Ambient Signals based Load Modeling with Combined Gradient-based Optimization and Regression Method

Xinran Zhang, Member, IEEE, David J. Hill, Life Fellow, IEEE, Chao Lu, Senior Member, IEEE, and Yue Song, Member, IEEE

Abstract—Load modeling has been an important issue in modeling a power system. Ambient signals based load modeling approach has recently been proposed to better track the time-varying changes of load models caused by the increasing uncertain factors in power loads. To improve the computation efficiency and the model structure complexity of the previous approaches, a combined gradient-based optimization and regression method is proposed in this paper to identify the load model parameters from ambient signals. An open static load model structure in which various static load models can be applied, together with the induction motor as the dynamic load model, are selected as the composite load model structure for parameter identification. Then, the static load model parameters are identified through regression, after which the induction motor parameters can be obtained through optimization with the regression residuals being the objective function. After the transformation of the induction motor model, the objective function is quasiconvex in most of the feasible region so that the gradient-based optimization algorithm can be applied. The case study results in Guangdong Power Grid have shown the effectiveness and the improvement in computation efficiency of the proposed approach.

Index Terms—Load modeling; parameter identification; ambient signals; gradient-based optimization.

I. INTRODUCTION

LOAD modeling has been an important and challenging issue in the analysis of power systems [1]. The accuracy of the power system time-domain simulation results relies on the accuracy of power system models, which means inappropriate power system models may lead to misleading simulation results [2]. Unlike most of the other power system models, load model is an aggregation of multiple different types of power components, which makes load modeling more complicated [3].

Two categories of load modeling methods have been proposed. i.e., the component based approaches and the measurement based approaches [4]. As a more widely used approach, measurement based load modeling is to consider all the components connected to one load bus as a composite load model, and then to identify the parameters from local measurement data with the load model structure pre-selected. In the previous measurement based load modeling approaches, the composite load model with the ZIP model as the static load part and the induction motor as the dynamic load part is widely used, forming the widely known ZIP+IM model [5]. In recent years, with the increasing penetration of renewable energy in the demand side of the power system, the distributed renewable energy components and power electronics load in load modeling is discussed in [6], [7]. The robust time-varying load modeling approach has been proposed in [8]-[10] to continuously track the load model parameters with the previously estimated parameters as the initial values.

In the previous measurement based load modeling approaches, the measurement data mostly used is the post large disturbance response (PLDR) data. Therefore, whether the PLDR based load modeling can be conducted depends on the occurrence of large disturbance events. Then, PLDR based load modeling cannot be frequently conducted because large disturbance events occasionally happen in the power systems. Recently, with the increasing variety of load model components and the increasing integration of uncertain power resources such as the renewables and the demand responses, it is more necessary to track the time-varying property of the load models, which cannot be achieved by the PLDR based load modeling.

In this paper, an ambient signals based load modeling approach is proposed to track the time-varying property of the load models. Ambient signals refer to the small disturbances contained in the power and voltage signals during power system daily operation [11]. With the ambient signals always existing in power system measurements, the ambient signals based load modeling can be conducted at any time without the dependence on the occurrence of large disturbance events. In this way, load models can be identified periodically to better track the time-varying changes of the load models. A common weakness in previous work about ambient signals based load modeling is that the static load model structure is simplified as constant Z to reduce the computation complexity [12], [13].

In previous load modeling approaches, optimization algorithms have been widely used to identify the load model parameters. Since the objective function of the induction motor model has been validated to be non-convex [13], the heuristic algorithms which have a stronger global optimum searching capability in the non-convex problems have been widely applied in previous work, such as genetic algorithm [5], support vector machines [4], simulated annealing [2], and differential evolution [6], [13]. Nevertheless, if the optimization problem can be approximately convexified, the computation efficiency can be significantly improved by adopting the gradient-based optimization algorithms.

In this paper, a new ambient signals based load model parameter identification algorithm is proposed, in which the regression method is combined with the gradient-based optimization algorithm to identify the load model parameters. The contribution of this paper is summarized as follows. Firstly, compared with the previous ambient signals based load
modeling approaches, an open static load model structure instead of the constant Z model is applied as the static load part of the composite load model, in which the ZIP model, the exponential model and so on can all be applied. Secondly, through the transformation of the induction motor model, the objective function of the newly defined decision variables is quasiconvex in most part of the feasible region, after which the gradient-based optimization algorithm can be applied in the identification of induction motor parameters for the first time. Moreover, the proposed approach can provide high quality initial iteration values that are required by the robust time-varying load modeling methodology [8]-[10] to further improve its applicability.

The rest of this paper is organized as follows. The load model structure for parameter identification is introduced in Section II. In Section III, the basic idea and framework of ambient signals based load modeling are proposed. In Section IV, the regression method to identify the static load model is proposed. The optimization method to identify the induction motor parameters is given in Section V. In Section VI, the case study results are presented to validate the proposed load model parameter identification algorithm. Section VII concludes this paper.

II. LOAD MODEL STRUCTURE

A. Composite Load Model Structure

![Load model structure](image)

Fig. 1 Load model structure of measurement based load modeling

For a bus in a power system, there are four measurement signals which are essential in power system dynamic analysis, i.e., the bus voltage magnitude (U), the bus voltage phase angle (θ), the active power consumption (P) and the reactive power consumption (Q). The function of load models is to calculate P and Q under a given group of load bus voltage phasors (U and θ), as shown in Fig. 1.

Generally, there are two parts of load models in a load model structure, i.e., the static load models and the dynamic load models. The P and Q of the static load models are only related to the current U and θ, which is only related to U in most static load models. Therefore, the relationship between P, Q and U of the static load models is described by algebraic equations. In contrast, the P and Q of the dynamic load models is not only related to the current U and θ, but also to the past states of the model. Therefore, the relationship between P, Q and U, θ of the dynamic load models is described by differential equations.

B. Dynamic Load Model: Induction Motor

1) Third-order Induction Motor Model

The third-order induction motor model is derived from the T-shape equivalent circuit [8]. In this equivalent circuit, Xi is the stator reactance, Xs is the rotor reactance, Xm is the excitation reactance, R is the rotor resistance and s is the slip ratio. The parameters can be simplified as X=Xi+Xm, Xs=Xi+sXm, Td=ωm/(ωmR), where X is the rotor open circuit reactance, Xs is the rotor transient reactance, Td is the rotor open-circuit time constant, ωm is the synchronous rotation angular speed. The third-order state-space formulae of the induction motor are given as follows,

\[
\begin{align*}
\frac{dE_d}{dt} &= -\frac{X}{T_{eq}}E_d + \frac{1}{T_{eq}} - 1)U_d \\
\frac{dE_q}{dt} &= -\frac{X}{T_{eq}}E_q - \frac{1}{T_{eq}} - 1)U_q \\
\frac{ds}{dt} &= \frac{1}{H_1}(T_e - T_f)
\end{align*}
\]

where \( E=E_d+jE_q \) is the phasor of the electromotive force, \( E_d \) is the d-axis electromotive force, \( E_q \) is the q-axis electromotive force, \( U_d \) is the d-axis component of the voltage phasor, \( U_q \) is the q-axis component of the voltage phasor, \( T_m \) is the mechanical torque, which is assumed to be constant during the identification period in this paper. \( T_e \) is the electromagnetic torque, which is calculated as follows,

\[
T_e = \frac{E_d U_q - E_q U_d}{X}
\]

Then, the power consumption \( P \) and \( Q \) can be calculated from \( E_d, E_q, U_d \) and \( U_q \) as follows, which form the output formulae of the induction motor:

\[
\begin{align*}
P &= \frac{E_d U_q - E_q U_d}{X} \\
Q &= \frac{U_d^2 + U_q^2 - U_d E_q - U_q E_d}{X}
\end{align*}
\]

2) Model Transformation

In this section, the third-order induction motor model is transformed to reduce the number of parameters to be identified. Firstly, new state variables are defined, i.e., \( E_d/X \) and \( E_q/X \), to replace the original state variables \( E_d \) and \( E_q \). Then, by defining the new parameters \( a = \frac{X}{T_{eq}} \) and \( b = \frac{X}{T_{eq}} \), the state space formulae are further transformed as follows,

\[
\begin{align*}
\frac{dE_d/X}{dt} &= -b(E_d/X) + aU_d \\
\frac{dE_q/X}{dt} &= -aE_q/X + aU_q \\
\frac{ds}{dt} &= \frac{1}{H_1}(T_e - U_d(E_d/X) + U_q(E_q/X))
\end{align*}
\]

It can be observed that the state space formulae of the induction motor third-order model have been simplified compared with the original form. Then, the next task is to calculate \( P \) and \( Q \) from the newly defined state variables \( E_d/X \) and \( E_q/X \). From (3), it can be observed that the first part of \( Q \), \( (U_d^2 + U_q^2)/X^2 \), has the same form with the static load models. If this part of \( Q \) is regarded as the reactive power consumption of the static load part, the rest part of \( Q \), together with \( P \), can be calculated from \( U_d, U_q, E_d/X \) and \( E_q/X \) as follows,

\[
\begin{align*}
P &= (E_d/X)U_q - (E_q/X)U_d \\
Q &= -U_d(E_d/X) - U_q(E_q/X)
\end{align*}
\]

In this way, the number of parameters is reduced. There are four parameters to be identified, i.e. \([a \ b \ H_1 \ T_m]\).

3) Standardization of Parameters

The parameters’ ranges are needed in the optimization process. However, in the induction motor model, the parameters’ ranges cannot be determined because the amount of induction motor load still impacts the parameters’ ranges. For example, if \( N \) same induction motors are paralleled, the parameters will
change as follows: \( X_0 = X/N, \ X'_0 = X'_N, \ H_{2, N} = H_N, \ T_{m, N} = T_m N, \ P_N = P N, \ Q_N = Q N \), where the subscribe \( N \) means the parameters’ equivalent values after parallel. Then, the parameters’ values will change as follows: \( a_N = a N, \ b_N = b, \ H_{2, N} = H \), \( T_{m, N} = T_m N \).

To deal with this problem, the definition of a standardized induction motor is proposed, which refers to the induction motors with \( T_m = 1 \). Then, the induction motors with \( T_m = N \) can be regarded as the parallel of \( N \) same standardized induction motors. In this way, the parameters’ ranges of \( a \) and \( H_2 \) are no longer impacted by the change of the total amount of the induction motor load. After testing 130,000 different randomly generated induction motors based on the original parameters’ ranges in [5], the ranges of \( \{a, b, H_2\} \) can be approximately given as follows: \( \{a, b, H_2\} \in [10, 3, 0.5], [80, 30, 20] \). The range of \( T_m \) is within \([0, P]\), where \( P \) is the total active power.

C. Static Load Model: Open Structure

According to the results of the survey in [14], there are two commonly used static load models, i.e., the ZIP model and the exponential model. In the ZIP model, the relationship between \( P, Q \) and \( U \) is described as follows,

\[
\begin{align*}
P &= PU^2 + PU + P_a \\
Q &= QU^2 + QU + Q_a
\end{align*}
\]

where there are altogether six parameters to be identified, i.e. \( P_a, P_b, P_c, Q_a, Q_b \) and \( Q_c \). In the exponential model, the relationship between \( P, Q \) and \( U \) is described as follows,

\[
\begin{align*}
P &= PU^{20} \\
Q &= QU^{15}
\end{align*}
\]

where there are four parameters to be identified, i.e. \( P_e, k_p, Q_e \) and \( k_Q \). The open static load model structure means both two static load models will be considered and identified in the identification process, and the one with better accuracy will be selected. Any other static load models whose parameters can be identified through regression can also be included.

III. AMBIENT SIGNALS BASED LOAD MODELING: BASIC IDEA AND FRAMEWORK

A. Ambient Signals based Load Modeling

As a branch of measurement based load modeling, ambient signals based load modeling is to identify the load model parameters from ambient signals measurement data. The spectrum of power system ambient signals is distributed within the bandwidth between 0.2 and 2.0 Hz, which belongs to the range of power system electromechanical dynamics [11]. Therefore, it is reasonable to ambient signals data to identify the load model parameters by analyzing the relationship between the changes in voltage signals and the changes in power signals.

The advantage of the ambient signals based load modeling over the PLDR based load modeling has been discussed in the introduction part, which is the achievement of more frequent and periodical identification of load models. In this section, the limitation is discussed. Since the disturbance magnitudes of ambient signals are relatively smaller, the ambient signals based approach is trying to build the load model based on the partial dynamics around the operating point. However, some load model properties can only be excited by large disturbance. Therefore, only the dynamics which can be excited in small disturbance can be reflected by the load model built from ambient signals. For example, the renewables and the power electronics load can only be identified as constant power loads.

Nevertheless, the load model identified from ambient signals is still practically useful even with the limitation for the following two reasons. Firstly, it is still better to apply the periodically updated load models which are identified from ambient signals than to apply the historical load models which are identified from PLDR because more time-varying changes can be tracked. Secondly, the proportion of the load with different dynamic properties under large disturbance is limited. For the traditional static and induction motor load, the models under both large and small disturbance situations are the same. Since these two parts of load take a very large proportion of the total power load in most common cases [14], the accuracy of the load model built from ambient signals is guaranteed.

B. Identification Framework

In this paper, a combined optimization and regression identification framework is designed to identify the load model parameters from ambient signals, as shown in Fig. 2.

The identification is conducted based on the measurement data of \( U, \theta, P_m \) and \( Q_m \), where \( P_m \) and \( Q_m \) are the measured active and reactive power consumption of the load bus. The parameters of the induction motor, \( \{a, b, H_2, T_m\} \), are selected as the decision variables of the optimization problem. For a given group of decision variables, the induction power consumption can be predicted based on the voltage measurement data \( U, \theta \), the prediction results of which are \( P_{p, im} \) and \( Q_{p, im} \).
and \( Q_{p,im} \). By deducting the predicted induction motor power consumption from \( P_m \) and \( Q_m \), the rest part is regarded to be consumed by the static load, which is the power for static load model regression. After the regression is conducted, the optimal static load model parameters for this group of decision variables is obtained. Then, the power prediction error can be calculated from the residuals of the regression, which is just the objective function of the optimization (where \( n \) is the data length).

To conclude, both the predicted error objective function and the static load model parameters depend on the induction motor parameters, which are just the decision variables of the optimization problem. Once the optimal induction motor parameters are got through optimization, the corresponding regression results of the static load model parameters will then be selected as the identification results of the static load model.

IV. IDENTIFICATION OF STATIC LOAD PARAMETERS: REGRESSION

A. Prediction of Induction Motor Power

| Decision Variables: | Predicted Induction Motor Power | Measured Voltage |
|---------------------|---------------------------------|-----------------|
| \( \{a, b, H_2, T_m\} \) | \( P_{p,im}, Q_{p,im} \) | \( U, \theta \) |

Fig. 3 Prediction of induction motor power

In order to predict the power consumption of an induction motor, three parts of data are required, i.e., the measured voltage, the induction motor parameters, and the initial state. In this case, the measured voltage can be obtained from measurement data, and the induction motor parameters are the decision variables. However, the initial state of the induction motor is unknown, the estimation of which is necessary.

From simulation experience, we have found out a fact which is useful to estimate the initial state of the induction motor. The fact is that, for two induction motors with the same voltage curves and the same parameters but different initial states, the dynamics of the state variables and the power consumption will be the same after about 0.5s. Based on this, the initial state of the induction motor is estimated by setting the differential formulae to be 0, which means the induction motor starts from a stable state. In addition, the first 0.5s data is not used for identification because the estimation of state variables has not been accurate during this period. In this way, the initial state of the induction motor is estimated, after which the power consumption \( P_{p,im} \) and \( Q_{p,im} \) can be predicted.

B. Regression of Static Load Model Parameters

After the prediction of \( P_{p,im} \) and \( Q_{p,im} \), the rest parts in the measured \( P_m \) and \( Q_m \) are regarded as the power consumption of the static load. Then, the static load model parameters can be identified through regression. The regression process is designed for the open static load model structure, in which various static load models are considered, and the one with more accurate regression results will be selected.

There are two steps in the regression process. In the first step, the regression is conducted on both the ZIP model and the exponential model, respectively, and the regression results and the residuals of both models are recorded. For the ZIP model, the active power related parameters are identified through \((P_P, P_P, P_P, r_{p1}) = \text{regress}(P_m - P_{p,im}, [1, U, U^2])\), where \( P_m - P_{p,im} \) is the dependent variable of the regression, \([1, U, U^2]\) are the independent variables of the regression. \( P_E \) is the coefficient of 1, \( P_2 \) is the coefficient of \( U \), \( P_3 \) is the coefficient of \( U^2 \), and \( r_{p1} \) is the series of regression residuals. Similarly, \((Q_p, Q_p, Q_p, r_{q1}) = \text{regress}(Q_m - Q_{p,im}, [1, U, U^2])\). For the exponential model, \((P_p, k_p), (r_{p2}) = \text{regress}(P_m - P_{p,im}, [1, \ln(U)])\) and \((Q_p, k_q), (r_{q2}) = \text{regress}(Q_m - Q_{p,im}, [1, \ln(U)])\).

![Fig. 4 Regression of static load model parameters](image)

In the second step, the regression results of the model with the minimal average power squared differences (APSDs) are selected as the identification results. To make the comparison, it is necessary to transform the regression residuals into the APSDs. For the ZIP model, it is simple that \( P_m - P_{p,im} - P_{p,sl} = r_{p1} \) and \( Q_m - Q_{p,im} - Q_{p,sl} = r_{q1} \). For the exponential model, \( r_{p2} = \text{ln}(P_m - P_{p,im}) - \text{ln}(P_{p,sl}) \), then \( r_{p2} \) can be estimated as \( (P_m - P_{p,im} - P_{p,sl}) / (P_m - P_{p,im}) \) according to Taylor Expansion. In this way, the power residuals can be estimated as \( P_m - P_{p,im} - P_{p,sl} = r_{p1}(P_m - P_{p,im}) \) and \( Q_m - Q_{p,im} - Q_{p,sl} = r_{q1}(Q_m - Q_{p,im}) \). Afterwards, the APSDs can be compared and the models with smaller APSDs are selected as the identification results of the static load models.

Apart from obtaining the static load model parameters, another critical task in the regression step is to provide the objective function for induction motor parameters optimization. Based on the regression residuals, the objective function is calculated as follows, where \( n \) is the length of the measurements,

\[
\begin{align*}
R_{r_{p,ex}} &= \frac{r_{p,ex}}{n}, R_{r_{q,ex}} = \frac{r_{q,ex}}{n} \\
R_{r_{p,xr}} &= \frac{r_{p,xr}}{n}, R_{r_{q,xr}} = \frac{r_{q,xr}}{n} \\
(\sum(P_m - P_{p,im} - P_{p,sl})^2 + \sum(Q_m - Q_{p,im} - Q_{p,sl})^2) / n &= (\min(R_{r_{p,ex}}, R_{r_{p,xr}}) + \min(R_{r_{q,ex}}, R_{r_{q,xr}}))
\end{align*}
\]

V. IDENTIFICATION OF DYNAMIC LOAD PARAMETERS: OPTIMIZATION

A. Decision Variables and Objective Function

In the optimization process, the decision variables are the induction motor parameters, \([a, b, H_2, T_m]\). The objective function has been calculated in the previous section, which is the APSDs after the prediction of induction motor power and the regression of static load model parameters. Then, the task in this section is to obtain the decision variables which can minimize the APSDs. This group of decision variables is just the identification results of the induction motor parameters, and
the regression results of this group of decision variables are the identification results of the static load model parameters.

![Decision Variables and Objective Function](image)

**Fig. 5** Decision variables and objective function of optimization

### B. Quasiconvexity

In this section, the concept of quasiconvexity is introduced. The quasiconvexity is defined as follows: a function \( f:S \subset \mathbb{R}^n \) defined on a convex subset \( S \) of a real vector space is quasiconvex if for all \( x, y \in S \) and \( \lambda \in [0,1] \), the following inequality holds:

\[
f(\lambda x + (1-\lambda) y) \leq \max\{f(x), f(y)\}
\]

This definition means that it is always true for \( f \) that a point directly between two other points does not give a larger function value than both other points do. For a quasiconvex function, if a local minimum exists, then it is ensured to be the global minimum. Based on these properties, iterative gradient-based optimization algorithms can be applied. The quasiconvexity of the objective function of the induction motor parameters will be tested in the case study section to ensure the rationality of applying gradient-based optimization algorithm.

### C. SQP Optimization Algorithm

Sequential quadratic programming (SQP) is a typical gradient-based iterative method for constrained nonlinear optimization problems (NOP). In this paper, SQP is selected to solve the optimization problem of the induction motor parameters. The procedures of the iteration process can be summarized as follows. Firstly, an initial feasible solution \( x^0 \) should be generated according to the constraints, which are the ranges of the load model parameters in the problem of this paper. Then, in each iterative step \( k, k \in \mathbb{N} \), the NOP is approximately modeled as a quadratic programming (QP) subproblem in the neighborhood region of the iterate \( x^k \). After the QP subproblem is solved in the neighborhood region, the solution can be used to construct a new iterate \( x^k \). With the increase of \( k \), the iterate sequence \( (x^k)_{k \geq 0} \) converges to a local minimum \( x^* \) of the NOP. If the NOP is quasiconvex, then any local minimum is ensured to be the global minimum. In this way, the global minimum of the NOP can be solved.

### VI. CASE STUDY

#### A. System Introduction

The 500kV network of the Guangdong Power Grid is used for the load model identification and validation, the structure of which is given in Fig. 6. This system includes 83 buses and 97 lines. The load models of 20 load buses are identified. The other load buses are also set to be the composite load model containing a static part and an induction motor. The simulation in this section is conducted in the power system analysis toolbox [15] in Matlab. The simulation time length of one case is 10s, with the time step being 0.01s. The base value of the system capacity is 100MVA. To generate the ambient signals in the system, practically measured power ambient signals are penetrated in the active power consumption of 30 load buses whose load models are not identified. 50 groups of ambient signals are penetrated to generate 50 simulation cases. In different simulation cases, the load models of the 20 load buses whose models are identified are also changed. Then, there are altogether 1000 load model identification cases.

![Structure of Guangdong Power Grid](image)

**Fig. 6** Structure of Guangdong Power Grid

#### B. Algorithm Validation: An Example

Firstly, an example of how the load model parameters are identified from ambient signals through the proposed method is given. The load model on Bus DG in Case 1 of all the 50 cases is identified in this example. The accurate values of the load model parameters are \([51.25 19.67 3.47 3.76]\) for \([a b H2 Tn]\) (after standardization) and \([1.78 2.33 2.95 0.92 1.71 -3.24 2.17 1.18 -2.02 1.92]\) for \([P, P, P, P, kp, Q, Q, Q, Q, Q, Q, Q, Q]\). It should be noted that the static load model in simulation is a combination of ZIP and exponential models. The measurement curves of Bus DG in Case 1 are given in Fig. 7. To validate the proposed algorithm, the measurement errors are not considered in this section, which will be discussed in the following sections.

1) **Prediction of Induction Motor Power**

In Section IV.A a method to estimate the initial state of the induction motor and then to predict the power consumption is proposed, which is validated here. In this section, the accurate induction motor parameters are used for predicting the induction motor power. In Fig. 8, the real values are the power consumption of Bus DG in Case 1 obtained from the simulation results. In order to get the model predicted values, the initial state is calculated according to (4) by setting differential formulae to be 0. After estimating the initial state, the power consumption can then be predicted from the induction motor parameters and the voltage curve, the results of which are given.
in Fig. 8 as the predicted values. It can be observed that although the estimated initial state is not accurate, the dynamics of the power consumption will be the same after about 0.5s. More similar cases are tested, in which the conclusion still holds. Therefore, the initial state estimation and power prediction method proposed in Section IV. A is validated.

![Fig. 8 Prediction results of induction motor power consumption](image)

2) Regression of Static Load Model Parameters

In this section, examples of the regression of static load model parameters are given based on the induction motor power prediction results. The data from 1s to 10s is used for identification. Since the APSDs are used as the objective function, several examples of the APSD results are given. With the parameters being [70 10 2 5], the APSD is 49.48. With the parameters being [60 15 3 4], the APSD is 2.39*10^4. If the parameters are the accurate induction motor parameters [51.25 19.67 3.47 3.76], the APSD is 3.76*10^8. It can be concluded that with the induction motor parameters approaching the accurate values, the APSD will descend. Therefore, the APSD is suitable to be selected as the objective function.

Apart from calculating the objective function, another task of the regression is to identify the static load model parameters. With the induction motor parameters being the accurate parameters, the results of the APSDs predicted by different static load models are given as follows: \( R_{P, ZIP} = 2.64 \times 10^8 \), \( R_{P, EX} = 3.33 \times 10^8 \), \( R_{Q, ZIP} = 9.82 \times 10^9 \), and \( R_{Q, EX} = 1.60 \times 10^8 \). With smaller APSDs, the ZIP model is selected for the models of both \( P \) and \( Q \), the parameters of which are: \([P, P, P, Q, Q, Q] = [3.01 1.39 3.58 10.22 0.97 1.71]\). Also, the APSD is obtained by adding \( R_{P, ZIP} \) with \( R_{Q, ZIP} \).

3) Optimization of Induction Motor Parameters

![Fig. 9 The logarithm to base 10 of the objective function](image)

In this section, an example of the optimization process is given. Firstly, an initial feasible solution is randomly generated within the ranges of the parameters. In this example case, the initial feasible solution is selected to be \( P_0 = [70 10 2 5] \). By setting \( P_0 = [51.25 19.67 3.47 3.76] \), the values of the logarithm to base 10 of the objective function of \( P = P_0 + k(P_0 - P_0), k \in [0, 1.5] \) are given in Fig. 9. It can be observed that the objective function has only one local optimal solution with the increase of \( k \). The local optimal solution appears at \( k = 1 \), which is just \( P_0 \), the accurate parameters' values. Starting from \( P_0 \) in the SQP optimization process, the process converges at \( P_0 \), which means that the accurate parameters can be identified. It should be noted that the curve in Fig. 9 is not quasiconvex because the values of the logarithm to base 10 are plotted. Further tests about the quasiconvexity of the objective function and the effectiveness of the SQP algorithm will be given in the following sections.

4) Static Load Model Validation under Large Disturbance

The static load model in the simulation process is the combination of a ZIP model and an exponential model. In contrast, the identification results of the static load models are ZIP models for \( P \) and \( Q \). Whether the identified ZIP model can still fit the dynamic performance of the accurate static load model under large disturbance should be validated. Here the \( Q \) consumed by \( X^* \) is included in the static load.

![Fig. 10 Validation of static load model under large disturbance](image)

A validation case is given by setting a three-phase to ground fault at Bus YC at \( t = 1 \)s, which is cleared at \( t = 1.1 \)s. The validation case is conducted in another operation point (OP2), in which the loads of 5 buses are increased by 20%. Firstly, the simulation results of the original case with the accurate static load models are recorded as the real values. Secondly, the load model on Bus DG is replaced by the identification results, and the simulation case is conducted again, the results of which are recorded as the predicted values. The results of the real values and the predicted values are given in Fig. 10, from which it can be observed that the dynamic performance of the identified ZIP models is still very similar to that of the accurate static load model even under large disturbance.

C. Algorithm Test of Quasiconvexity and SQP Algorithm

In this section, the quasiconvexity of the objective function and the effectiveness of the SQP optimization algorithm are further tested under more cases. Firstly, to test the effectiveness of the SQP algorithm, it is applied in all the 1000 load model identification cases. If the identification results of all the parameters are within 1% range of the accurate values, it is regarded as a successful case. According to the identification results, there are 15 failed cases among all the 1000 cases through one-time SQP optimization. Therefore, the successful percentage can be estimated as 98.5%.

![Fig. 11 The logarithm to base 10 of the objective function of a failed case](image)

The 17th simulation case of Bus PC is given as an example of the failed cases, in which the accurate parameters are \( P_{a, err} = [40.08 16.86 1.32 13.94] \), and the identification results
are $P_{	ext{desired}}=57.43 \pm 3.20 \pm 0.72$. The values of the logarithm to base 10 of the objective function of $P=P_{\text{real}}+k(P_{\text{desired}}-P_{\text{real}})$, $k \in [-0.1, 1.1]$ are given in Fig. 11, in which there are more than one local optimal solutions with the increase of $k$.

Then, a further test on the quasiconvexity of the objective function is given. A pair of points $(x_1, x_2)$ are randomly selected within the ranges of the parameters, and then whether $f((x_1+x_2)/2) < \max(f(x_1), f(x_2))$ holds is tested, where $f(x)$ is the objective function. The results show that for more than 96.5% pairs of points, the above inequation holds. Therefore, it can be concluded that in most of the feasible region, the objective function is quasi-convex, and the SQP algorithm can be applied. Several exceptions may exist around the boundaries of the feasible region, such as the case in Fig. 11. The identified values of $b$ and $H_2$ have reached the parameters’ boundaries.

One approach to improve the reliability of the SQP algorithm is to repeatedly conduct the SQP optimization algorithm from different randomly generated initial feasible solutions and select the best solution among multiple SQP results as the optimization results. Then, the effectiveness of the SQP algorithm is tested again by repeating the algorithm for 10 times, which is successful in all the 1000 cases. In the following sections, the results are all obtained through 10 times SQP.

D. Model Identification and Validation Results

In this part, the identification cases are conducted with the measurement errors added to the measurement data. The measurement errors are based on the experimental test results in [16], which include two parts, i.e., the systematical error (the offset) and the accidental error (the random variations).

1) An Example with the Impact of Measurement Error

![Fig. 12 P and Q dynamic response of Bus BJ in the 3rd case](image)

Fig. 12 P and Q dynamic response of Bus BJ in the 3rd case

Firstly, the concept of signal to noise ratio (SNR) is used to describe the relative magnitudes of signal and noise, as follows,

$$SNR = \frac{\text{Energy of signal}}{\text{Energy of noise}} = 10\log \left( \frac{\sum y_i^2}{\sum e_i^2} \right)$$

where $y_i$ is the signal and $e_i$ is the measurement error. In this section, an example of load model parameters identification from the measurement data with errors is given, which is the data of Bus BJ in the 3rd simulation case of all the 50 cases. The real values and the measured values with measurement errors are given in Fig. 12, (a) and (b). Then, a low pass filter with the cut off frequency being 3 Hz is applied on the measured values to restrain the impact of accidental errors. The filtered measured values are also given in Fig. 12, (a) and (b). The SNRs of P and Q in this case are 19.68 and 11.53, respectively.

Afterwards, the identification process is conducted based on the filtered measured values. The accurate parameters of the induction motor are $P_{\text{real}}=[64.02 \pm 22.46 \pm 3.73 \pm 5.51]$, while the identification results are $P_{\text{desired}}=[64.18 \pm 23.28 \pm 3.13 \pm 5.64]$. It can be observed that the measurement errors have led the identification results to deviate from the accurate values. Then, a large disturbance validation case is simulated on OP2, which is the three-phase to ground fault on Bus YC. The comparison between the real values and the model predicted values is given in Fig. 12 (c) and (d). The fitting degree (FD) is used to measure the accuracy of the cases, which is calculated as follows,

$$Fitting\ Degree = 1 - \frac{\sum (y_i - y_a)^2}{\sum (y_m - \text{mean}(y_m))^2}$$

where $y_p$ is the predicted value, $y_a$ is the measured value, and $\text{mean}(y_m)$ is the mean value of $y_m$. In this validation case, the FDs of P and Q are 0.9778 and 0.9383, respectively.

2) Identification Accuracy

In this section, all the 1000 identification cases are conducted under two disturbance levels (DLs). The previous example cases are noted as DL=1. In this section, the 50 simulation cases are conducted again to generate the data for identification, in which the penetrated ambient signals have larger disturbance magnitudes compared with the previous DL=1 cases. Afterwards, the 1000 load models are identified based on the newly generated data, the results of which are noted as DL=2. Since the disturbance magnitudes in DL=2 cases are larger than that in DL=1 cases, the SNRs are also larger.

After obtaining the identification results, they are firstly divided by the accurate parameter values for normalization, which means 1 is the accurate parameter value. The mean SNRs, the mean values ($\mu$) and the standard deviations ($\sigma$) of all the parameters of the 1000 cases for each DL are calculated, which are given in TABLE I. It can be observed that the identification results are still accurate even with the impact of measurement errors. In addition, the identification accuracy is better with a larger DL, judging from the fact that the mean values are getting closer to 1 and the standard deviations are getting smaller.

| TABLE I Mean Values and Standard Deviations of the Identification Results |
| DL | Mean SNR | $\sigma(a)$ | $\mu(a)$ | $\sigma(b)$ | $\mu(b)$ | $\sigma(H_3)$ | $\mu(H_3)$ | $\sigma(T_a)$ | $\mu(T_a)$ |
|----|----------|--------------|----------|--------------|----------|---------------|-------------|--------------|----------|
| 1  | 22.79    | 0.031        | 1.014     | 0.032        | 1.014    | 0.045         | 0.977       | 0.021        | 0.996    |
| 2  | 25.70    | 0.022        | 1.007     | 0.022        | 1.013    | 0.033         | 0.985       | 0.016        | 0.997    |

3) Model Validation under Large Disturbance

The ability of the identification results to predict the dynamic response under large disturbance is validated in this section. Two three-phase to ground faults on Bus YC and Bus YD are tested on OP2. The identification results from two different DLs are both used for validation. In each validation case, the real values are obtained from the simulation results with the accurate load models, while the predicted values are obtained from the simulation results with the load models of the 20 load buses whose models are identified replaced with the identification results. All the 50 simulation cases are tested under the two faults, after which 1000 P curves and 1000 Q
curves with real values and predicted values are obtained.

Then, the FDs of the P and Q curves are calculated. The median FDs, the percentage of the FDs larger than 0.95, and the percentage of the FDs larger than 0.9, are given in TABLE II, for the identification results from different DLs and different faults respectively. It can be observed from the results that the median FDs are very close to 1, and the percentages are also very large. In addition, the validation results of the identified models from DL=2 is more accurate than that from DL=1.

Two examples of the validation cases are given in Fig. 13. The first example is the P curve of Bus XI in the 12th case under YD fault, the model of which is identified with DL being 1. The second example is the Q curve of Bus HL in the 42nd case under YC fault, the model of which is identified with DL being 2. The FDs of these two cases are 0.9478 and 0.9978, respectively.

| DL | Fault | Median | % | % | Median | % | % |
|----|-------|--------|---|---|--------|---|---|
|    |       | FP     | FP>0.9 | FP>0.95 | FP     | FP>0.9 | FP>0.95 |
| 1  | YC    | 0.9968 | 91.7  | 90.1  | 0.9927 | 89.7  | 84.2  |
|    | YD    | 0.9967 | 94.2  | 91.8  | 0.9935 | 90.5  | 84.6  |
| 2  | YC    | 0.9990 | 99.6  | 99.2  | 0.9975 | 99.1  | 96.5  |
|    | YD    | 0.9989 | 95.5  | 92.2  | 0.9976 | 93.4  | 91.4  |

Fig. 13 Two validation examples

E. Comparison with the Previous Approach

Since the Z+IM model structure is applied in previous ambient signals based load modeling work, it cannot be directly compared with the proposed approach. However, the computation efficiency can still be compared. According to the results in [13], it takes about 30s to identify the Z+IM model parameters. In contrast, the average computation time of the 1000 DL=1 cases is 7.78s. Compared with the previous approach, the newly proposed approach can identify the parameters of a more complicated load model structure within shorter computation time. This can be explained from the following two aspects. Firstly, the static load model parameters are identified through regression instead of optimization, through which a more complicated model structure can be applied without increasing the computation cost. Secondly, through the transformation of the induction motor model, the objective function is quasiconvex in most parts of the feasible region, which makes it possible to apply the gradient-based optimization algorithm instead of the heuristic algorithms.

VII. CONCLUSION

This paper proposes a combined gradient-based optimization and regression method for ambient signals based load modeling. An open static load model structure is used to improve the model structure applicability. Then, with the induction motor parameters being the decision variables and the regression residuals being the objective function, the induction motor parameters are identified through gradient-based optimization, while the static load model parameters are identified through regression. The case study results have validated the following conclusions of the proposed approach. Firstly, the objective function is quasiconvex in most parts of the feasible region, so it is reasonable to apply the gradient-based optimization algorithm. Secondly, the proposed approach can accurately identify the load model parameters from ambient signals with the impact of measurement errors considered. Thirdly, the computation efficiency is improved compared with previous ambient signals based load modeling approach.

In our future work, the impact of renewables and power electronics load on the identification accuracy can be analyzed. In addition, machine learning approaches can be applied to provide the initial feasible solution for optimization.

REFERENCES

[1] A. Arif, Z. Wang, J. Wang, B. Mather, H. Bashshaudo, and D. Zhao, “Load modeling – a review,” IEEE Trans. Smart Grid, vol. 9, no. 6, pp. 5986-5999, Nov. 2017.
[2] Y. Zhu and J. V. Milanovic, “Automatic identification of power system load models based on field measurements,” IEEE Trans. Power Syst., vol. 33, no. 3, pp. 3162-3171, May 2018.
[3] 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 modeling,” IEEE Trans. Smart Grid, vol. 10, no. 6, pp. 6102-6114, Nov. 2019.
[4] J. Ma, R. He, and D. J. Hill, “Load modeling by finding support vectors of load data from field measurements,” IEEE Trans. Power Syst., vol. 21, no. 2, pp. 726-735, May 2006.
[5] R. He, J. Ma, and D. J. Hill, “Composite load modeling via measurement approach,” IEEE Trans. Power Syst., vol. 21, no. 2, pp. 663-672, 2006.
[6] X. Zhang, D. J. Hill, and C. Lu, “Identification of composite demand side model with distributed photovoltaic generation and energy storage,” IEEE Trans. Sustain. Energy, vol. 11, no. 1, pp. 326-336, Jan. 2020.
[7] C. Wang, Z. Wang, J. Wang, and D. Zhao, “SVM-based parameter identification for composite ZIP and electronic load modeling,” IEEE Trans. Power Syst., vol. 34, no. 3, pp. 182-193, Jan. 2019.
[8] C. Wang, Z. Wang, J. Wang, and D. Zhao, “Robust time-varying parameter identification for composite load modeling,” IEEE Trans. Smart Grid, vol. 10, no. 1, pp. 967-979, Jan. 2019.
[9] M. Cui, J. Wang, Y. Wang, R. Diao, and D. Shi, “Robust time-varying synthesis load modeling in distribution networks considering voltage disturbances,” IEEE Trans. Power Syst., vol. 34, no. 6, pp. 4438-4450, Nov. 2019.
[10] J. Zhao, Z. Wang, and J. Wang, “Robust time-varying load modeling for conservation voltage reduction assessment,” IEEE Trans. Smart Grid, vol. 9, no. 4, pp. 3304-3312, July 2018.
[11] J. W. Pierre, D. J. Trudnowski, and M. K. Donnelly, “Initial results in electromechanical mode identification from ambient data,” IEEE Trans. Power Syst., vol. 12, no. 3, pp. 1245-1251, Aug. 1997.
[12] Y. Zhang, C. Lu, Y. Han, S. Xu, J. Wang, H. Huang, and Y. Su, “Ambient signal based load model parameter identification using optimization method,” in Proc. IEEE PES General Meeting, Denver, CO, 2015, pp. 1-5.
[13] Y. Zhang, C. Lu, and X. Zhang, “Applicability comparison of different algorithms for ambient signal based load model parameter identification,” Int. J. Elec. Power Energy Syst., vol. 111, pp. 382-389, Oct. 2019.
[14] J. V. Milanovic, K. Yamashita, J. Martinez Villanueva, S. Z. Djkic, and L. M. Korunovic, “International industry practice on power system load modeling,” IEEE Trans. Power Syst., vol. 28, no. 3, pp. 3038-3046, Aug 2013.
[15] F. Milano, L. Vanfretti, and J. C. Morataya, “An open source power system virtual laboratory: The psat case and experience,” IEEE Trans. Edu., vol. 51, no. 1, pp. 17-23, Feb. 2008.
[16] X. Zhang, C. Lu, Y. Wang, and J. Lin, “Experimental measurement of PMU error distribution and its impact on load model identification,” in IEEE PES General Meeting, Boston, 2016, pp. 1-5.