COMPARATIVE ASSESSMENT OF ICA METHODS FOR BLIND SOURCE SEPARATION OF INSTANTANEOUS MIXTURES

Niva Das

Institute of Technical Education and Research, Deptt. of Electronics & Communication Engg., BBSR, Orissa, India
Email: nivadas@iter.ac.in

Abstract -The paper presents a comparative assessment of Blind Source Separation (BSS) methods for instantaneous mixtures based on namely generalized eigen-value decomposition, geometrical concepts, differential of mutual information and Kalman filtering applied to Nonlinear Principal Component Analysis (Nonlinear PCA). The methods highlight the independence concept underlying Independent Component Analysis (ICA). The methods have been tested on instantaneous mixtures of synthetic periodic signals, monotonous noise from electromechanical systems and speech signals. A comparison among the methods has been made on the basis of separation ability, processing time and accuracy. The quality of output, complexity of algorithms and simplicity (implementation) of the methods are some of the performance measures which have been highlighted with respect to the above signals.

Keywords: Blind source separation, ICA, Nonlinear PCA, Mutual information, Geometrical concept

I. INTRODUCTION

Blind Source Separation (BSS) has been a topic of interest in signal processing since last few years due to its potential applications in areas like speech processing, array signal processing, biomedical signal processing, image processing and telecommunications [13-17, 25, 27]. The problem of source separation concerns extracting source signals from their mixtures, which are observations at the output of a set of sensors each, receiving a different combination of the source signals. The term 'blind' is frequently used to indicate that no precise information is available of either the mixing process or the sources. This feature makes the BSS technique extremely versatile because it does not rely on modeling the underlying physical phenomena thus making it useful in many applications where the underlying physical phenomena are difficult or impossible to be modeled accurately.

We briefly describe four methods on which our study is based. These methods were selected because they are different from conventional and somewhat more recent among the approaches to solve the BSS problem. This paper is organized as follows. Section 2 discusses the basic BSS model. Section 3 presents the first BSS method where source separation problem is formulated as a generalized eigen-value decomposition problem under certain assumptions [1,21,26,28]. The second method on BSS based on geometrical concepts is described in Section 4 [2]. The third method derived from 'differential of mutual information' for blind source separation in linear mixtures is described in Section 5 [3, 22, 23, 24]. The last method in Section 6 describes a Kalman filtering algorithm applied to Nonlinear Principal Component Analysis for blind source separation of pre-whitened data [4,18-20]. Section 7 contains the results for both self generated data and real world data. The last section contains the conclusion and discussion on performance related issues.

II. BASIC BSS MODEL

Most of the linear BSS models in the simplest form can be expressed algebraically as:

$$x(k) = As(k) + n(k) \qquad for \quad k = 1,2,3,.......N \quad (1)$$

where, $x(k) = \left[x_1(k), x_2(k),......, x_m(k)\right]$ is a vector of observed signals at the discrete time instant k.

$s(k) = \left[s_1(k), s_2(k),.......s_n(k)\right]$ is a vector of the source components at the same time instant.

The source signals are assumed to be statistically independent.

n(k) represents the additive noise independent of sources.

The *N* columns of matrices *X* and *S* represent the samples in time.

A is a nonsingular matrix known as the mixing matrix having dimensions $m \times n$

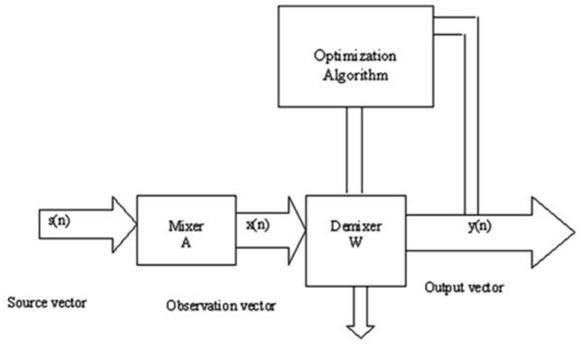

Fig. 1 The basic block diagram of the BSS process

The BSS problem can be stated as the estimation of *n* sources from *m* measurements that are unknown function of sources. The basic BSS model is shown in Fig.1. The BSS problem becomes underdetermined when the number of observations is less than the number of sources, i.e. *m<n*,. The problem becomes over-determined when the number of observations is more than the number of sources, i.e. *m>n*, .The solution to the BSS problem depends on issues like:

- mixture is linear or nonlinear.
- mixture is time varying or time-invariant.
- mixing operation is convolutive or instantaneous.
- sensors are noisy or noiseless.
- underdetermined or over-determined problem.

The source separation can be formulated as the computation of an unmixing matrix W which transforms the observed signal X to Y as:

Y=WX, Y being an estimate of X (2)

The basic BSS model considers as many sensors as sources *(m=n)*, instantaneous mixing and noise free observations. Instantaneous mixing can be seen in studio recordings, where audio signals are mixed using a mixing desk without any delay or reverberations. In biomedical applications such as fMRI and EEG, signals and images are almost instantaneous mixture problems [5]. For instantaneous noise free mixing, we have:

X=AS and Y=WX,

where the task is to recover the original sources by finding W, which is theoretically equal to the inverse of the unknown mixing matrix , i.e. $W = A^{-1}$, so that Y is as close as possible to S.

III. SOURCE SEPARATION AS A GENERALIZED EIGEN-VALUE PROBLEM

In this case the BSS problem for instantaneous mixtures is formulated as a generalized eigen-value problem under the assumption of independent sources being non-stationary, non-white or non-Gaussian. The solution for the separating matrix W is given by the generalized eigenvectors that simultaneously diagonalize the covariance matrix of the observations and an additional matrix which is selected on the basis of underlying statistical assumptions on the sources as stated above. The method provides a general and unified solution that verifies the statistical assumptions for successful source separation.

- Mathematical Formulation

The time averaged covariance matrix of the observations is given by:

$$R_x = \sum_t E\left|x(t)x^H(t)\right| \Rightarrow R_x = AR_sA^H \quad (3)$$

for a general case of complex valued variables.

H stands for Hermitian transpose.

R_s is diagonal if sources are independent or uncorrelated. The method states that for non-white, non-stationary or non-Gaussian sources in addition to the covariance matrix there exists another symmetrical matrix representing the cross-statistics Q_s, which have the same diagonalization property as equation(3) as given by:

$$Q_x = AQ_sA^H \quad (4)$$

To recover the sources from the observations X, the unmixing matrix W is to be determined such that $W^H A = I$ hence $\hat{S} = W^H X$ Further, Q_s is assumed to have non-zero diagonal values.

Multiplying equation(3) and equation(4) by W and equation(4) with Q_s^{-1} we get:

$$R_xW = AR_s \quad (5)$$

$$Q_xWQ_s^{-1} = A \quad (6)$$

Combining equations (5) and (6):

$$R_xW = Q_xWQ_s^{-1}R_s = Q_xWR_sQ_s^{-1} = Q_xWA \quad (7)$$

where by assumption, $A = R_sQ_s^{-1} = Q_s^{-1}R_s$ is a diagonal matrix.

Equation (7) constitutes a generalized eigen-value problem, where R_x *and* Q_x are square matrices and the eigen values represent the ratio of individual source statistics measured by the diagonals of R_s and Q_s. A summary of different statistical assumptions used in BSS and how they lead to different diagonal cross statistics is given below in tabular (Table.1) manner. A detailed explanation can be found in [1].

Table.1 Source characteristics ~ Diagonal Cross-Statistics

Assumptions On Source	Diagonal Cross-Statistics (Choice of Q)	Comments
1.Decorrelated (mixing matrix is orthogonal)	Q=I	The method reduces to a standard eigen-value decomposition problem
2.Non stationary and decorrelated (sources with non-stationary power)	$Q_x = R_x(t) = E\left\|x(t)x^H(t)\right\|$, $Q = Q_x$	Q is the covariance computed for a comparable period of stationarity time t. The signal is assumed to be stationary in a window.
3.Non-white and decorrelated	$Q = R_x(\tau) = E\left\|x(t)x^H(t+\tau)\right\|$	Q represents symmetric cross-correlation for time delayed $\tau.(\tau \neq 0)$
4.Non-gaussian and independent	$Q = \sum_k Cum(s_i, s_j, s_k, s_l$ $= E[x^H xxx^H] - R_x Trace(R_x) - E[xx^T]E[x^* x^H] - R_x R_x$	The method is based on ICA. is the sum over 4th order cumulants

IV. Separation Based On A Geometrical Concept

The main idea behind the approach is to project the concept of independence from a geometrical point of view. For a simple 2-source-sensor BSS problem, the sources are represented geometrically by their scatter plot, which is rectangular, if the sources are assumed to be independent. The mixing effect transforms the rectangle to a parallelogram, i.e., the scatter plot of the observed signals at the output of sensors is a parallelogram. The solution to the problem assumes same number of sources as sensors. The separation of sources can be achieved up to a permutation and a scale factor according to which, the global matrix defined by: G=WA should be equal to PD i.e. G=PD, where, P is a permutation matrix and D is a diagonal matrix [6,29-30]. The algorithm for separation of two sources from the observations of two sensors is presented here. For a more general case, explanation may be found in [2]. The algorithm separates the signals from their linear mixture in 2 steps, i.e., transformation and rotation. Fig.2 shows the block diagram of the proposed model.

Geometrically, signals are represented by their scatter diagram, which is a rectangle in case of independent signals, turning to a parallelogram when mixed through an unknown mixing matrix to be available as observed signals for a two dimensional separation problem. In the first step (transformation stage) , the parallelogram (observed signals) is transformed into a square representing orthogonal signals. Mathematically, the orthogonal signals are obtained from the observed signals by Cholesky decomposition. The covariance matrix of the observed signals and its square root become full rank matrices when sources are statistically independent and number of sources is same as number of sensors.

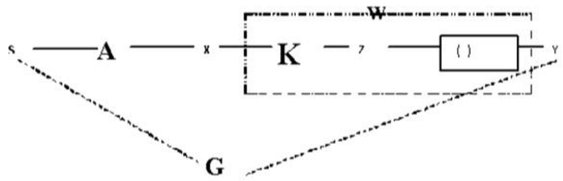

Fig. 2 Block diagram representation of BSS Model based on Geometrical Concepts

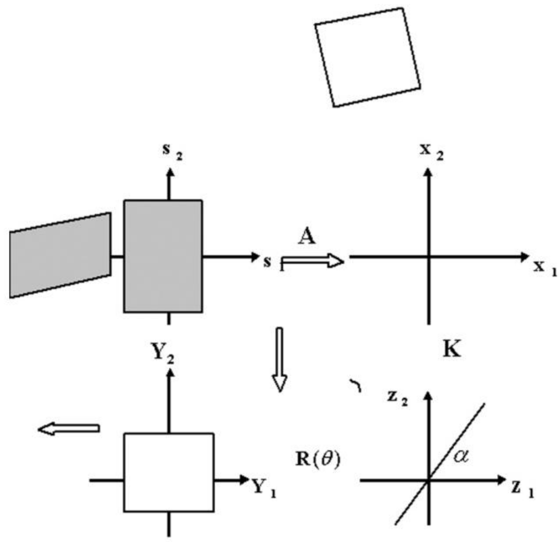

Fig. 3 Steps of BSS based on geometrical concepts

Steps of Transformation

Z= KX, K is a matrix that transforms the observed signals to orthogonal signals, Z is a matrix of orthogonal signals, X is a matrix of observed signals $K=L^{-1}$. L^{-1} is obtained through a Cholesky Decomposition of the conariance matrix of observed signals $R_X= LL^T$ (through Cholesky Decomposition), where $R_X= E[XX^T]$ represents the covariance matrix of the observed signals $R_Z=E[ZZ^T]=KE[XX^T]K^T =L^{-1} R_X$ L-T=I, where R_Z is the covariance matrix of orthogonal signals.

In the second step (rotation stage), the orthogonal signals (Z) are rotated through an angle θ to produce the estimated signals (Y).

- Mathematical steps of Rotation

$$Y = R(\theta)Z,\ where\ R(\theta) = \begin{bmatrix} \cos(\theta) - \sin(\theta) \\ \sin(\theta) \quad \cos(\theta) \end{bmatrix}$$

The steps of the algorithm are shown in Fig.3.

$'\alpha'$ represents the angle between the first diagonal and the horizontal axis and can be estimated from the coordinates of the farthest point from the origin. Finally to achieve separation, the orthogonal signals are rotated by an angle θ, which varies with the class of the probability density function (PDF) of the source as specified in Table.2 below.

Table.2 Rotation angle (θ) ~ Nature of Source PDF

Rotation angle (θ)	Nature of Source PDF
1. $\theta = \pi/4 - \alpha$	Sources have uniform PDF or close to uniform PDF
2. $\theta = -\alpha$	Sources have unimodal PDF like symmetrical Gamma PDF or close to symmetrical Gamma PDF like Laplace or Cauchy

V. Source Separation Using The Differential Of Mutual Information

This is a powerful gradient based approach for minimization of mutual information in a parametric model. The model is given by:

$$y(t) = \xi(x(t);\theta) \qquad (8)$$

Where ξ is a known separating system with unknown parameters θ, x(t) represents observed signals and y(t) represents estimate of source signals.

The condition of independence imposed on the output signals ensures separation. The degree of independence between the output components can be measured using mutual information *I*, which is defined as:

$$I(y) = \int_y p_y(y) ln \frac{p_y(y)}{\Pi_i p_{yi}(y_i)} dy \qquad (9)$$

where, . $y = [y_1, y_2, ..., y_n]^r$ Equation(9) also defines the Kullback-Leibler divergence between $p_y(y)$ and $\Pi_i p_{yi}(y_i)$. $I(y)$ is always non-negative and vanishes when the components of *y* become independent of each other. Hence the solution to the BSS problem in (8) reduces to finding the parameter vector θ, that minimizes $I(y)$ [7,8]. The gradient of mutual information can be expressed in terms of score functions as defined below:

Score Function: The score function of a scalar random variable *x* is defined as the negative of the log derivative of its density, i.e.

$$\psi_x(x) = -\frac{d}{dx} ln\ p_x(x) = -\frac{p_x^{'}(x)}{p_x(x)} \qquad (10)$$

where p_x is the probability density function (PDF) of *x*.

Marginal Score Function: The marginal score function (MSF) of a random vector $X = [x_1, x_2, ..., x_n]^7$ is the vector of score functions of its components, i.e.,

$$\psi_x(X) = [\psi_1(x_1), \psi_2(x_2), ...\psi_n(x_n)]^T \qquad (11)$$

where $\psi_i(x_i) = \frac{d}{dx_i} ln\ p_{xi}(x_i) = \frac{p_{xi}^{'}(x_i)}{p_{xi}(x_i)}$

Joint Score Function: The joint score function (JSF) of x is the gradient of $"-ln\ p_x(X)"$ i.e.,

$$\varphi(x) = [\varphi_1(x), \varphi_2(x), ...\varphi_n(x)]^T \qquad (12)$$

where

$$\varphi_i(x) = -\frac{\partial}{\partial x_i} ln\ p_x(x) = -\frac{(\partial/\partial x_i) p_x(x)}{p_x(x)} \qquad (13)$$

Score Function Difference: The score function difference (SFD) of x is the difference of its MSF and JSF, i.e.,

$$\beta_x(x) = \psi_x(x) - \varphi_x(x) \qquad (14)$$

The gradient based method can be developed from the knowledge of "differential" of the mutual information, i.e., its variation resulting from a small deviation in its argument [9].

Theorem-1: The differential of mutual information is stated as:

$$I(x + \Delta) - I(x) = E\{\Delta^T \beta_x(x)\} + o(\Delta) \qquad (15)$$

where, $\beta_x(x)$ is the SFD of x and $o(\Delta)$ denotes higher order terms in Δ .

Equation(15) may be stated in the following form:

$$I(x+\varsigma y)-I(x)=E\{(\varsigma y)^T\beta_x(x)\}+o(\varsigma) \quad (16)$$

where, x and y are bounded random vectors, ς is a matrix with small random entries and $o(\varsigma)$ stands for a term that converges to zero faster than $\|\varsigma\|$. Equation(16) is simpler than equation(15) and easier to be used in developing gradient based algorithms for optimizing mutual information. Any multivariable differentiable function *f(x)* can have a form

$$f(x+\Delta)-f(x)=\Delta^T(\Delta f(x))+o(\Delta) \quad (17)$$

A comparison of (15) with (17) implies that, the SFD can be viewed as the "stochastic gradient" of the mutual information.

Property of SFD: The components of a random vector are independent if and only if, its SFD is zero. Hence considering SFD as a gradient for mutual information and utilizing this, it can be stated that "gradient of mutual information must vanish" is a necessary and sufficient condition for *I* to be minimum.

Many methods are available for estimating JSF and SFD for minimizing mutual information .We have highlighted only the Histogram method, which estimates SFD directly. Though Histogram is not an accurate estimator of PDF, it works fine for instantaneous BSS problems since direct estimate of SFD does not require a very good estimate of PDF.

Histogram method for estimating SFD: A histogram can be used for estimating the joint PDF of x . For two-dimensional vectors, let $N(n_1,n_2)$ denote the number of observations in the bin (n_1,n_2), then the histogram estimation of p_x is:

$$p(n_1,n_2)=\frac{N(n_1,n_2)}{T} \quad (18)$$

where, T is the number of observations. From (18) a histogram estimation of is p_{x1} obtained by:

$$p_1(n_1)=\sum\nolimits_{n2}p(n_1,n_2) \quad (19)$$

From (18) and (19) an estimation of $p(x_2\mid x_1)$ is obtained as: $p(n_2\mid n_1)=\dfrac{p(n_1,n_2)}{p_1(n_1)}=\dfrac{N(n_1,n_2)}{N(n_1)}$

From 14) we have, . $\beta_1(x_2,x_1)=\dfrac{\frac{\partial}{\partial x_1}p(x_2\mid x_1)}{p(x_2\mid x_1)}$

A histogram estimation of $\beta(x_2\mid x_1)$ is given by:

$$\beta_1(n_1,n_2)=\frac{p(n_2\mid n_1)-p(n_2\mid n_1-1)}{p(n_2\mid n_1)} \quad (20)$$

$\beta_2(n_1,n_2)$ is estimated in a similar manner. The value of $\beta(n_1,n_2)$ is assigned to all the points in bin (n_1,n_2).

Gradient approach to mutual information minimization: In this approach, $\dfrac{\partial I(y)}{\partial\theta}$ is first calculated using Theorem-1. Then, the parameter vector is updated as per the steepest descent algorithm given by: $\theta=\theta-\mu\dfrac{\partial I(y)}{\partial\theta}$

For linear instantaneous mixtures y=Wx,. Let $\hat{W}=W+\varsigma$, where ς is a small matrix.

Then the new output is: $\hat{y}=\hat{W}x=Wx+\varsigma x+y+\varsigma x$

From Theorem-1 we obtain:

$$\hat{I}-I=E\{\beta_y^T(y)\varsigma x\}=\langle\varsigma,E\{\beta_y(y)x^T\}\rangle \quad (21)$$

where $\langle .,.\rangle$ stands for scalar product of matrices.

Equation(21) implies: $\dfrac{\partial I(y)}{\partial W}=E\{\beta_y(y)x^T\}$ (22)

The separation algorithm is given by:

$$W=W-\mu E\{\beta_y(y)x^T\} \quad (23)$$

where, μ is a small positive constant. To get a separation quality independent of the mixing matrix, it is preferable to use relative gradient instead of $\dfrac{\partial I(y)}{\partial W}$, i.e.,

$$\nabla_w I=\frac{\partial I}{\partial W}W^T=E\{\beta_y(y)y^T\} \quad (24)$$

$$W=(I-\mu\nabla_w I)W \quad (25)$$

The algorithm may be stated using the following steps:

- Initialization: W=I and y=x
- Loop:

1) Estimate $\beta_y(y)$ (using Histogram method)

2) $\nabla_w I=E\{\beta_y(y)y^T\}$

3) $W = (I - \mu \nabla_w I)W$

4) $y = Wx$

5) Normalization: $y_i = y_i / \sigma_i$, where σ_i^2 is the energy of y_i. Divide the i-th

row of W by σ_i.

- Repeat until convergence.

VI. Kalman Filtering Algorithm For BSS

A Kalman filtering based nonlinear PCA has been used for blind sourse separation of pre-whitened data [10-12]. Assuming same number of sources as sensors, the nonlinear PCA algorithm with pre-whitening may be stated as:

$$J(w) = E\left\{ \left\| v_t - W^T g(W v_t \right\|^2 \right\} \qquad (26)$$

$E\{.\}$ is the expectation operator, $\|.\|$ stands for Euclidean norm,

v_t, is the whitened vector and $y_t = W v_t$.

$$\phi(y_t) = [\phi_1(y_1(t)), \phi_2(y_2(t)), ...\phi_n(y_n(t))]^T$$

denotes a vector of nonlinearly modified output signals.

To develop Kalman filtering algorithm for BSS, the state equation and measurement equations are expressed as follows:

$$z_t = \phi(W_{t-1} v_t) \qquad (27)$$

$$\begin{aligned} \overrightarrow{W_{t+1}} &= \overrightarrow{W_t} && (state\ equation) \\ v_t &= \left[I \otimes z_t^T \right] \overrightarrow{W_t} + e_t && (measurement\ equation) \end{aligned} \qquad (28)$$

where, $\overrightarrow{W_t} = vec(W_t)$ is a vector obtained by stacking the columns of one W_t beneath the other. $\otimes$ stands for Kronecker product, I denotes $n \times n$ identity matrix, e_t models the measurement noise. The state equation has $n^2 \times n^2$ identity state transition matrix. The optimum weight matrix at equilibrium points is time-invariant. Equation(28) represents the standard state-space models of Kalman filters as given in [4].

Let C_t measurement matrix = $I \otimes z_t^T$

Q_t =covariance matrix of e_t.. The Kalman filter algorithm may be formulated as:

$$\begin{aligned} G_t &= K_{t-1} C_t^T \left[C_t K_{t-1} C_t^T + Q_t \right]^{-1} \\ K_t &= K_{t-1} - G_t C_t K_{t-1} \\ \overrightarrow{W_t} &= \overrightarrow{W_{t-1}} + G_t \left[v_t - C_t \overrightarrow{W_{t-1}} \right] \end{aligned} \qquad (29)$$

$K_t = E\left\{ \left(\overrightarrow{W_t} - \overrightarrow{W_{opt}} \right) \left(\overrightarrow{W_t} - \overrightarrow{W_{opt}} \right)^T \right\}$ is the state error correlation matrix. To reduce the heavy computational burden encountered in (29) a simplified approach is presented below:

Proposition: Let H_i be $n \times n$ symmetric matrices for $i = 0,1$ and p_t a scalar.

The state error correlation matrix at time t may be written as: $K_t = I \otimes H_t$, provided that $Q_t = p_t I$ and the initial matrix selected is given by: $K_0 = I \otimes H_0$

Using the Proposition and applying the identity, $C_t \overrightarrow{W_{t-1}} = W_{t-1}^T z_t$ a simplified and computationally efficient Kalman filtering algorithm is presented as:

$$\begin{aligned} h_t &= H_{t-1} z_t \\ g_t &= h_t / (z_t^T h_t + p_t) \\ H_t &= H_{t-1} - g_t h_t^T \\ W_t &= W_{t-1} + g_t \left[v_t - W_{t-1}^T z_t \right]^T \end{aligned} \qquad (30)$$

where $y_t = W_{t-1} v_t$ and $z_t = \phi(y_t)$

VII. Results

The performance of all the above methods have been tested on three varieties of data and the results are presented in this section. For each type, the source signal (available in these cases), observed signal and the separated signal are shown. The method used to measure the performance of the separating algorithms tests the diagonal property of the system matrix, G, i.e., G = WA (a product of demixing and mixing matrix).

Case1: Two synthetically generated sinusoids with different frequencies sampled at a sampling frequency as specified in the result sheet are mixed using a random mixing matrix.

Case2: Acoustic electromechanical signals generated from induction machines running at two different speeds are considered. The noise of each machine is recorded at a sampling frequency of 22050 Hz. The samples from two machines collected separately are mixed using a random mixing matrix.

Case3: Speech signals of two speakers (one male and one female) each recorded at a sampling frequency of 22050 Hz are considered. The speech samples collected from individual speakers are artificially mixed using a mixing matrix.

- Separation process
- Generalized Eigen-value Decomposition Method

Case1: (sinusoids) Two sinusoids with frequencies 500 Hz and 900 Hz are sampled at a sampling frequency of 10 KHz.

Number of samples considered: 10000,

Noise (additive white Gaussian): 15dB.

The separated signal, source signal and the mixed signal are shown in Fig.4-a.

Correlation between source signal and estimated signal are found to be: 0.9488 & -0.9491

Sinusoids with noise fall under the category of decorrelated and white source statistics having a close to flat spectrum. Hence the mixing matrix chosen is orthogonal and the diagonal cross statistics Q is chosen as an Identity matrix.

Case2: (Electromechanical noise) Source data is stationary but falls under the category of "non-white and decorrelated" source statistics. Hence, the diagonal cross statistics Q is chosen as:.

$$Q = R_x(\tau) = E\left|x(t)x^H(t+\tau)\right|$$

Number of samples considered: 48165.

Time lag selected $(\tau.)$ = 1.

Correlation between source signals and estimated signals are found to be: -0.9999 & -1

The separated signal, source signal and the mixed signal are shown in fig.4-b.

Case3: (speech signal) Speech can be classified either as non-stationary or as non-white or as non-Gaussian signal. Here speech is considered as a non-stationary decorrelated signal and consequently diagonal cross statistics Q is selected as:

$$Q_x = R_x(t) = E\left|x(t)x^H(t)\right|,$$

$$Q = Q_x$$

T (stationarity time): 10,000

Number of samples considered: 169248

Correlation between source signals and estimated signals are found to be: -0.9999 & 1.

The separated signal, source signal and the mixed signal are shown in fig.4-c.

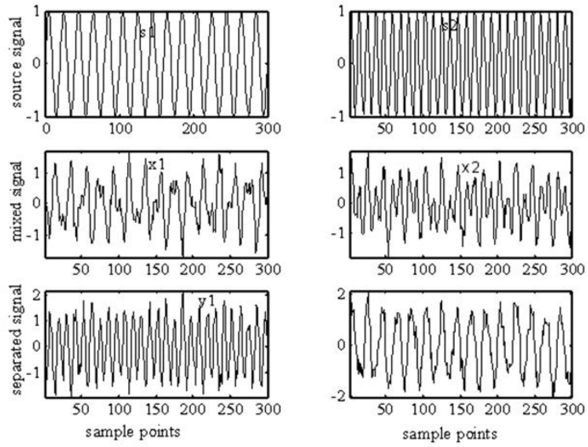

Fig. 4(a) BSS based on generalized eigen-value decomposition (sinusoids)

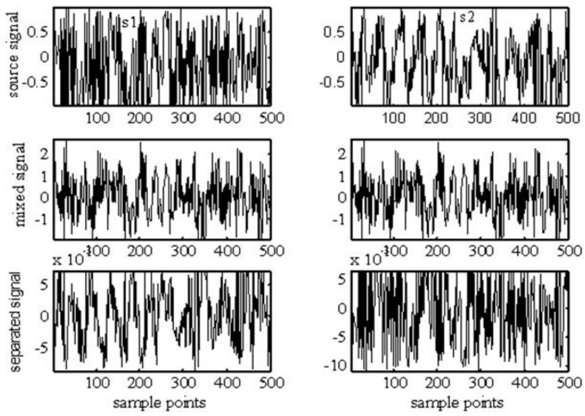

Fig. 4(b) BSS based on generalized eigen-value decomposition (electromechanical noise)

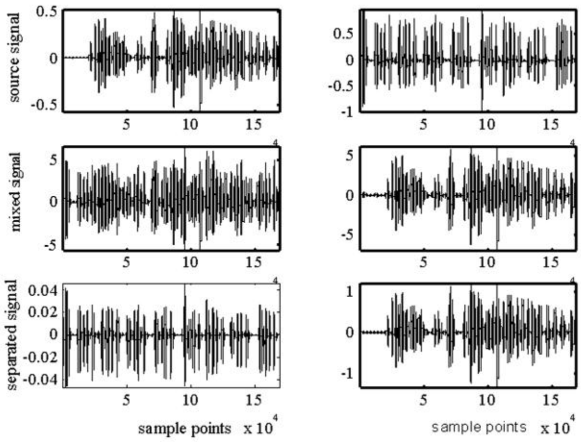

Fig. 4(c) BSS based on generalized eigen-value decomposition (speech signals)

- Algorithm based on Geometrical Concept

Case1: (sinusoids) Two sinusoids with frequencies 400 Hz and 900 Hz are sampled with a sampling frequency 10 KHz.

Number of samples considered: 20000.

Noise (additive white Gaussian): 13dB.

Since sinusoids with noise have a close to uniform PDF, the rotation angle θ is chosen as: $\theta = (\pi / 4) - \alpha$ as specified in the literature.

The correlation between source signals and estimated signals are found to be: -0.9606 & 0.9297. The separated signal, source signal and the mixed signal are shown in fig.5-a.

Case2: (electromechanical noise)

Number of samples considered: 24082.

The rotation angle $\theta = (\pi / 4) - \alpha$.

Correlation between source signals and estimated signals found to be: -0.9988 & 0.9971

The separated signal, source signal and the mixed signal are shown in fig.5-b.

Case3: (speech signals) Speech of one male and one female speaker are recorded at a sampling frequency of 22050 Hz.

Number of samples considered: 56416

Since speech signal exhibits a PDF close to symmetrical gamma PDF or Laplace PDF, the rotation angle is given by : $\theta - \alpha$.

Correlation between source signal and estimated signal are found to be: 0.9986 & 0.8711.

The separated signal, source signal and the mixed signal are shown in fig.5-c.

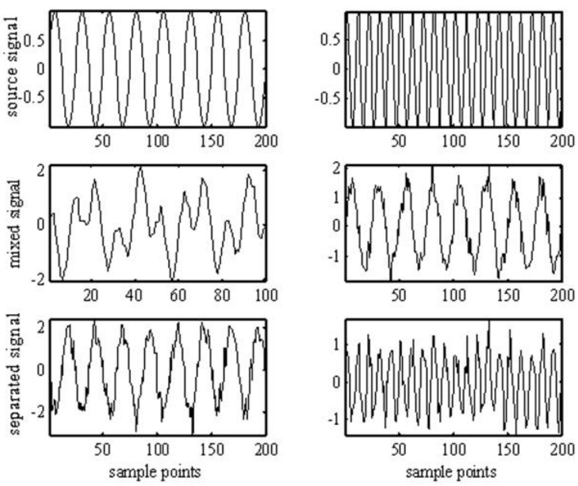

Fig. 5(a) BSS based on geometrical concepts (sinusoids)

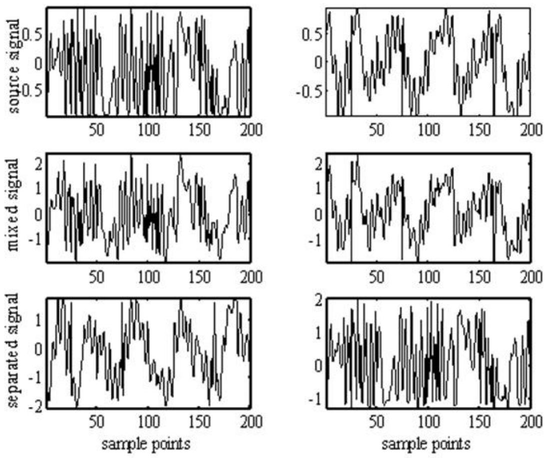

Fig. 5(b) BSS based on geometrical concepts (electromechanical noise)

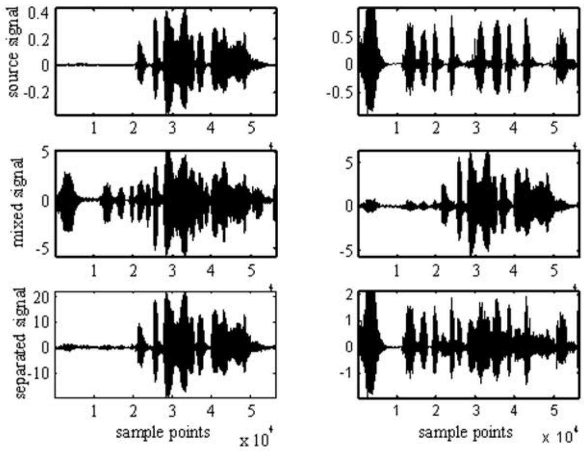

Fig. 5(c) BSS based on geometrical concepts (speech signals)

Gradient based Mutual Information Minimization

Case1: (Histogram method of PDF estimation): Two sinusoids with frequencies 400 Hz and 1 KHz are sampled at a sampling frequency of 5 KHz and mixed using a random mixing matrix.

Number of samples considered: 1000.

Histogram size: 10 x 10

Number of iterations required for convergence: 50

Correlation between source signal and estimated signal found are to be 0.989 & 0.99 without noise. Correlation between source signal and estimated signal are found to be 0.89 & 0.88 with 13dB noise.

The separated signal, source signal and the mixed signal are shown in fig.6-a.

Case2: (Electromechanical noise)

Number of samples considered: 2000.

Histogram size: 10 x 10

Number of iterations required for convergence: 50

Correlation between source signal and estimated signal are found to be 0.9998 & 0.9982.

The initial value of demixing matrix : Identity matrix

The separated signal, source signal and the mixed signal are shown in fig.6-b.

Case3: (speech signal)

Number of samples considered: 169248.

Histogram size: 50 x 50

Number of iterations required for convergence: 50

Correlation between source signal and estimated signal are found to be 0.99988 & 0.99988. The separated signal, source signal and the mixed signal are shown in fig.6-c.

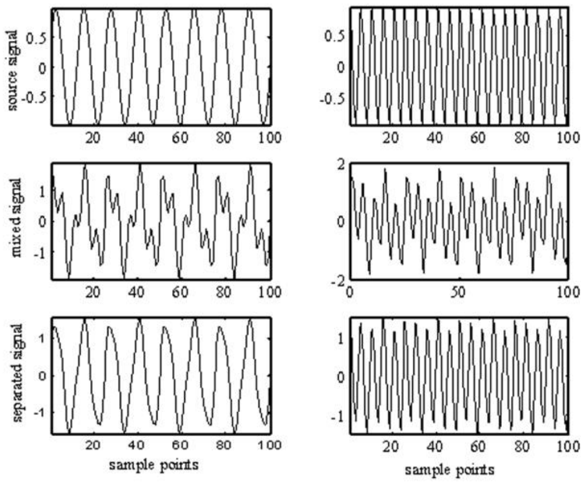

Fig. 6(a) BSS based on Differential of Mutual information (sinusoids)

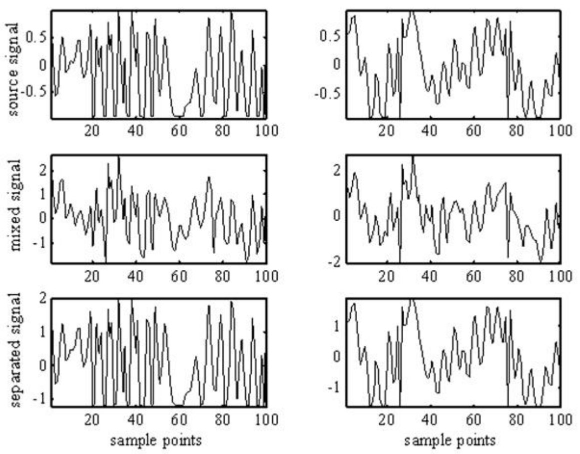

Fig. 6(b) BSS based on Differential of Mutual information (electromechanical noise)

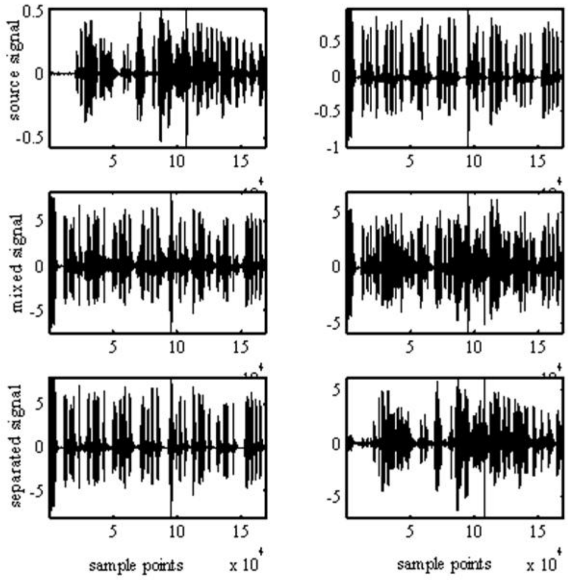

Fig. 6(c) BSS based on Differential of Mutual information (speech signal)

- Kalman Filtering with Non-linear PCA

Case1: (sinusoids) Sinusoids with frequencies 400 Hz and 900 Hz are sampled at a frequency 5 KHz and are mixed using a random mixing matrix.

The initial value of state-error-correlation matrix: $K_0 = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$

The initial value of demixing matrix W_0 is a randomly generated symmetric matrix.

Number of samples considered: 5000.

Correlation between source signals and estimated signals are found to be -0.9621 & -0.9713. The separated signal, source signal and the mixed signal are shown in fig.7-a.

Nonlinear function used: g(t) = t - tanh(t)

Case2: (electromechanical noise)

The initial value of state-error-correlation matrix : $K_0 = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}$

The initial value of Demixing matrix W_0 is a randomly generated symmetric matrix.

Number of samples considered: 8000.

Correlation between source signals and estimated signals are found to be 0.9665 & -0.9195. The separated signal, source signal and the mixed signal are shown in fig.7-b.

Nonlinear function used: g(t) = tanh(t)

Case3: (speech)

The initial value of state-error-correlation matrix : $K_0 = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}$

The initial value of Demixing matrix W_0 is a randomly generated symmetric matrix.

Number of samples considered: 56416.

Correlation between source signals and estimated signals are found to be 0.9677 & -0.9726. The separated signal, source signal and the mixed signal are shown in fig.7-c.

The nonlinear function used: g(t) = tanh(t).

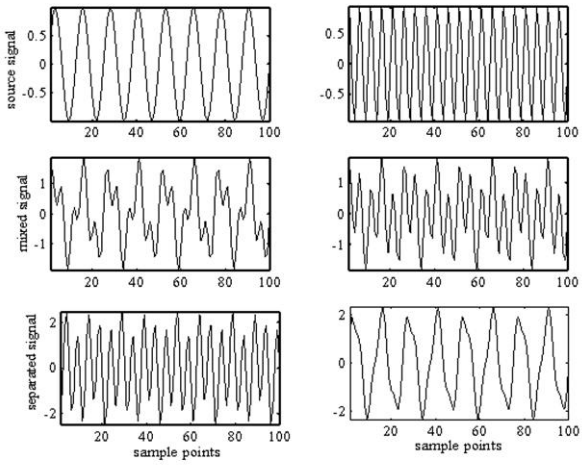

Fig. 7(a) BSS based on Kalman filtering applied to nonlinear PCA (sinusoids)

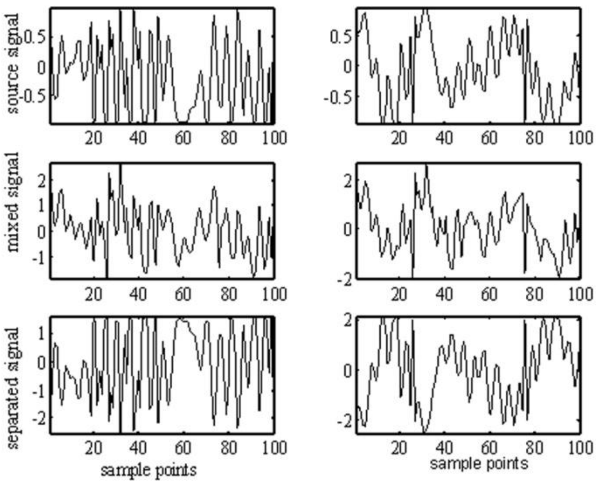

Fig. 7(b) BSS based on Kalman filtering applied to nonlinear PCA (electromechanical noise)

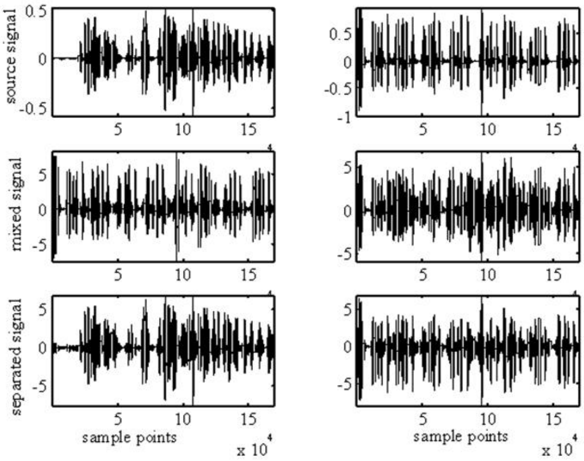

Fig. 7(c) BSS based on Kalman filtering applied to nonlinear PCA (speech signal)

Table 3 CPU Time taken for separation

BSS Method	Signal Type	CPU Time
Generalized Eigen Value Decomposition	Sinusoids with noise(15 dB) Fig.4-a	0.032 sec
	Induction machine data Fig.4-b	0.031 sec.
	Speech data Fig.4-c	0.031 sec.
Algorithm based on geometrical concepts	Sinusoids with noise(13 dB) Fig.5-a	5.579 sec.
	Induction machine data Fig.5-b	0.98 sec.
	Speech data Fig.5-c	57 sec.
Gradient based mutual information minimization	Sinusoid with noise(13 dB) Fig.6-a	7.5 sec.
	Induction machine data Fig.6-b	20.797 sec
	Speech data Fig.6-c	90.89 sec
Kalman Filtering with Nonlinear PCA	Sinusoids without noise Fig.7-a	6 sec.
	Induction machine data Fig.7-b	4.2 sec
	Speech data Fig.7-c	20.96 sec

VIII. Conclusion

This paper presented a review of some of the recently published BSS methods applied to speech, acoustic electromechanical noise and synthetic signals. All the separation algorithms performed well in all cases.

In Eigen value decomposition method, since the order and scale of these eigen vectors are arbitrary, recovered sources also get arbitrarily scaled and permuted. These ambiguities can be resolved by scaling the eigen vectors to unit norm and sorting them by the magnitude of their generalized eigen values. The method can be used for over determined BSS but doesn't address the under determined case. The method is also not robust to estimation error.

Experimental results from BSS based on geometrical concept demonstrate superior separation ability for stationary and non-stationary data. The algorithm is robust and performs well even under 10dB noise. For stationary and less number of source-sensor combination, the algorithm requires less number of samples for convergence. Convergence time of the algorithm is very competitive. One drawback of the algorithm is its inability to separate a mixture of sources having uniform and unimodal PDF. For more than three source-sensor combination the algorithm becomes computationally complex.

Gradient based mutual information minimization technique demonstrates superior separation ability, competitive convergence speed. One limitation of the method is the requirement of multivariate PDF estimation that requires large number of data when the number of variables is large. Therefore the maximum number of sources that can be estimated is limited to 3 or 4.

Kalman filtering algorithm shows its proficiency in separation ability and convergence time. Since the estimation progresses sample wise it has the best tracking ability. The algorithm is suitable for real time data and on-line processing.

References

[1] L.Parra, P. Sajda, Blind Source Separation via Generalized Eigenvalue Decomposition, Journal of Machine Learning Research, 2003 – MIT Press.

[2] A. Mansour, N. Ohnishi, C.G. Puntonet, Blind multiuser separation of instantaneous mixture algorithm based on geometrical concepts, Signal Processing 82 (2002) 1155-1175.

[3] M. Babaie-Zadeh, C. Jutten, A general approach for mutual information minimization and its application to blind source separation, Signal Processing 85(2005) 975-995.

[4] X. Zhu, X. Zhang and Y. Jia, A new Kalman filtering algorithm for Nonlinear Principal Component Analysis, Springer-Verlag Berlin Heidelberg, 2005, 1010-1015.

[5] M. McKeown, S. Makeig, S. Brown, T.-P. Jung, S. Kindermann, A.J. Bell, V. Iragui, and T. Sejnowski. Blind separation of functional magnetic resonance imaging (fMRI) data. Human Brain Mapping, 6(5-6):368-372, 1998.

[6] P. Comon, Independent Component Analysis, a new concept? Signal Processing 36(3), (April 1994) 287-314.

[7] A. J. Bell and T. Sejnowski, An information-maximization approach to blind separation and blind deconvolution, NeuralComputation, vol.7, pp.1129–1159, 1995.

[8] D.T. Pham, Mutual information approach to blind separation of stationary sources, IEEE Trans. Inform. Theory 48(7) (July 2002) 1-12.

[9] M. Babaie-Zadeh, C. Jutten, K. Nayebi, Differential of mutual information, IEEE Signal Processing Letters, 11(1), (January 2004) 48-51.

[10] Grewal, Mohinder S., Angus P.,Andrews (1993), Kalman Filtering –Theory and Practice, Prentice Hall, NJ USA.

[11] E. Oja, The nonlinear PCA learning in Independent Component Analysis, Neurocomputing, 17(1997) 25-46.

[12] Q. Lv, X.D. Zhang, Y. Jia, Kalman Filtering Algorithm for Blind Source Separation, ICASSP(2005).

[13] J. Karhunen, A. Hyvärinen, R. Vigario, J. Hurri, and E. Oja. Applications of neural blind separation to signal and image processing. In Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'97), pages 131-134, Munich, Germany, 1997.

[14] L.De Lathauwer,B. de Moor, J. Vandewalle. Fetal electrocardiogram extraction by source subspace separation. In Proc. HOS'95,pages 134-8, Aiguablava,Spain, June 1995.

[15] S. Makeig, A.J. Bell, T.-P. Jung, and T.-J. Sejnowski. Independent component analysis of electroencephalographic data. In Advances in Neural Information Processing Systems 8, pages 145-151. MIT Press, 1996.

[16] R. Vigário, V. Jousmäki, M. Hämäläinen, R. Hari, and E. Oja. Independent component analysis for identification of artifacts in magnetoencephalo graphic recordings. in Advances in Neural Information Processing 10 (Proc. NIPS'97), pages 229-235, Cambridge, MA, 1998. MIT Press.

[17] R. Vigário. Extraction of ocular artifacts from EEG using independent component analysis. Electroencephclin. Neurophysiol., 103(3) :395-404, 1997.

[18] E. Oja. The nonlinear PCA learning rule in independent component analysis. Neurocomputing, 17(1):25-46.

[19] J. Karhunen and J. Joutsensalo. Generalizations of principal component analysis, optimization problems, and neural networks. Neural Networks, 8(4):549-562, 1995.

[20] J. Karhunen, P. Pajunen, and E. Oja. The nonlinear PCA criterion in blind source separation: Relations with other approaches. Neurocomputing, 22:5-20, 1998.

[21] Belouchrani, K. Abed Meraim, J.-F. Cardoso, and E. Moulines. A blind source separation technique based on

second order statistics. IEEE Trans. on S.P., 45(2):434-44, 1997.

[22] J.J. Atick. Entropy minimization: A design principle for sensory perception? International Journal of Neural Systems, 3:81-90, 1992.

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

[24] M. Wax and T. Kailath. Detection of signals by information-theoretic criteria. IEEE Trans. on ASSP, 33:387-392, 1985.

[25] M. McKeown, S. Makeig, S. Brown, T.-P. Jung, S. Kindermann, A.J. Bell, V. Iragui, and T. Sejnowski. Blind separation of functional magnetic resonance imaging (fMRI) data. Human Brain Mapping, 6(5-6):368-372, 1998.

[26] A. Mansour, A. K. Barros and N. Ohnishi, "Blind separation of sources: Methods, assumptions and applications, "IEICE Trans. On Fundamental of Electronics, Communications and Computer Sciences, vol.E83-A, pp. 1498-1512, 2000.

[27] Martens S. M, Rabotti C, Mischi M, Sluijter RJ, "A robust fetal ECG detection method for abdominal recordings," Physiol Meas. 2007 Apr; 28(4): 373-88. Epub 2007 Mar 7.

[28] A.-J. van der Veen, S.Talvar, and A. Paulraj, "A subspace approach to blind space-time signal processing for wireless communication systems, "IEEE Trans. Signal Processing, vol.45, pp. 173-190, Jan. 1997.

[29] A. Hyvarinen and E. Oja, "Independent Component Analysis: algorithms and applications," Neural Netw.13, pp.411-430, 2000.

[30] S. Choi, A. Cichocki, H.-M. Park, and S.-Y. Lee, "Blind source separation and Independent component analysis: A Review," Neural Information Processing- Letters and Reviews, vol.6, No.1, January 2005 Systems, 3:81-90, 1992. Supp. 1992 on second order statistics. IEEE Trans. on S.P., 45(2):434-44, 1997. other approaches. Neurocomputing, 22:5-20, 1998. 1997.

◈◈◈