Constrained Genetic Algorithm Based Independent Component Analysis

D.P.Acharya[1], G.Panda[1], Senior Member IEEE, Y.V.S.Lakshmi[2]
[1]Department of Electronics & Communication Engineering, NIT, Rourkela -769008,
[2]Center for Development of Telematics, Bangalore – 560 100.
d_p_acharya@rediffmail.com, ganapati.panda@gmail.com

Abstract- **Independent Component Analysis, a computationally efficient statistical signal processing technique, has been an area of interest for researchers for many practical applications in various fields of science and engineering. The present paper proposes a constrained genetic algorithm optimization based independent component analysis assuming a noise free Independent Component Analysis (ICA) model. It investigates on the application and performance of the popular evolutionary computation technique GA in independent component analysis problem. It is observed that the proposed constrained genetic algorithm optimization based ICA overcomes the long standing permutation ambiguity and recovers the independent components in a fixed order which is dependent on the statistical characteristics of the signals to be estimated. The constrained GA based ICA has also been compared with the most popular fast ICA algorithm.**

I. INTRODUCTION

Independent Component Analysis (ICA) is a statistical signal processing technique in which the goal is to find a linear representation of nongaussian data so that the components are statistically independent or as independent as possible [1]. Such a representation seems to capture the essential structure of data in many emerging new practical application areas, such as blind separation of mixed voices or images, analysis of several types of data, feature extraction, speech recognition, data communication, sensor signal processing and biomedical signal processing [2, 3, 4]. In the classical application of ICA model in blind source separation (BSS) the independent components are called as source signals.

There has been several algorithms for ICA in literature since its inception by Comon [1]. One of the promising work by Jutten [5] is inspired by neural networks which is based on canceling the non linear cross correlations. However the algorithm converges only under severe restrictions [6]. A class of algorithms consists of those based on maximization of network entropy (infomax). These are based on stochastic gradient ascent of some functions and are equivalent to maximum likelihood approach [7, 8]. Here Newton method is used to maximize the likelihood. Though the method gives faster convergence but involves a matrix inversion in each iteration. In [9] non linear extensions of the neural Principle Component Analysis (PCA) based algorithms have been proposed. Above algorithms based on a stochastic gradient descent are adaptive in nature which creates problem in many practical applications where there is no necessity of adaptation. Apart from this they depend on choice of some learning rate parameter and has sometimes slow convergence. So batch fixed-point iteration algorithm popularly known as fastICA is introduced in [10]. This algorithm either estimates all the independent components (ICs) one-by-one using deflation decorrelation approach or estimates all ICs parallely using symmetric decorrelation. FastICA algorithm has become most popular due to its fastness and robustness. Tensor based algorithms use fourth order cumulant tensor for ICA estimation [11]. These non adaptive batch algorithms use tensorial techniques as eigenmatrix decomposition. However such algorithms are difficult to program for their sophisticated matrix manipulations and hence are practically complex.

Unfortunately all the existing methods discussed above do not find a global optimum once the algorithm reaches a local optimum. It gets stuck in the valley and is unable to jump the surrounding hills. The non linear objective functions for ICA being multimodal, this problem becomes more prominent. Therefore, good initial values are important in initializing these algorithms. In addition to the problem of getting trapped in a local optimum, these algorithms have the ambiguities like scaling and permutation. This means that the independent components (ICs) are estimated only up to a multiplicative scalar constant and the order in which the ICs are estimated are not known. In [12] attempts have been made to overcome the permutation ambiguity in the frequency domain only but the ambiguity remains still unresolved in the time domain.

The evolutionary computation based optimization techniques like Genetic Algorithms (GA) have an inherent characteristic to converge to the global optimum. In a preliminary study done by Yoshioka et al. [13] GA is applied to separate original images from noise corrupted images by directly minimizing Kullback Leibler (KL) divergence. However prior probability estimation of the sources is prerequisite of the method. Tan and Wang in [14] used GA to solve the nonlinear BSS problem using higher order statistics where the sources have been estimated regardless of the indeterminacies of permutation and scaling. Rajas et al. in [15] applied the GA based method for signal separation from their post nonlinear mixtures. GA has been directly applied to ICA problem for first time in [16] for denoising the electrocardiogram (ECG) signals where the method estimates only one independent component i.e. the ECG signal. However

2443

all the applications of GA to BSS problem still have the permutation indeterminacy. In practical applications where we need all the signals or independent components to be recovered simple application of GA seems to be inadequate.

In the present paper we propose a novel Independent Component Analysis algorithm using optimization based on the constrained GA (CGAICA) which recovers all the ICs using the popular deflation approach. Our method overcomes the long standing permutation ambiguity and recovers the ICs in a fixed order which is dependent on the statistical characteristics of the signals to be estimated. It offers almost comparable performance to the most popular fast ICA algorithm.

This paper is organized as follows. In section II, we present the Independent Component Analysis problem. Section III discusses different contrast functions of ICA for optimization. In section IV, we propose the constrained GA based ICA algorithm. In section V, the simulation experiments carried out to show the effectiveness of this algorithm has been described. Section VI the results has been illustrated and discussed. Finally, in section VII, we provide the conclusions.

II. INDEPENDENT COMPONENT ANALYSIS (ICA)

ICA is a computationally efficient statistical signal processing technique for revealing hidden factors that underlie sets of random variables, measurements or signals. The basic idea of the ICA is to minimize the dependency among the output components. A generative model for the observed multivariable data, which is typically given as a large database of samples is defined by ICA. The data variables in the model are assumed to be linear or non-linear mixtures of some unknown latent variables and the system of mixing is unknown. The extraction of source in this process is done based on the assumption that the latent variables are non-Gaussian and statistically independent [20].

Suppose a set of observations of random variables is $(x_1(t), x_2(t),..., x_n(t))$ where t is the time or the sample index and they are generated from a linear mixture of sources $(s_1(t)\ s_2(t)...s_n(t))$ that are statistically independent. This is expressed in the following form

$$[x_1(t)\ x_2(t)...x_n(t)]^T = A[s_1(t)\ s_2(t)...s_n(t)]^T \quad (1)$$

where A is some unknown mixing matrix and $^{\mathrm{T}}$ stands for the transpose operator of a matrix.

Independent component analysis estimates both A and $s_i(t)$ when only the observations $x_i(t)$ are at hand. Two fundamental restrictions of the model are that, firstly we can only estimate non-Gaussian ICs (except if just one of the ICs is Gaussian), and secondly, we must have at least as many observed linear mixtures as ICs. Moreover, the ICs and the columns of A can only be estimated up to a multiplicative constant and their order of appearance is not known. For

simplicity we assume a linear and noise free model of ICA here.

The estimation of the data model of ICA is usually performed by formulating an objective function and then minimizing or maximizing it. Such a function is often called a contrast function or cost function which in the language of evolutionary computational techniques is known as fitness function or nutrient function. So the ICA method combines an objective function and an optimization algorithm. The statistical properties such as consistency, asymptotic variance and robustness of the ICA depend on the choice of the contrast function. The algorithmic properties such as convergence speed, memory requirements and numerical stability depend on the optimization algorithm.

III. CONTRAST FUNCTIONS FOR ICA

A. Kurtosis:

An interesting characteristic of ICA is that maximizing non-Gaussianity of data allows the ICs to be obtained. Non-Gaussianity is measured by various methods. The first method uses kurtosis which is the fourth order moment of the random data. Given some random data, y, the kurtosis denoted by $kurt(y)$ is defined as

$$kurt(y) = E\{y^4\} - 3(E\{y^2\})^2 \quad (2)$$

where $E\{.\}$ is the statistical expectation operator. For Gaussian variables kurtosis is zero and for non-Gaussian variables it is non zero. Particularly when this value is positive the distribution is known as super Gaussian and when negative it is called sub Gaussian. The demixing matrix W is so chosen that $kurt(w\,x)$ gives a maximum where x is a column vector of the matrix in left side of (1) and w is column of W.

B. Negentropy:

A second method of measuring non-Gaussianity is negentropy which is based on the information theoretic differential entropy. The entropy of data is related to the information that is observed. The more random and unpredictable the data is, the larger entropy it will have. The entropy S of a random variable y with a density of $p(\eta)$ is

$$S(y) = -\int p_y(\eta) \log p_y(\eta) \, dp_y(\eta) \quad (3)$$

Of all observed random variables with unit variance a Gaussian variable has the largest entropy value. Differential entropy normalized with respect to Gaussian variables result in negentropy. Negentropy denoted by H is defined as

$$H(y) = S(y_{gauss}) - S(y) \quad (4)$$

y_{gauss} is a Gaussian random variable with the same correlation and covariance as y. Since the negentropy is normalized, it is always non-negative and is zero if y is Gaussian distributed.

C. Mutual Information:

Mutual information is a natural measure of dependence between random variables. If y is a n-dimensional random variable and $p_y(\eta)$ its pdf then vector y has mutually independent components if and only if

$$p_y(\eta) = p_{y_1}(\eta_1) \cdot p_{y_2}(\eta_2) \dots p_{y_n}(\eta_n) \qquad (5)$$

A natural way of checking whether y has ICs is to measure a distance between both sides of above equation

$$\delta(p_y, \prod_i^n p_{y_i}) \qquad (6)$$

Average mutual information of y as given by Comon [1] is given by

$$I(p_y) = \int p_y(\mu) \log\left(p_y(\mu) / \prod p_y(\mu)\right) d\mu \quad (7)$$

Average mutual information vanishes if and only if the variables are mutually independent and strictly positive otherwise. In terms of negentropy mutual information is written as

$$I(y_1, y_2, \dots, y_n) = J(y) - \sum_i J(y_i) \qquad (8)$$

The ICA of a random vector X for an invertible transformation Y=WX where the matrix W is determined so that the mutual information of the transformed components y_i is minimized.

D. Approximations of Negentropy:

Negentropy may be practically difficult and require complex computations. Hence the approximations to negentropy in terms of expectations of certain nonlinear functions G is expressed as

$$H(y) = K[E\{G(y_i)\} - E\{G(v)\}]^2 \qquad (9)$$

Where k is constant and v is a Gaussian variable of zero mean and unit variance. The key is to choose the nonlinear function G correctly. The frequent choices of G are

$$G_1(y) = \frac{1}{a_1} \log \cosh(a_1 y) \qquad (10)$$

$$G_2(y) = -\frac{1}{a_2} \exp(-a_2 y^2 / 2) \qquad (11)$$

$$G_3(y) = \frac{1}{4} y^4 \qquad (12)$$

where a_1 and a_2 are constants.
Under the approximation

$$E\{(w_i^T z)(w_j^T z) = \delta_{ij} \qquad (13)$$

above equation simplifies to

$$H(w) = E\{G(w^T z)\} \qquad (14)$$

ICs are revealed when the function $H(w)$ is maximized.
E. Choice of Practical Contrast Function:

The properties like consistency, asymptotic variance and robustness are considered for choosing the a contrast function for practical applications. The major drawback of kurtosis is that it is sensitive to outliers in data. G_1 is a good general purpose contrast function. For highly super Gaussian ICs and robustness G_2 is a better choice. For reducing computational overhead a piecewise linear approximations of G_1 and G_2 is good choice.

IV. CONSTRAINED GA BASED ICA ALGORITHM

A.   The Genetic Algorithm:

The operations like encoding of parameters, initialization of population, mate selection, crossover, mutation and population replacement constitute a canonical GA. These operations of a GA are described in the following manner
Encoding: The parameters of the independent component analysis system to be optimized are generally encoded into genes and chromosomes (also called individuals) as a string of binary digits using one's complement representation. The parameters are assumed to be bounded in the region.

$$|\theta_k| \le b_k \text{ for } k = 1, \dots, h$$

where h represents the number of parameters. The length of the gene and individuals can be computed as the length of the binary string B_k to be encoded θ_k based on b_k and the desired accuracy. Other encoding methods are also possible.
Initial population generation: The initial population is generated randomly in the range of each parameters. Therefore, at the beginning of the separating procedure, *N* individuals are generated as random binary string.
Evaluation of fitness: After the initial population generations, the fitness of each individual is determined. Fitness is a numeric index to measure the effectiveness of each individuals of the population as a solution, which is usually utilized to select members of the population for reproduction. For the ICA problem, we can define the fitness function based on defined cost function in (2) or (14).
Selection Operation: A pair of individuals is selected from the current population for mating using tournament selection.
Crossover Operation: A multipoint crossover with probability P_c is applied to the newly selected (parents) individuals to generate two offspring. Specifically, the number of crossover points in our application is equal to the number of the parameters to be optimized.
Mutation Operation: Random mutation operator is applied to the newly generated offspring to prevent from premature convergence. It randomly alters the gene from zero to one or from one to zero with a probability expressed by, P_m, where P_m is called mutation probability.
B.   Constrained GA based ICA Algorithm

With Genetic Algorithm based optimization we use kurtosis as the contrast function to be optimized. A random column

vector w which is represented as two chromosomes is used to find the linear transformation $w^T z$.

Step-1: Data Centering

The mean $X_m = (x_{1m}, x_{2m}..., x_{nm})^T$ of the observed mixed signal data $X = (x_1, x_2..., x_n)^T$ is computed and the mean is subtracted from the observed data set to make it zero mean.

$$X_c = X - X_m \tag{16}$$

Step-2: Whitening

The covariance matrix $CovX$ of the centered data X_c is computed. The eigenvalue decomposition of $CovX$ is performed. If D is the eigenvalue matrix and E is the eigenvector matrix then

$$Z = D^{-1/2} E * X_c \tag{17}$$

Step -3: GA-based ICA Iteration

The following iterative procedure is followed to implement GA based ICA algorithm.

1) An initial population $\{\hat{\theta}_i\} N$ of size N is created
from a random initial set of parameter. The encoding length of each parameter is 15 bits. By decoding the individual to get the parameter of the system, the fitness for each individual is evaluated.

2) Two mates are selected for reproduction with probabilities proportional to their using tournament selection.

3) The multipoint crossover operator with crossover operator with crossover probability P_c is applied to the two mates and a pair of offspring are generated.

4) The mutation operator with probability P_m is applied to the newly generated offspring.

5) The fitness value for the off spring are computed after they are decoded as the parameter sets of the parametric system.

6) Steps 2-5 are repeated until an entirely new population of individuals is generated.

7) The previous population is replaced with the new population with the addition of an elitist selection.

8) If the stopping criterion is satisfied, go to step 11.

9) If generation number is greater than a predetermined value go to step 2.

10) Reinitialize the population survival, go to step 2.

11) Output the individual with the best fitness value and terminate the iterative procedure.

Step-4: Evaluation of Second Independent Component

To estimate the other ICs step 3 of the algorithm is repeated for getting weight vectors $w_2,..., w_n$. To prevent different vectors from converging to the same optimum and hence the same IC, the weight vectors are decorrelated using Gram-Schmidt like orthogonaliztion. When p vectors $w_1,..., w_p$ have been estimated, step 3 is run for w_{p+1} and after every iteration step the following iteration steps are performed.

$$w_{p+1} = w_{p+1} - \sum_{j=1}^{p} (w_{p+1}^T w_j) w_j \tag{21}$$

$$w_{p+1} = \frac{w_{p+1}}{\sqrt{w_{p+1}^T w_{p+1}}} \tag{22}$$

Above equations constrain the GA based optimization process.

V. SIMULATION EXPERIMENT

In the experimental studies for the verification of the validity and performance of the proposed constrained GA optimization based ICA algorithm, programs for separating the signals blindly from their observed mixtures were written. In the simulation environment, the two signals were mixed by a known matrix A and the mixed signals were the inputs to the CGAICA algorithm for separation. Two different examples were taken to verify the separation capability of the proposed algorithm. For a particular example, the parameters such as a number of chromosomes N_c, number of bits of binary coded GA N_b, number of generations N_{re}, probability of crossover P_c and probability of mutation P_m parameters are tuned, to get the proper separation. In this simulation for CGAICA we have considered the following typical values: $N_c = 8, p = 2, N_b = 15, N_{re} = 10, P_c = 0.85$ and $P_m = 0.01$.

The separation performance parameter, the mean square error was evaluated. For the first example, separation was performed by using contrast functions kurtosis and approximations to negentropy with function as in (10).The minimum value of the reciprocal of fitness function J is plotted against the number of its evaluations.

VI. RESULTS & DISCUSSIONS

The separation capability of CGAICA was verified through following three examples.

Example 1:

A random binary wave and a sine wave with 400 samples as as shown in fig 1 are mixed by the mixing matrix

$$A = \begin{pmatrix} 0.9121 & 0.2292 \\ 0.4763 & 0.7348 \end{pmatrix} \tag{23}$$

Their mixtures are represented in fig 2. Using CGAICA the two signals or ICs are recovered in the decreasing order of the value of their contrast function. For the case of kurtosis as contrast function the separated signals are depicted in fig 3.Random binary signal has more kurtosis value than the sine wave used. So it is observed that the random binary is recovered first and the sine wave comes as the second IC. The same random binary signal and sine wave were considered for 400 samples each and mixed by mixing matrix A as in (23). Applying CGAICA algorithm the signals were separated

clearly shown in fig 3 in the case of 400 samples. Two sets of tuned values of the optimization parameters for CGAICA algorithm with kurtosis as the contrast function and the corresponding mean square error (MSE) for random binary signal is summarized in table 1.

Example 2:

In this example two speech signals (fig 4) with 50000 samples each were taken and instantaneously mixed by the artificial mixing matrix A as given in (23) which is shown in fig 5. The mixed signals were applied to CGAICA algorithm for separation using kurtosis as the contrast function. The estimated speech signals show a clear separation from the mixtures which is depicted in fig 6. first followed by the speech signal with lower value of kurtosis.

A. Comparison with Gradient Based Optimization Scheme

The signal separation as done in example 1 was also performed by the most popular fast ICA algorithm which is a gradient based scheme to have a comparison with our proposed CGAICA algorithm. Table 2 summarizes the typical MSE values estimated for random binary and sine wave components for fast ICA and CGAICA algorithms. We have chosen the GA with 15 bits which is most commonly considered for binary GAs. If we increase the number of bits for binary coding then MSE performance improves significantly which is very much natural for binary GAs as with higher number of bits they tend towards real coded GA. It is clearly observed that CGAICA yields almost comparable performance as that of the most popular fast ICA algorithm. The MSE performance of fastICA algorithm varies with varying initialization.

B. Convergence

In GA schemes, a *gene* is converged when 95% of the population share the same value. To know the convergence of CGAICA, we studied the variation of the best value of fitness function with the number of times the fitness function is evaluated. In the case of example 1 with same mixing matrix A, CGAICA was performed. Fig 7 shows the variation of the reciprocal of fitness function J values with the number of J evaluations for random binary component. Similar results were obtained by taking approximation to negentropy as a contrast function.

C. Permutation Ambiguity

From all the above two examples it was observed that using CGAICA ICs were recovered always in a fixed order in all runs of the simulation experiments. The IC for which the fitness function has a global maximum value appeared first and then appeared the IC with subsequent maximum value of the fitness function. So we can predict the order of the ICs if we relatively know about the value of their statistical property like kurtosis. Hence the permutation indeterminacy present in all

ICA techniques seems to be dissolved by use of our proposed technique. This can be extended without loss of generalization to cases of ICA with any number of independent components.

VII. CONCLUSION

The Genetic Algorithm based optimization is used in a constrained manner to estimate the independent components from their observed mixtures. This scheme is tested using several examples including the speech signals for instantaneous mixing cases. From the simulation results it is very clear that in the CGAICA algorithm the MSE of the estimated ICs decreases with increase in the number of chromosomes. Also for a fixed number of chromosomes the number of generations affects the MSE. The CGAICA algorithm has almost comparable MSE as the fast ICA algorithm. The permutation ambiguity present in ICA techniques is resolved by use of the CGAICA algorithm if we have relative knowledge of the statistical characteristics of the signals to be estimated. However care should be taken while adjusting the parameters for constrained Genetic Algorithm based optimization so that premature convergence in a local optimum does not occur. For this the parameters should be initially varied over a large range to ensure the proper convergence to a global optimum.

REFERENCES

[1] P. Comon, "Independent Component Analysis-A new concept?" *Signal Processing*, vol-36, pp.287-314,1994.

[2] L.Potanutis et al., "Independent Component Analysis applied to Feature Extraction for Robust Automatic Speech Recognition," *Electronics Lett.*, vol-36, No.23, pp.1977-1978, Nov.2000.

[3] S.Z.Li, et al., "Learing Multiview Face Subspaces and Facial Pose Estimation using Independent Component Analysis," *IEEE Trans. Image Processing*, vol-14, No.6, pp.705-712, June 2005.

[4] I.Dagher,R.Nachar, "Face Recognition using IPCA-ICA Algorithm," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol-28,No.6,pp.996-1000, June 2006.

[5] C.Jutten and J.Herault, "Blind separation of sources, part I: An adaptive algorithm based on neuromimetric architecture," *Signal Processing*, 24:1-10, 1991.

[6] N.Delfosse and P.Loubaton, "Adaptive blind separation of independent sources: a deflation approach," *Signal Processing*, 45:59-83, 1995.

[7] P.Pajunen and J. Karhunen, "A maximum likelihood approach to nonlinear blind source separation," *Proc. of Int. conf. on Artificial Neural Networks (ICANN'97)*, pp.541-546, Lausanne, Switzerland, 1997.

[8] J.F.Cardoso, "Infomax and maximum likelihood for sources separation," *IEEE Letters on Signal Processing*, 4:112-114, 1997.

[9] E.Oja, H.Ogawa, and J. Wangviwattana, "Learning in nonlinear constrained Hebbian networks," In T.Kohonen et al.,editor, *Artificial Neural Networks,Proc.ICANN'91*,pp385-390, Espoo, Finland, 1991. North-Holland, Amsterdam.

[10] A.Hyverinen, "Fast and Robust Fixed-point Algorithms for Independent Component Analysis," *IEEE Trans. On Neural Networks.*, pp.626-634, vol.10, no.3, May 1999.

[11] J.F. Cardoso, "Eigen-structure of the fourth-order cumulant tensor with application to the blind source separation problem," *Proc. ICASSP'90*, pp.2655-2658, Albuquerque, NM, USA, 1990.

[12] H.Sawada, R.Mukai, S.Araki and S.Makino, "Solving Permutation and Circularity problem in Frequency-Domain Blind Source Separation," Proc. International Conf. on ICA 2004, Japan

Table 1 (Parameters of CGAICA to recover Sine wave and Random Binary wave from its mixtures)

No of Chromosomes N_c	4	8	8	10	20
No of Generations N_{re}	5	10	20	20	20
No of Bits N_b	15	15	15	15	15
Probability of Crossover P_c	0.85	0.85	0.85	0.85	0.85
Probability of Mutation P_m	0.01	0.01	0.01	0.01	0.01
MSE for Random Binary IC	3.562×10^{-4}	4.5883×10^{-5}	2.7647×10^{-5}	1.2268×10^{-6}	2.1811×10^{-8}
MSE for Sine Wave IC	8.60×10^{-2}	8.61×10^{-2}	8.60×10^{-2}	8.58×10^{-2}	8.58×10^{-2}

[13] M.Yoshioka and S. Omatu, "Signal separation method using genetic algorithms," in *Proc. IEEE Int. Joint Conf. Neural Networks*, vol.2, pp.909-912, 1998.
[14] Y.Tan and J.Wang, "Nonlinear Blind Source Separation Using Higher Order Statistics and a Genetic Algorithm," *IEEE Trans. On Evolutionary Computation*, vol.5, No.6, pp.600-611, Dec. 2001.
[15] F.Rojas, C.G.Puntonet, M.Rodriguez-Alvarez, I.Rojas, and R.Martin-Clemente, "Blind Source separation in post-nonlinear mixtures using competitive learning, simulated annealing and a genetic algorithm," *IEEE Trans. On Systems, Man and Cybernatics –Part C: Applications and Reviews*, vol.34,no.4, pp407-416, November 2004.
[16] R.Palaniappan and C.N.Gupta, "Genetic Algorithm based independent component analysis to separate noise from Electrocardiogram signals," *Proc. IEEE* , 2006
[17] Hyverinen., J.Kahrunen. and E. Oja, " *Independent Component Analysis*", John Wiley & Sons, 2001.
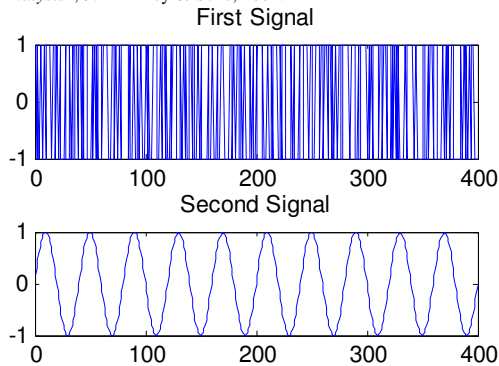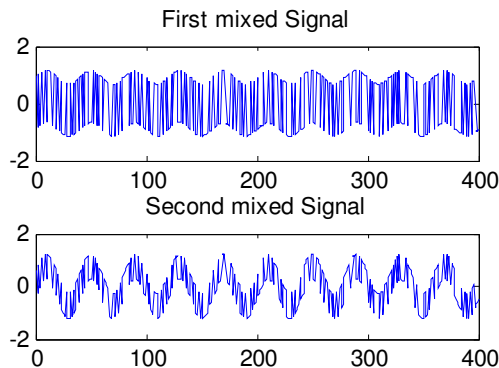
Fig.1 (Original Signals)
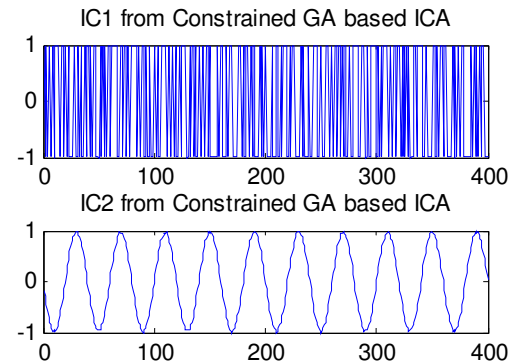

Fig.2 (Mixer of Random Binary & Sine wave)

Fig.3 (The Recovered Independent Components)

Table 2 (Comparison of fast ICA and CGAICA)

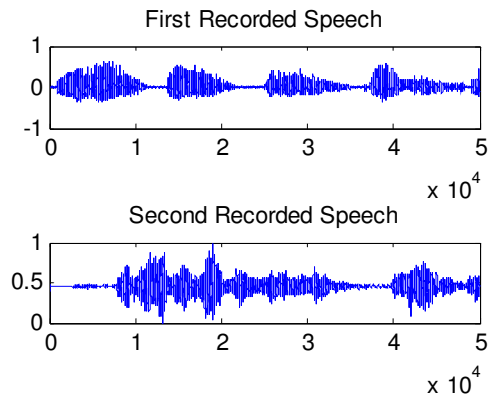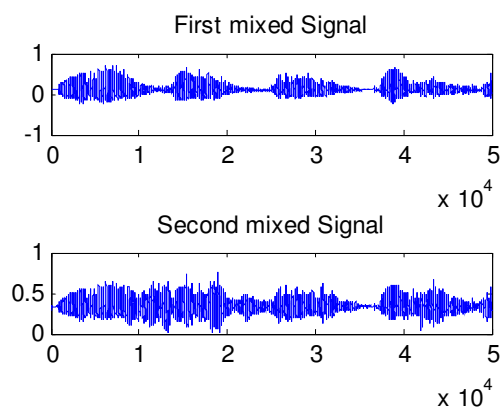| Algorithm	MSE	
	Random Binary	Sine wave
Fast ICA	2.5004×10^{-9}	8.58×10^{-2}
CGAICA	2.1811×10^{-8}	8.58×10^{-2}

Fig.4 (Recorded Speech Signals)

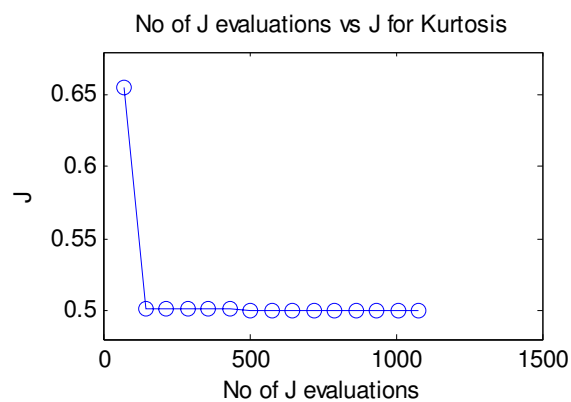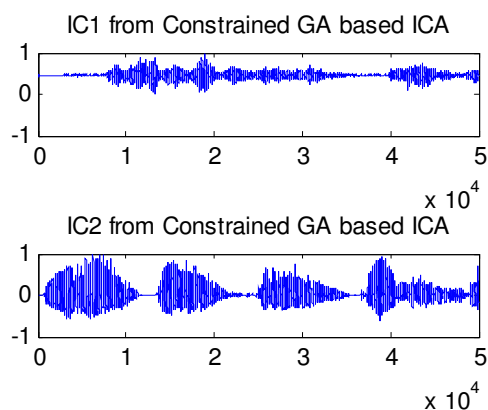Fig5. (Speech Signals mixed with artificial mixing matrix)

Fig. 7 (Convergence of CGAICA)

Fig.6 (Speech Signals recovered by CGAICA)