Enhance the Separation Performance of ICA via Clustering Evaluation and Its Applications

Wei Cheng*, Zhusuo Zhang, and Zhengjia He

State Key Laboratory for Manufacturing Systems Engineering, Xi’an Jiaotong University, Xi’an 710049, P. R. China

Traditional independent component analysis may produce different results in the repeated calculations for their statistical characteristics, and thus they are unreliable and unstable. This paper introduces a novel ICA algorithm which enhances the separation performance and stability by clustering evaluation. Firstly, the improved ICA algorithm runs a single ICA algorithm for several times with the different initial parameters; secondly, the separated components produced are clustered according to their relevance; lastly, the best separated components are selected as the optimal results by clustering evaluation. The effectiveness of the improved ICA algorithm is validated in the simulation by typical mechanical signals: The proposed method is also applied to extract the effective information of observed signals on the bulkhead of a ship, and the results show that most of the important information is well extracted. This research provides a novel approach for vibration reduction and control of ships.

Keywords: Independent Component Analysis, Clustering Evaluation, Improved ICA Algorithm, Source Separation, Vibration Reduction and Control.

1. INTRODUCTION

Independent component analysis (ICA) is a statistical and computational technique for revealing hidden factors that underlies sets of random variables, measurements, or signals. Now ICA is widely used in the analysis of multidimensional data, such as feature extraction of brain imaging, image feature extraction and recognition, biological signal analysis, fault feature extraction, astronomical data analysis, and data compression. However, because it is a stochastic signal processing method and the measured signals always have finite sample size, an important problem of most ICA algorithms is that the results may be different in repeated calculation, and thus the results cannot be trusted. This paper introduced a novel method that enhance the separation performance of ICA via clustering evaluation, and applied the method to analyze the mechanical vibration signals of a ship.

For any statistical method, it is necessary to analyze the statistical reliability of the results. It has been proved that finite sample size will induce statistical errors in the estimation, and this problem was solved using bootstrapping strategy. Typical ICA methods such as Fast ICA and natural gradient algorithms separate the multidimensional data by minimizing or maximizing of an objective function, such as likelihood, mutual information, or negentropy. However, all these algorithms have the problem of uncertainty results. To enhance the separation performance and the stability of ICA algorithms, many works have been devoted to the theoretical model and computing method. One of the promising methods for enhancing the stability is to run a single ICA algorithm for many times, and produce each independent component for a certain number, and then evaluate these components by clustering evaluation method. This paper detailed analyzes a clustering evaluation method based on multidimensional scaling (MDS) to measure the dissimilarities between different signals, and at last visualize the clustering results with the nonlinear 2-D projection.

Vibration and noise control is very important to large-scale equipments especially for warships and submarines. Due to the transmission influences of ship hulls and nonlinear mixing of source signals, it is a very challenging work to extract the effective information of sources from the vibration signals on the bulkhead. Many works has been devoted to transmission characteristics of structures, such as beams, girders, casings and shells. And some works dedicated to the signal processing method to extract the features. However, all these methods did not extract the complete effective information of vibration sources. Aiming at extracting complete effective information, this paper proposed a novel approach for source separation of mixed signals on the bulkhead, and thus the effective information can be extracted and the sources can be online identified and actively controlled.

This paper mainly focuses on vibration reduction and control based on source separation. Firstly, the basic theory of ICA and an improved ICA algorithm is introduced; secondly, the separa-
ration performance is validated by typical mechanical signals; lastly, the improved ICA algorithm is applied to source separation of vibration signals of a ship, and the separated components are further analyzed by correlation analysis and mark tracking method.

2. INDEPENDENT COMPONENT ANALYSIS

2.1. Basic Theory of ICA

Assume that \( n \) sources \( S = \{s_1, s_2, \ldots, s_n\} \) exist at the same time, and \( m \) mixed signals \( X = \{x_1, x_2, \ldots, x_m\} \) which are composed of these sources are obtained in different places. Thus each mixed signal can be described as:

\[
x_i = \sum_{j=1}^{n} a_{ij} s_j + n_i \quad i = 1, 2, \ldots, n, \quad j = 1, \ldots, m
\]

(1)

\[
y_i = \sum_{i=1}^{n} \omega_i x_i = \sum_{i=1}^{n} \omega_i a_{ij} s_j
\]

(2)

where \( y_i \) is the independent component, and \( \omega \) is the separating matrix.

2.2. Separation Criterion

\[
H(s) = -\int p(s) \log p(s) \, ds
\]

(3)

\[
NG(s) = H(s_{\text{Gauss}}) - H(s)
\]

(4)

\[
NG(s) \propto [E[G(s)] - E(G(s_{\text{Gauss}}))]^2
\]

(5)

where \( H(\cdot) \) is an entropy function; \( NG(\cdot) \) is a negentropy function; \( E(\cdot) \) is a mean function; and \( G(\cdot) \) is a nonlinear function. Generally,

\[
G(u) = \frac{1}{a} \log \cosh(au)
\]

(6)

\[
G(u) = -\exp(-u^2/2)
\]

(7)

2.3. Enhance the Separation Performance via Clustering

In essential, most ICA algorithms such as Fast ICA are statistical. The framework of Fast ICA is shown in Figure 1. Therefore, their results may be different in repeated executing of the algorithm, and thus the outputs of a single run of an ICA algorithm cannot be trusted.\(^{22}\) In this paper, the stability and effectiveness of the ICA are improved via clustering evaluation.

\[
d_{ij} = 1 - p_{ij}
\]

(8)

\[
I_E = \frac{1}{l} \sum_{i=1}^{l} S^2_{ij}
\]

(9)

\[
S^2_{ij} = \frac{1}{|CL_j||CL_i|} \sum_{i \in CL_j} \sum_{j \in CL_i} d_{ij}
\]

(10)

\[
S^2 = \min_{i \neq j} \frac{1}{|CL_i||CL_j|} \sum_{i \in CL_i} \sum_{j \in CL_j} d_{ij}
\]

(11)

where \( p_{ij} \) is correlation coefficient between signal \( i \) and signal \( j \); \( d_{ij} \) is a dissimilarity coefficient between signal \( i \) and signal \( j \); \( I_E \) is a cluster validity index; \( CL \) is a set of all the components; and \( l \) is the serial number of the components.

2.4. Waveform Correlation Coefficient \( \rho_{xy} \)

\[
\rho_{xy} = \frac{\sum_{k=1}^{n} s(k)x(k) + s^*(k)x^*(k)}{\sqrt{\sum_{k=1}^{n} s(k)^2} \sum_{k=1}^{n} x(k)^2}
\]

(12)

where \( k \) is the data sequence.

3. SIMULATION EXPERIMENT

In the simulation experiment, there are four source signals. Three of them are typical modulated signals, and the other one is the white noise. The generation functions of these sources are as Eq. (13), and the mixing matrix \( A \) is obtained randomly. The data length of source signals is 1000, and the step is 1.0.

\[
S(t) = \begin{cases} 
  s_1(t) = N(t) \\
  s_2(t) = \sin(0.2t) \cos(15t) + \sin(2t) \\
  s_3(t) = \sin(0.3t) \sin(5t + \sin t) \\
  s_4(t) = \sin(0.3t \sin(0.5t)) 
\end{cases}
\]

(13)

\[
A = \begin{bmatrix} 0.65 & 0.75 & 0.65 & 0.60 \\
 0.95 & 0.70 & 0.75 & 0.85 \\
 0.28 & 0.40 & 0.80 & 0.32 \\
 0.90 & 0.60 & 0.52 & 0.95 
\end{bmatrix}
\]
The waveforms of source signals are shown in Figure 2; the mixed signals are generated by the mixing matrix $A$ and the source signals $S(t)$, and their waveforms are shown in Figure 3. The mixed signals are respectively separated by Fast ICA and the proposed ICA algorithm, and the waveforms of the separated components are respectively shown in Figures 4 and 5. It should be mentioned that Figure 4 is one of the inaccurate results of Fast ICA.

From Table I, Figures 4 and 5, it can be clearly seen that the waveforms separated by Fast ICA are complex, and the effective information is not well extracted. This is because Fast ICA is a statistical signal processing method, and the contrast function of Fast ICA does not always converge to the same point. While the correlation coefficients between components separated by the improved ICA algorithm and respective sources are all more than 0.98, which means that the effective information is well extracted. The 2D CCA projection of clustering in Figure 6 clearly shows that all the components in the same cluster have very high relevance, and the components in different clusters have low relevance, which means the sources are independent; some components of clusters 2, 3 and 4 have high relevance for a single run, which means these components are unreliable. However, the optimal components can be well selected by clustering evaluation method, and it is proven in Figure 5 that all the optimal components are credible.

4. APPLICATIONS

Warships and submarines attacked by underwater weapons are mainly identified and locked by radiated noise, therefore, it is essential to identify the vibration sources so as to actively control over the vibration and radiated noise. In this paper, the improved ICA algorithm is applied to extract the source information from the mixed signals of a ship’s bulkhead, and the ship has two diesel engines which are generally seen as the main vibration sources. The sample frequency is 16384 Hz, and data length is 16384. The test diagram of the ship is shown in Figure 7.

The waveforms of the observed signals are shown in Figure 8. It can be clearly seen that there are many shocks in the waveforms, and that the waveforms of the observed signals are

<table>
<thead>
<tr>
<th>Table I. Waveform correlation coefficients.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_1(t)$</td>
</tr>
<tr>
<td>Components by fast ICA</td>
</tr>
<tr>
<td>Components by the improved ICA</td>
</tr>
</tbody>
</table>
complex due to the transmission influence of the ship’s hull. Figure 9 shows that the spectral bands of the major components of the observed signals range from 0 to 2000 Hz; some range in a narrow band (near 3000 and 5000 Hz). However, it is still a difficult task to reveal the effective information of each diesel engine from the observed signals.

By the DSS method, the single ICA algorithm is executed for 10 times with the different origin parameters, and 3 components are extracted one by one, while the first and the second separated components are selected as the effective components. Figure 10 shows that the two separated components are separated well. Besides only one component has the high relevance to the components in the other cluster, every component in one cluster has very high relevance to the others, and the correlation coefficients are all more than 0.90. By clustering evaluation, the separated
component which has the high relevance with the others in the same cluster is selected as the optimal component.

Figures 11 and 12 show the waveforms of the optimal separated components and source signals respectively, and it can be seen in the red circles that most important information of the source signals are well extracted by the improved ICA algorithm. The correlation coefficients between the separated components and respective sources are shown in the following matrix:

$$\rho_w = \begin{bmatrix} 0.85 & 0.28 \\ 0.35 & 0.83 \end{bmatrix}$$

The correlation coefficient also indicates that the improved ICA algorithm can effectively extract the source information from the mixed signals that observed on the bulkhead of the ship, and thus the radiated noise caused by the vibrations of diesel engines can be actively controlled by controlling the related sources.

To further analyze the vibration sources, some identical components of spectral band 0–2000 Hz are marked and tracked. The local spectrums of separated components and source signals are shown in Figures 13 and 14. Compared Figures 13 with 14, it can be seen that some identical components such as 231, 342, 590, 660, 1000, 1200, 1320, 1999 Hz that are extracted from the observed signals root in the source signal 1, which is caused by the diesel generator; other identical components such as 231, 301, 391, 540, 584, 735, 1000, 1200, 1360 Hz root in the source signal 2, which is caused by the propulsion diesel engine. By mark tracking method, all the important components can be marked and tracked, thus all the causes of these components can be traced to the source, and all the interesting components can be actively controlled from the source.

5. CONCLUSIONS

In this paper a novel active control over vibration and noise method based on an improved ICA algorithm is proposed. The improved ICA algorithm enhances the separation performance and stability by running a single ICA algorithm for several times, and chooses the optimal separated components by clustering evaluation method. In the simulation experiment, typical mechanical signals are used to validate the separation performance of the Fast ICA algorithm and the improved ICA algorithm. The result shows that all the coefficients between separated components and source signals are more than 0.98, which indicates that the improved ICA algorithm is effective and stable. The improved ICA algorithm is applied to extract the effective information of vibration signals observed on the bulkhead of a ship, and the result shows that most important information of the source signals are extracted and the effectiveness is validated by the mark tracking method. Therefore, it provides a new way to vibration and noise reduction and control.

Acknowledgment: This work is supported by the key project of National Nature Science Foundation of China (No. 50335030), National Nature Science Foundation of China (No. 50875197), the Fundamental Research Funds for the Central Universities, and the Project Sponsored Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry.

References and Notes


Received: 27 May 2010. Accepted: 6 August 2010.