A NOVEL PRE-PROCESSING TECHNIQUE OF BLIND SOURCE SEPARATION APPLYING Q-MODE FACTOR ANALYSIS

Yoshio Konno[a], Jianting Cao[bc] and Mamoru Tanaka[a].

[a] Department of Electrical and Electronics Engineering, Sophia University,
7-1 Kioicho, Chiyoda-ku, Tokyo 102-8554, Japan.
[b] Department of Electronics Engineering, Saitama Institute of Technology,
1690 Fusaiji, Okabe, Saitama 369-0293, Japan.
[c]Brain Signal Processing, RIKEN,
2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan.
Phone: +81-3-3238-3878, Fax: +81-3-3238-3321, E-mail : yo-konno@sophia.ac.jp

ABSTRACT

In this study, a novel way of processing observations before independent component analysis (ICA) using a Q-mode factor analysis (FA) was proposed for noisy blind source separation (BSS). The Q-mode analysis is a very efficient technique in classifying a data in cases where there are a large number of objects and where there is a little prior knowledge of the constituents. In the R-mode analyses, interrelationships between variables are analyzed. On the other hand, in the Q-mode analysis, interrelationships between objects are analyzed. Applying this approach to the experimental noisy data, we show that our proposed approach is more effective than the R-mode analysis for source separation of noisy data.

1. INTRODUCTION

Blind source separation has received a great deal of attention due to various applications in science and technology [1]-[4]. The problem of BSS has been studied by many researchers in neural networks and statistical signal processing and many interesting theoretical and practical results have been achieved.

The basic problem of noisy BSS or factor model in relation to the data matrix is defined as

$$\mathbf{X}_{(N\times m)} = \mathbf{S}_{(N\times n)}\mathbf{A}^T_{(n\times m)} + \mathbf{E}_{(N\times m)}. \tag{1}$$

Here $\mathbf{X}$ is the data matrix of observations. Any column vectors of $\mathbf{X}$ are variables (components) on which we have N objects (samples). A row vector of $\mathbf{X}$ represents one object on which m variables have been measured. $\mathbf{S}$ is the matrix of sources or factor scores: A column vector of $\mathbf{S}$ is a factor (the number of factors is n). $\mathbf{A}$ is the matrix of factor loadings, and $\mathbf{E}$ is the matrix of residuals or error terms.

For convenience, we take the transpose of Eq. (1) to get

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{e}, \tag{2}$$

where $\mathbf{x}$ is a column vector representing one of the objects of the data matrix. In the model, $\mathbf{s}$, $\mathbf{e}$, $\mathbf{A}$ and n are unknown but only $\mathbf{x}$ are accessible. It is assumed that the components of sources are mutually statistically independent, as well as being statistically independent of the noise components. Moreover, the noise components themselves are assumed to be mutually independent. Our goal is to estimate the independent sources under the challenging conditions or assumptions.

2. R-MODE FACTOR ANALYSIS

The R-mode factor analysis is one of the pre-processing technique for BSS, which is extended the principal component analysis (PCA) for pre-whitening of the observation with high-level noise reduction [1]-[3].

When the sample size N is sufficiently large, the covariance matrix of the observation in the mixing model $\mathbf{\Sigma}_{(m\times m)}$ can be written as

$$\mathbf{\Sigma} = \mathbf{A}\mathbf{A}^T + \mathbf{\Psi}, \tag{3}$$

where $\mathbf{\Psi}_{(m\times m)} = \mathbf{E}^T\mathbf{E}/N$ is a diagonal matrix. The covariance matrix of the observation recorded by sensors can also be given by $\mathbf{C}_{(m\times m)} = \mathbf{X}^T\mathbf{X}/N$.

In the R-mode FA, the matrix $\mathbf{A}$ can be estimated as

$$\widehat{\mathbf{A}} = \mathbf{U}_{\hat{n}}\mathbf{\Lambda}_{\hat{n}}^{\frac{1}{2}}, \tag{4}$$

by applying the standard PCA approach. Here $\mathbf{\Lambda}_{\hat{n}(\hat{n}\times\hat{n})}$ is a diagonal matrix whose elements are the n largest eigenvalues of $\mathbf{C}$, the columns of $\mathbf{U}_{\hat{n}(m\times\hat{n})}$ are the corresponding eigenvectors, and $\widehat{n}$ is the estimated number of sources.

To estimate $\mathbf{\Psi}$, we fit $\mathbf{A}\mathbf{A}^T$ to $\mathbf{C} - \mathbf{\Psi}$ using the eigenvalue decomposition (EVD) method. In this case, the cost function

$$L(\mathbf{A}, \mathbf{\Psi}) = \mathrm{tr}[\mathbf{C} - \mathbf{A}\mathbf{A}^{\mathrm{T}} - \mathbf{\Psi}]^2 \tag{5}$$

is employed and it is minimized by $\partial L(\mathbf{A}, \mathbf{\Psi})/\partial \mathbf{\Psi} = 0$, whereby the estimate noise variance $\mathbf{\Psi}$ is obtained as

$$\widehat{\mathbf{\Psi}} = \mathrm{diag}(\mathbf{C} - \widehat{\mathbf{A}}\widehat{\mathbf{A}}^{\mathrm{T}}). \tag{6}$$

Once the estimates $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{\Psi}}$ converge to stable values, we need to finally compute the score matrix, the pseudo-inverse matrix. Since the solution for a pseudo-inverse matrix is not unique, we employ the Bartlett method to determine the transform matrix $\mathbf{Q}_{(\hat{n}\times m)}$. In this method, the noise variance $\mathbf{\Psi}$ is included in the calculation, that is

$$\mathbf{Q} = [\widehat{\mathbf{A}}^T \widehat{\mathbf{\Psi}}^{-1} \widehat{\mathbf{A}}]^{-1} \widehat{\mathbf{A}}^T \widehat{\mathbf{\Psi}}^{-1}. \tag{7}$$

Using the above result, the new set of data transformed from the observations, that is the estimated factor score $\widehat{\mathbf{F}}_{(N\times\hat{n})}$, can be obtained as

$$\widehat{\mathbf{F}} = \mathbf{X}\mathbf{Q}^T, \quad \widehat{\mathbf{f}} = \mathbf{Q}\mathbf{x}. \tag{8}$$

Note that the covariance matrix is $\mathrm{E}[\widehat{\mathbf{F}}^{\mathrm{T}}\widehat{\mathbf{F}}] = \mathbf{I}_{\hat{\mathrm{n}}} + \mathbf{Q}\mathbf{\Psi}\mathbf{Q}^{\mathrm{T}}$, which implies that the source signals in a subspace are decorrelated.

3. Q-MODE FACTOR ANALYSIS

The Q-mode analysis is a very efficient technique in classifying a data in cases where there are a large number of objects and where there is a little prior knowledge of the constituents [1]. In the R-mode analyses, interrelationships between variables are analyzed. On the other hand, in the Q-mode analysis, interrelationships between objects are analyzed.

In the Q-mode analysis, the matrix $\mathbf{X}$ are row-normalized as

$$\mathbf{W}_{(N\times m)} = \mathbf{D}^{-\frac{1}{2}}\mathbf{X}, \tag{9}$$

where $\mathbf{D}_{(N\times N)} = \mathrm{diag}(\mathbf{X}\mathbf{X}^{\mathrm{T}})$. Using the row-normalized observation $\mathbf{W}$, the association matrix $\mathbf{H}$ is defined as the major product moment of row-normalized data:

$$\mathbf{H}_{(N\times N)} = \mathbf{W}\mathbf{W}^T. \tag{10}$$

The row-normalized data can be expressed approximately as the product of a factor-loading matrix $\mathbf{A}_{(N\times\hat{n})}$ and a factor-score matrix $\mathbf{F}_{(m\times\hat{n})}$, that is

$$\mathbf{W} \approx \mathbf{A}\mathbf{F}^T, \tag{11}$$

where $\hat{n}$ is the approximate rank of $\mathbf{W}$. The relationship between $\mathbf{W}$, $\mathbf{H}$, $\mathbf{A}$ and $\mathbf{F}$ are given by

$$\mathbf{H} = \mathbf{W}\mathbf{W}^T = \mathbf{A}\mathbf{F}^T\mathbf{F}\mathbf{A}^T. \tag{12}$$

The constraint that the matrix $\mathbf{F}$ be columwise orthonormal is expressed by $\mathbf{F}^T\mathbf{F} = \mathbf{I}$, which leads to $\mathbf{H} = \mathbf{A}\mathbf{A}^T$.

The square, symmetric matrix $\mathbf{H}$ can be factored according to

$$\mathbf{H} = \mathbf{U}_{\hat{n}}\mathbf{\Lambda}_{\hat{n}}\mathbf{U}_{\hat{n}}^T, \tag{13}$$

where $\mathbf{U}_{\hat{n}(N\times\hat{n})}$ is the matrix of eigenvectors and $\mathbf{\Lambda}_{\hat{n}(\hat{n}\times\hat{n})}$ is the diagonal matrix of associated eigenvalues. Here, one possible solution is

$$\widehat{\mathbf{A}} = \mathbf{U}_{\hat{n}}\mathbf{\Lambda}_{\hat{n}}^{\frac{1}{2}}. \tag{14}$$

It denotes that the matrix of factor loading is the matrix of eigenvectors, scaled by the square roots of the eigenvectors.

Once the estimated factor loading matrix $\widehat{\mathbf{A}}$ is obtained, the matrix of factor scores may be obtained by

$$\mathbf{W} \approx \widehat{\mathbf{A}}\mathbf{F}^T. \tag{15}$$

Premultiplying it by $\widehat{\mathbf{A}}^T$ gives $\widehat{\mathbf{A}}^T\mathbf{W} \approx \widehat{\mathbf{A}}^T\widehat{\mathbf{A}}\mathbf{F}^T$. Using $\widehat{\mathbf{A}}^T\widehat{\mathbf{A}} = \mathbf{\Lambda}_{\hat{n}}$, the factor score matrix can be obtained as

$$\widehat{\mathbf{F}} = \mathbf{W}^T\widehat{\mathbf{A}}\mathbf{\Lambda}_{\hat{n}}^{-1}. \tag{16}$$

In practice, when the number of objects N is much greater than the number of variables m, it is computationally more efficient to use the following procedure. In this version of the Q-mode analysis, the minor product moment

$$\mathbf{H}^*_{(m\times m)} = \mathbf{W}^T\mathbf{W} \tag{17}$$

is computed instead of Eq. (10). It should be noted that the order of $\mathbf{H}^*$ is $m \times m$, which is pretty smaller than that of $\mathbf{H}$. The eigenvalues $\mathbf{\Lambda}_{\hat{n}(\hat{n}\times\hat{n})}$ and eigenvectors $\mathbf{V}_{\hat{n}(m\times\hat{n})}$ of $\mathbf{H}^*$ are computed as

$$\mathbf{H}^* = \mathbf{V}_{\hat{n}}\mathbf{\Lambda}_{\hat{n}}\mathbf{V}_{\hat{n}}^T. \tag{18}$$

It should be noted that the positive eigenvalues of $\mathbf{H}^*$ are same as those of $\mathbf{H}$ and the factor scores matrix $\widehat{\mathbf{F}}$ is identical to $\mathbf{V}_{\hat{n}}$, that is,

$$\widehat{\mathbf{F}} = \mathbf{V}_{\hat{n}}. \tag{19}$$

Using this estimation, the factor loading matrix $\widehat{\mathbf{A}}$ can now be computed as

$$\widehat{\mathbf{A}} = \mathbf{W}\widehat{\mathbf{F}}. \tag{20}$$

It should be noted that this procedure is computationally simpler because m is smaller than N and no scaling of the columns of $\mathbf{A}$ and $\mathbf{F}$ is needed. In our simulation, we regard the column vectors of $\widehat{\mathbf{A}}$ as the components of signal on which we have N objects.

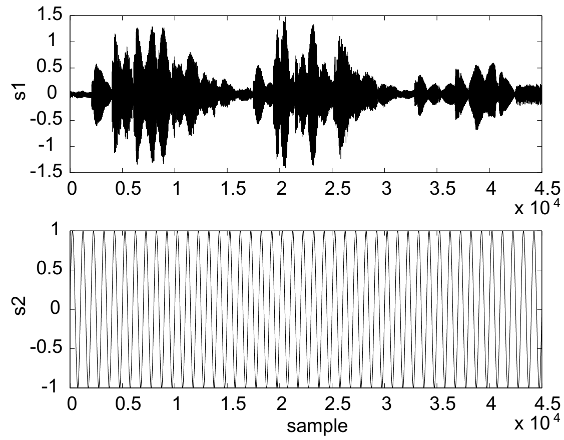

Fig. 1. Source signals.

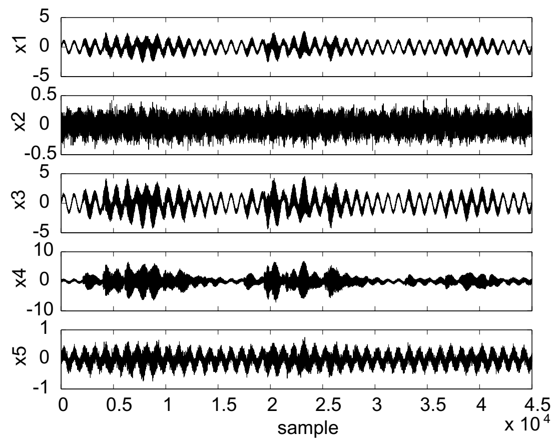

Fig. 2. Mixed signals with noises.

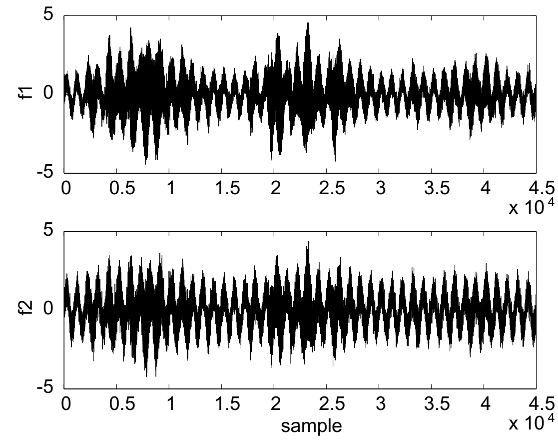

Fig. 3. Pre-processing with R-mode FA.

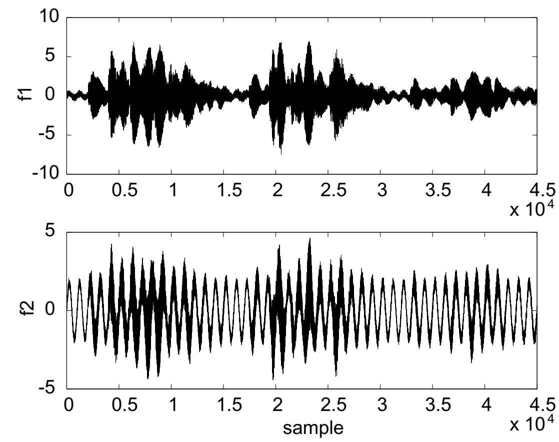

Fig. 4. Pre-processing with Q-mode FA.

4. FAST-ICA ALGORITHM

After pre-processing of the noisy observations, the transformed signals are obtained through a procedure in which the powers of noise, mutual correlation and dimensionality have been reduced. The decomposed independent sources can be obtained from a linear transformation as

$$\widehat{\mathbf{f}}_{ICA} = \mathbf{W}_{ICA}\widehat{\mathbf{f}}, \tag{21}$$

where $\mathbf{W}_{ICA(n\times n)}$ is termed the de-mixing matrix which can be computed by using the Fast-ICA algorithm.

The Fast-ICA algorithm has been proposed in [4]. This algorithm is based on fixed-point method and is represented by

$$\mathbf{w}^{+} = \mathbf{w}(t) - \eta\frac{E[fg(\mathbf{w}^T(t)f)] - \beta\mathbf{w}(t)}{E[g\prime(\mathbf{w}^T(t)f)] - \beta}, \tag{22}$$

$$\mathbf{w}(t+1) = \frac{\mathbf{w}^{+}}{\|\mathbf{w}^{+}\|}, \tag{23}$$

where $g(y) = y^3$, or $g(y) = \tanh(y)$.

The fixed-point algorithm has a higher speed convergence property since the Newton method in block mode is applied. It is easy to apply in data analysis since there is no learning rate parameter needs to be adjusted. Furthermore, we can extract independent sources one by one. This means the condition of the prior knowledge of source number will be more relaxed.

5. COMPUTER SIMULATIONS

In this section, we have performed a simulation experiment with one super-Gaussian source ($\mathrm{kurtosis} = 3.4274$), and one sub-Gaussian source ($\mathrm{kurtosis} = -1.5000$) (see Fig. 1). Two sources were artificially mixed by a 5×2 random numeric matrix. Five un-correlated Gaussian noises was added to an associated element of observations (see Fig. 2).

To compare the power of the source to that of the noise, the signal-to-noise ratio (SNR) was defined as

$$\mathrm{SNR}_i = 10\log\frac{E[(\sum_k a_{ik}s_k)^2]}{E[e_i^2]}, \tag{24}$$

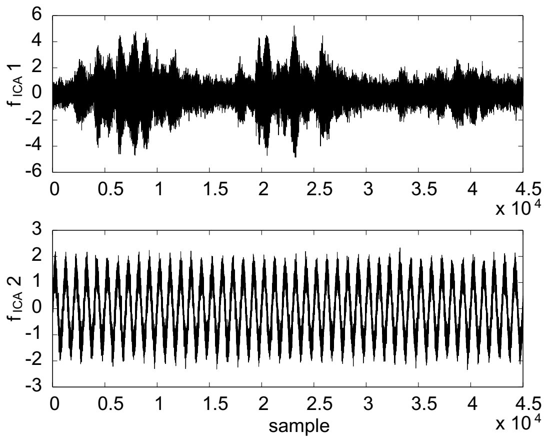

Fig. 5. Decomposed source signals applying R-mode FA and ICA.

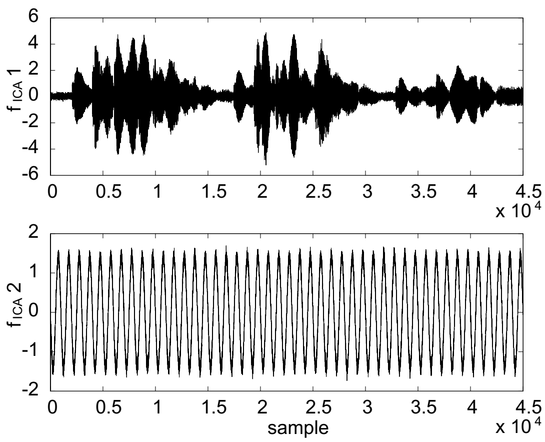

Fig. 6. Decomposed source signals applying Q-mode FA and ICA.

on sensor i ($i = 1, \cdots, 5$). Using this formula, the SNR became $-7.4795\,\mathrm{dB}$ on sensor 2.

The R-mode FA and the Q-mode FA were used for pre-processing of source separation. As seen from the result of the R-mode analysis (Fig. 3), some noises were not removed and the sources were still overlapping. On the other hand, in the results of the Q-mode analysis (Fig. 4), even though the sources were still overlapping, but the high-level noises were almost removed.

Following this result, the Fast-ICA algorithm was used to further separate the overlapped components. The results of the R-mode analysis with the Fast-ICA (Fig. 5) indicate that the sources were not overlapping but some noises were not removed. On the other hand, in the results of the Q-mode analysis with the Fast-ICA (Fig. 6), the source signals are accurately estimated and the high-level noises were almost removed.

In order to strictly compare the results of the R-mode and the Q-mode analysis, we defined the signal-to-error ratio (SER) as

$$\mathrm{SER}_i = 10 \log \frac{E[(\sum_k a_{ik} s_k)^2]}{E[(\sum_k a_{ik} s_k - \sum_k \widehat{a}_{ik} \widehat{f}_{ICA\ k})^2]}, \quad (25)$$

on the sensor spaces. Here, $\widehat{a}_{ik}$ denotes the coefficient to project k-th decomposed component into i-th sensor space, which is calculated by the R-mode or Q-mode FA and ICA. As for the result of the R-mode analysis, the averaged SER became $10.8383\,\mathrm{dB}$. On the other hand, in the result of the Q-mode analysis, the averaged SER became $22.5390\,\mathrm{dB}$. This result means that, when applying the Q-mode analysis, a higher value of SER is obtained. Given these results, we can confirm that the proposed Q-mode method is effective for noisy signal separation and pre-processing of ICA.

6. CONCLUSIONS

In this study, we proposed the novel approach using the Q-mode factor analysis for pre-processing of ICA. The R-mode analyses are designed to portray the interrelationships between variables. On the other hand, the Q-mode analysis are designed to portray the interrelationships between objects. Applying this approach to the experimental data, we confirmed that this technique is effective for noisy signal separation.

7. REFERENCES

[1] Richard Reyment, K. G. Jöreskog, "Applied Factor Analysis in the Natural Sciences", ISBN 0-521-41242-0, Cambridge University Press, 1993.

[2] S. Ikeda, K. Toyama, "Independent component analysis for noisy data - MEG data analysis," Neural Networks 13, pp. 1063-1074, 2000.

[3] J. Cao, N. Murata, S. Amari, A. Cichocki and T. Takeda, "A robust approach to independent component analysis of signals with high-level noise measurements," IEEE Trans. on Neural Networks, Vol. 14, No. 3, pp. 631-645, June 2003.

[4] A. Hyvaerinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," Neural Computation, Vol. 9, No. 7, pp. 1483-1492, 1997.