1085. An efficient optimized independent component analysis method based on genetic algorithm

Liangmin Li[1], Guangrui Wen[2], Jingyan Ren[3], Xiaoni Dong[4], Lin Liang[5]

[1]Key Laboratory of Automotive Transportation Safety Enhancement Technology of the Ministry of Communication, School of Automobile, Chang'an University, Xi'an 710064, China

[2, 4, 5]School of Mechanical Engineering, Xi'an Jiao Tong University, Xi'an 710049, China

[3]Grid Electric Power Science & Technology Co., Ltd., Xi'an 710075, China

[2, 5]Corresponding authors

E-mail: [1]*leelm@chd.edu.cn*, [2]*grwen@mail.xjtu.edu.cn*, [3]*ren.j.y@163.com*, [4]*birdy_dong@163.com*, [5]*lianglin@mail.xjtu.edu.cn*

Abstract. Three simulation experiments are designed to evaluate and compare the performance of three common independent component analysis implementation algorithms – FastICA, JADE, and extended-Infomax. Experiment results show that the above three algorithms can't separate the mixtures of super-Gaussian and sub-Gaussian precisely, and FastICA fails in recovering weak source signals from mixed signals. In this case an independent component analysis algorithm, which applies genetic algorithm to minimize the difference between joint probability and product of marginal probabilities of separated signals, is proposed. The computation procedure, especially the fitness evaluation when signals are in discrete form, is discussed in detail. The validity of the proposed algorithm is proved by simulation tests. Moreover, the results indicate that the proposed algorithm outperforms the above three common algorithms significantly. Finally the proposed algorithm is applied to separate the mixture of rolling bearing sound signal and electromotor signal, and the results are satisfied.

Keywords: independent component analysis, FastICA, JADE, extended-Infomax, genetic algorithm, rolling bearing.

1. Introduction

Blind source separation (BSS) [1] is one of the research focuses of signal processing area, its purpose is to recover source signals from mixed signals, without any prior knowledge (or with very little information) about the source signals. There are two main methods for BSS: principal component analysis (PCA) [2] and independent component analysis (ICA) [3]. PCA is based on the second-order statistics. Its main purpose is to eliminate the correlation between signals, so it is mainly used to compress the dimension of data. ICA considers the higher-order statistics of data, output signals of ICA are mutually independent. Owing to this property, ICA has been widely used in many areas such as audio processing [4], image processing [5] and vibration signal processing [6].

Algorithms for ICA include FastICA [7], JADE [8], Infomax [9] and many others. FastICA is an iterative algorithm maximizing non-Gaussianity as a measure of statistical independence. As FastICA algorithm has the advantage of quick convergence and free from setting iteration step length, it has become the most popular algorithm for ICA. But the update formula of FastICA includes activation functions that depend on the distribution of source signals, which is unknown for practical problems. Infomax algorithm proposed by Bell and Sejnowski is another widely used ICA implementation algorithm. This algorithm has the same drawback as FastICA – activation function selection problem. This disadvantage led to the appearance of an extension version – extended-Infomax algorithm [10], which is able blindly to separate mixed signals with sub- and super-Gaussian source distributions, and has been proved to outperform traditional Infomax algorithm. In 1999 another widely used ICA algorithm, JADE algorithm, appeared. It's a fourth order technique, involving the diagonalization of cumulant matrices. Lots of literatures proved the success of these algorithms in solving BSS problems. Note that these are mostly normal BSS

problems, i. e. there are only super-Gaussian sources or only sub-Gaussian sources existing, each source has similar percentage in the mixed signals. However in realistic practice there might be some unusual BSS problems appearing. For example, when diagnosing early failure of machine, as the fault phenomenon is faint, the observed signals are the mixtures of weak fault signals and strong noise signals. In order to detect early faults, we have to recover weak fault signal from mixed signals. Moreover sometimes we have to separate the mixed signals of super-Gaussians and sub-Gaussians. These are some unusual BSS problems and few papers discussed the abilities of ICA algorithms dealing with these problems.

This paper is organized as follows. Principle of ICA is introduced in Section 2. Some simulation experiments are designed to evaluate and compare the performances of three common ICA algorithms (FastICA, JADE and extended-Infomax) in separating mixtures of sub-Gaussians and super-Gaussians which cover different percentage of mixtures in Section 3. Furthermore, a novel ICA algorithm, which utilizes genetic algorithm to minimize the difference between joint probability and product of marginal probabilities of separated signals, is introduced in detail. In Section 5 the validity of the proposed ICA algorithm is tested and its performance is compared with those of above three algorithms. Section 6 gives an engineering application of our algorithm, and Section 7 concludes the text.

2. Principle of LSVM

Assume that there exist unknown source signals $s_i(t)$, $i = 1, ..., n$ which are mutually independent, and observed signals $x_j(t)$, $j = 1, ..., m$ which are the linear mixtures of source signals, that is:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t),$$
$$\mathbf{s}(t) = [s_1(t), s_2(t), ..., s_n(t)]^T, \tag{1}$$
$$\mathbf{x}(t) = [x_1(t), x_2(t), ..., x_m(t)]^T,$$

where $\mathbf{A}$ denotes the mixing matrix. The problem is to find a separating matrix $\mathbf{W}$ to retrieve the source signals from $x(t)$ without knowing the mixing vector $\mathbf{A}$. That is, let $\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t)$, so that $\mathbf{y}(t) = [y_1(t), ..., y_n(t)]$ is an estimate of $\mathbf{s}(t)$. The optimality criterion for $\mathbf{W}$ is that the components of $\mathbf{y}(t)$ are mutually independent.

3. Performance of common ICA algorithms

In order to compare the performances of three common ICA algorithms mentioned above, three experiments are designed with different test targets. The first experiment examines the performance in separating mixtures of sub-Gaussians and super Gaussians, the second one is a separation task of weak sources mixed with strong sources, and the final one simulates a BSS problem of strong super-Gaussian mixed with weak sub-Gaussian. The criteria used to evaluate the performances of ICA algorithms include observing and comparing the waveforms of source signals and separated ones intuitively, and a quantitative indicator $R_{SN}(y_i)$ which is the signal noise ratio of separated outputs [11]. $R_{SN}(y_i)$ is denoted as following:

$$R_{SN}(y_i) = 10\log\frac{\mathrm{E}(|s_i|^2)}{\mathrm{E}(|y_i - s_i|^2)}, \tag{2}$$

where s_i is the i-th source signal, y_i is the corresponding separated signal. The larger the $R_{SN}(y_i)$ value, the better the separation effect.

3.1. Sub-Gaussians mixed with super-Gaussians

Fig. 1 is the simulating source signals. The first signal is a voice signal and the second one is a sine wave. Kurtosis values of these signals are 1.1750 and –0.3770 respectively, so it's a separation problem of sub-Gaussian mixed with super-Gaussian. The mixing matrix $\mathbf{A} = \begin{bmatrix} 0.44, -0.53 \\ 0.80, 0.30 \end{bmatrix}$. Fig. 2 shows the mixed signals.
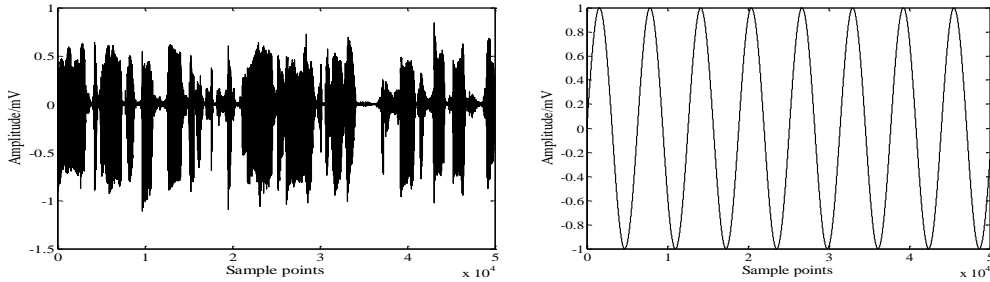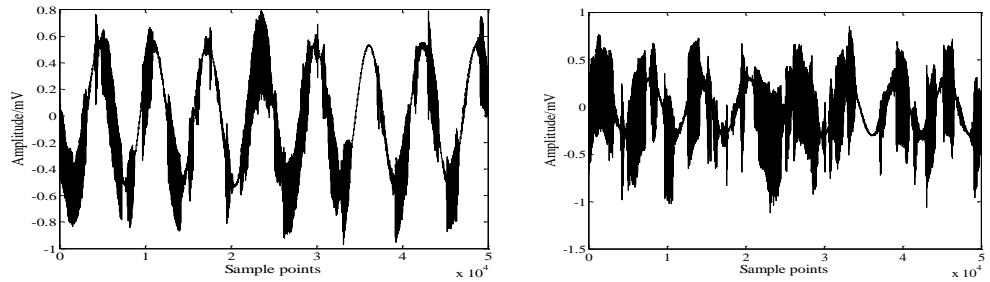

Fig. 1. Source signals

Fig. 2. Mixed signals

According to the distribution of source signals, we choose $G(u) = \frac{1}{a}\log\cosh(au)$ as the activation function of FastICA. Fig. 3-5 show signals recovered by FastICA, extended-Infomax and JADE. Table 1 gives $R_{SN}(y_i)$ of each algorithm, the bolded numbers are the highest values of $R_{SN}(y_i)$.

Table 1. Signal noise ratio of separated outputs
(y_1 corresponds to voice signal, y_2 corresponds to sine signal)

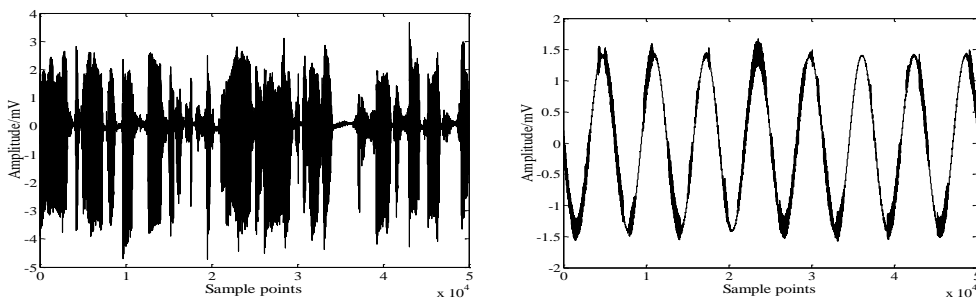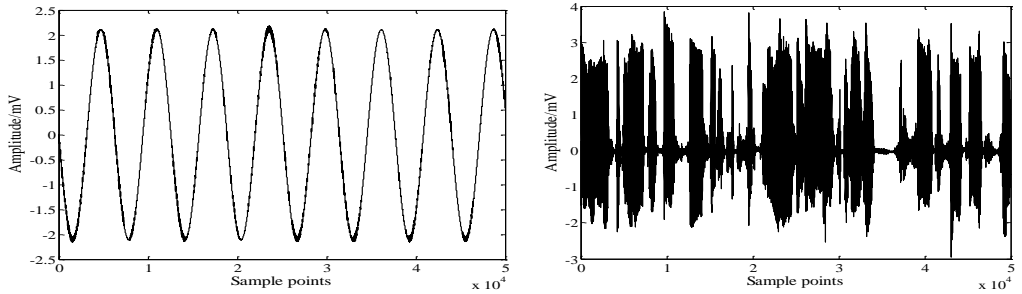| Algorithm	FastICA	Extended-Infomax	JADE
$RSN(y_1)$	81.8523	**87.6789**	80.2098
$RSN(y_2)$	48.8070	**83.8255**	47.0163

Fig. 3. Signals recovered by FastICA
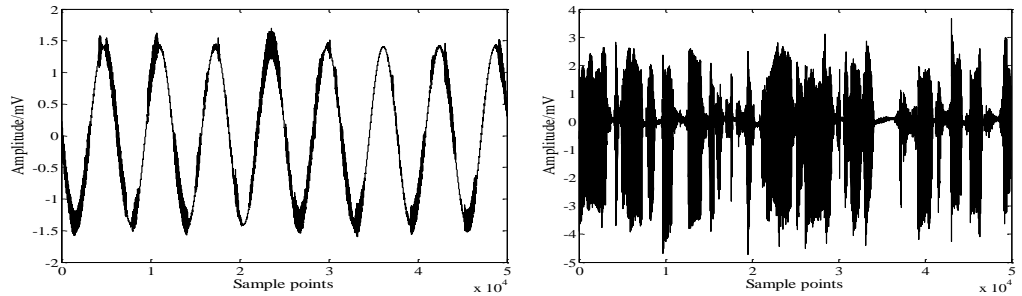
Fig. 4. Separated signals by extended-Infomax

Fig. 5. Results of JADE

The comparison reveals that extended-Infomax algorithm outperforms the other two algorithms, followed by FastICA, and the least effective one is JADE. However even for the best performed algorithm extended-Infomax, there are differences in the waveforms of separated signals and source signals, the separation result is unsatisfied.

3.2. Weak sources mixed with strong sources

In this experiment two sound signals shown in Fig. 6 are taken as source signals. Kurtosis values of these signals are 1.1750 and 5.9013 respectively. The mixing matrix $\mathbf{A} = \begin{bmatrix} 5 & 0.001 \\ 7 & 0.008 \end{bmatrix}$. Fig. 7 shows the mixed signals. Due to the great differences of the coefficients in mixing matrix $\mathbf{A}$, both mixed signals are similar to the first source signal. We consider this as a separation task of weak source and strong source.

Results of extended-Infomax and JADE are shown in Fig. 8 and Fig. 9. FastICA considers the two mixed signals as one and totally fails in recovering source signals. The $R_{SN}(y_i)$ of each algorithm are listed in Table 2.

Figures indicate that both JADE and extended-Infomax algorithms recover the source signals successfully, results of Table 2 reveal that JADE performs slightly better than extended-Infomax.
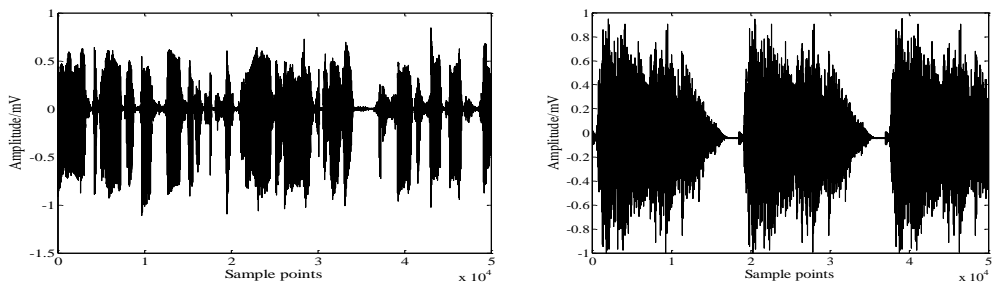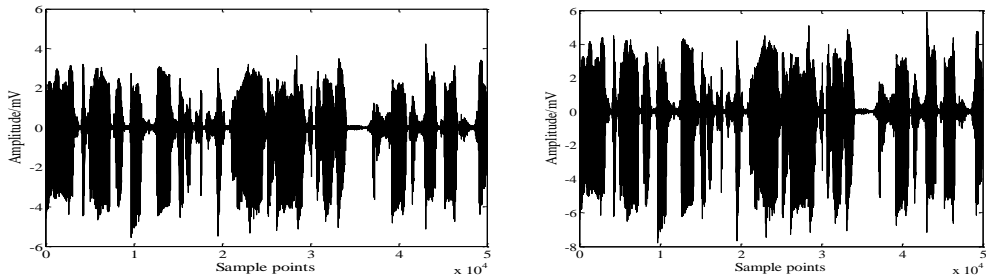

Fig. 6. Source signals
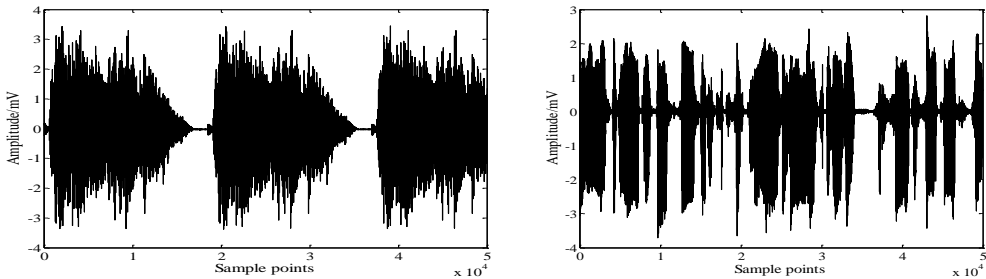
Fig. 7. Mixed signals
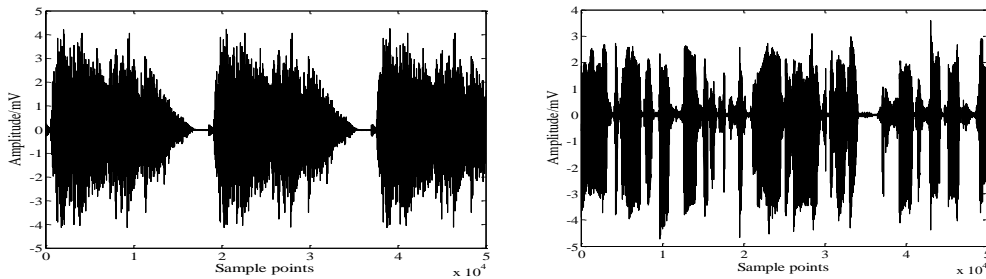

Fig. 8. Results of extended-Infomax

Fig. 9. Results of JADE

Table 2. $R_{SN}(y_i)$ of three algorithms
(y_1 corresponds to the first source signal, y_2 corresponds to the second one)

Algorithm	FastICA	Extended-Infomax	JADE
$RSN(y_1)$	——	137.8288	146.4628
$RSN(y_2)$	——	136.0017	139.9788

3.3. Strong super-Gaussian mixed with weak sub-Gaussian

In this experiment we use the same source signals as those in section 3.1 and the same mixing matrix **A** as that in section 3.2, simulating the BSS problem of strong super-Gaussian mixed with weak sub-Gaussian. Signals recovered by JADE and extended-Infomax algorithms are shown in Fig. 10 and Fig. 11, the $R_{SN}(y_i)$ of each algorithm are listed in Table 3. FastICA algorithm still fails completely in this experiment.

Table 3. $R_{SN}(y_i)$ of three algorithms
(y_1 corresponds to the voice signal, y_2 corresponds to sine signal)

Algorithm	FastICA	Extended-Infomax	JADE
$RSN(y_1)$	——	101.7624	80.1420
$RSN(y_2)$	——	79.9405	46.9442

Figures show that results of extended-Infomax and JADE are all unsatisfied, separated signals are still lightly mixed. Results of Table 3 demonstrate that the performance of extended-Infomax is superior to JADE.

Conclusions can be drawn from these experiments:

(1) All of three ICA algorithms mentioned above can't solve the BSS problem when super-Gaussian sources coexist with sub-Gaussian sources.

(2) FastICA fails in recovering weak source signals from mixed signals. This property makes FastICA not suit for weak signal detection problem, e. g. early fault diagnosis.

(3) The comprehensive performance of extended-Infomax is the best in these algorithms and FastICA performs worst. Actually, researches prove that the performance of FastICA is influenced by the selection of non-linear activation functions [12], thus efficiency of FastICA for practical problems is uncertain.
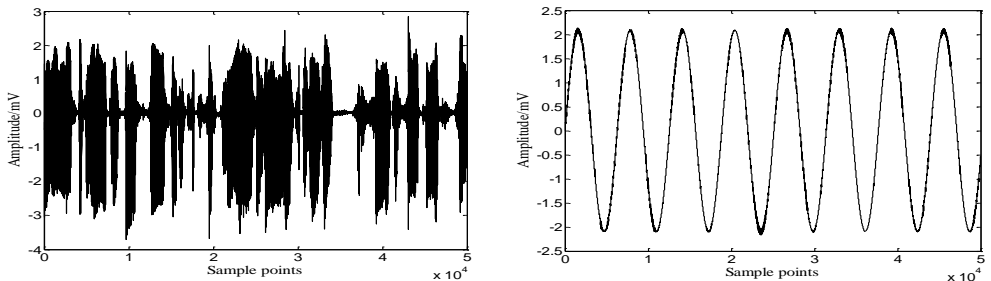

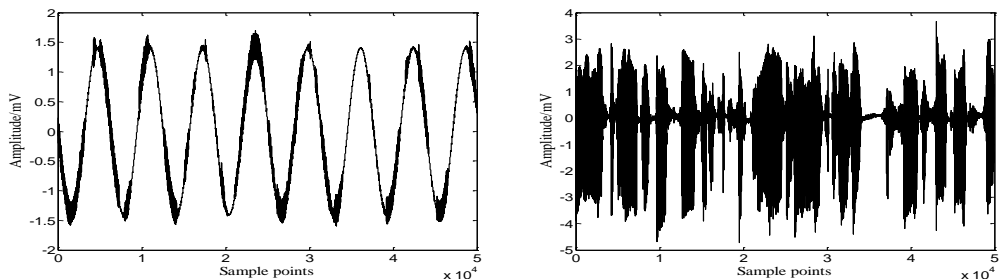**Fig. 10.** Results of extended-Infomax

Fig. 11. Results of JADE

4. Genetic algorithm optimized ICA algorithm

The essence of ICA is to build a separating matrix **W** to maximize the independence of output signals. As it's difficult to get the gradient information of most independent criterions, conventional gradient descent method is inappropriate to deal with this optimization problem, whereas genetic algorithm is a good choice.

4.1. Optimization model

In probability theory, if variables $y_1, y_2, \dots, y_m$ satisfy the following equation, we consider them to be statistically independent:

$$P(y_1, y_2, \dots, y_m) = \prod_{i=1}^{m} P(y_i), \tag{3}$$

where $P(y_1, y_2, \dots, y_m)$ is the joint probability of these variables and $P(y_i)$ is the marginal probability of variable y_i. So the difference between joint probability and product of marginal probabilities of variables is taken as a measure of statistical independence and then the

optimization model is built as following:

$$\text{Minimize } \left| P(y_1, y_2, \dots, y_m) - \prod_{i=1}^{m} P(y_i) \right| \text{ w.r.t. } \mathbf{W}, \tag{4}$$

s.t.: $\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t)$.

4.2. Optimization algorithm and optimization process

As it's difficult to give the derivative function of Eq. (4), a genetic algorithm optimized ICA algorithm is established in this study.

Three major steps in executing genetic algorithm to find the optimum matrix $\mathbf{W}$ are as following:

(I) Initialization.

As $\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t)$ exists, changing values in equal proportions, orders, or signs of row vectors in separating matrix $\mathbf{W}$ wouldn't influence the dependence of separated signals, the range of possible solutions is narrowed in $[-1, 1]$.

Note that if waveforms of mixed signals are similar to each other (just like the example in section 3.2), implying a BSS problem with strong sources and weak sources, there is a great difference in the values of $\mathbf{W}$, we should increase the length of chromosome to search in a finer grid.

Determine the population size and generate the initial individuals randomly.

(II) Iteratively perform the following sub-steps on the population until the termination criterion is satisfied. The maximum number of generation to be run is determined beforehand.

(1) Create new individuals by crossover and mutation operation.

(2) Assign a fitness value to each individual in the population using the fitness measure.

The calculation of objective function defined as Eq. (4) requires estimating the joint probability and marginal probabilities of each variable. Common probability estimation methods include histogram, Rosenblatt method, Parzen kernel estimation and nearest neighbor estimation method, etc., [13] and histogram is the most basic one. It's a rough assessment of the distribution of a given variable with the characteristic of easy calculation. Other methods have relatively higher estimation precision but take much more computing time. Actually the precision of distribution estimation has little influence on the fitness as it values the relative, not absolute, independence of separated signals. Considering all the factors histogram is the final choice of distribution estimation.

Assume there are 2 source signals, that is $n = 2$. Let $X(i)$, $i = 1,2,\dots,N$ and $Y(i)$, $i = 1,2,\dots,N$ denote the separated signals in discrete form, the objective function computation process is as following:

a) Assign a positive number to the variable h which is referred as the bin width of histogram, then the value range of signal X and Y is divided into N_x and N_y disjoint categories (known as bins) separately, and the whole value range is divided into $N_x \cdot N_y$ categories. There is an empirical equation for the bin width h [13]:

$$h = \frac{\max X - \min X}{1.87(N-1)^{0.4}}. \tag{5}$$

b) Count the number of discrete observations that falls into each bin, denoted by D_{ij}, $i = 1,\dots,N_x, j = 1,\dots,N_y$.

c) The frequency of observations in each bin is used to estimate the joint probability of signals X and Y, that is:

$$P(X_i, Y_j) = \frac{D_{ij}}{N}. \tag{6}$$

The marginal probability distribution of signals X and Y is calculated as following:

$$P(X_i) = \frac{\sum_j D_{ij}}{N}, \tag{7}$$

$$P(Y_j) = \frac{\sum_i D_{ij}}{N}. \tag{8}$$

d) Put the above results into Eq. (4) to get the objective function value. Smaller objective function value indicates a better individual.

Note that high estimation accuracy of histogram requires long discrete signals. Fortunately most signals are long enough for histogram, if not, Bootstrap [14] is a proper approach to increase the length of signal.

(3) Select better individuals to form a new population. In order to avoid premature, ranking selection is adopted.

(III) When evolution finished, the best individual of the final population is then the optimum **W**.

5. Performance of proposed algorithm

Same simulation signals as those in section 3 are used to test the performance of proposed algorithm which is denoted by GA-ICA to simplify description.

5.1. Sub-Gaussians mixed with super-Gaussians

Basic parameters used in genetic algorithm are listed in Table 4. The separated signals are shown in Fig. 12, $R_{SN}(y_1) = 160.1629$, $R_{SN}(y_2) = 186.3932$. Apparently performance of GA-ICA is far superior to other three ICA algorithms, $R_{SN}(y_1)$ of GA-ICA is 82.7 % higher than the best value of other three algorithms, and $R_{SN}(y_2)$ has an increase of 122.4 % over the best value of other three algorithms.
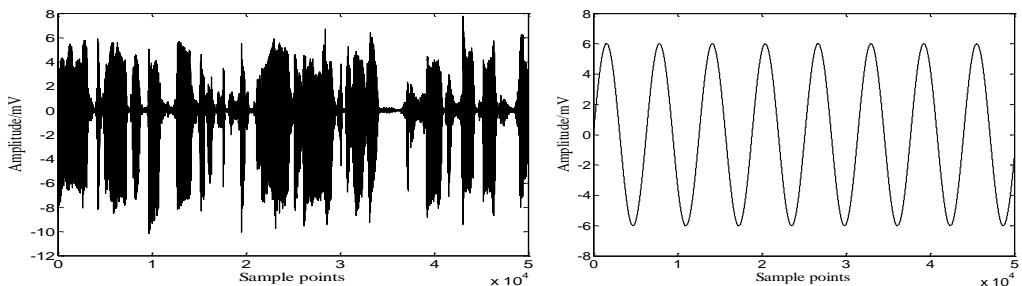

Fig. 12. Results of GA-ICA

Table 4. Basic parameters of genetic algorithm

Parameter name	Value	Parameter name	Value
Scale of population	60	Crossover probability	0.9
Mutation probability	0.01	Maximum generation	100
Chromosome length	40		

This experiment proves the validity of GA-ICA in dealing with sub-Gaussians mixed with super-Gaussians BSS problem.
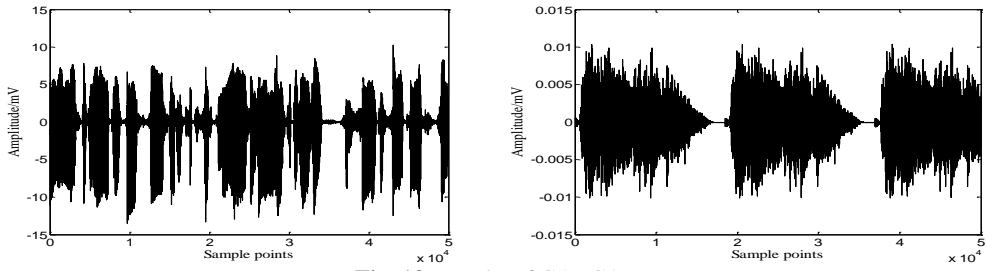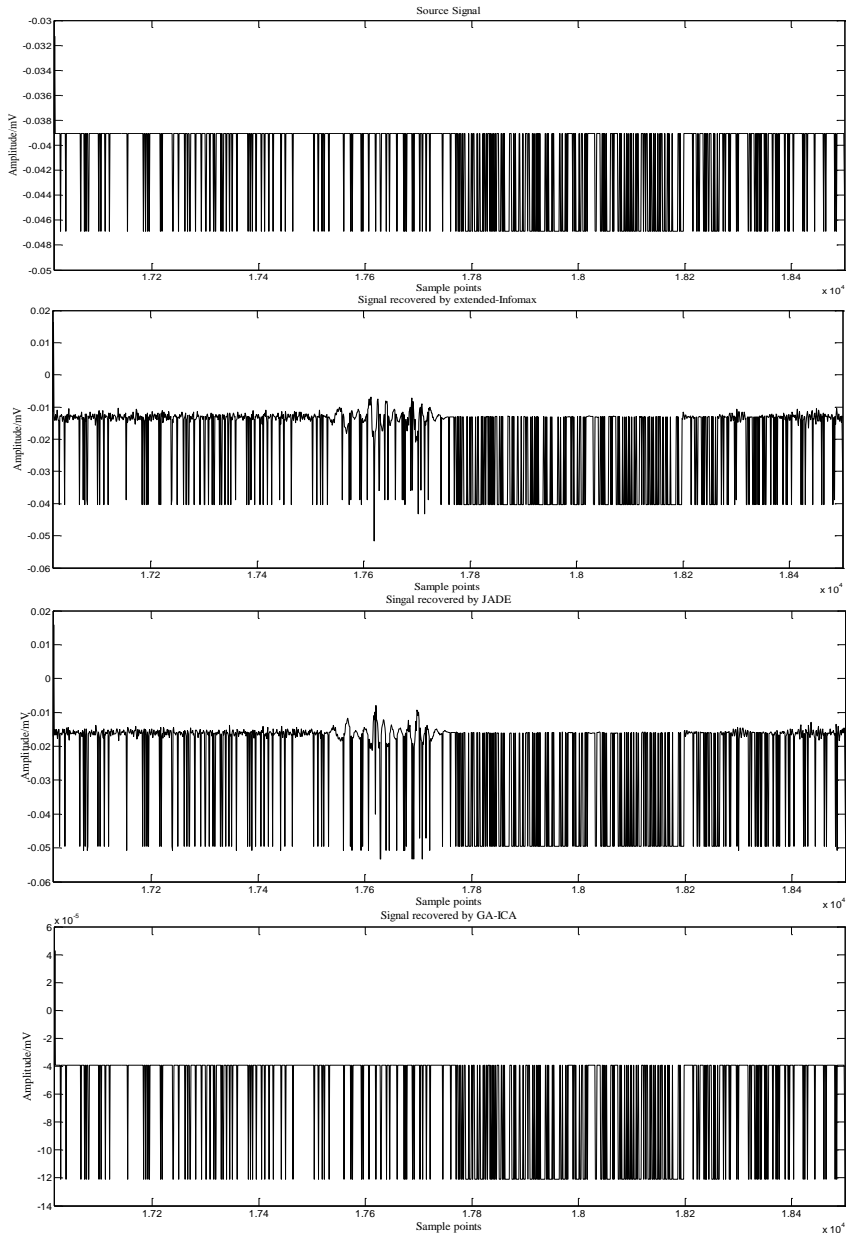
Fig. 13. Results of GA-ICA

Fig. 14. Partial enlarged details of source signal 2 and those recovered by different ICA algorithms

5.2. Weak sources mixed with strong sources

Most basic parameters of genetic algorithm used in this section are the same with those in section 5.1, except chromosome length. Just as mentioned above, this is a BSS problem with weak source and strong source, chromosome length is increased to 60.

Fig. 13 shows the results of GA-ICA. $R_{SN}(y_1) = 212.6588$, 45.2 % higher than the best value of other algorithms, $R_{SN}(y_2) = 571.8645$, an increase of 308.5 % over the other algorithms. Fig. 14 are the partial enlarged details (sample points: 17016-18500) of source signal 2 which simulates a weak source and those recovered by different ICA algorithms. Comparison reveals that there are some distortions existing in the signals recovered by extended-Infomax and JADE, while GA-ICA recovers source signals without any distortion, thus proves the outperformance of GA-ICA furtherly.

This example demonstrates that GA-ICA performs far more excellently in detecting weak useful signals (just like the signal y_2 in this example) from strong background signals (simulated by the signal y_1) than other three algorithms do. This property makes GA-ICA especially suitable for dealing with early fault diagnosis problem.

5.3. Strong super-Gaussian mixed with weak sub-Gaussian

Most basic parameters of genetic algorithm remain unchanged, chromosome length is 120. Separation results of GA-ICA are shown in Fig. 15. $R_{SN}(y_1) = 202.4397$, $R_{SN}(y_2) = 195.3203$, which increase 98.9 % and 144.3 % over the best value of other three algorithms respectively. The qualities of signals recovered by GA-ICA are much better than those of other three algorithms.
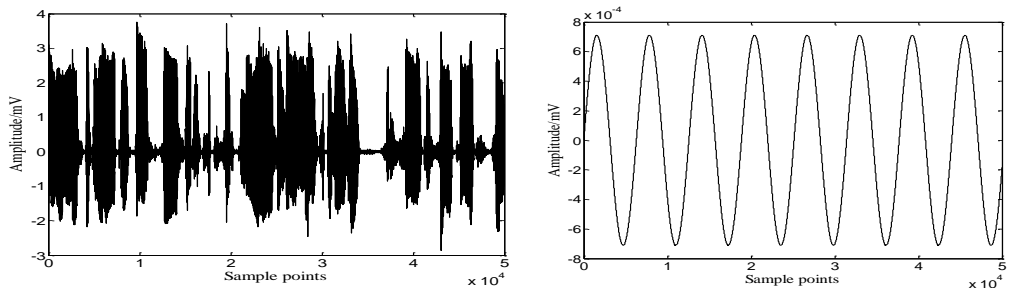

Fig. 15. Results of GA-ICA

Three simulation experiments reveal that GA-ICA outperforms three common ICA algorithms significantly and is universal for any BSS problem.

6. Engineering application

The proposed ICA algorithm is applied to deal with sound signals sampled from a rolling bearing test rig. Sound sources include a NSK 6308 rolling bearing, which is with a fatigue spall fault but the damage source is unkown, and an electric motor, which is in normal condition. Two sound level meters are placed between the motor and the rolling bearing with a distance of 25 cm to pick up the sound signals. The observed signals with sample frequancy $f_s = 40$ kHz are shown in Fig. 16.

As the velocity of sound is 340 m/s and the distance between two sound level meters is 25 cm, the time delay of two source signals is less than 1 ms. We consider the observed signals as the linear mixtures of the motor signal and the rolling bearing signal, i. e. a linear BSS problem.

Results of our algorithm are shown in Fig. 17. When surface damage failure occurred, the collisions of rolling bearing components are amplified, and then periodic pulses appear in the

sound signal produced by the rolling bearing. The sound signal produced by a normal motor may contain mostly low-frequency components. Apparently the first separated signal is the recovery of the signal produced by motor and the second one corresponds to the signal produced by faulty rolling bearing.
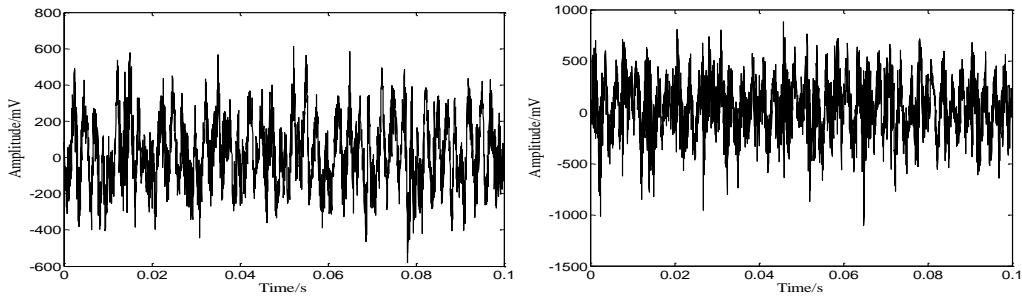

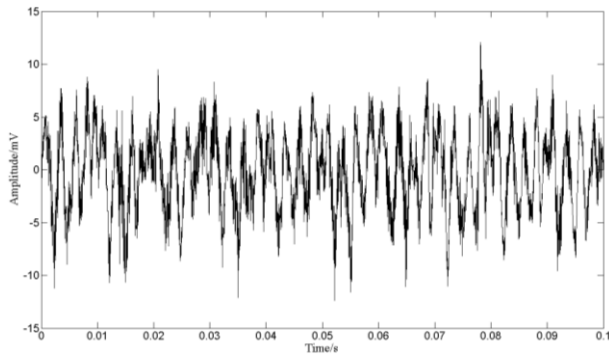**Fig. 16.** Sound signals picked up by sound level meters

Fig. 17. Sound signals recovered by the proposed ICA algorithm

The geometric parameters of the NSK 6308 rolling bearing are: the number of balls $n = 8$, the bearing ball diameter $d = 15$ mm, the pitch diameter $D_p = 65$ mm, the contact angle between the ball and the race $\theta = 0°$ and the rotational frequency of the shaft with rolling bearing mounted is 52.3 Hz. Then characteristic frequencies with fault on the inner race, outer race, or bearing balls are 257.6 Hz, 161.2 Hz and 107.2 Hz respectively. The period time of pulses in the signal of rolling bearing is 0.0062 s (see Fig. 17), which corresponds to the frequency of 161.2 Hz. It reveals that the fatigue spall damage occurs on the outer race. Inspecting bearing components carefully, a metal flaking, about 7 mm^2 large and 0.2 mm depth, is found in the middle of the outer race, indicating the correctness of our analysis and the effectiveness of the proposed algorithm.

7. Conclusion

A novel ICA algorithm is proposed in this paper. This algorithm applies genetic algorithm to minimize the difference between joint probability and products of marginal probabilities of separated signals. Three simulation experiments are designed to evaluate and compare the performance of FastICA, JADE, extended-Infomax and GA-ICA. Comparison reveals that for super-Gaussian mixed with sub-Gaussian problem only GA-ICA recovers the source signals precisely, there are some distortions existing in the signals recovered by the other three ICA algorithms; for strong source mixed with weak source problem, GA-ICA, extended-Infomax and JADE recover the source signals successfully, GA-ICA performs best, $R_{SN}(y_i)$ of GA-ICA has an increase of 45.2 % and 308.5 % over the best value of other algorithms respectively, while FastICA fails in this problem; for strong super-Gaussian mixed with weak sub-Gaussian problem,

only GA-ICA recovers the source signals precisely, FastICA still fails. So the conclusion is drawn that GA-ICA outperforms three common ICA algorithms significantly and is a universal technique for any BSS problem. Finally the proposed ICA algorithm is applied to deal with sound signals sampled from a rolling bearing test rig. Analysis of the separated signal reveals the damage source of rolling bearing, indicating the effectiveness of the proposed algorithm.

Acknowledgements

References

[1] **Jutten C., Herault J.** Blind separation of sources, Part 1: An adaptive algorithm based on neuromimetic architecture. Signal Processing, Vol. 24, Issue 1, 1991, p. 1-10.
[2] **Common P.** Independent component analysis, a new concept. Signal Processing, Vol. 36, Issue 3, 1994, p. 287-314.
[3] **Oja E.** Principal components, minor components, and linear neural networks. Neural Networks, Vol. 5, Issue 6, 1992, p. 927-935.
[4] **Mark D. Plumbley, Samer A., et al.** Sparse representations of polyphonic music. Signal Processing, Vol. 86, 2006, p. 417-431.
[5] **Wenjie Dun, Zhichun Mu, Hailong Zhao** Multi-modal recognition of face and ear images based on two types of independent component analysis. Journal of Computational Information Systems, Vol. 5, 2008, p. 1997-1983.
[6] **Haiqing Qin, Kejun Xu, Jianping Ou** Analysis of aero engine vibration signal based on blind source separation technology. Journal of Beijing University of Aeronautics and Astronautics, Vol. 11, 2010, p. 1307-1310.
[7] **Hyvarien A.** Fast and roubust fixed-point algorithms for independent component analysis. IEEE Transactions on Neural Networks, Vol. 10, Issue 3, 1999, p. 626-634.
[8] **J. F. Cardoso** High-order contrasts for independent component analysis. Neural Computation, Vol. 11, Issue 1, 1999, p. 157-192.
[9] **Bell A. J., Senjnowsk T. J.** An information-maximization approach to blind separation and blind deconvolution. Neural Computation, Vol. 7, Issue 6, 1995, p. 1129-1159.
[10] **T. Lee, M. Girolami, T. Sejnowski** Independent component analysis using an extended-Infomax algorithm for mixed sub-Gaussian and super-Gaussian sources. Neural Computation, Vol. 11, Issue 2, 1999, p. 417-441.
[11] **A. G. Westner** Object Based Audio Capture: Separating Acoustically Mixed Sounds. Master's Thesis, MIT, 1999.
[12] **Liangmin Li** New blind source separation method based on genetic algorithm. Journal of Xi'an Jiaotong University, Vol. 39, Issue 7, 2005, p. 740-743.
[13] **Chen Xiru, Chai Genxiang** Non-Parametric Statistics. Shanghai, East China Normal University Press, 1993, p. 247-255.
[14] **Zoubir A. M., Iskander D. R.** Bootstrap methods and applications: a tutorial for the signal processing practitioner. IEEE Signal Processing Magazine, Vol. 24, Issue 4, 2007, p. 10-19.