Parameter Estimation in Three-Phase Power Distribution Networks Using Smart Meter Data

Wenyu Wang
Department of Electrical and Computer Engineering
University of California, Riverside
Riverside, USA
wwang032@ucr.edu

Nanpeng Yu
Department of Electrical and Computer Engineering
University of California, Riverside
Riverside, USA
nyu@ece.ucr.edu

***Abstract*—Accurate estimates of network parameters are essential for advanced control and monitoring in power distribution systems. The existing methods for parameter estimation either assume a simple single-phase network model or require widespread installation of micro-phasor measurement units (micro-PMUs), which are cost prohibitive. In this paper, we propose a parameter estimation approach, which considers three-phase series impedance and only leverages readily available smart meter measurements. We first build a physical model based on the linearized three-phase power flow manifold, which links the network parameters with the smart meter measurements. The parameter estimation problem is then formulated as a maximum likelihood estimation (MLE) problem. We prove that the correct network parameters yield the highest likelihood value. A stochastic gradient descent (SGD) method with early stopping is then adopted to solve the MLE problem. Comprehensive numerical tests show that the proposed algorithm improves the accuracy of the network parameters.**

***Index Terms*—Distribution network, maximum likelihood estimation, parameter estimation, smart meter.**

NOMENCLATURE

$E_{m\times n}^{(i,j)}$	An $m \times n$ matrix, in which the ij-th element is 1 and the rest of elements are all 0.
I_n	$n \times n$ identity matrix.
$Im(\cdot)$	Imaginary part of a complex variable.
$Re(\cdot)$	Real part of a complex variable.
diag$(\cdot)$	diag$(\boldsymbol{x})$ of a vector $\boldsymbol{x}$ is a diagonal matrix with $\boldsymbol{x}$ on the main diagonal. diag$(X_1, ..., X_n)$ is a block diagonal matrix with diagonal matrices of $X_1, ..., X_n$.
$\mathbb{0}_{m\times n}$	An $m \times n$ all-0 matrix.

I. INTRODUCTION

Accurate modeling of three-phase power distribution systems is gaining importance with increasing penetration of distributed energy resources (DERs). To monitor and coordinate the operations of DERs in distribution networks, distribution system operators need key applications such as three-phase power flow, distribution system state estimation, three-phase optimal power flow, and distribution network reconfiguration. All of these applications rely on accurate models of three-phase distribution networks, which include the network topology and parameters. However, the distribution network parameters and topology in the geographic information system (GIS) may contain errors due to unreliable documentation during the system modifications and upgrades.

Though the topic of topology estimation for distribution networks has been studied extensively [1]–[5], the estimation of distribution network parameters such as line impedance has not received sufficient attention. The task of parameter estimation in power distribution networks is more challenging than that in transmission networks because the distribution lines are almost always not transposed. Untransposed lines will lead to unequal diagonal and off-diagonal terms in the phase impedance matrix. Thus, instead of single-phase models, three-phase line segment models need to be developed. Specifically, the elements of a 3×3 phase impedance matrix need to be estimated for each three-phase distribution line segment.

Many technical methods have been developed to estimate transmission network parameters. However, very few of them can be applied to the three-phase distribution networks with readily available sensor data. The existing parameter estimation literature can be roughly classified into three groups based on the sensor data used. The sensor data that were used for parameter estimation include supervisory control and data acquisition (SCADA) system information, phasor measurement unit (PMU) data, and smart meter data.

The first group of literature [6], [7] uses SCADA data such as power and current injections to estimate network parameters of the transmission system with a single-phase model. Most of these works perform joint state and parameter estimation by residual sensitivity analysis, state vector augmentation, and Kalman filter.

The second group of literature uses time synchronized measurements such as voltage and current phasors to estimate single-phase line models in transmission systems and three-phase line models in distribution networks [8]–[12]. Although these algorithms can achieve highly accurate network parameter estimates, they require widespread installation of PMUs, which are extremely costly.

The third group of literature uses readily available smart meter data to estimate network parameters of distribution systems [13]–[15]. Linear regression models are fitted based on voltage magnitude and complex power consumption measurements to estimate line parameters of single-phase secondary feeders [13], [15]. By solving power flow equations with voltage magnitude and complex power measurements, the parameters of a single-phase distribution line model can be estimated [14].

In this paper, we propose a data-driven algorithm to estimate the serial conductance and serial susceptance of the π equivalent model for three-phase distribution lines by using the readily available smart meter measurements of voltage magnitude, real power consumption, and reactive power consumption. The serial conductance and susceptance are the real and imaginary part of the inverse of a line's phase impedance matrix. By linearizing the three-phase power flow manifold, we first build a physical model, which links smart meter measurements and the three-phase serial conductance and susceptance. We then formulate the three-phase parameter estimation problem as a maximum likelihood estimation (MLE) problem and prove that the correct network parameters yield the highest likelihood value. At last, we adopt the stochastic gradient descent (SGD) algorithm with early stopping to solve the MLE problem.

Compared to the existing parameter estimation methods, our proposed algorithm has two advantages. First, our proposed approach is specifically designed to estimate network parameters of three-phase distribution networks, which takes unequal self and mutual serial conductance and susceptance into consideration. Second, our proposed approach only uses readily available smart meter data and can be easily applied in real-world distribution circuits.

The rest of the paper is organized as follows. Section II describes the problem setup and builds the physical model linking network parameters with smart meter measurements. Section III formulates the parameter estimation problem as an MLE problem and proposes an SGD algorithm to solve it. Section IV evaluates the performance of the proposed algorithm with a comprehensive numerical study. Section V states the conclusion and future work.

II. Problem Setup and the Model of Network Parameter Estimation

A. Problem Setup

We aim to estimate the serial conductance and susceptance (i.e., the real and imaginary part of the inverse of the phase impedance matrix) of three-phase primary lines of a distribution feeder's network. The network contains $\mathfrak{L}$ lines/edges and $N+1$ nodes, indexed as node 0 to N, in which node 0 is the substation. There are M loads connected to the primary lines through the non-substation nodes. The loads can be single-phase, two-phase, or three-phase.

B. Assumptions

1) Availability of Measurement Data and Network Model: First, for a single-phase load on phase i, we know its power injection (both real and reactive power) and voltage magnitude of phase i. Second, for a two-phase delta-connected load between phase i and j, we know its power injection and voltage magnitude across phase i and j. Third, for a three-phase load, we know its total power injection and the voltage magnitude of a known phase i. Fourth, for the source node, we know the voltage measurement. Fifth, it is assumed that each load's phase connection is known. Sixth, the topology of the primary line network is known. Seventh, we assume that the GIS has rough estimates of the network parameters, which are inaccurate but not far from the correct values. Finally, we assume that the distribution feeder is not severely unbalanced. Assumptions one to four are based on the typical measurement configurations of smart meters and SCADA. Assumptions five to seven are based on the available information in GIS. The last assumption holds for distribution feeders under normal operations. The task of network parameter estimation is to estimate the 3×3 serial conductance and susceptance matrices of the three-phase primary line segments.

2) Statistical Assumptions: First, we assume that the incremental changes in measured real, reactive power, and voltage magnitudes across different time intervals are independent. Second, we assume that the noise term which represents the model errors and the measurement errors is i.i.d. Gaussian. Note that the noise term will be derived later in Section II-C. Third, we assume that the noise term is independent of the incremental changes in smart meter measurements. These statistical assumptions have been verified in [16].

C. Linearized Power Flow Model of Distribution Feeders

In order to build the model of network parameter estimation, we first introduce a linearized three-phase power flow model [16] as shown in (1). This linearized model links three parts of a distribution system: the first difference of smart meter measurement time series ($\tilde{\boldsymbol{v}}(t)$, $\tilde{\boldsymbol{v}}^{\text{ref}}(t)$, $\tilde{\boldsymbol{p}}(t)$, and $\tilde{\boldsymbol{q}}(t)$), the load phase connection X, and the primary feeder's topology and parameters (U^1, U^2, $\hat{U}^1$, $\hat{U}^2$, P, and $\hat{A}$). Next, we will explain these three parts in detail. For the detailed derivation of the linearized three-phase power flow model, please refer to [16]. Note that $\boldsymbol{n}(t)$ is the noise term, which represents the model errors and the measurement errors and is assumed to be i.i.d. Gaussian.

$$\begin{aligned}\tilde{\boldsymbol{v}}(t) =& X\tilde{\boldsymbol{v}}^{\text{ref}}(t) + X\begin{bmatrix}U^1 & U^2\end{bmatrix} P\breve{A}^{-1}P^T \begin{bmatrix}\hat{U}^1 & \hat{U}^2 \\ -\hat{U}^2 & \hat{U}^1\end{bmatrix} \\ &\cdot \begin{bmatrix}X^T & \\ & X^T\end{bmatrix}\begin{bmatrix}\tilde{\boldsymbol{p}}(t) \\ \tilde{\boldsymbol{q}}(t)\end{bmatrix} + \boldsymbol{n}(t)\end{aligned} \tag{1}$$

1) The Smart Meter Measurements: The measurements are modeled as follows. Let $\hat{\boldsymbol{v}}(t)$, $\hat{\boldsymbol{p}}(t)$, and $\hat{\boldsymbol{q}}(t)$ denote $M\times1$ vectors of the measurement of voltage magnitude, real power, and reactive power of the M loads at time t. Let v_0^i denote the substation's voltage magnitude of phase i, and let v_0^{ij} denote the substation's voltage magnitude across phase ij. Define a $3M\times1$ vector $\hat{\boldsymbol{v}}^{\text{ref}}(t) \triangleq [\hat{\boldsymbol{v}}_1^{\text{ref}}(t),\dots,\hat{\boldsymbol{v}}_M^{\text{ref}}(t)]^T$, where $\hat{\boldsymbol{v}}_m^{\text{ref}}(t) = [v_0^a(t), v_0^b(t), v_0^c(t)]$ if load m is single-phase or three-phase; $\hat{\boldsymbol{v}}_m^{\text{ref}} = [v_0^{ab}(t), v_0^{bc}(t), v_0^{ca}(t)]$ if load m is two-phase. $\tilde{\boldsymbol{v}}(t) \triangleq \hat{\boldsymbol{v}}(t) - \hat{\boldsymbol{v}}(t-1)$. $\tilde{\boldsymbol{p}}(t)$, $\tilde{\boldsymbol{q}}(t)$, and $\tilde{\boldsymbol{v}}^{\text{ref}}(t)$ are defined in a similar way as $\tilde{\boldsymbol{v}}(t)$.

2) The Load Phase Connection: The $M\times3M$ block diagonal matrix $X \triangleq \text{diag}([x_1^1\ x_1^2\ x_1^3],...,[x_M^1\ x_M^2\ x_M^3])$ represents the loads' phase connections. $x_m^i = 0$ or 1, and $\sum_i x_m^i = 1,\ \forall\ m$. If load m is single-phase, then x_m^1, x_m^2, and x_m^3 represent AN, BN, and CN connections. If m is two-phase, then x_m^1, x_m^2, and x_m^3 represent AB, BC, and CA connections. If m is three-phase and one of AN, BN, and CN voltages is measured, then x_m^1, x_m^2, and x_m^3 represent which phase is measured.

3) The Primary Feeder's Topology and Parameters: The primary feeder's topology and parameters are modeled as follows. Let $\alpha = e^{-j\frac{2\pi}{3}}$ and let $I_{(N+1)}$ be an identity matrix of size $N+1$. Let's first define matrix Y as follows:

$$Y \triangleq \begin{bmatrix} Y^{aa} & Y^{ab} & Y^{ac} \\ Y^{ba} & Y^{bb} & Y^{bc} \\ Y^{ca} & Y^{cb} & Y^{cc} \end{bmatrix} \tag{2}$$

where Y^{ij} is the $(N+1)\times(N+1)$ nodal admittance matrix between phase i and j. Define a block diagonal matrix $\Phi \triangleq \text{diag}(I_{(N+1)},\ \alpha I_{(N+1)},\ \alpha^2 I_{(N+1)})$ and define matrix A as

$$A \triangleq \begin{bmatrix} Re(\Phi^{-1}Y\Phi) & -Im(\Phi^{-1}Y\Phi) \\ -Im(\Phi^{-1}Y\Phi) & -Re(\Phi^{-1}Y\Phi) \end{bmatrix} \tag{3}$$

By removing the rows and columns corresponding to the substation from A, we obtain a $6N \times 6N$ matrix $\breve{A}$. P is a known constant $6N \times 6N$ permutation matrix that regroups the rows and columns of $\breve{A}$ by nodes instead of by phases.

U^1 and U^2 are $3M \times 3N$ matrices. $\hat{U}^1$ and $\hat{U}^2$ are $3N \times 3M$ matrices. These four matrices represent which of the N non-substation nodes each load is connected to and how many phases each load is connected to. Please refer to [16] for the details on the calculation of these matrices. The elements of these matrices are determined once each load's location and the number of phases are given. In this paper, these four matrices are treated as constant matrices.

D. Explicit Model of Distribution Line Parameters in Linearized Power Flow Model

The distribution line parameters are implicitly considered in $\breve{A}^{-1}$ of the linearized power flow model (1) derived in Section II-C. In this subsection, we explicitly model the distribution line parameters in the linearized power flow model.

A three-phase line segment l's serial conductance and susceptance can be represented by two symmetric matrices, the serial conductance matrix $[g]_l$ and the serial susceptance matrix $[b]_l$:

$$[g]_l \triangleq \begin{bmatrix} g_l^{aa} & g_l^{ab} & g_l^{ac} \\ g_l^{ba} & g_l^{bb} & g_l^{bc} \\ g_l^{ca} & g_l^{cb} & g_l^{cc} \end{bmatrix},\ [b]_l \triangleq \begin{bmatrix} b_l^{aa} & b_l^{ab} & b_l^{ac} \\ b_l^{ba} & b_l^{bb} & b_l^{bc} \\ b_l^{ca} & b_l^{cb} & b_l^{cc} \end{bmatrix} \tag{4}$$

Since both matrices are symmetric, only 12 independent parameters need to be derived for each line segment. Define Λ as the set of all line parameters, i.e., $\Lambda \triangleq \{g_l^{ij}, b_l^{ij} \,|\, l \in \{1,\ldots,\mathfrak{L}\},\ ij \in \{aa,ab,ac,bb,bc,cc\}\}$. Define $\boldsymbol{g}^{ij}$ and $\boldsymbol{b}^{ij}$ as $\boldsymbol{g}^{ij} \triangleq [g_1^{ij},\ldots,g_{\mathfrak{L}}^{ij}]$ and $\boldsymbol{b}^{ij} \triangleq [b_1^{ij},\ldots,b_{\mathfrak{L}}^{ij}]$. Then, the serial conductances can be grouped in a $3\mathfrak{L} \times 3\mathfrak{L}$ matrix as:

$$\Lambda_g \triangleq \begin{bmatrix} \text{diag}(\boldsymbol{g}^{aa}) & \text{diag}(\boldsymbol{g}^{ab}) & \text{diag}(\boldsymbol{g}^{ac}) \\ \text{diag}(\boldsymbol{g}^{ab}) & \text{diag}(\boldsymbol{g}^{bb}) & \text{diag}(\boldsymbol{g}^{bc}) \\ \text{diag}(\boldsymbol{g}^{ac}) & \text{diag}(\boldsymbol{g}^{bc}) & \text{diag}(\boldsymbol{g}^{cc}) \end{bmatrix} \tag{5}$$

Λ_b can be defined in a similar way for serial susceptances. Next we define four $3\mathfrak{L} \times 3\mathfrak{L}$ matrices:

$$\begin{aligned} \sin(\Phi^{-1}) &\triangleq \text{diag}\big(\sin(0)\cdot I_{\mathfrak{L}}, \sin(\tfrac{2\pi}{3})\cdot I_{\mathfrak{L}}, \sin(-\tfrac{2\pi}{3})\cdot I_{\mathfrak{L}}\big) \\ \cos(\Phi^{-1}) &\triangleq \text{diag}\big(\cos(0)\cdot I_{\mathfrak{L}}, \cos(\tfrac{2\pi}{3})\cdot I_{\mathfrak{L}}, \cos(-\tfrac{2\pi}{3})\cdot I_{\mathfrak{L}}\big) \\ \sin(\Phi) &\triangleq \text{diag}\big(\sin(0)\cdot I_{\mathfrak{L}}, \sin(-\tfrac{2\pi}{3})\cdot I_{\mathfrak{L}}, \sin(\tfrac{2\pi}{3})\cdot I_{\mathfrak{L}}\big) \\ \cos(\Phi) &\triangleq \text{diag}\big(\cos(0)\cdot I_{\mathfrak{L}}, \cos(-\tfrac{2\pi}{3})\cdot I_{\mathfrak{L}}, \cos(\tfrac{2\pi}{3})\cdot I_{\mathfrak{L}}\big) \end{aligned} \tag{6}$$

Define two rotation matrices $R(\Phi^{-1})$ and $R(\Phi)$ as:

$$\begin{aligned} R(\Phi^{-1}) &\triangleq \begin{bmatrix} \cos(\Phi^{-1}) & \sin(\Phi^{-1}) \\ -\sin(\Phi^{-1}) & \cos(\Phi^{-1}) \end{bmatrix} \\ R(\Phi) &\triangleq \begin{bmatrix} \cos(\Phi) & \sin(\Phi) \\ -\sin(\Phi) & \cos(\Phi) \end{bmatrix} \end{aligned} \tag{7}$$

Let $\mathcal{A}$ denote the $(N+1)\times\mathfrak{L}$ incidence matrix representing the topology of the primary feeder. If line segment l connects node i and j ($i < j$), then $\mathcal{A}_{il} = 1$, $\mathcal{A}_{jl} = -1$, and $\mathcal{A}_{kl} = 0, \forall k \neq i,j$. By removing the row corresponding to the substation, we obtain a $N \times \mathfrak{L}$ matrix $\breve{\mathcal{A}}$. Define $\breve{\mathcal{A}}_{6N}$ as $\breve{\mathcal{A}}_{6N} \triangleq \text{diag}(\breve{\mathcal{A}},\breve{\mathcal{A}},\breve{\mathcal{A}},\breve{\mathcal{A}},\breve{\mathcal{A}},\breve{\mathcal{A}})$ and define Λ_y as:

$$\Lambda_y \triangleq \begin{bmatrix} \Lambda_g & -\Lambda_b \\ -\Lambda_b & -\Lambda_g \end{bmatrix} \tag{8}$$

Then, it can be shown that

$$\breve{A} = \breve{A}(\Lambda) = \breve{\mathcal{A}}_{6N} R(\Phi^{-1}) \Lambda_y R(\Phi)^T \breve{\mathcal{A}}_{6N}^T \tag{9}$$

By plugging (9) into (1), we can obtain an explicit model of network parameters in the linearized power flow model.

III. Maximum Likelihood Estimation of Distribution Network Parameters

In this section, we first show how to formulate the network parameter estimation problem using maximum likelihood estimation (MLE). Then, we derive the gradient of the negative log likelihood function with respect to network parameters. Lastly, we develop an SGD-based algorithm with early stopping to solve the MLE problem.

A. MLE Problem Formulation

Define $\tilde{\boldsymbol{v}}(t,\Lambda)$ as the theoretical value of $\tilde{\boldsymbol{v}}(t)$, i.e., the first difference of voltage time series with network parameters Λ:

$$\tilde{\boldsymbol{v}}(t,\Lambda) \triangleq X\tilde{\boldsymbol{v}}^{\text{ref}}(t) + X\begin{bmatrix} U^1 & U^2 \end{bmatrix} P\breve{A}(\Lambda)^{-1}P^T \begin{bmatrix} \hat{U}^1 & \hat{U}^2 \\ -\hat{U}^2 & \hat{U}^1 \end{bmatrix} \cdot \begin{bmatrix} X^T & \\ & X^T \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{p}}(t) \\ \tilde{\boldsymbol{q}}(t) \end{bmatrix} \tag{10}$$

Then, $\tilde{\boldsymbol{v}}(t) = \tilde{\boldsymbol{v}}(t,\Lambda) + \boldsymbol{n}(t)$, in which Λ is the set of network parameters to estimate.

As stated in Section II-B2, we assume that the noise $\boldsymbol{n}(t)$ is independent of $\tilde{\boldsymbol{v}}^{\text{ref}}(t)$, $\tilde{\boldsymbol{p}}(t)$, and $\tilde{\boldsymbol{q}}(t)$ and is i.i.d. Gaussian $\boldsymbol{n}(t) \sim \mathcal{N}(\mathbb{0}_{M\times 1}, \Sigma_n)$, where Σ_n is an unknown underlying covariance matrix. Given these conditions, $\boldsymbol{n}(t)$ is also independent of $\tilde{\boldsymbol{v}}(t,\Lambda)$. Thus, the likelihood of observing

$\{\tilde{\boldsymbol{v}}(t)\}_{t=1}^T$ given X, $\{\tilde{\boldsymbol{v}}^{\text{ref}}(t)\}_{t=1}^T$, $\{\tilde{\boldsymbol{p}}(t)\}_{t=1}^T$, and $\{\tilde{\boldsymbol{q}}(t)\}_{t=1}^T$ is a function of Λ:

$$Prob(\{\tilde{\boldsymbol{v}}(t)\}_{t=1}^T|X,\{\tilde{\boldsymbol{v}}^{\text{ref}}(t)\}_{t=1}^T,\{\tilde{\boldsymbol{p}}(t)\}_{t=1}^T,\{\tilde{\boldsymbol{q}}(t)\}_{t=1}^T;\Lambda) = \frac{|\Sigma_n|^{-\frac{T}{2}}}{(2\pi)^{\frac{MT}{2}}}\times\exp\Big\{-\frac{1}{2}\sum_{t=1}^{T}[\tilde{\boldsymbol{v}}(t)-\tilde{\boldsymbol{v}}(t,\Lambda)]^T\Sigma_n^{-1}[\tilde{\boldsymbol{v}}(t)-\tilde{\boldsymbol{v}}(t,\Lambda)]\Big\} \tag{11}$$

Taking the negative logarithm of (11), removing the constant term, and scaling by $\frac{2}{T}$, we get

$$f(\Lambda) \triangleq \frac{1}{T}\sum_{t=1}^{T}[\tilde{\boldsymbol{v}}(t)-\tilde{\boldsymbol{v}}(t,\Lambda)]^T\Sigma_n^{-1}[\tilde{\boldsymbol{v}}(t)-\tilde{\boldsymbol{v}}(t,\Lambda)] \tag{12}$$

It will be shown in Lemma 1 that the correct network parameters Λ maximize the likelihood function (11) and minimizes $f(\Lambda)$ under two mild assumptions.

Lemma 1. *Let Λ^* be the correct network parameters. If the following two conditions are satisfied, then as $T\to\infty$, Λ^* is a global minimizer of $f(\Lambda)$.*

1) *$\boldsymbol{n}(t_k)$ is i.i.d. and independent of $\tilde{\boldsymbol{v}}^{ref}(t_l)$, $\tilde{\boldsymbol{p}}(t_l)$, and $\tilde{\boldsymbol{q}}(t_l)$, for $\forall t_k, t_l \in Z^+$.*
2) *$\tilde{\boldsymbol{v}}^{ref}(t_k)$, $\tilde{\boldsymbol{p}}(t_k)$, and $\tilde{\boldsymbol{q}}(t_k)$ are independent of $\tilde{\boldsymbol{v}}^{ref}(t_l)$, $\tilde{\boldsymbol{p}}(t_l)$, and $\tilde{\boldsymbol{q}}(t_l)$, for $\forall t_k, t_l \in Z^+,\ t_k \neq t_l$*

For the proof of Lemma 1, please refer to Appendix E of Ref. [16]. The only difference is that in this paper, the decision variable is the network parameter Λ, while in Ref. [16], the decision variable is the phase connection $\boldsymbol{x}$. In real-world applications, Σ_n is unknown, so we can use I_M instead. With I_M, Lemma 1 still holds and the proof is similar.

By substituting (10) into (12), we can see that directly minimizing $f(\Lambda)$ is very difficult because it is nonconvex and highly nonlinear. Thus, we adopt SGD to solve the problem.

B. Derive Gradient of the Negative Log-likelihood Function

In this subsection, we derive the gradient of $f(\Lambda)$, which will be used to find the Λ that minimizes $f(\Lambda)$. To derive the gradient in matrix form, we define the following terms:

$$\boldsymbol{y}(t) \triangleq \tilde{\boldsymbol{v}}(t) - X\tilde{\boldsymbol{v}}^{\text{ref}}(t),\ \boldsymbol{z}(t) \triangleq \begin{bmatrix}\tilde{\boldsymbol{p}}(t)\\ \tilde{\boldsymbol{q}}(t)\end{bmatrix}$$
$$\mathcal{C} \triangleq X\begin{bmatrix}U^1 & U^2\end{bmatrix}P,\ \mathcal{D} \triangleq P^T\begin{bmatrix}\hat{U}^1 & \hat{U}^2\\ -\hat{U}^2 & \hat{U}^1\end{bmatrix}\begin{bmatrix}X^T & \\ & X^T\end{bmatrix} \tag{13}$$

Then $\tilde{\boldsymbol{v}}(t)-\tilde{\boldsymbol{v}}(t,\Lambda) = \boldsymbol{y}(t) - \mathcal{C}\check{A}(\Lambda)^{-1}\mathcal{D}\boldsymbol{z}(t)$. Using the chain rule, for $\forall\lambda\in\Lambda$, we have

$$\frac{\partial f(\Lambda)}{\partial\lambda} = \text{Tr}\left(\left[\frac{\partial f(\Lambda)}{\partial(\mathcal{C}\check{A}(\Lambda)^{-1}\mathcal{D})}\right]^T \times \frac{\partial(\mathcal{C}\check{A}(\Lambda)^{-1}\mathcal{D})}{\partial\lambda}\right) \tag{14}$$

where

$$\frac{\partial f(\Lambda)}{\partial(\mathcal{C}\check{A}(\Lambda)^{-1}\mathcal{D})} = -\frac{2}{T}\Sigma_n^{-1}\cdot\sum_{t=1}^{T}\big(\boldsymbol{y}(t)-\mathcal{C}\check{A}(\Lambda)^{-1}\mathcal{D}\boldsymbol{z}(t)\big)\boldsymbol{z}(t)^T \tag{15}$$

Calculating $\partial(\mathcal{C}\check{A}(\Lambda)^{-1}\mathcal{D})/\partial\lambda$ is equivalent to calculating every element's derivative $\partial[\mathcal{C}\check{A}(\Lambda)^{-1}\mathcal{D}]_{i,j}/\partial\lambda$, in which $[\mathcal{C}\check{A}(\Lambda)^{-1}\mathcal{D}]_{i,j}$ is the ijth element of $(\mathcal{C}\check{A}(\Lambda)^{-1}\mathcal{D})$, $i = 1,\ldots,M$ and $j = 1,\ldots,2M$. Using the chain rule, we have

$$\frac{\partial[(\mathcal{C}\check{A}(\Lambda)^{-1}\mathcal{D})]_{i,j}}{\partial\lambda} = \text{Tr}\left(\left[\frac{\partial[\mathcal{C}\check{A}(\Lambda)^{-1}\mathcal{D}]_{i,j}}{\partial\check{A}(\Lambda)}\right]^T \times \frac{\partial\check{A}(\Lambda)}{\partial\lambda}\right) \tag{16}$$

Define $E_{m\times n}^{(i,j)}$ as an $m\times n$ matrix, in which the ij-th element is 1 and the rest of elements are all 0. Using the rules of matrix derivatives [17], we have

$$\frac{\partial[\mathcal{C}\check{A}(\Lambda)^{-1}\mathcal{D}]_{i,j}}{\partial\check{A}(\Lambda)} = -\check{A}(\Lambda)^{-T}\mathcal{C}^T E_{M\times 2M}^{(i,j)}\mathcal{D}^T\check{A}(\Lambda)^{-T} \tag{17}$$

Let $[\check{A}(\Lambda)]_{i,j}$ be the ijth element of $\check{A}(\Lambda)$. Using the chain rule, we get

$$\frac{\partial[\check{A}(\Lambda)]_{i,j}}{\partial\lambda} = \text{Tr}\left(\left[\frac{\partial[\check{A}(\Lambda)]_{i,j}}{\partial\Lambda_y}\right]^T \times \frac{\partial\Lambda_y}{\partial\lambda}\right) \tag{18}$$

Using the rules of matrix derivatives [17], we have

$$\frac{\partial[\check{A}(\Lambda)]_{i,j}}{\partial\Lambda_y} = R(\Phi^{-1})^T\check{\mathcal{A}}_{6N}^T E_{6N\times 6N}^{(i,j)}\check{\mathcal{A}}_{6N}R(\Phi) \tag{19}$$

From (8), we have

$$\frac{\partial\Lambda_y}{\partial\lambda} = \begin{bmatrix}\partial\Lambda_g/\partial\lambda & -\partial\Lambda_b/\partial\lambda\\ -\partial\Lambda_b/\partial\lambda & -\partial\Lambda_g/\partial\lambda\end{bmatrix} \tag{20}$$

The calculation of $\partial\Lambda_g/\partial\lambda$ is straightforward. By (5), we have

$$\frac{\partial\Lambda_g}{\partial\lambda} = \begin{cases}\mathbb{0}_{3\mathfrak{L}\times 3\mathfrak{L}} & \text{if } \lambda\notin\Lambda_g\\ E_{3\mathfrak{L}\times 3\mathfrak{L}}^{(i,i)} & \text{if } \lambda \text{ is the } ii\text{-th diagonal element in } \Lambda_g\\ E_{3\mathfrak{L}\times 3\mathfrak{L}}^{(i,j)} + E_{3\mathfrak{L}\times 3\mathfrak{L}}^{(j,i)} & \text{if } \lambda \text{ is the } ij\text{-th and } ji\text{-th off-diagonal elements in } \Lambda_g\end{cases} \tag{21}$$

$\partial\Lambda_b/\partial\lambda$ can be calculated in a similar way. Based on the derivations above, we can calculate the gradient $\nabla f(\Lambda)$ for any given Λ by calculating $\partial f(\Lambda)/\partial\lambda$ for all $\lambda\in\Lambda$ as follows. First, calculate $\partial[\check{A}(\Lambda)]_{i,j}/\partial\lambda$ for $\forall i,j$ using (18), (19), (20), and (21). Next, calculate $\partial[(\mathcal{C}\check{A}(\Lambda)^{-1}\mathcal{D})]_{i,j}/\partial\lambda$ for $\forall i,j$ using (16) and (17). Lastly, calculate $\partial f(\Lambda)/\partial\lambda$ using (14) and (15).

C. The SGD Algorithm

We design an SGD-based algorithm with early stopping to minimize $f(\Lambda)$ and estimate the network parameters Λ. As shown in Algorithm 1, in step 1, the parameters Λ are initialized with their original values in the GIS. The initial values for the parameters are assumed to be not far from the correct values. In steps 2 to 17, we iteratively update Λ using SGD, in which we update Λ by descending $f(\Lambda)$'s gradient over a small group of samples (i.e., a batch) of size n_{batch}. We use patience $n_{patience}$ to decide when to stop the iterative process. That is to say, the algorithm will be stopped if $f(\Lambda)$ over all T samples is not improved in $n_{patience}$ epochs (an epoch goes through all T samples in batches). Steps 4 to 9 show the procedure of updating Λ over each batch of samples, in which we use the backtracking line search of parameters a_{step}, α, and β to determine the step size in each move. In step 18, the parameters Λ with the lowest $f(\Lambda)$ is selected as the output.

Algorithm 1 Network Parameter Estimation Algorithm

Input: First difference of smart meter measurements $\tilde{\boldsymbol{v}}(t)$, $\tilde{\boldsymbol{v}}^{\text{ref}}(t)$, $\tilde{\boldsymbol{p}}(t)$, and $\tilde{\boldsymbol{q}}(t)$, $t = 1, ..., T$; feeder constant matrices $\mathcal{C}$, $\mathcal{D}$, $R(\Phi)$, $R(\Phi^{-1})$; hyperparameters n_{batch}, $n_{patience}$, a_{step}, α, and β; an initial estimate of Λ for the $\mathfrak{L}$ primary line segments.

Output: Updated estimates of Λ.

```
Let Λ_best = Λ and f_best = f(Λ), in which f(Λ) is calculated over all T measurements. n_epoch = 0.
while n_epoch < n_patience do
    Randomly divide the T measurements into batches of size n_batch.
    for each batch do
        Calculate ∇f(Λ) over the batch following Section III-B. The descent direction is ΔΛ = −∇f(Λ). s = a_step.
        while f(Λ + sΔΛ) > f(Λ) + αs∇f(Λ)^T ΔΛ do
            s = βs
        end while
        Λ = Λ + sΔΛ
    end for
    Calculate f(Λ) over all T measurements.
    if f(Λ) < f_best then
        f_best = f(Λ), Λ_best = Λ, n_epoch = 0.
    else
        n_epoch = n_epoch + 1
    end if
end while
Output the Λ_best.
```

IV. Numerical Study

A. Setup for Numerical Tests

We evaluate the performance of our proposed parameter estimation algorithm with the modified IEEE 13-bus test feeder, which is shown in Figure 1. We modify the standard 13-bus test feeder by introducing loads with all 7 types of phase connections, AN, BN. CN, AB, BC, CA, and ABC. The test circuits' primary feeder contains 6 line segments and 7 nodes, which serve 10 loads.

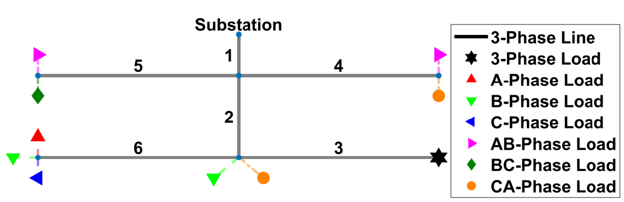


Fig. 1. Schematic of the modified IEEE 13-bus test feeder.

We aggregated the hourly average real power consumption data from the smart meters of a distribution feeder managed by an electric utility in North America, as the hourly loads on the test feeder. The length of the real power consumption time series is 2160, which represents 90 days of measurements. The reactive power time series of the lagging loads are calculated with power factors randomly sampled from a uniform distribution $\mathcal{U}(0.9, 1)$ (a typical range for distribution network loads). The peak load of the feeder is 3 MW. The power flow results are generated using OpenDSS. All smart meter measurements contain noise, which follows a zero-mean Gaussian distribution with three-sigma deviation matching 0.1% to 0.2% of the nominal values. The 0.1 and 0.2 accuracy class smart meters established in ANSI C12.20-2015 represent the typical noise levels in real-world systems. To make the parameter estimation task more challenging, we assume the smart meters have limited precision. That is to say, after adding measurement noise, the voltage measurements are rounded to the nearest 1V. The real and reactive power measurements are rounded to the nearest 0.1 kW and 0.1 kVar. We assume that the initial estimates for network parameters Λ are randomly sampled from a uniform distribution within $\pm 50\%$ of the correct values, which are very inaccurate starting values.

We set hyperparameters of the SGD algorithm as $n_{batch} = 10$, $n_{patience} = 10$, $a_{step} = 1e8$, $\alpha = 0.3$, and $\beta = 0.5$. These values are set empirically so that the algorithm updates $f(\Lambda)$ adequately and stops when it saturates. The SGD algorithm is implemented using MATLAB on a DELL workstation with 3.3 GHz Intel Xeon CPU and 16 GB RAM.

B. Performance of Parameter Estimation Algorithm

We demonstrate the effectiveness of our proposed network parameter estimation algorithm with two meter accuracy classes (0.1% and 0.2%) and two time windows (30 days and 90 days). We use the mean absolute deviation ratio (MADR) to measure the parameter estimation error. The MADR between the estimated Λ and the correct value Λ^* is defined in (22).

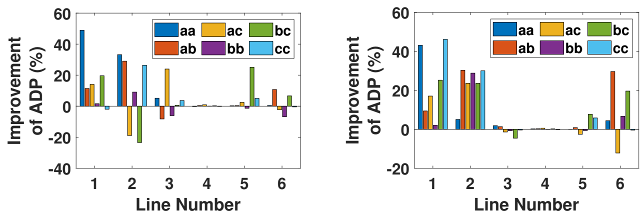
$$\text{MADR} \triangleq \frac{\sum_{i=1}^{12\mathfrak{L}} |\lambda_i - \lambda_i^*|}{\sum_{i=1}^{12\mathfrak{L}} |\lambda_i^*|} \times 100\% \tag{22}$$

The percentage of MADR improvements resulting from applying our proposed algorithm is reported in Table I. In other words, we report $(\text{MADR}_{initial} – \text{MADR}_{final})/\text{MADR}_{initial} \times 100\%$, where $\text{MADR}_{initial}$ and MADR_{final} represent the MADR of the initial and the final network parameter estimates. The maximum possible MADR improvement is 100% with perfect estimation, i.e., $\text{MADR}_{final} = 0\%$. As shown in the table, our proposed algorithm significantly reduces the network parameter estimation error. The improvement is more pronounced with longer periods of more accurate smart meter data.

TABLE I
Improvement in MADR of the Network Parameter Estimates of the Proposed Algorithm

Meter Class	Number of Days	MADR Improvement
0.1%	30	12.53%
	90	13.54%
0.2%	30	8.76%
	90	11.64%

To quantify the estimation error of each network parameter, we define the absolute deviation percentage (ADP) of a parameter λ_i as $|\lambda_i - \lambda_i^*|/|\lambda_i^*| \times 100\%$. Figure 2 shows the improvement in ADP due to the proposed algorithm, i.e., $\text{ADP}_{initial} - \text{ADP}_{final}$. As shown in the figure, our proposed algorithm reduces ADP for most of the network parameters. The improvement is more significant for line segments 1 and 2, which are the "backbones" of the feeder. Some parameters' estimation deteriorates with negative improvement, indicating that the algorithm may converge to a local minimum.



(a) Serial Conductance (b) Serial Susceptance

Fig. 2. Improvement in ADP of all network parameter estimates due to the proposed algorithm (0.1% meter class, 90 day data).

C. Performance with Different Smart Meter Penetration Levels

We also evaluate the performance of our proposed method with different smart meter penetration levels. In the 13-bus test feeder, there are 10 and 45 possible meter placement combinations with 90% and 80% smart meter penetration levels. The reduction in MADR are calculated for each case and the average reduction in MADR due to our proposed algorithm are reported in Table II. As shown in the table, the improvement in MADR decreases when the penetration level of the smart meters decreases. When the smart meter penetration level drops to around 80%, our proposed algorithm is no longer effective. This is because the linearized power flow model becomes inaccurate when we have incomplete smart meter measurements. Note that this limitation of our proposed algorithm will be less concerning as the penetration level of smart meters continues to increase.

TABLE II
IMPACT OF SMART METER PENETRATION LEVEL ON THE PERFORMANCE OF THE PROPOSED ALGORITHM

Meter Class	100% Penetration	90% Penetration	80% Penetration
0.1%	13.54%	6.24%	-1.41%
0.2%	11.64%	2.23%	-7.70%

V. CONCLUSION AND FUTURE WORK

In this paper, we develop a data-driven parameter estimation algorithm for three-phase power distribution networks. Our proposed algorithm uses only the readily available smart meter data to estimate the three-phase serial conductance and susceptance of the primary line segments. The network parameter estimation problem is first formulated as an MLE problem based on the linearized three-phase power flow. It can be proven that the correct network parameters yield the highest likelihood value. We design an SGD-based algorithm with early stopping to solve the MLE problem. The comprehensive numerical study results show that our proposed algorithm is capable of improving the accuracy of the parameter estimates.

In the future, we plan to develop parameter estimation methods that rigorously consider the prior knowledge of network parameters estimates. In addition, we will develop an algorithm to jointly estimate network parameters and phase connections of distribution feeders.

REFERENCES

[1] W. Wang, N. Yu, B. Foggo, J. Davis, and J. Li, "Phase identification in electric power distribution systems by clustering of smart meter data," in *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, Dec. 2016, pp. 259–265.

[2] W. Wang and N. Yu, "Advanced metering infrastructure data driven phase identification in smart grid," in *The Second International Conference on Green Communications, Computing and Technologies*, Sep. 2017, pp. 16–23.

[3] B. Foggo and N. Yu, "A comprehensive evaluation of supervised machine learning for the phase identification problem," *World Academy of Science, Engineering and Technology International Journal of Computer and Systems Engineering*, vol. 12, no. 6, pp. 419–427, 2018.

[4] Y. Liao, Y. Weng, G. Liu, Z. Zhang, C. W. Tan, and R. Rajagopal, "Unbalanced three-phase distribution grid topology estimation and bus phase identification," *arXiv preprint arXiv:1809.07192 [cs.SY]*, Sep. 2018. [Online]. Available: https://arxiv.org/pdf/1809.07192.pdf

[5] B. Foggo and N. Yu, "Improving supervised phase identification through the theory of information losses," *IEEE Transactions on Smart Grid*, 2019.

[6] W.-H. E. Liu and S.-L. Lim, "Parameter error identification and estimation in power system state estimation," *IEEE Transactions on Power Systems*, vol. 10, no. 1, pp. 200–209, 1995.

[7] P. Zarco and A. G. Exposito, "Power system parameter estimation: a survey," *IEEE Transactions on Power Systems*, vol. 15, no. 1, pp. 216–222, 2000.

[8] X. Bian, X. R. Li, H. Chen, D. Gan, and J. Qiu, "Joint estimation of state and parameter with synchrophasors—part ii: Parameter tracking," *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1209–1220, 2011.

[9] C. Rakpenthai and S. Uatrongjit, "Power system state and transmission line conductor temperature estimation," *IEEE Transactions on Power Systems*, vol. 32, no. 3, pp. 1818–1827, 2016.

[10] D. Ritzmann, P. S. Wright, W. Holderbaum, and B. Potter, "A method for accurate transmission line impedance parameter estimation," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 10, pp. 2204–2213, 2016.

[11] J. Yu, Y. Weng, and R. Rajagopal, "Patopa: a data-driven parameter and topology joint estimation framework in distribution grids," *IEEE Transactions on Power Systems*, vol. 33, no. 4, pp. 4335–4347, 2017.

[12] O. Ardakanian, V. W. S. Wong, R. Dobbe, S. H. Low, A. von Meier, C. J. Tomlin, and Y. Yuan, "On identification of distribution grids," *IEEE Transactions on Control of Network Systems*, vol. 6, no. 3, pp. 950–960, 2019.

[13] T. A. Short, "Advanced metering for phase identification, transformer identification, and secondary modeling," *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 651–658, Jun. 2013.

[14] A. M. Prostejovsky, O. Gehrke, A. M. Kosek, T. Strasser, and H. W. Bindner, "Distribution line parameter estimation under consideration of measurement tolerances," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 2, pp. 726–735, 2016.

[15] 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.

[16] W. Wang and N. Yu, "Maximum marginal likelihood estimation of phase connections in power distribution systems," *IEEE Transactions on Power Systems*, 2020. [Online]. Available: https://arxiv.org/pdf/1902.09686v2.pdf

[17] D. Zwillinger, *CRC Standard Mathematical Tables and Formulas*, 33rd ed. Chapman and Hall/CRC, 2018.