Improvement of Electromechanical Mode Identification from Ambient Data with Stochastic Subspace Method

Abstract—Stochastic subspace method utilizing ambient data to identify electromechanical modes acts as a useful approach to monitoring power system oscillation modes in real time. This paper proposes an approach based on stochastic subspace method and clustering algorithms to detect modes automatically with damping, frequency and mode shape considered. To improve the performance of mode identification further, a revised algorithm is also introduced, which can be divided into two parts. By estimating modes with different signals and clustering the results, the first part increases the estimation accuracy. By identifying modes in each signal and performing clustering technology to the results, the second part can detect missing modes observable in many measurements. IEEE 39-bus system is adopted to illustrate the effectiveness of the algorithms proposed.

Index Terms—Ambient data, Clustering algorithm, Electromechanical mode identification, Stochastic subspace method.

I. INTRODUCTION

Electromechanical modes with weak or even negative damping will bring great challenges and hazards to the safety and stability of power systems. Identifying low-frequency modes and corresponding mode shapes in real time can keep abreast of many characteristics of transient process under small perturbation, such as attenuation factor, frequency and the distribution of oscillatory transition. Such information can be of great help in judging whether the system is stable, calculating the attenuation velocity of oscillation, providing early warning information and formulating strategies to suppress oscillations.

The approaches to estimating modes can be divided into two categories: one is based on mathematical models and the other is based on measurement signals, which can be further separated into methods utilizing ring down data and methods using ambient data. Model based methods first establish the accurate dynamic model of the power system, then linearize it at the operating point, and finally carry out eigenvalue analysis to identify the oscillation modes [1]. Although all modes and all the characteristics of the modes can be obtained by these methods, there are still some issues: 1) the computational burden increases dramatically with the scale of system increasing; 2) it is difficult to model the system accurately; 3) the model and its parameters cannot keep up with the variation of the system in real time. Thus these methods are not suitable for on-line identification [2].

There are many kinds of methods based on ring down data, among which the Prony method [3] is one of the most successful approaches. But in practice, ring down data caused by large disturbances are rare and it is unrealistic to carry out large disturbance experiments in operating systems. Therefore, these methods can only identify oscillation modes after large disturbances occur and then give warnings, but cannot evaluate the dynamic stability of the system during normal operation.

Small stochastic disturbances such as random fluctuations in loads which are assumed to be white noise [4] will result in ambient data. Thus different from ring down data, ambient data is always available in power systems. There are several methods using ambient data to conduct system estimation. Methods based on ARMA model estimate the parameters of ARMA model or AR model by using least squares method [5], ITD [6], etc. The estimation accuracy of such methods is sensitive to sampling rate and noise, and in practical applications, the effect is often not as good as that of NEXT-ERA [7] and stochastic subspace method [8], which are based on state space model. Stochastic subspace method constructs the measurements into a matrix, and then identifies the system parameters from the row and column spaces of the matrix [8]. In addition to the above methods, the modes identification methods based on ambient data include peak picking technique [9], RDT [10], etc.

Among these methods, stochastic subspace method is one of the most mature methods in the field of system identification. With good robustness [11] and strong anti-noise performance, it has been successfully applied in civil engineering and other fields. In the field of power systems, it has been used in modes identification [12], parameter determination of PSS [13], oscillatory stability limit prediction [14] and other respects. Although stochastic subspace identification (SSI) has been widely used, there are still some drawbacks: 1) this method may produce spurious modes, which will confuse the extraction of real modes; 2) some modes may be missing; 3) the estimation error is relatively big in spite of the fact that this method has higher accuracy compared to other methods based on ambient data.

This paper aims at reducing the impacts of the above problems, improving the performance of mode identification, with the general idea to cluster a large number of identification results. The estimation results of SSI vary with the preset system orders and selected input signals, and a large number of results can be obtained by choosing different orders and input measurements. Conducting cluster technologies to these results can help discriminate real modes from spurious modes, find out missing modes and improve the identification accuracy.

The structure of this paper is organized as follows: In Section II, the modal identification using covariance-driven stochastic subspace identification (SSI-COV) is reviewed. In
Section III, an algorithm realizing automatic modes identification based on SSI is developed. In Section IV, a revised algorithm is forged to heighten the identification performance. In Section V, simulation data are used to evaluate the performance of the proposed approaches. Finally, Section VI concludes the paper.

II. STOCHASTIC SUBSPACE IDENTIFICATION

In this section, we describe the principle of SSI-COV. Firstly, a power system can be represented by an $n$-order discrete-time state-space model [2]

$$
\begin{align*}
    x_{k+1} &= Ax_k + w_k \\
    y_k &= Cx_k + v_k
\end{align*}
$$

where $x_k \in \mathbb{R}^{n \times 1}$, $y_k \in \mathbb{R}^{m \times 1}$ are state, output vectors, respectively; $A$ and $C$ are state and output matrices, respectively; $w_k \in \mathbb{R}^{n \times 1}$ and $v_k \in \mathbb{R}^{m \times 1}$ are the process noise and the measurement noise, respectively. $w_k$ and $v_k$ are assumed to be zero-mean Gaussian white noise and they are not correlated with each other. The covariance of output vectors is

$$
A_r = E[y_{k+1} y_k^T] = \lim_{j \to \infty} \frac{1}{j} \sum_{k=0}^{j-1} y_{k+1} y_k^T.
$$

It can be proved that

$$
A_r = CA^{-1}G
$$

where $G = E[x_{k+1} x_k^T]$. In reality, it is impossible to obtain infinite number of measurements, so an estimation of $A_r$ will be used as (4)

$$
A_r = \frac{1}{j} \sum_{k=0}^{j-1} y_{k+1} y_k^T
$$

where $j$ is the length of time series. Using the Hankel matric

$$
Y_{0 \nu_{i-1}} =
\begin{bmatrix}
    y_0 & y_1 & \cdots & y_{j-1} \\
    y_1 & y_2 & \cdots & y_j \\
    \vdots & \vdots & \ddots & \vdots \\
    y_{2i-1} & y_{2i} & \cdots & y_{2i+j-2}
\end{bmatrix}
$$

where $i$ is a parameter determined by users, the Toeplitz matric can be calculated as (6)

$$
T_{i,j} = \frac{1}{j} Y_{2i-1} Y_{0j-1}^T =
\begin{bmatrix}
    A_i & A_{i-1} & \cdots & A_1 \\
    A_{i+1} & A_i & \cdots & A_2 \\
    \vdots & \vdots & \ddots & \vdots \\
    A_{2i-1} & A_{2i-2} & \cdots & A_i
\end{bmatrix}
$$

Conduct SVD decomposition to $T_{i,j}$, we get:

$$
T_{i,j} = U_d S_d V_d
$$

where $S_d$ is a diagonal matrix composed of nonzero singular values. Compare (6) and (7), and $O_i$ can be obtained:

$$
O_i = U_d S_d i^{1/2}
$$

Then state matrix and output matrix can be calculated:

$$
A = O_i (1:m \times (i-1), :)^T O_i (m+1:m \times i, :)
$$

$$
C = O_i (1:m, :)
$$

Where $O_i$ denotes the Moor-Penrose pseudo-inverse. By eigenvalue analysis of $A$, we get the discrete eigenvalues $\lambda_i$ and the corresponding right eigenvectors $\phi_i$. Then the mode damping and frequencies can be obtained by (11) and the mode shapes can be calculated by (12)

$$
\lambda_i = \frac{\ln \lambda_i}{\Delta t}
$$

$$
V_i = C \phi_i
$$

where $\Delta t$ is the sampling interval; $\lambda_i$ are the eigenvalues in continuous time domain; $V_i$ are the mode shapes.

III. AUTOMATIC MODES IDENTIFICATION

An algorithm combining SSI and clustering technologies to detect low-frequency modes automatically is proposed in this section.

Due to the finite number of data samples $j$ and the fact that $w_k$ and $v_k$ may not be white noise, those singular values who should have been zero in theory are not zero, making it difficult to determine the system order. So it is usual to preset different orders and then apply SSI. By this way, massive estimation results can be obtained, among which there are real modes as well as spurious modes. In [16] and [20], clustering techniques are used to deal with these results, realizing automatic extraction of true modes. However, mode shapes being not considered in these two papers, the algorithms may confuse two modes with similar damping and frequencies but different mode shapes. The algorithm proposed in this section also employs clustering algorithms to deal with SSI results, and mode shapes will be considered.

A. Using of Clustering Algorithms

Considering the characteristics of the results of SSI and the performance of different clustering algorithms, we utilize DBSCAN [19] to deal with the damping and frequency estimation results while MAFIA [15] is applied to the mode shape estimates.

Increase preset system order $n$ from $n_{\text{min}}$ to $n_{\text{max}}$ ($n$ requires to be even), and for each $n$, $n/2$ mode estimation results will be obtained by SSI. Draw these results on the two-dimensional plane (called stabilization diagram) with damping plotted on the horizontal axis and frequency on the vertical axis, where those results belonging to a real mode will concentrate in a small area while most spurious modes scatter. Because real modes will appear in the results as the order changes, the number of results standing for one real mode will get close to $(n_{\text{max}} - n_{\text{min}})/2 + 1$. Due to the reasons above, it is easy to determine the input parameters of DBSCAN, and DBSCAN will show good clustering performance under such dataset. By proper selection of the input parameters, real modes will be presented in the form of the clusters of DBSCAN, while most spurious modes will be judged as noisy data.
Mode shapes are m-dimensional vectors, whose dimensions grow as the number of measured signals increases. The accuracy of mode shape estimation is inferior to that of damping and frequency estimation in SSI. Therefore, for even those results belonging to the same modes, the shape estimation results are not tightly concentrated, and there are even some dimensions irrelevant. That is why this paper chooses MAFIA who handles high dimensional data well and uncovers clusters in different subspaces to deal with the mode shape estimation results.

B. Procedure of the Approach

The algorithm firstly removes the trends in the signals. Secondly, SSI is performed in different orders to get massive results. Next, DBSCAN is applied to find out results alike in damping and frequency. And then, conduct MAFIA to find out results with not only similar damping and frequencies, but also similar mode shapes. Finally, average these similar results as the final estimation results. For convenience, call this method SSI-Clustering.

1) Data preprocessing. Besides ambient data, measured signals contain trends, which are useless for SSI. To advance SNR, the detrending technology in [17] is adopted to remove trends.

2) Apply SSI. Vary system order n from \( n_{\text{min}} \) to \( n_{\text{max}} \), and for each n, conduct SSI to the output signals preprocessed. Then we will get massive estimation results. Considering that in practice electromechanical modes with lower damping are more important, only modes with damping ratios below 6% and frequencies between 0.1 and 2.5 Hz can remain, and other results will be weed out.

3) Classify results according to damping and frequency. Draw those results screened on the stabilization diagram, and conduct DBSCAN to interpret the stabilization diagram. Then the results in the same cluster produced by DBSCAN will have similarity in damping ratio and frequency.

4) Consider mode shapes. Firstly, mode shapes should be normalized. And then for those estimates with close damping ratios and frequencies obtained in the previous step, employ MAFIA to their mode shape vectors. To save the time, only several dimensions with the largest magnitudes will be selected to be the inputs of MAFIA.

5) Get final estimation results. For each cluster found in step 3) and step 4), average those estimates in them as the final results.

IV. APPROACH IMPROVEMENT

To improve the identification function, a revised algorithm is forged in this section, which can be split into two parts, with the first part aiming at increasing the accuracy and the second part intended for detecting missing modes.

A. Improving Accuracy

A and C produced by SSI are only estimates, but not the exact state and output matrices. Consequently, error exists in the mode estimation results. Choosing different measured signals as the inputs of SSI-Clustering and averaging the results standing for the same modes can further enhance the accuracy of SSI. Need to be noted is that the massive results in this section are obtained by selecting diverse signals to be the inputs of SSI-Clustering, instead of by choosing different orders. By this way, those spurious modes who concentrate on the stabilization diagram just as the real modes, thus being mixed with the real ones, can be picked out, because spurious modes brought by mathematical error will disappear when choosing other signals as the inputs of SSI. Call this part Accurate Identification Algorithm, which improves the accuracy of estimation and at the same time eliminates more spurious modes. As follow is the concrete implementation flow:

1) Find out the modes. All measured signals are detrended and input into SSI-Clustering. Assume that k tentative estimation results are \( \{\text{mode}_1,\text{mode}_2,\ldots,\text{mode}_k\} \).

2) Obtain a large number of results corresponding to the same modes. For each \( \text{mode}_i \), by comparing the magnitudes of the corresponding mode shapes, the M signals with the strongest observability, i.e. the largest mode shape magnitudes, are found. Randomly select several signals among these M signals as the inputs of SSI-Clustering. Keep the results that are similar to \( \text{mode}_i \) in damping, frequency and mode shape, regarding them as other estimates of \( \text{mode}_i \). Repeat random selection of input signals and conduct SSI-Clustering N times. Then a number of results calculated by utilizing different signals but estimating the same mode \( \text{mode}_i \) are obtained. The reason why the signals with the strongest observability are chosen is that reference [14] mentioned that the modes concerned should be observable in the signals input into SSI.

3) Obtain final accurate results. DBSCAN is adopted again to remove those bad results with great error in the form of outliers. And then average the rest valid results as the accurate estimate of each \( \text{mode}_i \).

B. Searching Missing Modes

Those modes that are not found by SSI when using some signals as inputs may be identified when other signals are chosen as inputs. Searching Algorithm (i.e. the second part of the revised algorithm) intended to find out missing modes firstly inputs each signal separately to SSI, and after screening out modes who have been identified, applies clustering technology to the results. By doing so, some missing modes will appear in the form of clusters. This procedure includes four steps.

1) Find out modes existing in each signal. For each signal, it will be input into SSI-Clustering separately. Then the identification results of the modes with strong observability in each signal are obtained.

2) Remove modes identified. The results produced in step 1) will be compared with the accurate estimates found by Accurate Identification Algorithm, and those results similar to accurate results in damping and frequency, thus considered to have been identified, will be eliminated.

3) Find out missing modes. Conduct another density-based clustering algorithm CFSFDP [18] to the results screened in step 2) and each cluster is matched with a missing mode. The reason why CFSFDP is adopted here instead of DBSCAN is...
that experiments indicate that the distribution of the points obtained in step 1) and step 2) belonging to the same missing modes are not centralized well, in which case DBSCAN performs badly.

4) Calculating accurate estimate of missing modes. For each cluster of CFSFDP, input the signals corresponding to the points located in the cluster into Accurate Identification Algorithm to get final accurate results.

V. TEST RESULTS

In this section, IEEE 39-bus system is used to evaluate the performance of the algorithms proposed. For the purpose to bring in modes with low damping levels, generator parameters are changed. 1% of all the active load is modeled to be zero-mean Gaussian white noise to emulate the effect of random load fluctuations. The system is simulated for 5 min with time step 0.02s, and time series of active power of all the lines and active outputs of all the generators with sampling rate 50Hz are collected. Finally, the zero-mean Gaussian white noise with SNR of 60 dB are applied to the measured signals to simulate the measurement error.

First of all, find out the modes. Fig. 1 shows the stabilization diagram obtained after conducting SSI-Clustering to all the signals collected.

![Figure 1. The stabilization diagram](image1)

From Fig. 1, we can see that while eliminating spurious modes, SSI-Clustering discovers three real modes, which are exactly all the weakly damped modes in the simulation system.

![Figure 2. Estimation results for mode 1](image2)

The damping ratios and frequencies of the tentative estimation results compared with the actual values are shown in Table. 1. For each mode, many estimation results by choosing diverse input signals are obtained after completing Accurate Identification Algorithm. Due to limit space, only estimates for mode 1 are given (see Fig. 2). By averaging those valid results, final accurate results are obtained and given in Table. II. Comparing the results in Table. I and Table. II, we can find that Accurate Identification Algorithm can heighten the estimating accuracy of SSI indeed (especially the damping ratio of mode 3).

To demonstrate the validly of Searching Algorithm, remove mode 2 from results of Accurate Identification Algorithm, thus simulating the case that mode 2 is not identified by Accurate Identification Algorithm, and then apply Searching Algorithm. When the first three steps of Searching Algorithm are completed, the estimation results of the missing mode are presented as a cluster of CFSFDP in Fig.3, with other results as halos. Finally, choose the signals corresponding to the points located in the cluster as the input of Accurate Identification Algorithm. Fig. 4 shows the massive results for the missing mode obtained by Accurate Identification Algorithm, and the final accurate damping ratio and frequency estimation results of the missing mode are 3.85% and 1.0877Hz, which are close to the actual values of damping ratio and frequency of mode 2.

![Figure 3. Missing modes](image3)

![Figure 4. Estimation results for the missing mode](image4)
VI. CONCLUSION

This paper proposed an approach based on SSI and clustering algorithms to automatically detect low-frequency modes in power systems. The stochastic subspace method gives estimation of system modes and DBSCAN and MAFIA discriminate real modes from spurious modes. To reduce the estimation error, Accurate Identification Algorithm was proposed. By utilizing different signals as the inputs of SSI-Clustering and averaging the results of the same modes, it enhances the accuracy of estimation and further removes spurious modes. Searching Algorithm was introduced to find missing modes. It firstly finds out modes existing in each signal and then uses CFSFDP to detect missing modes which are observable in many signals. Test shows that they can meet the needs of use, there may be some other improving the performance of mode identification.

DBSCAN and MAFIA used in this paper are both conventional clustering algorithms. Although the current results show that they can meet the needs of use, there may be some other advanced clustering algorithms, which are better than the two algorithms in performance and speed. Finding such advanced clustering algorithms will be a direction to improve this research in the future. Also, Accurate Identification Algorithm is time-consuming for it runs SSI many times. Improving the speed is another point of future research.

REFERENCES