. We show that the dual of the constrained SVR model can be formulated as a quadratic program (QP) and solved efficiently with off-the-shelf solvers. Testing our method on standard IEEE test cases, we demonstrate that our proposed method significantly outperforms existing state-of-the-art SVR methods in learning the true network topology, even in the presence of measurement noise, outliers, and missing data.**

I. Introduction

With the adoption of new smart grid technologies such as smart meters, distributed and renewable generation, and an increased prevalence of phasor measurement units (PMUs) in power networks, the optimization and monitoring methods of legacy power systems will need to evolve [1]. Additionally, as the risk of cyberattacks grows, many of the classical optimization problems such as state estimation (SE) take on increased importance in ensuring the reliability of the electric grid [2], [3]. The safe and effective operation of the power grid depends on solving a variety of optimization and control problems, including SE, power flow (PF), optimal power flow (OPF), unit commitment (UC), false data detection (FDD), and voltage control [4]. The fundamental bases for many of these problems are the nonlinear AC power flow equations, which define the physics of power flow in a network. Using the power flow equations in most applications relies on an awareness of the power network topology and line parameters. However, knowledge of the network topology and/or the line parameters can be limited due to cyberattacks, real-time topology switching, or other data collection deficiencies and inaccuracies [5]. In order to deal with these uncertainties, we consider a system identification problem that aims to learn the power flow mapping. By learning the power flow mapping, the goal is to recover the true system parameters and topology.

A variety of machine learning and optimization methods have been proposed to learn the power flow mapping in a generic network, including neural networks (NN) [6], [7], graph neural networks (GNNs) [8], and support vector regression (SVR) [5], [9]. Many of these data-driven methods exploit the abundance of power system data from both traditional SCADA measurements and PMUs to learn the mapping [7]. However, these black- and gray-box methods suffer from overfitting and lack a physical representation in the power network. These methods do not explicitly make use of the sparsity inherent in power networks, which has been effectively exploited in other power applications such as OPF to efficiently solve hard, nonconvex problems [10].

While recent papers show that SVR can be effective at learning forward and inverse mappings between power system inputs and outputs, these papers apply classic SVR methods that fail to recover the true system parameters because they do not account for power network sparsity [5], [9]. In [5], it is shown that the power flow equations can be exactly written as a quadratic kernel within the reproducing kernel Hilbert space (RKHS). However, we will demonstrate later that the feature vector corresponding this RKHS contains many features that do not contribute to the power flow equations and should be associated with parameters equal to zero. However, in [5] and [9], the authors' methods do not ensure that these parameters are zero and will thus recover a dense parameter set that has no realistic physical meaning. In this paper, we fix these shortcomings in the existing literature by proposing a new constrained SVR method that considers the actual power network sparsity. We show that this method can recover the true physical parameters and topology of a power system with high accuracy.

A. Support Vector Regression (SVR) and Prior Knowledge

SVR was developed in the 1990s as an extension of the support vector machine (SVM) classification learning algorithm [11], [12]. The idea of SVM is to find a hyperplane decision boundary that maximizes the margin between differently classified sets of data [11]. The fast implementation of SVM in nonlinear settings relies on the kernel trick which maps nonlinear features into a high-dimensional space that corresponds to a linear classifier. SVR is the regression extension of SVM. While SVM aims to find a classification hyperplane that minimizes data proximity to the plane, SVR

Elizabeth Glista is with Lawrence Livermore National Laboratory (LLNL), and Somayeh Sojoudi is with the Department of Mechanical Engineering at the University of California, Berkeley. Emails: glista1@llnl.gov, sojoudi@berkeley.edu.

This work was supported by grants from NSF, AFOSR and ONR.

aims to find a linear estimator that maximizes data proximity around the estimator, penalizing data points outside of an ϵ-tube around the estimator [13]. Like SVM, SVR makes use of the kernel trick to efficiently estimate nonlinear functions.

SVM and SVR have been shown to be effective in a variety of nonlinear applications, including image classification and load forecasting [14], [15]. However, these methods do not make use of prior information known about the space of the estimator, which may result in estimators that do not accurately represent the corresponding systems. Recent research has considered a constrained SVR problem in the case where both the kernel and the constraints are linear [16]. The authors of [16] show that their constrained SVR method can recover better estimators in terms of root-mean-square error (RMSE) compared to both classic SVR and other constrained regression methods, such as constrained least squares, in various biomedical and weather data settings where information is known *a priori* about the space of the estimator. Building on the work in [16], we extend the constrained SVR method to a nonlinear kernel, which allows for the use of constrained SVR in a much broader range of realistic applications. While our main focus is the power flow mapping application, the proposed constrained SVR methodology could be applied to other network mapping problems as well as more general supervised learning settings in which information is known *a priori* about the system.

B. Notations

The symbols $\mathbb{R}$ and $\mathbb{C}$ denote the sets of real and complex numbers, respectively. $\mathbb{R}^N$ and $\mathbb{C}^N$ denote the spaces of N-dimensional real and complex vectors, respectively. The symbol $\mathbb{R}^N_+$ denotes the space of real vectors with non-negative entries. The symbol $\mathbb{S}^N$ denotes the space of $N \times N$ symmetric real matrices. The symbols $(\cdot)^T$ and $(\cdot)^*$ denote the transpose and conjugate transpose of a vector or matrix. $\mathrm{Re}\{\cdot\}$ and $\mathrm{Im}\{\cdot\}$ denote the real and imaginary part of a given scalar or matrix. The symbol $|\cdot|$ is the absolute value operator if the argument is a scalar, vector, or matrix; otherwise, it is the cardinality of a measurable set. The imaginary unit is denoted by $\mathbf{j} = \sqrt{-1}$.

II. Problem Background

In this section, we present the mathematical formulation of the power flow mapping problem.

A. Alternating Current (AC) Power Flow Mapping

Let the power network be defined by a graph $(\mathcal{B}, \mathcal{L})$, where $\mathcal{B}$ is the set of buses and $\mathcal{L}$ is the set of transmission or distribution lines. Let $\mathcal{G} \subseteq \mathcal{B}$ be the buses that are attached to generators. The equations that govern AC power flow between buses in the network are given as:

$$p_{ij} = |v_i||v_j|\big(G_{ij}\cos\theta_{ij} + B_{ij}\sin\theta_{ij}\big) \quad \forall (i,j) \in \mathcal{L} \quad (1a)$$
$$q_{ij} = |v_i||v_j|\big(G_{ij}\sin\theta_{ij} - B_{ij}\cos\theta_{ij}\big) \quad \forall (i,j) \in \mathcal{L} \quad (1b)$$

where the complex voltage at each bus i is given as $v_i \triangleq |v_i|e^{\mathbf{j}\theta_i} \in \mathbb{C}$. The expressions p_{ij} and q_{ij} respectively represent the real and reactive power flows between buses i and j. The expression θ_{ij} is the difference in voltage angle between buses i and j, given as $\theta_{ij} \triangleq \theta_i - \theta_j$. The network parameters G_{ij} and B_{ij} are respectively the conductance and susceptance for the line between buses i and j, where the complex admittance is given as $Y_{ij} = G_{ij} + \mathbf{j}B_{ij}$.

In the power flow mapping problem, the conductance G_{ij} and susceptance B_{ij} are taken to be part of the unknown parameter set, and some subset of possible SCADA measurements $\mathcal{M}$ are assumed to be available. These measurements can consist of real power flows p_{ij}, reactive power flows q_{ij}, real power injections p_i, and reactive power injections q_i. Additionally, we have PMU readings at some buses that provide estimates for voltage magnitudes $|v_i|$ and voltage angles θ_i. We allow for the case where the network may be only partially observable as discussed in Section IV-B.

Note that while this formulation does assume some awareness of the power network topology, provided by the graph $(\mathcal{B}, \mathcal{L})$, we can model uncertainty in some portion of the topology using only the line parameters. To accomplish this, we consider $\mathcal{L}$ to include all possible connected lines in the network. If the line (i,j) is switched off or does not actually exist in the network, the line parameters G_{ij} and B_{ij} will be zero. Thus, we can model uncertainty in the network topology by considering only the line parameters as unknowns. However, we will assume that there is some baseline understanding of the network topology, including some awareness of the number of buses in the network and how they are interconnected. This is a reasonable assumption for most realistic test cases in which a system operator would have full understanding of the baseline network topology but might not be aware of lines that have switched open or closed due to information delay or cyberattacks.

In order to formulate the power flow equations (1) as a kernel within the RKHS so that the kernel trick can be used for SVR, we introduce new variables $d_i \triangleq |v_i|\cos\theta_i$ and $e_i \triangleq |v_i|\sin\theta_i$ for all $i \in \mathcal{B}$ that correspond to the rectangular coordinates of the complex voltage. Then, using trigonometric identities, we can rewrite the power flow equations as:

$$p_{ij} = G_{ij}(d_i d_j + e_i e_j) + B_{ij}(e_i d_j - d_i e_j) \quad (2a)$$
$$q_{ij} = G_{ij}(e_i d_j - d_i e_j) - B_{ij}(d_i d_j + e_i e_j) \quad (2b)$$
$$\forall (i,j) \in \mathcal{L}$$

We can do the same for real and reactive power injections to get the following relations:

$$p_i = G_{ii}d_i^2 + G_{ii}e_i^2 + \textstyle\sum_{j\neq i} G_{ij}(d_i d_j + e_i e_j) + B_{ij}(e_i d_j - d_i e_j) \quad (3a)$$
$$q_i = -B_{ii}d_i^2 - B_{ii}e_i^2 + \textstyle\sum_{j\neq i} G_{ij}(e_i d_j - d_i e_j) - B_{ij}(d_i d_j + e_i e_j) \quad (3b)$$

where G_{ii} is the self conductance at bus i, composed of the shunt conductance G_i^{sh} and line conductances G_{ij} as $G_{ii} = G_i^{sh} - \sum_{j\neq i} G_{ij}$. Similarly, B_{ii} is the self susceptance at bus i, composed of the shunt susceptance B_i^{sh} and line susceptances B_{ij} as $B_{ii} = B_i^{sh} - \sum_{j\neq i} B_{ij}$. These parameters are also assumed to be unknown for the network.

B. Availability of PMU Data

There is strong interest in using increasingly available PMU data in state estimation and false data detection [17]–[19]. PMUs provide measurements of voltage and current phasors and have been shown to be helpful in improving the reliability of grid monitoring and detection tools [17]. While PMU penetration in the grid is growing but still somewhat limited due to high PMU installation costs, PMUs provide a larger quantity of real-time data than existing SCADA systems. While SCADA systems collect samples about every 4 seconds, PMUs typically collect about 30-60 samples per second [17], [20]. The large amount of PMU is well-suited for machine learning applications such as the power flow mapping problem presented in the preceding section.

We consider the case where we obtain voltage magnitude $|v_i|$ and angle data θ_i from PMUs on some subset of buses in the network. The PMU data serve as the input data for our mapping problem (appearing as $\mathbf{x}$ in Equation (12)), and the SCADA data serve as the output (appearing as $\mathbf{y}$ in Equation (12)). In order to reconcile the synchronization gap between PMU and SCADA data, we associate some set of PMU data to each SCADA measurement, i.e. we take 10 PMU samples collected both before and after a given SCADA measurement and associate those to the SCADA measurement. Thus, for one time step of SCADA data, we have created duplicate $\mathbf{y}_t$ SCADA measurements corresponding to each $\mathbf{x}_t$ PMU measurement. Other types of time synchronization smoothing methods could also be used to reconcile the PMU and SCADA datasets such as averaging the PMU data or the estimation fusion approach in [21].

III. Constrained SVR Problem Formulation

In this section, we present the primal SVR formulation of the power system mapping problem and show how sparsity-enforcing constraints can be added to the classic SVR problem. Then, we find the dual of this constrained SVR problem and show that it is a convex quadratic program.

A. Power Flow as Represented by the Quadratic Kernel

In [5], it was shown that AC power flow can be represented exactly by the quadratic kernel, which is given as:

$$K(\mathbf{x}_1, \mathbf{x}_2) = (\langle \mathbf{x}_1, \mathbf{x}_2 \rangle)^2 \qquad \forall \mathbf{x} \in \mathcal{X} = \mathbb{R}^{2n} \tag{4}$$

where $n \leq |\mathcal{B}|$ corresponds to the number of buses in the network where PMU data is available and $\mathbf{x} \in \mathbb{R}^{2n}$ corresponds to the real and reactive components in the complex voltage vector $\mathbf{v} \in \mathbb{C}^n$, where $\mathbf{x}$ is given as:

$$\mathbf{x} \triangleq \begin{bmatrix} d_1 & d_2 & \dots & d_n & e_1 & e_2 & \dots & e_n \end{bmatrix}^T \tag{5}$$

Then, the feature mapping corresponding to the quadratic kernel $\phi(\mathbf{x}) \in \mathbb{R}^D$ is given by:

$$\phi(\mathbf{x}) = [d_1^2 \ \dots \ d_n^2 \ e_1^2 \ \dots \ e_n^2 \ \sqrt{2}d_1d_2 \ \dots \ \dots \ \sqrt{2}d_1e_2 \ \dots \ \sqrt{2}e_{n-1}e_n]^T \tag{6}$$

where we have that $D = \binom{2n+1}{2} = 2n^2+n$. By construction, we have that $K(\mathbf{x}_1, \mathbf{x}_2) = \phi(\mathbf{x}_1)^T\phi(\mathbf{x}_2)$, a relation that allows for the kernel trick in the dual SVR formulation, i.e. the replacement of $\phi(\mathbf{x}_1)^T\phi(\mathbf{x}_2)$ terms by $K(\mathbf{x}_1, \mathbf{x}_2)$.

Based on this formulation, we can rewrite each of the real and reactive power flow and injection measurement relations given by Equations (2) and (3) as a dot product of the quadratic feature mapping $\phi(\mathbf{x})$ given in (6) and a specific parameter vector with known structure and unknown values. The power flow measurement equations (2a) and (2b) can be written as:

$$p_{ij} = \langle \mu_{p_{ij}}, \phi(\mathbf{x}) \rangle, \quad \forall (i,j) \in \mathcal{L} \tag{7a}$$
$$q_{ij} = \langle \mu_{q_{ij}}, \phi(\mathbf{x}) \rangle, \quad \forall (i,j) \in \mathcal{L} \tag{7b}$$

where $\mu_{p_{ij}}, \mu_{q_{ij}} \in \mathbb{R}^D$ with the k^{th} entries of $\mu_{p_{ij}}$ and $\mu_{q_{ij}}$ are defined as:

$$(\mu_{p_{ij}})_k \triangleq \begin{cases} G_{ij}, \text{ if } \phi(\mathbf{x}^{ij})_k = \sqrt{2}d_id_j \text{ or } \sqrt{2}e_ie_j \\ B_{ij}, \text{ if } \phi(\mathbf{x}^{ij})_k = \sqrt{2}e_id_j \\ -B_{ij}, \text{ if } \phi(\mathbf{x}^{ij})_k = \sqrt{2}d_ie_j \\ 0, \text{ otherwise} \end{cases} \tag{8a}$$

$$(\mu_{q_{ij}})_k \triangleq \begin{cases} -B_{ij}, \text{if } \phi(\mathbf{x}^{ij})_k = \sqrt{2}d_id_j \text{ or } \sqrt{2}e_ie_j \\ G_{ij}, \text{ if } \phi(\mathbf{x}^{ij})_k = \sqrt{2}e_id_j \\ -G_{ij}, \text{ if } \phi(\mathbf{x}^{ij})_k = \sqrt{2}d_ie_j \\ 0, \text{ otherwise} \end{cases} \tag{8b}$$

Similarly, the power injection measurement equations (3a) and (3b) can be written as:

$$p_i = \langle \mu_{p_i}, \phi(\mathbf{x}) \rangle, \quad \forall i \in \mathcal{B} \tag{9a}$$
$$q_i = \langle \mu_{q_i}, \phi(\mathbf{x}) \rangle, \quad \forall i \in \mathcal{B} \tag{9b}$$

where $\mu_{p_i}, \mu_{q_i} \in \mathbb{R}^D$ with the k^{th} entries μ_{p_i} and μ_{q_i} are defined as:

$$(\mu_{p_i})_k \triangleq \begin{cases} G_{ii}, & \text{if } \phi(\mathbf{x})_k = d_i^2 \text{ or } e_i^2 \\ (\mu_{p_{ij}})_k, & \text{otherwise} \end{cases} \tag{10a}$$

$$(\mu_{q_i})_k \triangleq \begin{cases} -B_{ii}, & \text{if } \phi(\mathbf{x})_k = d_i^2 \text{ or } e_i^2 \\ (\mu_{p_{ij}})_k, & \text{otherwise} \end{cases} \tag{10b}$$

The dot product relations (7) and (9) can be shown to be respectively equivalent to (2) and (3) by expanding the equations (7) and (9) with the defined μ-parameters given by (8) and (10) and using trigonometric relations to simplify the equations. Note that the G_{ij} and B_{ij} entries should be equal to zero in the case where there is no line that connects buses i and j. However, in the case where the line topology is only partially known, we can associate G_{ij} and B_{ij} parameters to any possible line in the network.

We can observe that due to the sparsity inherent in power networks [10], most of the entries in the μ-parameters should be zero. We formalize this observation in the following lemma.

Lemma 1. *The maximum ratio of non-zero entries to zero entries in $\mu_{p_{ij}}$ or $\mu_{q_{ij}}$ is $4 : 2n^2 + n - 4$, and the maximum ratio of non-zero entries to zero entries in μ_{p_i} or μ_{q_i} is $4|\mathcal{L}(i)| + 2 : 2n^2 + n - 4|\mathcal{L}(i)| - 2$ where $\mathcal{L}(i)$ is the set of lines attached to bus i.*

Proof: In the case of real (or reactive) power flows, only 4 of the entries in each $\mu_{p_{ij}}$ (or $\mu_{q_{ij}}$) could possibly be non-zero by definition in (8). Similarly, in the case of real (or reactive) power injections, only $4|\mathcal{L}(i)|+2$ of the entries in each μ_{p_i} or μ_{q_i} could possibly be non-zero by definition in (10). The minimum number of zero entries is found by subtracting these from the number of features, given by $D = 2n^2 + n$. □

We will exploit the inherent sparsity of the μ-parameters in order to learn the network line parameters and topology via a constrained SVR method described in the sections below.

B. Constrained SVR with Multiple Measurement Types

For a complete background on support vector regression methods, see [13]. Most of the SVR methods in the literature try to learn a mapping between vectors $\mathbf{x}_t$ and scalars y_t, for multiple time steps $t = 1, \ldots, T$. However, for power flow mapping, we want to learn the relationship between vectors $\mathbf{x}_t \in \mathbb{R}^{2n}$ and vectors $\mathbf{y}_t \in \mathbb{R}^M$, for multiple time steps $t = 1, \ldots, T$, where the vector $\mathbf{y}_t$ contains multiple types of p_{ij}, q_{ij}, p_i, and q_i SCADA measurements and the mapping $\mathbf{x}_t \rightarrow \mathbf{y}_t$ is given by the equations (7) and (9). Given a set of measurements $\mathcal{M}$, where $M \triangleq |\mathcal{M}|$, and a set of data collection time steps $1, \ldots, T$, we can write the corresponding state equation model in concise form as:

$$\mathbf{y}_t = W\phi(\mathbf{x}_t) \qquad \forall t \in \{1, \ldots, T\} \tag{11}$$

where the weight matrix $W \in \mathbb{R}^{M \times D}$ relates the features corresponding to PMU voltage measurements to the SCADA measurements in $\mathcal{M}$. Each row in matrix W corresponds to a μ-parameter defined in Equations (8) and (10). In Lemma 1, we observed that these μ-parameters are very sparse, which translates to W having this same sparse structure. We will take this sparsity pattern to be $E \in \mathbb{R}^{M \times D}$, which is composed of a sparse set of variables as defined in the next section, noting that E corresponds to the baseline grid model that is flexible enough to allow for unknown topology as discussed in Section II-A. We want to enforce this sparsity pattern on the weight matrix W, which provides the constraint $W - E = 0$. Combining this constraint with the SVR model, we arrive at the primal version of the constrained SVR problem:

$$\min_{W,\xi,E} \ \frac{1}{2}||W||_F^2 + C\sum_{m=1}^{M}\sum_{t=1}^{T}(\xi_{m,t} + \xi^*_{m,t}) \tag{12a}$$

$$\text{s.t. } \mathbf{y}_t - W\phi(\mathbf{x}_t) \leq \epsilon + \xi_t, \ \ \forall t \in \{1, \ldots, T\} \tag{12b}$$

$$W\phi(\mathbf{x}_t) - \mathbf{y}_t \leq \epsilon + \xi^*_t, \ \ \forall t \in \{1, \ldots, T\} \tag{12c}$$

$$\xi_t, \xi^*_t \geq 0, \ \ \forall t \in \{1, \ldots, T\} \tag{12d}$$

$$W - E = 0 \tag{12e}$$

where $\epsilon \in \mathbb{R}^M$, $\epsilon_i \geq 0$ for all $i \in \{1, \ldots, M\}$ defines the ϵ-tube around the estimator inside which errors are not penalized. The variables $\xi_t, \xi^*_t \in \mathbb{R}^M$ defined for all $t \in \{1, \ldots, T\}$ are the penalty terms for violating the state equations (11) outside of the ϵ-tube. A linear penalty on these errors, scaled by the hyperparameter $C > 0$, is added to the Frobenius norm of the weight matrix W to form the objective function. The inclusion of the Frobenius norm in the objective serves to encourage flatness in the weight parameters, thus promoting parameters that are physically realistic. The hyperparameter C determines the trade-off between the bi-objectives of promoting flatness and minimizing violations of the state equations.

Our modification from the classical SVR problem that maps $\mathbf{x}_t$ to $y_t \in \mathbb{R}$ is that the weight vector becomes a matrix given by W and ξ_t, ξ^*_t and ϵ are now vectors in $\mathbb{R}^M$. If we ignore the sparsity constraint (12e), we can decouple this matrix version of the SVR problem (12) into the standard vector-version SVR problems for each measurement type y_m, for all $m = 1, \ldots, M$. Thus, the matrix $E \in \mathbb{R}^{M \times D}$ serves to couple the different SVR problems that correspond to each measurement type in the vector $\mathbf{y}_t$. Intuitively, this matrix E allows us to use information obtained about the parameters corresponding to one measurement type to learn the parameters corresponding to a different measurement type. For example, the line conductance G_{ij} of some line $(i,j) \in \mathcal{L}$ appears in the measurement equations for both p_{ij} and q_i (as well as q_{ij} and p_i), thus its predicted value should be the same whether we predict the mapping from $\mathbf{x}$ to p_{ij} or from $\mathbf{x}$ to q_i.

C. Defining the Sparsity Pattern on a Two-Bus Network

To define structure of the sparsity pattern E, let us first consider a two-bus network, with one line connecting buses 1 and 2. We consider the case where $\mathbf{y}_t \in \mathbb{R}^M$ with $M = 4$ that corresponds to the following measurement vectors for each time step $t \in \{1, \ldots, T\}$:

$$\mathbf{y} = \begin{bmatrix} p_{12} & q_{12} & p_1 & q_1 \end{bmatrix}^T \tag{13}$$

Using the quadratic feature vector given by (6), where $\phi(\mathbf{x}) \in \mathbb{R}^D$ with $D = 10$ corresponding to the two-bus network, and the sparse relations for the μ-parameters given by (8) and (10), we know that W will have the following sparsity pattern E:

$$E \triangleq \begin{bmatrix} 0 & 0 & 0 & 0 & G_{12} & 0 & B_{12} & -B_{12} & 0 & G_{12} \\ 0 & 0 & 0 & 0 & B_{12} & 0 & -G_{12} & G_{12} & 0 & B_{12} \\ G_{11} & 0 & G_{11} & 0 & G_{12} & 0 & B_{12} & -B_{12} & 0 & G_{12} \\ -B_{11} & 0 & -B_{11} & 0 & B_{12} & 0 & -G_{12} & G_{12} & 0 & B_{12} \end{bmatrix} \tag{14}$$

We have used Equations (8a), (8b), (10a), and (10b) to respectively construct rows 1, 2, 3, and 4 of matrix E corresponding to the measurements p_{12}, q_{12}, p_1, and q_1. Construction of the matrix E depends solely on the choice of measurement set and the baseline network topology, both of which are known *a priori*. Note that all the G and B terms in matrix E are unknown parameters. Using this example definition of E, we can formulate the dual of (12) for the two-bus network example. We will then generalize our formulation to larger networks.

D. Formulating the Dual of the Constrained SVR Problem

In the primal constrained SVR problem (12), the constraints (12b) and (12c) will slow down a generic quadratic program solver due to the high dimension of the feature

vector $\phi(\mathbf{x})$, which scales with n^2. Thus, it is more computationally useful to consider the dual form of (12), making use of the kernel trick to eliminate most instances of $\phi(\mathbf{x})$.

In order to take the dual of (12), we introduce the Lagrange multipliers $\alpha_t, \alpha_t^* \in \mathbb{R}_+^M$ corresponding to the (12b) and (12c) inequality constraints, $\beta_t, \beta_t^* \in \mathbb{R}_+^M$ corresponding to the (12d) inequality constraints, and $\lambda \in \mathbb{R}^{M\times D}$ corresponding to the (12e) equality constraints. Then, the Lagrangian of (12) can be written as:

$$\begin{aligned} L = \frac{1}{2}||W||_F^2 + C\sum_{m=1}^{M}\sum_{t=1}^{T}(\xi_{m,t}+\xi_{m,t}^*) - \sum_{t=1}^{T}(\alpha_t+\alpha_t^*)^T\epsilon \\ + \sum_{t=1}^{T}(\alpha_t-\alpha_t^*)^T(\mathbf{y}_t - W\phi(\mathbf{x}_t)) + \text{trace}\{\lambda^T(E-W)\} \\ - \sum_{t=1}^{T}\left(\alpha_t^T\xi_t + {\beta_t}^T\xi_t + (\alpha_t^*)^T\xi_t^* + (\beta_t^*)^T\xi_t^*\right) \quad (15) \end{aligned}$$

For the two-bus network example, given Equation (14) defining E, we can expand the trace$\{\lambda^T E\}$ term as:

$$\begin{aligned} \text{trace}\{\lambda^T E\} &= \lambda_{1,5}G_{12} + \lambda_{1,7}B_{12} - \lambda_{1,8}B_{12} + \lambda_{1,10}G_{12} \\ &+ \lambda_{2,5}B_{12} - \lambda_{2,7}G_{12} + \lambda_{2,8}G_{12} + \lambda_{2,10}B_{12} \\ &+ \lambda_{3,1}G_{11} + \lambda_{3,3}G_{11} + \lambda_{3,5}G_{12} + \lambda_{3,7}B_{12} \\ &- \lambda_{3,8}B_{12} + \lambda_{3,10}G_{12} - \lambda_{4,1}B_{11} - \lambda_{4,3}B_{11} \\ &+ \lambda_{4,5}B_{12} - \lambda_{4,7}G_{12} + \lambda_{4,8}G_{12} + \lambda_{4,10}B_{12} \quad (16) \end{aligned}$$

The equations ensuring stationarity of the Lagrangian (15) with respect to the G and B unknown parameters are:

$$\begin{aligned} \partial L/\partial G_{11} &= \lambda_{3,1} + \lambda_{3,3} &&= 0 \quad (17a) \\ \partial L/\partial B_{11} &= -\lambda_{4,1} - \lambda_{4,3} &&= 0 \quad (17b) \\ \partial L/\partial G_{12} &= \lambda_{1,5} + \lambda_{1,10} - \lambda_{2,7} + \lambda_{2,8} \\ &\quad + \lambda_{3,5} + \lambda_{3,10} - \lambda_{4,7} + \lambda_{4,8} &&= 0 \quad (17c) \\ \partial L/\partial B_{12} &= \lambda_{1,7} - \lambda_{1,8} + \lambda_{2,5} + \lambda_{2,10} \\ &\quad + \lambda_{3,7} - \lambda_{3,8} + \lambda_{4,5} + \lambda_{4,10} &&= 0 \quad (17d) \end{aligned}$$

The equations (17) can be written in more concise form as:

$$\text{trace}\{\lambda^T L_r\} = 0, \quad \forall r \in \{1,\dots,R\} \quad (18)$$

where $R = 4$ for the given two-bus example and measurement set. More generally, R is a known integer corresponding to the number of unique unknown line parameters in the network, as defined by the chosen measurement set $\mathcal{M}$ and the given baseline network topology. The matrices $L_r \in \mathbb{R}^{M\times D}$ for all $r \in \{1,\dots,R\}$ are also explicitly known given the chosen measurement set and line topology.

For example, on the given two-bus example, Equation (17c) corresponding to the stationarity of the Lagrangian with respect to G_{12} can be written as trace$\{\lambda^T L_3\} = 0$ by defining L_3 as:

$$L_3 \triangleq \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \end{bmatrix} \quad (19)$$

The equations ensuring stationarity of the Lagrangian (15) with respect to the error penalty variables ξ_t, and ξ_t^* for all $t \in \{1,\dots,T\}$ are:

$$\begin{aligned} \partial L/\partial \xi_{m,t} &= C - \alpha_{m,t} - \beta_{m,t} &&= 0 \quad (20a) \\ \partial L/\partial \xi_{m,t}^* &= C - \alpha_{m,t}^* - \beta_{m,t}^* &&= 0 \quad (20b) \end{aligned}$$

Finding the stationarity of the Lagrangian (15) with respect to the weight matrix W yields the relationship between the primal weight matrix W and the dual support vectors α_t and α_t^* for all $t \in \{1,\dots,T\}$:

$$W = \textstyle\sum_{t=1}^{T}(\alpha_t - \alpha_t^*)\phi(\mathbf{x}_t)^T + \lambda \quad (21)$$

Combining these constraints and plugging the definition of W in (21) back into the Lagrangian (15), we arrive at the dual form of the constrained SVR problem, given in the theorem below.

Theorem 1. *The dual of the constrained SVR problem (12) corresponding to the power mapping problem can be written as the convex quadratic program:*

$$\begin{aligned} \min_{\alpha,\alpha^*,\lambda} \ & f(\alpha,\alpha^*,\lambda) && (22a) \\ \textit{s.t.} \ & \alpha_{m,t}, \alpha_{m,t}^* \in [0,C], \ \forall m \in \{1,\dots,M\} \\ & \qquad \forall t \in \{1,\dots,T\} && (22b) \\ & \textit{trace}\{\lambda^T L_r\} = 0, \ \forall r \in \{1,\dots,R\} && (22c) \end{aligned}$$

where we have $\alpha, \alpha^* \in \mathbb{R}^{M\times T}$, $\lambda \in \mathbb{R}^{M\times D}$, and the objective function:

$$\begin{aligned} f(\alpha_t,\alpha_t^*,\lambda) = \frac{1}{2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\alpha_t-\alpha_t^*)^T(\alpha_s-\alpha_s^*)K(\boldsymbol{x}_s,\boldsymbol{x}_t) \\ + \sum_{t=1}^{T}(\alpha_t+\alpha_t^*)^T\epsilon - \sum_{t=1}^{T}(\alpha_t-\alpha_t^*)^T\boldsymbol{y}_t + \frac{1}{2}||\lambda||_F^2 \\ + \sum_{t=1}^{T}(\alpha_t-\alpha_t^*)^T\lambda\phi(\boldsymbol{x}_t) \quad (23) \end{aligned}$$

Proof: The dual objective is given by expanding (15) using (20) and (21), observing that (18) cancels out the trace$\{\lambda^T E\}$ term, and by substituting $K(\mathbf{x}_s,\mathbf{x}_t)$ for the $\phi(\mathbf{x}_s)^T\phi(\mathbf{x}_t)$ terms for all $s,t \in \{1,\dots,T\}$ using the kernel trick. The (22b) constraints of the dual problem follow from the stationarity relations (20) plus the dual feasibility constraints $\alpha_{m,t}, \alpha_{m,t}^*, \beta_{m,t}, \beta_{m,t}^* \geq 0$ for all $t \in \{1,\dots,T\}$, $m \in \{1,\dots,M\}$.

In order to show that the dual problem (22) is a convex quadratic program, we introduce the variable $Z \triangleq \begin{bmatrix}\alpha & \alpha^* & \lambda\end{bmatrix}^T \in \mathbb{R}^{(2T+D)\times M}$ and rewrite the problem in stacked form as:

$$\begin{aligned} \min_{Z} \ & \frac{1}{2}\text{trace}\{Z^T A_0 Z\} + \text{trace}\{A_1^T Z\} && (24a) \\ \text{subject to:} \ & \underline{Z} \leq Z \leq \overline{Z} && (24b) \\ & \text{trace}\{Z^T\tilde{L}_r\} = 0, \ \forall r = 1,\dots,R && (24c) \end{aligned}$$

where the lower and upper bounds $\underline{Z}$ and $\overline{Z}$ follow from (22b) and the matrices $\tilde{L}_r$ are expanded versions of L_r such

that trace$\{Z^T \tilde{L}_r\}$ = trace$\{\lambda^T L_r\}$ for all $r \in \{1, \ldots, R\}$. Additionally, we define the matrices $A_0 \in \mathbb{S}^{(2T+D)}$ and $A_1 \in \mathbb{R}^{(2T+D)\times M}$ as:

$$A_0 \triangleq \begin{bmatrix} Q & -Q & \Phi^T \\ -Q & Q & -\Phi^T \\ \Phi & -\Phi & I_D \end{bmatrix}, A_1 \triangleq \begin{bmatrix} [(\epsilon^T)_{\times T}] - Y^T \\ [(\epsilon^T)_{\times T}] + Y^T \\ 0_{D\times M} \end{bmatrix} \quad (25)$$

where we have $\Phi \in \mathbb{R}^{D\times T}$ as the matrix composed of columns $\phi(\mathbf{x}_t) \in \mathbb{R}^D$ and $Y \in \mathbb{R}^{M\times T}$ as the matrix composed of columns $\mathbf{y}_t \in \mathbb{R}^D$ for all $t \in \{1, \ldots, T\}$. We define the kernel matrix $Q \in \mathbb{R}^{T\times T}$ as having entries $Q_{s,t} = K(\mathbf{x}_s, \mathbf{x}_t)$ for all $s, t \in \{1, \ldots, T\}$, noting that $Q = \Phi^T\Phi$. The notation $[(\epsilon^T)_{\times T}]$ indicates a $T \times M$ matrix where every row equals ϵ^T, and the notation $0_{D\times M}$ indicates a $D \times M$ matrix composed of all zeros. In this stacked formulation, we can observe that $A_0 = aa^T$ where $a \triangleq \begin{bmatrix}\Phi & -\Phi & I_D\end{bmatrix}^T$, thus $A_0 \succeq 0$ and the program is convex. □

IV. Analysis of Constrained SVR Approach

A. Strong Duality of Constrained SVR Problem

Similar to the classic SVR approach where the dual problem is used as a solution for the primal, we show that the proposed dual of the constrained SVR problem can also be used as a solution to the primal.

Theorem 2. *The dual of the constrained SVR problem given by (22) is exact. Therefore, solving (22) will recover the solution to (12).*

Proof: Since the primal problem (12) is a convex quadratic program, we just need to show that it satisfies the weak Slater's condition to prove strong duality. To show that there exists some point strictly within the feasible space of (12), we can set $W = E = 0$ and $\xi_{m,t} > \max\{0, (\mathbf{y}_t - \epsilon)_m\}$ and $\xi^*_{m,t} > \max\{0, (-\mathbf{y}_t - \epsilon)_m\}$ for all $t \in \{1, \ldots, T\}$, $m \in \{1, \ldots, M\}$. Then, we can see that the inequality constraints (12b),(12c), and (12d) are all strictly feasible for this point and weak Slater's holds. □

B. Effect of Measurement Availability on Constrained SVR

The availability of measurements, both in terms of SCADA measurements and PMU measurements, directly affects which line parameters we are able to accurately learn with the constrained SVR method. For example, in the case where two buses i and j both have PMUs and the SCADA measurements p_{ij} and q_{ij} are the only available measurements, we are able to exactly recover G_{ij} and B_{ij} when there is no measurement noise and are able to recover good estimates for these parameters when there is noise. This is evident by considering Equations (8a) and (8b) in relation to the constrained SVR problem (12) where C is sufficiently high. Similarly, if we have all possible PMU and SCADA measurements available in the network, we can recover all the exact G and B parameters in the network when there is no measurement noise and good estimates when there is noise. However, in the more realistic case where PMU measurements are not available at every bus, the formulation (12) attributes the discrepancy due observed and unobserved PMU measurements to the penalty terms ξ and ξ^*. When there is no available PMU at bus i, it will be impossible to recover the parameters G_{ij}, B_{ij}, G_{ii} and B_{ii} since the features d_i^2, e_i^2, $\sqrt{2}d_id_j$, $\sqrt{2}e_id_j$, $\sqrt{2}d_ie_j$, and $\sqrt{2}e_ie_j$ do not appear in the feature vector $\phi(\mathbf{x})$. Thus, SCADA measurements p_{ij} and q_{ij} are not helpful to learn G_{ij} and B_{ij} if either bus i or bus j is not equipped with a PMU. On the contrary, the real and reactive power injection measurements at bus i can still be useful to learn some parameters even if not all buses attached to bus i are equipped with PMUs.

While limited access to PMU measurements does prevent the constrained SVR problem from accurately learning all the network parameters, there are many realistic cases where most of network parameters and topology are known but there is some uncertainty in parts of the network. For example, some line $(i, j) \in \mathcal{L}$ might often switch from open to closed in real-time such that its status is often unknown. By placing PMUs at buses i and j and solving the constrained SVR problem, this method could accurately recover G_{ij} and B_{ij} for the line and thus determine the status of the line. Future work should consider how strategic PMU placement can improve power flow mapping recovery.

V. Simulations

The simulations are run on a standard laptop (2.6 GHz 6-Core Intel Core i7 with 16 GB 2400MHz RAM). The software MATPOWER is used to import test networks, formulate admittance matrices, and generate sample data points based on solving power flow problems with the known system parameters [22]. Then, the constrained SVR model is formulated using the Pyomo modeling language in Python 3.8. The convex quadratic program (22) is solved with the Gurobi solver.

In the following simulations, we test our method on cases with various signal-to-noise (SNR) ratios in the SCADA measurements, cases with outliers in the SCADA or PMU measurements, and networks where measurements are only partially observed. For all of these test cases, we assume that the noise in PMU magnitude measurements follows a zero-mean Gaussian distribution with 0.005 p.u. standard deviation (SNR of 46 dB) and that the noise in PMU angle measurements follows a zero-mean Gaussian distribution with a SNR of 40 dB. These values are consistent with the existing literature on PMU errors [23]. For these simulations, we consider the case where we have 10 PMU measurements associated to each SCADA measurement at any time step, as discussed in Section II-B.

A. Performance Metrics

We compare the proposed constrained SVR method with the classic SVR methods in [5] and [9], which were shown to outperform various NN-based methods in [9]. For this comparison, we consider the following performance metrics. Root-mean-square error (RMSE) measures how well the estimator fits the SCADA measurements $\mathbf{y}$ by

penalizing the squared error discrepancy, i.e. $RMSE = \sqrt{\frac{1}{MT}\sum_{t=1}^{T}\sum_{i=1}^{M}(y_{m,t}-\hat{y}_{m,t})^2}$, where $\hat{y}_{m,t}$ is output predicted by the estimator for some time step t and measurement type m. Mean absolute error (MAE) measures how well the estimator fits the SCADA measurements $\mathbf{y}$ by penalizing the absolute error discrepancy, i.e. $MAE = \frac{1}{MT}\sum_{t=1}^{T}\sum_{i=1}^{M}|y_{m,t}-\hat{y}_{m,t}|$. In terms of RMSE and MAE, the classic SVR method has been shown to perform well in the power flow mapping problem [5], [9]. This is because the classic SVR method has good performance learning the overall mapping between $\mathbf{x}_t$'s and $\mathbf{y}_t$'s, and RMSE and MAE do not penalize overfitting. Thus, a more useful metric to see if the method really learns the true line parameters is to consider the error between the actual line parameters and the estimated line parameters. Note that the classic SVR method could be coupled with a post-processing estimation step to get a better estimation of the line parameters, such as using a least-squares estimator to find a set of sparse line parameters consistent with the expected mapping. However, this would involve solving a second optimization problem, further increasing the solution time for large systems. In these results we consider only a single-stage classic SVR problem, thus the results presented below are just used as a baseline.

In the case of line conductance, denoted by the subscript LC, we take $\mathbf{G}_{\text{LC}}$ to be the vector of true line conductances G_{ij} for all lines (i,j) where we have obtained estimates $\hat{G}_{ij}$ using the SVR method and take $\hat{\mathbf{G}}_{\text{LC}}$ to be the corresponding vector of $\hat{G}_{ij}$'s. Then, the normalized estimation error for line conductance over all estimated lines is given by:

$$\Gamma_{\text{LC}} \triangleq \frac{||\mathbf{G}_{\text{LC}} - \hat{\mathbf{G}}_{\text{LC}}||_2}{||\mathbf{G}_{\text{LC}}||_2} \tag{26}$$

We have similar normalized error relations for the line susceptances (given as Γ_{LS}), self conductances (given as Γ_{SC}), and self susceptances (given as Γ_{SS}).

B. Effectiveness in the Presence of Noise

We start with the case that all buses are equipped with PMUs so that the voltages in the network are fully observable. For the IEEE 14-bus test case, we generate PMU measurements based on the initial voltage state of the MATPOWER case file, adding Gaussian-distributed random noise as described in the methodology above. We consider a SCADA measurement set that consists of all real and reactive power flow measurements as well as real and reactive power injection measurements at buses 2 and 3. For this test case setup, we consider the scenario where we have zero noise (SNR=∞) and 50 simulations at various SNRs from 45 to 0 dB. The results of these experiments are given in Table I. Note that the normalized estimation errors for line conductance and self conductance, i.e. Γ_{LC} and Γ_{SC}, are omitted from Table I for concision. These experiments demonstrate that the constrained SVR method outperforms the classic SVR method in terms of both line parameter recovery and solution speed and has similar performance to the classic SVR method in terms of RMSE and MAE. In

TABLE I

COMPARISON OF METHODS ON 14-BUS NETWORK WITH NOISE

Model	SNR	Avg. RMSE	Avg. MAE	Avg. solve time (s)	Avg. Γ_{LS}	Avg. Γ_{SS}
Classic SVR	∞	0.0013	0.0010	46.86	1.0004	1.0000
	45	0.0009	0.0009	44.26	1.0004	1.0000
	30	0.0009	0.0008	44.49	1.0004	1.0000
	15	0.0009	0.0009	45.25	1.0004	1.0000
	0	0.0001	0.0009	42.71	1.0004	1.0000
Constr. SVR	∞	0.0535	0.0272	5.25	0.0020	0.0051
	45	0.0535	0.0272	5.21	0.0048	0.0052
	30	0.0536	0.0272	5.17	0.0249	0.0141
	15	0.0539	0.0273	5.19	0.1392	0.0672
	0	0.0664	0.0346	5.24	0.7952	0.4202

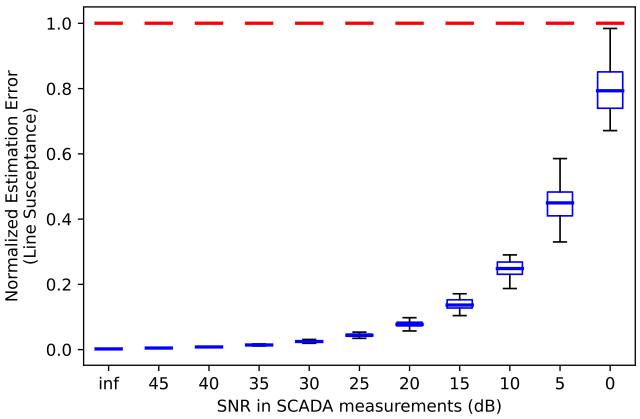

Fig. 1. The normalized estimation error for line susceptance (B_{ij}'s) as a function of signal-to-noise ratio (SNR) is presented for the 14-bus test case. Each box plot corresponds to 50 simulations of Gaussian-distributed random noise added to the SCADA measurements of this test case. The red lines are box plots of normalized estimation errors Γ_{LS} for the classic SVR method, and the blue box plots correspond to the normalized estimation errors Γ_{LS} for our proposed constrained SVR method.

Figure 1, we have plotted Γ_{LS} as a function of the SNR of the SCADA measurements, noting that similar plots were obtained for the normalized estimation errors of the other line parameters (omitted for concision). Even when SCADA measurement errors are present, the constrained SVR method can recover decent estimates for the line parameters (within 14% of the true values on average) as long as the SNR in the measurements is greater than or equal to 15 dB. Conversely, the classic SVR method consistently fails to recover the true line parameters even when no noise is present in the SCADA measurements.

C. Robustness in the Case of Outliers

Next, we consider the same 14-bus test case as above with a SNR of 40 dB in the SCADA measurements but allow for outliers to be present in 0 to 8% in some of the PMU or SCADA measurement data. The subset of outlier-affected measurements within the PMU and SCADA datasets is chosen on a uniform distribution over all possible measurements, and the values of the outlier-affected measurements are chosen on a uniform distribution from $[-2, 2]$ per unit, scaled in relation to the chosen measurement type. In practice, outlier values may be arbitrarily large, so we

TABLE II

COMPARISON OF METHODS ON 14-BUS NETWORK WITH OUTLIERS

Model	% outlier	Avg. sol. time (s)	Avg. Γ_{LC}	Avg. Γ_{SC}	Avg. Γ_{LS}	Avg. Γ_{SS}
Classic SVR	0.00	43.42	0.9906	1.0003	1.0005	1.0000
	2.04	47.46	0.9914	1.0004	1.0004	1.0000
	4.08	48.21	0.9922	1.0003	1.0003	1.0000
	6.12	47.74	0.9917	1.0004	1.0004	1.0000
	8.16	47.79	0.9925	0.9999	1.0002	0.9999
Constr. SVR	0.00	5.31	0.0089	0.0064	0.0083	0.0049
	2.04	5.89	0.6211	0.2000	0.4078	0.0853
	4.08	5.84	0.9450	0.3332	0.3823	0.1284
	6.12	5.80	1.8727	0.5187	0.8018	0.2609
	8.16	5.87	2.0223	0.5204	0.9563	0.2962

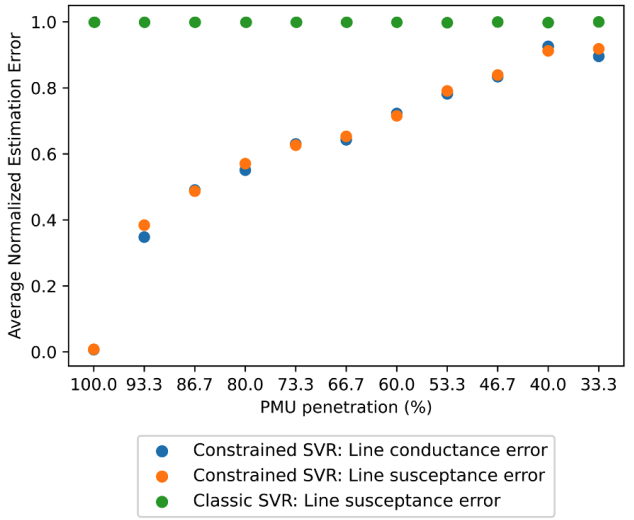

Fig. 2. The average normalized estimation error of line conductance and line susceptance parameters for the 30-bus network, with varying amounts of PMU penetration in the network. During each simulation, some subset of buses is randomly selected to be equipped with PMUs. Each data point corresponds to the average over 20 random simulations.

take this outlier distribution simply as an assumption used to compare our model performance with classic SVR. The results from these simulations are given in Table II.

D. Effectiveness in Partially-Observable Networks

Finally, we consider the case where only a subset of buses in the network are equipped with PMUs. For these simulations, we consider the IEEE 30-bus network and perform a sweep over varying PMU penetration levels in the network. For these simulations, we take the SNR of the SCADA noise as 40 dB with no outliers. Some results from these simulations are given in Figure 2. We see that the sparse SVR model provides better average normalized estimation error at all PMU penetration levels than the classic SVR method but that its benefits are more apparent at higher PMU penetrations.

VI. CONCLUSIONS

In this paper, we proposed a new constrained SVR method that can learn the true power network topology of a network. Based on simulations on IEEE test cases, we showed that the proposed constrained SVR method is much better at recovering the true line parameters of a network, even in the presence of SCADA measurement noise, outliers, and missing data, than existing SVR methods.

REFERENCES

[1] U.S. Department of Energy, "Grid modernization and the smart grid." [Online]. Available: https://www.energy.gov/oe/grid-modernization-and-smart-grid

[2] M. Jin, J. Lavaei, and K. H. Johansson, "Power grid AC-based SE: Vulnerability analysis against cyber attacks," *IEEE Transactions on Automatic Control*, vol. 64, no. 5, pp. 1784–1799, 2019.

[3] M. Jin, J. Lavaei, S. Sojoudi, and R. Baldick, "Boundary defense against cyber threat for power system SE," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 1752–1767, 2021.

[4] A. Wood, B. Wollenberg, and G. Sheblé, *Power Generation, Operation, and Control*. Wiley, 2013.

[5] J. Yu, Y. Weng, and R. Rajagopal, "Robust mapping rule estimation for power flow analysis in distribution grids," in *2017 North American Power Symposium (NAPS)*, 2017, pp. 1–6.

[6] M. Delimar, I. Pavic, and Z. Hebel, "Artificial neural networks in power system topology recognition," in *The IEEE Region 8 EUROCON 2003. Computer as a Tool.*, vol. 2, 2003, pp. 287–291 vol.2.

[7] X. Hu, H. Hu, S. Verma, and Z.-L. Zhang, "Physics-guided deep neural networks for power flow analysis," *IEEE Transactions on Power Systems*, vol. 36, no. 3, pp. 2082–2092, 2021.

[8] C. Wang, J. An, and G. Mu, "Power system network topology identification based on knowledge graph and graph neural network," *Frontiers in Energy Research*, 2021.

[9] J. Yuan and Y. Weng, "Support matrix regression for learning power flow in distribution grid with unobservability," *IEEE Transactions on Power Systems*, vol. 37, no. 2, pp. 1151–1161, 2022.

[10] S. Sojoudi and J. Lavaei, "Physics of power networks makes hard optimization problems easy to solve," *2012 IEEE Power and Energy Society General Meeting*, pp. 1–8, 2012.

[11] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*. Association for Computing Machinery, 1992, p. 144–152.

[12] H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola, and V. Vapnik, "Support vector regression machines," in *Advances in Neural Information Processing Systems*, vol. 9. MIT Press, 1996.

[13] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004.

[14] O. Chapelle, P. Haffner, and V. Vapnik, "Support vector machines for histogram-based image classification," *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 1055–1064, 1999.

[15] B.-J. Chen, M.-W. Chang, and C.-J. lin, "Load forecasting using support vector machines: a study on EUNITE competition 2001," *IEEE Transactions on Power Systems*, vol. 19, no. 4, pp. 1821–1830, 2004.

[16] Q. Klopfenstein and S. Vaiter, "Linear SVR with linear constraints," *Machine Learning*, vol. 110, no. 7, pp. 1939–1974, 2021.

[17] H. Wu and Giri, "PMU impact on state estimation reliability for improved grid security," in *2005/2006 IEEE/PES Transmission and Distribution Conference and Exhibition*, 2006, pp. 1349–1351.

[18] A. Ashok, M. Govindarasu, and V. Ajjarapu, "Online detection of stealthy false data injection attacks in power system state estimation," *IEEE Transactions on Smart Grid*, vol. 9, no. 3, pp. 1636–1646, 2018.

[19] E. Glista and S. Sojoudi, "A MILP for optimal measurement choice in robust power grid state estimation," in *2022 IEEE Power and Energy Society General Meeting (PESGM)*, 2022.

[20] L. Vanfretti, M. Baudette, and A. D. White, "Monitoring and control of renewable energy sources using synchronized phasor measurements," in *Renewable Energy Integration*, 2017, pp. 419–434.

[21] A. Simões Costa, A. Albuquerque, and D. Bez, "An estimation fusion method for including phasor measurements into power system real-time modeling," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1910–1920, 2013.

[22] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas, "MATPOWER: Steady-state operations, planning and analysis tools for power systems research and education," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12–19, February 2011.

[23] M. Brown, M. Biswal, S. Brahma, S. J. Ranade, and H. Cao, "Characterizing and quantifying noise in PMU data," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*, 2016, pp. 1–5.