INDEPENDENT COMPONENT ANALYSIS BASED BLIND SIGNAL SEPARATION FOR MIXED SPEECH SIGNAL

A THESIS SUBMITTED IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF

BACHELOR OF TECHNOLOGY

IN

ELECTRICAL ENGINEERING.

By

G.Karthikeyan
And
Radharaman sahoo

Under the guidance of

Dr. Dipti Patra

Department of Electrical Engineering
National Institute of Technology, Rourkela

National Institute of Technology
Rourkela

CERTIFICATE

This is to certify that the thesis entitled **"Independent Component analysis based Blind Signal Separation For Mixed Speech Signal"** submitted by Sri G.Karthikeyan and Sri Radharaman Sahoo in partial fulfillment of the requirements for the award of Bachelor of Technoloy Degree in Electrical Engineering at National Institute of Technology, Rourkela (Deemed University) is an authentic work carried out by him under my supervision and guidance.

To the best of my knowledge, the matter embodied in the thesis has not been any other University/ Institute for the award of any degree or diploma.

Date: 02/05/07

Dr. Dipti Patra
Asst Professor
Department of Electrical Engg.
National Institute of Technology
Rourkela - 769008

ACKNOWLEDGEMENT

No thesis is created entirely by an individual, many people have helped to create this thesis and each of their contribution has been valuable. My deepest gratitude goes to my thesis supervisor, Dr.Dipti Patra, Professor, Department of Electrical Engineering, for his guidance, support, motivation and encouragement through out the period this work was carried out. His readiness for consultation at all times, his educative comments, his concern and assistance even with practical things have been invaluable. I would also like to thank all professors and lecturers, and members of the department of Electrical Engineering for their generous help in various ways for the completion of this thesis. A vote of thanks to my fellow students for their friendly co-operation.

G.Karthikeyan
Roll. No. 10302066
Dept. Of Electrical Engineering
NIT Rourkela

Radharaman Sahoo
Roll.No. 10302025
Dept. Of Electrical Engineering
NIT Rourkela

LIST OF FIGURES

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

Independent component analysis (ICA) is a new technique to statistically extract independent components from the observed multidimensional mixture of data. Many successful examples of ICA application in the filed of signal processing are reported recently. Independent component analysis (ICA) was originally developed to deal with problems that are closely related to cocktail- party problems.ICA is a powerful and useful statistical tool for extracting independent source given only observed data that are mixtures of the unknown sources. ICA has been studied by many researchers in neural networks and statistical approaches during the past 10 years. Independent component analysis is a signal processing technique whose goal is to express a set of random variables as linear combinations of statistically independent component variables.

ICA can reveal interesting information on sensor signals by giving access to its independent components. Independent component analysis is a new class of analysis method developed in recent years to solve these problems. By ICA we can separate the original sources blindly only by their mixtures. The application of ICA in extracting the characteristic signals is based on the difference of higher-order statistical characteristics.

This is represented compactly by the mixing equation x(t)=As(t) where s(t) is a column vector collecting the source signals, vector x(t) similarly collects the n observed signals, and the square n×n "mixing matrix" contains the mixture coefficients. The ICA problem exists in recovering the source vector using only the observed data the assumption of independence between the entries of the input vector, and possibly some prior information about the probability distribution of the inputs. It can be formulated as the computation of an n×n separating matrix B whose output y(t) =Bx (t) is an estimate of the vector s(t) of the source signal.

Figure 1.1: Mixing and Separating

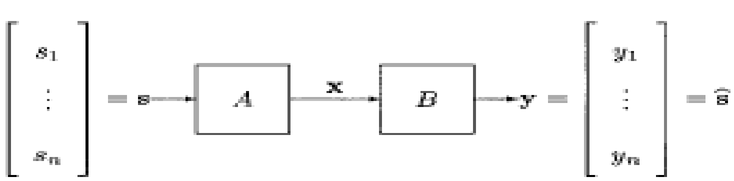


Fig. 1. Mixing and separating. Unobserved signals *s*; observations *x*, estimated source signals *y*.

Chapter 2

LITERATURE ON INDEPENDENT COMPONENT ANALYSIS

Introduction to ICA

Nongaussian is Independent

Statistical Independence

Measure of Nongaussianity

2.1 Introduction to ICA:

This section presents the generic mixture model used in blind source separation and independent component analysis.

Consider the following linear model,

$$\mathbf{X}_{orig} = \mathbf{AS} + \nu, \qquad (4.1)$$

where

$$\mathbf{X}_{orig} = \begin{bmatrix} \mathbf{x}_{orig,1} \\ \mathbf{x}_{orig,2} \\ \vdots \\ \mathbf{x}_{orig,K} \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \\ \vdots \\ \mathbf{s}_N \end{bmatrix}, \quad \nu = \begin{bmatrix} \nu_1 \\ \nu_2 \\ \vdots \\ \nu_K \end{bmatrix}.$$

This model consists of N sources of T samples, i.e., $S_i = [s_i(1) \dots s_i(t) \dots s_i(T)]$.The symbol t represents time. The observations Xorig consists of K mixtures of the sources, where, Xorig, i = $[x_i(1) \dots x_i(t) \dots x_i(T)]$. Usually it is assumed that there are at least as many observations as sources i.e., K >= N. The sources and the observations are related by a K × N matrix $A = [a_1\ a_2 \dots a_n]$ consisting of the vectors $a_i = [a_{1ii}\ a_{2i} \dots a_{ki}]^T$. This linear mapping is called the mixing matrix. The model assumes some noise V considered to be Gaussian. Solution to the linear source separation problem is not possible, if there is no information on some of the variables A or S, in addition to the observed (known) data X_{orig}. If the mixing A is known and the noise is negligible, the sources can be estimated by finding the (unmixing) matrix B,the (pseudo) inverse of the mixing matrix A, for which BX_{orig} = BAS = S .If there are as many observations as sources, then A is square and has full-rank, hence B = A−1. The full-rank assumption is the necessary and sufficient condition for the existence of the pseudo-inverse of A. When there are More observations than sources, there exist several matrices B that satisfy the Condition BA = I. In this case, the choice of B depends on the components of s that we are interested in.

For cases where there are fewer observations than Sources, a solution does not exist unless further assumptions are made. Now the Rank of A is less than the number of sources. There are some redundancies in the mixing matrix, and hence further information is required. On the other hand, if no non-trivial prior information about the mixing matrix A is known or assumed, this problem of estimating the matrices A and S is referred to as blind source separation (BSS).

The model with Negligible noise is then separable under the following fundamental restrictions.

(R1) the components of S are statistically independent;

(R2) at most one component of S is Gaussian distributed;

(R3) the mixing matrix A is of full rank.

One very popular technique for solving the BSS problem is independent component analysis (ICA). In BSS, the main focus is to determine the underlying independent sources. The best known applications of ICA are in

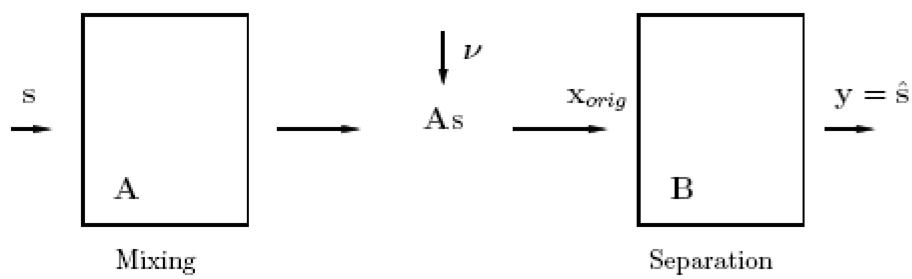


FIGURE 2.1 SCHEMATIC ILLUSTRATION OF MIXING AND SEPARATION. A IS CALLED THE MIXING MATRIX AND ITS ESTIMATED INVERSE B IS THE UNMIXING MATRIX.

The field of signal and image processing e.g., biomedical engineering speech processing multispectral image processing etc. Hence, ICA can be defined as the computation method for separating a multivariate signal into its subcomponents assuming that all of these subcomponents are mutually independent. Alternatively, ICA can also be defined as a linear transformation on a multivariate signal X_{orig}: $S = BX_{orig}$, so that the components are as independent as possible, in the sense of maximizing some function $F(s_1, . . . , s_n)$, that measures independence. This section first reviews the concepts of independence and various measures to quantify independence. Then some preprocessing techniques usually used in ICA are reviewed briefly in subsection. Subsection ends this section with a brief discussion of the ambiguities of ICA.

2.2 Nongaussian is Independent:

Intuitively speaking, the key to estimating the ICA model is nongaussianity. Actually, without nongaussianity the estimation is not possible at all. This is at the same time probably the main reason for the rather late resurgence of ICA research: In most of classical statistical theory, random variables are assumed to have Gaussian distributions, thus precluding any methods related to ICA.

The Central Limit Theorem, a classical result in probability theory, tells that the distribution of a sum of independent random variables tends toward a Gaussian distribution, under certain conditions. Thus, a sum of two independent random variables usually has a distribution that is closer to Gaussian than any of the two original random variables.

Let us now assume that the data vector **X** is distributed according to the ICA data model, i.e. it is a mixture of independent components. For simplicity, let us assume in this section that all the independent components have identical distributions. To estimate one of the independent components, we consider a linear combination of the x_i; let us denote this by$\mathbf{Y=W^TX=\sum_i W_i X_i}$, where Wis a vector to be determined. If W were one of the rows of the inverse of A, this linear combination would actually equal one of the independent components. The question is now: How could we use the Central Limit Theorem to determine Wso that it would equal one of the rows of the inverse of A? In practice, we cannot determine such a Wexactly, because we have no knowledge of matrix A, but we can find an estimator that gives a good approximation.

To see how this leads to the basic principle of ICA estimation, let us make a change of variables, defining$Z=A^TX$. Then we have$\mathbf{Y=W^TX=W^TAS=Z^TS}$. Y is thus a linear combination of s_i, with weights given by z_i. Since a sum of even two independent random variables is more Gaussian than the original variables, $\mathbf{Z^TS}$ is more Gaussian than any of the s_i and becomes least Gaussian when it in fact equals one of the s_i. In this case, obviously only one of the elements z_i of Z is nonzero. (Note that the s_i were here assumed to have identical distributions.)

Therefore, we could take as Wa vector that *maximizes the nongaussianity* of $\mathbf{W^TX}$. Such a vector would necessarily correspond (in the transformed coordinate system) to a $\mathbf{z}$ which has only one nonzero component. This means that $\mathbf{W^TX=Z^TS}$ equals one of the independent components!

Maximizing the nongaussianity of $\mathbf{W^TX}$ thus gives us one of the independent components. In fact, the optimization landscape for nongaussianity in the n-dimensional space of vectors $\mathbf{W^TX}$ has 2 N local maxima, two for each independent component, corresponding to s_i and $-s_i$ (recall that the independent components can be estimated only up to a multiplicative sign). To find several independent components, we need to find all these local maxima. This is not difficult, because the different independent components are uncorrelated: We can always constrain the search to the space that gives estimates uncorrelated with the previous ones. This corresponds to orthogonalization in a suitably transformed (i.e. whitened) space.

2.3 Statistical Independence:

Two random variables s1 and s2 are said to be independent if their joint probability density is a product of their respective marginal densities. Intuitively, this appears to be correct: having observed one random variable, all the statistics of the other, independent random variables remain unchanged. Formally, a random vector $s = [s_1, \cdots, s_N]^T$, with a multivariate density p(s) has statistically. Independent components if the density can be factorized as

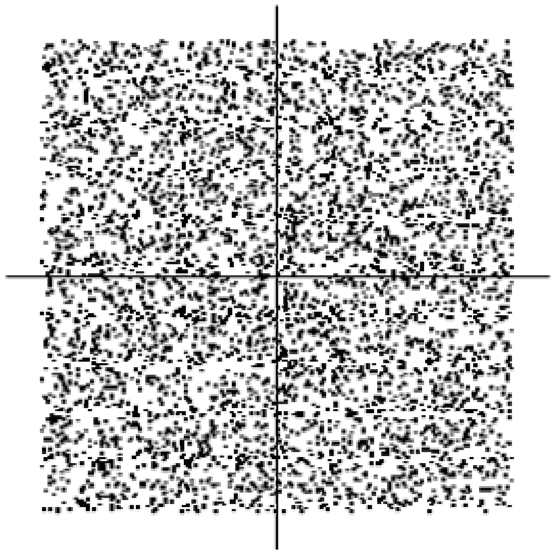
$$p(\mathbf{s}) = \prod_{i=1}^{N} p_i(s_i).$$

It is assumed that a density function for each random variable exists. This
definition leads to the notion of independence based on conditional densities.
Hence, for two random variables s1 and s2,

p (s1|s2) = p(s1, s2)/p(s2) = p(s1)p(s2)/p(s2) = p(s1).

In other words, the density of s1 is unaffected by observing s2, when the two
variables are independent.

Figure 2.2: The joint distribution of the independent components s_1 and s_2 with uniform distributions. Horizontal axis: s_1, vertical axis: s_2.



2.3 Measure of Nongaussianity:

Cumulants provide a practical way to describe distributions using simple Scalar functions. Consider a zero mean random variable s. The characteristic function of s is defined as ˆh(t) = E{exp(its)}. Expanding the logarithm of the characteristic function as a Taylor series gives,

$$\log \hat{h}(t) = \kappa_1(it) + \frac{\kappa_2(it)^2}{2} + \dots + \frac{\kappa_r(it)^r}{r!} + \dots,$$

where the Kr are some constants. These constants are called the cumulants (of the distribution) of s. In particular, the first few (four) cumulants have simple expressions,

$$k_1=E\{S\}$$

$$k_2=E\{S^2\}$$

$$k_3=E\{S^3\}$$

Classically this is achieved using Kurtosis, defined by: $kurt(Y) = E(Y^4) - 3$ for Y such that $E(Y) = 0$ and $E(Y^2) = 1$. This statistic is zero if Y ~N(0,1) and is non-zero for most (but not all) non-normal variables.

- If Kurt>0, Y is said to be leptokurtic, and is typically 'spiky', with long tails.
- If Kurt<0, Y is said to be platykurtic, and is typically flat.

However |kurt| or kurt^2 are not good statistics to sample as they are very sensitive to outliers among the data, and so are not used for ICA.

In fact the measure that is used is known as Negentropy. Entropy is defined by: $H(Y) = -\int f_Y(y)\log(f_Y(y))dy$.

A result of Information Theory is that of all random variables of equal variance the normal one has the largest entropy. Thus normals are the 'least structured' of all random variables, and entropy is generally small for spiky or clearly clustered variables.

Then Negentropy is defined by: $J(Y) = H(Y_{gauss}) - H(Y)$,where Y_{gauss} is a multivariate normal with the same covariance matrix as Y. Hence we have that $J(Y) \geq 0$ and $J(Y) = 0 \Leftrightarrow Y$ is normal.

J may be approximated by: $J(Y) \approx k[E(G(Y)) - E(G(\nu))]^2$, for some $k \geq 0, \nu$ ~N(0,1) and when Y is standardized.G here is a non-quadratic function and is picked to avoid the problems of kurtosis with outliers. Examples include $G(u) = \tanh(u)$ and $G(u) = e^{-u^2/2}$.

Before getting stuck into the algorithm the data should be preprocessed. Firstly centre each variable, so $E(X) = 0$ and 'whiten' so $E(XX^t) = I$. To achieve this take $X' = PD^{-1/2}P^tX$ where $E(XX^t) = PDP^t$ and $P^tP = I$ and D is diagonal.As a result, after this transformation, the A we seek is orthogonal. $I = E(XX^t) = E(ASS^tA^t) = AA^t$, by assumption on S.

Chapter 3

PREPROCESSING FOR ICA

Central Limit Theorem

Whitening

Gaussian Variables are Forbidden

Ambiguities of ICA

PREPROCESSING FOR ICA :

In the preceding section, we discussed the statistical principles underlying ICA methods. Practical algorithms based on these principles will be discussed in the next section. However, before applying an ICA algorithm on the data, it is usually very useful to do some preprocessing. In this section, we discuss some preprocessing techniques that make the problem of ICA estimation simpler and better conditioned

3.1 Central Limit Theorem:

The most basic and necessary preprocessing is to center X, i.e. subtract its mean vector m=E{X},so as to make X a zero-mean variable. This implies that S is zero-mean as well, as can be seen by taking expectations on both sides. This preprocessing is made solely to simplify the ICA algorithms: It does not mean that the mean could not be estimated. After estimating the mixing matrix A with centered data, we can complete the estimation by adding the mean vector of s back to the centered estimates ofS. The mean vector of S is given by $A^{-1}m$, where m is the mean that was subtracted in the preprocessing.

3.2 Whitening:

Another useful preprocessing strategy in ICA is to first whiten the observed variables. This means that before the application of the ICA algorithm (and after centering), we transform the observed vector X linearly so that we obtain a new vector X'which is white, i.e. its components are uncorrelated and their variances equal unity. In other words, the covariance matrix of X'equals the identity matrix:

$$\mathbf{E\{X'X'^T\}=I}$$

The whitening transformation is always possible. One popular method for whitening is to use the eigen-value decomposition (EVD) of the covariance matrix $E\{XX^T\}=EDE^T$, where E is the orthogonal matrix of eigenvectors of $E\{XX^T\}$ and D is the diagonal matrix of its eigenvalues $D=diag(d_1,d_2...d_n)$. Note that $E\{XX^T\}$ can be estimated in a standard way from the available sample. x(1),x(2),…x(n),whitening can now be done by

$$\mathbf{X'=E\,D^{-1/2}\,E^{T}X} \quad (31)$$

where the matrix $\mathbf{D^{-1/2}}$ is computed by a simple component-wise operation as $D^{-1/2}=diag(d_1^{-1/2},d_2^{-1/2}...d_n^{-1/2})$. It is easy to check that now. $E\{X'X'^T\}$

Whitening transforms the mixing matrix into a new one,A'.

$$\tilde{\mathbf{x}} = \mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T\mathbf{A}\mathbf{s} = \tilde{\mathbf{A}}\mathbf{s} \quad (32)$$

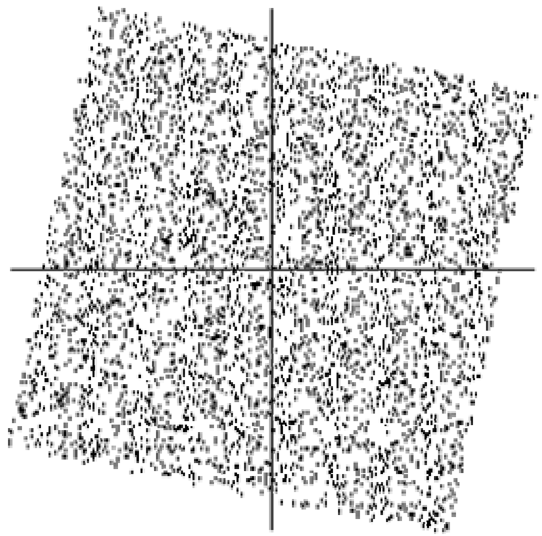
The utility of whitening resides in the fact that the new mixing matrix $\tilde{\mathbf{A}}$ is orthogonal. This can be seen from

$$E\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T\} = \tilde{\mathbf{A}}E\{\mathbf{s}\mathbf{s}^T\}\tilde{\mathbf{A}}^T = \tilde{\mathbf{A}}\tilde{\mathbf{A}}^T = \mathbf{I}. \quad (33)$$

Here we see that whitening reduces the number of parameters to be estimated. Instead of having to estimate the n2 parameters that are the elements of the original matrix A, we only need to estimate the new, orthogonal mixing matrix A. An orthogonal matrix contains n(n-1)/2degrees of freedom. For example, in two dimensions, an orthogonal transformation is determined by a single angle parameter. In larger dimensions, an orthogonal matrix contains only about half of the number of parameters of an arbitrary matrix. Thus one can say that whitening solves half of the problem of ICA. Because whitening is a very simple and standard procedure, much simpler than any ICA algorithms, it is a good idea to reduce the complexity of the problem this way. It may also be quite useful to reduce the dimension of the data at the same time as we do the whitening. Then we look at the eigenvalues dj of $E\{XX^T\}$ and discard those that are too small, as is often done in the statistical technique of principal component analysis. This has often the effect of reducing noise. Moreover, dimension reduction prevents over learning, which can sometimes be observed in ICA. A graphical illustration of the effect of whitening can be seen in Figure, in which the data in previous Figure has been whitened. The square defining the distribution is now clearly a rotated version of the original square. All that is left is the estimation of a single angle that gives the rotation.

Figure3.1 The joint distribution of the

whitened mixtures.



In the rest of this tutorial, we assume that the data has been preprocessed by centering and whitening. For simplicity of notation, we denote the preprocessed data just byX, and the transformed mixing matrix byA, omitting the tildes.

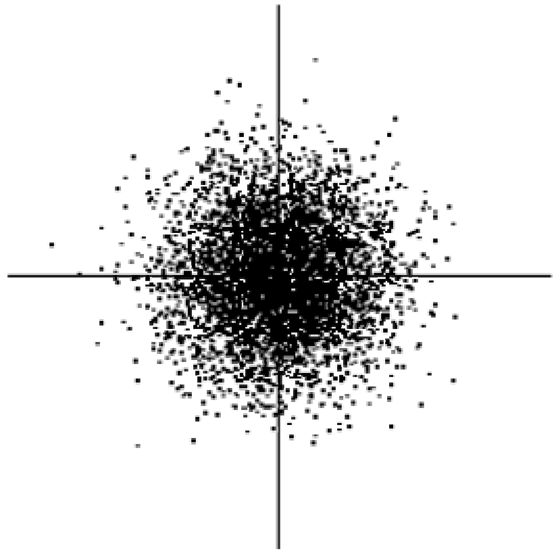
3.3 Gaussian Variables are Forbidden:

The fundamental restriction in ICA is that the independent components must be nongaussian for ICA to be possible. To see why Gaussian variables make ICA impossible, assume that the mixing matrix is orthogonal and the si are Gaussian. Then x1 and x2 are Gaussian, uncorrelated, and of unit variance. Their joint density is given by

$$p(x_1, x_2) = \frac{1}{2\pi} \exp\left(-\frac{x_1^2 + x_2^2}{2}\right)$$

This distribution is illustrated in Fig. The Figure shows that the density is completely symmetric. Therefore, it does not contain any information on the directions of the columns of the mixing matrix A. This is why A cannot be estimated.

Figure3.2: The multivariate distribution of two independent Gaussian variables.



3.4 Ambiguities of ICA:

In the ICA model, it is easy to see that the following ambiguities will hold:

1. We cannot determine the variances (energies) of the independent components. The reason is that, both Sand Abeing unknown, any scalar multiplier in one of the sources s_i could always be cancelled by dividing the corresponding column a_i of A by the same scalar. As a consequence, we may quite as well fix the magnitudes of the independent components; as they are random variables, the most natural way to do this is to assume that each has unit variance$E\{S_i^2\}=1$. Then the matrix A will be adapted in the ICA solution methods to take into account this restriction.

2. We cannot determine the order of the independent components.The reason is that, again both S and Abeing unknown, we can freely change the order of the terms in the sum in and call any of the independent components the first one.

Chapter 4

ICA ALGORITHM AND ITS APPLICATIONS

Kurtosis Based Fixed Point Algorithm

Drawbacks of Kurtosis

Negentropy Based Fixed Point Algorithm

Applications

4.1 Kurtosis Based Fixed Point Algorithm:

For finding maximum nongaussianity of wT x.

1. Choose an initial weight vector W

2. Then $\mathbf{W=E\{Z(W^TZ)^3\}-3}$

3. Let $\mathbf{W=W^+/\|W\|}$

4. If not converged go back to repeat from 2.

4.2 Drawbacks of Kurtosis:

- Its value may be depend on only few observations in tails of the observation.
- Kurtosis is not a robust measure of nongaussianity

4.3 Negentropy Based Fixed Point Algorithm:

To begin with, we shall show the one-unit version of FastICA. By a "unit" we refer to a computational unit, eventually an artificial neuron, having a weight vector W that the neuron is able to update by a learning rule. The FastICA learning rule finds a direction, i.e. a unit vector Wsuch that the projection $\mathbf{W^TX}$ maximizes nongaussianity. Nongaussianity is here measured by the approximation of negentropy $\mathbf{J(W^TX)}$. Recall that the variance of $\mathbf{W^TX}$ must here be constrained to unity; for whitened data this is equivalent to constraining the norm of W to be unity.

The FastICA is based on a fixed-point iteration scheme for finding a maximum of the nongaussianity of $\mathbf{W^TX}$. It can be also derived as an approximative Newton iteration. Denote by g the derivative of the nonquadratic function G for example the derivatives of the functions are:

$$\mathbf{G(u)=Tanh(au)}$$

where $1<a<2$is some suitable constant, often taken as $a_1=1$. The basic form of the FastICA algorithm is as follows:

1. Choose an initial (e.g. random) weight vector **W**.

2. Let $\mathbf{W^{+}=E\{Wg(W^{T}X)- E\{Wg'(W^{T}X)\}W}$

3. Let $\mathbf{W=W^{+}/\|W\|}$

4. If not converged, go back to 2.

Note that convergence means that the old and new values of W point in the same direction, i.e. their dot-product is (almost) equal to 1. It is not necessary that the vector converges to a single point, since W and –W define the same direction. This is again because the independent components can be defined only up to a multiplicative sign. Note also that it is here assumed that the data is prewhitened.

4.4 Applications:

1. **Power system**:

   We can suppress the effect of harmonics in the power system, by separating the harmonics from the sinusoidal current.

2. **Telecommunication**:

   We can suppress the interference in the spread spectrum communication. It's also used for Array processing, i.e. in Blind beamforming applications

3. **Speech processing**:

   In cock-tail party problems, from the mixture of speech signals we can separate the speech signal of the individuals.

4. **Finance / Econometrics:**

   It is a tempting alternative to try ICA on financial data. There are many situations in that application domain in which parallel time series are available, such as currency exchange rates or daily returns of stocks, that may have some common underlying factors

5 **Bio-Medical science**:

The EEG (electroencephalogram) data consists of mixture of different components of brain activity. ICA can reveal interesting information on brain activity by giving access to its independent components. It can be also used for MEG data.

6. **Digital image processing**:

Like other applications, from the mixture of digital images the individual component can be separated. ICA also finds application in feature extraction techniques

Chapter 5

SIMULATION AND RESULTS

Software Implementation

Hardware Implementation

Sample Waveforms

ComparisonsofICAAlgorithms

5.1 SOFTWARE IMPLEMENTATION:

It is done in MATLAB 7.0, the MATLAB code for separating the speech signal is given below.

```
clc;
close all;
clear all;
max_iteration=10;
epsilon=0.00001;
n=3;
T=32768;
A=wavread('C:\Documents and Settings\user\Desktop\speech\latest code wid corrected error\s1.wav');
B=wavread('C:\Documents and Settings\user\Desktop\speech\latest code wid corrected error\s2.wav',length(A));
figure;
subplot(2,1,1); plot(A);
title('Speech1 wave'),xlabel('Time(secs)'),ylabel('Amplitude')
subplot(2,1,2); plot(B, 'r');
title('Speech2 wave'),xlabel('Time(secs)'),ylabel('Amplitude')
wavplay(A,11000);
wavplay(B,11000);
M1 = A' - 2*B'+3*C';
M2 = 1.73*A'+3.41*B'-2.15*C';
figure;
subplot(2,1,1); plot(M1);
title('Mixture1'),xlabel('Time(secs)'),ylabel('Amplitude')
subplot(2,1,2); plot(M2, 'r');
title('Mixture2'),xlabel('Time(secs)'),ylabel('Amplitude')

wavplay(M1,11000);
```

```
wavplay(M2,11000);
x = [M1;M2];
h=zeros(size(x));
h=[A';B'];
[E,c]=eig(cov(x',1))
sq=inv(sqrtm(c));
mx=mean(x');
xx=x-mx'*ones(1,T);
xx=sq*E'*xx;
cov(xx')
figure; plot(xx(1,:), xx(2,:), '.'); grid on; axis square;
title('joint dist. of comps after whitening');
W=algo(epsilon,sq,E,max_iteration,xx,T);
output=W*x;
h=h-output;
figure; subplot(2,1,1);plot(h(1,:));
title('error for signal 1'),xlabel('Time(secs)'),ylabel('error value')
subplot(2,1,2);plot(h(2,:));
title('error for signal 2'),xlabel('Time(secs)'),ylabel('error value')
figure;
subplot(2,1,1),plot(output(1,:)/max(abs(output(1,:))))
title('Ind. comp. 1'),xlabel('Time(secs)'),ylabel('Amplitude');
subplot(2,1,2),plot(output(2,:)/max(abs(output(2,:))),'r')
title('Ind. comp. 2'),xlabel('Time(secs)'),ylabel('Amplitude');
wavplay(output(1,:)/max(abs(output(1,:))),11000);
wavplay(output(2,:)/max(abs(output(2,:))),11000);
```

5.2 HARDWARE IMPLEMENTATION:

The hardware implementation can be done in DSP-kit

5.3 SAMPLE WAVEFORMS:

Figure 5.1: *Original speech signals*

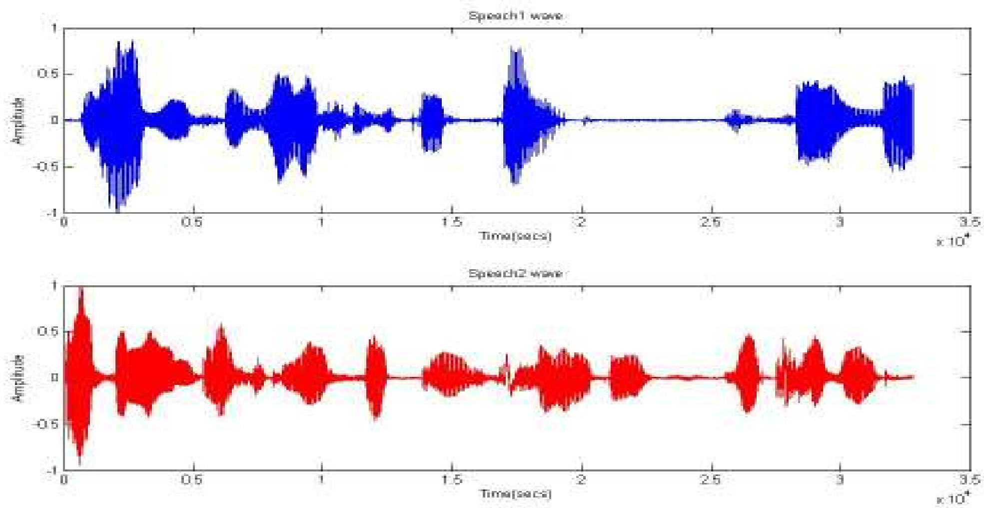


Figure 5.2: *Recovered by fixed point algorithm*

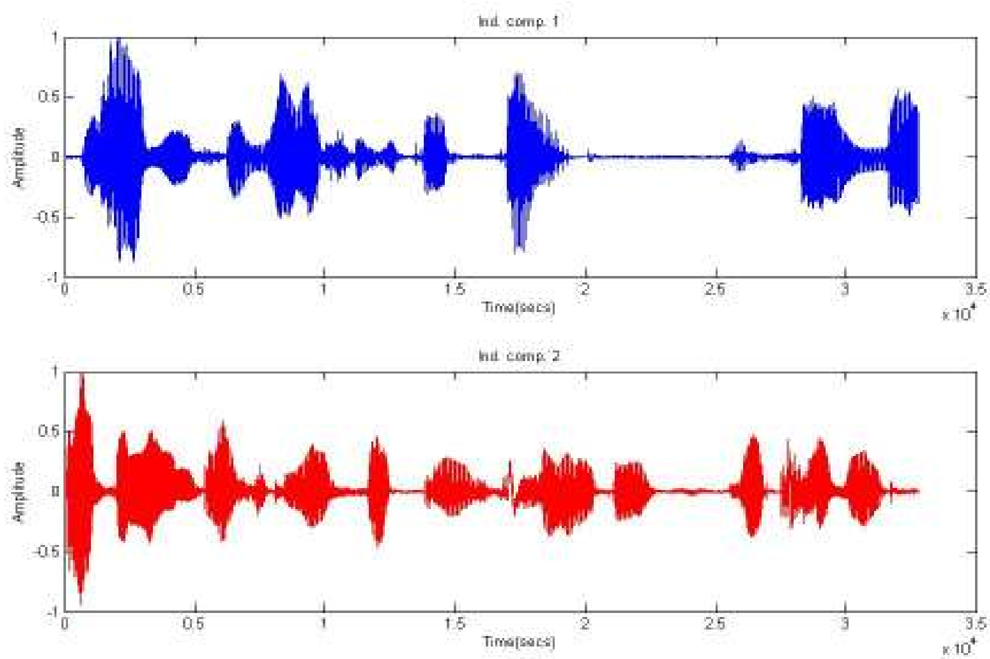


Figure 5.3: *Original speech signals*

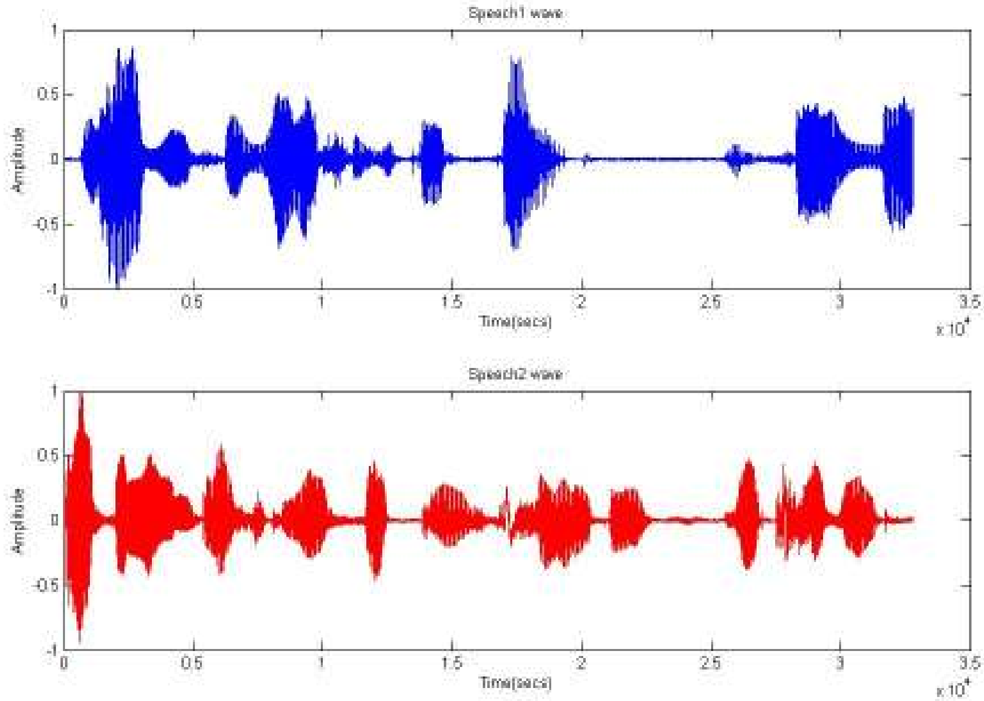
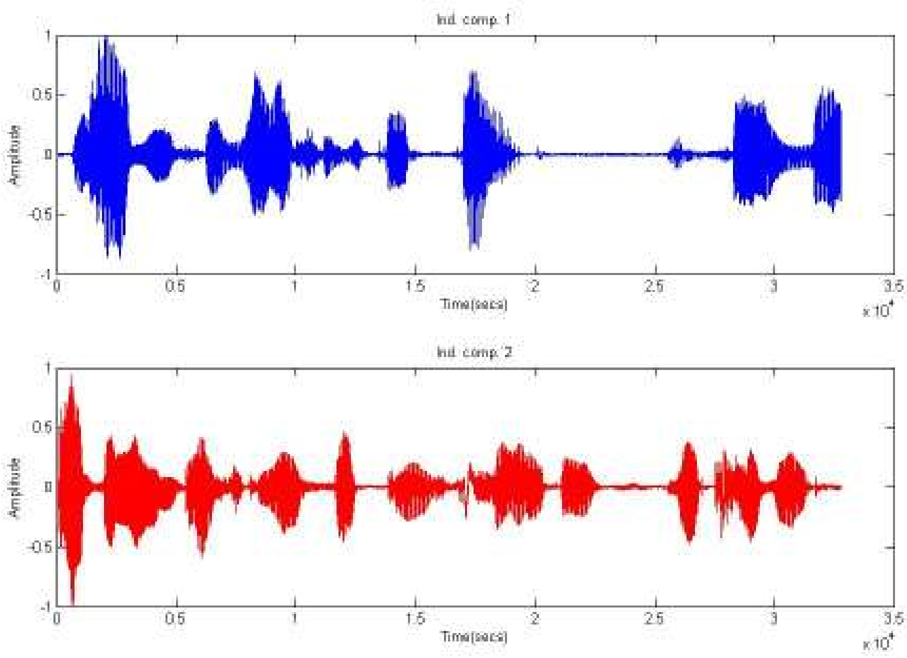


Figure 5.4: *Recovered by fixed point algorithm with tanh optimization and orthogonal symmetry*



Error Plots

Figure 5.5: *from first method (fixed point algorithm)*

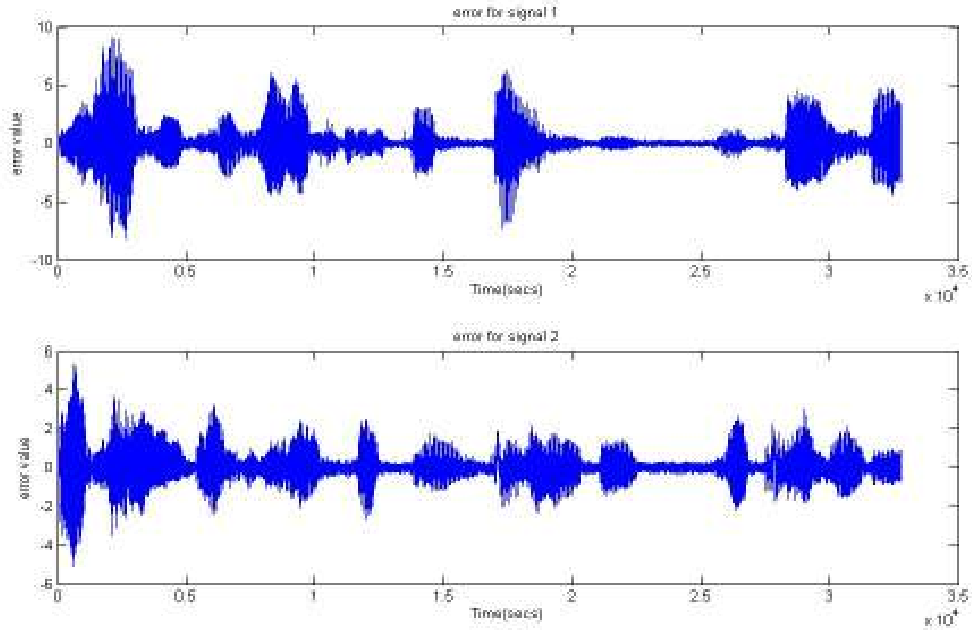
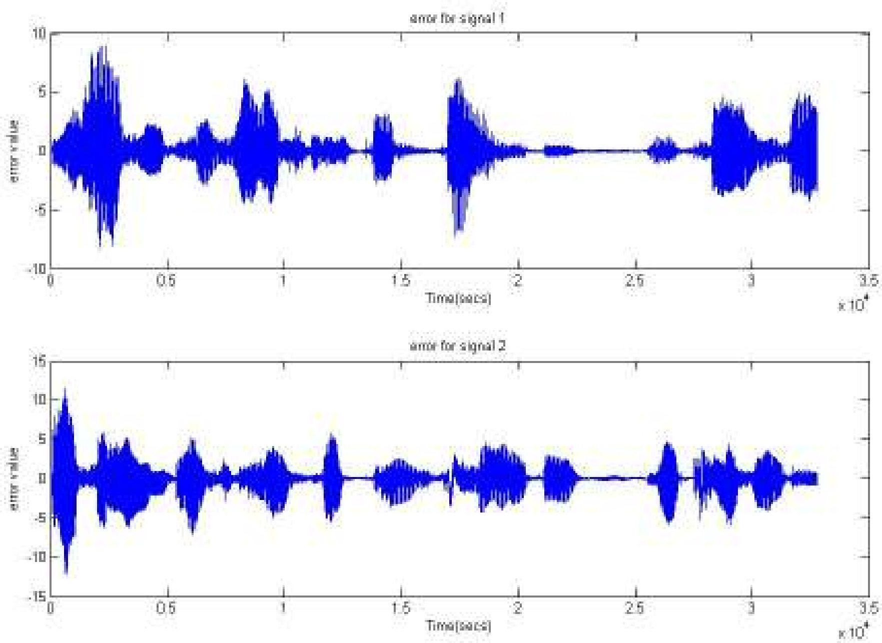


Figure 5.6: *from second method (tanh optimization)*



5.4 COMPARISON OF ICA ALGORITHMS:

The basic experiment measures the computational load and statistical performance of the tested algorithms. As for statistical performance the best results are obtained by using tanh nonlinearity. the statistical performance is based on nonlinearity and not on optimization method. All algorithms using tanh have pretty much the same statistical performance. The kurtosis based FastICA is clearly inferior to this. Looking at the computational load, one sees clearly that FastICA requires the smallest amount of computation. These ordinary gradient type algorithms have a computational load that is about 20-50 times larger than for Fast ICA.

Chapter 6

CONCLUSION

CONCLUSION:

- ICA is a very general-purpose statistical technique in which observed random data are linearly transformed into components that are maximally independent from each other, and simultaneously have “interesting” distributions. ICA can be formulated as the estimation of a latent variable model. The intuitive notion of maximum nongaussianity can be used to derive different objective functions whose optimization enables the estimation of the ICA model. Alternatively, one may use more classical notions like maximum likelihood estimation or minimization of mutual information to estimate ICA; somewhat surprisingly, these approaches are (approximately) equivalent. Although there are many approaches to solve ICA, the separation of complex signals like CDMA remains a challenge.
- This tanh with symmetric orthogonalization based fixed point algorithm is the future prospect of ICA.
- And this algorithm can also be used in complex applications like harmonic separation in power system and interference suppression in CDMA systems.

REFERENCES:

- Independent Component Analysis
  - By Apo Hyvarien and Erkki Oja
- Hyvärinen,A. (1999a). Fast and robust fixed-point algorithms for independent component analysis. IEEE Transactions on Neural Networks, 10(3):626–634
- Cardoso, J.-F. (1997). Infomax and maximum likelihood for source separation. IEEE Letters on Signal Processing,4:112–114..
- Jean-Francois Cardoso, "Blind Signal Processing: Statistical Principles" vol-86 page 2009-2023, 1998.
- M. Gaeta and j-l Lacoume "Source separation without Prior knowledge: The maximum likelihood solution" in Proc. EUSIPCO 1990, pp 621-624.
- Hyvärinen, A. (1999b). The fixed-point algorithm and maximum likelihood estimation for independent component analysis. Neural Processing Letters, 10(1):1–5.
- Hyvärinen, A. and Oja, E. (1997). A fast fixed-point algorithm for independent component analysis. Neural Computation, 9(7):1483–1492.
- Hyvärinen, A. and Oja, E. (1998). Independent component analysis by general nonlinear Hebbian-like learning rules. Signal Processing, 64(3):301–313.
- ICA demo step-by-step
  - http://www.cis.hut.fi/projects/ica/icademo/
- Lots of links
  - http://sound.media.mit.edu/~paris/ica.html
- object-based audio capture demos
  - http://www.media.mit.edu/~westner/sepdemo.html
- Demo for BBS with „CoBliSS" (wav-files)
  - http://www.esp.ele.tue.nl/onderzoek/daniels/BSS.html
- Tomas Zeman's page on BSS research
  - http://ica.fun-thom.misto.cz/page3.html
- Virtual Laboratories in Probability and Statistics
  - http://www.math.uah.edu/stat/index.html