Power System Oscillation Mode Prediction Based on the LASSO Method

WEIKE MO\(^1\), (Student Member, IEEE), JIAQING LV\(^2\), MIROSŁAW PAWLAK\(^{2,3}\), (Member, IEEE), UDAYA D. ANNAKKAGE\(^3\), (Senior Member, IEEE), AND HAOYONG CHEN\(^{1,1}\), (Senior Member, IEEE)

\(^1\)School of Electric Power, South China University of Technology, Guangzhou 510641, China
\(^2\)Department of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering, AGH University of Science and Technology, 30-059 Kraków, Poland
\(^3\)Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, MB R3T 2N2, Canada

Corresponding author: Haoyong Chen (eehychen@scut.edu.cn)

Weike Mo was supported by the National Natural Science Foundation of China (51937005). Jiaqing Lv was supported by the Polish National Centre of Science under Grant DEC-2018/31/N/ST7/03977.

\textbf{ABSTRACT} This paper utilizes modern statistical and machine learning methodology to predict the oscillation mode of interest in complex power engineering systems. The damping ratio of the electromechanical oscillation mode is formulated as a function of the power of the generators and loads as well as bus voltage magnitudes in the entire power system. The celebrated Lasso algorithm is implemented to solve this high-dimension modeling problem. By the nature of the $L_1$ design, the Lasso algorithm can automatically render a sparse solution, and by eliminating redundant features, it provides desirable prediction power. The resultant model processes a simple structure, and it is easily interpretable. The precision of our sparse modeling framework is demonstrated in the context of an IEEE 50-Generator 145-Bus power network and an online learning framework for the power system oscillation mode prediction is also provided.

\textbf{INDEX TERMS} Small-signal stability, electromechanical oscillations, system identification, mode damping prediction, Lasso, sparse modeling machine learning, online learning.

\section{I. INTRODUCTION}
Inter-area oscillations are one of the main concerns of small-signal stability in power systems [1]–[4]. These oscillations are inherent to power systems and can limit the power exchanged between areas if they are insufficiently damped [5]. Conventionally, the oscillation mode information can be calculated with the power system model linearized on a specific operation point (equilibrium point) [6]. However, this method is not practical in further online applications because the system changes on different operating points, and the correlation between the damping of an electromechanical oscillation mode and operating scenario is not explicit. Our goal in this paper is to bridge this gap by investigating how real-time changes in operating scenarios impact the damping of the oscillation mode – the prediction of mode damping on different operating scenarios.

Power system oscillation mode prediction is a helpful tool for day-ahead planning activities and online generator re-dispatch. Using the correlations between the damping of an electromechanical oscillation mode and operating scenarios, transmission system operators (TSOs) can optimize the 24h ahead load flow operating points with the desired power exchanged between areas and sufficient damping. On the other hand, by using this correlation and the operation data from the Phasor Measurement Units (PMUs), TSOs can monitor the electromechanical oscillation mode in real-time, and adjust system operation (generator re-dispatch) in an emergency for satisfied damping.

The aforementioned problem becomes increasingly attractive as a result of the increasing quantity of available measurements, such as Wide Area Monitoring Systems (WAMS). Various works have attempted to explore the possibility of establishing a bridge between the mode damping and the system operating condition. Moreover, these works split into two categories:

1) Find the direct relationship $f(\cdot)$ between the mode damping $\zeta$ and the system operation point $P_0$, that is $\zeta = f(P_0)$;

2) Find the sensitivity $g(\cdot)$ of the mode damping $\zeta$ on the system operation point $P_0$, that is $\Delta \zeta = g(\Delta P)\big|_{P_0}$, which is called power system online sensitivity identification (OSI).
between these two categories, the former approach dominates the other. For instance, linear regression was implemented in [7], [8] to identify the damping sensitivity at an operation point, while weighted linear regression was employed in [9] to find critical variables in the identification procedure. Apart from the regression-type methods, in [10] the problem was formulated as a pattern classification problem, and decision trees were used to predict well or poorly damped oscillations using power flow data. In [11], neural networks were also applied to find the relationship between system variables and a dominant mode.

Some works concentrating on the power system online sensitivity identification were also brought to our attentions. A locally weighted linear regression (LWLR) method was applied in [12] to predict damping ratio of a dominant mode online. A k-nearest neighbors locally weighted ridge regression was used in [13] to predict electromechanical oscillation mode. Furthermore, the locally weighted ridge regression integrated with norm-2 Tikhonov-Phillips regularization was examined in [14] to overcome the collinearity problem encountered in power system electromechanical oscillation mode prediction.

However, most of these papers were dealing with relatively small-size power systems. Actually, when the power system of interest is of hundreds of buses or even more, the aforementioned methods such as multiple linear regression or ridge regression can hardly work, due to the well-known “curse of dimensionality” phenomenon [15] known to data scientists and statisticians. In order to overcome this problem, in this paper, we apply a sparse machine learning algorithm called the Lasso (least absolute shrinkage and selection operator), and statisticians. In order to overcome this problem, in this paper, we apply a sparse machine learning algorithm called the Lasso (least absolute shrinkage and selection operator), which was widely applied to power system community, such as power system transient stability problems [16], [17], total transfer capability online estimation [18], and voltage stability margin online monitoring [19]. This paper examines this approach in assessing, understanding, and predicting the small-signal behavior of large interconnected power systems and estimates the direct relationship between the mode damping and the system operation point. The validity of our approach is demonstrated by an IEEE 50-Generator 145-Bus power network, and the prediction accuracy is compared with two competing methods recently applied in the field. Namely, they are the multiple linear regression, and the ridge regression methods.

This paper is organized as follows: Section II presents a brief overview of the relationship between electromechanical oscillation mode damping and operating scenario. System identification methods using the Lasso are introduced in Section III. Test cases and rules for generating multiple random power flow study cases for extensive eigenvalue analysis and eventually for system identification are discussed in Section IV. Results are presented and discussed in the context of two dominant electromechanical modes in Section V. Section V also shows the extension to apply our methodology in the online power system oscillation mode prediction. Section VI concludes the method on this paper.

II. THE RELATIONSHIP BETWEEN ELECTROMECHANICAL OSCILLATION MODE DAMPING AND OPERATING SCENARIO

To explain the relationship between electromechanical oscillation mode damping and operating scenario clearly, we use the classical electromechanical model [6], [20]–[22] with loads modeled as constant impedances. Consider an n-Machine, N-Bus power system. Machine $i$ is modeled as a constant voltage $E_i$ behind a transient reactance $x_{di}^t$. The motion of the $i$-th machine rotor angle $\delta_i$ is modeled as

$$m_i \ddot{\delta}_i = P_{mi} - P_{el} = P_{mi} - \frac{E_i V_i \sin(\delta_i - \theta_j)}{x_{di}^t} - D_i \dot{\delta}_i$$

$$= f_i(\delta, \omega, \dot{\theta}_j)$$

$\dot{\delta}_i = 2\pi f_0 \dot{\omega}_i$ (1)

where

$$V_j = \sqrt{V_{j, re}^2 + V_{j, im}^2}$$

$$\theta_j = \tan^{-1}\left(\frac{V_{j, im}}{V_{j, re}}\right)$$

$P_{mi}$ is the input mechanical power, $P_{el}$ is the output electrical power, $V_{j, re}$ and $V_{j, im}$ are the real and imaginary parts of the bus voltage phasor at bus $j$ which is the terminal bus of Machine $i$. $\dot{\theta}_j = V_j / \omega_m = V_{j, re} + jV_{j, im}$ is the load bus voltage phasor, $\delta_i$ is the $i$-th machine angle, $\omega_m$ is the $i$-th machine speed, $m_i = 2H_i$ where $H_i$ is the inertia of the $i$-th machine most commonly used in power system stability simulation programs, and $D_i$ is the $i$-th machine’s damping coefficient. Note that $j$ denotes the imaginary sign if it is not used as an index.

For each load bus $i$, the power flow equation is:

$$\sum_{k=1, k \neq i}^{N} (\dot{V}_i - \dot{V}_k) \left( \frac{V_i}{R_{lik} + jX_{lik}} \right)^* + V_i^2 G_i - P_{el}$$

$$+ j[V_i^2(B_i - \frac{B_{lik}}{2}) - Q_{el}] = g_i(\delta, V) = 0$$

(5)

where $R_{lik}$, $X_{lik}$, and $B_{lik}$ are the resistance, reactance, and line charging, respectively, of the line connecting buses $i$ and $k$. Moreover, $P_{el}$ and $Q_{el}$ are the $i$-th generator’s active and reactive electrical output power, respectively, if bus $i$ is a generator bus. $G_i$ and $B_i$ are the load conductance and susceptance at bus $i$. $V$ is the $N$-vector of the load bus voltage phasors. Equations (1), (2), and (5) can be combined into a vector form

$$M \dot{\omega} = f(\delta, \omega, V)$$

$$\dot{\delta} = 2\pi f_0 \dot{\omega}$$

$$0 = g(\delta, V)$$

(8)

where $M$ is the diagonal machine inertia matrix, $f$ is a vector of acceleration torques, and $g$ is the power flow equation of the transmission network. We linearize (8) on the operating scenario $\mathcal{P}_0(\delta_0, V_0)$ (steady-state/equilibrium point) to obtain
the linear model

\[
\mathbf{M} \Delta \omega = \left( \frac{\partial f(\delta, \omega, V)}{\partial \delta} + \frac{\partial f(\delta, \omega, V)}{\partial \omega} + \frac{\partial f(\delta, \omega, V)}{\partial V} \right) \bigg|_{\mathcal{P}_0} \\
= K_1 \Delta \delta + D \Delta \omega + K_2 \Delta V
\]

\[
\Delta \dot{\delta} = 2 \pi f_0 \Delta \omega = C \Delta \omega
\]

\[
0 = \frac{\partial g(\delta, V)}{\partial \delta} \bigg|_{\mathcal{P}_0} + \frac{\partial g(\delta, V)}{\partial V} \bigg|_{\mathcal{P}_0} = K_3 \Delta \delta + K_4 \Delta V
\]

(9) (10) (11)

where 0\(\Delta \delta\) is an \(n\)-vector of the machine angle deviations from \(\delta_0\), \(\Delta \omega\) is an \(n\)-vector of the machine speed deviations from \(\omega_0\), and \(\Delta V\) is an \(N\)-vector of the load bus voltage phasor deviations from \(V_0\). \(\mathbf{D}\) is an \(n \times n\) diagonal matrix with elements being the damping coefficients. \(\mathbf{C}\) is an \(n \times n\) diagonal matrix with elements being the constant \(2\pi f_0\). The \(n \times n\) matrix \(\mathbf{K}_1, \mathbf{K}_2, \mathbf{K}_3, \) and \(\mathbf{K}_4\) are related to the partial derivatives of the power transfer between the machines and their terminal buses. The \(N \times N\) matrix \(\mathbf{K}_4\) is the admittance matrix of the network. \(\mathbf{K}_1, \mathbf{K}_2, \mathbf{K}_3, \) and \(\mathbf{K}_4\) are nonsingular, we can substitute (11) into (9) to eliminate \(\Delta V\) to obtain a linearized electromechanical model reduced to the machine internal nodes:

\[
\begin{bmatrix}
\frac{\partial \omega}{\partial \delta} \\
\frac{\partial g(\delta, V)}{\partial \delta} \\
\frac{\partial g(\delta, V)}{\partial V}
\end{bmatrix} = \begin{bmatrix}
\mathbf{M}^{-1} \mathbf{D} & \mathbf{M}^{-1} \mathbf{K} \\
\mathbf{C} & 0_{n \times n}
\end{bmatrix}
\begin{bmatrix}
\Delta \omega \\
\Delta \delta
\end{bmatrix} = \mathbf{A}_{2n \times 2n} \begin{bmatrix}
\Delta \omega \\
\Delta \delta
\end{bmatrix}
\]

(12)

where

\[
\mathbf{K} = \mathbf{K}_1 - \mathbf{K}_2 \mathbf{K}_4^{-1} \mathbf{K}_3
\]

(13)

We denote \(\mathbf{A}_{2n \times 2n}\) as system state matrix in power system small-signal analysis. The damping ratio \(\zeta_i\) of each mode \(\lambda_i\) (for \(i = 1, 2, \ldots, n\)) is determined by the state matrix \(\mathbf{A}_{2n \times 2n}\), which is formulated by physical laws described above. As long as the grid topology is steady, the value of \(\mathbf{A}_{2n \times 2n}\) is determined by component parameters (such as matrix \(\mathbf{M}^{-1} \mathbf{D}\)) and system operating conditions (matrix \(\mathbf{M}^{-1} \mathbf{K}\)). Therefore, in the case of a developed power grid whose grid topology and physical parameters are rarely changed, it is reasonable that the following link holds:

\[
\zeta_i \leftarrow \lambda_i \leftarrow \mathbf{A}_{2n \times 2n} \leftarrow \mathbf{K} \leftarrow (\mathbf{K}_1, \mathbf{K}_2, \mathbf{K}_3, \mathbf{K}_4^{-1}) \leftarrow \mathcal{P}_0
\]

(14)

where \(\mathcal{P}_0\) is the operating scenario. In other words, given an operating scenario \(\mathcal{P}_0\), \((\mathbf{K}_1, \mathbf{K}_2, \mathbf{K}_3, \mathbf{K}_4^{-1}) \leftarrow \mathbf{K} \leftarrow \mathbf{A}_{2n \times 2n} \leftarrow \lambda_i \rightarrow \zeta_i\) can be computed numerically step by step. And the corresponding \(\zeta_i\) is unique to the operating scenario \(\mathcal{P}_0\). Therefore, the link in (14) can be reformulated as

\[
\zeta_i = f(\mathcal{P}_0)
\]

(15)

where \(f(\cdot)\) is a function with input \(\mathcal{P}_0\) and output \(\zeta_i\).

The eigenvalue calculation of a matrix can refer to root calculation of this matrix’s characteristic polynomial. Galois’ theory indicates that there is no solution in radicals to general polynomial equations of degree five or higher with arbitrary coefficients [23]. This means that the eigenvalue of a \(n\) order matrix, for \(n \geq 5\), cannot be solved algebraically. Therefore, examining the inverse of matrix \(\mathbf{K}_4\) and the eigenvalue calculation of matrix \(\mathbf{A}\), it’s impossible to get the explicit symbolic expression of the function \(f(\cdot)\) with input \(\mathcal{P}_0\) and output \(\zeta_i\).

Thus the relationship between equilibrium point \(\mathcal{P}_0\) and \(i\)-th mode’s damping ratio \(\zeta_i\) can extend to a data-driven damping ratio prediction problem. The machine learning algorithms we use in modeling \(\zeta_i = f(\mathcal{P}_0)\) will be described in Section III.

III. SYSTEM IDENTIFICATION METHODS USING THE LASSO

In system identification, the training data of the form

\[
D_n = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}
\]

is available, where \(x_i \in \mathbb{R}^p\) is the \(p\)-dimensional input variable, whereas \(y_i \in \mathbb{R}^1\) is the output variable. It is worth noting that the input variable is also sometimes referred to as the “observational variable”, and the output variable is called the “response variable”. Then system identification tries to learn the mapping between \(x_i\) and \(y_i\), i.e., \(f : \mathbb{R}^p \rightarrow \mathbb{R}^1\).

In the power system oscillation mode prediction problem, the input variable \(x_i = [x_{i1}, \ldots, x_{ip}]\) corresponds to the voltage magnitudes together with the active and reactive power of the generators and loads in the entire power system, whereas the output variable \(y_i\) corresponds to the damping ratio of a certain oscillation mode of interest. The aim of using artificial intelligence methods is to learn the relationship between them, so that with new observations, one can use the model to predict the responses accordingly, and the prediction error would be as small as possible. In order to achieve this task, one may use a large class of possible models. It is, however, practical to start with a simple yet very useful model based on the linear mapping

\[
y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i,
\]

(16)

where \(\varepsilon_i\) is the noise process. Here we refer to the discussions in the Subsection V-D for more detailed justification of this simple model in the application of ultra-high-dimensional learning problems.

For the model in (16), one can write

\[
\mathbf{Y} = \begin{bmatrix}
\mathbf{Y}_1 \\
\vdots \\
\mathbf{Y}_n
\end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix}
1 & x_{11} & \cdots & x_{1p} \\
\vdots & \vdots & \ddots & \vdots \\
1 & x_{n1} & \cdots & x_{np}
\end{bmatrix}, \quad \mathbf{\beta} = [\beta_0, \beta_1, \ldots, \beta_p]^	op, \quad \mathbf{\varepsilon} = [\varepsilon_1, \ldots, \varepsilon_n]^	op
\]

\[
\text{then the linear model (16) is equivalent to}
\]

\[
\mathbf{Y} = \mathbf{X} \mathbf{\beta} + \mathbf{\varepsilon},
\]

(17)

The most straightforward approach to solve (16) is through minimizing the mean square error

\[
\hat{\mathbf{\beta}}_{LS} = \arg \min_{\mathbf{\beta}} \frac{1}{n} ||\mathbf{Y} - \mathbf{X} \mathbf{\beta}||_2^2,
\]

(18)
where \( \frac{1}{n} ||Y - X\beta||_2^2 \) is equivalent to \( \frac{1}{n} \sum_{i=1}^{n} (Y_i - \beta_0 - \sum_{j=1}^{p} \beta_j X_{ij})^2 \), and the solution is the well-known least square (LS) solution [24] expressed by

\[
\hat{\beta}_{LS} = (X^T X)^{-1}(X^T Y).
\]

The least-square formula provides a solution to the linear model, yet it naturally has the drawback of numerical instability, i.e., a small change in the data, such as when new observations come to be available, may result in a large variation of the estimate \( \hat{\beta} \). Also, the matrix \( X^T X \) might not be full rank, creating a problematic scenario in the matrix inversion process.

In order to overcome the defects in the classical least square approach, the so-called ridge regression was developed in the 1970s [25], [26]. It is to minimize the least square criterion together with a penalty term on the weights. Namely,

\[
\hat{\beta}_{ridge} = \arg \min_{\beta} \left( \frac{1}{n} ||Y - X\beta||_2^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right),
\]

where \( \lambda \) is a regularization parameter that needs to be specified. It controls the amount of shrinkage in the minimization criterion. In practice, one can select \( \lambda \) according to some re-sampling methods, i.e., cross validation.

It is worth mentioning that the ridge regression also has close-form solution

\[
\hat{\beta}_{ridge} = (X^T X + n\lambda I_p)^{-1}(X^T Y),
\]

where \( I_p \) is a \( p \times p \) identity matrix.

The ridge regression method processes the numerical stability in the solution and has been widely applied in engineering and applied science. It was applied in the small signal stability problem in [12].

In the late 1990s, the so-called “least absolute shrinkage and selection operator” (Lasso) was developed [27]. It is also based on minimization of the least square criterion with a penalty term. However, the penalty term is the \( L_1 \) norm of the weights, rather than the \( L_2 \) norm case as in the ridge regression scenario. Specifically, the Lasso has the following form

\[
\hat{\beta}_{Lasso} = \arg \min_{\beta} \left( \frac{1}{n} ||Y - X\beta||_2^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right),
\]

where \( \lambda \) is the regularization term that needs to be specified.

It is worth mentioning that according to the dual optimization theory, the ridge regression and the Lasso also have the following equivalent primal forms, respectively,

\[
\hat{\beta}_{ridge, primal} = \arg \min_{\beta_{|\beta|_2 < s}} \left( \frac{1}{n} ||Y - X\beta||_2^2 \right),
\]

\[
\hat{\beta}_{Lasso, primal} = \arg \min_{\beta_{|\beta|_1 < s}} \left( \frac{1}{n} ||Y - X\beta||_2^2 \right),
\]

where \( || \cdot ||_2 \) and \( || \cdot ||_1 \) denote the \( L_2 \) norm and the \( L_1 \) norm. Specifically, \( ||\beta||_2 = (\sum_{j=1}^{p} \beta_j^2)^{1/2} \) and \( ||\beta||_1 = \sum_{j=1}^{p} |\beta_j| \).

Furthermore, \( s \) is a term that can be expressed alternatively in terms of \( \lambda \).

The difference between Lasso and ridge regression can be shown in the following illustration in Fig. 1 [28].

In Fig. 1, it shows an example of a bivariate case. The sum of squares is shown, and the shaded area corresponds to the weight constraints. The left panel corresponds to the \( L_2 \) case, and the shaded area is a circular shape. In higher dimensional cases, it would be a hyper-ball. The right panel corresponds to the \( L_1 \) case, and the shaded area is a rectangular shape. It can be seen that sometimes the optimal solution can happen at the corner points in the Lasso case, suggesting that part of all the features can be set to zero weight. Such a scenario can never happen in the ridge case, meaning that ridge regression can never render a sparse solution. In high-dimension problems, it is known that the proper sparse solution can greatly enhance the model prediction accuracy due to the well known “curse of dimensionality” [15] in data science. As a consequence, it is expected that in our damping ratio prediction problem, the implementation of the Lasso algorithm can greatly increase the model prediction performance.

Unfortunately, the solution of Lasso does not have closed-form formula as the least square case (19) or the ridge case (21). In practice, after specification of regularization parameter \( \lambda \), people use coordinate optimization method to calculate the solution for Lasso. This algorithm is shown in Algorithm 1, and it is also called the “shooting algorithm” for the Lasso.

Note that in Algorithm 1, \((c)_+ \) equals to \( c \) if \( c > 0 \), and equals to 0 otherwise. Besides the \( \hat{\beta}_j, j = 1, \cdots, p \), \( \beta_0 \) can be estimated by taking the empirical mean of the sample \( \{Y_1, \cdots, Y_n\} \). It is also worth mentioning, in order for each feature to have equal strength, the user should standardize the data so that each covariate of \( X \) (except for the first column) should have zero mean and unit variance. This can be done by pre-processing data when the user subtracts the mean and then divides by the standard deviation for each feature. It is worth mentioning that it is standard practice for statistical researchers to pre-process data in such a way, before applying the machine learning / system identification methods.

In practice, the specification of the regularization parameter \( \lambda \) plays a critical role. It can be seen that larger \( \lambda \) will
shrink more weights to zero. In practice, one can use cross-validation methods to select $\lambda$. Namely, on a grid of candidate $\lambda$ values, estimate the model according to Algorithm 1, find the cross-validation errors, and specify in the final modeling the value of $\lambda$ according to the one leads to the smallest cross-validation error.

**IV. DATA GENERATION**

**A. GENERATING MULTIPLE POWERFLOW CASES FOR EIGENVALUE ANALYSIS**

The system base case powerflow is randomly adjusted to simulate the real power system. The training data were generated in the following way to randomly generate powerflow cases whilst not exceeding the system limits that maintain realistic operation of the components. First, every load is randomly adjusted within $\pm 25\%$ variation from the base case. For the $k$-th steady-state point, active and reactive power consumption on the $i$-th load bus are given by

$$P_{L}^{(i)}(k) = P_{L0}^{(i)}(1 + \Delta P_{L}^{(i)}(k)),$$

$$Q_{L}^{(i)}(k) = Q_{L0}^{(i)}(1 + \Delta Q_{L}^{(i)}(k)),$$

where $P_{L0}^{(i)}$ and $Q_{L0}^{(i)}$ are nominal active and reactive power setting of the $i$-th load bus in the base case. Moreover, $\Delta P_{L}^{(i)}$ and $\Delta Q_{L}^{(i)}$ are the percentage of perturbation of active and reactive power allowed in the $i$-th load bus. In this case they are both $25\%$. Also $\epsilon_{PL}^{(i)}(k), \epsilon_{QL}^{(i)}(k)$ are i.i.d. uniformly distributed random variables on the interval $(-1, 1)$ and this is denoted as $U(-1, 1)$.

Secondly, the active power injection and voltage at the $i$-th generator bus are given by the result of the optimal power flow (OPF) with the loads’ new active and reactive power. The generator cost data are provided by Matpower software package [29]. If the optimal power flow converges, the OPF solution will be the $k$-th steady-state point.

After the $k$-th steady-state point is carried out, the initial value of each component can be calculated to form the system state matrix. So we can get the eigenvalue and eigenvector matrix of the system state matrix.

**B. MODE TRACKING**

Let row vector $W_{i}^{0}$ and column vector $V_{i}^{0}$ denote the normalized left and right eigenvector of the specific mode $\lambda_{i}$ at the base case $P_{0}$, respectively. After each adjustment of the loads, the optimal power flow is solved, and an eigenvalue analysis is performed. Then we can denote column vector $V_{i}^{*}$ as the normalized right eigenvector of mode $\lambda_{k}$ ($k = 1, 2, \ldots, n$) at the new steady-state point $P_{0}^{*}$.

For one specified electromechanical mode $\lambda_{i}$, the following method has been implemented to make sure the mode has been tracked correctly after each steady-state point $P_{0}^{*}$ was produced:

1) Exclude those modes whose highest participation factors are not the angle and speed states;
2) Use a reasonable range of values to bound the frequency of the electromechanical mode $\lambda_{i}$ (for instance, we can use $0.5 \pm 0.2$ Hz on the 0.5-Hz mode), and exclude those modes which fall outside this range;
3) Calculate $\hat{k} = \arg \min_{k=1,2,\ldots,m} ||W_{i}^{0}V_{k}^{*} - 1||$, where $m$ is the number of mode set after excluding on Step 1 and 2, arg min is argument of the minimum. $\lambda_{\hat{k}}$ is the specific electromechanical mode we want to track.

For the aforementioned Step 3, in an equilibrium point, the normalized left and right eigenvectors corresponding to different eigenvalues are orthogonal, $W_{i}V_{j} = 0$ for $i \neq j$. In the cases that eigenvectors correspond to the same eigenvalue, we have $W_{i}V_{i} = 1$. Physically speaking, the right eigenvectors $V_{i}^{0}$ and $V_{i}^{*}$ of the same electromechanical mode $\lambda_{i}$ at the power base $P_{0}$ and the new steady-state point $P_{0}^{*}$ respectively, must be similar, i.e., $V_{i}^{0} \approx V_{i}^{*}$. Thus, according to the aforementioned expression $W_{i}^{0}V_{i}^{0} \approx 1$, we have $W_{i}^{0}V_{i}^{*} \approx 1$.

Therefore $\hat{k} = \arg \min_{k=1,2,\ldots,m} ||W_{i}^{0}V_{k}^{*} - 1||$ is the number of the mode $\lambda_{\hat{k}}$ we want to track at the new steady-state point $P_{0}^{*}$.

**C. DATE RECORD**

For each adjustment of the loads, the optimal power flow is solved, and an eigenvalue analysis is performed. The active and reactive power generation, as well as the bus voltage magnitudes (the bus voltage angles are not recommended as features) are recorded. The frequency and damping ratio of specific electromechanical mode are also recorded.

**V. SIMULATION STUDIES**

**A. IEEE 50-GENERATOR 145-BUS SYSTEM TEST CASE**

The performance of our proposed modeling methodology is evaluated using an IEEE 50-Generator 145-Bus system test case. Its network topology structure is shown in Fig. 2. This test system consists of 145 buses, 50 generating units (7 machines use sub-transient reactance model and 43 machines use classical model), 64 impedance loads, 453 transmission lines. We refer to [29] for a detailed description of this power system.
In this 145 bus system, we have 373 features:
1) Features #1 to #145 correspond to magnitudes of bus voltage.
2) Features #146 to #195, and Features #196 to #245 correspond to active power and reactive power of generator units sorted by bus number, respectively.
3) Features #246 to #309, and Features #310 to #373 correspond to active power and reactive power of loads sorted by bus number, respectively.

Four hundred and ninety-nine (499) different operating scenarios are generated by the method proposed in Section.IV. Two different modes following are adapted to demonstrate this method. These two modes are:
1) The intra-area 1.3-Hz mode related to Generators within Area-1.
2) The inter-area 0.33-Hz mode in which generators in Area-2 oscillates against those in Area-3.

We denote these modes by Mode 1 and Mode 2. It is worth noting that Mode 1 is a simple oscillation mode corresponding to an intra-area oscillation, while Mode 2 is a more complexed oscillation mode corresponding to an inter-area oscillation. In the following parts, we refer to Mode 1 as the simple mode, and Mode 2 as the hard mode. We use the first 250 cases as the training set, and the rest 249 cases as the testing set.

### B. THE INTRA-AREA MODE

From the small-signal analysis, one can see that the 1.3-Hz mode is an intra-area mode in which generators on bus-98, 100 and 103 oscillate against generator on bus-89 within Area-1. The top 10 participation factor states of this mode are shown in Table 1, and the mode shape is shown in Fig. 3.

After the proper selection of regularization parameter $\lambda$ described in Section.III, the final model consists of only 8 features. The strength of these features are visually demonstrated in Fig. 4. We refer to the beginning of Section V-A for the description of these features. Besides, we also show these features in terms of their physical meaning in Table 2. Furthermore, Fig. 5 plots the intra-area 1.3-Hz mode’s damping ratio versus the top 6 dominant features using the sample data.

**Table 1.** Top 10 participation factor states of the intra-area 1.3-Hz mode on the power base.

| State | $\omega_1$ | $\omega_2$ | $\omega_3$ | $\omega_4$ | $\omega_5$ |
|-------|------------|------------|------------|------------|------------|
| PF(1) | 1          | 1          | 0.82       | 0.82       | 0.36       |

**FIGURE 3.** The mode shape for the intra-area 1.3-Hz mode on the power base.

By applying the model to the testing data set, we find the root mean square (RMSE) testing error is 0.00007359, or 0.007359%. Alternatively, if we evaluate the testing error by the mean absolute error (MAE), it is 0.00005518, or 0.005518%. The comparison between the first 100 predicted values and their true values are shown in Fig. 6. We can see that there is almost no noticeable difference between the two sets of values.

From Fig. 5, the correlation between intra-area 1.3-Hz mode damping and the top 6 dominant features look consistent with the weight of corresponding features acquired.
by Lasso in Table 2. Moreover, the results from Lasso indicate that the active power of generators on bus-89 and bus-108 (close to bus-98) are negatively correlated with the intra-area 1.3-Hz mode damping, which are compatible with the premise derived from small-signal analysis that this mode is an intra-area mode in which generators on bus-98, 100 and 103 oscillate against generator on bus-89 within Area-1.

The intra-area modes are typically determined by only a few number of machines. This explains why we observe very accurate prediction results. Our modeling also identify those most important features automatically.

C. THE INTER-AREA MODE

From the small-signal analysis, one can see that the 0.33-Hz mode is an inter-area mode in which the generators in Area-2 (generators on bus-128, 104, 105, 121, etc) oscillate against those in Area-3 (generators on bus-135, 136, 137 and 139). The top 10 participation factor states of this mode are shown in Table 3 and the mode shape is shown in Fig. 7.

After the proper selection of regularization parameter $\lambda$ described in Section III, the final model consists of only 24 features. The strength of these features are visually
FIGURE 8. The weights in the final model for the Case 2.

TABLE 4. Top 10 dominant features for the Case 2.

| No. | Features                                      | Weights |
|-----|-----------------------------------------------|---------|
| 1   | Active power of generator on bus-131         | -0.3672 |
| 2   | Reactive power of generator on bus-121       | -0.1933 |
| 3   | Reactive power of generator on bus-115       | 0.1307  |
| 4   | Active power of load on bus-141             | 0.1162  |
| 5   | Active power of load on bus-134             | 0.1150  |
| 6   | Active power of generator on bus-136        | -0.1072 |
| 7   | Active power of load on bus-128             | -0.0988 |
| 8   | Reactive power of generator on bus-67        | -0.0986 |
| 9   | Active power of load on bus-136             | -0.0671 |
| 10  | Active power of load on bus-139             | -0.0553 |

FIGURE 9. The inter-area mode damping versus the top 6 dominant features: $P_{G, bus-131}$, $Q_{G, bus-121}$, $Q_{G, bus-115}$, $P_{L, bus-141}$, $P_{L, bus-134}$, $P_{G, bus-136}$.

Consistent with the weight of corresponding features acquired by Lasso in Table 4. And results from Lasso indicate that active and reactive power of generators/loads within Area-2 and Area-3 are correlated with the inter-area 0.33-Hz mode damping, which are compatible with the premise derived from small-signal analysis that this mode is an inter-area mode in which generators in Area-2 oscillate against those in Area-3.

D. DISCUSSIONS AND COMPARISON WITH OTHER ALGORITHMS

In Section V-C, we have applied the Lasso algorithms in the damping ratio prediction problem of an inter-area oscillation mode (a hard mode). In comparison, we also apply the multiple linear regression and the ridge regression using all the 373 candidate features. The results are shown in Table 5.

TABLE 5. Comparison with other methods.

| Methods            | Multiple linear regression | Ridge regression  | Lasso |
|--------------------|----------------------------|-------------------|-------|
| RMSE               | 342.6                      | 1.70%             | 1.34% |
| MAE                | 69.6                       | 1.16%             | 0.83% |
| # Features         | 373                        | 373               | 24    |

From Table 5, we can see that multiple linear regression can hardly do the work in our problem, in which the dimensionality equals to 373. This confirms the discussions on its drawbacks in Section III. Actually, when this method is applied in [7], the researchers only apply the modeling using a subset of all the features based on their subjective choice originated from the industry experience. Comparing to this method, the Lasso can automatically decide which features should exist in the final model, thus circumvent the subjectivity.

Our method also outperforms the ridge regression approach. It is conjectured that in larger power networks (e.g., thousands of buses, or more), the superiority of our method will be better demonstrated.

So far we have applied high-dimensioned linear modeling in the power system damping mode prediction problem.
One may come to the natural question how good is it to represent an unknown nonlinear system by a linear model. Generally speaking, such representation will incur an irreducible error called “approximation error”, and [30] shows that under a wide range of conditions, the approximation error is quite affordable. This is reflected by the fact that our final modeling result. A 1.34% RMSE for the Inter-area Mode and a negligible prediction error for the Intra-area Mode have achieved the modeling accuracy confirming the acceptable level of the approximation error due to the linear model. As matter of fact, the “curse of dimensionality” phenomenon [15] suggests that in ultra-high-dimensionality problem, the dimension-reduction is always much more crucial compared to the adoption of linear structures to represent the general nonlinear system. This reflects the merits of our paper.

E. IMPLEMENTATION IN ONLINE LEARNING

So far we have introduced the Lasso approach in the power system oscillation mode prediction problem. Generally speaking, the classical Lasso approach is an offline methodology. Nonetheless, Combing with the so-called Kalman filter [31], this approach can be used for online learning scenarios.

The Kalman filter is a type of techniques that allow the model to be constantly updated, as new training data come to availability. Due to the recursive nature of the Lasso shotting algorithm, the Kalman filter for Lasso might be hard to derive. On the other hand, since the Lasso technique quite often reduces the dimensionality dramatically, one can always use Lasso as the first step, and on the reduced feature space, apply Kalman filter version of classical linear regression instead.

With a bit of abuse of notations, let

\[
Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & X_{1,1} & \cdots & X_{1,q} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n,1} & \cdots & X_{n,q} \end{bmatrix} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix},
\]

\[
\beta = [\beta_0, \beta_1, \ldots, \beta_q]^{\top},
\]

where \{(1),(2),\ldots,(q)\} are a subset of index, and have been selected optimally by the Lasso method. Let \(\hat{\beta}\) be the model based on dataset \(D_n = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}\). Now new data comes to the availability, then

\[
Y' = \begin{bmatrix} Y_1 \\ \vdots \\ Y_{n+1} \end{bmatrix}, \quad X' = \begin{bmatrix} 1 & X_{1,1} & \cdots & X_{1,q} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n,1} & \cdots & X_{n,q} \end{bmatrix} = \begin{bmatrix} X_1 \\ \vdots \\ X_{n+1} \end{bmatrix}.
\]

\(Y'\) and \(X'\) are the updated data sets. Let \(\beta'\) denote the updated model parameters. Then following Equation (19):

\[
(X'^{\top}X')\beta' = X'^{\top}Y'.
\]

Then we can easily derive \((X'^{\top}X + X_{n+1}^{\top}X_{n+1})\beta' = X'^{\top}Y + X_{n+1}^{\top}Y_{n+1} \). Since \(X'^{\top}X\beta = X'^{\top}Y\).

Then \(\beta' = [X'^{\top}X + X_{n+1}^{\top}X_{n+1}]^{-1}(X'^{\top}X\beta' + X_{n+1}^{\top}Y_{n+1})\) is the online update model.

VI. CONCLUSION

In this paper, we have applied the Lasso method to model the power system oscillation mode damping prediction problem. The sparse nature of our methodology enables the modeling to be applied in large-scale power systems, and automatic feature selection is achieved simultaneously with the regression process. The performance of the method is evaluated using the IEEE 50-Generator 145-Bus testing case. Our method shows a stronger prediction accuracy comparing to two competing methods recently applied in the field. The correlations between the electromechanical oscillation mode damping and operating scenarios can enable the power engineering users to monitor and control some power system quantities to ensure the desirable level of damping, even if she/he processes little background understanding about the system. Importantly, the automatic feature selection property of the proposed approach can largely reduce the range of these power system quantities affecting the specific oscillation mode in a large power network. The framework of industry online application using the Lasso algorithm was shown in Fig. 11. Instead of using Prony analysis at the first step in Fig. 11, some novel power system oscillation mode estimation methods [32]–[34] can also calculate the power oscillation mode of interest from power system sampled signal in real-time.

![Diagram of industry online application framework](image-url)

FIGURE 11. The framework of industry online application.

In a practical system, noises exist in measurement data. Therefore, the sensitivity analysis of the training data can be the further work of this study.

ACKNOWLEDGMENT

(Weike Mo and Jiaqing Lv contributed equally to this work.)
MIROSLAW PAWLAK (Member, IEEE) received the Ph.D. and D.Sc. degrees in computer engineering from the Wroclaw University of Technology, Wroclaw, Poland. He is currently a Professor with the Department of Electrical and Computer Engineering, University of Manitoba, Canada. His research interests include statistical signal processing, machine learning, and nonparametric modeling. Among his publications in these areas are the books Image Analysis by Moments (Wroclaw Univ. Technol. Press, 2006), and Nonparametric System Identification (Cambridge Univ. Press, 2009), coauthored with Dr. Greblicki. He has been an Associate Editor of the Journal of Pattern Recognition and Applications, Pattern Recognition, the International Journal on Sampling Theory in Signal and Image Processing, Opuscula Mathematica, and the Statistics in Transition-New Series.

UDAYA D. ANNAKKAGE (Senior Member, IEEE) received the B.Sc. Eng. degree in electrical engineering from the University of Moratuwa, Moratuwa, Sri Lanka, in 1982, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of Manchester Institute of Science and Technology, Manchester, U.K., in 1984 and 1987, respectively. He is currently a Professor with the University of Manitoba, Winnipeg, MB, Canada. He was the Head of the Electrical and Computer Engineering Department, University of Manitoba, from 2008 to 2012. He is currently the Convener of CIGRE Working Group on Application of Phasor Measurement Units for monitoring power system dynamic performance. His research interests include power system stability and control, security assessment and control, operation of restructured power systems, and power system simulation. He was an Editor of the IEEE TRANSACTIONS ON POWER SYSTEMS from 2009 to 2012.

HAOYONG CHEN (Senior Member, IEEE) received the B.S., M.S.E., and Ph.D. degrees in electrical engineering from Xi’an Jiaotong University, in 1995, 1997, and 2000, respectively. He was an Influential Professor of electrical engineering (in both academia and industry). He has 14 years’ research and teaching experience as a University Faculty Member of electrical engineering. He is currently the Director of the Institute of Power Economics and Electricity Markets, South China University of Technology. He has been the Principal of many key national projects of China and has close connections with the Chinese Power Industry. His current research interests include excellent research record on modeling/optimization/control of power systems and electricity markets.