A Data-driven Linearization Approach to Analyze the Three-phase Unbalance in Active Distribution Systems

Jiaqi Chen and Line A. Roald
Department of Electrical and Computer Engineering
University of Wisconsin-Madison, Madison, USA
{jiaqi.chen, roald}@wisc.edu

***Abstract*—Due to unevenly distributed energy resources between phases, it is expected that active distribution networks (ADNs) may experience a high degree of three-phase unbalance. Accurate analysis tools play an essential role in addressing the issues caused by the three-phase unbalance. To address existing computational challenges in the assessment and mitigation of three-phase unbalance in ADNs, we propose a *direct* and an *indirect* linear approximation models of three-phase voltage unbalance metrics (*VUF*, *LVUR* and *PVUR*), which are obtained using a support vector regression (SVR)-based three-phase voltage unbalance metric linearization approach. The linearized three-phase voltage unbalance metrics are incorporated in an improved optimal dispatch problem with voltage unbalance constraints. The resulting optimization problem is a convex quadratic program with low computational complexity and high accuracy. Numerical analyses demonstrate that the proposed approach achieves satisfactory accuracy, which indicates its promise for inclusion in three-phase unbalance analysis in ADNs.**

***Index Terms*-- Data-driven linearization, three-phase unbalance.**

I. Introduction

With new operational challenges and opportunities arising from the integration of large-scale distributed generations (DGs), traditional distribution networks are transformed into active distribution networks (ADNs). ADNs are typically three-phase unbalanced systems with single-phase connected rooftop photovoltaics (PVs) and single, double or three-phase loads [1]. Because rooftop PV may be unevenly distributed among phases, ADNs may experience a high degree of three-phase unbalance, which can cause increased power loss, overloads on distribution lines and transformers and voltage deviations of neutral points, which will affect the secure operation of ADNs [2].

Accurate analysis tools are necessary to address the issues caused by the three-phase unbalance. The voltage unbalance is defined by various organizations such as IEC [3], NEMA [4], and IEEE [5]. However, these definitions are different and partially inconsistent, and voltage unbalance levels considered acceptable by one standard may violate limits defined by another standard [6]. There are a number of papers that mitigate the voltage unbalance in ADNs with a three-phase optimal power flow (OPF) model [7]-[11]. However, they use only one type of voltage unbalance definition either as objective or constraints in the problem formulation. The authors of [12] create a flexible three-phase OPF formulation that includes three different voltage unbalance metrics defined by IEC, NEMA, and IEEE (*VUF*, *LVUR* and *PVUR*) as both constraints and objective functions. However, in the OPF models in [12], the equations describing the three-phase power flow (PF) and the voltage unbalance metrics are nonlinear. These non-linear equations lead to a non-convex optimization problem, which can be hard to solve due to non-convergence problems and high computational burdens. This issue can be resolved by deriving a linear representation of the power flow and voltage unbalance metrics, however, the highly non-linear nature of the voltage unbalance metrics makes this challenging.

In general, there are two types of linearization approaches: model-based and data-driven-based approaches. The model-based approach uses assumptions based on physical knowledge to approximate nonlinear terms with a linear expression. Many existing works have proposed linearizations of the PF with a model-based approach, including [13]-[15]. The voltage unbalance metric is linearized in [16] by assuming the product of two voltage magnitudes at different phases is equal to the square of the voltage magnitude at one phase. Unfortunately, the physical assumptions underpinning such model-based linearizations may not be applicable for practical ADNs under variable operating conditions, and as a result, the model-based approach will involve errors. Moreover, paper [16] only derives the linear approximation for *PVUR*, the other three-phase unbalance definitions are not considered. In contrast to model-based approaches, a data-driven linearization is essentially a linear regression problem, which uses training data to derive the linear relationship between the input variables (such as the power generation and consumption) and the independent variables (such as voltage magnitudes or voltage unbalance) without physical assumptions. With the increasing amount of measurement devices, the data-driven linearization approach using the measurement data is a promising approach and can achieve high accuracy [17]. Different data-driven PF linearization approaches are proposed in [18]-[20] based on different regression algorithms to tackle different issues. However, to the best knowledge of the authors, there is no work on linearizing the three-phase voltage unbalance metrics using the data-driven approach. Furthermore, the aforementioned data-driven linear PF models have not been leveraged to solve optimal dispatch problems.

Submitted to the 22nd Power Systems Computation Conference (PSCC 2022).
This work is funded by the National Science Foundation (NSF) under Grant. No. ECCS-2045860.

Therefore, this paper proposes a support vector regression (SVR)-based approach to linearize the three-phase unbalance metrics from a data-driven point of view. To obtain the most accurate linear representation of the unbalance metrics, we propose two types of linear approximation models (*direct* and *indirect approximation*). Then the obtained linear approximations are applied to an improved optimal dispatch problem to alleviate the three-phase unbalance in ADNs.

The contribution of this paper is three-fold:

1) An SVR-based three-phase voltage unbalance metric linearization approach is proposed for *VUF*, *LVUR* and *PVUR*. Two *direct* and *indirect* linear approximation models are proposed, where the *direct approximation* directly expresses *VUF*, *LVUR* and *PVUR* as a function of the input variables, and the *indirect approximation* uses an intermediate step to express the voltage deviation for the individual phases and then uses this value to express *LVUR* and *PVUR*. Our results show that the proposed *indirect* approach is significantly more accurate than the more straightforward *direct* approach.

2) To alleviate the three-phase unbalance in ADNs with low computational complexity and high accuracy, the optimal dispatch model in [16] is improved by using a data-driven linear PF model as the power flow constraint, and applying the proposed linear three-phase voltage unbalance metric models of the three definitions (*VUF*, *LVUR* and *PVUR*) as unbalance constraints.

3) The accuracy and computational efficiency of the proposed method is demonstrated in the case study. Moreover, based on the *direct* linear approximation model, the sensitivity of the three-phase unbalance metrics with respect to the power injections is analyzed, which can help the system operator identify the most important locations to monitor to track and alleviate the three-phase unbalance in ADNs.

The remainder of this paper is organized as follows. Section II proposes an SVR-based three-phase voltage unbalance metric linearization approach. Section III applies the proposed linear three-phase voltage unbalance metric models to an improved optimal dispatch problem. In Section IV, numerical simulations are conducted to analyze the effectiveness of the proposed method, and conclusions are drawn in Section V.

II. SVR-based Three-phase Voltage Unbalance Metric Linearization Approach

In this section, we first propose two linear approximation models of the three-phase unbalance metrics from a data-driven perspective. Then we propose a data-driven approach to obtain the linear three-phase voltage unbalance metric models based on the SVR algorithm.

A. Linear Three-phase Voltage Unbalance Metric Model

Voltage unbalance occurs when the phase voltage magnitude is asymmetric or the phase angle displacement is not equal to $\angle 120°$ in the three-phase power system. There are three commonly used definitions (*VUF*, *LVUR* and *PVUR*) to evaluate the voltage unbalance. Since the three definitions are nonlinear, which will involve computational burden and cause non-convexity in the optimization application, we propose the following linear three-phase unbalance metric models.

1) IEC definition (VUF) The VUF definition is given by

$$VUF[\%] = |V_n| / |V_p| \times 100, \text{ where}$$
$$V_p = (V_a + a \cdot V_b + a^2 \cdot V_c)/3, V_n = (V_a + a^2 \cdot V_b + a \cdot V_c)/3. \tag{1}$$

where V_n and V_p are the negative and positive sequence voltage phasors, respectively; $a = 1\angle 120°$; and V_a, V_b, V_c are the three-phase line-to-ground voltage phasors. We build the *direct* linear relationship between the power injection and *VUF* to approximate *VUF* at one node as follows.

$$VUF_{direct} = \boldsymbol{A}^{VUF} \cdot \boldsymbol{p} + \boldsymbol{B}^{VUF} \cdot \boldsymbol{q} + \boldsymbol{C}^{VUF} \cdot \boldsymbol{u}_{root} + D^{VUF} \tag{2}$$

where $\boldsymbol{A}^{VUF}$, $\boldsymbol{B}^{VUF}$, $\boldsymbol{C}^{VUF}$ are coefficients vectors, and D^{VUF} is a constant parameter. $\boldsymbol{p}$, $\boldsymbol{q}$ are the vectors of three-phase active and reactive power injections. The root node is the slack bus and $\boldsymbol{u}_{root}$ is the vector of squared line-to-ground voltage magnitude at the root node. The coefficients $\boldsymbol{A}^{VUF}$, $\boldsymbol{B}^{VUF}$ in this *direct approximation* model represent the sensitivity of the three-phase unbalance metrics with respect to the power injections, which can be used to analyze the impact of the power injections on the unbalance metrics.

2) NEMA definition (LVUR) The LVUF definition is calculated by line-to-line voltage magnitudes $\mathbf{V}_{ab}$, $\mathbf{V}_{bc}$ and $\mathbf{V}_{ca}$ as follows,

$$LVUR[\%] = |\Delta V_L^{\max}| / |V_L^{avg}| \times 100,$$
$$\text{where} \quad V_L^{avg} = (V_{ab} + V_{bc} + V_{ca})/3,$$
$$\Delta V_L^{\max} = \max\{|V_{ab} - V_L^{avg}|, |V_{bc} - V_L^{avg}|, |V_{ca} - V_L^{avg}|\} \tag{3}$$

It can be observed that $\Delta V_L^{\max}$ is calculated by combining the calculation of absolute value and the maximum value. Fig.1 illustrates the corresponding calculation. The red line is the final calculation result. It can be observed that for the linear functions which have negative values at some points, the calculation of the maximum value of the absolute value causes the relationship to be highly non-linear. Therefore, if we use a *direct* linear approximation to approximate *LVUR*, the accuracy might be low. To address this, we define the following two types of linear approximations for the *LVUR*:

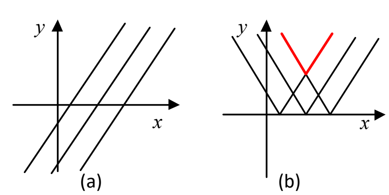

Figure 1. Illustration of the maximum and absolute value calculation. (a) the original functions, (b) the final calculation result.

a) Direct approximation. In this form, we directly build the linear relationship between the power injection and LVUR as follows,

$$LVUR_{direct} = \boldsymbol{A}^{LVUR_{direct}} \cdot \boldsymbol{p} + \boldsymbol{B}^{LVUR_{direct}} \cdot \boldsymbol{q} + \boldsymbol{C}^{LVUR_{direct}} \cdot \boldsymbol{u}_{root} + D^{LVUR_{direct}}. \quad (4)$$

where $\boldsymbol{A}^{LVUR_{direct}}$, $\boldsymbol{B}^{LVUR_{direct}}$ and $\boldsymbol{C}^{LVUR_{direct}}$ are coefficients vectors, and $D^{LVUR_{direct}}$ is a constant parameter.

b) Indirect approximation. In this approximation, we define the auxiliary variable R^{LVUR}, and first build the linear relationship between the power injection and the auxiliary variable R^{LVUR} as follows,

$$\begin{aligned} R_{ab}^{LVUR} = V_{ab}/V_L^{avg} &= \boldsymbol{A}^{R_{ab}^{LVUR}} \cdot \boldsymbol{p} + \boldsymbol{B}^{R_{ab}^{LVUR}} \cdot \boldsymbol{q} + \boldsymbol{C}^{R_{ab}^{LVUR}} \cdot \boldsymbol{u}_{root} + D^{R_{ab}^{LVUR}} \\ R_{bc}^{LVUR} = V_{bc}/V_L^{avg} &= \boldsymbol{A}^{R_{bc}^{LVUR}} \cdot \boldsymbol{p} + \boldsymbol{B}^{R_{bc}^{LVUR}} \cdot \boldsymbol{q} + \boldsymbol{C}^{R_{bc}^{LVUR}} \cdot \boldsymbol{u}_{root} + D^{R_{bc}^{LVUR}} \\ R_{ca}^{LVUR} = V_{ca}/V_L^{avg} &= \boldsymbol{A}^{R_{ca}^{LVUR}} \cdot \boldsymbol{p} + \boldsymbol{B}^{R_{ca}^{LVUR}} \cdot \boldsymbol{q} + \boldsymbol{C}^{R_{ca}^{LVUR}} \cdot \boldsymbol{u}_{root} + D^{R_{ca}^{LVUR}} \end{aligned} . \quad (5)$$

where $\boldsymbol{A}^{R_{\varphi}^{LVUR}}$, $\boldsymbol{B}^{R_{\varphi}^{LVUR}}$ and $\boldsymbol{C}^{R_{\varphi}^{LVUR}}$ are coefficients vectors, and $D^{R_{\varphi}^{LVUR}}$ is a constant parameter. Then the *LVUR* can be calculated as follows.

$$LVUR_{indirect} = \max\left\{\left|R_{ab}^{LVUR}-1\right|, \left|R_{bc}^{LVUR}-1\right|, \left|R_{ca}^{LVUR}-1\right|\right\}. \quad (6)$$

Since R_{ab}^{LVUR}, R_{bc}^{LVUR} and R_{ca}^{LVUR} are affine functions, and $LVUR_{indirect}$ is the pointwise maximum of $\left|R_{ab}^{LVUR}-1\right|, \left|R_{bc}^{LVUR}-1\right|$ and $\left|R_{ca}^{LVUR}-1\right|$, $LVUR_{indirect}$ is convex and piecewise affine.

3) IEEE definition (PVUR) The PVUR definition is calculated by line-to-ground voltage magnitudes $\mathbf{V}_a$, $\mathbf{V}_b$, and $\mathbf{V}_c$ as follows,

$$\begin{aligned} & PVUR[\%] = \Delta V_P^{\max} / V_P^{avg} \times 100, \\ \text{where } & V_P^{avg} = (V_a + V_b + V_c)/3, \\ & \Delta V_P^{\max} = \max\left\{\left|V_a - V_P^{avg}\right|, \left|V_b - V_P^{avg}\right|, \left|V_c - V_P^{avg}\right|\right\}. \end{aligned} \quad (7)$$

Similar to the approximation of *LVUR*, the use of the maximum across the absolute values leads to significant non-linearity. We, therefore, define the following two approximations for *PVUR*:

a) Direct approximation. This is given by

$$PVUR_{direct} = \boldsymbol{A}^{PVUR_{direct}} \cdot \boldsymbol{p} + \boldsymbol{B}^{PVUR_{direct}} \cdot \boldsymbol{q} + \boldsymbol{C}^{PVUR_{direct}} \cdot \boldsymbol{u}_{root} + D^{PVUR_{direct}}. \quad (8)$$

where $\boldsymbol{A}^{PVUR_{direct}}$, $\boldsymbol{B}^{PVUR_{direct}}$ and $\boldsymbol{C}^{PVUR_{direct}}$ are coefficients vectors, and $D^{PVUR_{direct}}$ is a constant parameter.

b) Indirect approximation. Similar to LVUR, we define the auxiliary variable R^{PVUR}, and build the linear relationship as follows,

$$\begin{aligned} R_{a}^{PVUR} = V_{a}/V_P^{avg} &= \boldsymbol{A}^{R_{a}^{PVUR}} \cdot \boldsymbol{p} + \boldsymbol{B}^{R_{a}^{PVUR}} \cdot \boldsymbol{q} + \boldsymbol{C}^{R_{a}^{PVUR}} \cdot \boldsymbol{u}_{root} + D^{R_{a}^{PVUR}} \\ R_{b}^{PVUR} = V_{b}/V_P^{avg} &= \boldsymbol{A}^{R_{b}^{PVUR}} \cdot \boldsymbol{p} + \boldsymbol{B}^{R_{b}^{PVUR}} \cdot \boldsymbol{q} + \boldsymbol{C}^{R_{b}^{PVUR}} \cdot \boldsymbol{u}_{root} + D^{R_{b}^{PVUR}} \\ R_{c}^{PVUR} = V_{c}/V_P^{avg} &= \boldsymbol{A}^{R_{c}^{PVUR}} \cdot \boldsymbol{p} + \boldsymbol{B}^{R_{c}^{PVUR}} \cdot \boldsymbol{q} + \boldsymbol{C}^{R_{c}^{PVUR}} \cdot \boldsymbol{u}_{root} + D^{R_{c}^{PVUR}} \end{aligned} , \quad (9)$$

where $\boldsymbol{A}^{R_{\varphi}^{PVUR}}$, , $\boldsymbol{B}^{R_{\varphi}^{PVUR}}$ and $\boldsymbol{C}^{R_{\varphi}^{PVUR}}$ are coefficients vectors, and $D^{R_{\varphi}^{PVUR}}$ is a constant parameter. Then the *PVUR* can be calculated using equation (10)

$$PVUR_{indirect} = \max\left\{\left|R_{a}^{PVUR}-1\right|, \left|R_{b}^{PVUR}-1\right|, \left|R_{c}^{PVUR}-1\right|\right\}, \quad (10)$$

which is also a convex and piecewise affine function.

B. Three-phase Voltage Unbalance Metric Linearization based on SVR

We train the linear three-phase voltage unbalance metric models using the data-driven linear regression. Based on the data from different scenarios, the linear regression builds a relationship between the input variables and output variables by identifying the parameters of the linear functions. All the linear relationships of the proposed linear three-phase voltage unbalance metric model can be summarized as (11), where vector $\boldsymbol{x} = \left[\boldsymbol{p}^T \quad \boldsymbol{q}^T \quad \boldsymbol{u}_{root}^T\right]^T$ represents the input variable, vector $\boldsymbol{y}$=[VUF_{direct} $LVUR_{direct}$ R_{ab}^{LVUR} R_{bc}^{LVUR} R_{ca}^{LVUR} $PVUR_{direct}$ R_{a}^{PVUR} R_{b}^{PVUR} R_{c}^{PVUR} $]^T$ represents the output variable; $\boldsymbol{w}$ is the coefficient matrix; $\boldsymbol{b}$ is a constant vector. The target of data-driven linearization is to obtain the parameters $\boldsymbol{w}$ and $\boldsymbol{b}$.

$$\boldsymbol{y} = \boldsymbol{w}^T \cdot \boldsymbol{x} + \boldsymbol{b}. \quad (11)$$

The SVR method can address the issues caused by the measurement outliers and data collinearity, which are common in the training in power systems. Therefore, we use the SVR method to obtain the above linear approximation. With the input data of $\{(\boldsymbol{x}_1, \boldsymbol{y}_1), \ldots, (\boldsymbol{x}_N, \boldsymbol{y}_N)\}$, the SVR model can be obtained by solving the following optimization problem.

$$\begin{aligned} & \underset{\boldsymbol{w}, \xi_i, \xi_i^*, b}{\text{minimize}} \quad \frac{1}{2}\|\boldsymbol{w}\|^2 + \beta \sum_{i=1}^{N} \left(\xi_i + \xi_i^*\right) \\ & \text{s.t.} \begin{cases} y_i - \boldsymbol{w}^T \boldsymbol{x}_i - b_i \le \varepsilon + \xi_i \\ \boldsymbol{w}^T \boldsymbol{x}_i + b_i - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases} , \\ & |\xi|_\varepsilon = \begin{cases} 0, & \text{if } |\xi| \le \varepsilon \\ |\xi| - \varepsilon, & \text{otherwise} \end{cases} \end{aligned} \quad (12)$$

where $\boldsymbol{w}$, ξ_i, ξ_i^*, b are variables. Parameters β and ε can be determined by cross-validation.

The SVR optimization problem is a convex optimization problem, which can be solved by optimization problem solvers. When the regression process finishes, the final coefficient matrix $\boldsymbol{w}$ and the constant vector $\boldsymbol{b}$ are obtained. Thus, the linear approximations of *VUF*, *LVUR* and *PVUR* are obtained. Since the regression is an offline process, it will not impact the efficiency of the online application.

III. Optimal Dispatch Problem with Linear Three-phase Voltage Unbalance Metric

The optimal dispatch problem is essential in the power system operation analysis. To consider the three-phase unbalance in the ADNs, we apply the obtained linear three-phase voltage unbalance metric model to the optimal dispatch problems. Paper [16] proposes an optimal dispatch model that uses three-phase distribution locational marginal prices as the price signal to induce the behavior of the prosumers. We improve this optimal dispatch model by i) using the data-driven linear PF model in [20] instead of the model-based linear PF model as the power flow constraint, and ii) applying the proposed linear three-phase voltage unbalance metric models of *VUF*, *LVUR* and *PVUR* as the unbalance constraints. With this transformation, the optimization problem is converted into the convex quadratic problem, which significantly reduces the computational burden. The following sections describe the mathematical models for the objective functions and constraints.

A. Social Welfare

The distribution system can be regarded as a price taker that purchases electricity from the main grid with a constant locational marginal price π_{LMP} at the root node [21]. The cost function $C\left(p_{g,root}^{\varphi}\right)$ for the root node at phase φ is

$$C\left(p_{g,root}^{\varphi}\right)=\pi_{LMP}\cdot p_{g,root}^{\varphi}, \tag{13}$$

where $p_{g,root}^{\varphi}$ is the active power at the root node. The utility function $U_i^{\varphi}(p_{d,i}^{\varphi})$ and cost function $C_i^{\varphi}(p_{g,i}^{\varphi})$ for the prosumer and generator *i* at phase φ are

$$\begin{aligned} U_i^{\varphi}(p_{d,i}^{\varphi})&=c_{1,i}^{\varphi}\left(p_{d,i}^{\varphi}\right)^2+c_{2,i}^{\varphi}p_{d,i}^{\varphi}+c_{3,i}^{\varphi} \\ C_i^{\varphi}(p_{g,i}^{\varphi})&=c_{4,i}^{\varphi}\left(p_{g,i}^{\varphi}\right)^2+c_{5,i}^{\varphi}p_{g,i}^{\varphi}+c_{6,i}^{\varphi} \end{aligned}, \tag{14}$$

where $p_{d,i}^{\varphi}$ and $p_{g,i}^{\varphi}$ is the prosumer's active power demand and the DG's active power output, respectively; parameters $c_{1,i}^{\varphi}$ to $c_{6,i}^{\varphi}$ are coefficients. The social welfare (15) is the difference between the total utility $U(\boldsymbol{p_d})$ of prosumers and the total cost $C(\boldsymbol{p_g})$ of the system, which is a combination of the costs of DGs generation and the electricity purchased from the main grid. Vectors $\boldsymbol{p_d}$, $\boldsymbol{p_g}$ are the active power demand and generators' active power output at all nodes, respectively.

$$U(\boldsymbol{p_d})-C(\boldsymbol{p_g}) \tag{15}$$

B. System Constraints

To avoid the issues caused by the model-based linearization approach, based on our previous work, we use the following SVR-based three-phase LPF model in paper [20] as the power flow constraint.

$$\begin{aligned} &\boldsymbol{p}_b=\boldsymbol{A}^{p_b}\cdot\boldsymbol{p}+\boldsymbol{B}^{p_b}\cdot\boldsymbol{q}+\boldsymbol{C}^{p_b}\cdot\boldsymbol{u}_{root}+\boldsymbol{D}^{p_b}, \\ &\boldsymbol{q}_b=\boldsymbol{A}^{q_b}\cdot\boldsymbol{p}+\boldsymbol{B}^{q_b}\cdot\boldsymbol{q}+\boldsymbol{C}^{q_b}\cdot\boldsymbol{u}_{root}+\boldsymbol{D}^{q_b}, \\ &\boldsymbol{u}=\boldsymbol{A}^{u}\cdot\boldsymbol{p}+\boldsymbol{B}^{u}\cdot\boldsymbol{q}+\boldsymbol{C}^{u}\cdot\boldsymbol{u}_{root}+\boldsymbol{D}^{u}, \\ &\boldsymbol{p}=\boldsymbol{p_g}-\boldsymbol{p_d},\ \boldsymbol{q}=\boldsymbol{q_g}-\boldsymbol{q_d}, \end{aligned} \tag{16}$$

where $\boldsymbol{p}_b$, $\boldsymbol{q}_b$ are the vectors of three-phase branch active and reactive power flow, respectively; vector $\boldsymbol{u}$ is the square of line-to-ground voltage magnitude of all nodes except for the root node; $\boldsymbol{A}^{p_b},\boldsymbol{A}^{q_b},\boldsymbol{A}^{u},\boldsymbol{B}^{p_b},\boldsymbol{B}^{q_b},\boldsymbol{B}^{u},\boldsymbol{C}^{p_b},\boldsymbol{C}^{q_b},\boldsymbol{C}^{u}$ are the coefficient matrixes. $\boldsymbol{D}^{p_b},\boldsymbol{D}^{q_b},\boldsymbol{D}^{u}$ are constant parameters. $\boldsymbol{p}_{root}$, $\boldsymbol{q}_{root}$ are the vectors of three-phase active power and reactive power injection at the root node.

The square of line-to-ground voltage magnitude $\boldsymbol{u}$ has the lower and upper limits $\underline{\boldsymbol{u}}$ and $\overline{\boldsymbol{u}}$, while the squared line-to-ground voltage magnitude at the root node $\boldsymbol{u}_{root}$ should equal to the reference value $\boldsymbol{u}_{ref}$.

$$\underline{\boldsymbol{u}}\le\boldsymbol{u}\le\overline{\boldsymbol{u}},\ \boldsymbol{u}_{root}=\boldsymbol{u}_{ref} \tag{17}$$

The branch apparent power should not exceed the maximum apparent power capacity $s_{b,\max}$ as follows,

$$p_{b,i}^2+q_{b,i}^2\le s_{b,\max}^2. \tag{18}$$

The above constraint can be linearized using the quadratic constraint linearization method [22], giving rise to the following constraints

$$\begin{aligned} &-s_{b,\max}\le p_b\le s_{b,\max} \\ &-s_{b,\max}\le q_b\le s_{b,\max} \\ &-\sqrt{2}\cdot s_{b,\max}\le p_b+q_b\le\sqrt{2}\cdot s_{b,\max} \\ &-\sqrt{2}\cdot s_{b,\max}\le p_b-q_b\le\sqrt{2}\cdot s_{b,\max} \end{aligned} \tag{19}$$

The generators' active power output and reactive power output have lower and upper limits $\underline{\boldsymbol{p}}_{\boldsymbol{g}},\overline{\boldsymbol{p}}_{\boldsymbol{g}}$ and $\underline{\boldsymbol{q}}_{\boldsymbol{g}},\overline{\boldsymbol{q}}_{\boldsymbol{g}}$, i.e.,

$$\underline{\boldsymbol{p}}_{\boldsymbol{g}}\le\boldsymbol{p_g}\le\overline{\boldsymbol{p}}_{\boldsymbol{g}},\ \underline{\boldsymbol{q}}_{\boldsymbol{g}}\le\boldsymbol{q_g}\le\overline{\boldsymbol{q}}_{\boldsymbol{g}} \tag{20}$$

Constraint (21) is the constant power factor constraint that the proportion of the reactive power demand to active power demand is a constant η [23], and also enforces the upper and lower limits $\underline{\boldsymbol{p}}_{\boldsymbol{d}}$ and $\overline{\boldsymbol{p}}_{\boldsymbol{d}}$ of the active power demand.

$$\underline{\boldsymbol{p}}_d\le\boldsymbol{p}_d\le\overline{\boldsymbol{p}}_d,\ \eta=q_{d,i}/p_{d,i} \tag{21}$$

Equations (22) represent power balance at the root node, where L_{root} is the set of branches that connect to the root node.

$$\boldsymbol{p}_{root}=\sum_{j\in L_{root}}\boldsymbol{p}_{rootj},\ \boldsymbol{q}_{root}=\sum_{j\in L_{root}}\boldsymbol{q}_{rootj} \tag{22}$$

C. Unbalance Constraint of VUF

We apply the proposed unbalance metric approximation VUF_{direct} as an unbalance constraint with an upper bound $\bar{\delta}_{VUF}$,

$$VUF_{direct} \le \bar{\delta}_{VUF} \tag{23}$$

D. Unbalance Constraint of LVUR

For the unbalance metric *LVUR* in the unbalance constraint, according to the proposed two forms of *LVUR* linear approximations, we have the following two different forms of constraints:

1) Direct approximation

We use the *direct approximation* of *LVUR* as a constraint with the upper bound $\bar{\delta}_{LVUR}$,

$$LVUR_{direct} \le \bar{\delta}_{LVUR} \tag{24}$$

2) Indirect approximation

When we use the *indirect approximation* of *LVUR* as a constraint, the *indirect approximation* has the upper limit:

$$LVUR_{indirect} \le \bar{\delta}_{LVUR} \tag{25}$$

We define two additional variables $z_{1\varphi}^{LVUR}, z_2^{LVUR}$ to convert the above absolute value and maximum value calculation into linear forms, then (25) can be converted as follows.

$$\begin{aligned}
&z_{1ab}^{LVUR} \ge R_{ab}^{LVUR} - 1, \quad z_{1ab}^{LVUR} \ge -\left(R_{ab}^{LVUR} - 1\right), \\
&z_{1bc}^{LVUR} \ge R_{bc}^{LVUR} - 1, \quad z_{1bc}^{LVUR} \ge -\left(R_{bc}^{LVUR} - 1\right), \\
&z_{1ca}^{LVUR} \ge R_{ca}^{LVUR} - 1, \quad z_{1ca}^{LVUR} \ge -\left(R_{ca}^{LVUR} - 1\right), \\
&z_2^{LVUR} \ge z_{1ab}^{LVUR}, z_2^{LVUR} \ge z_{1bc}^{LVUR}, z_2^{LVUR} \ge z_{1ca}^{LVUR}, z_2^{LVUR} \le \bar{\delta}_{LVUR}.
\end{aligned} \tag{26}$$

E. Unbalance Constraint of PVUR

Similar to *LVUR*, for the unbalance metric *PVUR*, we build the following two forms of constraints.

1) Direct approximation

The *direct approximation* of *PVUR* should be less than the upper bound $\bar{\delta}_{PVUR}$,

$$PVUR_{direct} \le \bar{\delta}_{PVUR} \ . \tag{27}$$

2) Indirect approximation

The *indirect approximation* of *PVUR* has the upper limit:

$$PVUR_{indirect} \le \bar{\delta}_{PVUR} \tag{28}$$

We define two additional variables $z_{1\varphi}^{PVUR}$, z_2^{PVUR} to convert (28) into linear forms as follows,

$$\begin{aligned}
&z_{1a}^{PVUR} \ge R_a^{PVUR} - 1, \quad z_{1a}^{PVUR} \ge -\left(R_a^{PVUR} - 1\right), \\
&z_{1b}^{PVUR} \ge R_b^{PVUR} - 1, \quad z_{1b}^{PVUR} \ge -\left(R_b^{PVUR} - 1\right), \\
&z_{1c}^{PVUR} \ge R_c^{PVUR} - 1, \quad z_{1c}^{PVUR} \ge -\left(R_c^{PVUR} - 1\right), \\
&z_2^{PVUR} \ge z_{1a}^{PVUR}, z_2^{PVUR} \ge z_{1b}^{PVUR}, z_2^{PVUR} \ge z_{1c}^{PVUR}, z_2^{PVUR} \le \bar{\delta}_{PVUR}.
\end{aligned} \tag{29}$$

F. Optimal Dispatch Problem Formulations

The objective of the optimal dispatch problem is to maximize social welfare. To guarantee the secure operation of the system, all three types of unbalance constraints (*VUF*, *LVUR* and *PVUR*) should be satisfied. To compare the performance of the *direct* and *indirect approximation*, we build the following two optimal dispatch models.

(P1) max Social Welfare (15)

s.t. System constraints (16)-(17), (19)-(22)
direct unbalance constraints for *VUF*, *LVUR* and *PVUR*: (2), (23), (4), (24), (8), (27)

(P2) max Social Welfare (15)

s.t. System constraints (16)-(17), (19)-(22)
direct unbalance constraints for *VUF*: (2), (23)
indirect unbalance constraints for *LVUR* and *PVUR*: (5), (26), (9), (29)

IV. Numerical Test

The proposed approach is tested on a modified unbalanced three-phase IEEE 33-bus distribution system. The information about branch parameters and load profiles is available online [24]. The result of the back-forward sweep (BFS) algorithm is used as "real" values for the power flow. We first discuss how to train the linear approximation of *VUF*, *LVUR* and *PVUR*, and analyze the regression accuracy and the sensitivity. Then we apply the obtained linear three-phase voltage unbalance metric models as the unbalance constraints in the optimal dispatch model. All simulations were implemented using MATLAB on a personal laptop with an Intel Core i7-7500M 2.70-GHz processor and 16 GB of RAM.

A. Results of the Three-phase Voltage Unbalance Metric Linear Regression

We first test the linear regression accuracy of the three-phase voltage unbalance metric models. The BFS algorithm is used to calculate the power flow solutions. To generate input data for the BFS power flow, we generate active load consumption from the preset load consumption multiplied by a factor randomly drawn from a uniform distribution over the interval [0.5, 1.5]. The reactive load consumption is calculated from the active load consumption multiplied by the constant power factor η . Then, we solve the PF using BFS, and combine the input data and the PF solutions to create the input data $\{\boldsymbol{x}, \boldsymbol{y}\}$ in (11) as the training dataset. The testing dataset is generated in the same way. The linear approximations of the three-phase unbalance metrics are trained with the training dataset, and the accuracy is tested using the newly generated testing dataset. The value of parameters β and ε in (12) in the

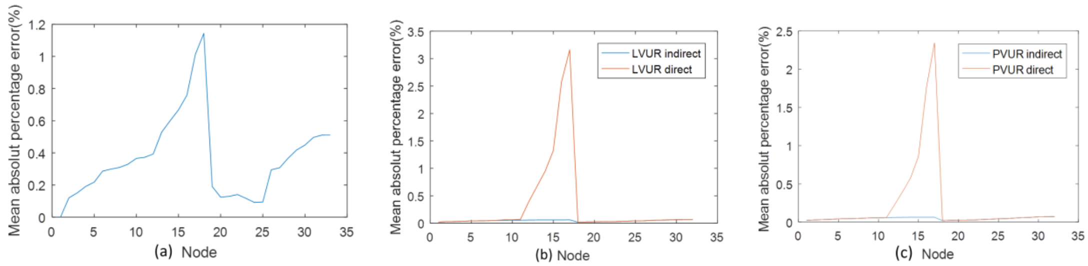

Figure 2. Mean absolute percentage error. (a) *VUF*, (b) *LVUR*, (c) *PVUR*.

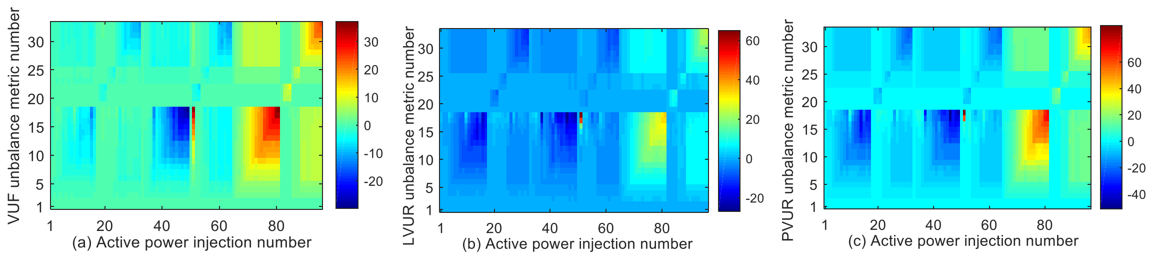

Figure 3. Sensitivity of the three-phase unbalance metrics to the power injection. (a) *VUF*, (b) *LVUR*, (c) *PVUR*

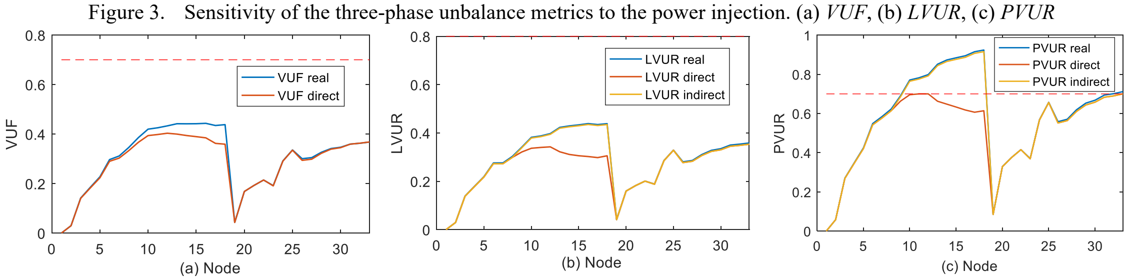

Figure 4. Unbalance metrics value of P1. (a) *VUF*, (b) *LVUR*, (c) *PVUR*

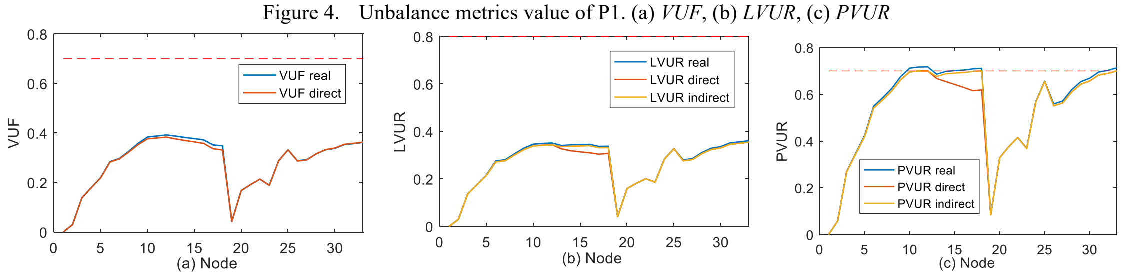

Figure 5. Unbalance metrics value of P2. (a) *VUF*, (b) *LVUR*, (c) *PVUR*

SVR approach is determined by cross-validation, respectively. The size of the training data and the testing data is 600 and 400, respectively.

1) Linear regression errors of VUF, LVUR and PVUR

The accuracy of the three-phase voltage unbalance metric linear regression is measured by the mean absolute percentage regression error of each node, which is illustrated in Fig.2. It can be observed that the regression errors for *VUF*, *LVUR* and *PVUR* are within 1.2%, 3.2%, and 2.4%, which illustrates the accuracy of the regression. Moreover, it can be observed that for *LVUR* and *PVUR*, the regression errors of the *indirect approximation* are within 0.1%, which is much lower than those of the *direct approximation*; therefore, the *indirect* linear approximation is more accurate than the *direct* one.

2) Sensitivity analysis.

Next, we analyze the impact of the power injection of each node at each phase on the three-phase unbalance metrics. The impact can be reflected by the values of the linear coefficients of the *direct* linear approximation model of *VUF*, *LVUR,* and *PVUR*. Since the reactive power injection is proportional to the active power injection under our test case, we just illustrate the sensitivity of the value of three-phase unbalance metrics to the active power injection in Fig.3. The *x*-axis is the active power injection number, among which [1, 32], [33, 64], and [65, 96] are for the nodes of phase *a*, *b* and *c*, respectively. The *y*-axis is the voltage unbalance metric number. The color represents the sensitivity, which is the value of the linear coefficients of the *direct* linear approximation model for each node with respect to the active power injection of each node at each phase. It can be observed that the color is different at different points, which illustrates that the impact of power injection of each node at each phase on the value of the three-phase unbalance metric is different. For some points, the values of the coefficient are very high or very low, which illustrates the three-phase unbalance metrics are more sensitive to the active power injection at these points. For example, we observe that active power injections in phase *c* at node 16-18 have a strong positive relationship with *VUF*, *LVUR* and *PVUR* at the same nodes (red values), whereas power injections in phase *b* have a strong negative relationship (blue values). If the system operator wants to alleviate the three-phase unbalance in the power system in practice, the point with a high absolute linear coefficient value should be paid more attention to.

B. Results of the Optimal Dispatch Problem

Next, we applied the obtained linear three-phase unbalance metrics model to the optimal dispatch problem. We assume the utility functions are different for prosumers at different phases and nodes. Node 1 is connected to the main grid and is regarded as a conventional generator with infinite capacity. Six different DGs are connected to nodes 3, 6, 12, 18, 22, and 33, respectively, and the active output limit of DGs is [0, 0.1] MW. The output of DGs is assumed phase-independent. To demonstrate the benefits of the proposed method, it is important to have a test case where the unbalance limits are binding. Therefore, the unbalance limits are set to be 0.7, 0.8 and 0.7 for *VUF*, *LVUR* and *PVUR*, respectively.

We analyze the accuracy of the *direct* and the *indirect approximation*. Fig.4 illustrates the unbalance metric values obtained from the solution of the optimal dispatch model (P1) with the *direct approximation* constraints. Fig.5 illustrates the unbalance metric value obtained from the solution of the optimal dispatch model (P2) with the *indirect approximation* constraints. It can be observed that under the two scenarios, the *PVUR* unbalance constraint is the binding constraint. Fig.4 illustrates that when we use the *direct approximation* as constraints, the real value of the *PVUR* (plotted in blue) significantly exceeds the unbalance limit of 0.7 at some nodes, even though the *direct approximation* (in red) is below the limit. When we use the *indirect approximation* as constraints in Fig.5, we observe that the real value of *PVUR* (in blue) is only a little over the limit, which illustrates that the *indirect PVUR approximation* (shown in yellow) is a more accurate representation of the unbalance constraint in this setting. Generally, from Fig.4 and Fig.5, it can be observed that the value of the *indirect approximation* of *LVUR* and *PVUR* is much closer to the real value when compared with the *direct approximation*. This illustrates that the *indirect approximation* is a much more accurate approximation of the three-phase *LVUR* and *PVUR* unbalance metrics.

C. Results of the Computational Time

We test the computational time of the *direct* and *indirect* approximations. TABLE I illustrates the computational time to calculate a *PVUR* value and to solve the optimal dispatch problem for *direct* and *indirect* approximations, respectively. It can be observed that since both the *direct* and *indirect* approximations are linear, the difference in the computational time is not very significant. Although the computation time of the *indirect approximation* is slightly larger than that of the *direct approximation*, the accuracy can be improved significantly by implementing the *indirect approximation* as demonstrated in the results in Section IV. B and Section IV. C. Therefore, the *indirect approximation* is better to calculate the unbalance metric value and to be applied in the optimal dispatch problem.

Computational Time	***PVUR calculation***	***Optimal dispatch problem***
direct	0.01544s	0.26938s
indirect	0.03727s	0.30544s

TABLE I. COMPUTATIONAL TIME

V. CONCLUSION

This paper proposes a data-driven linearization approach to analyze three-phase voltage unbalance in ADNs. To avoid the non-convexity, possible non-convergence issues and high computational burden associated with the evaluation of the non-linear three-phase voltage unbalance metrics (*VUF*, *LVUR* and *PVUR*), we propose the *direct* and *indirect* linear approximation models. The linear approximations are trained with an SVR-based three-phase voltage unbalance metric linearization approach. The resulting linear three-phase voltage unbalance metric models are incorporated in an optimal dispatch problem, which implements the PF using a linear data-driven PF model and integrates the three types of unbalance metrics constraints (*VUF*, *LVUR* and *PVUR*). We demonstrate the proposed metrics and their use in the optimization on the modified three-phase unbalanced IEEE 33-bus distribution system. The results show that the proposed linear approximations can provide accurate representations of the three-phase unbalance metrics. We also observe that the *indirect approximation* models for *PVUR* and *LVUR* are more accurate compared with the *direct approximation,* with a minimal increase in computational time. Since the proposed three-phase voltage unbalance metrics are linear, the computational complexity is reduced and the three-phase optimal dispatch problem with voltage unbalance constraints can be solved very efficiently.

REFERENCES

[1] W. Kong, K. Ma, and Q. Wu, “Three-phase power imbalance decomposition into systematic imbalance and random imbalance,” *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3001–3012, 2017.

[2] N. C. Woolley and J. V. Milanovic, "Statistical Estimation of the Source and Level of Voltage Unbalance in Distribution Networks," *IEEE Trans. Power Deliv.*, vol. 27, no. 3, pp. 1450-1460, July 2012.

[3] IEC 61000-2-2, EMC – Part 2-2: Environment – Compatibility levels for low frequency conducted disturbances and signaling in public low-voltage power supply systems, (2002).

[4] Motors and generators (ANSI/NEMA Standard MG1-1993).

[5] IEEE, Recommended practice for electric power distribution for industrial plants, IEEE Standard 141–1993, (1994), pp. 1–768, https://doi.org/10.1109/IEEESTD. 1994.121642.

[6] P. Pillay, M. Manyage, Definitions of voltage unbalance, *IEEE Power Eng. Rev*. 21 (5) (2001) 50–51.

[7] X. Su, M.A. Masoum, P.J. Wolfs, Optimal PV inverter reactive power control and real power curtailment to improve performance of unbalanced four-wire LV distribution networks, *IEEE Trans. Sustain. Energy,* 5 (3) (2014) 967–977.

[8] S. Weckx, C. Gonzalez and J. Driesen, "Reducing grid losses and voltage unbalance with PV inverters," *2014 IEEE PES General Meeting Conference & Exposition*, 2014, pp. 1-5.

[9] S. Karagiannopoulos, P. Aristidou and G. Hug, "A Centralised Control Method for Tackling Unbalances in Active Distribution Grids," *2018 Power Systems Computation Conference (PSCC)*, 2018, pp. 1-7.

[10] K. P. Schneider et al., "Analytic Considerations and Design Basis for the IEEE Distribution Test Feeders," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3181-3188, May 2018.

[11] C. G. Bajo, S. Hashemi, S. B. Kjsær, G. Yang and J. Østergaard, "Voltage unbalance mitigation in LV networks using three-phase PV systems," *2015 IEEE International Conference on Industrial Technology (ICIT)*, 2015, pp. 2875-2879.

[12] K. Girigoudar and L. A. Roald, “On the impact of different voltage unbalance metrics in distribution system optimization,” *Electric Power Systems Research*, vol. 189, p. 106656, 2020.

[13] H. Li, H. Zhu, and L. Pan, "A three-phase linear load flow solution based on loop-analysis theory for distribution system," *Int. J. Comput. Math. Elect. Electron. Eng.*, vol. 38, no. 2, pp. 703–723, 2019.

[14] H. Li, X. Yan, J. Yan, A. Zhang, and F. Zhang, "A three–phase unbalanced linear power flow solution with PV bus and ZIP load, " *IEEE Access*, vol. 7, pp. 138879–138889, 2019.

[15] K. Liu, C. Wang, W. Wang, Y. Chen, H. Wu, "Linear power flow calculation of distribution networks with distributed generation", *IEEE Access*, vol. 7, pp. 44686-44695, 2019.

[16] J. Chen, Y. Guo and W. Wu, "Optimal dispatch scheme for DSO and prosumers by implementing three-phase distribution locational marginal prices." *IET Gener, Transmis. Distrib*., Vol.4, no.11, pp. 2138 – 2146, Jun. 2020.

[17] P. Huynh, H. Zhu, Q. Chen and A. E. Elbanna, "Data-Driven Estimation of Frequency Response From Ambient Synchrophasor Measurements," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 6590-6599, Nov. 2018.

[18] Y. Liu, N. Zhang, Y. Wang, J. Yang and C. Kang, "Data-Driven Power Flow Linearization: A Regression Approach," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 2569-2580, May 2019.

[19] Y. Liu, Y. Wang, N. Zhang, D. Lu and C. Kang, "A Data-Driven Approach to Linearize Power Flow Equations Considering Measurement Noise," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 2576-2587, May 2020..

[20] J. Chen, W. Li, W. Wu, T. Zhu, Z. Wang, and C. Zhao, “Robust data-driven linearization for distribution three-phase power flow,” in *2020 IEEE Conf. Energy Internet and Energy System Integration*, pp. 1527–1532.

[21] S. de la Torre, A. J. Conejo and J. Contreras, "Simulating oligopolistic pool-based electricity markets: a multiperiod approach," *IEEE Trans. Power Syst*, vol. 18, no. 4, pp. 1547-1555, Nov. 2003.

[22] X. Chen, W. Wu and B. Zhang, "Robust Restoration Method for Active Distribution Networks," *IEEE Trans. Power Syst*., vol. 31, no. 5, pp. 4005-4015, Sept. 2016.

[23] W. Zheng, W. Wu, B. Zhang and C. Lin, "Distributed optimal residential demand response considering operational constraints of unbalanced distribution networks", *IET Gener. Transm. Distrib.*, vol. 12, no. 9, pp. 1970-1979, May 2018.

[24] Supplementary Material of Unbalanced IEEE 33-bus Distribution System for DG Capacity Assessment, Jan. 2017 [Online]. Available: https://drive.google.com/file/d/0BzBNHekT7PzcR1V4eG1NZC1EOVU/view?usp=sharing.