A Bilevel Voltage Regulation Operation for Distribution Systems with Self-Operated Microgrids

Tianqi Hong, *Member, IEEE,* Dongbo Zhao*, Senior Member, IEEE,* Yichen Zhang*, Senior Member, IEEE*, Zhaoyu Wang, *Senior Member, IEEE*

***Abstract*— The emerging of microgrids in distribution systems has significantly enhanced the resilience of power grids. However, the operators of a distribution system and microgrids therein can be different and have accessibility to different devices. To model the operation of such a grid, this paper proposes a bilevel formulation and probes into the voltage regulation operation, considering the interaction between different systems. The proposed bilevel formulation considers the cooperation of active energy resources (AER), transformer tap-changers, and capacitor banks that are controlled by different operators. To facilitate the solution time of the target bilevel optimization, the lower-level problems with different objectives are modeled using deep neural networks (DNNs) which are then converted into a set of constraints. Hence, the bilevel problem can be reformed to a single-level problem. Lastly, the proposed solution procedures are validated using a customized joint system constructed by the IEEE 123-bus system and a real distribution system in Iowa. According to the numerical validation results, the solution time of the proposed nonlinear activation function based DNN model is 69 times faster than other methods in solving voltage regulation with a bilevel structure.**

***Index Terms*—Bilevel optimization, deep neural network, multiphase distribution system, voltage control**

I. Introduction

Pursuing more flexible distribution systems is always one objective of utilities and researchers. Nowadays, the flexibility of distribution systems mainly comes from controllable devices such as voltage regulators, controllable capacitor banks, and active energy resources (AERs). Besides, the management model of distribution systems starts to step from a centralized one to hierarchical or decentralized structures. Such evolution in the management model inevitably breaks the availability of global information and the controllability of active devices. To accommodate such paradigm shifts, various management problems have been formulated as bilevel or even higher-level optimization problems [1]-[3].

This work is supported by the U.S. Department of Energy Office of Electricity – Advanced Grid Modeling Program.

T. Hong, D. Zhao, Y. Zhang are with Energy Systems Division at Argonne National Laboratory, Lemont, IL 60439 USA (e-mails: yichen.zhang@anl.gov).

Z. Wang is with Electrical and Computer Engineering Department of Iowa State University, Ames, IA 50011.

Among all possible operational schemes at the distribution level, voltage regulation problems can be considered as the most critical one for distribution system operators since voltage deviations are directly related to the power qualities [4]-[5] and end-user experience. Various progress has been made during the past decade related to voltage regulation problems [6]-[8]. In [5]-[8], full load models, power electronics links, and system uncertainties have been considered and integrated into voltage regulation problems and solved effectively. For multi-zone systems or the system with self-operated microgrids, the management objectives and the targets of each entity can be different. Voltage regulation problems for such hierarchical structures will be naturally modeled as bilevel/multilevel optimization problems in which suboptimization problems are inserted in an optimization formulation. However, this type of voltage regulation (or volt-var control) problem is seldomly studied in the literature due to its complexity.

The multilevel optimization problem is naturally an extremely complicated problem. Computing a multilevel problem as well as validating the solution optimality of such problem are NP-hard tasks [9]. Although bilevel problems are the simplest problem among the multilevel optimization family, solving bilevel problems is still time-consuming when considering the nonlinearity and discreteness of transformer taps and capacitor banks.

A bilevel formulation of a voltage regulation problem is proposed in [4] to tackle bilevel voltage regulation problems at the distribution level. The objective functions of both levels regulate the voltage profile of the entire system. Besides, the interactions between upper and lower-level problems have been decoupled.

Apart from ignoring the coupling between each level, there are two tracks to handle bilevel optimization problems, and heuristic searching methods are favorable for solving highly nonlinear tasks. Another way is to reformulate a multilevel problem into a single-level problem, and the common analytic reformulation strategies are the Karush–Kuhn–Tucker (KKT) condition and primal-dual approaches [10]-[11]. However, the major drawback of the KKT and primal-dual modeling approaches is that they preserve the nonlinearity of the original lower-level problem, and the optimality of the lower-level problem cannot be guaranteed when problems are nonlinear

and nonconvex. Besides, the KKT complementary condition may introduce additional integer variables [12], which complicates the original formulation.

Different from analytic and heuristic searching approaches, this paper aims at applying data-driven approaches for modeling the lower-level problems and converting bilevel problems to single-level problems. The data-driven based modeling methods are widely applied to power system applications [13]-[15]. Among all existing data-driven approaches, deep neural network (DNN) is the most effective way to model systems with high nonlinearity.

To explore the possibility of deploying DNN in solving bilevel voltage regulation problems, this paper performs a comprehensive study on DNN structures considering different types of activation functions. Regarding the training data generation, effective relaxation methods are proposed, which can significantly reduce the size and complexity of DNNs. The proposed method is validated using a joint distribution system between the IEEE 123-bus system and an actual Iowa distribution network. Finally, the proposed DNN based approaches are compared with a heuristic searching method to highlight their effectiveness. The comparison between centralized operation and bilevel operation in voltage regulation is also discussed in the numerical case study section.

The contributions of this paper can be concluded as (1) the DNN based data-driven approach is first introduced to solve the bilevel voltage regulation problem at the distribution level; (2) both mixed integer model and smooth nonlinear model of DNN are studied and compared for modeling lower-level problems; (3) the proposed methods are compared with a heuristic searching method in a distribution system with self-operated microgrids to show the effectiveness of the data-driven modeling approaches.

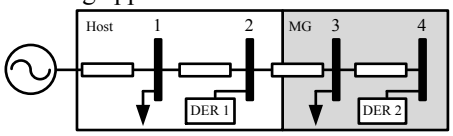


Fig. 1. Example system with a hierarchical structure.

II. Voltage Regulation Problem Formulation

Voltage regulation problems can be generally classified as either global regulation problems or local regulation problems. The global regulation problem is equivalent to the centralized regulation problem, in which the system operator controls all controllable devices to regulate the voltages of the entire system. For example, the operator in Fig. 1 can control both DERs in the host system and the microgrid (MG) to regulate voltages at all four buses. In a local regulation problem (decentralized), different operators only manage their systems. For example, the operator in the host system controls DER 1 to regulate voltages at buses 1 and 2. Meanwhile, the microgrid operator uses DER 2 to regulate the voltage at buses 3 and 4. In a system operated under local regulation, if the host system wishes to consider the operations of the subsystems (the microgrid in this case), a bilevel problem will be formulated. The operational performances under different control or management structures are further demonstrated in Section V.

Global voltage regulation problems can be solved or realized using centralized or distributed algorithms [16], [17]. Ideally, it is also possible to solve or achieve local regulation using distributed algorithms. However, the existing distributed algorithms still cannot directly be applied to solve the voltage regulation problem with bilevel operation structures [17]-[18].

The main focus of this paper is to address the voltage regulation problem for the distribution systems with bilevel operation scenarios from the host system perspective. The target control frequencies are daily or hourly. A generalized system structure of a distribution system with a bilevel operation structure is shown in Fig. 2, where different operators only manage their own systems. The host system can be a utility-owned distribution network with most customers. The lower-level systems can be microgrids, smart communities, university campuses.

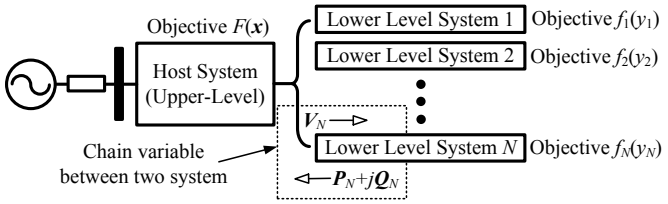


Fig. 2. One-line diagram of bilevel operated system.

Considering the system shown in Fig. 2, this paper investigates the scenario that the operator cannot manage the controllable devices in other levels. For example, the operator in the host system cannot dispatch the AERs in lower-level systems. Under this condition, the terminal voltage and the power exchanges become the only connection between the host system and subsystems during the operation. A practical example system can be found in [19], in which the microgrid of Illinois Institute of Technology (IIT) is attached to a distribution system feeder, and the microgrid can control its AERs.

A. Bilevel Optimization Structure

The mathematical model of the problem presented in Fig. 2 can be formulated as a bilevel optimization problem as

$$\min_{\boldsymbol{x}} F(\boldsymbol{x}, \boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_N) \tag{1a}$$

s.t.

$$\boldsymbol{G}(\boldsymbol{x}, \boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_N) \le \boldsymbol{0} \tag{1b}$$

$$\boldsymbol{\alpha}_i = O_i(\boldsymbol{x}, \boldsymbol{y}_1^{\blacksquare}, \dots, \boldsymbol{y}_N^{\blacksquare}) \quad \forall i \in \mathcal{N} \tag{1c}$$

$$\boldsymbol{y}_i^{\blacksquare} = \operatorname*{argmin}_{\boldsymbol{y}_i} f_i(\boldsymbol{\beta}_i, \boldsymbol{y}_i) \quad \forall i \in \mathcal{N} \tag{1d}$$

$$\boldsymbol{g}_i(\boldsymbol{\beta}_i, \boldsymbol{y}_i) \le \boldsymbol{0} \quad \forall i \in \mathcal{N} \tag{1e}$$

$$\boldsymbol{\beta}_i = o_i(\boldsymbol{x}, \boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_N) \quad \forall i \in \mathcal{N} \tag{1f}$$

where $\boldsymbol{x}$, $\boldsymbol{y}_i$, and $\boldsymbol{y}_i^{\blacksquare}$ are vectors; i is the index of the subsystems (or lower-level systems), $\mathcal{N}$ is the set that contains all index i, and the size of $\mathcal{N}$ is N; the upper-level problem is defined by (1a)-(1b) and the lower-level problem is defined by (1d)-(1e); $F(\cdot)$ and $\boldsymbol{G}(\cdot)$ are the cost and constraints of the upper-level problem, $f_i(\cdot)$ and $\boldsymbol{g}_i(\cdot)$ are the cost and constraints of the ith lower-level problem; $o_i(\cdot)$ and $O_i(\cdot)$ are the boundary constraints that connect the upper and lower systems; $\boldsymbol{y}_i^{\blacksquare}$ is the solution to the ith lower-level problem.

In the context of the voltage regulation problem, the deci-

sion variable $\boldsymbol{x}$ is the host system owned transformer tap positions, capacitor banks, and AERs. The decision variable $\boldsymbol{y}_i$ is the subsystem owned transformer tap positions, capacitor banks, and AERs. The boundary variable $\boldsymbol{\alpha}_i$ is the power injection from the host system to subsystem i and $\boldsymbol{\beta}_i$ is the terminal voltage of the subsystem i; see Fig. 2. According to (1c) and (1f), $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$ couple the upper and lower-level problems.

B. Objectives of Bilevel Problem

In this paper, the objective function of the host system is fixed to minimize the voltage deviation through transformer tap positions $\boldsymbol{t}_T$, controllable capacitor banks $\boldsymbol{t}_C$, and dispatchable power $\boldsymbol{s}_A$ from the AERs. Mathematically, we have

$$F(\boldsymbol{x}, \{\boldsymbol{\alpha}_i\}_i^N) = \sum_{\tau} w_\tau \left(U_\tau - \breve{U}_\tau\right)^2 \tag{2}$$

where $\boldsymbol{x} = [\boldsymbol{t}_T, \boldsymbol{t}_C, \boldsymbol{s}_A]$; $\{\boldsymbol{\alpha}_i\}_i^N = [\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_N]$ represents the power injections from all subsystems; w_τ is the weight assigned to each node and controlled by the operator; U_τ and $\breve{U}_\tau$ are the voltage amplitude and voltage reference of node τ in the host distribution system.

Remark that this paper aims at finding an optimal voltage regulation solution for host systems (or utility). Hence, the optimal value of the decision variable $\boldsymbol{x}$ is the target, and the objective function of the host system does not consider the voltages inside each subsystem.

For the lower-level problem, this paper models two possible operation objectives of subsystems. The first one is voltage regulation, and the second one is line loss minimization. We can model the operation objectives as

$$f_i^A(\boldsymbol{U}_{t,i}, \boldsymbol{y}_i) = \sum_j \breve{w}_{j,i} \left(\mathcal{U}_{j,i} - \breve{\mathcal{U}}_{j,i}\right)^2 \tag{3a}$$

$$f_i^B(\boldsymbol{U}_{t,i}, \boldsymbol{y}_i) = \boldsymbol{e}_i^T \boldsymbol{g}_i \boldsymbol{e}_i + \boldsymbol{f}_i^T \boldsymbol{g}_i \boldsymbol{f}_i \tag{3b}$$

where $\boldsymbol{y}_i = [\boldsymbol{t}_{T,i}, \boldsymbol{t}_{C,i}, \boldsymbol{s}_{A,i}]$, and $\boldsymbol{t}_{T,i}$, $\boldsymbol{t}_{C,i}$, $\boldsymbol{s}_{A,i}$ are the tap changers, capacitor banks, and the AERs in the subsystem i; $\breve{w}_{j,i}$ is the weight assigned to node j in subsystem i; $\mathcal{U}_{j,i}$ and $\breve{\mathcal{U}}_{j,i}$ are the voltage amplitude and voltage reference of node j in subsystem i; $\boldsymbol{U}_{t,i}$ is the voltage amplitudes of the bus that connects the host system and the subsystem i; $\boldsymbol{e}_i$ and $\boldsymbol{f}_i$ are the real and imaginary vectors of the node voltage; $\boldsymbol{g}_i$ is the real part of the admittance matrix. The dimension of $\boldsymbol{U}_{t,i}$ indicates single-, two-, and three-phase cases, and the voltage phase angle unbalances of the interconnecting bus are neglected in this paper. However, the rest of the loads and systems is still unbalanced. Note that each subsystem can operate with only one objective at a timeslot.

C. Constraints

Power flow constraints are physically identical for both host and subsystems. Hence, this subsection does not differentiate the variables in the power flow constraints of the systems at different levels, and the variable notation inherits from the host system. In this paper, the power flow model in rectangular coordinates is applied to describe the nodal power balance constraints, and we can write

$$\boldsymbol{V}_L = -\boldsymbol{Y}_{LL}^{-1}\boldsymbol{Y}_{LS}\boldsymbol{V}_S - \boldsymbol{Y}_{LL}^{-1}\mathrm{diag}^{-1}(\boldsymbol{V}_L^*)\boldsymbol{S}_L^* \tag{4}$$

where index $L \in \mathcal{L}$ and $S \in \mathcal{S}$ in which $\mathcal{L}$ stands for the set of load bus indexes and $\mathcal{S}$ stands for the set of substation bus indexes; $\boldsymbol{V}_L$ and $\boldsymbol{V}_S$ are the voltage vectors of load and substation buses (interconnecting buses of subsystems), respectively; $\boldsymbol{S}_L$ is the apparent power vector representing ZIP loads and dispatchable power injections; diag(·) is the operator to expand a vector to a diagonal matrix; $\boldsymbol{Y}_{LL}$ and $\boldsymbol{Y}_{LS}$ are submatrices of the system admittance matrix $\boldsymbol{Y}$. Note that $\boldsymbol{Y}_{LL}$ and $\boldsymbol{Y}_{LS}$ are functions of $\boldsymbol{t}_T$, $\boldsymbol{t}_C$.

The node current I_k at node k can be expressed as

$$I_k = \left(\frac{S_k}{V_k}\right)^* = \frac{K_{P,k}^*}{V_k^*} + \frac{s_{A,k}^*}{V_k^*} + K_{I,k}^* \frac{V_k}{U_k} + K_{Z,k}^* V_k \tag{5}$$

where $K_{P,k}$, $K_{I,k}$, and $K_{Z,k}$ are ZIP parameters for the loads at node k, $s_{A,k}$ is the complex dispatchable power; V_k is the complex node voltage at node k. The nonlinear relationship in (5) can be linearized using regional regression [20], and we have

$$I_k = \frac{S_k^*}{V_k^*} \approx K_{e,k} e_k + K_{f,k} f_k + K_{p,k} p_{A,k} + K_{q,k} q_{A,k} + C_{I,k} \tag{6}$$

where $K_{e,k}$, $K_{f,k}$, $K_{p,k}$, $K_{q,k}$, and $C_{I,k}$are linearized parameters; e_k and f_k are real and imaginary parts of V_k; $p_{A,k}$ and $q_{A,k}$ are dispatchable active and reactive powers at node k.

Apart from the nodal power balance constraints, the branches with tap changeable transformers can be modeled using controllable sources as shown in Fig. 3, and we have

$$\boldsymbol{I}_k = \boldsymbol{I}_{in} - \boldsymbol{A}_i \boldsymbol{I}_{km} = \boldsymbol{I}_{in} - \boldsymbol{A}_i \boldsymbol{Y}_{km} (\boldsymbol{A}_v \boldsymbol{V}_k - \boldsymbol{V}_m) \tag{7a}$$

$$\boldsymbol{I}_m = \boldsymbol{I}_{km} - \boldsymbol{I}_{out} = \boldsymbol{Y}_{km} (\boldsymbol{A}_v \boldsymbol{V}_k - \boldsymbol{V}_m) - \boldsymbol{I}_{out} \tag{7b}$$

where $\boldsymbol{I}_k$ and $\boldsymbol{I}_m$ are the multiphase node currents of nodes k and m; $\boldsymbol{I}_{in}$ and $\boldsymbol{I}_{out}$ are the input and output currents of the transformer and they represent the rest of the network; $\boldsymbol{I}_{km}$ is the transformer current w.r.t node m; $\boldsymbol{A}_v$ and $\boldsymbol{A}_i$ stand for the amplify matrices on node voltage and current; $\boldsymbol{Y}_{km}$ is the line impedance of the transformer. The relationship in (7) can be re-organized to a matrix form, and we have

$$\begin{bmatrix} \boldsymbol{I}_k \\ \boldsymbol{I}_m \end{bmatrix} = \begin{bmatrix} -\boldsymbol{A}_i \boldsymbol{Y}_{km} \boldsymbol{A}_v & \boldsymbol{A}_i \boldsymbol{Y}_{km} \\ \boldsymbol{Y}_{km} \boldsymbol{A}_v & -\boldsymbol{Y}_{km} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_k \\ \boldsymbol{V}_m \end{bmatrix} + \begin{bmatrix} \boldsymbol{I}_{in} \\ -\boldsymbol{I}_{out} \end{bmatrix} \tag{8}$$

where $\boldsymbol{A}_v = \boldsymbol{A}_i^T$ for transformer in wye, closed-delta, and open-delta connections.

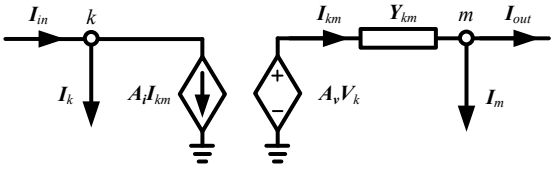


Fig. 3. One-line diagram of a multiphase transformer.

According to [21], the effective regulator ratio a_R for single-phase tap changeable transformer is defined as

$$a_R = 1 \mp 0.00625 t_{T,k} \tag{9}$$

where $t_{T,k}$ is an integer and an example in $\boldsymbol{t}_T$ which interlinks the tap positions to nodal power balance constraints (4). Note that the detailed relationship between a_R and $\boldsymbol{A}_i^T$ can be founded in [21] based on the connectivity types of the transformer. For a real transformer, $\boldsymbol{t}_T$ will always be bounded as

$$\underline{\boldsymbol{t}}_T \le \boldsymbol{t}_T \le \overline{\boldsymbol{t}}_T \tag{10}$$

where $\underline{\boldsymbol{t}}_T$ and $\overline{\boldsymbol{t}}_T$ represent the lower and upper bounds of $\boldsymbol{t}_T$.

Similar to a single-phase transformer, the susceptance of a

switched capacitor bank at node k can be modeled as

$$B_{C,k} = t_{C,k} b_{C,k} \tag{11}$$

where $B_{C,k}$ is the total susceptance offered by the capacitor bank; $t_{C,k}$ is an example in $\boldsymbol{t}_C$; $b_{C,k}$ is the incremental susceptance of the capacitor bank. The capacitor bank tap $\boldsymbol{t}_C$ is also constrained by

$$\underline{\boldsymbol{t}}_C \le \boldsymbol{t}_C \le \overline{\boldsymbol{t}}_C \tag{12}$$

where $\underline{\boldsymbol{t}}_C$ and $\overline{\boldsymbol{t}}_C$ are the lower and upper bounds of $\boldsymbol{t}_C$.

Additional to power flow constraints, the constraints on voltage amplitude also exist in lower- and upper-level problems, and we have

$$\underline{\boldsymbol{U}} \le \boldsymbol{U} \le \overline{\boldsymbol{U}} \tag{13}$$

where $\boldsymbol{U}$ is the vector of all node voltage amplitudes; $\overline{\boldsymbol{U}}$ and $\underline{\boldsymbol{U}}$ are the upper and lower bounds of the voltage amplitudes. Remark that it is essential to include voltage amplitude constraints in the lower-level problems since the upper-level decision may cause infeasibilities of lower-level problems.

In all systems, the dispatchable active power and reactive power of AERs are bounded by box constraints, yielding

$$\underline{\boldsymbol{p}}_A \le \boldsymbol{p}_A \le \overline{\boldsymbol{p}}_A \tag{14a}$$

$$\underline{\boldsymbol{q}}_A \le \boldsymbol{q}_A \le \overline{\boldsymbol{q}}_A \tag{14b}$$

where $\boldsymbol{s}_A = \boldsymbol{p}_A + j\boldsymbol{q}_A$.

According to the problem formulation, the voltage regulation problem considered in this paper is a bilevel mixed-integer nonlinear problem, and it is an NP-hard problem. In the next section, a hybrid method is proposed to solve the proposed voltage regulation problems.

III. Lower-Level Problem Modeling Through Data-Driven Approach

Compared with other neural network based methods, DNN is one of the most popular, fundamental, and flexible methods to model smooth nonlinear systems [22]. Hence, this paper proposes to use DNNs with different activation functions to model the behaviors of each lower-level problem with some relaxation.

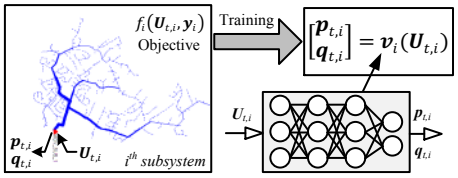


Fig. 4. Lower-level system modeling.

A. Deep Neural Network Representation

Apart from direct computation using heuristic search approaches, one intuitive way of solving bilevel problems is to convert the bilevel formulation into a single-level problem. A successful single-level problem conversion relies on the accuracy and complexity of the lower-level problem. Apart from traditional KKT conditions or primal-dual approaches, a promising method is to use a deep neural network to model the behavior of the lower-level system; see Fig. 4.

With a constant terminal voltage vector $\boldsymbol{U}_{t,i}$, the ith lower-level problem can be solved using existing solvers effectively. Consider the lower-level problem as a system with input $\boldsymbol{U}_{t,i}$ and output $\boldsymbol{\alpha}_i$, we can define a function to represent such relationships between input and output, and we write as

$$\boldsymbol{\alpha}_i = \begin{bmatrix}\boldsymbol{p}_{t,i}\\ \boldsymbol{q}_{t,i}\end{bmatrix} = \boldsymbol{v}_i(\boldsymbol{U}_{t,i}) \tag{15}$$

where $\boldsymbol{v}_i(\cdot)$ represents the relationship described by (1c)-(1e), $\boldsymbol{p}_{t,i}$ and $\boldsymbol{q}_{t,i}$ are the active and reactive power vectors at the terminal buses.

In this paper, the function $v_i(\cdot)$ is approximated using a set of parametrized functions. The input training data $\boldsymbol{\mathcal{X}}$ of the DNN is a collection of the random sampled $\boldsymbol{U}_{t,i}$, where the lth sample $\boldsymbol{x}_l = \boldsymbol{U}_{t,i,l}$ and $\boldsymbol{\mathcal{X}} = \left[\boldsymbol{x}_1^T, \boldsymbol{x}_2^T, \dots \boldsymbol{x}_l^T, \dots, \boldsymbol{x}_{N_s}^T\right]^T$. The labeled output data $\boldsymbol{\mathcal{Y}}$ consists of the power injection $\boldsymbol{\alpha}_{i,l}$ from the ith subsystem under given $\boldsymbol{x}_i$, we can write

$$\boldsymbol{\mathcal{Y}} = \left[\boldsymbol{\alpha}_{i,1}^T, \boldsymbol{\alpha}_{i,2}^T, \dots, \boldsymbol{\alpha}_{i,l}^T, \dots, \boldsymbol{\alpha}_{i,N_s}^T\right]^T \tag{16}$$

where N_s is the number of training data pairs. Now we can construct the DNN considering a fully connected neural network with N_L hidden layers. Each layer follows a specific activation function denoted as $\sigma(\cdot)$, and the network ends with an output layer that uses a linear activation function. The approximated power injection $\boldsymbol{\alpha}_i$ can be expressed as

$$\boldsymbol{z}_1 = \boldsymbol{x}_l \boldsymbol{W}_1 + \boldsymbol{b}_1 \tag{17}$$

$$\hat{\boldsymbol{z}}_h = \boldsymbol{z}_{h-1} \boldsymbol{W}_h + \boldsymbol{b}_h \tag{18}$$

$$\boldsymbol{z}_h = \sigma(\hat{\boldsymbol{z}}_h) \tag{19}$$

$$\hat{\boldsymbol{\alpha}}_{i,l} = \boldsymbol{z}_{N_D} \boldsymbol{W}_{N_D+1} + \boldsymbol{b}_{N_D+1} \tag{20}$$

where $\hat{\boldsymbol{\alpha}}_{i,l}$ is the approximated power injection w.r.t input $\boldsymbol{x}_l$; $\boldsymbol{W}_h$ and $\boldsymbol{b}_h$ for h = 1, …, N_D are the weight and bias of all hidden layers; $\boldsymbol{W}_{N_D+1}$ and $\boldsymbol{b}_{N_D+1}$ are the weight and bias of the output layer. All the weights and biases are selected to minimize the squared $L2$-norm of the difference between labeled power injections and the approximated injections, yielding

$$[\boldsymbol{W}_h, \boldsymbol{b}_h] = \underset{\boldsymbol{W}_h, \boldsymbol{b}_h}{\operatorname{argmin}} \frac{1}{N_s} \left\| \hat{\boldsymbol{\mathcal{Y}}} - \boldsymbol{\mathcal{Y}} \right\|_2^2 \tag{21}$$

where $\hat{\boldsymbol{\mathcal{Y}}} = \left[\hat{\boldsymbol{\alpha}}_{i,1}^T, \hat{\boldsymbol{\alpha}}_{i,2}^T, \dots, \hat{\boldsymbol{\alpha}}_{i,l}^T, \dots, \hat{\boldsymbol{\alpha}}_{i,N_s}^T\right]^T$ and $\|\cdot\|_2$ is the $L2$-norm operator. By replacing the lower-level problem using constraints (17) – (20), the proposed bilevel voltage regulation problem can be converted to a single-level problem.

B. Activation Function $\sigma(\cdot)$

The selection of activation function is the key to DNN based lower-level problem reformulation. The most popular activation function is the rectified linear unit (ReLU), which is the simplest activation function that can be represented by mixed-integer linear constraints in an exact way [14]. Compared with ReLU, other nonlinear and smooth activation functions such as tanh(·), sigmoid(·), and Gaussian(·) can also be selected to form a DNN model. However, the nonlinear and smooth activation functions do not have exact mixed integer representations.

Clearly, the selection of the activation functions affects the mathematical model in the final single-level formulation. To thoroughly study the impact of the activation functions in solving the bilevel voltage regulation problem, the ReLU activation function and the tanh(·) activation function are select-

ed in this paper to build the DNN based models. Mathematically, the ReLU function is defined as

$$\sigma_R(x) = \max(x, 0) \tag{22}$$

where $\max(x, 0)$ returns a larger value between input x and 0. The ReLU function can be considered as a representative of all the other activation functions, which can be converted to mixed-integer linear constraints. In comparison, the tanh(·) function represents the nonlinear smooth functions, and it can be expressed as

$$\sigma_T(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{23}$$

Examples of the ReLU and tanh(·) are plotted in Fig. 5 for comparison purposes. The detailed process of converting the ReLU function into mixed-integer linear constraints is provided in Appendix A.

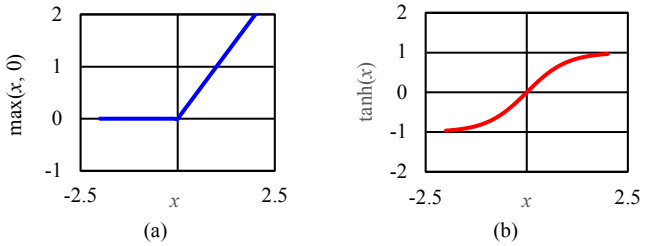


Fig. 5. Activation functions in DNN. (a) ReLU function; (b) tanh(·) function.

Note that the authors do compare the performance between tanh(·) function with other popular activation functions such as sigmoid(·), and Gaussian(·). Since the prediction accuracies of sigmoid(·) and Gaussian(·) are slightly lower than or equivalent to tanh(·), we select tanh(·) as a representative.

C. Training Data Generation

Training data generation strategy is crucial to DNN model training. Unsuccessful training data set often leads to more complicated network structures, lower prediction accuracy, and longer training time. Remark that this paper assumes that the subsystem operators have sufficient information about their system to generate training data. Either subsystems or host systems can do the model training using the generated data.

To reduce the complexity of the training data, the voltage constraints (13) in the lower-level problem are relaxed as

$$\begin{gathered} \underline{\boldsymbol{U}} - \Delta\underline{u} \le \boldsymbol{U} \le \overline{\boldsymbol{U}} + \Delta\overline{u} \\ \Delta\underline{u} > 0 \text{ and } \Delta\overline{u} \ge 0 \end{gathered} \tag{24}$$

where $\Delta\underline{u}$ and $\Delta\overline{u}$ are scalar slack variables. The objective function of the subsystem is adjusted correspondingly as

$$\tilde{f}_i(\boldsymbol{U}_{t,i}, \boldsymbol{y}_i) = f_i(\boldsymbol{U}_{t,i}, \boldsymbol{y}_i) + \mathcal{M}_L(\Delta\underline{u} + \Delta\overline{u}) \tag{25}$$

where $f_i(\boldsymbol{U}_{t,i}, \boldsymbol{y}_i)$ is the original objective function for lower-level problem shown in (3) and $\mathcal{M}_L$ is a big number. Note that the voltage constraint relaxation proposed in (24)-(25) can reduce the data generating time significantly.

In addition, the proposed relaxation eliminates the requirement of an additional DNN model for feasibility classification. As mentioned in Section II, improper host system operation often causes the infeasibility of the lower-level problem. Without the relaxation (24)-(25), a feasibility classification model is necessary to avoid the bilevel problem being trapped in the infeasible region during the problem-solving process. An extra classification model results in a more complex problem and increases the training time and training data. Although (24)-(25) are straightforward relaxation ideas, it is critical for the DNN modeling approach with power system applications.

D. Lower-Level Problem Update Frequency

To ensure the optimality and accuracy of the proposed bilevel operation, the DNN model needs to be updated in response to the system events. The target update frequency can be designed according to the operation strategies of the host system (minutes, hourly, or day-ahead timeframes). The trained DNN model will be updated during every dispatch if necessary. For example, in a 15-mins dispatch frequency, the DNN model can be updated every 15-mins to handle the significant system topology changes or loads changes.

Similar to the voltage regulation problem for centralized operated distribution systems [4]-[6], the proposed voltage regulation operation cannot cover all unexpected system events such as faults. Hence, the DNN model does not need to be able to predict the system behavior accurately under unexpected events. A robust bilevel formulation might be able to provide certain performance guarantees for unexpected events. However, the robust bilevel voltage regulation is beyond the scope of this paper.

IV. Upper-Level Problem Reformulation and Benchmark

After converting the lower-level problem into a DNN model, the bilevel problem can be reformulated as a single-level voltage regulation problem, while the subsystems are considered as multiphase nonlinear voltage-dependent loads. Different DNN models create different types of single-level problems.

In Section III, the lower-level problem has been relaxed for DNN training data generation, and the upper-level problem needs to be adjusted accordingly. To limit the error of the relaxed problem within an acceptable range, additional constraints are added in the upper-level problem as

$$\Delta\underline{u} \le \mathcal{M}_U \text{ and } \Delta\overline{u} \le \mathcal{M}_U \tag{26}$$

where $\mathcal{M}_U$ is a small number that controls the exactness of lower-level voltage limit constraints. $\mathcal{M}_U$ can be set to zero to achieve an exact relaxation of the original bilevel problem. The overall solution algorithm is concluded in Algorithm 1.

Algorithm 1 DNN Model based Bilevel Problem

Identify subsystem problem (1d)-(1f) and chain variables $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$ for all i;
Select activation function from DNN model;
for i = 1: N:
  Relax subsystem problem using (24)-(25)
  Generate $\boldsymbol{X}$ randomly and compute $\boldsymbol{Y}$ using solver (i.e. Gurobi, IPOPT);
  Training DNN model using data pair ($\boldsymbol{X}$, $\boldsymbol{Y}$);
  Building i[th] DNN model using (17)-(20);
end for
Form single-level problem by replacing (1d)-(1f) by DNN models;
Add constraint (26) to single-level problem;
Solving single-level problem using solver (i.e. Gurobi, IPOPT).

According to Algorithm 1, each subsystem can train its

DNN model separately and provide the encoded DNN model to the host system without providing network and control information. Besides, the subsystem training processes are not coupled with other subsystems. Hence, the training processes can be fully distributed or parallel to reduce the training time.

For N subsystems, each subsystem can learn their system by themselves. After the N DNN models are well-trained, the operator of each subsystem can send the model to the host system operator, and the operator of the host system will solve the DNN-based single-level problem.

It is worth pointing out that each DNN model does not need to synchronize with other subsystems during training. The DNN just represents the behaviors of each subsystem. The synchronization between each subsystem and the host system will be done during the solving process of the upper-level problem.

A. Mixed-Integer Quadratic Programming Formulation

Since the ReLU function enabled DNN model can be formulated as mixed-integer linear constraints, the single-level voltage regulation problem becomes a mixed-integer programming (MIP) problem. Due to the nonlinear term $\boldsymbol{A}_i\boldsymbol{Y}_{km}\boldsymbol{A}_v\boldsymbol{V}_k$ in the transformer model shown in (8), the highest polynomial degree is three.

The problem with the MIP components and degree three polynomials cannot be solved efficiently by commercial solvers. Hence, the integer variable $\boldsymbol{t}_T$ is modeled by a set of binary variables, and the third-degree polynomial term can be exactly relaxed as a set of mixed-integer constraints. The detailed process of relaxation process can be found in Appendix B. However, the proposed binary variable representation of the transformer taps cannot be applied to closed-delta and open-delta transformers with phase-independent taps. Hence, this paper only considers multiphase tap-changeable transformers in the wyn connection and the open-delta connection with phase-dependent taps.

Finally, the single-level voltage regulation problem becomes a problem with mixed-integer linear constraints and quadratic cost that can be computed by many existing commercially available solvers.

B. Nonlinear Programming Formulation

Different from the ReLU-based DNN model, the tanh(·) enabled DNN model is established by smooth functions that cannot be represented by mixed integer constraints exactly. Hence, we formulate the tanh(·) enabled DNN model as a nonlinear constraint set.

Due to the nonlinearity of tanh(·), introducing integer variables such as $\boldsymbol{t}_T$ and $\boldsymbol{t}_C$ increase the complexity of the single-level problem. Therefore, the integer variables (tap positions) have been relaxed to continuous variables, and the final single-level problem becomes a pure nonlinear problem without any integer variables. Although the global optimum of nonconvex problems cannot be guaranteed, different off-the-shelf solvers can still be applied to compute the local optimal solution. In this paper, the final nonlinear problems are solved effectively using IPOPT [23]. After solving the nonlinear single-level problem, quantization processes will be applied to converting the tap position variables to integers.

C. Heuristic Searching Benchmark

Although KKT condition and primal-dual approaches are widely used methods for converting bilevel formulations into single-level problems. The complementary constraint in the KKT condition of the lower-level problem may include integer variables and high degree nonlinear constraints. Modeling the lower-level problem using the KKT condition can neither ensure the optimality of the lower-level problem nor reduce the complexity of this original problem. Hence, the KKT condition method is not selected as a benchmark to compare with the proposed DNN based method.

Algorithm 2 Heuristic Searching Method

for particle $j = 1,\ldots, J$:
  Initialize $\boldsymbol{x}_j = [\boldsymbol{t}_{T,j}, \boldsymbol{t}_{C,j}, \boldsymbol{s}_{A,j}]$ for jth particle and particle best position;
  Solve (1b) using $\boldsymbol{x}_j$ and identify $\boldsymbol{U}_{t,i}$ for all subsystems;
  Solve low-level problems (1d)-(1f) with $\boldsymbol{U}_{t,i}$ and compute $[\boldsymbol{p}_{t,i}, \boldsymbol{q}_{t,i}]$;
  Update $\boldsymbol{U}_{t,i}$ with $\boldsymbol{x}_j$ and $[\boldsymbol{p}_{t,i}, \boldsymbol{q}_{t,i}]$;
  Repeat until changes of $\boldsymbol{U}_{t,i}$ is small;
  Evaluate cost (2) and update the global best position.
  Initialize the particle's velocity $\boldsymbol{v}_j$;
end for
while termination criterion is not met **do**:
  for particle $j = 1,\ldots, J$:
    Update velocity $\boldsymbol{v}_j$;
    Update position $\boldsymbol{x}_j$;
    Repeat to compute a stable $[\boldsymbol{p}_{t,i}, \boldsymbol{q}_{t,i}]$ with updated $\boldsymbol{x}_j$
    Update particle's best position;
    Update global best position;
  end for

To evaluate the performance of the proposed DNN based methods for solving bilevel problems, a heuristic searching method is applied to solve the proposed bilevel problem directly. The lower-level problem (with upper-level constraints and fixed decision variables) will be solved by commercially available solvers during each searching step, and the upper-level problem will be solved using particle swarm optimization (PSO). The algorithm for using PSO to solve the proposed bilevel problem is provided in Algorithm 2.

V. Numerical Examples

In this section, a simple distribution system shown in Fig.1 is first investigated to illustrate the difference between the global voltage regulation problem, local voltage regulation, and bilevel voltage regulation problem. Later, comprehensive numerical studies using the IEEE 123-bus test system and a realistic distribution system in Iowa [24] are presented to compare the proposed data-driven modeling approach with the existing bilevel problem-solving method.

A. Simple Example of Voltage Regulation Problem

The system topology of the first example is shown in Fig. 1. Two DERs are installed at bus 2 and bus 4, respectively. The upper and lower bounds for the active power of each DER are 0.3 p.u. and -0.3 p.u. The upper and lower bounds for the reactive power are 0.1 p.u. and -0.1 p.u. The substation voltage is set to 1.02 p.u. and the line impedance is 0.002+j0.04 p.u. The

loads assigned to all four buses are 0.2+*j*0.05 p.u.

Four different voltage regulation scenarios are computed in this example. This first scenario is the base case in which the outputs of DERs are zero. The rest three scenarios are the global regulation, the local regulation, and the bilevel voltage regulation. In the global problem, the host system operator controls two DERs to regulate all four bus voltage amplitudes to 1 p.u. In the local regulation problem, the host system operation controls the DER 1 to regulate the voltages of buses 1 and 2 to 1 p.u. The subsystem operator controls the DER 2 to minimize the loss of the subsystem consisted of buses 3 and 4. The subsystem operator does not communicate with the host system operator in the localized scenario.

In the last scenario, the bilevel problem is studied. The subsystem still minimizes its line loss, and the host system operator knows the strategy of the subsystem. The voltage profiles under four different scenarios are presented in Fig. 6, and the corresponding dispatch solutions and the system losses are concluded in Table I.

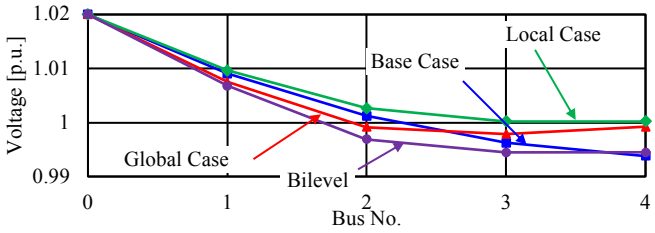


Fig. 6. Voltage profiles for four management scenarios.

TABLE I
DISPATCH SOLUTIONS AND SYSTEM LOSSES FOR FOUR-BUS SYSTEM

	Base Case	Global	Local	Bilevel
DER 1 [p.u.]	0 + *j*0	0.3 + *j*0.076	0.3 + *j*0.029	0.3 + *j*0.1
DER 2 [p.u.]	0 + *j*0	0.089 – *j*0.1	– 0.2 – *j*0.05	– 0.2 – *j*0.05
Loss [p.u.]	0.0026	0.0056	0.0028	0.0029

According to the voltage profiles shown in Fig. 6, different management structures lead to different dispatch solutions. The global operation results in the smallest voltage derivation and largest line loss. Compared with the fully localized operation, the bilevel voltage regulation balances the minimizations of voltage deviation and the line losses in the subsystem. Hence, voltage regulation problems with the bilevel structure are different from traditional voltage regulation problems.

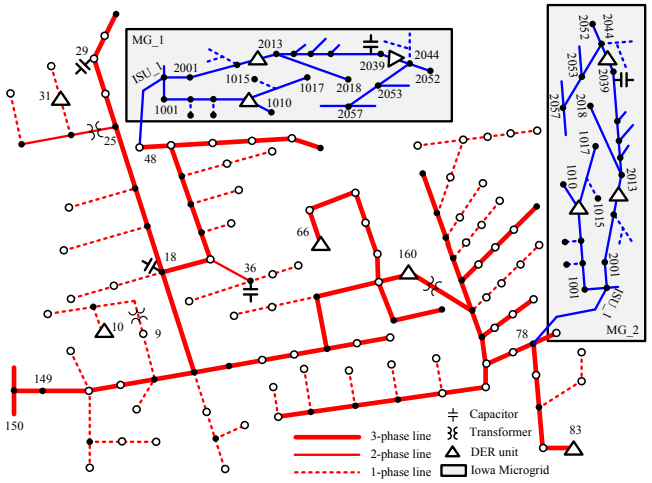


Fig. 7. One-line diagram of the IEEE123-Iowa240 joint system that is under investigation. Two Iowa240 systems are denoted as MG_1 and MG_2.

B. IEEE 123-bus System with Single Microgrid

The test system topologies of the second and third examples are plotted in Fig. 7, in which the Iowa distribution systems are considered as a self-managed microgrid (denoted as MG_1 and MG_2) and attached to the host IEEE 123-bus test system. To avoid host system overloading, the feeder C from the original Iowa distribution system is removed, and the total loads of the Iowa system are reduced to 430.45 kW and 154.06 kVar. Besides, the line lengths in the Iowa system are extended to create sufficient voltage deviations in the microgrids. The active energy resources managed by the Iowa system are located at buses 1009, 2011, and 2043. Two Iowa systems with identical configurations are connected to the IEEE 123-bus system at buses 48 and 78. The original loads at buses 48 and 78 are removed from the IEEE 123-bus system. The information about other controllable devices such as tap-changeable transformers, capacitor banks are collected and presented in Table II. The voltage weights for both host and subsystems are set to 1 in the following numerical examples.

TABLE II
SYSTEM SETUPS FOR JOINT DISTRIBUTION SYSTEM

System	No.	Bus	Type (Total)	Bounds
IEEE 123	1-6	13.abc, 22.b, 48.b, 56.abc, 100.abc, 151.abc	DER (14)	±100 kW±100 kVar
	7-9	160.a - 160r.a, 160.b - 160r.b, 160.c - 160r.c	Tap (3)	0.9325 - 1.0675
	10-12	25.a - 25r.a, 25.c - 25r.c, 9.a - 9r.a	Tap (3)	
	13-15	36.a, 29.b, 18.c	Cap (3)	0 - 0.2 p.u.
Iowa 240	16-18	1009.abc, 2011.abc, 2043.abc	DER (9)	±100 kW±100 kVar
	19-21	2038.abc	Cap (3)	0 - 0.2 p.u.

In this example, the joint system is simplified to enable the comparison between different algorithms for bilevel problems. To limit the complexity of the target system, the MG_2 microgrid is disconnected from the host system and the capacitor banks in the MG_1 are disabled. Only No. 7, 10, and 12 transformer taps in the host IEEE 123-bus system are enabled.

As described in Section II, the MG_1 only controls its AERs to minimize its voltage deviation (3a) without considering the impact on the host system. Meanwhile, the host system (IEEE 123-bus system) operates its transformers, capacitor banks, and AERs to minimize the voltage deviation of the host system. To study the effectiveness of the proposed approaches, this bilevel operation problem is solved using the DNN based approaches as well as a PSO-based heuristic searching approach.

The DNN structure applied to model the microgrid voltage regulation problem is plotted in Fig. 8. According to the structure of the DNN, the number of hidden layers is one, and the number of neurons per hidden layer is set to 100. Mean-

while, the number of training data pairs for training the DNN networks is 1000. The total data generation time is 45.48 s for both DNN models. Other information related to the training process is provided in Table III.

TABLE III
TRAINING PROCESS OF IEEE 123-BUS SYSTEM WITH SINGLE MICROGRID

σ(·)	No. of Data	Generation time	Training time	MSE [%]
ReLU	1000	45.48 [s]	161 [s]	9.41e-4
tanh(·)	1000	45.48 [s]	170 [s]	1.16e-3

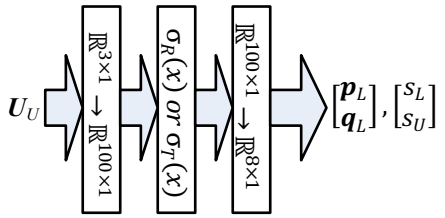


Fig. 8. Structure of DNN applied in this paper.

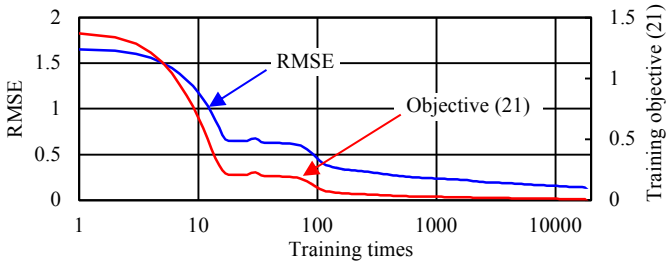


Fig. 9. Example root mean square error (RMSE) and training objective (loss) function value during the training process of tanh(·).

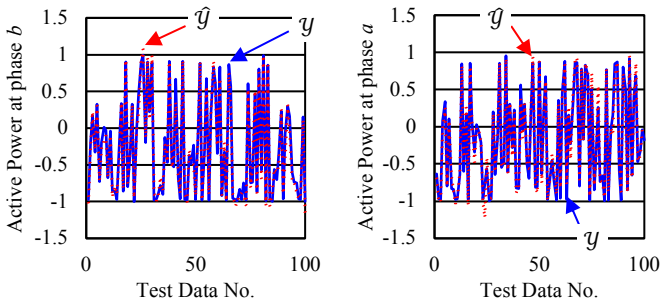


Fig. 10. Prediction results of different DNN models after normalization. (a) ReLU based DNN model. (b) tanh(·) based DNN model.

TABLE IV
ROOT MEAN SQUARE ERROR COMPARISON BETWEEN DIFFERENT MACHINE LEARNING METHODS

	DNN	SVM	Gaussian Kernel
RMSE	0.108	0.253	0.145

The training process of the tanh(·) based DNN model for MG_1 is shown in Fig. 9. According to Fig. 9, the training objective function value reduces to 1.16×10^{-3} at the end of the training. The ending root-mean-square error (RMSE) of the training data is around 0.2. The graphic examples to show the accuracies of the DNN model are presented in Fig. 10. Although two different DNN models are trained by the same data set, the performances of the DNN models are different from each other. The worst prediction of the ReLU based DNN model occurs at the active power of phase *b*. In comparison, the worst prediction of the tanh(·) based DNN model is the active power at phase *a*. However, the prediction errors of both DNN models are acceptable for lower-level problem modeling. The DNN model is also compared with other machine learning methods such as support vector machine (SVM) and Gaussian Kernel regression, as shown in Table IV. According to the comparison results, the DNN model has the highest accuracy in the active power prediction at phase *b* of the microgrid.

Remark that the prediction error of both the DNN model can be further reduced by complicating the network structure and increasing the number of neurons, which will unfortunately complicate the single-level problems.

For comparison and benchmarking purposes, we have applied four different methods to solve the proposed problem. The first three methods are two DNN based methods and the heuristic searching method that solve the voltage regulation problem with bilevel structure. The fourth method solves a global voltage regulation problem for using PSO only since the objectives for both systems are identical. After solving the bilevel voltage regulation problem, the corresponding voltage profiles are plotted in Fig. 11. All four methods can fix the voltage problem in the joint system. According to Fig. 11, the voltage profiles computed from DNN based methods are almost overlapped with the centralized solution. However, the voltage profile of the heuristic method is not as good as the DNN based methods. To compare the performances of different methods numerically, the optimality data and the corresponding solution information are concluded in Table V.

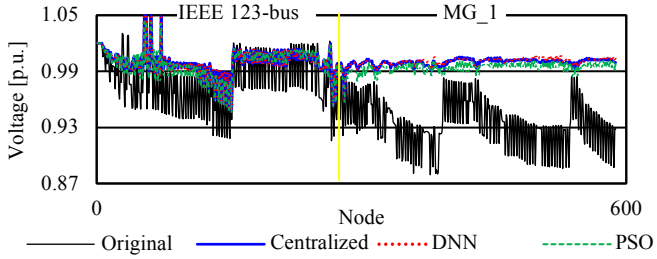


Fig. 11. Voltage profile comparison of host IEEE system computed by DNN model approach, heuristic searching, and centralized operation.

TABLE V
SOLUTION OF IEEE 123-BUS SYSTEM WITH SINGLE MICROGRID

	DNN(ReLU)	DNN(tanh)	Heuristic	Central
Lower-level solver	Gurobi(QP)	Gurobi(QP)	IPOPT(NP)	PSO
Upper-level solver	Gurobi (MIQP)	IPOPT(NP)	PSO(NP)	
ε_U[1] [%]	0.132	0.24	-	-
Objective$_L$ (3a)	0.0036	0.0036	0.0157	0.0032
Objective$_U$ (2)	0.06	0.06	0.0705	0.0596
$\|\boldsymbol{U}\|_2^2$	0.0635	0.0635	0.0862	0.0628
Loss Reduction	25.9%	25.7%	29.6%	22.5%
Solution Time [s]	2178.4	31.72	2499.9	2424.2
Remark	5% gap	-	Max Iteration	-

According to Table V, different DNN models can generate similar solutions for the bilevel voltage regulation problem and the voltage errors are negligible. However, the ReLU

[1] ε_U is the maximum node voltage different between the linear power flow and the nonlinear power flow results computed from OpenDSS.

based DNN model requires 36.3 minutes to compute the solution with 5% gaps, which is almost 69 times longer than the tanh(·) based DNN model. The performance of the heuristic method is even worse, which takes 42 minutes to reach a non-optimal solution. As expected, the voltage deviation of the centralized operation is the smallest one and different from the bilevel operated system. Although loss minimization is not the operational objective in this example, the regulation operation also affects the system loss reduction. According to the system losses results in Table V, the global voltage regulation has a smaller loss reduction compared with bilevel operations.

Another advantage of the DNN based method is that the upper-level operator does not require any details of the subsystems during the bilevel problem-solving process from the host system perspective. The only information they need is the DNN encoded model. Hence the privacy and security between different systems can be enhanced.

C. *IEEE 123-bus System with Multiple Microgrids*

In the third example, the MG_2 microgrid is reconnected to the host IEEE 123-bus system at bus 78, and all the transformer taps listed in Table I are controllable. In addition, the capacitor banks in both microgrids are activated. The total node number in the joint system increased to 910. Different from scenario one, the objective of the MG_2 is minimizing its system line losses (3b), and the objective function of the MG_1 is still to minimize its own voltage deviation (3a). The DNN structures for both microgrids are identical to the one shown in Fig. 8. Note that the ReLU based DNN method and the heuristic searching method are not selected to solve the problem in this example since both methods take days to solve the proposed bilevel problem.

Based on the proposed tanh(·) based DNN method, the data generation and training time of two microgrids are shown in Table VI. Based on the results shown in Table VI, the data generation time with capacitor banks is longer than the second example since the lower-level problem shifts from quadratic programming problems to MIQP problems.

TABLE VI
TRAINING PROCESS OF IEEE 123-BUS SYSTEM WITH TWO MICROGRID

	No. of Data	Generation time	Training time	MSE [%]
MG_1	1000	533.77 [s]	167 [s]	1.43E-03
MG_2	1000	286.90 [s]	169 [s]	6.80E-04

TABLE VII
SOLUTION OF IEEE 123-BUS SYSTEM WITH TWO MICROGRID

	Lower-level solver	Upper-level solver	ε_U [%]	Solution Time [s]
	Gurobi (MIQP)	IPOPT(NP)	0.087	8.18
	Objective_{L1} (3a)	Objective_{L2} (3b)	Objective_{U} (2)	$\|\boldsymbol{U}\|_2^2$
Before [p.u.]	0.109	0.0122	0.2512	0.4693
After [p.u.]	0.0057	0.0023	0.0194	0.0343

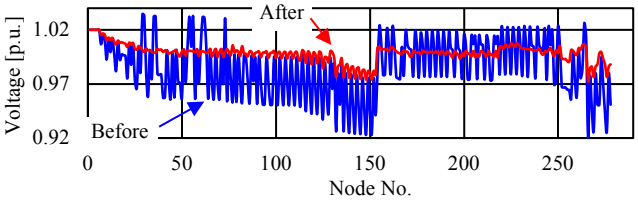


Fig. 12. voltage profile of the host IEEE 123 system before and after optimization.

The solutions to the bilevel problem with two microgrids are shown in Table VII. According to Table VII, the computational time of the DNN based approach is 8.18 s, which is acceptable for all near real-time operations. Besides, the voltage deviation of the MG_1 is reduced by 94.7%, and the line loss of the MG_2 is reduced from 12.28 kW to 2.3 kW. They are satisfied with their operational objectives. The voltage deviation of the host IEEE 123-bus system is reduced by 92.28%, which is also better than the second example since more transformer taps are available this time. A voltage profile comparison study between the pre-optimization and post-optimization is shown in Fig. 12. According to Fig. 12, the voltages of the host system are well-controlled around 1.0 p.u. compared with the unregulated condition. The maximum voltage profile error generated by linearization and quantization processes is only 0.087%.

VI. CONCLUSION

This paper explores the potential of using DNN based modeling approaches to solve bilevel optimization problems in distribution systems. Compared with the existing methods, the DNN based model can be applied to reformulate subproblems as constraints and integrated into upper-level problems. Although additional time is spent in data generation and training processes, the DNN based model is reusable and more secure than other existing methods because of its non-disclosure features.

Because of the complexity of the MIP problem, the tanh(·) based DNN model is more suitable in solving the voltage regulation problems than the ReLU based DNN model. However, it is worth pointing out that the tanh(·) based DNN model may cause the final single-level problem trapped in a local minimum.

According to the numerical study in this paper, the tanh(·) based DNN model can solve the bilevel voltage regulation problem with fast speed and excellent optimality. Remark that the proposed DNN based approach can also be applied to other bilevel problems with different management objectives.

APPENDIX A

MILP MODEL OF TRANSFORMER AND CAPACITOR

According to the transformer model (8) and the capacitor bank model (11), the system admittance matrix $\boldsymbol{Y}$ becomes a function of tap change positions. Hence, multiplications between integer and node voltage can be found in the power flow formulations. In the following reformulation process, the

tap positions of the transformer and the capacitor bank are generalized as integer variable t. The multiplication between the real part of a node voltage e and integer variable t is selected for illustration.

To simplify the nonlinear constraints to mixed integer linear constraints, the integer tap position t with power m can be modeled by a set of binary variables as

$$t^m = \sum_j^{N_I} t_j^m k_j \tag{A1}$$

$$\sum_j^{N_I} k_j = 1 \; and \; k_j \in \{0,1\} \tag{A2}$$

where t_j is the jth integer constant in the set of t, N_I is the size of the set; k_j is a binary variable; m is a continuous variable in $\mathbb{R}$. After converting integer t_j^m to a set of binaries $\{k_j\}_1^{N_I}$, the multiplication between k_j and e can be modeled by μ_j using big-M as

$$\underline{e} \le \mu_j \le \overline{e} \tag{A3}$$

$$\underline{e} k_j \le \mu_j \le \overline{e} k_j \tag{A4}$$

$$e - \overline{e}(1 - k_j) \le \mu_j \le e - \underline{e}(1 - k_j) \tag{A5}$$

where $\overline{e}$ and $\underline{e}$ are the upper and lower bound of e; μ_j represents $k_j e$. In this paper, the upper and lower bounds of e and f are ±1.5 p.u.

APPENDIX B

MILP ENCODING OF RELU BASED DNN MODEL

A binary vector $\boldsymbol{a}_h$ represents the activation status of ReLU at hth hidden layer, and $a_{h,n}$ is the status of the nth neuron at hth layer. Assume an element $\hat{z}_{h,n}$ in $\hat{\boldsymbol{z}}_h$ can be bounded by $\underline{Z}_{h,n}$ and $\overline{Z}_{h,n}$, the relationship shown in (22) can be represented as [14]

$$z_{h,n} \le \hat{z}_{h,n} - \underline{Z}_{h,n}(1 - a_{h,n}) \tag{B1}$$

$$z_{h,n} \ge \hat{z}_{h,n} \tag{B2}$$

$$z_{h,n} \le \overline{Z}_{h,n} a_{h,n} \tag{B3}$$

$$z_{h,n} \ge 0 \tag{B4}$$

$$a_{h,n} \in \{0,1\} \tag{B5}$$

According to (B1)-(B5), the nonlinear relationship shown in (22) can be converted to a set of mixed-integer linear constraints in an exact manner. Hence, the entire DNN-based model can be rewritten as a large set of mixed-integer linear constraints.

REFERENCES

[1] M. Razmara, G. R. Bharati, M. Shahbakhti, S. Paudyal and R. D. Robinett, "Bilevel Optimization Framework for Smart Building-to-Grid Systems," in IEEE Transactions on Smart Grid, vol. 9, no. 2, pp. 582-593, March 2018, doi: 10.1109/TSG.2016.2557334.

[2] F. Samadi Gazijahani and J. Salehi, "Optimal Bilevel Model for Stochastic Risk-Based Planning of Microgrids Under Uncertainty," in IEEE Transactions on Industrial Informatics, vol. 14, no. 7, pp. 3054-3064, July 2018, doi: 10.1109/TII.2017.2769656.

[3] R. Mohammadi, H. R. Mashhadi and M. Shahidehpour, "Market-Based Customer Reliability Provision in Distribution Systems Based on Game Theory: A Bi-Level Optimization Approach," in IEEE Transactions on Smart Grid, vol. 10, no. 4, pp. 3840-3848, July 2019, doi: 10.1109/TSG.2018.2839598.

[4] R. R. Jha, A. Dubey, C. Liu and K. P. Schneider, "Bi-Level Volt-VAR Optimization to Coordinate Smart Inverters With Voltage Control Devices," in IEEE Transactions on Power Systems, vol. 34, no. 3, pp. 1801-1813, May 2019, doi: 10.1109/TPWRS.2018.2890613.

[5] Z. Wang, J. Wang, B. Chen, M. M. Begovic and Y. He, "MPC-Based Voltage/Var Optimization for Distribution Circuits With Distributed Generators and Exponential Load Models," in IEEE Transactions on Smart Grid, vol. 5, no. 5, pp. 2412-2420, Sept. 2014, doi: 10.1109/TSG.2014.2329842.

[6] P. Li et al., "Coordinated Control Method of Voltage and Reactive Power for Active Distribution Networks Based on Soft Open Point," in IEEE Transactions on Sustainable Energy, vol. 8, no. 4, pp. 1430-1442, Oct. 2017, doi: 10.1109/TSTE.2017.2686009.

[7] F. U. Nazir, B. C. Pal and R. A. Jabr, "Distributed Solution of Stochastic Volt/VAr Control in Radial Networks," in IEEE Transactions on Smart Grid, doi: 10.1109/TSG.2020.3002100.

[8] F. U. Nazir, B. C. Pal and R. A. Jabr, "A Two-Stage Chance Constrained Volt/Var Control Scheme for Active Distribution Networks With Nodal Power Uncertainties," in IEEE Transactions on Power Systems, vol. 34, no. 1, pp. 314-325, Jan. 2019, doi: 10.1109/TPWRS.2018.2859759.

[9] A. Sinha, P. Malo and K. Deb, "A Review on Bilevel Optimization: From Classical to Evolutionary Approaches and Applications," in IEEE Transactions on Evolutionary Computation, vol. 22, no. 2, pp. 276-295, April 2018, doi: 10.1109/TEVC.2017.2712906.

[10] D. A. Arias, A. A. Mota, L. T. M. Mota and C. A. Castro, "A bilevel programming approach for power system operation planning considering voltage stability and economic dispatch," 2008 IEEE/PES Transmission and Distribution Conference and Exposition: Latin America, Bogota, 2008, pp. 1-6, doi: 10.1109/TDC-LA.2008.4641718.

[11] H. Haghighat and B. Zeng, "Bilevel Conic Transmission Expansion Planning," in IEEE Transactions on Power Systems, vol. 33, no. 4, pp. 4640-4642, July 2018, doi: 10.1109/TPWRS.2018.2835663.

[12] S. Pineda and J. M. Morales, "Solving Linear Bilevel Problems Using Big-Ms: Not All That Glitters Is Gold," in IEEE Transactions on Power Systems, vol. 34, no. 3, pp. 2469-2471, May 2019, doi: 10.1109/TPWRS.2019.2892607.

[13] T. Lu, Z. Wang, J. Wang, Q. Ai and C. Wang, "A Data-Driven Stackelberg Market Strategy for Demand Response-Enabled Distribution Systems," in IEEE Transactions on Smart Grid, vol. 10, no. 3, pp. 2345-2357, May 2019, doi: 10.1109/TSG.2018.2795007.

[14] Y. Zhang, C. Chen, G. Liu, T. Hong and F. Qiu, "Approximating Trajectory Constraints with Machine Learning Microgrid Islanding with Frequency Constraints," in IEEE Transactions on Power Systems, doi: 10.1109/TPWRS.2020.3015913.

[15] Y. Du, F. Li, T. Zheng and J. Li, "Fast Cascading Outage Screening Based on Deep Convolutional Neural Network and Depth-First Search," in IEEE Transactions on Power Systems, vol. 35, no. 4, pp. 2704-2715, July 2020, doi: 10.1109/TPWRS.2020.2969956.

[16] T. Hong and F. de León, "Centralized Unbalanced Dispatch of Smart Distribution DC Microgrid Systems," in IEEE Transactions on Smart Grid, vol. 9, no. 4, pp. 2852-2861, July 2018, doi: 10.1109/TSG.2016.2622681.

[17] Z. Tang, D. J. Hill and T. Liu, "Fast Distributed Reactive Power Control for Voltage Regulation in Distribution Networks," in IEEE Transactions on Power Systems, vol. 34, no. 1, pp. 802-805, Jan. 2019, doi: 10.1109/TPWRD.2018.2868158.

[18] R. Sadnan and A. Dubey, "Distributed Optimization Using Reduced Network Equivalents for Radial Power Distribution Systems," in IEEE Transactions on Power Systems, vol. 36, no. 4, pp. 3645-3656, July 2021, doi: 10.1109/TPWRS.2020.3049135.

[19] Microgrid at Illinois Institute of Technology – IIT's Microgrid, online available: http://www.iitmicrogrid.net/microgrid.aspx, Jul. 12, 2021.

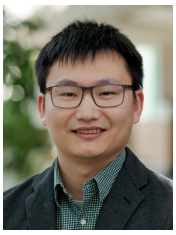
[20] T. Hong, D. Zhao, Y. Zhang, B. Cui and Y. Tian, "Optimal Voltage Reference for Droop-Based DERs in Distribution Systems," in IEEE Transactions on Smart Grid, vol. 11, no. 3, pp. 2357-2366, May 2020, doi: 10.1109/TSG.2019.2953154.

[21] M. Bazrafshan and N. Gatsis, "Comprehensive Modeling of Three-Phase Distribution Systems via the Bus Admittance Matrix," in IEEE Transactions on Power Systems, vol. 33, no. 2, pp. 2015-2029, March 2018, doi: 10.1109/TPWRS.2017.2728618.

[22] D. Elbrächter, D. Perekrestenko, P. Grohs and H. Bölcskei, "Deep Neural Network Approximation Theory," in IEEE Transactions on Information Theory, vol. 67, no. 5, pp. 2581-2623, May 2021, doi: 10.1109/TIT.2021.3062161.

[23] A. Wächter, and L. T. Biegler, “On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming,” Mathematical Programming, vol. 106, no. 1, pp.25-57, Apr. 2005.

[24] F. Bu, Y. Yuan, Z. Wang, K. Dehghanpour and A. Kimber, "A Time-Series Distribution Test System Based on Real Utility Data," 2019 North American Power Symposium (NAPS), Wichita, KS, USA, 2019, pp. 1-6, doi: 10.1109/NAPS46351.2019.8999982.



Tianqi Hong (S'13–M'16) received the B.Sc. degree in electrical engineering from Hohai University, China, in 2011, and the M.Sc. degree in electrical engineering from the Southeast University, China and the Engineering school of New York University in 2013. He received a Ph.D. degree from New York University in 2016. His main research interests are power system analysis, power electronics systems, microgrid, and electromagnetic design.

Currently, he is an Energy Systems Scientist at Argonne National Laboratory. Prior to this, he was a Postdoc Fellow in the Engineering school of New York University and a Senior Research Scientist at Unique Technical Services, LLC, responsible for heavy-duty vehicle electrification, battery energy storage integration, and medium capacity microgrid. Dr. Hong is an active reviewer for multiple international journals in the power engineering area, and he serves as an Editorial Board Member of International Transactions on Electrical Energy Systems and IEEE Transactions on Power Delivery. He also serves as Special Activity Co-Chair of the IEEE IAS Industrial Power Converters Committee (IPCC).

Dongbo Zhao (SM'16) received his B.S. degrees from Tsinghua University, Beijing, China, the M.S. degree from Texas A&M University, College Station, Texas, and the Ph.D. degree from Georgia Institute of Technology, Atlanta, Georgia, all in electrical engineering. He has worked with Eaton Corporation from 2014 to 2016 as a Lead Engineer in its Corporate Research and Technology Division, and with ABB in its US Corporate Research Center from 2010 to 2011. Currently he is a Principal Energy System Scientist with Argonne National Laboratory, Lemont, IL. He is also an Institute Fellow of Northwestern Argonne Institute of Science and Engineering of Northwestern University. His research interests include power system modeling and control, protection, reliability analysis, transmission and distribution automation, and electric market optimization.

Dr. Zhao is a Senior Member of IEEE, and a member of IEEE PES, IAS and IES Societies. He is the editor of IEEE Transactions on Power Delivery, IEEE Transactions on Sustainable Energy and IEEE Power Engineering Letters. He is the subject editor of subject "Power system operation and planning with renewable power generation" of IET Renewable Power Generation and the Associate Editor of IEEE Access.

Yichen Zhang (S'13–M18'–SM'21) received B.S. degree from Northwestern Polytechnical University, Xi'an, China, in 2010, M.S. degree from Xi'an Jiaotong University, Xi'an, China, in 2012, and Ph.D. degree in the Department of Electrical Engineering and Computer Science (EECS) at the University of Tennessee, Knoxville, TN, USA, all in Electrical Engineering. He is currently an Energy Systems Scientist in the Energy Systems Division at Argonne National Laboratory.

His research interests include power system dynamics, renewable energy, control systems and hybrid systems.

Zhaoyu Wang (S'13–M'15–SM'20) is the Northrop Grumman Associate Professor with Iowa State University. He received the B.S. and M.S. degrees in electrical engineering from Shanghai Jiaotong University, and the M.S. and Ph.D. degrees in electrical and computer engineering from Georgia Institute of Technology. His research interests include optimization and data analytics in power distribution systems and microgrids. He is the Principal Investigator for a multitude of projects focused on these topics and funded by the National Science Foundation, the Department of Energy, National Laboratories, PSERC, and Iowa Economic Development Authority. Dr. Wang is the Chair of IEEE Power and Energy Society (PES) PSOPE Award Subcommittee, Co-Vice Chair of PES Distribution System Operation and Planning Subcommittee, and Vice Chair of PES Task Force on Advances in Natural Disaster Mitigation Methods. He is an editor of IEEE Transactions on Power Systems, IEEE Transactions on Smart Grid, IEEE Open Access Journal of Power and Energy, IEEE Power Engineering Letters, and IET Smart Grid. Dr. Wang has been a recipient of the National Science Foundation (NSF) CAREER Award, the IEEE PES Outstanding Young Engineer Award, and the Harpole-Pentair Young Faculty Award Endowment.