Hindawi
International Transactions on Electrical Energy Systems
Volume 2022, Article ID 9300522, 14 pages
https://doi.org/10.1155/2022/9300522

Research Article

Adaptive Gradient-Based Optimization Method for Parameter Identification in Power Distribution Network

Bin Li,[1] Yehai Jiang,[1] Ke Hu,[2] Xiangyi Zhou,[3] Haoran Chen,[4] Shihe Xu,[5] Hao Jiao,[6] and Jinming Chen[6]

[1]*Nanjing Institute of Technology, Nanjing 211167, Jiangsu, China*
[2]*Chongqing University of Posts and Telecommunications, Chongqing 400065, China*
[3]*Chongqing University, Chongqing 401331, China*
[4]*National University of Defense Technology, Wuhan 430000, Hubei, China*
[5]*University of Science and Technology of China, Hefei 230027, Anhui, China*
[6]*State Grid Jiangsu Electric Power Co., Ltd., Research Institute, Nanjing 210096, Jiangsu, China*

Correspondence should be addressed to Ke Hu; huke@cqupt.edu.cn

Received 10 August 2022; Revised 30 August 2022; Accepted 5 September 2022; Published 15 October 2022

Academic Editor: B. Rajanarayan Prusty

Accurate device parameters play a critical role in the calculation and analysis of power distribution networks (PDNs). However, device parameters are always affected by the operating status and influenced by manual entry. Besides, the distribution area of PDN is very wide, which brings more challenges to parameter identification work. Therefore, developing appropriate algorithms for accurately identifying PDN parameters has attracted much more attention from researchers recently. Most of the existing parameter identification algorithms are gradient-free and based on heuristic schemes. Herein, an adaptive gradient-based method is proposed for parameter identification in PDN. The analytical expressions of the gradients of the loss function with respect to the parameters are derived, and an adaptive updating scheme is utilized. By comparing the proposed method and several heuristic algorithms, it is found that the errors in both three criteria via our solution are much lower with a much smoother and more stable convergence of loss function. By further taking a linear transformation of the loss function, the method of this work significantly promotes the parameter identification performance with much lower variance in repeat experiments, indicating that the proposed method in this work achieves a more robust performance to identify PDN parameters. This work gives a practical demonstration by utilizing the gradient-based method for parameter identification of PDN.

1. Introduction

Obtaining reliable and accurate device parameters is one of the top priorities in the power distribution network (PDN) in many aspects, such as security analysis, system control, state estimation, line loss calculation, power flow calculation, protection setting, and fault analysis [1]. Nevertheless, due to the lack of in-situ measurement techniques, real-time information is hard to obtain directly for PDNs under security and stability situations. Some newly introduced PDN parameters, such as line resistance, line reactance, transformer resistance, transformer reactance, transformer conductance, and transformer electrical susceptance, are in general assumed to be static in a real situation. As a result, it usually leads to poor estimation for parameter identification in PDN [2]. Some new approaches focusing on improving numerical efficiency and error reduction have been developed in many fields of PDN, such as supervisory control and data acquisition, power management unit (PMU), power management unit, and advanced metering infrastructure. These methods can be classified into full-scale approach [3], PSOSR [4], normalized Lagrange multiplier (NLM) test [5], finite-time algorithm (FTA) [6], residual method, sensitivity analysis method, Lagrange multiplier method [7] and Heffron-Phillips method [8]. Furthermore, beneficial to the development of machine learning or deep learning techniques, some smart methods have been proposed recently, such as graph convolution network (GCN) [9], support vector machine (SVM) [10], multihead attention network [11], deep reinforcement learning [12], estimation using synchrophasor data [13], PSCAD simulation [14], and multimodal long short-term memory deep learning [15]. These methods are proved effective with simulation data. However, they are also required to assemble several specialized measuring devices in the grid.

The power flow model with identification equations in PDN provides a mathematical approach to solve the problems of lacking required data and measuring devices. It builds relationships between PDN's parameters and the data which can be obtained easily, such as active power, reaction power, and voltage. Generally, some important parameters are lacking in practice, such as transformer electrical, transformer resistance, transformer conductance, transformer electrical susceptance, line resistance, and line reactance. The ranges of these parameters are always optimized by algorithms combined with the static parameters like active power, reaction power and voltage on the low-voltage side. The calculating values of voltage on the high-voltage can be computed by power flow model with the parameters mentioned before. The residuals between calculated and true values are always used to build loss functions for optimization algorithms. Therefore, the problems of parameter identification can be easily solved by model-free methods, such as least squares (LS) [16] and Markov Chain Monte Carlo (MCMC) [17]. Another kind of methods called heuristic algorithms can also be utilized to identify PDN's parameters, such as particle swarm optimization (PSO), ant colony, and genetic algorithms (GA) combined with transient measurements [18]. Some algorithms based on global optimization have been published lately for parameter identification in the field of PDN calculation [19], including random search approach (RS) [20], tree-structured Parzen estimator approach (TPE) [21] and simulated annealing (SA) [22], which have shown satisfactory performance and provided novel ideas for PDN analysis. Moreover, some methods based on machine learning and deep learning have been published for parameter identification, such as support vector regression [23], and convolutional neural network (CNN) [24]. These methods are effective for simulation data with the precise values of voltage and voltage phase angle. Nevertheless, only part of the parameters can be obtained.

The methods mentioned above for parameter identification can be generally classified as gradient-free methods. In addition, the searching methods in these gradient-free algorithms are largely dependent on the initialization. For example, in the genetic algorithm, to evaluate the designed fitness function, it is first needed to initialize the population via the encoding of the individual. The initialization is usually randomly generated, and different initializations can lead to results with large differences. To tackle these problems in parameter identification, it is crucial to develop more robust and stable numerical methods. In this work, based on the analysis of the physical model in PDN, we derive an analytical theoretical model based on the gradients of the loss function with respect to the parameters. Based on the abovementioned points, this paper mainly has the following contributions:

(1) The analytical expressions of the gradients of the loss function with respect to the parameters to be optimized in PDN are derived in detail, which are rarely studied and neglected by other investigations.

(2) The adaptive scheme for parameter updating based on the loss function gradients is utilized in this work, and it is found that the error during the numerical experiments is remarkably reduced compared with several heuristic algorithms. In addition, the loss function decay during the optimization is much smoother and more stable compared with other algorithms.

(3) The variance in the numerical calculation is much smaller than that of several heuristic algorithms, indicating that the methods proposed in this work are more robust in numerical performance.

This paper is organized as follows: Section 2 introduces the identification equations of the power flow model in PDN and proposes the gradient-based optimization method. The experimental data and calculation details are given in Section 3. The results and relevant discussions are given in Section 4. Finally, Section 5 gives a brief conclusion.

2. Materials and Methods

2.1. Power Flow Model Calculation. The basic theory of analysis in PDN can be found in ref [17, 25]. To simplify the computation, the three-phase is assumed to be balanced as the premise for calculating the power flow in this work. The schematic diagram of the power-flow calculation circuit model is shown in Figure 1.

In Figure 1, P_d, Q_d and U_d represent the active power, reaction power, and voltage on the high-voltage side of the transformers at bus D, respectively. These three parameters can be obtained directly by real-time measurements. Other parameters, such as transformer electrical R_d, transformer resistance X_d, transformer conductance G_d, transformer electrical susceptance B_d, line resistance R_{cd} and line reactance X_{cd}, are in general hard to detect in PDN calculation, and satisfy equations (1)–(3):

$$P_d = P_{Ld} + \frac{P_{Ld}^2 + Q_{Ld}^2}{U_{Ld}^2} R_d^T + U_{Ld}^2 G_d^2, \quad (1)$$

$$Q_d = Q_{Ld} + \frac{P_{Ld}^2 + Q_{Ld}^2}{U_{Ld}^2} X_d^T + U_{Ld}^2 B_d^2, \quad (2)$$

$$U_d = \sqrt{\left(U_{Ld} + \Delta U_d^T\right)^2 + \left(\delta U_d^T\right)^2}, \quad (3)$$

where ΔU_d^T and δU_d^T in equation (3) are the longitudinal and transverse components of the transformer impedance voltage drop at bus D in V. P_{Ld}, Q_{Ld} and U_{Ld} represent the active power, reaction power, and voltage on the low-voltage

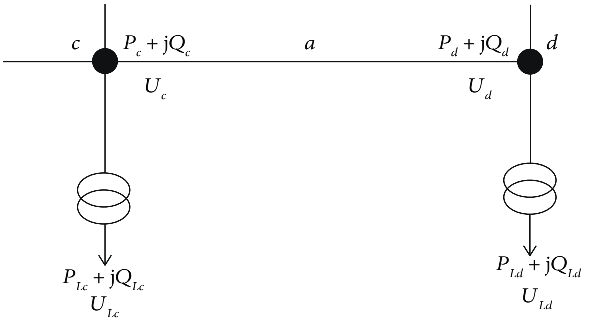


Figure 1: Power flow calculation circuit model.

side of the transformers at bus D, respectively. ΔU_d^T and δU_d^T can be expressed by equations (4) and (5), respectively.

$$\Delta U_d^T = \frac{P_{Ld} R_d^T + Q_{Ld} X_d^T}{U_{Ld}}, \tag{4}$$

$$\delta U_d^T = \frac{P_{Ld} X_d^T - Q_{Ld} R_d^T}{U_{Ld}}. \tag{5}$$

The equation of bus C can be expressed as equations (6)~(8):

$$U_c = \sqrt{\left(U_d + \Delta U_{cd}^T\right)^2 + \left(\delta U_{cd}^T\right)^2}, \tag{6}$$

$$\Delta U_{cd}^T = \frac{P_d R_{cd}^T + Q_d X_{cd}^T}{U_{Ld}}, \tag{7}$$

$$\delta U_{cd}^T = \frac{P_d X_{cd}^T - Q_d R_{cd}^T}{U_d}, \tag{8}$$

where ΔU_{cd}^T and δU_{cd}^T in equation (6) are the longitudinal and transverse components of the transformer impedance voltage drop at bus C in V. Once we have the above quantities, the final result is calculated by the following equation.

$$f_c = \sqrt{\left(U_d + \Delta U_{cd}^T\right)^2 + \left(\delta U_{cd}^T\right)^2}. \tag{9}$$

The parameters in the line and transformer can be calculated based on equations (1)~(9) with the measured power and voltage data.

2.2. Gradient-Based Algorithm for Parameter Identification in PDN.

2.2. Gradient-Based Algorithm for Parameter Identification in PDN. To utilize the gradient-based algorithms for PDN calculations, first the surrogate function (or, namely, the objective function) should be defined. As mentioned above, the inputs of PDN can be defined as: $\mathbf{X} = (P_{Ld,i}, Q_{Ld,i}, U_{Ld,i})$ representing the set of the active power, reaction power, and voltage on the high-voltage side of the transformers, and the subscript i represents the sample points for $i = (1, 2, \ldots, N)$. While other parameters, namely, $R_d, X_d, G_d, B_d, R_{cd}$ and X_{cd} are seen as the initialized parameters, which are needed to be optimized. Herein, we consider a linear regression of the target value $\mathbf{Y}$ with the combination of the inputs as follows:

$$\mathbf{Y} = \theta_1 \mathbf{X} + \theta_0 + \eta, \tag{10}$$

where θ_0 represents the bias term and η represents the noise term. In this work, we assume that the noise term is subject to the distribution where the mean value and variance are 0 and 1, respectively, namely, $\eta \sim N(0, 1)$. Therefore, the mean value and variance of the target value $\mathbf{Y}$ are

$$E[\mathbf{Y}] = E[\theta_1 \mathbf{X} + \theta_0 + \eta] = \theta_1 \mathbf{X} + \theta_0, \tag{11}$$

$$\mathrm{Var}[\mathbf{Y}] = \mathrm{Var}[\theta_1 \mathbf{X} + \theta_0 + \eta] = 1. \tag{12}$$

Therefore, the target values are subject to the distribution where its mean value is $\theta_1 \mathbf{X} + \theta_0$ and the variance is 1. Then the probability of the sample point (x_i, y_i) is

$$p(y_i|x_i) = \exp\left[-\frac{(y_i - (\theta_1 x_i + \theta_0))^2}{2}\right]. \tag{13}$$

The likelihood probability of the whole sample sets is

$$L(\mathbf{X}, \mathbf{Y}) = \prod_{i=1}^{N} p(y_i|x_i) = \prod_{i=1}^{N} \exp\left[-\frac{(y_i - (\theta_1 x_i + \theta_0))^2}{2}\right]. \tag{14}$$

Take the logarithm of the likelihood function above and then, we obtain

$$l(\mathbf{X}, \mathbf{Y}) = \log L(\mathbf{X}, \mathbf{Y}) = -\frac{1}{2} \sum_{i=1}^{N} [y_i - (\theta_1 x_i + \theta_0)]^2. \tag{15}$$

Extend the above logarithm of the likelihood function, and the loss function in this work is designed as

$$\mathrm{loss} = \frac{1}{2N} \sum_{i=1}^{N} \left[y_i - \widehat{y}_i\right]^2 = \frac{1}{2N} \sum_{i=1}^{N} \left[\left(\theta_{1,i} f_{c,i} + \theta_{0,i}\right) - \widehat{U}_{c,i}\right]^2, \tag{16}$$

where y_i and $\widehat{y}_i$ represents the calculation value and true value, respectively. Herein, we also introduce a linear transform with parameter $\theta_{1,i}$ and $\theta_{0,i}, \widehat{U}_{c,i}$ represents the true value of voltage at the on the high-voltage side.

2.2.1. Adaptive Gradient-Based Optimization Methods. Up to now, we have known that the optimization mission in this work is to minimize the loss function mentioned above with respect to the following six parameters, namely, $R_d, X_d, G_d, B_d, R_{cd}$ and X_{cd}:

$$\text{minimize}\,\mathrm{loss}\left(\theta \in \mathbb{R}^d\right) = \mathrm{loss}(R_d, X_d, G_d, B_d, R_{cd}, X_{cd}). \tag{17}$$

To utilize the gradient-based optimization methods, it is generally needed to calculate the gradient of the loss function with respect to the parameters, namely, $g_t = \nabla_\theta \mathrm{loss}(\theta)$. In addition, the historical information during the optimization in each step can be utilized to accelerate the optimization, obtaining the first-order and second-order moments as follows:

$$m_t = \phi(g_1, g_2, \ldots, g_t), \tag{18}$$

$$v_t = \psi(g_1, g_2, \ldots, g_t). \tag{19}$$

Then, the parameters are updated with the following rule:

$$\theta_{t+1} = \theta_t - \frac{1}{\sqrt{v_t + \varepsilon}} m_t, \tag{20}$$

where ε represents a smooth term to prevent the denominator equaling to zero. In equation (20), if we ignore the second-order moment v_t and set it as $v_t = I^2, m_t = \eta g_t$ (where η is often known as the learning rate), then the optimization scheme is recovered as the stochastic gradient descent approach. The convergence of the stochastic gradient descent approach is usually slow and it is easy to oscillate at the saddle point. In addition, different parameters should adopt different updating schemes, which means some parameters may not update so frequently during the optimization, thus, it is expected to utilize a relatively large step to speed up the convergence. Oppositely, for those parameters updating frequently, generally it is needed to take a small step for achieving a more stable optimization process. Based on these analyses, the following update scheme is utilized:

$$v_t = \mathrm{diag}\left(\sum_{i=1}^{t} g_{i,1}^2, \sum_{i=1}^{t} g_{i,2}^2, \ldots, \sum_{i=1}^{t} g_{i,d}^2\right), \tag{21}$$

where $v_t \in \mathbb{R}^{d\times d}$ is a diagonal matrix. Equation (21) indicates that for frequently updating parameters before the moment t, its second-order momentum is relatively large and results in a small learning rate. The learning rate in the gradient-based optimization method is also called the hyper-parameter, which can be dynamically changed with an appropriate schedule. To test the influence of hyper-parameters such as the learning rate on the performance of parameter identification in PDN. To utilize the gradient-based method to optimize six parameters, namely, $R_d, X_d, G_d, B_d, R_{cd}$ and X_{cd}, the gradient of the loss function with respect to these parameters needs to be derived in the next section.

2.2.2. Gradients of the Loss Function Based on Chain-Rule. Once we have the loss function, the next step is to calculate the gradients of the loss function with respect to the six parameters to be optimized. With the help of the chain-rule, the gradients can be derived. First, $\partial\mathrm{loss}/\partial R_d$ and $\partial\mathrm{loss}/\partial X_d$ are given as follows:

$$\frac{\partial \mathrm{loss}}{\partial R_d} = \frac{\partial \mathrm{loss}}{\partial f_c} \cdot \frac{\partial f_c}{\partial U_d} \cdot \frac{\partial U_d}{\partial \Delta U_d^T} \cdot \frac{\partial \Delta U_d^T}{\partial R_d} + \frac{\partial \mathrm{loss}}{\partial f_c} \cdot \frac{\partial f_c}{\partial U_d} \cdot \frac{\partial U_d}{\partial \delta U_d^T} \cdot \frac{\partial \delta U_d^T}{\partial R_d}, \tag{22}$$

$$\frac{\partial \mathrm{loss}}{\partial X_d} = \frac{\partial \mathrm{loss}}{\partial f_c} \cdot \frac{\partial f_c}{\partial U_d} \cdot \frac{\partial U_d}{\partial \Delta U_d^T} \cdot \frac{\partial \Delta U_d^T}{\partial X_d} + \frac{\partial \mathrm{loss}}{\partial f_c} \cdot \frac{\partial f_c}{\partial U_d} \cdot \frac{\partial U_d}{\partial \delta U_d^T} \cdot \frac{\partial \delta U_d^T}{\partial X_d}, \tag{23}$$

where the derivatives $(\partial\mathrm{loss}/\partial f_c)\cdot(\partial f_c/\partial U_d)$ is shown in the both above two equations, and it is given as follows:

$$\frac{\sum_{i=1}^{N} \theta_{1,i}\left[\left(\theta_{1,i} f_{c,i} + \theta_{0,i}\right) - \widehat{U}_{c,i}\right]}{N} \cdot \frac{\left(U_{d,i} + \Delta U_{cd,i}^T\right)}{\sqrt{\left(U_{d,i} + \Delta U_{cd,i}^T\right)^2 + \left(\delta U_{cd,i}^T\right)^2}}. \tag{24}$$

Also, the derivatives $(\partial U_d/\partial\Delta U_d^T)\cdot(\partial\Delta U_d^T/\partial R_d)$ and $(\partial U_d/\partial\delta U_d^T)\cdot(\partial\delta U_d^T/\partial R_d)$ are derived as

$$\frac{\partial U_d}{\partial \Delta U_d^T} \cdot \frac{\partial \Delta U_d^T}{\partial R_d} = \frac{U_{Ld,i} + \Delta U_{d,i}^T}{\sqrt{\left(U_{Ld,i} + \Delta U_{d,i}^T\right)^2 + \left(\delta U_{d,i}^T\right)^2}} \cdot \frac{P_{Ld,i}}{U_{Ld,i}}, \tag{25}$$

$$\frac{\partial U_d}{\partial \delta U_d^T} \cdot \frac{\partial \delta U_d^T}{\partial R_d} = \frac{-\delta U_{d,i}^T}{\sqrt{\left(U_{Ld,i} + \Delta U_{d,i}^T\right)^2 + \left(\delta U_{d,i}^T\right)^2}} \cdot \frac{Q_{Ld,i}}{U_{Ld,i}}. \tag{26}$$

Substituting (19)–(21) into (17) and (18), then we obtain

$$\begin{aligned}\frac{\partial \text{loss}}{\partial R_d} = {} & \frac{\sum_{i=1}^{N} \theta_{1,i}\left[\left(\theta_{1,i} f_{c,i} + \theta_{0,i}\right) - \widehat{U}_{c,i}\right]}{N} \cdot \frac{\left(U_{d,i} + \Delta U_{cd,i}^{T}\right)}{\sqrt{\left(U_{d,i} + \Delta U_{cd,i}^{T}\right)^{2} + \left(\delta U_{cd,i}^{T}\right)^{2}}} \\ & \cdot \frac{\delta U_{d,i}^{T}}{\sqrt{\left(U_{Ld,i} + \Delta U_{d,i}^{T}\right)^{2} + \left(\delta U_{d,i}^{T}\right)^{2}}} \frac{P_{Ld,i}\left(U_{Ld,i} + \Delta U_{d,i}^{T}\right) - Q_{Ld,i}\delta U_{d,i}^{T}}{U_{Ld,i}},\end{aligned} \tag{27}$$

$$\begin{aligned}\frac{\partial \text{loss}}{\partial X_d} = {} & \frac{\sum_{i=1}^{N} \theta_{1,i}\left[\left(\theta_{1,i} f_{c,i} + \theta_{0,i}\right) - \widehat{U}_{c,i}\right]}{N} \cdot \frac{\left(U_{d,i} + \Delta U_{cd,i}^{T}\right)}{\sqrt{\left(U_{d,i} + \Delta U_{cd,i}^{T}\right)^{2} + \left(\delta U_{cd,i}^{T}\right)^{2}}} \\ & \cdot \frac{\delta U_{d,i}^{T}}{\sqrt{\left(U_{Ld,i} + \Delta U_{d,i}^{T}\right)^{2} + \left(\delta U_{d,i}^{T}\right)^{2}}} \frac{P_{Ld,i}\left(U_{Ld,i} + \Delta U_{d,i}^{T}\right) + Q_{Ld,i}\delta U_{d,i}^{T}}{U_{Ld,i}}.\end{aligned} \tag{28}$$

Similar to the derivation of (22) and (23), the gradient of the loss function with respect to the other parameters is

$$\begin{aligned}\frac{\partial \text{loss}}{\partial G_d} & = \frac{\partial \text{loss}}{\partial f_c} \cdot \frac{\partial f_c}{\partial \Delta U_{cd}^{T}} \cdot \frac{\partial \Delta U_{cd}^{T}}{\partial P_d} \cdot \frac{\partial P_d}{\partial G_d} + \frac{\partial \text{loss}}{\partial f_c} \cdot \frac{\partial f_c}{\partial \delta U_{cd}^{T}} \cdot \frac{\partial \delta U_{cd}^{T}}{\partial P_d} \cdot \frac{\partial P_d}{\partial G_d} \\ & = \frac{\sum_{i=1}^{N} \theta_{1,i}\left[\left(\theta_{1,i} f_{c,i} + \theta_{0,i}\right) - \widehat{U}_{c,i}\right]}{N} \cdot \frac{2U_{Ld,i}^{2} G_{d,i}}{\sqrt{\left(U_{d,i} + \Delta U_{cd,i}^{T}\right)^{2} + \left(\delta U_{cd,i}^{T}\right)^{2}}} \cdot \left[\left(U_{d,i} + \Delta U_{cd,i}^{T}\right) \cdot \frac{R_{cd,i}^{T}}{U_{Ld,i}} + \delta U_{cd,i}^{T} \cdot \frac{X_{cd,i}^{T}}{U_{d,i}}\right],\end{aligned} \tag{29}$$

$$\begin{aligned}\frac{\partial \text{loss}}{\partial B_d} & = \frac{\partial \text{loss}}{\partial f_c} \cdot \frac{\partial f_c}{\partial \Delta U_{cd}^{T}} \cdot \frac{\partial \Delta U_{cd}^{T}}{\partial Q_d} \cdot \frac{\partial Q_d}{\partial B_d} + \frac{\partial \text{loss}}{\partial f_c} \cdot \frac{\partial f_c}{\partial \delta U_{cd}^{T}} \cdot \frac{\partial \delta U_{cd}^{T}}{\partial Q_d} \cdot \frac{\partial Q_d}{\partial B_d} \\ & = \frac{\sum_{i=1}^{N} \theta_{1,i}\left[\left(\theta_{1,i} f_{c,i} + \theta_{0,i}\right) - \widehat{U}_{c,i}\right]}{N} \cdot \frac{2U_{Ld,i}^{2} B_{d,i}}{\sqrt{\left(U_{d,i} + \Delta U_{cd,i}^{T}\right)^{2} + \left(\delta U_{cd,i}^{T}\right)^{2}}} \cdot \left[\left(U_{d,i} + \Delta U_{cd,i}^{T}\right) \cdot \frac{X_{cd,i}^{T}}{U_{Ld,i}} - \delta U_{cd,i}^{T} \cdot \frac{R_{cd,i}^{T}}{U_{d,i}}\right],\end{aligned} \tag{30}$$

$$\begin{aligned}\frac{\partial \text{loss}}{\partial R_{cd}} & = \frac{\partial \text{loss}}{\partial f_c} \cdot \frac{\partial f_c}{\partial \Delta U_{cd}^{T}} \cdot \frac{\partial \Delta U_{cd}^{T}}{\partial R_{cd}} + \frac{\partial \text{loss}}{\partial f_c} \cdot \frac{\partial f_c}{\partial \delta U_{cd}^{T}} \cdot \frac{\partial \delta U_{cd}^{T}}{\partial R_{cd}} \\ & = \frac{1}{N} \sum_{i=1}^{N} \theta_{1,i}\left[\left(\theta_{1,i} f_{c,i} + \theta_{0,i}\right) - \widehat{U}_{c,i}\right] \cdot \frac{\left(U_{d,i} + \Delta U_{cd,i}^{T}\right) \cdot \left(P_{d,i}/U_{Ld,i}\right) - \left(\delta U_{cd,i}^{T}\right) \cdot \left(Q_{d,i}/U_{d,i}\right)}{\sqrt{\left(U_{d,i} + \Delta U_{cd,i}^{T}\right)^{2} + \left(\delta U_{cd,i}^{T}\right)^{2}}},\end{aligned} \tag{31}$$

$$\begin{aligned}\frac{\partial \text{loss}}{\partial X_{cd}} & = \frac{\partial \text{loss}}{\partial f_c} \cdot \frac{\partial f_c}{\partial \Delta U_{cd}^{T}} \cdot \frac{\partial \Delta U_{cd}^{T}}{\partial X_{cd}} + \frac{\partial \text{loss}}{\partial f_c} \cdot \frac{\partial f_c}{\partial \delta U_{cd}^{T}} \cdot \frac{\partial \delta U_{cd}^{T}}{\partial X_{cd}} \\ & = \frac{1}{N} \sum_{i=1}^{N} \theta_{1,i}\left[\left(\theta_{1,i} f_{c,i} + \theta_{0,i}\right) - \widehat{U}_{c,i}\right] \cdot \frac{\left(U_{d,i} + \Delta U_{cd,i}^{T}\right) \cdot \left(Q_{d,i}/U_{Ld,i}\right) - \left(\delta U_{cd,i}^{T}\right) \cdot \left(P_{d,i}/U_{d,i}\right)}{\sqrt{\left(U_{d,i} + \Delta U_{cd,i}^{T}\right)^{2} + \left(\delta U_{cd,i}^{T}\right)^{2}}}.\end{aligned} \tag{32}$$

Now, from (17)–(27), we have the six gradients of the loss function with respect to the parameters to be optimized.

Once we have the above gradients of the loss function with respect to the parameters, then the gradient-based optimization can be implemented. The pseudo-code of the optimization method of this work is shown in Algorithm 1:

2.3. Evaluation Functions of the Algorithm. The underlying three functions are employed to estimate the performance of the algorithm proposed in this work:

(1) Mean absolute error (MAE):

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \widehat{y}_i\right|. \tag{33}$$

(2) Root mean square error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \widehat{y}_i\right)^2}. \tag{34}$$

(3) Mean absolute percentage error (MAPE):

$$\text{MAPE} = \frac{100\%}{N}\sum_{i=1}^{N}\left|\frac{y_i - \widehat{y}_i}{y_i}\right|, \tag{35}$$

where y_i and $\widehat{y}_i$ represent the ground true value and prediction value, respectively.

3. Data and Calculation Details

3.1. Raw Dataset Description. In this paper, 1499 samples in the raw dataset were collected by SCADA with a sampling period of 15 minutes [25, 26]. The three-phase first section voltages on the high-voltage side (denoted as U_a, U_b and U_c) are shown in Figure 2, and the low-voltage sides of them (denoted as u_a, u_b, and u_c) are displayed in Figure 3.

It can be found in Figures 2 and 3 that the high-voltage sides in the dataset are closed to the three-phase balance, and this dataset satisfies the requirements of the equations in Section 2.1. In addition, the active power (denoted as P_a, P_b, and P_c) and reactive power (denoted as Q_a, Q_b, and Q_c) of three-phase on the low voltage side are shown in Figures 4 and 5, respectively.

The trend of changes of active power and reactive power on the low voltage side is consistent in Figures 4 and 5, and it indicates that the samples in this dataset are stable and can be used to perform parameter identification.

3.2. Evaluation and Calculation. In this paper, 75% of the samples (1124) are split randomly as a training set to identify PDN's parameters. The best parameters are used to calculate voltage per unit in C bus (denoted as U_{cal}) by a power flow model. After that, the rest of the 25% samples (375) are used to evaluate the performance of parameter identification as a test set through the three metrics as equations (28)–(30). Instead of directly calculating these metrics, the linear regression should be applied in this paper, the values of U_c and U_{cal} are regarded as dependent variable and independent variable, respectively. The output values of linear regression are signed as U^*_{cal}, and the final evaluations of parameter identification are gained between U_c and U^*_{cal} as equation (31):

$$U^*_{\text{cal}} = aU_{\text{cal}} + b, \tag{36}$$

where a and b are denoted as the slope and bias of linear regression, and they can also be optimized like other parameters by MCMC and GA. In the following discussion, the

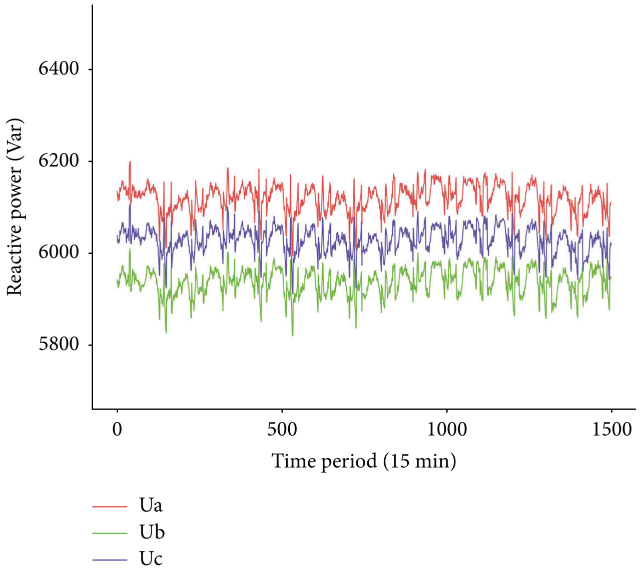


FIGURE 2: U_a, U_b, and U_c on the high-voltage side.

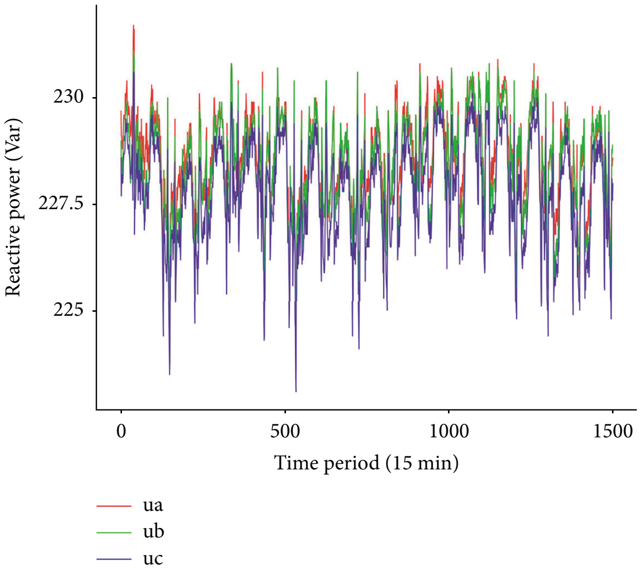


FIGURE 3: u_a, u_b, and u_c on the high-voltage side.

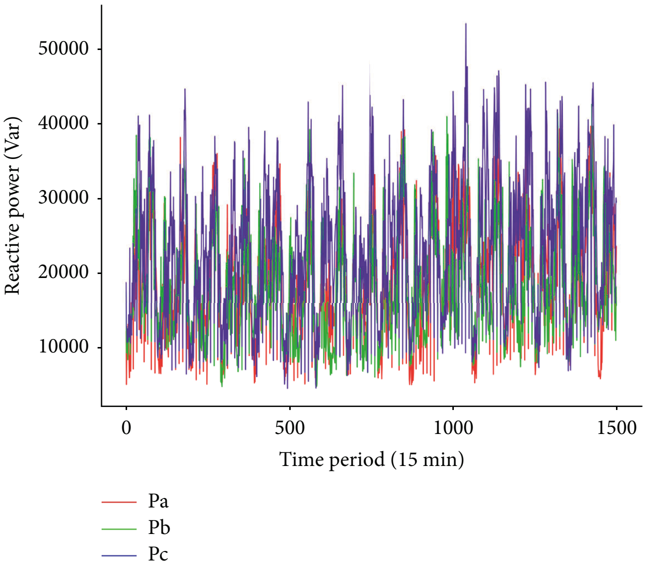


Figure 4: P_a, P_b, and P_c on the low voltage side.

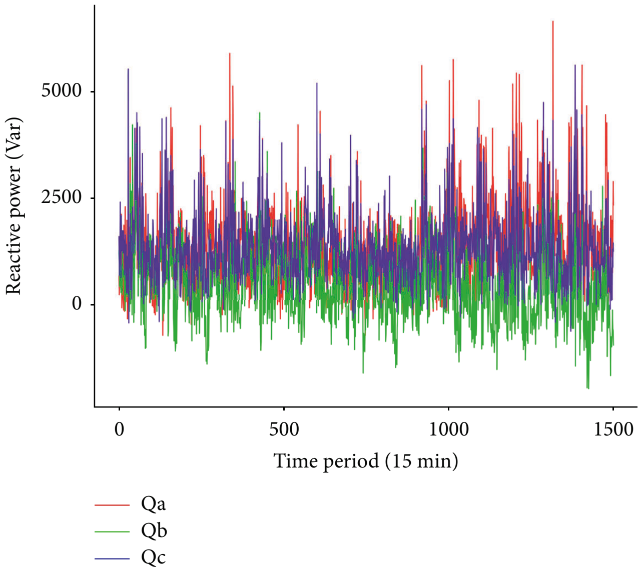


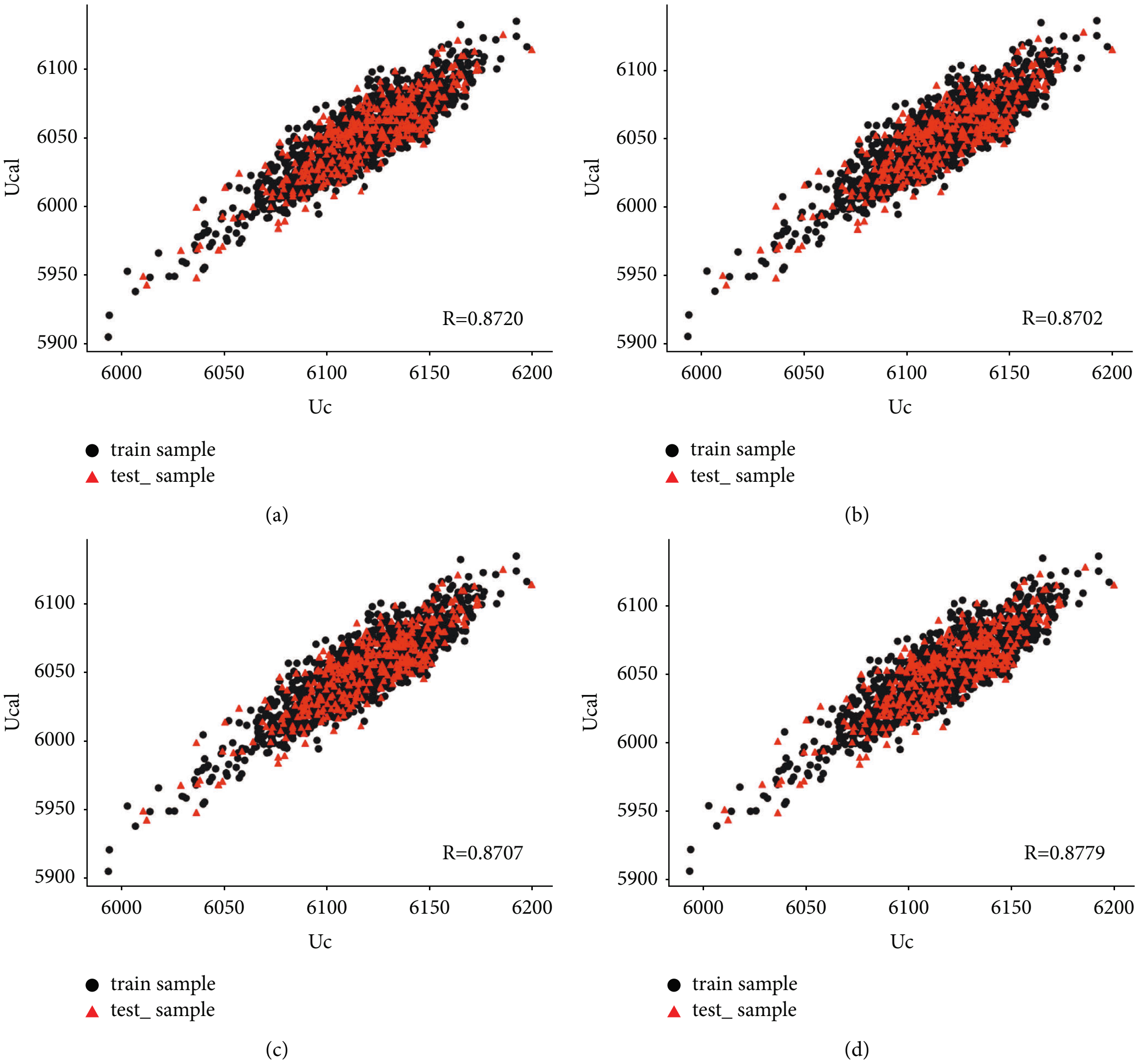
Figure 5: Q_a, Q_b, and Q_c on the low voltage side.

```
**Input**: $\theta_0$ initial parameters
$f(\theta)$ objective function to be optimized
$\beta_1, \beta_2$ decay rates for moment estimates
$m_0$ initialized first-order moment
$v_0$ initialized second-order moment
$t$ time step
$\eta$ learning rate
**While** $\theta_t$ is not converged **do**
$t \leftarrow t + 1$
$g_t \leftarrow \nabla_\theta f_t(\theta_{t-1})$ gradient w.r.t parameters at time step $t$
$m_t \leftarrow \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$ update first-order moment
$v_t \leftarrow \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2$ update second-order moment
$\widehat{m}_t \leftarrow m_t / (1 - \beta_1^t)$ biased-corrected first-order moment
$\widehat{v}_t \leftarrow v_t / (1 - \beta_2^t)$ biased-corrected second-order moment
$\theta_t \leftarrow \theta_{t-1} - \eta \widehat{m}_t / (\sqrt{\widehat{v}_t} + \varepsilon)$
End while
Return $\theta_t$
```

Algorithm 1: Adaptive gradient-based optimization methods.

Table 1: The upper and lower bounds of MCMC and GA.

Parameter name	Abbreviation	Upper bound	Lower bound	Unit
Line resistance	R_{cd}	0.5	0.005	Ω/km
Line reactance	X_{cd}	0.5	0.005	Ω/km
Transformer resistance	X_d	20	5	Ω/km
Transformer reactance	R_d	10	0.8	Ω/km
Transformer conductance	G_d	8e-6	4e-6	S
Transformer electrical susceptance	B_d	8e-5	2e-5	S
Slope of linear regression	a	−5	5	∗
Bias of linear regression	b	500	1000	∗



(a)

(b)

(c)

(d)

Figure 6: The scatter plot of U_c and U_{cal} by RS (a), TPE (b), SA (c), and AGBO (d).

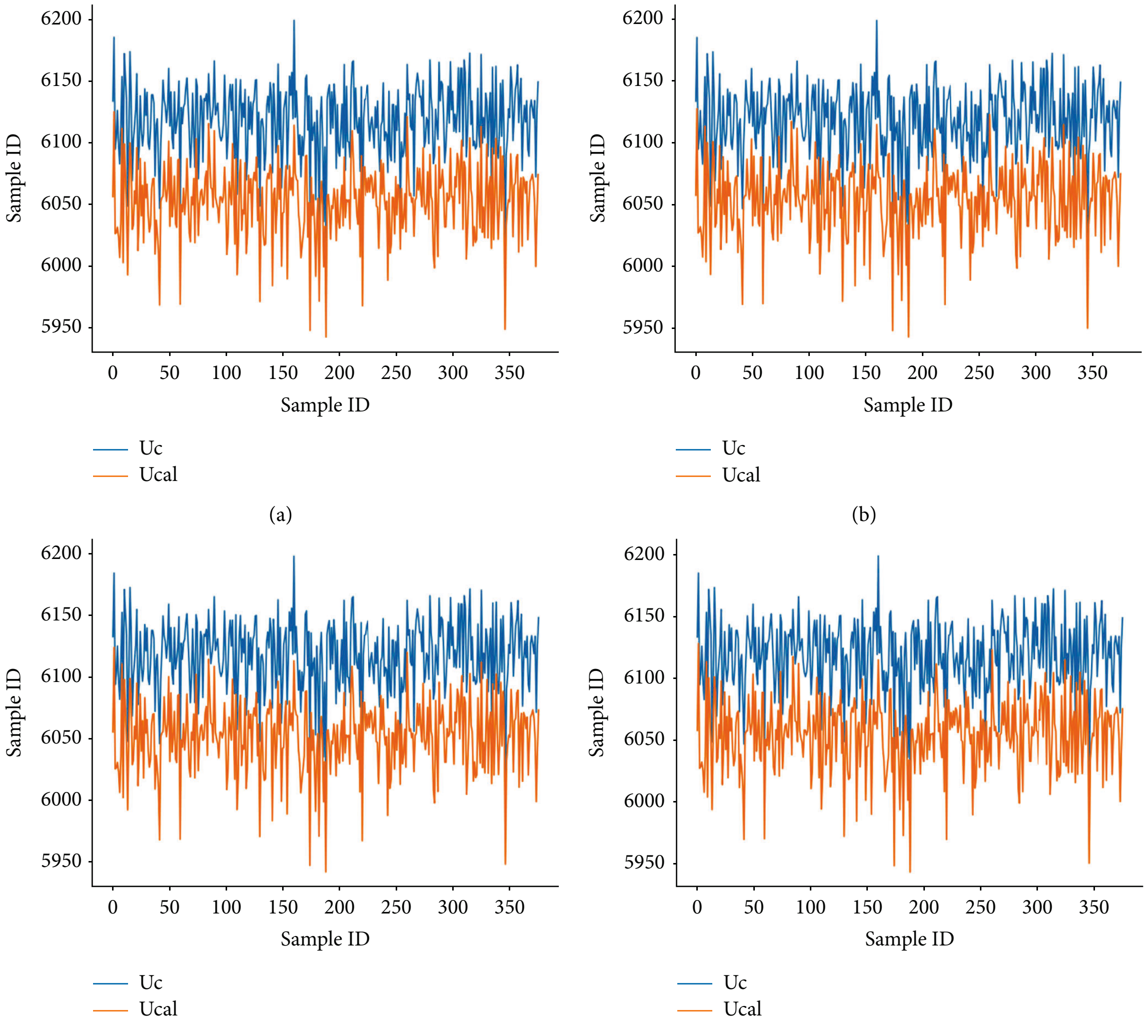
parameters of linear regression optimized by MCMC and GA are signed as MCMC-LR and GA-LR, respectively. The upper bounds and lower bounds of MCMC and GA should be determined before parameter identification, and they are listed in Table 1:

To avoid the impact of randomness on GA and MCMC, each method is repeated 25 times to guarantee the correctness and stability of results.

4. Results and Discussion

The parameter identification results of AGBO and three SMBO algorithms based on mean square error between U_c and U_{cal} are discussed first in this section. The prior weight and number of started jobs are set as 1 and 20 for the TPE method. The rate of reduction in simulated annealing is 0.1 as a default value [25]. The learning rate and weight decay are 5e-4 and 0 in AGBO. The maximum iteration step is 1000. Their prediction performances are shown in Table 2, Figures 6 and 7, respectively.

It can be found in Table 2 that AGBO has the lowest values in MAE, RMSE, and MAPE compared with the results of RS, TPE, and SA. However, their results only have minor differences, and Figures 6 and 7 also show that the prediction results by these algorithms do not have remarkable differences. The reason can be attributed to the fact that calculating mean square error of U_c and U_{cal} neglecting the statistical relationship between them and has a negative impact on parameter identification. Therefore, based on the previous study [26], the line



(a)

(b)

(c)

(d)

Figure 7: The line plot of U_c and U_{cal} by RS (a), TPE (b), SA (c),and AGBO (d).

Table 2: The results of parameter identification with the loss function of mean square error.

Algorithm	MAE	RMSE	MAPE
RS	65.950	67.500	1.078
TPE	64.648	66.340	1.056
SA	66.003	67.550	1.078
AGBO*	64.384	66.091	1.052

transformation can be implemented to U_{cal} before calculating the loss function. The parameters of line transformation are optimized with other PDN parameters simultaneously, and their results are displayed in Table 3, Figures 8 and 9.

All algorithms perform better in Table 3 than the results in Table 2, which indicates the statistical property between U_c and U_{cal}, such as linear relationship, has contributed to identifying PDN's parameters. The prediction errors in Table 3 are almost one-tenth of those in Table 2. Moreover, it can be found that AGBO-LR has a significantly better prediction performance evaluated by all of the three metrics compared with RS-LR, TPE-LR, and SA-LR in Table 3, and the scatter points are closer to a line of 45 degrees than the

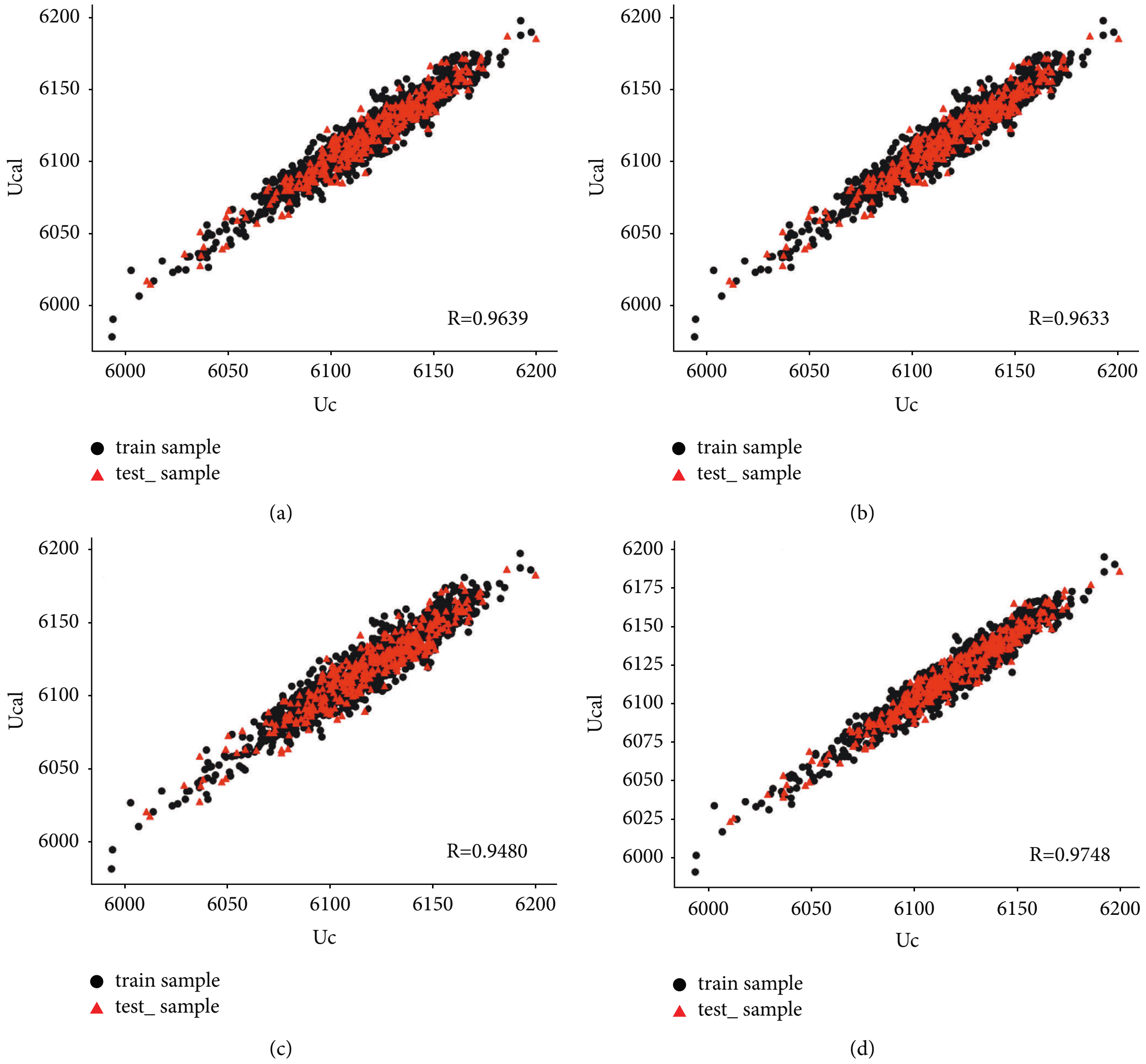


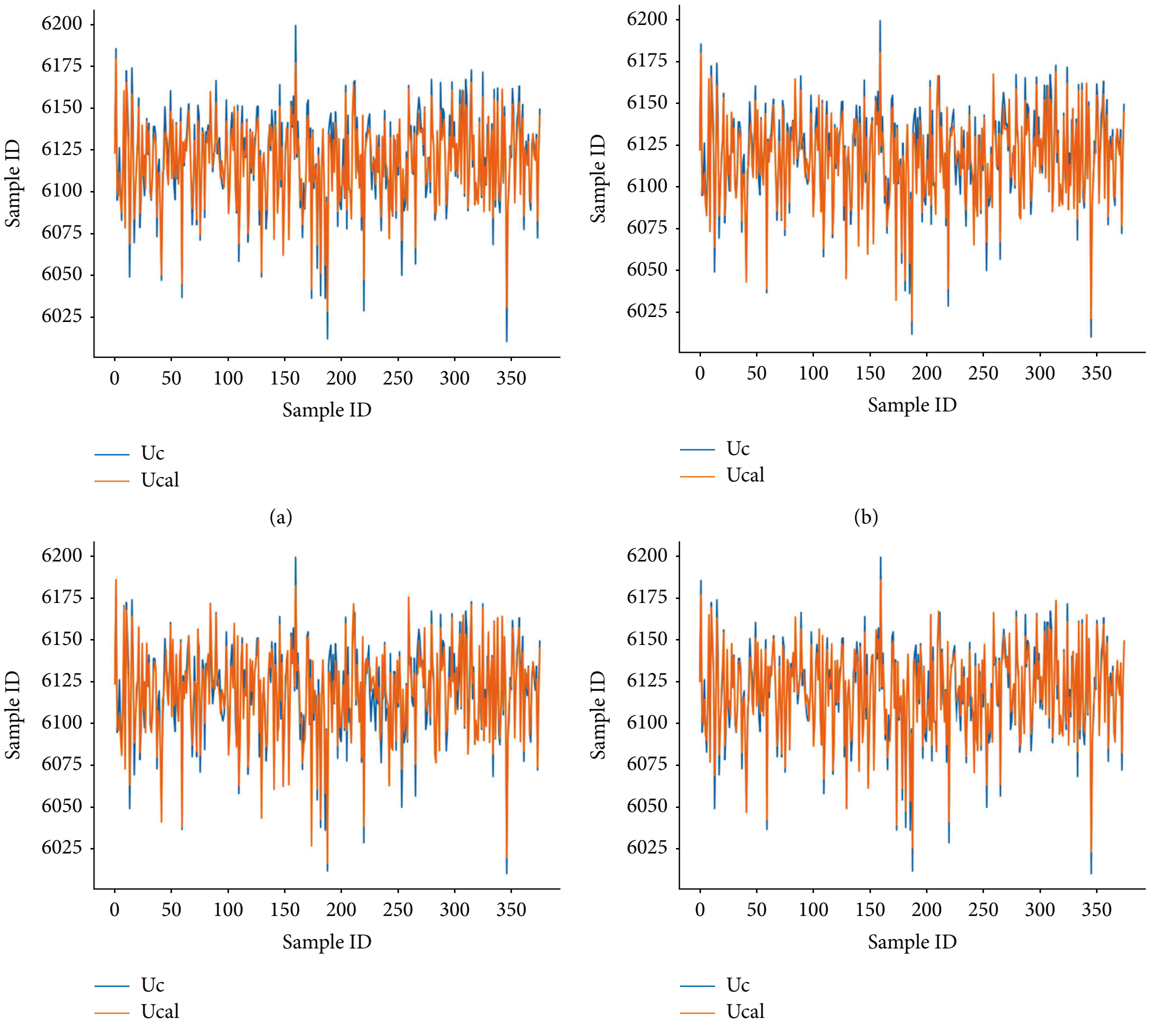
Figure 8: The scatter plot of U_c and U_{cal} by RS (a), TPE (b), SA (c) , and AGBO (d) with linear transformation.

others, which mean the prediction error of AGBO-LR is lower.

Learning rate and weight decay are two important hyperparameters in the optimization proposed in this work. Therefore, the performance for various values of these parameters is also investigated. The value of weight decay is set as 0 first, and the performances of AGBO with different learning rates are displayed in Table 4:

It is found in Table 4 that the variation of the learning rate has a large influence on the performance of optimization, and the value of the learning rate between 5e-3 and 5e-4 is suggested in this paper. The influence of weight decay on AGBO is investigated subsequently with the optimal value of learning rate; the results are listed in Table 5:

Compared with the learning rate, the values of MAE, RMSE, and MAPE of AGBO only have slight differences and the optimal values of weight decay are 1e-6. The convergence plot of AGBO with the optimal parameters is shown in Figure 10, and the loss value is converged to 0 after approximately 400 iteration steps. Compared with three SMBO algorithms, the convergence curve of AGBO is much smoother and stabler. This result can be attributed to the fact that in the gradient-based optimization method proposed in this work, the searching direction for parameter update is deterministic, while for other gradient-free methods, the searching direction is heuristic and largely dependent on the initialization, which usually leads to an oscillation in the loss function, making the convergence much more difficult.



(a)

(b)

(c)

(d)

Figure 9: The line plot of U_c and U_{cal} by RS (a), TPE (b), SA (c), and AGBO (d) with linear transformation.

Table 3: The results of parameter identification with linear transformation.

Algorithm	MAE	RMSE	MAPE
RS-LR	6.338	7.859	0.104
TPE-LR	6.000	7.616	0.098
SA-LR	7.450	9.328	0.122
AGBO-LR	**5.390**	**6.741**	**0.088**

Table 4: The performance of AGBO under different learning rates.

Learning rate	MAE	RMSE	MAPE
1e-5	50.282	51.300	0.822
5e-5	15.127	16.953	0.247
1e-4	6.987	8.693	0.114
5e-4	5.652	7.097	0.092
1e-3	5.397	6.771	0.088
5e-3	5.273	6.645	0.086
1e-2	10.859	12.670	0.177
5e-2	14.107	16.092	0.231

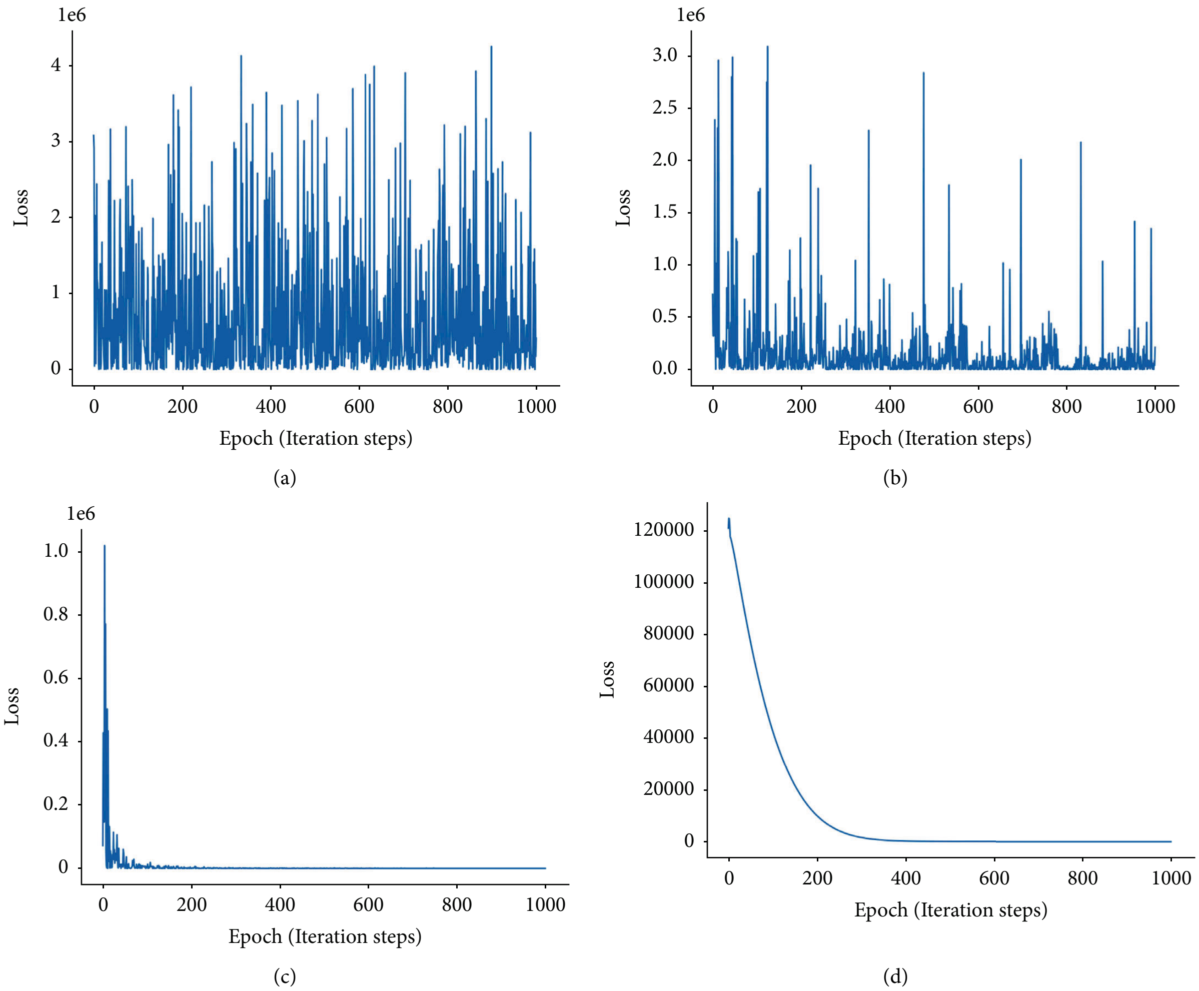


Figure 10: The convergence plot of RS-LR (a), TPE-LR (b), SA-LR (c), and AGBO-LR (d).

Table 5: The performance of AGBO with different values of weight decay.

Weight decay	MAE	RMSE	MAPE
1e-7	5.302	6.659	0.087
1e-6	5.247	6.593	0.086
1e-5	5.266	6.621	0.086
1e-4	5.300	6.678	0.087
1e-3	5.363	6.743	0.088

Since, RS, TPE, SA, and AGBO all have certain degrees of randomness, each algorithm is repeated 25 times and then averaged to guarantee the correctness and stability of numerical performance, the results in three evaluation functions are shown in Table 6 and Figure 11fig11. The repetition result indicates that

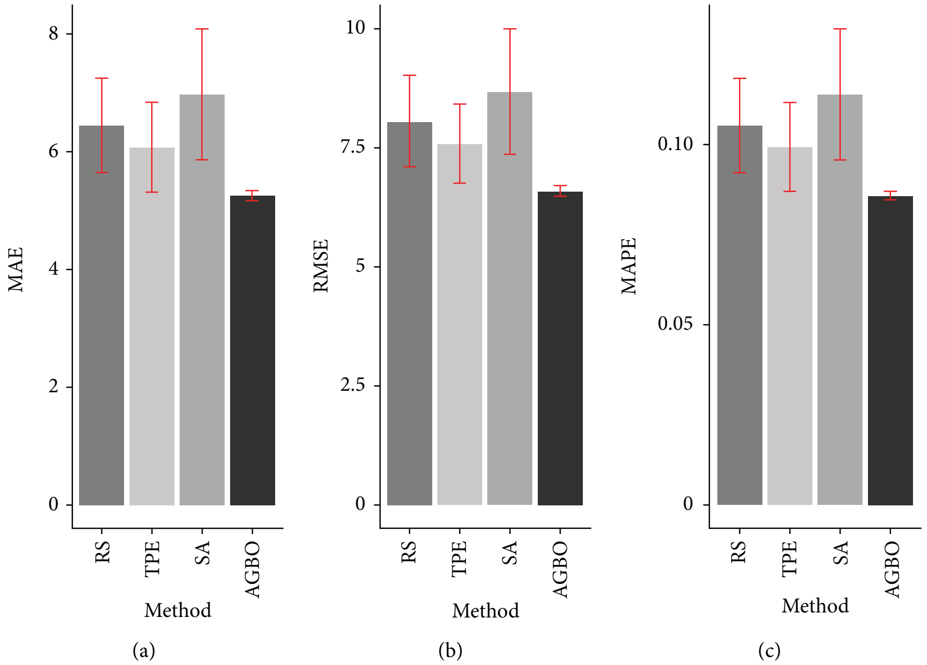


Figure 11: The bar chart with error bars from results of MAE (a), RMSE (b), and MAPE (c).

Table 6: The results of parameter identification repeated 25 times.

Algorithm	MAE	RMSE	MAPE
RS-LR	6.447 ± 0.801	8.054 ± 0.958	0.105 ± 0.013
TPE-LR	6.078 ± 0.753	7.589 ± 0.830	0.099 ± 0.012
SA-LR	6.970 ± 1.111	8.682 ± 1.318	0.114 ± 0.018
AGBO-LR	**5.247 ± 0.079**	**6.593 ± 0.111**	**0.086 ± 0.001**

AGBO-LR not only performs better in PDN's parameter identification compared with gradient-free algorithms but also has better numerical stability and robustness with the lowest variants during repeating experiments.

5. Conclusions

Parameter identification plays a key role in PDN calculation and analysis; therefore, some methods have been proposed to improve the accuracy of parameters in PDN, such as LS, MCMC and sequential model-based global optimization. However, most of the existing algorithms are classified as gradient-free methods. Therefore, in this work, an adaptive gradient-based optimization method is proposed for parameter identification in PDN. The analytical expressions of the gradients of the loss function with respect to the parameters are derived, and an adaptive updating scheme is utilized. We compare the proposed method with several heuristic algorithms such as RS, TPE, and SA. It is found that the errors via adaptive gradient-based methods are lower in all three evaluation functions, namely, MAE, RMSE, and MAPE, with smooth and stable convergence of the loss function. By further taking a linear transformation of the loss function, the method of this work significantly promotes the parameter identification performance with a much lower variance in 25 repeat experiments. In addition, the variations in hyperparameters of optimization methods such as learning rate and weight decay are also investigated, indicating that the method proposed in this work achieves more stable and robust performance to identify PDN parameters. It should be noted that the gradient-based optimization method can also be further explored in future work, such as the updating schedule of the learning rate for more stable and faster convergence of the loss function, the interpretability of the gradient-based method for parameter identification of PDN.

Data Availability

The data used to support the findings of this study are currently under embargo while the research findings are commercialized. Requests for data, 6/12 months after publication of this article, will be considered by the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Nanjing Institute of Technology Scientific Research Start-Up Fund for High-Level Introduced Talents, Grant no. YKJ202046, and the State Grid Jiangsu Electric Power Co., Ltd., Grant no. J2020097.

References

[1] J. Li, H. Ma, and H. Dong, "Present development condition and trends of energy storage technology in the integration of distributed renewable energy," *Transactions of China Electrotechnical Society*, vol. 31, pp. 1–10, 2016.

[2] P. M. Anderson, *Analysis of Faulted Power Systems*, Wiley-IEEE Press, Hoboken, NJ, USA, 1995.

[3] M. Lave, M. J. Reno, and J. Peppanen, "Distribution system parameter and topology estimation applied to resolve low-voltage circuits on three real distribution feeders," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 3, pp. 1585–1592, 2019.

[4] Z. Wang, C. Hu, W. Wang, X. Kong, and W. Zhang, "A prognostics-based spare part ordering and system replacement policy for a deteriorating system subjected to a random lead time," *International Journal of Production Research*, vol. 53, no. 15, pp. 4511–4527, 2014.

[5] Y. Lin and A. Abur, "Strategic use of synchronized phasor measurements to improve network parameter error detection," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5281–5290, 2018.

[6] A. Arancibia, C. A. Soriano-Rangel, F. Mancilla-David, R. Ortega, and K. Strunz, "Finite-time identification of the Thévenin equivalent parameters in power grids," *International Journal of Electrical Power & Energy Systems*, vol. 116, Article ID 105534, 2020.

[7] B. V. Krishna and N. P. Srinivasu, "A direct approach for distribution system load flow solutions," *International Journal of Engineering and Advanced Technology*, vol. 8, p. 6s3, 2019.

[8] B. Zaker, G. B. Gharehpetian, and M. Karrari, "Small signal equivalent model of synchronous generator-based grid-connected microgrid using improved Heffron-Phillips model," *International Journal of Electrical Power & Energy Systems*, vol. 108, pp. 263–270, 2019.

[9] Z. Wang, M. Xia, M. Lu, L. Pan, and J. Liu, "Parameter identification in power transmission systems based on graph convolution network," *IEEE Transactions on Power Delivery*, vol. 37, no. 4, pp. 3155–3163, 2022.

[10] C. Wang, Z. Wang, J. Wang, and D. Zhao, "SVM-based parameter identification for composite ZIP and electronic load modeling," *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 182–193, 2019.

[11] Z. Wang, L. Weng, M. Lu, J. Liu, and L. Pan, "Application of multi-attention mechanism in power system branch parameter identification," *Complexity*, vol. 2021, Article ID 1834428, 12 pages, 2021.

[12] Z. Yan and Y. Xu, "Data-driven load frequency control for stochastic power systems: a deep reinforcement learning method with continuous action search," *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 1653–1656, 2019.

[13] J. Song, E. Dall'Anese, A. Simonetto, and H. Zhu, "Dynamic distribution state estimation using synchrophasor data," *IEEE Transactions on Smart Grid*, vol. 11, no. 1, pp. 821–831, 2020.

[14] Z. Ma, S. Zhang, J. Li, and L. I. Junge, "A single-phase grounding fault section location method for power distribution network based on parameter identification," *Guangdong Electric Power*, vol. 32, pp. 65–71, 2019, (in Chinese).

[15] M. Cui, M. Khodayar, C. Chen, X. Wang, Y. Zhang, and M. E. Khodayar, "Deep learning-based time-varying parameter identification for system-wide load modelling," *IEEE Transactions on Smart Grid*, vol. 10, no. 6, pp. 6102–6114, 2019.

[16] K. Yuan and Z. Wei, "Power line parameter identification based on multi-innovation least square algorithm," *Electric Power Engineering Technology*, vol. 39, pp. 55–160, 2020, (in Chinese).

[17] H. Wang, H. Jiao, J. Chen, and W. Liu, "Parameter identification for a power distribution network based on MCMC algorithm," *IEEE Access*, vol. 9, pp. 104154–104161, 2021.

[18] W. Dong, P. Zanchetta, and D. W. Thomas, "Identification of electrical parameters in a power network using genetic algorithms and transient measurements," *COMPEL: The International Journal for Computation & Mathematics in Electrical & Electronic Engineering*, vol. 29, no. 1, pp. 235–249, 2010.

[19] F. Hutter, H. H. Hoos, and K. Leyton-Brown, "Sequential model-based optimization for general algorithm configuration," *Lecture Notes in Computer Science*, vol. 6683, pp. 507–523, 2011.

[20] B. James and B. Yoshua, "Random search for hyper-parameter optimization," *Journal of Machine Learning Research*, vol. 13, pp. 281–305, 2012.

[21] B. James, B. Rémi, B. Yoshua, and K. Balázs, "Algorithms for hyper-parameter optimization," *Advances in Neural Information Processing Systems*, pp. 2546–2554, 2011.

[22] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[23] Z. Qu, M. Li, Z. Zhang, M. Cui, and Y. Zhou, "Dynamic optimization method of transmission line parameters based on grey support vector regression," *Frontiers in Energy Research*, vol. 9, 2021.

[24] J. Sun, M. Xia, and Q. Chen, "A classification identification method based on phasor measurement for distribution line parameter identification under insufficient measurements conditions," *IEEE Access*, vol. 7, pp. 158732–158743, 2019.

[25] B. Li, J. Y. Ma, K. Hu et al., "A method for parameter identification of distribution network equipment based on sequential model-based optimization," *International Transactions on Electrical Energy Systems*, vol. 2022, Article ID 9880284, 12 pages, 2022.

[26] B. Li, K. Hu, J. Ma et al., "The new definitions of loss function for the model-based parameter identification method in power distribution network," *International Transactions on Electrical Energy Systems*, vol. 2022, Article ID 4197043, 11 pages, 2022.