Support Matrix Regression for Learning Power Flow in Distribution Grid with Unobservability

Jingyi Yuan, Student Member, IEEE, and Yang Weng, Senior Member, IEEE,

Abstract—Increasing renewable penetration in distribution grids calls for improved monitoring and control, where power flow (PF) model is the basis for many advanced functionalities. However, unobservability makes the traditional way infeasible to construct PF analysis via admittance matrix for many distribution grids. While data-driven approaches can approximate PF mapping, direct machine learning (ML) applications may suffer from several drawbacks. First, complex ML models like deep neural networks lack the degradability and explainability to the true system model, leading to overfitting. There are also synchronization issues among different meters without GPS chips. Last but not least, bad data is quite common in the distribution grids. To resolve these problems all at once, we propose a variational support matrix regression (SMR). It provides structural learning to (1) embed kernels to regularize physical form in observable area while achieving good approximation at unobservable area, (2) integrate temporal information into matrix regression for asynchronous data imputation, and (3) define support matrix for margins to be robust against bad data. We test the performance for mapping rule via IEEE test systems and a utility distribution grid. Simulation results show high accuracy, degradability from data-driven model to physical model, and robustness to data quality issues.

Index Terms—Power flow mapping recovery, distribution grids, support matrix regression, unobservability, data-driven method

I. INTRODUCTION

The distribution grids undergo significant changes due to increasing distributed energy resources (DERs). Such DER deployment not only contributes to cleaner energy but also brings in significant challenges for sustainable operation [1]. Therefore, new technologies are developed for advanced and automated distribution management, e.g., hosting capacity analysis of DERs, fault identification [2], coordinated control among smart inverters [3], etc. However, many of these new technologies require to know an accurate power flow model, difficult to obtain in many distribution grids with increasing uncertainties due to fast DER integration [4].

Although the system information is unavailable directly, researchers try to rebuild the power flow model leveraging historical data from smart meters, micro-phasor measurement units (PMUs), and DER-equipped sensors, thanks to the investment of grid modernization [4]–[9]. However, these data-driven studies are often limited by unobservability issues, e.g., lack of measurements, to work properly. For example, some works assume to initially know a mostly correct topology and use real-time data for topology identification and parameter estimation with similar setups as the transmission grid [4]–[7]. Others treat the power flow analysis as a mapping-rule learning-task by assuming sufficient measurements of loads, generation, voltages, and currents [8], [9]. Their preconditions are hard to satisfy, as many utilities do not have a good knowledge of their distribution grids after decades of operational changes. For example, bus measurements of voltages and power injections are not fully available due to limited sensor deployment. Even worse, there are relatively more frequent topology/parameter changes in distribution due to plug and play of DERs [10]–[12]. Associated with these DERs, some customer-owned active controllers and related control rules are also not monitored by the utility. To resolve the problem in explicit model reproduction, the deep neural networks (NN) are proposed to implicitly approximate a high-dimensional mapping [13]–[18]. NNs are well-known for their universal function approximation capability, which provides sufficient freedom to fit the complexity in data [19], [20]. Although numerical results can be good for some specific examples, the model complexity has the potential weakness of overfitting. For example, their nested nonlinear structures fail to provide a degradability from NN model to the underlying physical model for confidences or guarantees. Such a lack of degradability easily results in overfitting problems in the observable areas and causes large errors when the operating points are beyond the historical data range. This is especially important for the case when distribution systems undergo rapid changes due to PVs, loads, and inverter controls. Learning models without physical degradability can be sensitive to such changes and performs poorly for extrapolation. Since there is no physical degradability to explain the obtained solution, system operators find it difficult to trust the learning methods for decision-making in planning and operations [21]–[23]. Hence, it is desirable to have a learning model that can approximate power flow mapping regarding unobservability while providing the degradability for the observable region as performance guarantees.

In addition to observability/explainability issues, the data quality problem is also a big issue in distribution grids, e.g., data synchronization and other bad data. The measurements are collected from different data sources with or without GPS chips, with limited or no effort to calibrate clocks for synchronization [7]. The time synchronization errors may also result from randomness, infinite accuracy of meters and the telecommunication medium, and even malicious packet attacks [24], [25]. Although synchrophasor technologies (i.e., µ-PMUs) can solve such a problem, it is still too costly for widespread use in distribution grids [25]–[28]. Besides, data errors in the distribution grid can deteriorate learning performance. For example, communication disturbances, measurement outliers, and missing data are common in distribution grids as various types of sensors are deployed, and some can be outdated [29], [30]. If these problems are treated separately, errors can propagate. Therefore, it will be great if a power flow learning model can simultaneously provide robustness against...
data quality problems.

To resolve these issues, we propose structural designs in machine learning that efficiently extracts information from limited data. We provide physical degradability in the observable region and good approximation in the unobservable region by embedding kernels into a learning model. The goal is to balance the physical embedding and universal approximation for the mapping recovery with unobservability. However, data quality issues such as asynchronization can distort the learning process in each time slot. We propose to expand a one-snapshot-based learning process into an integrated learning experience over consecutive time slots, so that time delay in sensors can be learned and reversed. Specifically, we design to extend measurement vectors in one-snapshot with a matrix of time-series inputs and use a regularization on the time span for enhancement. Such a design facilitates the applicability of power flow mapping recovery to asynchronization scenarios. Besides, bad data is another big issue to decrease learning performance, where we propose to define tolerances via the concept of margins rather than equally treating all the data [31].

Thus, to combine the physically meaningful kernel function, temporally meaningful matrix form, and margin for tolerance, we propose support matrix regression (SMR) that recovers the power flow mapping from historical measurements when grid information is unknown or partially unknown.

For validation, we use IEEE test systems (8-bus and 123-bus) and a utility distribution network to validate the proposed approach on recovering power flow mapping. SMR is tested with the existence of unmeasured bus nodes in the system, which is compared with both simple regression and more complex methods, e.g., multilayer perceptron (MLP), radial basis function neural network (RBF NN) [6], [8], [9], [13]–[18], [32], [33]. We not only show the high accuracy of power flow mapping but also demonstrate that the mapping coefficient can be regarded as recovered line parameters. The experiments show that these parameters are the same as the true parameter values for the observable region, demonstrating that SMR can reflect the underlying physical power flow model. Afterward, various asynchronization errors and different levels of bad data are added to validate the robustness of SMR design against poor data quality, when compared to other data-driven approaches. Simulation results demonstrate that support matrix regression is an effective tool for reproducing power flow mapping in the distribution grids.

The contributions of the work lie in providing structural designs of the learning model for power flow recovery, which 1) embeds kernels to regularize physical degradability, 2) extends time-series matrix inputs for asynchronization compensation, and 3) maximizes margins against data errors.

The rest of the paper is organized as follows: Section II forms the power flow mapping recovery problem, Section III proposes SMR with key designs. Section IV illustrates how specific designs tackle challenges in reproducing power flow mapping. Section V validates the proposed SMR via extensive simulations, and Section VI concludes the paper.

II. REPRODUCE THE POWER FLOW MODEL FROM DATA

In order to reproduce power flow model from historical data, we start with the classic power flow (PF) model in a single time slot. With \( n \) buses, the PF equations are

\[
\begin{align*}
    p_i &= \sum_{k=1}^{n} |v_k| |v_k| (g_{ik} \cos \theta_{ik} + b_{ik} \sin \theta_{ik}), \\
    q_i &= \sum_{k=1}^{n} |v_k| |v_k| (g_{ik} \cos \theta_{ik} - b_{ik} \sin \theta_{ik}),
\end{align*}
\]

where \( p_i \) and \( q_i \) are the active and reactive power injections at bus \( i \), and \( \theta_i \) and \( \theta_k \) are the voltage magnitude and angle at bus \( i \) (the angle difference \( \theta_{ik} = \theta_i - \theta_k \)). \( g_{ik} + j \beta_{ik} \) is the element of the admittance matrix \( Y_{bus} \in \mathbb{C}^{n \times n} \) to represent the line admittance between bus \( i \) and \( k \).

In distribution grids, accurate system knowledge, e.g., \( Y_{bus} \), is often unknown. While directly using the power flow model is typically infeasible, historical measurements of system information are available. Therefore, we rewrite (1), from a data-driven perspective, as a compact function of voltages and power injections \( [p_1; q_1] = f([v_1; \theta]), \) where \( [v] = [v_1; \ldots; v_n], [\theta] = [\theta_1; \ldots; \theta_n] \).

Subsequently, recovering \( f(\cdot) \) is a task of mapping-rule approximation: given data of voltages and power injections recorded in time sequence \( T: D_v = \{[v_1; \ldots; v_T]; D_{\theta} = \{\theta_1; \ldots; \theta_T\}, D_p = \{p_1; \ldots; p_T\}, D_q = \{q_1; \ldots; q_T\} \), we aim to recover the forward power flow mapping \( f(\cdot) \).

Many machine learning models can be implemented to approximate \( f(\cdot) \). However, prior works suffer from several drawbacks. For example, straightforward methods like linear regression can identify system parameters in the physical model. However, the validity relies on an oversimplified model with hard-to-achieve assumptions of full system observability, e.g., all bus nodes are measured. On the other hand, when some neighbor buses are unmeasured, deep neural networks can approximate complex relationships among available data. Nevertheless, the black-box structure can not explicitly express the physical model to possess degradability to (1). In addition to limited measurements, asynchronization errors and bad data within these limited measurements can lead to a distorted mapping \( f(\cdot) \).

To address these concerns, the learning tool has to balance between the explicit expression of the physical model in the observable region and a good approximation in the unobservable area. Meanwhile, it needs robust designs to guarantee accurate reproduction against data quality issues, e.g., asynchronization and bad data.

III. PROPOSED SUPPORT MATRIX REGRESSION

To tackle physical explainability and data quality concerns altogether, we propose in this section a method on support matrix regression via various machine learning designs.

A. Appropriate Power Flow Mapping Function Approximation

To reproduce the mapping from data, one needs to have a candidate model \( h : x \rightarrow y \), which approximates a target function for the input-output relationship [34]. One fundamental question in learning theory is how to choose an
For this purpose, we would like to choose an effective function \( h(\cdot) \) correlated with the underlying physical model \( f(\cdot) \). We also consider the possibility of degradability from data-driven to physical model, because many machine learning models lack physical interpretation as a guarantee of the results for system operators to trust.

Therefore, we look into (1) with a focus on the coupling between quadratic and sinusoidal functions that are difficult to represent. To simplify the functional types from two to one for kernel embedding, we convert (1) in polar coordinates to rectangular coordinates. Specifically, let \( v_{i,rc} = |v_i| \cos \theta_i \) and \( v_{i,im} = |v_i| \sin \theta_i \). Consecutively, (1) becomes

\[
\begin{align*}
\dot{p}_i &= \sum_{k=1}^{n} g_{ik}(v_{i,rc} v_{k,rc} + v_{i,im} v_{k,im}) \\
&\quad + b_{ik}(v_{i,im} v_{k,rc} - v_{i,rc} v_{k,im}), \\
\dot{q}_i &= \sum_{k=1}^{n} g_{ik}(v_{i,im} v_{k,rc} - v_{i,rc} v_{k,im}) \\
&\quad - b_{ik}(v_{i,rc} v_{k,rc} + v_{i,im} v_{k,im}).
\end{align*}
\]

In such an equivalent expression, power injections are the linear combination of the quadratic function of voltages \([v_{rc}:v_{im}]\). Therefore, a straightforward way to learn the forward mapping \( f(\cdot) \) is linear regression (LR) over the \( 2nd \)-degree polynomials of voltages as candidates. Input voltages are mapped to quadratic feature space by \( \phi(x) \), e.g., containing features \( v_{i,rc} v_{k,rc} \) and \( v_{i,im} v_{k,im} \). Then, LR is trained by simply minimizing the difference between real and approximated function outputs,

\[
\min_{\mathbf{W}} \sum_{t=1}^{T} (y_t - h_{W}(x_t))^2 = \sum_{t=1}^{T} (y_t - \langle \mathbf{w}, \phi(x_t) \rangle)^2.
\]

The sample input at time \( t \) is \( x_t = [v_{rc}:v_{im}]_t \) and output \( y_t = p_t \) or \( q_t \). Since it is a convex optimization [35], the resulted regression coefficients \( \mathbf{w} \) are optimal. With perfect measurements, the solution can exactly reflect line parameters of underlying power flow equations.

Although LR is fully transparent for operators to understand how the solutions are obtained, such successful implementation relies on strong assumptions hard to achieve. For example, it is difficult to acquire complete and clean data all around the system because there are unmeasured nodes in distribution systems. Also, LR is too simple to work against data quality issues, e.g., asynchronizations and outliers, common in real smart meter measurements.

### B. Support Matrix Regression

Thus, in this paper, we propose different designs to address the challenges in learning power flow mapping [31], [36]–[38]. They are integrated into the flexible structure of the novel support matrix regression (SMR) below.

\[
\begin{align*}
\min_{\mathbf{W}, \xi^+, \xi^-} & \quad \frac{1}{2} ||\mathbf{W}||_F^2 + c_1 \sum_{t=1}^{T} (\xi_t + \xi_t^+) + c_2 ||\mathbf{W}||_F \\
\text{s.t.} & \quad y_t - h_{W}(X_t) \leq \epsilon + \xi_t, \\
& \quad h_{W}(X_t) - y_t \leq \epsilon + \xi_t^+, \\
& \quad \xi_t, \xi_t^+ \geq 0.
\end{align*}
\]

In the designed model objective, minimizing the squared Frobenius norm \( ||\mathbf{W}||_F^2 \) of matrix is equivalent to maximizing margin (see Appendix A for illustration) that makes the regression less sensitive to noisy data. Such insensitivity is quantified by \( \epsilon \) to tolerate small measuring perturbations within the bounds. When a measured active power value has deviations much larger than \( \epsilon \), simply ignoring it causes learning errors. That’s why the slack variables \( \xi_t, \xi_t^+ \) are introduced to specify such value. In this way, constraints (4)-(6) are able to cover all data points in the feasible region, i.e., \( \epsilon \) set perturbation tolerance band and \( \xi_t, \xi_t^+ \) specify large deviations.

The summation of \( \xi_t, \xi_t^+ \) in the objective is to balance between bad data tolerance and identifying the correct mapping.

In case some measurements at snapshot \( t \) are missing due to asynchronized sensors in distribution grids, data imputation is designed in SMR. We feed the matrix input \( X_t \in \mathbb{R}^{(2r+1) \times 2n} \) that includes the voltage measurements from \( \tau \) time slots before to \( \tau \) time slots after time \( t \), instead of one snapshot vector \( x_t \) for each data sample. Naturally, the regression coefficients \( \mathbf{W} \) are also described in matrix form, and thus the approximated function is \( h_{W}(X) = \langle \mathbf{W}, X \rangle \), which allows embedding kernels for flexible feature mapping. For example, a \( 2nd \)-degree polynomial kernel is equivalent to the quadratic voltage features in (2). When unmeasured neighbors make perfect reconstruction impossible, a \( 3^{rd} \) or higher degree kernel can automatically be used for good approximation within our highly flexible framework. Moreover, the nuclear norm \( ||\mathbf{W}||_F \) couples with the kernel, using time span to enhance the data imputation [37]. Thus, the design avoids mapping distortion under asynchronization and the learning model is robust against data issues.

In addition, \( c_1 \) and \( c_2 \) are regularization parameters in the objectives. Several methods like \( k \)-fold cross validation and Aikake information criterion can be used to select the best parameters based on [39], and we use \( k \)-fold cross validation. We select \( c_1 \) from \{1 \times 10^{-3}, 5 \times 10^{-3}, 1 \times 10^{-2}, 5 \times 10^{-2}, 1 \times 10^{-1}, 5 \times 10^{-1}, 1 \times 10^{0}, 5 \times 10^{0}, 1 \times 10^{1}\}, \) and for each \( c_1 \), we tune \( c_2 \) to make the rank of \( W \) varied from 1 to the matrix size. The general formula of SMR not only integrates the aforementioned designs to address difficulties but also preserves convexity to find the optimal fitting.

According to [36], such a convex optimization problem can be solved easier through a dual problem with Lagrangian construction. Consequently, the optimal solutions of regression coefficients and the mapping function admit the formula

\[
\begin{align*}
W &= \sum_{t=1}^{T} (\alpha_t - \alpha_t^*) \phi(X_t), \\
y &= h_{W}(X) = \sum_{t=1}^{T} (\alpha_t - \alpha_t^*) (\phi(X_t), \phi(X)),
\end{align*}
\]

where \( \alpha \) and \( \alpha^* \) are the corresponding Lagrangian multipliers of constraints (4) and (5). They represent the weights assigned to different types of data samples that characterize the support matrices for optimal \( W \) and the power flow mapping. A detailed illustration of their solution format is included in Appendix B. Next, we show how SMR’s properties addressed unobservability and different data issues in the power flow mapping reproduction.
IV. Key Properties of SMR to Resolve Issues for Power Flow Mapping Reproduction

A. Guarantee Physical Mapping Recovery with Efficiency

Degradability is the ability of a data-driven model to degrade to the underlying physical model, providing a physical guarantee. In the following, we will show that SMR with flexible kernel embedding can efficiently reveal the underlying power flow equations of the observable region to guarantee learning performance.

For degradability, one intuitive way is to follow the highly resembled candidate model, e.g., LR, which prepossesses the quadratic features $\phi(x)$ to stay consistent with (2). Such a way is easy to understand but suffers from limited approximation capability with unobservability in distribution grids. As explained in Section III, we adopt kernels for flexible function representation. Kernels can be embedded because the solution formula in (8) is described in terms of proper inner products $\langle \cdot, \cdot \rangle$. Such a formula satisfies the restriction of kernel trick to directly map the input space to a high-dimensional space,

$$K(X_1, X_2) := \langle \phi(X_1), \phi(X_2) \rangle_{H},$$

which is a simple expression of the feature mapping $\phi(\cdot)$ for power flow function approximation. Therefore, the solution formula (8) is rewritten as $y = \sum_{t=1}^{T} (\alpha_t - \alpha_t^*) K(X_t, X)$, and there is no need for redundant feature prepossessing. Though the regression coefficient matrix $W$ is not derived explicitly, we show that using the kernel in SMR can reveal the underlying power flow equations with the following proposition. This provides a guarantee on physical mapping recovery.

**Proposition 1.** The physical power flow model in (2) can be interpreted by the function approximation of support matrix regression using the $2^{nd}$-degree polynomial kernel function.

**Proof.** As illustrated in III. A, The power flow equations in (2) can be represented as a linear combination of the quadratic functions of voltages, which is $p_i = \langle [g_i; b_i], \phi(X) \rangle$. According to Mercer’s theorem [40], the alternative kernel function for the same space shift is $2^{nd}$-degree polynomial kernel $K(X_1, X_2) = \langle X_1, X_2 \rangle^2$. Hence, the recovered power flow mapping from (8) is $\hat{p}_i = \sum_{t=1}^{T} (\alpha_t - \alpha_t^*) (X_t, X)^2 = \sum_{t=1}^{T} (\alpha_t - \alpha_t^*) \phi(X_t), \phi(X_i))$. Suppose the data perfectly matches the power flow equations, and we could reconstruct the physical model by inferring line parameters from SMR solutions, where $[g_i; b_i] = \sum_{t=1}^{T} (\alpha_t - \alpha_t^*) \phi(X_i)$.

By choosing appropriate function approximation, the physical information gain is maximized. It guarantees that for distribution systems undergoing changes such as gradually increased DER, the proposed SMR still learns the power flow mapping from historical data. Then, considering the fact that bus measurements are not guaranteed to be fully available in distribution grids, the data may be insufficient to reflect the entire power flow equations explicitly. In this case, we show in the following that SMR with kernel embedding enhances mapping approximation associated with unobservability.

As illustrated in Section II, suppose the goal is to recover the power flow mapping of bus $i$, which has some unmeasured neighbor buses. Hence, the underlying power flow model can be rewritten as $\hat{p}_i = f_{1,i}(\|v\|; \theta) + f_{2,i}(\|v\|; \theta)$, where $\|v\|; \theta$ includes complete bus voltages in the grid. With measurable voltages $x$ out of $\|v\|; \theta$, $f_{1,i}(\cdot)$ denotes the partial mapping that can be exactly represented so that $f_{1,i}(\|v\|; \theta) = f_{1,i}(x)$. On the other hand, $f_{2,i}(\|v\|; \theta)$ lacks information for explicit expression. Instead, it can be viewed as some implicit function approximation $f_{2,i}(x)$ of available measurements $x$, usually a complex relationship far beyond linear.

**Remark.** To illustrate with a toy example, suppose a test system is 3-bus network and bus 3 is unobservable. Thus, the complete bus voltages $\|v\|; \theta = [v_1, v_2, v_3; \theta_1, \theta_2, \theta_3]$, and the available voltages $x = [v_1, v_2; \theta_1, \theta_2]$. The target output $y = p_2 = |v_2|^2 g_{22} + |v_1| |v_2| (g_{12} \cos \theta_{12} + b_{12} \sin \theta_{12}) + |v_3| |v_2| (g_{13} \cos \theta_{13} + b_{13} \sin \theta_{13})$. Since bus 3 is unobservable, $f_1 = [v_1] [v_2] (g_{12} \cos \theta_{12} + b_{12} \sin \theta_{12}) + |v_2|^2 g_{22}$ and $f_2 = [v_3] [v_2] (g_{13} \cos \theta_{13} + b_{13} \sin \theta_{13})$. Because $f_3$ only contains the variables $[v_1], [v_2], [\theta_1; \theta_2]$, we can obtain $f_1(x) = f_1(\|v\|; \theta)$ with guaranteed equality. Meanwhile, $f_2(\|v\|; \theta)$ is approximated by $f_2^2(x)$.

Besides, many customers can own ancillary devices with active controllers to support reactive power for voltage regulation, but the dynamic model and control rules may be unavailable to the utility [2]. For example, the Volt/Var droop control of a smart inverter follows a combined curve of several piece-wise linear sections [41], which is different from the function types in basic power flow model.

Therefore, we leverage the flexibility of kernel function to expand the learning model complexity for equivalent power flow mapping. To approximate the mapping in these cases, we start from the proposed $2^{nd}$-degree polynomial kernel and try the $3^{rd}$-degree for a higher-dimensional polynomial approximation of unobservability. Moreover, commonly used kernels for SVM-based methods [36] include: radial basis function (RBF) kernel and hyperbolic tangent kernel. The RBF kernel is $K(X_1, X_2) = e^{-\gamma \|X_1 - X_2\|^2}$, where $\gamma$ is a hyper-parameter greater than 0, and it has a potentially infinite number of dimensions in feature space. For hyperbolic tangent kernel, the function is $K(X_1, X_2) = \tanh(\gamma \|X_1 - X_2\| + c)$, where $c$ is a hyper-parameter. They have shown excellent function approximation capability in widespread applications. Therefore, the different kernels are tested for the best regression performance in experiments.

In addition, another benefit of using kernels is computational efficiency. For example, for an $n$-bus system, the number of quadratic features is $\binom{2n+1}{2}$, and using the time-series inputs in matrix enlarges the number to $\binom{2n+1}{2} \times (2 \tau + 1)$, which is approximately $2.1 \times 10^9$ for the IEEE 123-bus system with $\tau = 3$. For a large-scale system with a more complicated topology, the feature number is close to infinite, inefficient in computation. On the contrary, the equivalent kernel function is easily computed to avoid the heavy computation burden of preprocessing features $\phi(X)$ while guaranteeing the learning performance.

B. Regularized Matrix Extension for Asynchronization

To ensure the kernel embedding mapping function is accurately recovered under the sensor asynchronization scenario,
we expand snapshot inputs to matrix form and enhance such imputation by regularization.

In distribution grids, as PMUs are not widely used, time-series data from different sources could bring in synchronization issues. There are time delays among different nodes’ measurement sequences resulted from multiple factors. In this case, coping with data in each snapshot is unable to find the correct power flow mapping as crucial information is missing. To compensate for such error, the intuitive way is to expand snapshot inputs to time-series that cover more information for machine learning. However, the difficulty lies in how to efficiently expand considering both the large input feature numbers and time correlations of the power system. Therefore, we consider extending the one snapshot input of each sample $x_t$ to a time sequence in a matrix form $X_t$. The matrix inputs compose time-series inputs and retain separated formats for different timestamps in order. Such a matrix extension is flexible that it can be degraded to the vector form without accounting for asynchronization (see details in Appendix A). For matrix input $X_t$, the columns represent the voltage inputs of different buses, and the rows are different time slots (e.g., $2\tau + 1$ time span around time $t$). Given the input voltages of past, present, and after time slots, it imputes asynchronized measurements to avoid the distortion of the recovered mapping.

When expanding inputs to matrix form, the data is arranged in a time sequence. Such temporal correlation can not be disordered, otherwise resulting in many possible results due to margin maximization (minimizing $\|W\|_F^2$ in (3)). Therefore, we enhance the matrix imputation by a regularization of the nuclear norm $\|\|_n$, [36].

In addition to addressing the asynchronization issue, such a regularization also benefits the degradability of SMR. Distribution systems are mostly radial in structure, coming with a sparse admittance matrix. Therefore, the whole input matrix $X_t$ includes many redundant features. Considering the asynchronization scenario, the corresponding coefficient matrix $W$ is sparse. Furthermore, the power system operates in a dynamic manner, e.g., loads and PV generation vary continuously with time. Exploiting correlations between continuous time slots is beneficial to learning performance [42]. Thus, the nuclear norm constrains the rank and sparsity of the regression coefficients, exploiting structure information in input rows and columns to facilitate the recovery of power flow mapping.

C. Margin Maximization for Bad Data

Despite asynchronized sensors, the measurement errors are unavoidable, either slight noises or outliers. Notice that data points with missing data conditions (zero value or null) can be seen as outliers, far away from the regular data points. We show how SMR uses margins to generalize error bounds for bad data during the learning process.

When fitting data, a common way of minimizing empirical risk is to penalize all the observing errors, e.g., LR. In this way, even small measuring disturbances create many fitting possibilities, unable to find the optimal coefficients. To maintain accuracy with the existence of bad data, we shift the objective to margin maximization [36], which is minimizing the squared Frobenius norm $\frac{1}{2}\|\|_F^2$ in (3).

Fig. 1: It is hard for common methods like LR to find the correct hyper-plane (blue dotted line) with noisy data, which has many possibilities (light blue lines). SMR, however, tolerates data within the $\epsilon$-tube. Data with exact $\pm \epsilon$ deviations contribute to locate the hyperplane accurately.

To illustrate, consider a two-bus system where bus 0 is a reference bus and connects to bus 1 through a resistive line, where $v_0 = 1$ p.u. and $\theta_{01} = 0$. Our goal is to recover power flow mapping at bus 1 using noisy measurements of $v_1$ and $p_1$. For simple visualization, each input data sample is converted to a vector version of one time slot in this case (convertible as shown in Appendix A). In Fig. 1, the blue dashed lines denote the correct mapping $p_1 = g_{01}(v_1 - v_1^2)$ and the measurements (shown as data points) have deviations around it. While common methods, which use rigorous error minimization, find it difficult to locate the exact hyper-plane, SMR targets margins. The margin means the distance from each data point to the correct hyper-plane. In the flexible SMR design, it is maximized to cover most noisy measurements in a tube (light blue shade), where $\epsilon$ quantifies the maximum deviation from noiseless values. Then, since the error values of data points on the tube boundaries are known precisely, they can be transferred to error-free data and locate the correct power flow mapping.

However, the $\epsilon$-tube could not describe data outliers with larger deviations, leading to learning errors. To account for that, slack variables $\xi, \xi^+$ are used to identify the data points beyond $\epsilon$ deviations. Notice that outliers are limited compared to normal data, and there is a trade-off between including more data points in the $\epsilon$-tube and identifying outliers with errors larger than $\epsilon$. Therefore, the constant $c_1 > 0$ is used to balance the trade-off in the learning process. In this way, we infer the correct hyper-plane to recover $g_{01}$ from noisy measurements.

Remark. The proposed SMR is a structural extension of support vector regression (SVR) for learning power flow mapping [36]. In specific, there are three major problems to address: lack of physical degradability, bad data like measurement...
noises or outliers, and time asynchronization issues. Based on the previous discussion, SVR can solve the first two by the inherent kernel embedding and margin maximization but fails to solve the last one.

To compensate for asynchronization issues, we adopt matrix inputs as an extension of SVR to bring another dimension of time-series information in learning. The matrix extension avoids a very high-dimensional vector input and retains the separated formats for different timestamps in order. Moreover, we propose the nuclear norm regularization that imposes the matrix rank as the dependency of coefficient matrix for maintaining the structural information, differing from the vector form. In this way, the regularized matrix extension not only benefits the model degradability with sparsity but also retains the model convexity for an optimal solution.

V. Numerical Tests

The proposed support matrix regression for power flow mapping recovery is validated extensively on various test cases. Here we illustrate the results on test distribution grids, including the single-phase IEEE 8-bus system [6], IEEE 123-bus system that is customized into multi-phase format as in [5], and a utility distribution network with 362 bus nodes (150 are equipped with PVs) modified from [43]. To make the simulation closer to reality, we adopt historical load and PV (photovoltaics) generation data from a local utility that contains the smart meter readings (recorded per 15 minutes) from 2017 March to 2018 March. The corresponding voltage data is prepared by running load flow in MATPOWER [44], [45] and OpenDSS [46] for IEEE test systems and the utility feeder, respectively. For comparison, representative learning models from the literature are selected. The proposed method is compared with representative learning models in literature with the same setups as our experiments: linear regression (LR) with interpretability of underlying physics [8], [9], Lasso with \( L_1 \) norm penalty to regularize the fitting [6], and neural networks (NNs) with variable architectures [18], [32], [33]. Specifically, we use a 3-layer multilayer perceptron (MLP) with ReLu activation functions, a 3-layer RBF NN that has Gaussian transfer function for the hidden layer neurons to fully exploit system inputs, and Bilinear NN to encode properties of power flow equations into the bilinear mapping rules from bus voltages to power injections. For the MLP and RBF NN, the number of hidden neurons are the same as the input dimension [18], [32]. For the implementation of machine learning methods, the prepared data samples are split as 60% for training the model, 15% for validation of model parameters, and 25% for testing.

Since our goal is to recover the power flow mapping from data, we test the cases when grid topology and line parameters are unavailable during the training process. It is a fair assumption, as the system model may not be useful if there are many uncertainties. In order to evaluate the recovery with unobservability issues in distribution grids, we randomly selected 0-20% of buses to be unmeasurable (no measurements for voltages and power injections) during learning and recover the power flow mapping for the rest. To demonstrate SMR’s performance with matrix extension design, we simulate sensor asynchronizations scenarios by setting the time-series measurements of different buses to arbitrarily delay for 0 – 3 time slots. Besides, to consider the measurement errors, random noises that follow Gaussian distribution (zero mean and 1% relative standard deviation) are added to data, if not specified. 1% of the data are randomly changed to outliers by increasing their values 5 – 10 times larger. Later, the noise relative standard deviation and outlier proportions are changed to different levels to test SMR’s robustness against data issues using margin loss.

A. Degradability of SMR for Accuracy Guarantee

In this section, we show the performance of machine learning-based power flow mapping recovery. To focus on the degradability of the data-driven model, we assume no asynchronization error or large value deviations in data and leave them for later analysis. For evaluation, one intuitive way is to calculate the error between real outputs \( y \) and learning model outputs \( \hat{y} \). The mean absolute percentage error is used, \( \text{MAPE} = \frac{1}{N} \sum_{m=1}^{M} \left| \frac{y_m - \hat{y}_m}{y_m} \right| \), where the lower MAPE means the higher accuracy. The results with respect to different unobservability scenarios are shown in Table I. With full measurements, Table I(a) shows good performances for LR, Lasso, Bilinear NN, which have physical degradability. When system observability becomes limited, their performance decreases significantly because they all require full measurements of buses with power injections to fit the preprocessed function basis. In contrast, the NNs (MLP and RBF NN) result in much smaller errors, especially when there are more unmeasured buses in the system. The proposed SMR always performs better than NNs as NNs overfit training data and has larger testing errors. With sufficient observability, the 2\(^{nd}\)-degree kernel works best because it has the highest similarity with the underlying physical function. While in lack of observability scenario, using other kernels can outperform the 2\(^{nd}\)-degree kernel SMR to approximate the mapping regarding unmeasured nodes. After many times of training and testing, we find that the RBF kernel-based SMR is slightly better than using other kernels, especially when testing with a larger system (utility network) and training with more data. For example, MAPEs of using RBF kernel are smaller with more unmeasured buses, as shown in Table I(c). RBF kernel has a potentially infinite feature space, which is more flexible than the polynomial kernels for approximation under unobservability.

Since PMU are not widely deployed in distribution grids, we demonstrate the results with no voltage angle data for training in Table II. Compared to using voltage angle data in Table I (a), the learning errors have an overall slight increase. This is because the phase angle differences are not large (within 0.4 -1.1 degrees), which can be seen as noises in data. Especially, Lasso performs better than other methods in comparison due to its \( L_1 \) norm regularization for noises. Similar to Table I (a), SMR with 2\(^{nd}\)-degree polynomial kernel outperforms others. This is because, with measurements of all the buses, the 2\(^{nd}\)-degree polynomial kernel better aligns with the underlying power flow mapping. In contrast, more complicated kernels easily overfit due to data deviations.
TABLE I: MAPE results of recovering the power flow mapping under different observability.

| Test Model | IEEE-8-bus | IEEE-123-bus | Utility Network |
|------------|------------|--------------|-----------------|
| LR         | 0.065 ± 0.010 | 0.154 ± 0.013 | 0.176 ± 0.022 |
| Lasso      | 0.015 ± 0.010 | 0.042 ± 0.019 | 0.105 ± 0.047 |
| MLP        | 0.076 ± 0.012 | 0.115 ± 0.042 | 0.141 ± 0.063 |
| RBF NN     | 0.080 ± 0.015 | 0.124 ± 0.041 | 0.143 ± 0.026 |
| Bilinear NN| 0.059 ± 0.007 | 0.104 ± 0.035 | 0.147 ± 0.026 |
| SMR 2nd    | 0.016 ± 0.008 | 0.037 ± 0.022 | 0.086 ± 0.054 |
| 3rd        | 0.033 ± 0.012 | 0.077 ± 0.020 | 0.129 ± 0.037 |
| RBF        | 0.058 ± 0.009 | 0.073 ± 0.014 | 0.136 ± 0.014 |
| tanh       | 0.062 ± 0.009 | 0.078 ± 0.019 | 0.159 ± 0.032 |

(b) 10% unmeasured bus nodes.

| Test Model | IEEE-8-bus | IEEE-123-bus | Utility Network |
|------------|------------|--------------|-----------------|
| LR         | 0.114 ± 0.024 | 0.301 ± 0.056 | 0.337 ± 0.017 |
| Lasso      | 0.136 ± 0.031 | 0.199 ± 0.041 | 0.287 ± 0.036 |
| MLP        | 0.057 ± 0.012 | 0.108 ± 0.009 | 0.157 ± 0.010 |
| RBF NN     | 0.064 ± 0.033 | 0.099 ± 0.025 | 0.184 ± 0.023 |
| Bilinear NN| 0.168 ± 0.022 | 0.231 ± 0.005 | 0.314 ± 0.015 |
| SMR 2nd    | 0.053 ± 0.011 | 0.088 ± 0.028 | 0.182 ± 0.027 |
| 3rd        | 0.041 ± 0.021 | 0.093 ± 0.014 | 0.155 ± 0.042 |
| RBF        | 0.051 ± 0.005 | 0.117 ± 0.026 | 0.144 ± 0.036 |
| tanh       | 0.049 ± 0.007 | 0.123 ± 0.045 | 0.149 ± 0.068 |

(c) 20% unmeasured bus nodes.

| Test Model | IEEE-8-bus | IEEE-123-bus | Utility Network |
|------------|------------|--------------|-----------------|
| LR         | 0.262 ± 0.034 | 0.335 ± 0.011 | 0.413 ± 0.032 |
| Lasso      | 0.244 ± 0.027 | 0.304 ± 0.072 | 0.380 ± 0.059 |
| MLP        | 0.092 ± 0.031 | 0.146 ± 0.027 | 0.202 ± 0.041 |
| RBF NN     | 0.077 ± 0.041 | 0.142 ± 0.021 | 0.197 ± 0.039 |
| Bilinear NN| 0.174 ± 0.036 | 0.276 ± 0.012 | 0.401 ± 0.022 |
| SMR 2nd    | 0.098 ± 0.023 | 0.154 ± 0.049 | 0.241 ± 0.013 |
| 3rd        | 0.078 ± 0.032 | 0.132 ± 0.031 | 0.193 ± 0.014 |
| RBF        | 0.076 ± 0.029 | 0.129 ± 0.033 | 0.208 ± 0.017 |
| tanh       | 0.083 ± 0.031 | 0.136 ± 0.027 | 0.222 ± 0.032 |

For power flow mapping recovery, a more important metric is the degradability to the underlying physical model, which serves as a performance guarantee to avoid numerical overfitting. Since back-box NN fails to express physical function, we show the comparison of recovered line parameters for LR and SMR with respect to different levels of measurement errors in Fig. 2. For visualization, all the system line parameters are normalized, meaning the closer to 1, the more accurate expression of the power flow mapping. The solid curve represents the average recovered $g_{ik}$ and $b_{ik}$ values of all normalized line parameters in observable areas, and the shaded area is their standard deviations. It is clear that, given noiseless measurements, both LR and SMR can learn accurate line parameters due to the degradable function approximation of power flow mapping. However, when disturbances increase in data, LR causes errors as not only the average values deviate from the true values but also the standard deviations become larger. As an example, the worst case can be less than 80% recovery of the true values for IEEE 123-bus system. In this scenario, SMR could still achieve stable $g_{ik}$’s and $b_{ik}$’s, providing the guarantee of the performance in Table I(a). Thus, by using two evaluation metrics, we show SMR possesses the degradability of observable areas as well as robust approximation regarding unobservability.

B. Applicability to Asynchronization Scenario

The measurements from different sources bring in uncertainties of data synchronization, which could perturb the learning process. Therefore, the performance of power flow mapping recovery is tested with asynchronized data sequences, which have an arbitrary delay for $0 - 3$ time slots. For SMR, the time sequence $\tau$ of each matrix is selected by the validation process, which is 3 in this test. The comparison of different methods’ recovery accuracy is shown in Table III.

Full measurements are considered under this test scenario, and the results are compared with the base case without the
TABLE III: MAPE results of recovering the power flow mapping from asynchronized data for different systems.

| Test Model | IEEE 8-bus | IEEE 123-bus | Utility Network |
|------------|------------|--------------|----------------|
| LR         | 0.139 ± 0.024 | 0.207 ± 0.029 | 0.312 ± 0.023 |
| Lasso      | 0.109 ± 0.031 | 0.176 ± 0.045 | 0.288 ± 0.041 |
| MLP        | 0.112 ± 0.031 | 0.149 ± 0.040 | 0.259 ± 0.035 |
| RBF NN     | 0.099 ± 0.025 | 0.170 ± 0.033 | 0.79 ± 0.027  |
| Bilinear   | 0.121 ± 0.013 | 0.189 ± 0.042 | 0.300 ± 0.021 |

PRBNN 2° 0.022 ± 0.017 0.039 ± 0.016 0.092 ± 0.029
SMR 3°rd 0.061 ± 0.022 0.079 ± 0.018 0.134 ± 0.033
RBF 0.054 ± 0.011 0.082 ± 0.008 0.143 ± 0.037
tanh 0.049 ± 0.017 0.085 ± 0.012 0.132 ± 0.046

C. Robustness Against Noisy Data and Varying Energy Levels

In poorly measured distribution grids, the ability to withstand errors in data is crucial. To validate the robust design of SMR, we vary the noise relative standard deviation and data outlier proportion from 1% to 5%. Despite the associated recovery of line parameters in Fig. 2, we show the compared learning performance (MAPE) with respect to noise levels in Fig. 3.

![Fig. 3: Compare the performance with respect to different levels of errors in measurements for IEEE 123-bus system.](image)

It is obvious that the MAPE of LR has a fast increase along with larger measurement errors. On the contrary, the proposed SMR model, no matter which kernel is chosen, shows stable MAPEs with increasing measurement noises. The inherent margin maximization design tolerates these data deviations during the learning process.

Moreover, we explore the literature and study representative works for handling bad data [6], [24], [47]. In specific, the conventional way is bad data detection, e.g., the Chi-squares ($\chi^2$) test [24]. Once bad data is detected, they need to be identified and eliminated before learning. To avoid error propagation, Lasso uses $L_1$ norm penalty to regularize the fitting [6]. The basic idea is still to filter out the bad data during learning process. In contrast, the proposed SMR treats bad data differently in learning, using margin maximization to identify support matrices with $\epsilon$-deviations and thus find the correct mapping. In addition, if the time asynchronization issues are treated as bad data, they are usually filtered out before learning. Since this issue may cover several time slots, sequences of data can not lead to the correct mapping so that the regularized matrix extension design of SMR is beneficial.

TABLE IV: Comparison of different methods in handling data issues.

| Test Model | IEEE 8-bus | IEEE 123-bus | Utility Network |
|------------|------------|--------------|----------------|
| $\chi^2$ Test + LR | 0.013 ± 0.0010 | 0.038 ± 0.0112 | 0.130 ± 0.0122 |
| Lasso      | 0.015 ± 0.0100 | 0.042 ± 0.0119 | 0.105 ± 0.0147 |
| PRBNN 2°  | 0.016 ± 0.0088 | 0.037 ± 0.0222 | 0.086 ± 0.0544 |
| SMR RBF    | 0.058 ± 1.0099 | 0.073 ± 0.0114 | 0.136 ± 0.0114 |

(a) Without asynchronization errors.

| Test Model | IEEE 8-bus | IEEE 123-bus | Utility Network |
|------------|------------|--------------|----------------|
| $\chi^2$ Test + LR | 0.092 ± 0.0209 | 0.193 ± 0.0144 | 0.295 ± 0.0439 |
| Lasso      | 0.109 ± 0.0311 | 0.176 ± 0.0454 | 0.288 ± 0.0414 |
| PRBNN 2°  | 0.022 ± 0.0117 | 0.039 ± 0.0160 | 0.092 ± 0.0298 |
| SMR RBF    | 0.054 ± 0.0111 | 0.082 ± 0.0088 | 0.113 ± 0.0373 |

(b) With asynchronization errors.
enhance the learning model’s robustness under noisy scenarios. The design of margin maximizing and error specifications can be acceptable. In this way, the experiment results illustrate that for a system with unobservability or short of degradability, errors from measurements, the data-driven approach is invalid. In addressing some concerns. For example, the historical rules from historical data, existing works have limitations in power flow models. Nevertheless, unavailable system topology integration of DER allocations and violation criteria. For future work, we will investigate the proposed method accommodating different types of time synchronization errors practically. On the other hand, we will extend the SMR-based power flow model for further studies in distribution systems. Take hosting capacity analysis as an example, if the system is fully observable, we can recover the complete power flow model and conduct the analysis. If unobservability exists, it is worth extending the proposed SMR work with the consideration of DER allocations and violation criteria.

VI. CONCLUSION
The sustainable operation of distribution grids needs to implement advanced functionalities, many of which are based on power flow models. Nevertheless, unavailable system topology and parameters make it infeasible to conduct conventional analysis. While one can learn the parameters or mapping rules from historical data, existing works have limitations in addressing some concerns. For example, the historical measurements have practical issues such as noises and asynchronization to deteriorate learning performance. Even without errors from measurements, the data-driven approach is invalid for a system with unobservability or short of degradability to the true physical power flow model. In this paper, we propose effective designs to resolve these issues for power flow mapping reproduction in distribution grids, which are integrated into the support matrix regression. On the one hand, by matrix imputation, we regularize expanded time-sequence measurements against the sensor asynchronization issues. On the other hand, its flexible kernel embedding not only enables the degradability to the underlying power flow model as a performance guarantee but also enhances the approximation for highly nonlinear mapping regarding the unobservability. In addition, the inherent margin maximizing loss saves the extra efforts on denoising data, which significantly reduces the overfitting risk. Numerical validation on several test systems verifies SMR’s superiority compared to other learning methods in recovering power flow mapping.

For future work, we will investigate the proposed method accommodating different types of time synchronization errors practically. On the other hand, we will extend the SMR-based power flow model for further studies in distribution systems. Take hosting capacity analysis as an example, if the system is fully observable, we can recover the complete power flow model and conduct the analysis. If unobservability exists, it is worth extending the proposed SMR work with the consideration of DER allocations and violation criteria.

REFERENCES
[1] Y. Weng, Y. Liao, and R. Rajagopal, “Distributed energy resources topology identification via graphical modeling,” IEEE Transactions on Power Systems, pp. 2682–2694, Jul 2017.
[2] M. Vadari, “The future of distribution operations and planning: The electric utility environment is changing,” IEEE Power and Energy Magazine, pp. 18–25, Feb 2020.
[3] B. Zhang, A. Y. S. Lam, A. D. Domínguez-García, and D. Tse, “An optimal and distributed method for voltage regulation in power distribution systems,” IEEE Transactions on Power Systems, pp. 1714–1726, Jul 2015.
[4] J. Yu, Y. Weng, and R. Rajagopal, “Robust mapping rule estimation for power flow analysis in distribution grids,” 2017 North American Power Symposium (NAPS), pp. 1–6, Sep 2017.
[5] Y. Liao, Y. Weng, G. Liu, Z. Zhao, C.-W. Tan, and R. Rajagopal, “Unbalanced multi-phase distribution grid topology estimation and bus phase identification,” IET Smart Grid, pp. 557–570, Dec 2019.
[6] Y. Liao, Y. Weng, G. Liu, and R. Rajagopal, “Urban mv and lv distribution grid topology estimation via group lasso,” IEEE Transactions on Power Systems, pp. 12–27, Jan 2019.
[7] X. Miao, M. Ilić, X. Wu, and U. Münz, “Distribution grid admittance estimation with limited non-synchronized measurements,” 2019 IEEE Power Energy Society General Meeting, pp. 1–5, Aug 2019.
[8] Y. Liu, N. Zhang, Y. Yang, J. Yang, and C. Kang, “Data-Driven Power Flow Linearization: A Regression Approach,” IEEE Transactions on Smart Grid, pp. 2569–2580, May 2019.
[9] Y. Yuan, O. Ardakanian, S. Low, and C. Tomlin, “On the inverse power flow problem,” arXiv preprint, Feb 2020.
[10] E. A. Bueno, C. Lyra, and C. Cavallucci, “Distribution network reconfiguration for loss reduction with variable demands,” IEEE/PES Transmission and Distribution Conference and Exposition, pp. 384–389, Nov 2004.
[11] O. F. Fajardo and A. Vargas, “Reconfiguration of MV distribution networks with multiset and multipoint alternative supply, part II: Reconfiguration plan,” IEEE Transactions on Power Systems, pp. 1401–1407, Aug 2008.
[12] R. A. Jabr, “Minimum loss operation of distribution networks with photovoltaic generation,” IET Renewable Power Generation, pp. 33–44, Jan 2014.
[13] E. Serhat Berat and M. CEBECI, “Power flow analysis by artificial neural network,” International Journal of Energy and Power Engineering, pp. 204–208, Nov 2013.
APPENDIX

A. Illustrating the Effectiveness of Matrix Extension in SMR

In the following, we show that expanding the input data to the matrix formula remains the key properties. First, the margin maximization is commonly implemented over vectors. Hence, let $w = \text{vec}(W^T)$, where $\text{vec}()$ is the vectorization operator that transfers a matrix into a vector by stacking its columns on top of one another. Then, we derive

$$\frac{1}{2}||W||_F^2 = \frac{1}{2} \text{tr}(W^T W),$$

$$= \frac{1}{2} \text{vec}(W^T)^T \text{vec}(W^T),$$

$$= \frac{1}{2} w^T w,$$

$$= \frac{1}{2} ||w||^2.$$

Therefore, minimizing the squared Frobenius norm of matrix $W$ equals to minimizing the norm of vector $w$, which is equivalent to maximizing margin over vectors. Similarly, the approximated function $h_W(X) = \{W, X\}$ is equivalent to $h_{W, X} = \langle w, [x_{t-\tau}, \ldots, x_t, \ldots, x_{t+\tau}] \rangle$. Without loss of generality, SMR can be degraded to the vector version by setting $c_{\tau} = 0$. Thus, SMR possesses the elegant properties from the vector version to admit the solution pattern in (7) and (8). On the other hand, the added nuclear norm is a regularization on the rank, which is unique for matrices to consider the correlation in rows and columns. Such a norm preserves the convexity of (3) that we can find the optimal recovery of power flow mapping.

B. Illustrating SMR Solution of Different Data Scenarios

In general, $\alpha$ and $\alpha^*$ are weights to characterize support matrices from all the data points for power flow mapping recovery. Specifically, $\alpha$ and $\alpha^*$ are the Lagrangian multipliers of constraints (4) and (5) to simplify the original problem (3) to a maximization dual problem. After solving the dual
problem, the regression coefficients are recovered from the solutions of $\alpha$ and $\alpha^*$ using $W = \sum_{t=1}^{T} (\alpha_t - \alpha^*_t) \phi(X_t)$. Accordingly, they are the weights assigned to characterize which data samples become support vectors so that the located hyperplane is optimal [48].

Specifically, consider each data sample as a point in high-dimensional feature space, where a simplified example is visualized in Fig. 1. The dual problem requires $\sum_{t=1}^{T} (\alpha_t - \alpha^*_t) = 0$ and $\alpha_t, \alpha^*_t \in [0, c_1]$ [36]. First, it is obvious that one data point can not lie on both sides of error bounds for $\alpha$ and $\alpha^*$ to be jointly non-zero. Then, when the data point is inside the $\epsilon$-tube, $\alpha = \alpha^* = 0$. The data deviation is implicit so that it is unnecessary to describe $W$ and, in turn, contributes to the sparsity of solution. When the data point lies on the $\epsilon$-tube boundaries, $\alpha$ or $\alpha^* \in (0, c_1)$, these points can become error-free data as the error is explicitly $\pm \epsilon$. Thus, they can precisely identify the hyper-plane, i.e., the accurate power flow mapping, and they are called support matrices. Besides, when the data point is outside the $\epsilon$-tube, $\alpha$ or $\alpha^* = c_1$ and slack variables $\xi, \xi^*$ are used to specify such large error without impacting the hyper-plane. Such data points are also called bounded support matrices.

Jingyi Yuan (Student Member, IEEE) received the B.Sc. degree in electrical engineering from the Huazhong University of Science and Technology, Wuhan, China, and the M.Sc. degree in electrical engineering from Arizona State University (ASU). She is currently serving as a Research Assistant and pursuing a Ph.D. degree in electrical, computer and energy engineering at Arizona State University, Tempe, AZ, USA. In recognition of her extraordinary academic achievements, she won University Graduate Fellowship and Engineering Graduate Fellowship. Her research interests are in the areas of machine learning and big data applications in power systems, explainable data-driven modeling of cyber-physical systems, and dynamic hosting capacity of photovoltaics. She won the Best Paper (Presentation) Award of the 51st IEEE North American Power Symposium in 2019.

Yang Weng (Senior Member, IEEE) He received the B.E. degree in electrical engineering from the Huazhong University of Science and Technology, Wuhan, China, the M.Sc. degree in statistics from the University of Illinois at Chicago, Chicago, IL, USA, and the M.Sc. degree in machine learning of computer science and the M.E. and Ph.D. degrees in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, USA. He joined Stanford University, Stanford, CA, USA, as the TomKat Fellow for sustainable energy. He is currently an Assistant Professor of electrical, computer, and energy engineering with Arizona State University, Tempe, AZ, USA. His research interest is in the interdisciplinary area of power systems, machine learning, and renewable integration. He received the CMU Dean’s Graduate Fellowship in 2010, the Best Paper Award at the International Conference on Smart Grid Communication (SGC) in 2012, the First Ranking Paper of SGC in 2013, the Best Papers at the Power and Energy Society General Meeting in 2014, the ABB Fellowship in 2014, the Golden Best Paper Award at the International Conference on Probabilistic Methods Applied to Power Systems in 2016, and the Best Paper Award at IEEE Conference on Energy Internet and Energy System Integration in 2017, IEEE North American Power Symposium in 2019, and the IEEE Sustainable Power and Energy Conference in 2019.