International Journal of Control and Automation
Vol. 9, No. 7 (2016), pp.147-158
http//dx.doi.org/10.14257/ijca.2016.9.7.14

Improved Independent Component Analysis Based on Epanechnikov Kernel Function

Liang-ju Yu[1], Gen-ke Yang [1*] and Yue Chen [1]

[1]Department of Automation and Key Laboratory of System Control and Information Processing, Shanghai Jiao Tong University, China
E-mail: yuliangju19920521@sjtu.edu.cn, gkyang@sjtu.edu.cn, chenyue@sjtu.edu.cn

Abstract

Traditionally, the key idea of estimating independent component analysis (ICA) model is to maximize the non-Gaussianity, however, often with the assumption that density of data is near the standardized Gaussian density. To avoid the unsuitable assumption, this article uses the nonparametric density estimating method. A nonparametric independent component analysis algorithm based on Epanechnikov kernel function is proposed in this paper. This algorithm uses the Epanechnikov kernel estimator to estimate random variable distribution, meanwhile, employs the hypothesis test to derive the nonparametric likelihood ratio (NLR) objective function. For optimizing the nonparametric density estimation, the selection of kernel function and bandwidth is crucial. From the perspective of minimizing the mean integrated square error (MISE), this paper discusses the optimal selection and conducts experiments for further study. To increase the algorithmic convergence rate and reduce the running time, the quasi-newton method has been used to optimize the objective function. Compared with previous nonparametric ICA algorithm, the simulation results demonstrate that the proposed method offers better performance both on speech separation and computing capability.

Keywords: *independent component analysis, Epanechnikov kernel function, nonparametric likelihood ratio, quasi-newton method, speech separation*

1. Introduction

Finding a suitable representation of multivariate data is the fundamental task in many fields. ICA is a recently developed method in which the goal is to find a linear representation of nongaussian data so that the components are statistically independent. Such a representation seems to capture the essential structure of the data in many applications, including feature extraction and signal separation [1]. Since the appearance of the first papers on ICA and blind source separation several decades ago, the field has made tremendous progress in terms of theory, algorithms, and applications [2]. Independent component analysis is widely applied to speech recognition[3],[4], speech synthesis[5], time-frequency analysis of cabin noise[6], communication systems[7], hyperspectral image processing[8],[9], and so on.

ISSN: 2005-4297 IJCA
Copyright © 2016 SERSC

Because of the unknown distributions of the components, Hyvarinen et al made the assumption that the density was not far from Gaussian distribution [10] and developed new approximations of negentropy as contrast function. A robust ICA algorithm called FastICA [11]was then attained by using fixed-point iteration for optimization of the contrast functions. In 2002, Bach et al proposed a class of kernel-based algorithms named Kernel ICA that utilized contrast functions based on canonical correlations in a reproducing kernel Hilbert space, and formulated Kernel ICA as a semiparametric model [12]. Cardoso showed in [13] that incorrect assumptions on such distributions could result in poor estimation performance, sometimes in a complete failure to obtain the source separation. To tackle this issue, in [14] Boscolo et al proposed a nonparametric model where the probability density functions were directly estimated from the data, by using a kernel density estimation technique. Also with Nonparametric Density Estimator, Chien et al presented a novel NLR objective function for ICA and applied the ICA method for speech feature extraction as well as for speech recognition [4]. However, since the computational cost was excessively high, it would be tremendous to do some computational improvement on this method.

In this paper, we use the NLR function proposed in [4] as the objective function. From the perspective of minimizing MISE, also for the sake of calculation, this paper discusses the selection of kernel Function and bandwidth, and chooses the Epanechnikov kernel function instead of Gaussian kernel to estimate random variable distribution. To increase the algorithmic convergence rate and reduce the running time, the quasi-newton method is used to optimize the objective function. In Section 2, firstly, the test of independence of random variables using hypothesis testing is briefly introduced. Then, the choice of Epanechnikov kernel and bandwidth is discussed from the perspective of MISE. Lastly, the derivation of NLR objective function using the DFP method is provided. In Section 3, a set of simulation experiments is conducted in order to demonstrate the performance improvement obtained with the proposed technique, for the case of mixtures of two signals. Finally, the conclusions drawn from this study are given in Section 4.

2. NLR-ICA Model Based on Epanechnikov Kernel Function

The conventional ICA separation model is usually based on the minimization of the mutual information between the reconstructed signals. The mutual information is difficult to approximate and optimize on the basis of a finite sample, and much research on ICA has focused on alternative contrast functions [12]. By using the hypothesis testing theory and Epanechnikov kernel, this paper derives nonparametric likelihood ratio which can be calculated directly to construct ICA model.

2.1. ICA Model and Separation Principle

The conventional ICA model assumes that independent source signals $\mathrm{s}=[s^{(1)},...,s^{(D)}]^T$ are mixed by an unknown, full-rank mixing matrix A (size $D\times D$), resulting in a set of mixtures given by $\mathrm{x}=A\mathrm{s}$ [1]. It aims to recover the original components with the transformed signals $\mathrm{y}=[y^{(1)},...,y^{(D)}]^T$ under $\mathrm{y}=W\mathrm{x}$, where W (size $D\times D$) is the demixing matrix.

Since ICA model focuses on finding a linear representation of non-Gaussian data so that the components are as independent as possible, the null hypothesis H_0 is naturally set that the components $y^{(1)},y^{(2)},...,y^{(D)}$ are mutually independent. Accordingly, the alternative hypothesis H_1 is that the components $y^{(1)},y^{(2)},...,y^{(D)}$ are dependent. When y is assumed to be Gaussian distributed, the null hypothesis equals to the uncorrelatedness which

 Copyright © 2016 SERSC

implies covariance between two components $y^{(i)}$ and $y^{(j)}$ is zero. Then the problem is easily solved.

However, the key to estimate the ICA model is nongaussianity. Without assuming any distribution of y , a nonparametric method where the variable distributions are directly estimated from the data using a kernel density estimation technique is introduced in this section. With transformed signal samples $\mathrm{Y}=\{\mathrm{y}_1,\mathrm{y}_2,...,\mathrm{y}_T\}$, the densities of transformed signals $y^{(d)}$ are presented by

$$p(y^{(d)})=\frac{1}{Th}\sum_{t=1}^{T}K(\frac{y^{(d)}-y_t^{(d)}}{h}),\quad d=1,\ldots,D \tag{1}$$

Here, $K(x)$ is the kernel function, satisfying $\int_{-\infty}^{\infty}K(x)dx=1$. T is the sample size and h is the kernel bandwidth, also called the smoothing parameter or bandwidth.

Using kernel method for multivariate data, the joint distribution density of D-dimensional vector y is given by

$$p(\mathrm{y})=\frac{1}{Th^D}\sum_{t=1}^{T}K_D(\frac{\mathrm{y}-\mathrm{y}_t}{h}) \tag{2}$$

The nonparametric likelihood ratio function is then derived from the maximum likelihood function measured respectively under the two hypothesis distribution [4]. When the null hypothesis is true, $y^{(1)},y^{(2)},...,y^{(d)}$ are mutually independent, the joint distribution can be factored into the product of distributions of individual components $p(\mathrm{y}\,|\,H_0)=\tilde{p}(\mathrm{y})=\prod_{i=1}^{d}p(y^{(i)})$. Then the objective maximum likelihood ratio for transformed signals becomes

$$\lambda_{NLR}=\frac{p(Y\,|\,H_0)}{p(Y\,|\,H_1)}=\frac{\prod_{t=1}^{T}\prod_{d=1}^{D}p(y_t^{(d)})}{\prod_{t=1}^{T}p(\mathrm{y}_t)} \tag{3}$$

The objective function is the logarithm of λ_{NLR}, which is a difference of log likelihoods between null hypothesis $L_0(W)$ and alternative hypothesis $L_1(W)$, defined by.

$$\underset{W}{Max}\log\lambda_{NLR}(W)\overset{\Delta}{=}L_0(W)-L_1(W)$$
$$Here,\ L_0(W)=\sum\nolimits_{t=1}^{T}\sum\nolimits_{d=1}^{D}\log(p(y_t^{(d)}));\ L_1(W)=\sum\nolimits_{t=1}^{T}\log(p(\mathrm{y}_t)) \tag{4}$$

With a level of significance λ_α , the null hypothesis is verified when λ_{NLR} is located in the acceptance region $\lambda_{NLR}\geq\lambda_\alpha$ [4].

2.2. The Selection of Kernel Function and Bandwidth

Undoubtedly, the estimation of signal distribution by means of nonparametric method is crucial for ICA model. It'll be necessary to discuss the selection of kernel function K and bandwidth h . Just as the naive estimator can be considered as a sum of 'boxes' centered at the observations, the kernel estimator is a sum of 'bumps' placed at the observations. The kernel function K determines the shape of the bumps while the window width h determines their width [15]. An illustration is given in Figure 1, composed of Gaussian kernel estimates of three different h. The individual bumps $1/Th*\varphi\{(x-X_i)/h\}$ as well as the estimate $\hat{f}$ constructed by adding them up are shown, with a small sample of size 7.

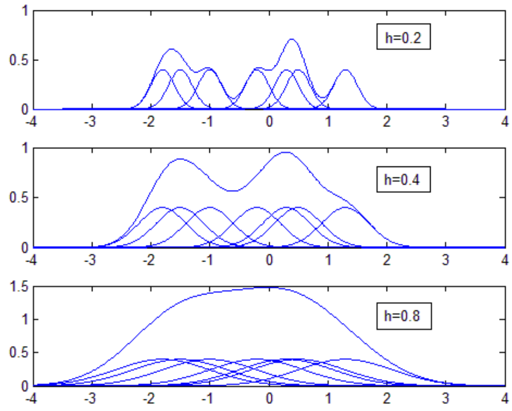

Figure 1. Kernel Estimates with Different Bandwidth H

It can be seen that if h is chosen too small then spurious fine structure becomes visible, however, if h is too large then the characteristic nature of the distribution is obscured. The choice of bandwidth is of great important to get better probability density function $p(y)$.

From the perspective of minimizing the MISE, there exists the ideal h and best kernel function. MISE is the most widely used way of measuring the accuracy of $\hat{f}(x)$ as an estimator of the true density $f(x)$, defined by $MISE(\hat{f})=E\int\{\hat{f}(x)-f(x)\}^2dx$. By minimizing the approximate MISE, the optimal value h_{opt} will be $h_{opt}=k_2^{-2/5}\{\int K(t)^2dt\}^{1/5}\{\int f^{"}(x)^2dx\}^{-1/5}n^{-1/5}$, where $k_2=\int t^2K(t)dt\neq 0$. Substituting h_{opt} back into the approximate MISE, then the approximate value of MISE will be $\frac{5}{4}C(K)\{\int f^{"}(x)^2dx\}^{1/5}n^{-4/5}$, where constant $C(K)$ is $C(K)=k_2^{2/5}\{\int K(t)^2dt\}^{4/5}$ [15].

This formula indicates that the value of MISE depends solely upon $C(K)$. Therefore, we choose $K(t)$ to get a smaller value of $C(K)$. The minimization of $C(K)$ turns into minimizing $\int K(t)^2$ subject to constrains $\int K(t)=1, \int t^2K(t)=1$. When $K(t)$ is set to be Epanechnikov kernel $K_e(t)$, the problem above is solved.

Adopting Epanechnikov kernel introduced in [15], the kernel functions $K(t)$ and $K_D(\mathrm{x})$ are defined as follows

$$K(t)=\begin{cases}\frac{3}{4\sqrt{5}}(1-\frac{1}{5}t^2) & -\sqrt{5}\le t\le\sqrt{5}\\ 0 & otherwise\end{cases} \tag{5}$$

And

$$K_D(\mathrm{x})=\begin{cases}\frac{1}{2}c_D^{-1}(D+2)(1-\mathrm{x}^T\mathrm{x}) & if \quad \mathrm{x}^T\mathrm{x}<1\\ 0 & otherwise\end{cases} \tag{6}$$

 Copyright © 2016 SERSC

Where c_D is the volume of the unit *D*-dimensional sphere: $c_1 = 2, c_2 = \pi, c_3 = 4\pi/3$. After substituting $K(t)$ back into the approximation of h_{opt} and making some coefficient adjustment, the estimated value of h_{opt} is set to be $\hat{h}_{opt} = 1.05\sigma n^{-1/5}$. With constrains of random variables, the in-depth simplification of $\hat{h}_{opt}$ is $1.05n^{-1/5}$, which is uniquely a function of the sample size.

For the consideration of computational effort involved and the degree of differentiability required, the selection of kernel function is essential. In the kernel method for multivariate data, one important factor in reducing the computational time is the choice of a kernel that can be calculated very quickly. In particular, it will be bad to use the Gaussian kernel, which requires calculating the exponential function each time [15]. Obviously, the choice of Epanechnikov kernel will avoid the weakness.

2.3. Objective Function Derivation

For the sake of smaller MISE and better computing capability, the Epanechnikov kernel is adopted to estimate the density of component $y^{(d)}$ and the joint distribution density of *D*-dimensional vector y. The nonparametric likelihood ratio function is then formed by

$$\lambda_{NLR} = \frac{p(Y \mid H_0)}{p(Y \mid H_1)} = \frac{\prod_{t=1}^{T}\prod_{d=1}^{D}[\frac{1}{Th}\sum_{k=1}^{T}K(\frac{y_t^{(d)} - y_k^{(d)}}{h})]}{\prod_{t=1}^{T}[\frac{1}{Th^D}\sum_{k=1}^{T}K_D(\frac{\mathrm{y}_t - \mathrm{y}_k}{h})]}$$

The objective function is equivalent to $\underset{W}{Max} \log \lambda_{NLR}(W)$, where

$$\log \lambda_{NLR}(W) = \sum_{t=1}^{T}\sum_{d=1}^{D}\log(\frac{1}{Th}\sum_{k=1}^{T}K(\frac{y_t^{(d)} - y_k^{(d)}}{h})) - \sum_{t=1}^{T}\log(\frac{1}{Th^D}\sum_{k=1}^{T}K_D(\frac{\mathrm{y}_t - \mathrm{y}_k}{h})) \quad (7)$$

$y_k^{(d)}$ is the kernel centroid, also the d th component of sample y_k.

$$y_k^{(d)} = \mathrm{w}_d \mathrm{x}_t = \sum_{j=1}^{D} w_{dj} x_k^{(j)} \quad (8)$$

w_d is the d th row of the separation matrix W, where $W = [w_{dj}]_{D\times D} = [\mathrm{w}_1^T \ldots \mathrm{w}_D^T]^T$.

The quasi-newton method is employed to maximize the objective function. The basic principle of quasi-newton method is to approximate the inverse of Hesse matrix in newton method without derivative of second order, which makes it more effective than newton method. Same with gradient descent algorithm, each step of iteration just needs the gradient information. By measuring the variation of gradient, a model of the objective function is generated to produce superlinear convergence. This method is vastly superior to gradient descent algorithm, especially for difficult issues. There are various quasi-newton methods by constructing different approximate matrix. Here we choose the DFP algorithm and the approximate matrix is defined by

$$H_{k+1} = H_k + \frac{p^{(k)} p^{(k)T}}{p^{(k)T} q^{(k)}} - \frac{H_k q^{(k)} q^{(k)T} H_k}{q^{(k)T} H_k q^{(k)}} \quad (9)$$

The objective function defined by $\Gamma_{NLR}(X, W) = L_0(W) - L_1(W)$ is maximized to find the demixing matrix W. Two gradient terms $\nabla_W L_0(W)$ and $\nabla_W L_1(W)$ should be determined. For $\nabla_W L_0(W)$, the elements of its jacobian matrix are determined as

$$\frac{\partial L_0(W)}{\partial w_{dj}} = \begin{cases} \sum_{t=1}^{T} \frac{-\sum_{k=1}^{T} \frac{3}{10\sqrt{5}} \cdot \frac{1}{h^2} \mathrm{w}_d\left(\mathrm{x}_t - \mathrm{x}_k\right)\left(x_t^{(j)} - x_k^{(j)}\right)}{\sum_{k=1}^{T} K(\frac{\mathrm{w}_d\left(\mathrm{x}_t - \mathrm{x}_k\right)}{h})} & -\sqrt{5} \le \frac{y_t^{(d)} - y_k^{(d)}}{h} \le \sqrt{5} \\ 0 & otherwise \end{cases} \tag{10}$$

$\nabla_W L_1(W)$ is derived as

$$\nabla_W L_1(W) = \begin{cases} \sum_{t=1}^{T} \frac{-\sum_{k=1}^{T} \frac{D+2}{c_D} \cdot \frac{1}{h^2} W\left(\mathrm{x}_t - \mathrm{x}_k\right)(\mathrm{x}_t - \mathrm{x}_k)^T}{\sum_{k=1}^{T} K_D(\frac{W\left(\mathrm{x}_t - \mathrm{x}_k\right)}{h})} & if \ (\mathrm{y}_t - \mathrm{y}_k)^T(\mathrm{y}_t - \mathrm{y}_k) < 1 \\ 0 & otherwise \end{cases} \tag{11}$$

Before applying the ICA algorithm on the data, it is usually very necessary to do centering and whitening preprocessing. The ICA procedure using quasi-newton method is as follows.

- Centering

$\mathrm{x}_t \leftarrow \mathrm{x}_t - E[\mathrm{x}]$

- Whitening

1. $E[\mathrm{x}\mathrm{x}^T] = \Phi D \Phi^T$
2. $\mathrm{x}_t \leftarrow \Phi D^{-1/2} \Phi^T \mathrm{x}_t$

- The computing steps of DFP algorithm are as follows

1. Initialize $W^{(1)}$ & set parameters η , h and ε
2. Set $H_1 = I_n$, calculate the gradient at $W^{(1)}$, by

$\nabla_W L_0(W^{(1)}) - \nabla_W L_1(W^{(1)})$

3. $d^{(k)} = -H_k g_k, W^{(k+1)} \leftarrow W^{(k)} + \eta d^{(k)}$
4. $\mathrm{w}_d \leftarrow \mathrm{w}_d / \|\mathrm{w}_d\|, d = 1, ..., D$
5. Check the convergence criteria, if $\left\|\nabla_W L_0(W^{(k+1)}) - \nabla_W L_1(W^{(k+1)})\right\| \le \varepsilon$, then stop iteration, make $\bar{W} = W^{(k+1)}$; otherwise, go to step 6
6. If k=n, make $W^{(1)} = W^{(k+1)}$ and go back to step 2; otherwise, conduct step 7
7. Compute $g_{k+1} = \nabla_W L_0(W^{(k+1)}) - \nabla_W L_1(W^{(k+1)})$, $p^{(k)} = W^{k+1} - W^k$, $q^{(k)} = g_{k+1} - g_k$, use the approximate matrix to calculate H_{k+1} and set $k = k + 1$, go back to step 3.

3. Simulation Experiments

In this paper, an improved ICA algorithm for blind separation of speech and audio signals is proposed. The simulation experiments have been conducted in order to investigate the performance of the proposed nonparametric method. The separation performance is evaluated in terms of signal-to-interference ratio (SIR), defined as

$$SIR = 10\log_{10}(\sum_{m=1}^{M} s_m^2 / \sum_{m=1}^{M} (\hat{s}_m - s_m)^2 (dB)) \tag{12}$$

Where the original signal is s_m and the reconstructed signal is $\hat{s}_m$. For comparison, we also implement FastICA algorithm and NLR-ICA based on Gaussian kernel [4].

 Copyright © 2016 SERSC

In the first experiment, the sensitivity of the algorithm to the choice of the bandwidth parameter h has been evaluated. The speech signal from a speaker and the noise signal are shown in Figure 2. The two signals are mixed by a 2×2 matrix A, as given in Figure 3.

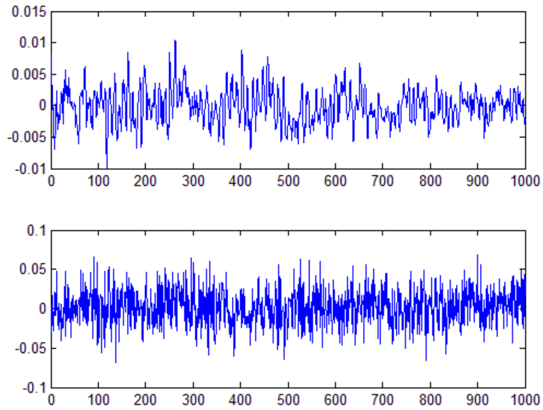

Figure 2. Source Speech (Upper) and Noise (Lower) Signals

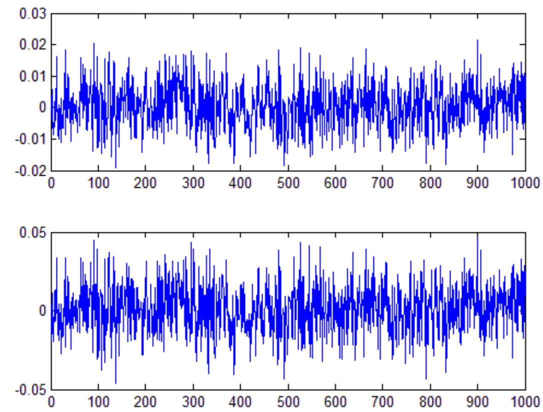

Figure 3. Mixed Signals

For sample size $T = 1000$, the estimated value is $\hat{h}_{opt} = 0.264$. The bandwidth parameter has been allowed to vary up to about 50% from the estimated value, from 0.1 to 0.4. In Figure 4, the theoretical bandwidth value $\hat{h}_{opt} = 0.264$ where the SIR is 18.74 dB and the experimental optimal value $h = 0.28$ where the SIR is as high as 35.19 dB are both marked. Since $\hat{h}_{opt}$ is obtained at the approximating formulation, it is understandable that the maximum SNR is got at a value which is near $\hat{h}_{opt}$. Also, the experiment seems to suggest that the separation performance is relatively sensitive to the particular choice of this parameter in a broad range of values.

Copyright © 2016 SERSC

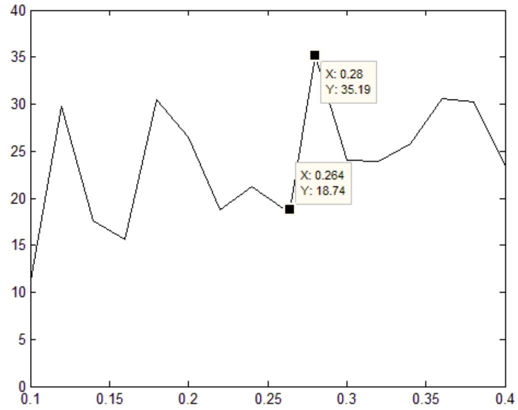

Figure 4. ICA Separation Results with Changing Values of Bandwidth H

In the second experiment, the signal separation has been attempted with the following three algorithms: FastICA, NLR-ICA based on Gaussian kernel (denoted as GNLR-ICA) and NLR-ICA based on Epanechnikov kernel (denoted as ENLR-ICA). As illustrated in Figure 5, Figure 6 and Figure 7, the demixed signals using FastICA, GNLR-ICA and ENLR-ICA algorithms are obtained, for sample size equal to 1000.

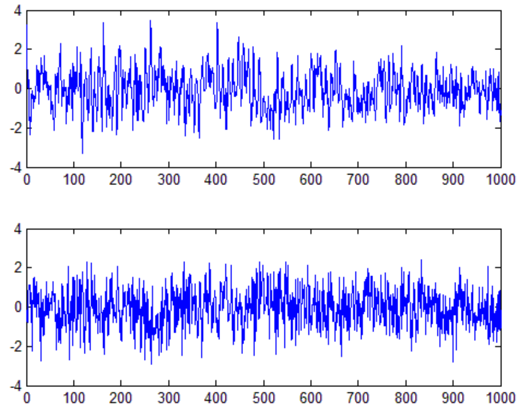

Figure 5. Demixed Speech (Upper) and Noise (Lower) Signals Using Fastica Algorithm

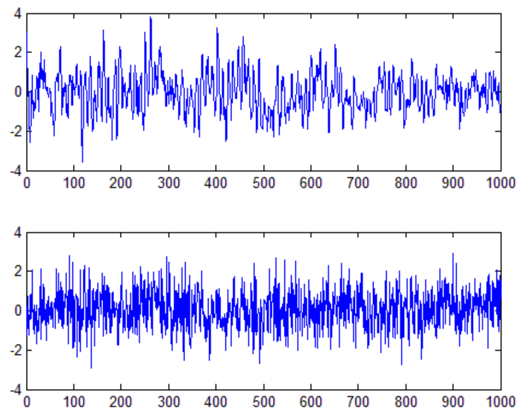

Figure 6. Demixed Speech (Upper) and Noise (Lower) Signals Using GNLR-ICA Algorithm

Copyright © 2016 SERSC

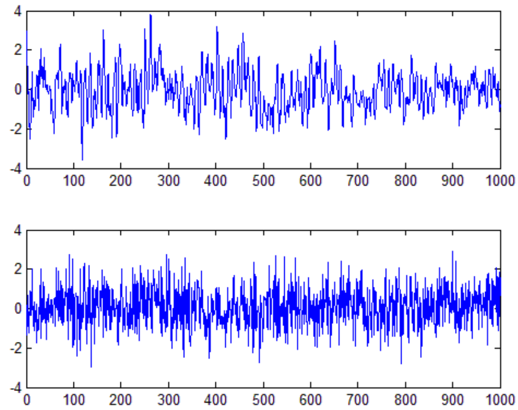

Figure 7. Demixed Speech (Upper) and Noise (Lower) Signals Using ENLR-ICA Algorithm

For different ICA methods, we compare the SIRs for the cases of mixed signals (without ICA), separated signals with FastICA, GNLR-ICA and ENLR-ICA. The results displayed in Table 1 show the obtained SIR. It's noted that the NLR-ICA methods get better performance than the traditional FastICA. Especially, by using the Epanechnikov kernel and setting the optimal value of smoothing parameter, this algorithm will be more powerful both on blind source separation and computing capability. Compared with the GNLR-ICA method, the iteration time has reduced about 30%.

Table 1. Comparison of SIR Using Three Different ICA Methods

	Without ICA	FastICA	GNLR-ICA	ENLR-ICA
SIR(dB)	-3.06	6.17	29.55	35.19

4. Conclusion

In this paper, a novel nonparametric independent component analysis algorithm based on Epanechnikov kernel function is presented. The proposed algorithm uses the kernel estimator to estimate random variable distribution directly and drives the quasi-newton method to optimize the likelihood ratio objective function. In the experiments on blind separation of speech and noise signals, this improved approach has obtained better separation performance compared to conventional FastICA and previous proposed NLR-ICA. The capability of modeling sources, combined with the good convergence properties for small sample size, makes the proposed approach a particularly attractive alternative to current ICA algorithms. In the future, the issues of controlling model size and reducing computational load will be studied.

Acknowledgements

The research work was mainly supported by National Natural Science Foundation of China under Grant No. 61290323, 61203178, 61304214. The authors would like to thank the anonymous reviewers for their valuable comments.

References

[1] A. Hyvärinen and E. Oja, "Independent Component Analysis: Algorithms and Applications". Neural networks, vol.13, no.4, **(2000)**, pp.411-430.
[2] P. Common and C. Jutten, Editors, Handbook of Blind Source Separation: Independent component analysis and applications, Academic Press **(2010)**.
[3] L. H. Hsieh, T. J. Chien, K. Shinoda and S. Furui, "Independent component analysis for noisy speech recognition. Acoustics", Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on. IEEE, **(2009)**, pp. 4369-4372.
[4] T. J. Chien and C. B. Chen, "A new independent component analysis for speech recognition and separation", Audio, Speech, and Language Processing, IEEE Transactions on, vol.14, no.4, **(2006)**, pp.1245-1254.
[5] H. Wei, X. Shi and J. Yang, "Speech Independent Component Analysis", 2010 International Conference on Measuring Technology and Mechatronics Automation, **(2010)**, pp. 445-448.
[6] W. Li, Z. Yi-qi, Y. Gang and M. Yong-zhen, "Time-Frequency Analysis of Cabin Noise Using EEMD-ICA Approaches", International Journal of Control & Automation, vol.7, no.5, **(2014)**,pp.269-280.
[7] K. A. Nandi, J. Gao and X. Zhu, "Independent component analysis-an innovative technique for wireless MIMO OFDM systems", Computers and Devices for Communication, 2009, CODEC 2009. 4th International Conference on. IEEE, **(2009)**, pp. 1-8.
[8] N. Falco, J. Benediktsson and L. Bruzzone, "A study on the effectiveness of different independent component analysis algorithms for hyperspectral image classification", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 7, no. 6, **(2014)**, pp. 2183-2199.
[9] N. Wang, B. Du and L. Zhang, "An Abundance Characteristic-Based Independent Component Analysis for Hyperspectral Unmixing", Geoscience and Remote Sensing, IEEE Transactions, vol. 53, no. 1, **(2015)**, pp.416-428.
[10] A. Hyvärinen, "New approximations of differential entropy for independent component analysis and projection pursuit", Neural Information Processing Systems, vol. 10, **(1998)**, pp. 273-279.
[11] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis", Neural Networks, IEEE Transactions on, vol. 10, no. 3, **(1999)**, pp. 626-634.
[12] R. F. Bach and I. M. Jordan, "Kernel independent component analysis", The Journal of Machine Learning Research**,** vol.3, **(2003)**, pp.1-48.
[13] F. J. Cardoso, "Blind signal separation: statistical principles", Proceedings of the IEEE, vol. 86, no. 10, **(1998)**, pp.2009-2025.
[14] R. Boscolo, H. Pan and P. V. Roychowdhury, "Independent component analysis based on nonparametric density estimation", Neural Networks, IEEE Transactions on, vol. 15, no. 1, **(2004)**, pp. 55-65.
[15] W. B. Silverman, Editor, Density estimation for statistics and data analysis, CRC press, **(1986)**.

Authors

Liang-ju Yu, She is a Master. Department of Automation and Key Laboratory of System Control and Information Processing, Shanghai Jiao Tong University, China.

Research Filed: Blind Source Separation, Speech Signal Processing, Optimization Methods, *etc*.

Gen-ke Yang, he is a Professor, Ph. D. Department of Automation and Key Laboratory of System Control and Information Processing, Shanghai Jiao Tong University, China.

Research Filed: Supply Chain Management, Logistics, Production Planning and Scheduling, Discrete Event Dynamics Systems, Hybrid Systems, Computer Integrated Manufacturing.

 Copyright © 2016 SERSC

Yue Chen, he is a M.E., Ph.D. (Control Theory and Control Engineering), currently, assistant researcher of Shanghai Jiao Tong University, China.

Research Filed: Multi-objective Decision Analysis, Optimization Methods, Anesthesia Process Control, *etc.*

Copyright © 2016 SERSC

 Copyright © 2016 SERSC