Machine Learning for Scalable and Optimal Load Shedding Under Power System Contingency

Yuqi Zhou, *Member, IEEE*, and Hao Zhu, *Senior Member, IEEE*

Abstract—Prompt and effective corrective actions in response to unexpected contingencies are crucial for improving power system resilience and preventing cascading blackouts. The optimal load shedding (OLS) accounting for network limits has the potential to address the diverse system-wide impacts of contingency scenarios as compared to traditional local schemes. However, due to the fast cascading propagation of initial contingencies, real-time OLS solutions are challenging to attain in large systems with high computation and communication needs. In this paper, we propose a decentralized design that leverages offline training of a neural network (NN) model for individual load centers to autonomously construct the OLS solutions from locally available measurements. Our learning-for-OLS approach can greatly reduce the computation and communication needs during online emergency responses, thus preventing the cascading propagation of contingencies for enhanced power grid resilience. Numerical studies on both the IEEE 118-bus system and a synthetic Texas 2000-bus system have demonstrated the efficiency and effectiveness of our scalable OLS learning design for timely power system emergency operations.

Index Terms—Optimal load shedding, decentralized control, deep learning, line outages, grid emergency operations.

I. INTRODUCTION

Prompt and effective responses to emergency events are crucial for enhancing power system resilience. Cascading failures could be mitigated via corrective actions by restoring power balance and maintaining network limits. Load shedding is such an emergency response commonly employed by grid operators via disconnecting the non-critical loads. Unlike normal operations, the decision-making of emergency responses is timing-critical. Due to the computation and communication needs for effective load shedding, machine learning (ML) approaches provide a unique opportunity for designing timely emergency responses by leveraging their online prediction capability.

Traditional load shedding approaches rely on localized automatic control mechanisms. To quickly restore *power balance*, Under Frequency Load Shedding (UFLS) [1]–[3] and

Y. Zhou is with the Power Systems Engineering Center, National Renewable Energy Laboratory, Golden, CO, 80401, USA (email: yuqi.zhou@nrel.gov).

H. Zhu is with the Chandra Family Department of Electrical & Computer Engineering, The University of Texas at Austin, Austin, TX, 78712, USA (email: haozhu@utexas.edu).

Under Voltage Load Shedding (UVLS) [4]–[6] are designed to respond to the drops in locally observed frequency or voltage using a fixed rate of reduction; see also the NERC standards in [7]. This proportional rule is easy to implement locally but fails to account for the grid-wide heterogeneous effects of the contingency scenario such as severe congestion on certain lines. To address this issue, the optimization-based load shedding scheme (e.g., [8]–[10]) has become popular, relying on centralized dispatch using advanced monitoring and communications across the system. The optimal load shedding (OLS) can effectively mitigate the risks of cascading failures through strategically targeting system-wide congestion or *network limits*. For example, convex relaxation is used in [8], [11] to address the nonlinearity of AC power flow, while the fairness of power shut-offs is considered in [12]. Nevertheless, solving this optimization in real time is still challenged by the high computational complexity and communication needs.

Similar to the rising adoption of machine learning (ML) for optimal power flow (OPF) [13], there is a surge of interest in using ML for online AC-OLS implementation. In a stressed system with power imbalance, failure to perform effective load shedding to stabilize the grid could trigger cascading failures within seconds to a few minutes [14]–[16]. Thus, to fully leverage the benefits of a centralized load shedding scheme while addressing the time-sensitive nature of such implementation, ML is ideally suited to enable prompt decision-making. A graph neural network-based approach is developed in [17] that requires the system-wide contingency information and observability, while reinforcement learning is considered in [18]–[20] for obtaining dynamic load shedding policies using the grid-wide system states as the input. Please see Table I for a detailed comparison of existing approaches. Although a centralized optimization/ML is suitable for grid decision-making during normal conditions, it is less attractive for timing-critical emergency responses due to the stringent requirements on sensing and communications capabilities. Therefore, it remains to develop autonomous OLS solutions that can cope up with the real-time needs at distributed load centers by providing timely corrective actions in response to sudden contingencies.

This paper aims to develop a learning-for-OLS framework to enable individual load centers to promptly form their own OLS decisions in a decentralized fashion. We first formulate the AC-OLS problem of determining the amount of nodal load shedding, to restore the system-wide power balance, and most importantly, mitigate the violations of network limits. Upon solving this problem under a range of loading conditions and contingency scenarios, one can learn the decision rules

TABLE I: Comparisons of Existing Approaches and Our Work

	Objective	Method	Decentralization	Techniques
[3]	Power balance	Optimization	Centralized	Particle swarm optimization
[8], [11]	Network limits	Optimization	Centralized	Convex relaxation
[9]	Network limits	Optimization	Centralized	Deterministic optimization
[24], [25]	Network limits	Optimization	Centralized	Genetic algorithm
[26]–[28]	Power balance	Multi-agent system	Distributed	Consensus-based algorithm
[17]	Network limits	Machine learning	Centralized	Graph neural networks
[29], [30]	Power balance	Machine learning	Centralized	Support vector machine
[31], [32]	Power balance	Machine learning	Centralized	Decision tree
[18], [20], [33]	Power balance	Machine learning	Centralized	Reinforcement learning
Our work	Network limits	Machine Learning	Decentralized	NN for a decentralized design

that map from input system conditions to target OLS outputs through offline training of NN models to capitalize on their excellent approximation capability. One notable feature is the *decentralized* design of the proposed NN-based OLS decision rules, which have been constructed from the *locally available measurements* at each load center only. Different from our preliminary work [21], this paper has judiciously designed the NN prediction outputs to be the nodal Lagrange multipliers of the OLS problem. The latter is known to enjoy a local dependence property [22], [23], thus ideally suited for the decentralized design. We have further developed the identifiability analysis based on the simplified DC power flow for understanding the inference capability of using local data only. The scalability of our design can significantly reduce the offline training time as compared to a centralized framework. Most importantly, these resultant local OLS decision rules allow for each load center to promptly and autonomously react to the specific contingency in real time, without the need to communicate to the control center. This way, it effectively eliminates the negative impacts of communication latency and reliability issues that are critical for emergency operations. The main contributions of our work include the following:

- A decentralized learning-for-OLS framework is proposed, allowing each load center to promptly predict its own OLS solution in response to emergency events.
- The ML algorithm is designed to use local measurements to quickly predict local dual variables of the OLS problem, enabling accurate mapping to corrective actions during real-time operations.
- An identifiability analysis is formally performed to discuss the use of solely local measurements and support the validity of our decentralized OLS design.

This paper is organized as follows. Section II formulates the AC-OLS as a nonlinear optimization problem. Section III presents the design of our proposed scalable OLS method, building upon offline training and computation to attain autonomous online responses. The identifiability analysis is presented in Section IV, to showcase the capability of differentiating contingencies using local data. In Section V, numerical simulations on both the IEEE 118-bus case and the synthetic Texas 2000-bus system have demonstrated the accuracy and efficiency of the scalable OLS design for real-time emergency operations. Last, the paper is concluded in Section VI.

Notation: $(\cdot)^{\mathsf{T}}$ stands for transposition; $(\cdot)^{\star}$ stands for the conjugate of a complex number; $(\cdot)^{*}$ stands for the optimal solution; $|\cdot|$ denotes the absolute value of a complex number;
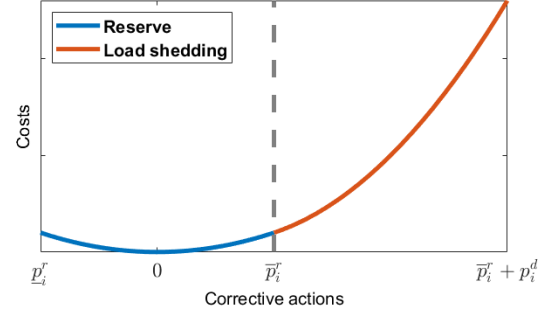

Fig. 1: A piece-wise quadratic cost function for the active power flexibility (both reserve and load shedding) per bus i.

$\|\cdot\|$ denotes the vector Euclidean norm; $\mathbb{R}$ denotes real numbers; $\mathbb{R}^+$ denotes positive real numbers.

II. AC-OLS PROBLEM FORMULATION

We formulate the optimal load shedding (OLS) problem under the nonlinear AC power flow model. Consider a transmission system with N buses and L lines collected in the sets $\mathcal{N}$ and $\mathcal{L}$, respectively. Let $\boldsymbol{Y} = \boldsymbol{G} + \mathrm{j}\boldsymbol{B}$ denote the complex network admittance matrix of size $N \times N$. Per bus i, we denote the complex power output from its connected generation by $p_i^g + \mathrm{j}q_i^g$, and from load demand by $p_i^d + \mathrm{j}q_i^d$. In addition, let $V_i\angle\theta_i$ denote the bus voltage phasor with the magnitude V_i and phase angle θ_i. Per line $(i, j) \in \mathcal{L}$, the complex power flow from bus i to j becomes:

$$S_{ij} = Y_{ij}^{\star}V_i^2 - Y_{ij}^{\star}V_iV_j\angle(\theta_i - \theta_j) \tag{1}$$

using the (i, j)-th entry of matrix $\boldsymbol{Y}$.

The AC-OLS problem is essentially an extension of the optimal power flow (OPF). The latter determines the optimal set-points of controllable resources under *normal operations* by accounting for the grid operation limits. During *emergency operations* arising from large contingencies, AC-OPF can become infeasible with insufficient resource flexibility or transfer capability. Thus, using generation reserves, and more drastic corrective actions like load shedding, are imperative for restoring power balance while satisfying operational limits. With known nominal injection $p_i := p_i^g - p_i^d$ and similarly q_i per bus i, the OLS problem aims to determine the amount of generation reserves and the load demand to shed, as

$$\min \ \sum_{i=1}^{N} c_i^p(p_i^s) \tag{2a}$$

$$\text{s. t. } V_i \in \mathbb{R}^+, \theta_i \in \mathbb{R}, p_i^s \in \mathbb{R}, q_i^s \in \mathbb{R}, \forall i \in \mathcal{N} \tag{2b}$$

This article has been accepted for publication in IEEE Transactions on Power Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2025.3528434

3

$$\underline{V}_i \leq V_i \leq \overline{V}_i, \ \underline{\theta}_i \leq \theta_i \leq \overline{\theta}_i \tag{2c}$$

$$\underline{p}_i^r \leq p_i^s \leq \overline{p}_i^r + p_i^d, \ \underline{q}_i^r \leq q_i^s \leq \overline{q}_i^r + q_i^d \tag{2d}$$

$$|S_{ij}| \leq \overline{S}_{ij}, \ \forall (i,j) \in \mathcal{L} \tag{2e}$$

$$\theta_{ij} = \theta_i - \theta_j, \ \forall (i,j) \in \mathcal{L} \tag{2f}$$

$$p_i + p_i^s = \sum_{j=1}^{N} V_i V_j \left(G_{ij} \cos \theta_{ij} + B_{ij} \sin \theta_{ij} \right) \tag{2g}$$

$$q_i + q_i^s = \sum_{j=1}^{N} V_i V_j \left(G_{ij} \sin \theta_{ij} - B_{ij} \cos \theta_{ij} \right). \tag{2h}$$

For the decision variables given in (2b), constraints in (2c)-(2e) enforce the operational limits or resource budget. Note that the line thermal limit in (2e) in terms of apparent power flow can also include other metrics such as line current or active power flow. Finally, the equality constraints in (2f)-(2h) represent the AC power flow model in (1). Note that our AC-OLS formulation includes both active and reactive powers as decision variables, with the cost function (2a) focusing on active power only. Similar to AC-OPF, the balance of reactive power in (2) is crucial for maintaining voltage stability, and it could automatically adjusted via local voltage control in practice. Thus, while (2) follows from earlier work [3], [8], [11] to include reactive power flexibility, the proposed method does not need to consider the prediction of these decisions. A simplified DC-OLS formulation with active power decisions only is also possible; see e.g., [10], [34]. Note that most of the cases, the AC-OLS problem (2) will automatically search for an optimal operating point that ensures feasibility, thanks to the flexibility provided by load shedding and reserves. However, in very extreme cases, it is possible that the contingency can cause infeasibility of the problem, and such scenarios can be handled by further allowing for disconnection of other components [8] in the network.

Remark 1 (Cost function design)**.** *The decision variables $\{p_i^s\}$ in (2) represent the flexibility from both utilizing generation reserves and performing load shedding at each bus. As shown in Fig. 1, we consider a piece-wise quadratic cost:*

$$c_i^p(p_i^s) := \begin{cases} a_{i,1}(p_i^s)^2, & \underline{p}_i^r \leq p_i^s \leq \overline{p}_i^r \\ a_{i,2}(p_i^s)^2 + b_{i,2}p_i^s + c_{i,2}, & \overline{p}_i^r \leq p_i^s \leq \overline{p}_i^r + \overline{p}_i^d \end{cases} \tag{3}$$

Here, $\underline{p}_i^r$ and $\overline{p}_i^r$ represent the limits of downward and upward reserves, respectively. Once all the reserves are utilized, the more expensive load-shedding option can be employed, capped at a maximum of $\overline{p}_i^d$. The latter can be smaller than the full loading of p_i^d, due to the presence of non-flexible critical loads. The two quadratic functions vary in incremental costs, with load shedding being more costly to initiate, and thus reserves are prioritized over load shedding.

Under such cost function design, the optimization problem (2) will automatically determine the use of reserves and load shedding. Specifically, when no correction actions are needed, the cost function is minimized at $p_i^s = 0$. When minor corrections such as reserves are necessary, the variable p_i^s takes values in the range of $[\underline{p}_i^r, 0]$ (downward reserves) or $[0, \overline{p}_i^r]$ (upward reserves). Whenever the optimal solution falls within $[\underline{p}_i^r, \overline{p}_i^r]$ (blue piece in Fig. 1), it indicates that only reserves are utilized. When all the reserves are utilized, the more expensive load shedding operations are activated. Solutions p_i^{s*} in this scenario (orange piece in Fig. 1) indicate that both reserves and load shedding are utilized, where the reserves are equal to $\overline{p}_i^r$ and the optimal load shedding equals $p_i^{s*} - \overline{p}_i^r$.

In addition, the coefficient a_2 can vary among the buses based on the priority of the connected loads. For the function (3), the following conditions hold at the point $\overline{p}_i^r$ in general:

$$a_{i,1}(\overline{p}_i^r)^2 = a_{i,2}(\overline{p}_i^r)^2 + b_{i,2}\overline{p}_i^r + c_{i,2}, \tag{4}$$

$$2a_{i,1}\overline{p}_i^r \leq 2a_{i,2}\overline{p}_i^r + b_{i,2}. \tag{5}$$

The first condition guarantees the continuity of c_i^p, while the second ensures the incremental cost of reserves will not exceed that of load shedding. Hence, the AC-OLS problem mainly differs from AC-OPF in terms of the level of flexibility that each load center can provide. It is also worth pointing out that while our OLS problem (2) only considers load shedding and regulating reserves, other corrective actions such as topology optimization and full de-energization of system components can be included, as well; see e.g., [8], [9]. Notably, in this work, we focus on optimization models that allow for rapid adjustments to the system's operating point to prevent power imbalances. After the load shedding actions are initiated, the dynamic behavior of the system will be managed with frequency control [35], [36]. The process generally involves primary control, secondary control (e.g., Automatic Generation Control), and tertiary control. These mechanisms are essential for managing frequency deviations [37] and thus ensure system stability.

Remark 2 (Solving AC-OLS)**.** *As a nonlinear program (NLP), the AC-OLS problem is nonconvex and generally NP-hard, similar to the AC-OPF [38]. Some convex relaxation methods can be adopted to tackle the nonconvexity of AC power flow equations, see the OPF work in [39], [40] and similarly the OLS one [8]. Fortunately, open-source packages such as MATPOWER and JuMP can be used to solve the NLP form of AC-OLS for practically sized power systems.*

III. SCALABLE OLS DESIGN

Albeit solvable in a centralized fashion, real-time OLS decision-making can significantly benefit from a decentralized learning approach to accelerate the responses to emergency events. During the onset of contingency, corrective actions are pressingly needed in order to mitigate severe damages and the chances of cascading failures to the rest of the grids. Nonetheless, solving the centralized OLS problem in a timely manner could be impractical, due to its needs of communications and network coordination. In particular, an effective online implementation of centralized OLS critically requires high-rate low-latency communications, such that the control center can promptly acquire system-wide information and further dispatch decentralized resources. However, the availability of such communication capability is challenging, especially during weather disasters or other disruptive events. Hence, to enable fast and effective OLS actions in real time, we put forth a scalable learning framework by leveraging offline training to facilitate prompt online decision-making.
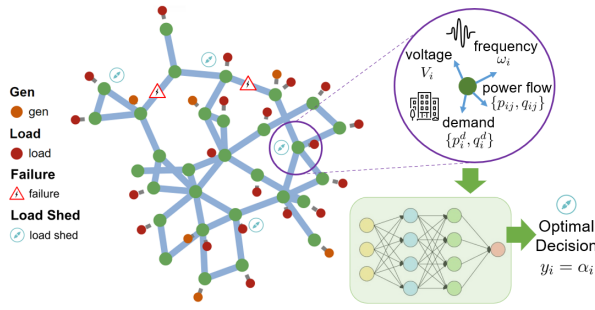
Fig. 2: Illustration of the proposed scalable OLS design.

Our proposed design aims to enable each load center to predict its own OLS decisions from locally available measurements only. For the prediction output, we choose the local *marginal costs* of AC-OLS, namely the sensitivity factors of the overall cost over each individual nodal flexibility. This is related to the Lagrange multipliers-based duality analysis for the active and reactive power balance constraints in (2f)-(2h). As shown by the recent OPF learning work [23], [41], [42], predicting the locational marginal prices (LMPs), as the sensitivity counterpart for the OPF problem, can be much more effectively done using localized information only. This is because the LMPs enjoy the *locality* property [22]; and similarly for the sensitivity analysis for generalized grid optimization problems like OLS. Inspired by these OPF-learning approaches, our decentralized OLS learning design will similarly predict the local marginal costs instead of the primal variables.

Let α_i represent the Lagrange multipliers for constraint (2g) per bus i. The optimality conditions by taking the upper/lower limits in (2d) as implicit constraints, state that the optimal solutions for p_i^s become:

$$\hat{p}_i^s = \underset{\underline{p}_i^r \leq p_i^s \leq \overline{p}_i^r + p_i^d}{\arg\min} \; c_i^p(p_i^s) - \alpha_i p_i^s, \qquad (6)$$

The relation is in line with the LMP-based OPF duality analysis, where the prices fully determine the generator dispatch. While the duality analysis of AC-OLS or AC-OPF is an important topic, it goes beyond the scope of our work; see e.g., earlier work [43]–[45] for more details. Using the piecewise quadratic cost function in (3), we can express the unique optimal solutions in closed-form:

$$\hat{p}_i^s = \begin{cases} \left(\dfrac{\alpha_i - b_{i,2}}{2a_{i,2}}\right)_{[\overline{p}_i^r, \overline{p}_i^r + p_i^d]}, & \text{if } \alpha_i > 2a_{i,2}\overline{p}_i^r + b_{i,2} \\ \left(\dfrac{\alpha_i}{2a_{i,1}}\right)_{[\underline{p}_i^r, \overline{p}_i^r]}, & \text{if } \alpha_i \leq 2a_{i,2}\overline{p}_i^r + b_{i,2} \end{cases} \qquad (7)$$

where the subscript indicates the projection to the corresponding box interval. Upon predicting the marginal costs, namely $\hat{\alpha}_i$, the load center at bus i can uniquely determine its own load shedding decisions $\hat{p}_i^s$. This equivalent condition will be utilized in our OLS learning design.

Fig. 2 illustrates the overall architecture of our scalable OLS design using an example system of the IEEE 30-bus test case. At each load center (or possibly a sub-area), local measurements such as voltage, frequency, power demand, and incident line flows are readily available from localized meters. Using these data during both pre- and post-contingency, the load center aims to predict the aforementioned OLS outputs, the marginal costs, and accordingly, optimal decisions, based upon a pre-trained neural network (NN). This way, no real-time communications are required between the control center and distributed loads. Of course, the feasibility of such a direct prediction from limited data within a local area comes into question. While there are no communications to exchange information, the post-contingency data at each load center reflect the physics of the underlying interconnection and contingency scenario according to the network power flow coupling. Thus, if each potential contingency exhibits a unique pattern of power flow changes at any load center, it becomes possible to infer the associated OLS decisions from its subset of measurements. We will formally analyze this inference capability in Section IV.

To enable the scalable OLS design, extensive offline computations are utilized to obtain the best mapping from the available measurements to real-time OLS decisions. We consider the standard feedforward NN for this scalable mapping, as the input-output variables are localized within each load center i. To formalize this prediction, we consider the input feature vector collecting all possible local measurements, given by:

$$\boldsymbol{x}_i = \left[p_i^d, q_i^d, V_i, \omega_i, \{p_{ij}\}_{j\in\mathcal{N}_i}, \{q_{ij}\}_{j\in\mathcal{N}_i}\right], \; \forall i, \qquad (8)$$

which includes local active/reactive demands, voltage magnitude, electric frequency, and incident active/reactive line flows. Note that both pre- and post-contingency values are used, except for the demands which could be unchanged throughout the contingency. Using both conditions increases the possibility of inferring the underlying contingency and thus the associated OLS decisions, as the changes from pre- to post-contingency power flows can differentiate the specific contingency. For example, both V_i and ω_i are good indicators of the system loading stress level. Furthermore, the changes in line flows can effectively provide the identifiability on the location of outaged lines due to the power flow redistribution. As discussed earlier, we consider the NN outputs to be the local marginal costs of the AC-OLS problem, as given by:

$$y_i = \alpha_i. \qquad (9)$$

This output per bus i enables us to directly compute the optimal corrective actions by comparing to the local OLS costs as in (6). Convenient implementation of this step is possible during the online phase based upon (7), thus attaining a fully decentralized OLS decision-making as long as the load center at bus i can predict y_i.

We can obtain the decision rule that maps from local data $\boldsymbol{x}_i$ to local outputs y_i, namely, $f_i(\boldsymbol{x}_i; \boldsymbol{\varphi}_i) \rightarrow y_i$ with $\boldsymbol{\varphi}_i$ denoting the NN parameters to be trained. To pursue a powerful decision rule $f_i(\cdot; \boldsymbol{\varphi}_i)$, massive data samples will be generated during the offline training to cover the possible operating conditions and contingency scenarios. Thus, it is beneficial to randomize the system loading with a sufficiently large range, as well as to consider all possible contingency scenarios that are likely to occur in the upcoming operating horizon (next hour or day). Upon drawing samples of system

---

Algorithm 1 Decentralized Learning for Scalable OLS

Input: $\{C_j\}, \forall j = 1, \cdots, J; \{\boldsymbol{p}_k^d\}, \forall k = 1, \cdots, K$
Output: $\hat{p}_i^s, \forall i = 1, \cdots, N$

    Collect optimization results (offline)
1: **for** $j = 1$ to J **do**
2:   **for** $k = 1$ to K **do**
3:     Solve optimization (2), and collect training samples $\boldsymbol{x}_i = \left[p_i^d, q_i^d, V_i, \omega_i, \{p_{ij}\}, \{q_{ij}\}\right], \boldsymbol{y}_i = \alpha_i, \forall i.$
4:   **end for**
5: **end for**
    Training the ML model (offline)
6: **for** $i = 1$ to N **do**
7:   Training for each load center to obtain the local decision rule $f_i(\cdot; \varphi_i)$.
8: **end for**
    Real-time load shedding prediction (online)
9: Each load center i collects its real-time local measurements $\hat{\boldsymbol{x}}_i$.
10: Calculate the predicted dual variable $\hat{\alpha}_i = f_i(\hat{\boldsymbol{x}}_i; \varphi_i)$.
11: Obtain the optimal corrective actions $\hat{p}_i^s$ via (7).
12: **return** $\hat{p}_i^s, \forall i = 1, \cdots, N$

---

loading and contingency scenarios, the offline computation generates the corresponding input-output pairs using model-based simulations. Specifically, the input $\boldsymbol{x}_i$ can be computed by power system simulators such as MATPOWER through steady-state power flow analysis, except for the frequency input ω_i. The latter could be approximated instead, using the difference between the total generated powers during pre- and post-contingency; see e.g., [46, Ch. 12]. Meanwhile, the output y_i can be obtained using grid optimization solvers based on MATPOWER. Basically, the AC-OLS problem (2) can be solved to find the dual variables in y_i based on the loading and contingency information. Upon completing the training of bus i's OLS rule using the samples of $\{\boldsymbol{x}_i, y_i\}$, the updated NN model will be sent to the load center on a regular basis (from hourly to daily), enabling a fully decentralized online implementation. In a nutshell, the proposed scalable OLS design leverages the extensive computation capability at the control center by performing offline training of OLS decision rules in order to empower the online implementation of decentralized OLS corrective actions.

For the parameter φ_i per load center i, the feedforward NN model consists of multiple hidden layers between the input and output. With the first layer $\boldsymbol{z}_i^0 = \boldsymbol{x}_i$ incorporating the input feature, each layer k can be represented as:

$$\boldsymbol{z}_i^k = \sigma\left(\boldsymbol{W}^k \boldsymbol{z}_i^{k-1} + \boldsymbol{b}^k\right), \quad \forall k = 1, \ldots, K \qquad (10)$$

where the final layer $\boldsymbol{z}_i^K \to y_i$ predicts the output target. Basically, the NN parameter vector φ_i includes the weight matrices $\{\boldsymbol{W}^k\}$ and bias vectors $\{\boldsymbol{b}^k\}$ for every layer k. There is also a nonlinear activation function $\sigma(\cdot)$, used to achieve an expressive functional mapping that goes beyond linearity. Common choices of $\sigma(\cdot)$ include sigmoid and ReLU.

To determine φ_i, we use the mean squared error (MSE) metric as the loss function to minimize, which can be easily trained by popular NN solvers like PyTorch. The detailed implementation of the scalable OLS algorithm is summarized in Algorithm 1. The whole process starts with generating a prioritized list of contingencies $\{C\}$ and a sufficient number of load samples $\{\boldsymbol{p}^d\}$. Following this, the OLS problem is solved repeatedly to generate training samples. The training is conducted individually for each load center to obtain local decision rules. During real-time operations, each load center uses local measurements to predict its own dual variable $\hat{\alpha}_i$, then determines corrective actions through optimality condition (7). Note that both optimization solving and ML training are conducted offline, which enables accurate prediction of OLS decisions for each load center in real time.

Remark 3 (Safety of OLS learning). *Corrective actions such as load shedding need to be effective and safe during online implementation. Such considerations can be incorporated into the offline training phase of our proposed OLS learning design. For example, the loss function could have a larger weight coefficient for prediction errors that are above a certain threshold (e.g., [47]). Alternatively, one could directly incorporate the constraint violation into a regularized loss function as in learning-for-OPF work [23], [41], [42]. For both of these mismatches, it is also possible to consider its worst-case risk by using e.g., the conditional value-at-risk (CVaR) metric; see e.g., [48]. These modifications of loss functions would only affect the training phase, and thus still ensure a decentralized OLS in the online implementation.*

IV. IDENTIFIABILITY ANALYSIS

An important aspect of our decentralized OLS is to ensure that using locally available measurements only can differentiate two distinct contingencies. Otherwise, if the local measurements are identical, the ML algorithm may be unable to distinguish which contingency triggered the inputs, potentially causing ambiguity in local load-shedding decisions. Therefore, this capability to differentiate between contingency scenarios enables more accurate predictions, translating local data into well-informed local decisions. To this end, we investigate the conditions for attaining this identifiability using the simple DC power flow model [49]. This analysis will provide useful insights into the realistic scenarios in practical systems. The analysis primarily focuses on contingencies that cause topology changes in the power system. For contingencies such as short circuits and bus faults, their impact on system topology is temporary, as these events typically trigger protective devices to isolate faults automatically [50]. Even so, if any fault in the category leads to the isolation of lines for fault clearance, then the analysis is still applicable.

We start the analysis with the DC model which linearly relates the active power injection $\mathbf{p}$ to nodal angle $\boldsymbol{\theta}$ and line flow $\mathbf{f}$, as given by

$$\mathbf{p} = \mathbf{B}\boldsymbol{\theta}, \text{ and } \mathbf{f} = \mathbf{K}\boldsymbol{\theta} = \mathbf{S}\mathbf{p} \qquad (11)$$

where $\mathbf{B}$ is the B-bus matrix, and matrix $\mathbf{S}$ is the so-termed injection shift factor (ISF) matrix. It allows for a very fast analysis of the post-contingency power flow. This is because the
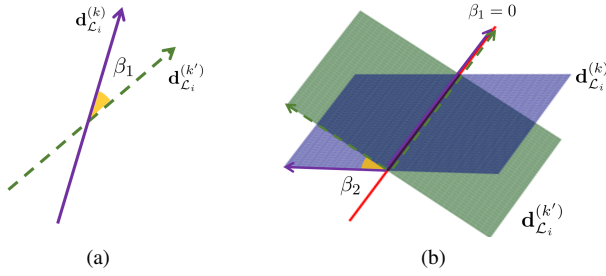
Fig. 3: The principal angles between the two subspaces spanned by $\mathbf{d}_{\mathcal{L}_i}^{(k)}$ and $\mathbf{d}_{\mathcal{L}_i}^{(k')}$, for the case of (a) 1-D subspaces with $\beta_1 > 0$; and (b) 2-D subspaces with $\beta_1 = 0$ and $\beta_2 > 0$.

B-bus matrix under any given contingency scenario indexed by k can be updated as $\mathbf{B}^{(k)}$, by eliminating the outaged lines in contingency k. Assuming the post-contingency system stays connected with no islanding, we can express the difference in line flows as

$$\Delta\mathbf{f} := \mathbf{S}^{(k)}\mathbf{p} - \mathbf{f} = \mathbf{K}([\mathbf{B}^{(k)}]^{-1} - \mathbf{B}^{-1})\mathbf{p}$$
$$= \mathbf{d}^{(k)}\mathbf{f}^{(k)} \qquad (12)$$

where the vector $\mathbf{f}^{(k)}$ contains the pre-contingency flows of the outaged lines under contingency k. The latter can be used to fully determine the line flow differences everywhere in the system by using the matrix $\mathbf{d}^{(k)}$. This result builds upon the special structure between $[\mathbf{B}^{(k)}]^{-1}$ and $(\mathbf{B}^{-1})$, based on the well-known matrix inverse lemma [51]. Taking the example of a single-line contingency k with the line (i,j) in outage, the vector $\mathbf{f}^{(k)}$ is a scalar with only the pre-outage flow for line (i,j). In this case, matrix $\mathbf{d}^{(k)}$ becomes a vector with the length as the total number of lines and is known as the line outage distribution factor (LODF), given by

$$\mathbf{d}^{(k)} := \frac{(\mathbf{S}^{(k)} - \mathbf{S})\mathbf{p}}{f_{ij}} = \frac{\mathbf{S}_{:i} - \mathbf{S}_{:j}}{1 - \mathbf{S}_{ki} + \mathbf{S}_{kj}} \qquad (13)$$

where the subscript of matrix $\mathbf{S}$ is used to index its entry with $:i$ indicating the i-th column. Generalizing (13) to a contingency of multiple-line outages is possible but would be more complicated; see e.g., [52], [53].

We can utilize the matrix $\mathbf{d}^{(k)}$ to determine whether there is sufficient identifiability of different contingencies. Note that only limited measurements are available to each load center, thus we consider a subset of lines denoted by $\mathcal{L}_i \subset \mathcal{L}$, as the local incident lines to bus i, for this analysis. This way, the change of power flow pattern in the local region of bus i is given by:

$$\Delta\mathbf{f}_{\mathcal{L}_i} = \mathbf{d}_{\mathcal{L}_i}^{(k)}\mathbf{f}^{(k)} = \left[\{\mathbf{d}_{\ell:}^{(k)}\}_{\ell \in \mathcal{L}_i}\right]\mathbf{f}^{(k)} \qquad (14)$$

by selecting only the lines $\ell \in \mathcal{L}_i$ to form the local flow vector $\Delta\mathbf{f}_{\mathcal{L}_i}$ and submatrix $\mathbf{d}_{\mathcal{L}_i}^{(k)}$ of sensitivity factors. Thus, the key for bus i to be able to differentiate two contingency scenarios, namely k and k', requires a sufficient level of dissimilarity in some sense, between the two corresponding matrices $\mathbf{d}_{\mathcal{L}_i}^{(k)}$ and $\mathbf{d}_{\mathcal{L}_i}^{(k')}$. We will formalize this condition.

First, consider the simple case of single-line outage scenarios, where the two matrices are essentially vectors of the same length. In this case, as long as the vector $\mathbf{d}_{\mathcal{L}_i}^{(k)}$ is not in parallel with vector $\mathbf{d}_{\mathcal{L}_i}^{(k')}$, their corresponding subspaces of local flow vectors will not overlap. As illustrated in Fig. 3(a), this is equivalent to requiring the acute angle (formally defined as the principal angle soon) between these two subspaces to be *non-zero*, which is satisfied by having

$$\frac{\left|\left[\mathbf{d}_{\mathcal{L}_i}^{(k)}\right]^{\mathsf{T}}\mathbf{d}_{\mathcal{L}_i}^{(k')}\right|}{\left\|\mathbf{d}_{\mathcal{L}_i}^{(k)}\right\| \cdot \left\|\mathbf{d}_{\mathcal{L}_i}^{(k')}\right\|} \neq 1. \qquad (15)$$

This condition guarantees that the change of local flows in bus i under the two contingencies k and k' will not be the same. As a result, the local flow measurements alone are sufficient for distinguishing the specific contingency, and accordingly, for determining the corresponding OLS corrective actions.

To consider multiple-line outage contingencies, we can generalize (15) to the matrix case. Given two matrices $\mathbf{d}_{\mathcal{L}_i}^{(k)}$ and $\mathbf{d}_{\mathcal{L}_i}^{(k')}$, the identifiability depends on the principal angles between the corresponding subspaces [54]. To this end, let us use the QR decomposition to obtain the orthonormal bases of the two subspaces, as given by

$$\mathbf{d}_{\mathcal{L}_i}^{(k)} = \mathbf{Q}^{(k)}\mathbf{R}^{(k)}, \text{ and } \mathbf{d}_{\mathcal{L}_i}^{(k')} = \mathbf{Q}^{(k')}\mathbf{R}^{(k')} \qquad (16)$$

where $\mathbf{Q}^{(k)}$ or $\mathbf{Q}^{(k')}$ is the orthornormal basis, while $\mathbf{R}^{(k)}$ or $\mathbf{R}^{(k')}$ is an upper triangular matrix. Clearly, the dimension of each subspace can be as large as the number of outaged lines in the contingency. We will use $|k|$ and $|k'|$ to denote the dimensions of these two subspaces, respectively.

We can further characterize the separation between the two subspaces via the singular value decomposition (SVD), as [55]

$$\left[\mathbf{Q}^{(k)}\right]^{\mathsf{T}}\mathbf{Q}^{(k')} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{T}} \qquad (17)$$

with all the singular values stored in the diagonal matrix $\boldsymbol{\Sigma}$. Here, the number of singular values $s = \min(|k|, |k'|)$, which is the minimum of the two subspaces' dimensions. The ordered singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_s \geq 0$ allow us to find the vector of principal angles in a *non-decreasing order*, as denoted by

$$\boldsymbol{\beta} := [\beta_1, \cdots, \beta_s]^{\mathsf{T}} = [\arccos(\sigma_1), \cdots, \arccos(\sigma_s)]^{\mathsf{T}}. \quad (18)$$

This vector $\boldsymbol{\beta}$ can fully characterize the angular separation between the two subspaces of interest. For the aforementioned single-line outages with $|k| = |k'| = 1$, the only singular value σ_1 exactly equals the ratio term given in (15). As shown in Fig. 3(a), the orthonormal bases of vectors $\mathbf{d}_{\mathcal{L}_i}^{(k)}$ and $\mathbf{d}_{\mathcal{L}_i}^{(k')}$ are essentially formed by normalizing each of them to have unit norm. Hence, if (15) holds, we have $\sigma_1 \neq 1$, and accordingly, the only principal angle $\beta_1 = \arccos(\sigma_1)$ is nonzero. Interestingly, extending this result to higher-dimensional subspaces requires having at least one non-zero principal angle in $\boldsymbol{\beta}$, or equivalently, at least one σ_ℓ in (18) to be different from 1. Fig. 3(b) shows the example of two 2-D subspaces, where the smaller angle $\beta_1 = 0$ but the larger one $\beta_2 \neq 0$ can ensure a sufficient level of subspace separation. This result directly follows from [55] and allows to formalize the identifiability for different contingencies, as stated below.
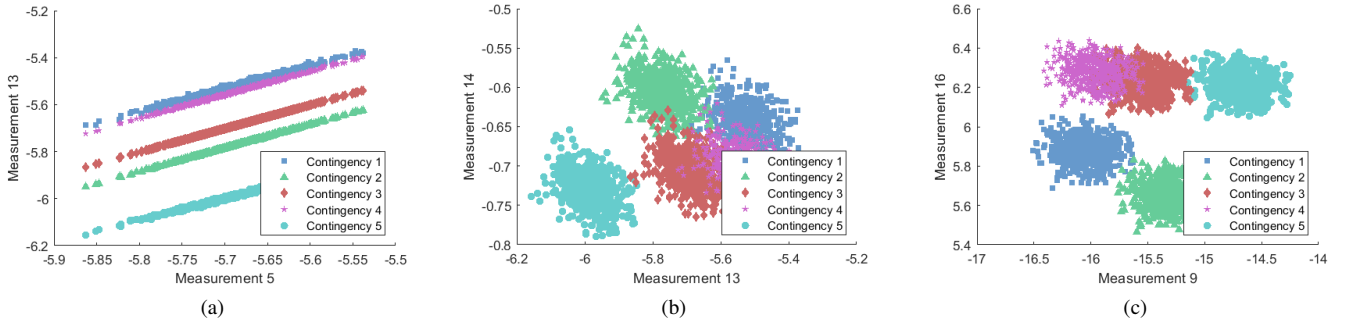
Fig. 4: The scatter plots of the local measurements at Bus 34 for the 5 contingencies considered for the IEEE 118-bus system based on a combination of only two measurements indexed by: (a) 5 and 13; (b) 13 and 14; and (c) 9 and 16.

Proposition 1 (Identifiability for two contingencies). *To guarantee that the local line flow data at bus i are sufficient for differentiating the two contingencies k and k', we need*

$$\exists \ell \in \{1, \ldots, s\}, \text{ such that } \sigma_\ell \neq 1 \qquad (19)$$

for at least one singular value found by (17). *This ensures a sufficient separation between the two subspaces spanned by $\mathbf{d}_{\mathcal{L}_i}^{(k)}$ and $\mathbf{d}_{\mathcal{L}_i}^{(k')}$ with at least one of the principal angles $\beta_\ell \neq 0$.*

Corollary 1 (Identifiability for a contingency set). *For a given contingency set $\mathcal{K}$, the local data can yield unique outputs for each scenario in $\mathcal{K}$ as long as every two contingency scenarios $k, k' \in \mathcal{K}$ can be differentiated per Proposition 1.*

Proposition 1 and Corollary 1 have provided the analytical conditions for using only local data to differentiate the possible contingencies. Specifically, (19) follows directly from the classical results on the angular separation between two linear subspaces in [54], [55], while Corollary 1 naturally extends it to the case of multiple subspaces. Note that Proposition 1 is also equivalent to

$$\min \{\sigma_1, \sigma_2, \cdots, \sigma_s\} < 1, \qquad (20)$$

which ensures the existence of at least one non-zero principal angle. This way, neither of the two subspaces can be entirely contained by the other one, and thus it is possible to distinguish them using a sufficient sampling of the two subspaces.

These analytical conditions can support the feasibility of our scalable OLS learning concept, by ensuring that local OLS decision rules can potentially infer the specific contingency and accordingly determine the best OLS actions. Of course, it can be sometimes challenging to satisfy these conditions. With only one measurement at bus i, it is almost impossible to differentiate any two contingencies. This is because in this case, both $\mathbf{d}_{\mathcal{L}_i}^{(k)}$ and $\mathbf{d}_{\mathcal{L}_i}^{(k')}$ in (15) would become scalars, and thus the ratio term therein has to be 1 which yields $\beta_1 = 0$. Having multiple measurements at bus i can, in most of the circumstances, lead to a non-zero β_1 for two vectors. The only circumstance having $\beta_1 = 0$ is when the two subspaces fully coincide with each other, which could be more unlikely when the number of measurements increases. In this sense, it would be "easier" to satisfy the condition (19) with multiple line-outages, because with an increasing dimension of subspaces, we only need one of the principal angles to be non-zero. In

general, the more diverse measurements bus i could access, the more likely the identifiability condition would hold to support our scalable OLS design. Note that the number of contingencies for a given networked system can be extremely high, especially when we consider multiple simultaneous component failures. However, in actual power system operations, ISOs typically restrict their analysis to N-1 and N-2 contingencies [56], [57], with N-3 considered only in rare cases. In addition, ISOs normally have a prioritized list of contingencies that they use to ensure reliable grid operations. These contingency scenarios are part of their contingency analysis and planning procedures [58], which help them maintain system stability and prepare for possible disruptions. Contingency prioritization [59] also takes into account events with specific risks, such as extreme weather events, and seasonal load fluctuations. Based on grid conditions, congestion, or maintenance plans, the ISOs will also adjust their priorities in real time. Without loss of generality, this work assumes that a subset of prioritized contingencies have been defined by system operators, which will be the primary focus of the learning process. Then again, since the learning process algorithm is mainly conducted offline using massive datasets, it can effectively cover most common system contingencies (e.g., N-1, N-2) to support real-time decision-making.

A. Numerical Example

We present one specific numerical example to corroborate that local data can be sufficient for distinguishing different contingency scenarios using the IEEE 118-bus system. This system consists of 186 lines and 19 conventional generators, with the remaining 99 buses as load centers. Specifically, we want to demonstrate that local data available at the load center of interest, bus 34, are sufficient for training a very accurate classifier that can distinguish all the scenarios in a contingency list. We have selected 5 contingency scenarios, including 3 single-line outage scenarios and 2 double-line outage ones. The outaged lines are chosen by prioritizing those with the highest nominal line flows, and thus the resultant contingencies tend to have a very large impact on the overall system safety and would require a high level of load shedding. To train this classifier, the input features include all the incident line flows (active/reactive powers) during both pre- and post-contingency operations provided by the power flow simulators. For each

contingency scenario, we randomly generate 1000 samples by adding some small perturbations to the steady-state loads. A simple feed-forward NN is trained using 70% of the total samples for this classification task. It is shown to exhibit an outstanding prediction performance when tested on the remaining 30% of samples, achieving a perfect accuracy of 100%. To illustrate this result, Fig. 4 plots the data samples for those 5 contingencies in a 2-D space, based on selected measurement pairs. We can observe that even with just two channels of local data, each of the contingencies manifests themselves with differentiable patterns of clusters or shapes in these 2-D distribution plots. These unique patterns have well supported our analysis that local data could provide sufficient identifiability for different contingency scenarios. Notably, this capability is truly important for our decentralized OLS design. It can ensure that even if trained with local data only, the resultant OLS decision rules are able to predict the specific contingency, possibly elsewhere in the system, and further effectively predict the globally optimal OLS actions. The ensuing section will present the comprehensive validation of the proposed scalable OLS design in terms of directly predicting OLS-related decisions.

V. SIMULATION RESULTS

This section presents the numerical results for our proposed AC-OLS learning approach using the IEEE 118-bus system and a larger 2000-bus system representing the Texas system [60]. The AC-OLS problem (2) is implemented in the MATPOWER simulator [61] by its primal-dual interior point method solver, while the post-contingency measurements are obtained by MATPOWER's power flow solver. The load shedding and reserve cost coefficients have been specified by modifying the quadratic cost functions in the original test cases to satisfy (4)-(5). Basically, the marginal cost of load shedding has been set to be higher than that of reserve flexibility, to ensure the latter is fully utilized before shedding the load. Note that the AC-OLS solver can also return the multiplier α_i for the active power balance equation (2g), which will be used as the output label for the NN training. The feedforward NN for our OLS design has been implemented using the MATLAB® deep learning toolbox with the Bayesian regularization. We have adopted a standard NN setup with 3 hidden layers. For the 2000-bus system, a 3-layer setup turns out to be adequate, even though each load center has dozens of input features (local measurement channels). All the simulations have been conducted on a regular laptop with Intel® CPU @ 2.60 GHz and 16 GB of RAM.

A. 118-bus System Tests

We have used the IEEE 118-bus system [62] for validating the scalable OLS design in addition to classifying the contingencies in Section IV-A. To solve the AC-OLS problem (2), we consider a total of 6 contingency scenarios. The majority of them are single-line or double-line outages, with one scenario having 3 outaged lines. We have chosen the outaged lines by prioritizing those with the highest nominal line flows. This is because the resultant contingencies would have a significant
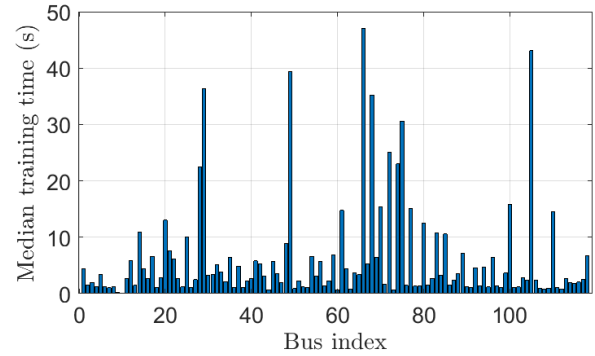

Fig. 5: The median value of training time (in seconds) for each bus in the 118-bus system.

impact on the overall safety of the system, necessitating a higher level of load shedding. For each contingency scenario, we generate a total of 2000 samples by randomly perturbing the pre-outage load profile. Specifically, the load demand at each bus has been changed randomly within the range of $[0.95, 1.05]$ times its nominal value. To train the OLS decision rules, the pre- and post-contingency local data serve as the input features at each load center, including local demand and incident line flows (real/reactive). The dimension of these local inputs ranges from 5 to 49, depending on the specific bus location. The output label is the corresponding multiplier α_i at bus i, as returned by the AC-OLS solver. A feedforward NN with 3 hidden layers (with 40, 30, and 20 neurons) has been trained using 70% of the total samples, with the remaining 30% reserved for testing. It turns out this standard NN setup can attain very high prediction performance later on.

To first demonstrate the computational efficiency, we have conducted multiple runs of NN training for every bus, and plot the median training time in Fig. 5. Clearly, the majority of buses enjoy extremely fast training time within a few seconds, with the maximum training time being less than 1 minute. This high computation speed is achieved thanks to the proposed scalable OLS design, under which the input features only depend on the local measurements available at each bus and thus are much less than the total number of measurements throughout the overall system.

We further demonstrate the effectiveness of our OLS design in predicting both the multiplier α_i and the OLS decision $\hat{p}_i^s$. Recall that the proposed OLS learning utilizes the KKT optimality condition in (6), where the value of α_i can fully determine $\hat{p}_i^s$. Fig. 6 compares the actual value with the predicted one using one representative instance of the testing sample. Note that while the α_i prediction can be performed for every bus, the prediction of load shedding amount $\hat{p}_i^s$ is only considered for the load bus only, and thus the x-axis of Fig. 6(b) only lists the 99 loads in the 118-bus system. Fig. 6(a) shows that the multipliers do have some strong locational dependence. While at most buses α_i is around 40$/MWh, some of the buses have very different values that can go as low as 0$/MWh. This is because the system congestion pattern could make certain buses much more preferred than the other for load shedding, and those preferable buses would
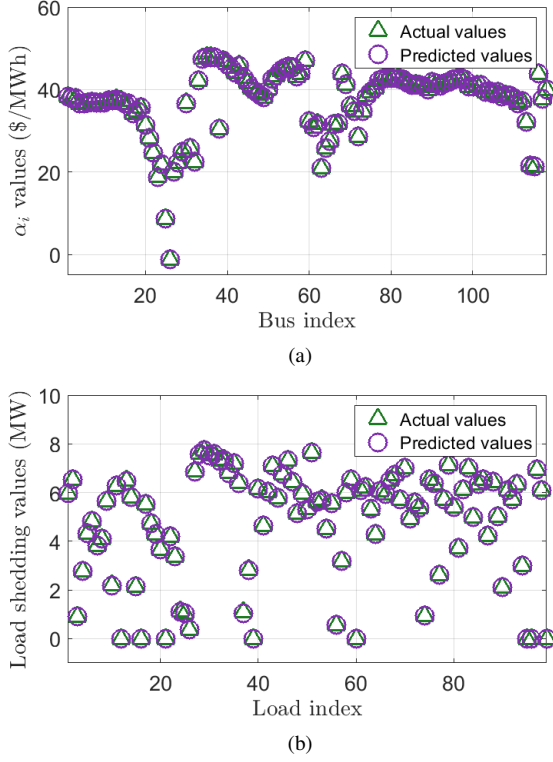
(a)

(b)

Fig. 6: Prediction on (a) Lagrange multipliers; and (b) load shedding; across different buses in the IEEE 118-bus system.
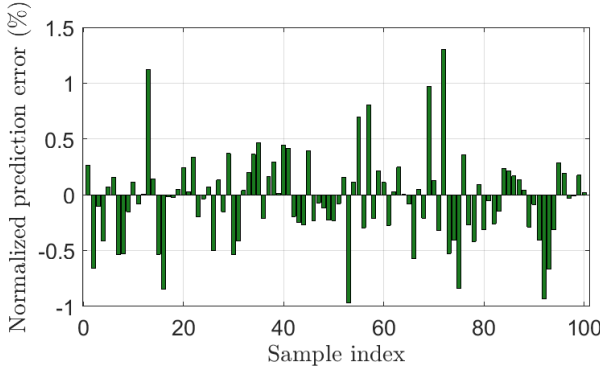

Fig. 7: Normalized prediction error of Lagrange multiplier at Bus 43 under randomly selected testing samples.

have relatively lower α_i values. We can observe that even with this high variability of α_i values, our OLS design can attain a very high prediction accuracy at all the buses. The average prediction error is 0.2716\$/MWh out of the average multiplier value of 36.8865\$/MWh, with an error percentage of 0.74%. Thanks to this excellent performance in predicting α_i, the OLS decisions p_i^s can be also well predicted as shown in Fig. 6(b). Similarly, there is also a high level of locational difference, as certain loads could more critically reduce the system-wide congestion during the post-contingency operations than the others. But still, the average prediction error for $\hat{p}_i^s$ is very low at 0.0621MW, with a maximum error of 0.48MW. Note that all these prediction results have been obtained from local data only available at each bus, and thus they have verified
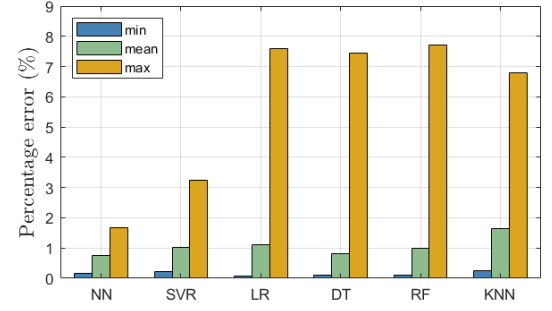

Fig. 8: Comparison of percentage errors over all buses in the 118-bus system under selected machine learning algorithms.

the effectiveness of our scalable OLS design.

Moreover, Fig. 7 shows the prediction performance across all samples by plotting the normalized prediction error attained at bus 43. This load center is a very representative one in the system which has consistently non-zero $\hat{p}_i^s$ across all samples. Since the α_i prediction is very critical for our overall design, we only plot the error percentage as normalized by the actual value for the task of predicting α_i only. Only 100 samples from the testing dataset have been selected here due to the size limit. But for all testing samples have very minimal prediction error at a similar level to those shown in Fig. 7. This prediction error tends to fall within the range of $[-1\%, \ 1\%]$, and its absolute value has an average of only 0.29%. This result again shows a very high accuracy in predicting α_i across all contingency scenarios and operating conditions.

Finally, we present a comprehensive analysis of the OLS decision prediction performance in Fig. 9, which plots the error statistics of predicting $\hat{p}_i^s$ for each load bus. Specifically, a box plot is used to mark the median, first, and third quartiles, as well as the minimum and maximum, of the absolute values of prediction errors across all testing samples. The median values range from almost 0MW to 0.28MW, with an average of 0.054MW across all load buses. For the first and third quartiles, they respectively have an average of 0.026MW and 0.092MW across all load buses. Based on an average load shedding amount of 4.81MW, this prediction has a very high accuracy for a majority of the load buses. Notably, a few of the load buses exhibit relatively higher errors than the average case, likely due to having a larger variability of the OLS decisions. Overall, the prediction accuracy for $\hat{p}_i^s$ has been pretty consistently maintained across all load buses, demonstrating the effectiveness of our proposed OLS learning on the 118-bus system. To showcase that the NN model provides an accurate prediction performance, we further include some comparison results from other regression models. Specifically, we have considered support vector regression (SVR), linear regression (LR), decision tree (DT), random forest (RF), and k-nearest neighbors (KNN), to compare with the benchmark NN model, in terms of prediction performance. The results are shown in Fig. 8. Across all buses, the NN model achieves an average prediction error of 0.74%, while other ML algorithms have slightly higher prediction errors which are 1.01%, 1.10%, 0.80%, 0.99%, 1.64%, respectively.
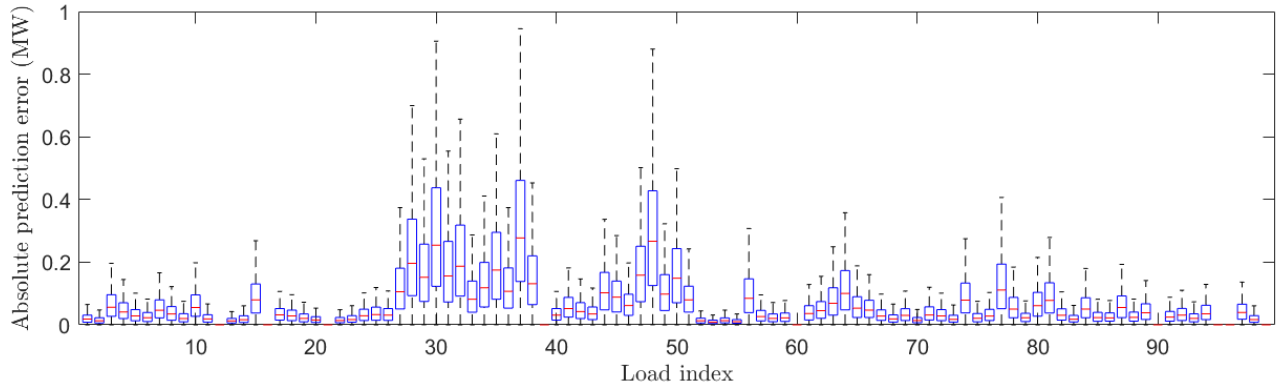
Fig. 9: Absolute value of prediction errors for p_i^s in box plot at every load bus in the 118-bus system.

In addition, the NN model maintains minimal variance in prediction errors compared to the other models, which makes it well-suited under the decentralized learning setup. Next, we will assess the performance of the scalable OLS algorithm using a more realistic 2000-bus test case.

B. 2000-bus System Tests

The ACTIVSg2000 test case [60] is a synthetic model, constructed using public data and statistical analyses of actual power systems. Developed to represent the geographical coverage of Texas, this system consists of 2000 buses and 3206 transmission lines. We once again consider a total of 6 contingency scenarios, with most of them being single-line or double-line outages, while one scenario includes 3 outaged lines. Similarly, the outaged lines are chosen by prioritizing those with the highest nominal line flows. To maintain the sample size, the samples are generated by focusing on the top 400 most heavily loaded buses by randomly perturbing the load demands therein within $[0.97, 1.03]$ of the nominal values. The training and testing of local OLS policies follow similarly from the 118-bus tests. The dimension of local input features at each bus increases to as high as 69, due to the presence of a more complicated network topology in this system. Thus, we use a feedforward NN with 3 hidden layers, with each of them having 50, 30, and 20 neurons, respectively. With the presence of thousands of load centers, we have selected a subset of them to showcase the test performance. Specifically, the load centers exhibiting the largest variability of the multiplier α_i have been selected, as they pose higher challenges in the OLS prediction as compared to load centers with pretty constant multipliers. The OLS training for the selected load centers takes 116.57s on average, with the fastest and slowest being 28.37s and 669.42s, respectively. This training time is slightly higher than that in the 118-bus system, mainly due to increased dimensionality. However, this increase of time is reasonable compared to the system size, corroborating the scalability of our proposed OLS design.

We first demonstrate the prediction performance across samples at bus 68, which features a comparably small set of 9 local inputs only. This particular load center is representative, as characterized by consistently non-zero, highly variable, load shedding amount. Similar to Fig. 7 for the 118-bus system,
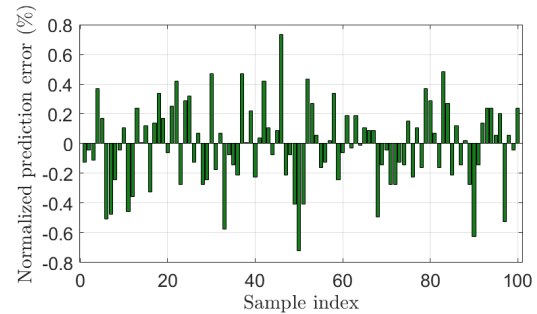

Fig. 10: Normalized prediction error of Lagrange multiplier at Bus 68 under randomly selected testing samples.
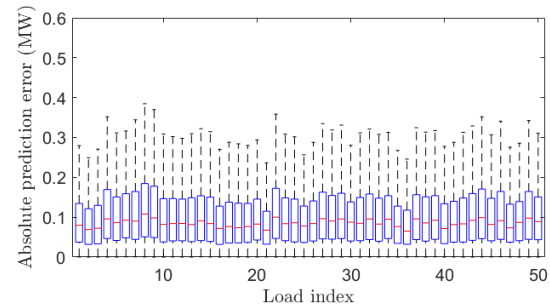

Fig. 11: Absolute value of prediction errors for p_i^s in box plot at selected load buses in the 2000-bus system.

Fig. 10 here plots the error percentage normalized by the actual value in predicting α_i, based on a total of 100 samples randomly chosen from the testing dataset. The prediction performance for the majority of the testing samples exhibits a similar error level to those in Fig. 10. We observe that the prediction errors lie within the range of $[-0.8\%, \ 0.8\%]$, with the absolute error percentages averaging at only 0.22%. This prediction error for the 2000-bus test case is slightly smaller than that in the 118-bus case. These results demonstrate exceptionally high accuracy in predicting the multiplier α_i and confirm that our proposed OLS design can well maintain the scalability with respect to system size.

Furthermore, we present a comprehensive validation of the prediction performance at multiple load buses. Specifically, we select the top 50 load buses with the highest variability

in their local multiplier. Accordingly, the variability of load shedding amounts at these selected buses are also among the highest due to the relation between these two variables in (6). Fig. 11 shows the box plot of OLS prediction error for $\hat{p}_i^s$, with the median, first and third quartiles, and the minimum and maximum of the absolute prediction errors. The median prediction errors range from 0.065MW to 0.108MW, with an average of 0.085MW across these load buses. In addition, the first and third quartiles have averages of 0.041MW and 0.148MW, respectively. Using the median error statistic, the OLS prediction across these 50 load centers has an extremely high accuracy of around 0.3% based on an average load shedding amount of 28.17MW. Overall, the consistency in prediction accuracy has been validated, demonstrating the effectiveness of our proposed OLS learning approach for the large 2000-bus system.

VI. CONCLUSION

In this paper, we developed a decentralized framework for performing real-time optimal load shedding to recover power balance and prevent cascading failures under emergency events. We start by formulating an AC power flow-based optimal load shedding (AC-OLS) problem. By solving AC-OLS repeatedly across diverse contingency scenarios, we propose a learning framework that maps from the local measurements of each load center to the corresponding Lagrange multiplier that exhibits locality property. The OLS decisions can be accurately determined by mapping from the predicted Lagrange multiplier using optimality conditions. The main contribution of this work lies in the scalable decentralized OLS design which enables each load center to react promptly to emergencies without reliance on central supervision. Numerical studies on the IEEE 118-bus and Texas 2000-bus systems have demonstrated the effectiveness of our proposed learning approach in accurately predicting the OLS solutions. Specifically, the offline training time is shown to scale well with the system size thanks to the decentralized design. Meanwhile, the prediction accuracy is very high on average (at a 0.22% of absolute error for the 2000-bus system). By increasing the number of training samples and selecting the candidate contingencies, the potential impact on prediction accuracy could be further mitigated. For future work, we plan to investigate topology-aware OLS learning approaches and work on improving the dependability of OLS predictions through robust risk-aware design.

REFERENCES

[1] Y. Omar, I. Abidin, S. Yusof, H. Hashim, and H. A. Rashid, "Under frequency load shedding (UFLS): Principles and implementation," in *Proc. IEEE Intl. Conf. on Power and Energy*, 2010, pp. 414–419.

[2] A. Ketabi and M. H. Fini, "An underfrequency load shedding scheme for hybrid and multiarea power systems," *IEEE Trans. Smart Grid*, vol. 6, no. 1, pp. 82–91, 2014.

[3] R. M. Larik, M. W. Mustafa, M. N. Aman, T. A. Jumani, S. Sajid, and M. K. Panjwani, "An improved algorithm for optimal load shedding in power systems," *Energies*, vol. 11, no. 7, p. 1808, 2018.

[4] R. M. Larik, M. W. Mustafa, and M. N. Aman, "A critical review of the state-of-art schemes for under voltage load shedding," *International Trans. Electrical Energy Systems*, vol. 29, no. 5, p. e2828, 2019.

[5] C. Mozina, "Undervoltage load shedding," in *2007 Power Systems Conference: Advanced Metering, Protection, Control, Communication, and Distributed Resources*. IEEE, 2007, pp. 39–54.

[6] T. Amraee, A. Ranjbar, B. Mozafari, and N. Sadati, "An enhanced under-voltage load-shedding scheme to provide voltage stability," *Electric power systems research*, vol. 77, no. 8, pp. 1038–1046, 2007.

[7] Automatic underfrequency load shedding. [Online]. Available: https://www.nerc.com/pa/Stand/Reliability%20Standards/PRC-006-2.pdf

[8] C. Coffrin, R. Bent, B. Tasseff, K. Sundar, and S. Backhaus, "Relaxations of AC maximal load delivery for severe contingency analysis," *IEEE Trans. Power Systems*, vol. 34, no. 2, pp. 1450–1458, 2018.

[9] N. Rhodes, L. Ntaimo, and L. Roald, "Balancing wildfire risk and power outages through optimized power shut-offs," *IEEE Trans. Power Systems*, 2020.

[10] F. Lin and C. Chen, "An ADMM algorithm for load shedding in electric power grids," in *American Control Conf.* IEEE, 2016, pp. 5002–5007.

[11] T. S. Fernandes, J. Lenzi, and M. A. Mikilita, "Load shedding strategies using optimal load flow with relaxation of restrictions," *IEEE Trans. Power Systems*, vol. 23, no. 2, pp. 712–718, 2008.

[12] A. Kody, A. West, and D. K. Molzahn, "Sharing the load: Considering fairness in de-energization scheduling to mitigate wildfire ignition risk using rolling optimization," in *Proc. IEEE Conf. on Decision and Control*, 2022, pp. 5705–5712.

[13] M. Chen and S. Low, "Machine learning for solving optimal power flow problems," 2023. [Online]. Available: https://www.mhchen.com/papers/OPF.DeepOPF.SGC.23.pdf

[14] S. Pahwa, C. Scoglio, S. Das, and N. Schulz, "Load-shedding strategies for preventing cascading failures in power grid," *Electric Power Components and Systems*, vol. 41, no. 9, pp. 879–895, 2013.

[15] M. Sinha, M. Panwar, R. Kadavil, T. Hussain, S. Suryanarayanan, and M. Papic, "Optimal load shedding for mitigation of cascading failures in power grids," in *Proceedings of the Tenth ACM International Conference on Future Energy Systems*, 2019, pp. 416–418.

[16] J. Song, E. Cotilla-Sanchez, G. Ghanavati, and P. D. Hines, "Dynamic modeling of cascading failure in power systems," *IEEE Trans. Power Systems*, vol. 31, no. 3, pp. 2085–2095, 2015.

[17] C. Kim, K. Kim, P. Balaprakash, and M. Anitescu, "Graph convolutional neural networks for optimal load shedding under line contingency," in *IEEE Power & Energy Society General Meeting*, 2019, pp. 1–5.

[18] T. L. Vu, S. Mukherjee, T. Yin, R. Huang, J. Tan, and Q. Huang, "Safe reinforcement learning for emergency load shedding of power systems," in *IEEE Power & Energy Society General Meeting*, 2021, pp. 1–5.

[19] Q. Huang, R. Huang, W. Hao, J. Tan, R. Fan, and Z. Huang, "Adaptive power system emergency control using deep reinforcement learning," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1171–1182, 2019.

[20] Y. Zhang, M. Yue, J. Wang, and S. Yoo, "Multi-agent graph-attention deep reinforcement learning for post-contingency grid emergency voltage control," *IEEE Trans. Neural Networks and Learning Systems*, 2024.

[21] Y. Zhou, J. Park, and H. Zhu, "Scalable learning for optimal load shedding under power grid emergency operations," in *IEEE Power & Energy Society General Meeting (PESGM)*, 2022, pp. 1–5.

[22] L. Jia, J. Kim, R. J. Thomas, and L. Tong, "Impact of data quality on real-time locational marginal price," *IEEE Trans. Power Systems*, vol. 29, no. 2, pp. 627–636, 2013.

[23] S. Liu, C. Wu, and H. Zhu, "Topology-aware graph neural networks for learning feasible and adaptive AC-OPF solutions," *IEEE Trans. Power Systems*, vol. 38, no. 6, pp. 5660–5670, 2022.

[24] A. S. Awad, T. H. El-Fouly, and M. M. Salama, "Optimal distributed generation allocation and load shedding for improving distribution system reliability," *Electric Power Components and Systems*, vol. 42, no. 6, pp. 576–584, 2014.

[25] M. Moradi and M. Abedini, "Optimal load shedding approach in distribution systems for improved voltage stability," in *IEEE International Power Engineering and Optimization Conference*, 2010, pp. 196–200.

[26] Z. Luo, H. Liu, N. Wang, T. Zhao, and J. Tian, "Optimal adaptive decentralized under-frequency load shedding for islanded smart distribution network considering wind power uncertainty," *Applied Energy*, vol. 365, p. 123162, 2024.

[27] W. Gu, W. Liu, J. Zhu, B. Zhao, Z. Wu, Z. Luo, and J. Yu, "Adaptive decentralized under-frequency load shedding for islanded smart distribution networks," *IEEE Trans. Sustainable Energy*, vol. 5, no. 3, 2014.

[28] Y. Xu, W. Liu, and J. Gong, "Stable multi-agent-based load shedding algorithm for power systems," *IEEE Trans. Power Systems*, vol. 26, no. 4, pp. 2006–2014, 2011.

[29] L. Zhu, C. Lu, and Y. Han, "Load shedding against short-term voltage instability using random subspace based SVM ensembles," in *IEEE Power & Energy Society General Meeting*, 2017, pp. 1–5.

This article has been accepted for publication in IEEE Transactions on Power Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2025.3528434

12

[30] H. Yang, W. Zhang, J. Chen, and L. Wang, "PMU-based voltage stability prediction using least square support vector machine with online learning," *Electric Power Systems Research*, vol. 160, pp. 234–242, 2018.

[31] E. Voumvoulakis, A. Gavoyiannis, and N. Hatziargyriou, "Decision trees for dynamic security assessment and load shedding scheme," in *IEEE Power Engineering Society General Meeting*, 2006.

[32] L. Zhu, C. Lu, and Y. Luo, "Response-based adaptive under-voltage load shedding using oblique energy decision tree," in *IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia)*, 2019, pp. 3657–3662.

[33] R. Huang, Y. Chen, T. Yin, Q. Huang, J. Tan, W. Yu, X. Li, A. Li, and Y. Du, "Learning and fast adaptation for grid emergency control via deep meta reinforcement learning," *IEEE Trans. Power Systems*, vol. 37, no. 6, pp. 4168–4178, 2022.

[34] D. Xu and A. A. Girgis, "Optimal load shedding strategy in power systems with distributed generation," in *IEEE Power Engineering Society Winter Meeting*, 2001, pp. 788–793.

[35] J. Machowski, Z. Lubosny, J. W. Bialek, and J. R. Bumby, *Power system dynamics: stability and control*.   John Wiley & Sons, 2020.

[36] H. Bevrani, *Robust power system frequency control*.   Springer, 2014, vol. 4, no. 2014.

[37] M. Garcia and R. Baldick, "Requirements for interdependent reserve types providing primary frequency control," *IEEE Trans. Power Systems*, vol. 37, no. 1, pp. 51–64, 2021.

[38] A. Castillo and R. P. O'Neill, "Survey of approaches to solving the ACOPF," *Federal Energy Regulatory Commission, Tech. Rep*, 2013.

[39] S. H. Low, "Convex relaxation of optimal power flow—Part I: Formulations and equivalence," *IEEE Trans. Control of Network Systems*, vol. 1, no. 1, pp. 15–27, 2014.

[40] B. Kocuk, S. S. Dey, and X. A. Sun, "Strong SOCP relaxations for the optimal power flow problem," *Operations Research*, vol. 64, no. 6, pp. 1177–1196, 2016.

[41] Y. Chen, L. Zhang, and B. Zhang, "Learning to solve DCOPF: A duality approach," *Electric Power Systems Research*, vol. 213, p. 108595, 2022.

[42] M. K. Singh, V. Kekatos, and G. B. Giannakis, "Learning to solve the AC-OPF using sensitivity-informed deep neural networks," *IEEE Trans. Power Systems*, vol. 37, no. 4, pp. 2833–2846, 2021.

[43] A. J. Conejo, E. Castillo, R. Mínguez, and F. Milano, "Locational marginal price sensitivities," *IEEE Trans. Power Systems*, vol. 20, no. 4, pp. 2026–2033, 2005.

[44] F. Li and R. Bo, "DCOPF-based LMP simulation: Algorithm, comparison with ACOPF, and sensitivity," *IEEE Trans. Power Systems*, vol. 22, no. 4, pp. 1475–1485, 2007.

[45] M. Garcia, H. Nagarajan, and R. Baldick, "Generalized convex hull pricing for the AC optimal power flow problem," *IEEE Trans. Control of Network Systems*, vol. 7, no. 3, pp. 1500–1510, 2020.

[46] J. D. Glover, M. S. Sarma, and T. Overbye, *Power system analysis & design*.   Cengage Learning, 2012.

[47] J. Mukhoti, V. Kulharia, A. Sanyal, S. Golodetz, P. Torr, and P. Dokania, "Calibrating deep neural networks using focal loss," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[48] S. Lin, S. Liu, and H. Zhu, "Risk-aware learning for scalable voltage optimization in distribution grids," *Electric Power Systems Research*, vol. 212, p. 108605, 2022.

[49] B. Stott, J. Jardim, and O. Alsac, "DC power flow revisited," *IEEE Trans. Power Systems*, vol. 24, no. 3, pp. 1290–1300, 2009.

[50] I. Kasikci, *Short circuits in power systems: A practical guide to IEC 60909-0*.   John Wiley & Sons, 2018.

[51] D. J. Tylavsky and G. R. Sohie, "Generalization of the matrix inversion lemma," *Proceedings of the IEEE*, vol. 74, no. 7, pp. 1050–1052, 1986.

[52] T. Guler, G. Gross, and M. Liu, "Generalized line outage distribution factors," *IEEE Trans. Power Systems*, vol. 22, no. 2, pp. 879–881, 2007.

[53] J. Guo, Y. Fu, Z. Li, and M. Shahidehpour, "Direct calculation of line outage distribution factors," *IEEE Trans. Power Systems*, vol. 24, no. 3, pp. 1633–1634, 2009.

[54] J. Miao and A. Ben-Israel, "On principal angles between subspaces in Rn," *Linear algebra and its applications*, vol. 171, pp. 81–98, 1992.

[55] G. H. Golub and H. Zha, *The canonical correlations of matrix pairs and their numerical computation*.   Springer, 1995.

[56] H. Chen and F. S. Bresler, "Risk management in PJM's market operation," in *2008 IEEE Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century*. IEEE, 2008, pp. 1–5.

[57] ISO New England Inc., System Planning, "Transmission planning technical guide," *Tech. Rep*, 2023.

[58] R. L.-Y. Chen, A. Cohn, N. Fan, and A. Pinar, "Contingency-risk informed power system design," *IEEE Trans. Power Systems*, vol. 29, no. 5, pp. 2087–2096, 2014.

[59] S. Fliscounakis, P. Panciatici, F. Capitanescu, and L. Wehenkel, "Contingency ranking with respect to overloads in very large power systems taking into account uncertainty, preventive, and corrective actions," *IEEE Trans. Power Systems*, vol. 28, no. 4, pp. 4909–4917, 2013.

[60] A. B. Birchfield, T. Xu, K. M. Gegner, K. S. Shetye, and T. J. Overbye, "Grid structural characteristics as validation criteria for synthetic networks," *IEEE Trans. Power Systems*, vol. 32, no. 4, 2016.

[61] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Trans. Power Systems*, vol. 26, no. 1, pp. 12–19, 2010.

[62] R. Christie. (1993) Power systems test case archive. [Online]. Available: https://labs.ece.uw.edu/pstca/pf118/pg_tca118bus.htm
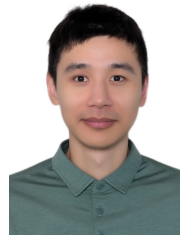
Yuqi Zhou (Member, IEEE) received the B.E. degree in information engineering from Xi'an Jiaotong University, Xi'an, China, in 2015. He received the M.S. degree in electrical engineering from Texas A&M University, College Station, TX, USA, in 2018, and the Ph.D. degree in electrical and computer engineering from the University of Texas at Austin, Austin, TX, USA, in 2022. From 2023 to 2024, he was a Postdoctoral Researcher with the National Renewable Energy Laboratory, Golden, CO, USA. He is currently a Researcher with Power Systems Engineering Center, National Renewable Energy Laboratory, Golden, CO, USA. His current research interests include optimization and machine learning, with applications in power system operation and control.

Hao Zhu (Senior Member, IEEE) is an Associate Professor of Electrical and Computer Engineering (ECE) at The University of Texas at Austin. She received the B.S. degree from Tsinghua University in 2006, and the M.Sc. and Ph.D. degrees from the University of Minnesota in 2009 and 2012, all in electrical engineering. From 2012 to 2017, she was a Postdoctoral Research Associate and then an Assistant Professor of ECE at the University of Illinois at Urbana-Champaign.

Her research focus is on developing algorithmic solutions for learning and optimization problems in future energy systems. Her current interests include machine learning for power system operations and resilience enhancements, and the design of energy management systems under cyber constraints. She is a recipient of the NSF CAREER Award and the faculty advisor for three Best Student Papers awarded at the North American Power Symposium. She is currently an Editor for the *IEEE Transactions on Smart Grid*, and *IEEE Transactions on Control of Network Systems*. .