RESEARCH ARTICLE OPEN ACCESS

Fast Independent Component Analysis Algorithm for Blind Separation of Non-Stationary Audio Signal

Sadiqua Khan*, Y. Rama Krishna **, Dr. Panakala Rajesh Kumar***
*(Department of Electronics and Communications, PVPSIT, Vijayawada.
** (Department of Electronics and Communications, PVPSIT, Vijayawada.
***(Department of Electronics and Communications, PVPSIT, Vijayawada.

ABSTRACT
FastICA is a statistical method for transforming an observed multidimensional random vector into components that are statistically as independent from each other as possible. Acoustic signals recorded simultaneously in a reverberant environment can be described as sum of differently convolved sources. The task of source separation is to identify the multiple channels and possibly to invert those in order to obtain estimates of the underlying sources. We tackle the problem by explicitly exploiting the non-stationary components of the acoustic sources. Using maximum entropy approximations of differential entropy, we introduce a family of new contrast (objective) functions for ICA. Here we propose an algorithm for blind source separation in which frequency domain ICA and time domain ICA are used for successful separation of signals
***Keywords* -** Blind source separation, Entropy, Independent Component Analysis, Non-guassianity.

I. INTRODUCTION.

Blind source separation (BSS) has been proposed for various fields in recent years [1]. It is used to extract individual signals from observed mixed signals. It can be potentially used in communication systems, biomedical signal processing, image restoration and the classical cocktail party problem. In the communication field, it is a promising tool for the design of multi-input multi-output (MIMO) equalizers for suppression of intersymbol interference, co-channel and adjacent channel interference and multi-access interference. In biomedical signal processing, BSS can be used to process electrocardiography (ECG), electroencephalography (EEG), electromyography (EMG) and magneto encephalograph (MEG) signals. In the image signal processing field, it can be used for image restoration and understanding. The cocktail party problem is our focus, where the target is to mimic in a machine the ability of a human to separate one speaker from a mixture of sounds. We focus on audio signal processing in a room environment, which can for example be used for teleconferencing.

During the past decades, there has been considerable research performed in the field of convolutive blind source separation (CBSS). Initially, research was aimed at solutions based in the time domain. In real room recording, however, where the impulse response is on the order of thousands of samples in length, the time domain algorithm would be computationally very expensive to separate the sources. To overcome this problem, a solution in the frequency domain was proposed. As convolution in the time domain corresponds to multiplication in the frequency domain, the transformation into the frequency domain converts the convolutive mixing problem to that of independent complex instantaneous mixing operations at each frequency bin provided the block length is not too large. In realization, moreover, care is necessary to overcome circular convolution effects.

In this paper, we present the implementation of blind source separation using FastICA (independent component analysis). The aspiration of this paper is to recover two independent source signals composed of unknown linear combinations [2]. Through BSS, we have successfully separated the two signals apart with and without background noise.

II. Entropy

A central problem in BSS is cocktail party, as well as in statistics and signal processing, is finding a suitable representation or transformation of the data. For computational and conceptual simplicity, the representation is often sought as a linear transformation of the original data. Let us denote by x = (x1,x2, ..., xm)T a zero-mean m-dimensional random variable that can be observed, and by s = (s1, s2, ..., sn)T its n-dimensional transform. Then the problem is to determine a constant (weight) matrix W so that the linear transformation of the observed variables has some suitable properties.

S= Wx (1)

Several principles and methods have been developed to find such a linear representation, including principal component analysis, factor analysis, projection pursuit, independent component analysis etc. The transformation may be defined using such criteria as optimal dimension reduction, statistical 'interestingness' of the resulting components 's' simplicity of the transformation, or other criteria, including application-oriented ones.

We treat in this paper the problem of estimating the transformation given by (linear) independent component analysis (ICA). Thus this method is a special case of redundancy reduction. One popular way of formulating the ICA problem is to consider the estimation of the following generative model for the data.

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad (2)$$

where x is an observed m-dimensional vector, s is an n-dimensional (latent) random vector whose components are assumed mutually independent, and A is a constant m × n matrix to be estimated . It is usually further assumed that the dimensions of x and s are equal, i.e., m = n; we make this assumption in the rest of the paper.

A noise vector may also be present. The matrix W defining the transformation as in (1) is then obtained as the (pseudo)inverse of the estimate of the matrix A. Non-Gaussianity of the independent components is necessary for the identity ability of the model. General formulation for ICA that does not need to assume an underlying data model. This definition is based on the concept of mutual information. First, we define the differential entropy H of a random vector y = (y1,... yn)T with density f (.):

$$H(y) = - \int f(y) \log f(y)\, dy \quad (3)$$

Differential entropy can be normalized to give rise to the determination of negentropy, which has the appealing property of being invariant for linear transformations. The definition of negentropy J is given by:

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

where y gauss is a Gaussian random vector of the same covariance matrix as y. Negentropy can also be interpreted as a measure of nongaussianity. Using the concept of differential entropy, one can define the mutual information I between the n (scalar) random variables yi, i = 1...n [8, 7]. Mutual information is a natural measure of the dependence between random variables. It is particularly interesting to express mutual information using negentropy, constraining the variables to be uncorrelated. In this case, we have [7]

$$I(y_1, y_2, ..., y_n) = J(y) -\sum J(y_i). \quad (5)$$

Since mutual information is the information-theoretic measure of the independence of random variables, it is natural to use it as the criterion for finding the ICA transform. The ICA of a random vector x as an invertible transformation **s = Wx** where the matrix W is determined so that the mutual information of the transformed components si is minimized.

Two promising applications of ICA are blind source separation and feature extraction. In blind source separation, the observed values of x correspond to a realization of an m-dimensional discrete-time signal x(t), t = 1, 2, ... Then the components s(t) are called source signals, which are usually original, uncorrupted signals or noise sources. Often such sources are statistically independent from each other, and thus the signals can be recovered from linear mixtures x by n finding a transformation in which the transformed signals are as independent as possible as in ICA.

III. Functions for ICA

3.1 ICA data model, minimization of mutual information.

One popular way of formulating the ICA problem is to consider the estimation of the following generative model for the data [1, 3, 5, 6] From (2) x is an observed m-dimensional vector, s is an n-dimensional (latent) random vector whose components are assumed mutually independent, and A is a constant m × n matrix to be estimated. It is usually further assumed that the dimensions of x and s are equal, i.e., m = n; we make this assumption in the rest of the paper. A noise vector may also be present. The matrix W defining the transformation as in (1) is then obtained as the (pseudo)inverse of the estimate of the matrix A. Non-Gaussianity of the independent components is necessary for the identibility of the model (2), see [7].

Comon [7] showed how to obtain a more general formulation for ICA that does not need to assume an underlying data model. This definition is based on the concept of mutual information. First, we define the differential entropy H of a random vector y = (y1 , yn)T with density f (.) as follows:

$$H(y) = -\int f(y) \log f(y)\, dy \quad (6)$$

Differential entropy can be normalized to give rise to the definition of negentropy, which has the appealing property of being invariant for linear transformations. The definition of negentropy J is given by

$$J(y)=H(y_{gauss}) - H(y) \quad (7)$$

where ygauss is a Gaussian random vector of the same covariance matrix as y. Negentropy can also be interpreted as a measure of nongaussianity [7]. Using the concept of differential entropy, one can define the mutual information I between the n (scalar) random variables yi, i = 1...n [8, 7].

Mutual information is a natural measure of the dependence between random variables. It is particularly interesting to express mutual information using negentropy, constraining the variables to be uncorrelated. In this case, we have [7]

$$I(y_1, y_2, ..., y_n) = J(y) - \sum J(y_i). \quad (8)$$

Since mutual information is the information-theoretic measure of the independence of random variables, it is natural to use it as the criterion for finding the ICA transform. Thus we define in this paper, following [7], the ICA of a random vector x as an invertible transformation as in (1) where the matrix W is determined so that the mutual information of the transformed components si is minimized. This constraint is not strictly necessary, but simplifies the computations considerably. Because negentropy is invariant for invertible linear transformations is now obvious from (8) that finding an invertible transformation W that minimizes the mutual information is roughly equivalent to directions in which the negentropy is maximized.

3.2 Approximations of Negentropy

To use the definition of ICA given above, a simple estimate of the negentropy (or of differential entropy) is needed. We use here the new approximations developed based on the maximum entropy principle. In the simplest case, these new approximations are of the form:

$$J(y_i) \approx c[E\{G(y_i)\} - E\{G(v)\}]^2 \quad (9)$$

where G is practically any non-quadratic function, c is an irrelevant constant, and is a Gaussian variable of zero mean and unit variance (i.e., standardized). The random variable yi is assumed to be of zero mean and unit variance. For symmetric variables, this is a generalization of the cumulant-based approximation in [7], which is obtained by taking $G(y_i) = y^4_i$
i . The choice of the function.

The approximation of negentropy given above in (9) gives readily a new objective function for estimating the ICA transform in our framework. First, to find one independent component, or projection pursuit direction as $y_i = \mathbf{w}^T\mathbf{x}$, we maximize the function JG given by

$$J_G(\mathbf{w}) = [E\{G(\mathbf{w}^T\mathbf{x})\} - E\{G(\mathrm{v})\}]^2 \quad (10)$$

where w is an m-dimensional (weight) vector constrained so that $E\{(\mathbf{w}^T\mathbf{x})^2\} = 1$ (we can fix the scale arbitrarily). Several independent components can then be estimated one-by-one using a scheme, see Section 4. Second, using the approach of minimizing mutual information, the above on e-unit contrast function can be simply extended to comp - ute the whole matrix W in (1). To do this, recall from (8) that mutual information is minimized (under the constraint of decorrelation) when the sum of the negentropies of the components in maximized.

IV. FIXED POINT ALGORITHM

To begin with, we shall derive the fixed-point algorithm with sphered data. First note that the maxima of $J_G(\mathbf{w})$ are obtained at certain optima of $E\{G(\mathbf{w}^T\mathbf{x})\}$. According to the Kuhn-Tucker cond-itions[18], the optima of $E\{G(\mathbf{w}^T\mathbf{x})\}$ under the constraint $E\{(\mathbf{w}^T\mathbf{x})^2\}=\mathrm{k}\mathbf{w}\mathrm{k}^2=1$ are obtained at points where

$$E\{\mathbf{x}g(\mathbf{w}^T\mathbf{x})\} - \beta\mathbf{w} = 0 \quad (11)$$

where ß is a constant that can be easily evaluated to give $\beta = E\{\mathbf{w}^T_0\,\mathbf{x}g(\mathbf{w}^T_0\,\mathbf{x})\}$, where w0 is the value of w at the optimum. Let us try to solve this equation by Newton's method. Denoting the function on the left-hand side of (11) by F, we obtain its Jacobian matrix JF (w) as

$$JF(\mathbf{w}) = E\{\mathbf{x}\mathbf{x}^T g'(\mathbf{w}^T\mathbf{x})\} - \beta\mathbf{I} \quad (12)$$

To simplify the inversion of this matrix, we decide to approximate the first term in (12). Since the data is sphered, a reasonable approximation seems to be $E\{\mathbf{x}\mathbf{x}^T g'(\mathbf{w}^T\mathbf{x})\} \approx E\{\mathbf{x}\mathbf{x}^T\}E\{g'(\mathbf{w}^T\mathbf{x})\} = E\{g'(\mathbf{w}^T\mathbf{x})\}\mathbf{I}$. Thus the jacobian matrix becomes diagonal and can easily inverted. We also approximate ß using the current value of w instead of w0. Thus we obtain the following approximative newton iteration:

$$\mathbf{w}^+ = \mathbf{w} - [E\{\mathbf{x}g(\mathbf{w}^T\mathbf{x})\} - \beta\mathbf{w}]/[E\{g'(\mathbf{w}^T\mathbf{x})\} - \square]$$
$$\mathbf{w}^* = \mathbf{w}^+/\mathrm{k}\mathbf{w}^+\mathrm{k} \quad (13)$$

where w*denote the value of w $\beta = E\{\mathbf{w}^T\mathbf{x}g(\mathbf{w}^T\mathbf{x})\}$ and the normalization has been added to improve the stability. This algorithm can be further simplified by multiplying both sides of the equation in (16) by $\beta - E\{g'(\mathbf{w}^T\mathbf{x})\}$. This gives the following fixed point algorithm:

$$\mathbf{w}^+ = E\{\mathbf{x}g(\mathbf{w}^T\mathbf{x})\} - E\{g'(\mathbf{w}^T\mathbf{x})\}\mathbf{w} \quad (14)$$
$$\mathbf{w}^* = \mathbf{w}^+/\mathrm{k}\mathbf{w}^+\mathrm{k}$$

which was introduced in [17] using a more heuristic derivation. It is well-known that the convergence of the Newton method may be rather uncertain. To ameliorate this, one may add a step size in (16), obtaining the stabilized fixed-point algorithm

$$\mathbf{w}^+ = \mathbf{w} - \mu\,[E\{\mathbf{x}g(\mathbf{w}^T\mathbf{x})\} - \beta\mathbf{w}]/[E\{g'(\mathbf{w}^T\mathbf{x})\} - \square]$$
$$\mathbf{w}^* = \mathbf{w}^+/\mathrm{k}\mathbf{w}^+\mathrm{k} \quad (15)$$

where $\beta = E\{\mathbf{w}^T\mathbf{x}g(\mathbf{w}^T\mathbf{x})\}$ as above, and μ is a step size parameter that may change with the iteration count. Taking a μ that is much smaller than unity (say, 0.1 or 0.01), the algorithm (15) converges with much more certainty. In particular, it is often a good strategy to start with μ = 1, in which case the algorithm is equivalent to the original fixed-point algorithm in (17). If convergence seems problematic, μ may then be decreased gradually until convergence is satisfactory.

The fixed-point algorithms may also be simply used for the original, that is, not sphered data. Transforming the data back to the non-sphered variables, one sees easily that the following modification of the algorithm (14) works for non-sphered data:

$$\mathbf{w}^+ = \mathbf{C}^{-1}E\{\mathbf{x}g(\mathbf{w}^T\mathbf{x})\} - E\{g'(\mathbf{w}^T\mathbf{x})\}\mathbf{w}$$
$$\mathbf{w}^* = \mathbf{w}^+ / \ (\mathbf{w}^+)^T\,\mathbf{C}\mathbf{w}^+ \quad (16)$$

where $\mathbf{C} = E\{\mathbf{x}\mathbf{x}^T\}$ is the covariance matrix of the data. The stabilized version, algorithm (15), can also be modified as follows to work with no n-sphered data:

$$\mathbf{w}^+ = \mathbf{w} - \mu\,[\mathbf{C}^{-1}E\{\mathbf{x}g(\mathbf{w}^T\mathbf{x})\} - \beta\mathbf{w}]/[E\{g'(\mathbf{w}^T\mathbf{x})\} - \square]$$
$$\mathbf{w}^* = \mathbf{w}^+ / \ (\mathbf{w}^+)^T\,\mathbf{C}\mathbf{w}^+ \quad (17)$$

Using this algorithm, one obtains directly an independent component as the linear combination $\mathbf{w}^T\mathbf{x}$, where x need not be sphered (pre-whitened). These modifications presuppose, of course, that the covariance matrix is not singular. If it is singular or near-singular, the dimension of the data must be reduced, for example with PCA [7].

V. Experimental results

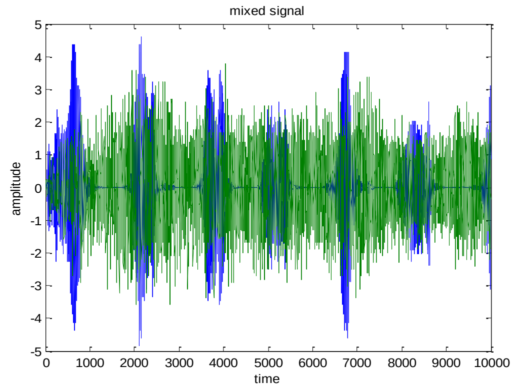

Fig.1: Mixed audio signals of a bird and horn.

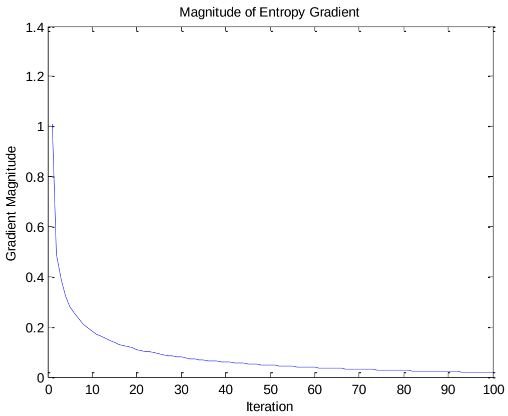

Fig.2 calculation of entropy.

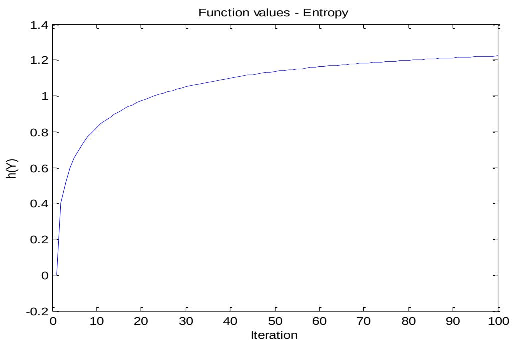

Fig.3: Calculation of gradient entropy.

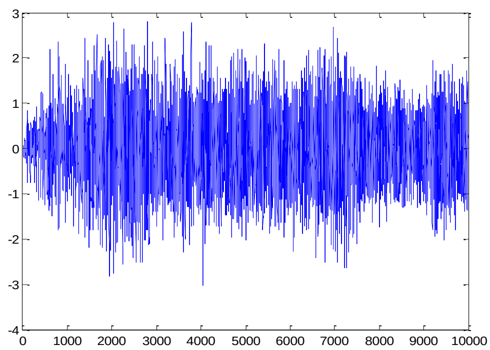

Fig.4: separated source audio of horn.

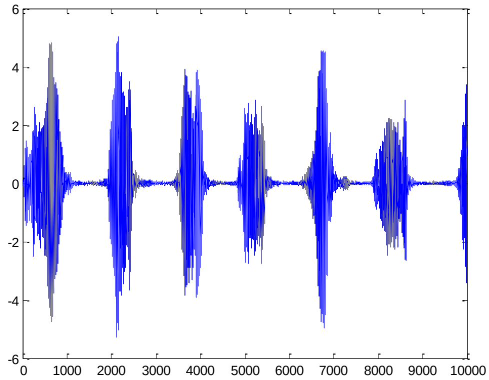

Fig.5: separated source audio of bird.

VI. Conclusion

The problem of linear independent component analysis (ICA), which is a form of redundancy reduction, was addressed. The main advantage of the fixed-point algorithms is that their convergence can be shown to be very fast (cubic or at least quadratic). Combining the good statistical properties (e.g. robustness) of the new contrast functions, and the good algorithmic properties of the fixed-point algorithm, a very appealing method for ICA was obtained. Simulations as well as applications on real-life data have validated the novel contrast functions and algorithms introduced. Some extensions of the methods introduced in this paper are present in which the problem of noisy data is addressed which deals with the situation where there are more independent components than observed variables..

References

[1] S.-I. Amari, A. Cichocki, and H.H. Yang. A new learning algorithm for blind source separation. In *Advances in Neural Information Processing Systems 8*, pages 757–763. MIT Press, 1996.

[2] H. B. Barlow. Possible principles underlying the transformations of sensory messages. In W. A.Rosen-blith, editor, *Sensory Communication*, pages 217–234. MIT Press, 1961.

[3] A.J. Bell and T.J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159, 1995.

[4] A.J. Bell and T.J. Sejnowski. The 'independent components' of natural scenes are edge filters. *Vision Research*, 37:3327–3338, 1997.

[5] J.-F. Cardoso and B. Hvam Laheld. Equivariant adaptive source separation. *IEEE Trans. on Signal Processing*, 44(12):3017–3030, 1996.

[6] A. Cichocki and R. Unbehauen. *Neural Networks for Signal Processing and*

Optimization. Wiley, 1994.

[7] P. Comon. Independent component analysis—a new concept ? *Signal Processing*, 36:287–314, 1994.

[8] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, 1991.

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

[10] The FastICA MATLAB package. Available at http://www.cis.hut.fi/projects/ica/fastica/.

[11] J. H. Friedman and J. W. Tukey. A projection pursuit algorithm for exploratory data analysis. *IEEE Trans. of Computers*, c-23(9):881–890, 1974. J.H. Friedman. Exploratory projection pursuit. J. of the American Statistical Association, 82(397):249– 266, 1987.

[12] J. H. Friedman and J. W. Tukey. A projection pursuit algorithm for exploratory data analysis. *IEEE Trans. of Computers*, c-23(9):881–890, 1974.

[13] X. Giannakopoulos, J. Karhunen, and E. Oja. Experimental comparison of neural ICA algorithms. In *Proc. Int. Conf. on Artificial Neural Networks (ICANN'98)* , pages 651–656, Skövde, Sweden, 1998.

[14] F.R. Hampel, E.M. Ronchetti, P.J. Rousseuw, and W.A. Stahel. *Robust Statistics*. Wiley, 1986.

[15] H. H. Harman. *Modern Factor Analysis*. University of Chicago Press, 2nd edition, 1967.

[16] P.J. Huber. Projection pursuit. *The Annals of Statistics*, 13(2):435–475, 1985.

[17] A. Hyvärinen. A family of fixed-point algorithms for independent component analysis. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'97)*, pages 3917–3920, Munich, Germany, 1997.

[18] D. Luenberger. *Optimization by Vector Space Methods*. Wiley, 1969.