ESANN'2003 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 23-25 April 2003, d-side publi., ISBN 2-930307-03-X, pp. 131-136

# Comparison of Neural Algorithms for Blind Source Separation in Sensor Array Applications

Guillermo Bedoya, Sergio Bermejo & Joan Cabestany

Department of Electronic Engineering, Technical University of Catalonia.
Jordi Girona 1-3, C4 building, 08034 Barcelona, Spain.
{bedoya, sbermejo, cabestan}@eel.upc.es

**Abstract -** A test bed of experiments with real and artificially generated data has been designed to compare the performance of three well-known algorithms for BSS. The main goal of these experiments was to extract some guidelines for their use in practical applications concerning their efficiency, accuracy, convergence speed, stability, and robustness under the presence of Gaussian noise and in presence of a large number of source signals.

## 1. Introduction

Blind source separation (BSS) is a technique, which allows separating a number of source signals from observed mixtures of those sources without a previous knowledge of the mixing process [1]. A lot of attention have been aroused in these techniques in recent years with an increasing number of existing approaches. However, much less attention has been received the comparison of performance between BSS algorithms for sensor array applications since researchers usually focus their comparison on particular aspects using simple data sets.

In this work, we propose a test bed of experiments to compare BSS algorithms based on the linear BSS model. In particular, we compare three well-known BSS algorithms (Fast ICA, JADE and eeA) applied to artificial and real data including data extracted from an Ion-sensitive field effect transistor (ISFET) sensor array.

The rest of this paper is organised as follow: the principles of BSS, the evaluated algorithms and the comparison criteria are introduced in section 2. Performance measures for the BSS algorithms are introduced in section 3. In section 4 the benchmarking and experiments are presented. Discussion is included in section 5.

## 2. Blind Source Separation

Assume that M is the number of source signals, they are unknown and mutually independent of zero mean and are denoted by a random vector $\mathbf{s} = [s_1,\ldots,s_M]$; the observations $\mathbf{x}$, are samples from the linear transformation of these independent sources via an unknown matrix, such a:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \qquad \textbf{(2.1)}$$

Where **A** is the unknown mixing matrix. The goal of BSS is to estimate the original sources **s** given only the observations **x**, from the realistic assumption of the independence of the sources. The estimation of independent sources **y** is done by means of the demixing matrix **W** as follows:

$$\mathbf{y} = \mathbf{W}\mathbf{x} = \mathbf{W}\mathbf{A}\mathbf{s} \qquad \textbf{(2.2)}$$

Where $\mathbf{W} = \mathbf{A}^{-1}$. However, the simple independence condition is not enough to recover the exact order or scale of the sources, so the demixing matrix has the form $\mathbf{W}=\mathbf{P}\mathbf{D}\mathbf{A}^{-1}$**,** where **P** is a permutation matrix and **D** = diag ($d_1$, $d_2$, $d_3$, …$d_n$) is a diagonal matrix for arbitrary scaling. Following we present the compared algorithms.

**Fixed-point (FP or Fast ICA) algorithm [2]:** Based on maximising non-gaussianity by means of minimising negentropy (or maximising the absolute value of negentropy). Thus, under the whiteness constraint, minimising the mutual information between entries of ***y*** is equivalent to minimising the sum of the entropies of the entries of ***y***. Mixing the entries of ***s*** tends to increase their entropies. So, we need to find separated source signals as those with minimum marginal entropies. The algorithm is based on the Newton's optimisation method, which make very fast the convergence, and different non-linearities (tanh (u), $u^3$ and $\exp(u^2/\sigma)$) can be employed in order to compute the independent components

**Joint Approximate Diagonalisation of Eigenmatrices (JADE) algorithm [3]***:* While the fixed-point algorithm optimises a transform of the data, JADE optimises the transform of a particular set of statistics about the data. JADE measure the mutual information between cross-cumulants. Second order cumulant is used to decorrelate the data, i.e., to obtain a whitening matrix $\hat{\mathbf{W}}$. The separation matrix is estimated as $\mathbf{V}'\cdot\hat{\mathbf{W}}$, where $\mathbf{V}'$ is a rotation matrix used to make the cumulant matrices as diagonal as possible according to an specific contrast function, which is the sum of squared fourth order cross cumulants from the cumulants matrix. To make the cumulants diagonal as possible is the same that makes the data as independent as possible, so the matrix that performs the diagonalisation of cumulants can be used to perform the separation on the mixed data.

**Extreme Event Analysis (eeA) [4]:** based on nonholonomic-nested Newton's methods for ICA. This method factorises the Newton's method to minimise sum of squared fourth order cross cumulants between components. It does not prewhiten the observed data and uses solely the fourth order statistics. It is related to the nonholonomic method (extension of the natural gradient method introducing the so-called nonholonomic constraints). The nonholonomic constraints in the natural gradient method allow resolving the non-stationary source signals problem. It is expected a fast convergence when the number of observation channels be quite large because of the method to determine the demixing matrix, and a better computational load.

### 2.2. Comparison criteria and Measures

ESANN'2003 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 23-25 April 2003, d-side publi., ISBN 2-930307-03-X, pp. 131-136

To ensure that an algorithm can be implemented in a sensor array-based microsystem, there where taken into account the measures on table 1.

| Comparison criteria | Performance Measure |
|---|---|
| Accuracy | PI (as defined by equation 3.1) and SNR |
| Robustness under the presence of added Gaussian noise | PI and SNR |
| Robustness under the presence of Gaussian sources | PI and SNR, when Gaussian sources replace the original sources |
| Stability | PI and SNR at different stopping criterion |
| Efficiency | PI and SNR at varying number of sources |
| Computational Load | Convergence speed |
| Number of tuneable parameters | |

Table 1. Comparison criteria and measures.

## 3. Performance Measures

Once the separating matrix **W** was computed, we can check the separation quality using the measure so-called performance index (PI) proposed in [1] as:

$$\mathrm{PI}(k)=\frac{1}{m}\sum_{i=1}^{m}\left\{\sum_{j=1}^{m}\frac{|p_{ij}|^2}{\max_q |p_{iq}|^2}-1\right\}+\frac{1}{m}\sum_{j=1}^{m}\left\{\sum_{i=1}^{m}\frac{|p_{ij}|^2}{\max_q |p_{qj}|^2}-1\right\} \qquad \textbf{(3.1)}$$

Where $p_{ij}$ denotes the component of the i$^{th}$ row and the j$^{th}$ column of **P**=(pij)=**AW**.
The second measure of separation quality used was the SNRs of the separated outputs. The equation describing the SNR is:

$$SNR = 10\log_{10}\left(\frac{\sum\left[s(t)^2\right]}{\sum\left[n(t)^2\right]}\right) \qquad \textbf{(3.2)}$$

Where s (t) is the desired signal and n (t) = y (t) – s (t) is the noise indicating the undesired signal. y (t) is the estimated source signals.

## 4. Benchmarking and comparison criteria

**Accuracy and computational load:** comparison of the dependence of the separation quality and convergence time on the training data size. Data was introduced by means a partition into sixteen no overlapped data sets, each one with a different size using a window method. Permutation of the demixed signals was made after every iteration (computing PI, SNR and convergence time for each data set) in order to ensure that the first separated source be on the first output vector. To study the behaviour of batch algorithms, the partition was repeated overlapping the data sets at 20%, 50% and 90% overlapping ratios.
**Noisy mixtures:** To study the effect of additive noise on the separation quality, we

ESANN'2003 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 23-25 April 2003, d-side publi., ISBN 2-930307-03-X, pp. 131-136

added to the mixtures the white Gaussian noise vector **n**, according to **x=As+n**, with variance from 0 to 1 in 0.01 step size. The comparison was done for SNR > 30db (zero noise limit) and for SNR less than 20db. The noise influence in the mixtures was controlled in two different ways, introducing four different noise components to each sensor signal and introducing the same component of noise to all signals.

**Efficiency and Stability:** The number of sources was increased to compare the efficiency. Since its assumed that signals are non-Gaussian, we compared the results when sub-Gaussians and super-Gaussians source signals were introduced. Stability was compared varying the number of iterations (500-1000) and the stopping criterion for FP and eeA algorithms from 0.0001 to 0.001. JADE it was not compared with this criteria.

### 4.2 Numerical Experiments

**Experiment 1. Performance on artificially generated data:** four artificially generated source signals (5000 points), two periodic deterministic signals, one impulsive noise and a random Gaussian signal were mixed using a randomly generated matrix to obtain a 4 dimensional observation vector.

**Experiment 2. Performance on real data with artificial mixing:** A set of 17000 samples of real recording sounds were mixed using a randomly generated mixing matrix.

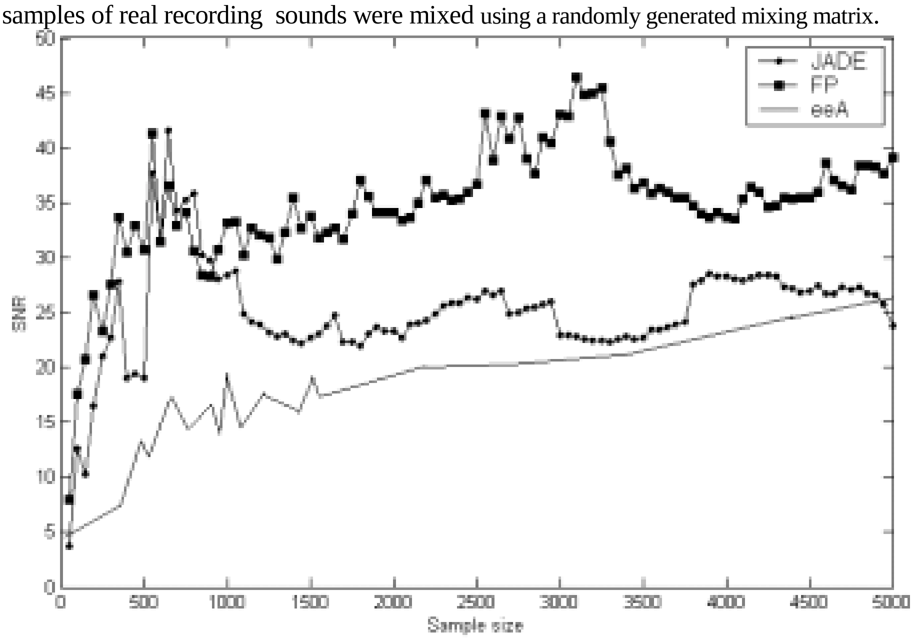


Figure 1. SNR of separated signal (sub-Gaussian source) at different data set sizes for exp. 1. FP using tanh (squares), JADE (circles), eeA(solid line).

**4.3 Experiment 3. Performance on real data with real mixing:** Using the response of three ISFETs sensors (5000 points), where the source signals are the ionic concentrations in an aqueous solution. Simulation was done according to the device

Electro-chemical model [5] [6]. The ISFETs array mixing system have been considered linear instantaneous, i.e., biasing the devices in the linear region and working in the range of ionic concentrations, which generates a linear response [5].

## 5. Simulations and Conclusion

A comparative study between the three algorithms for BSS reviewed in section 2 has been performed using the test set proposed in Section 3 considering (number of sensors ≥ number of sources). The artificial mixing matrix was chosen to be nearly singular simulating the situation where sensors are located very close to each other. From the results of the experiments, which are summarised on fig. 1 and tables 2 and 3, the below considerations can be made.

| **Algorithm** Non-linearity used | **PI 5000 points** ($x10^{-3}$) | **Mean SNR (dB)** | **Max. SNR (dB)** |
|---|---|---|---|
| FP - (Tanh) | 9.34 | 1.49 | 3.26 |
| FP - ($u^3$) | 5.82 | 1.90 | 6.24 |
| FP- $\exp(u^2/\sigma)$ | 2.20 | 2.68 | 4.64 |
| JADE | 6.3 | 2.87 | 6.21 |
| eeA | 1.84 | 4.19 | 8.94 |

Table 2. Accuracy from experiment 2 for noisy Mixtures when SNR equal to 15db.

| **Algorithm** | **PI 5000 points** ($x10^{-3}$) | **Mean SNR (dB)** | **Max SNR (dB)** |
|---|---|---|---|
| FP - (Tanh) | 3.11 | 21.03 | 33.4 |
| FP - ($u^3$) | 4.75 | 16.88 | 28.5 |
| FP- $\exp(u^2/\sigma)$ | 4.28 | 15.70 | 36 |
| JADE | 5.7 | 13.02 | 34 |
| eeA | 82.7 | 8.21 | 24 |

Table 3. Accuracy from Experiment 3.

***Accuracy and efficiency***: the SNR (see fig. 1) increases slow and PI decreases slow when the number of samples is larger than 3000 points for sub-Gaussian sources. However, the convergence time increases considerably when sample size is larger than 4000 points. The better relation between accuracy and convergence time can be achieved using data sets with 2500 points for sub-Gaussian sources while 8800 points (for JADE) and 10000 points (for FP($\exp(u^2/\sigma)$) and eeA) are good for speech signals (super-Gaussian). When the number of sources is increased, the efficiency of the eeA algorithm is lightly better than the other algorithms, that's because of the determination of the demixing matrix is computationally much cheaper for large number of sources $N$. JADE requires an eigenvalue decomposition of the $N^2 \ x \ N^2$ cumulant matrix and the storage required for all the 4th-order moments become impractical for implementations with large data sets (depending on the available memory). FP using sequential extraction approach shows the poorest efficiency under the presence of more than 6 sources, its accuracy improves when symmetric approach is used. From table 3, FP using gauss ($\exp(u^2/\sigma)$) nonlinearity and JADE presented high accuracy compared with eeA (even at reduced sample sizes). There was not significant accuracy improvement ovelapping the data sets in the batch training.

***Robustness under Gaussian noise****:* All algorithms for BSS that require whitening are sensitive to additive noise because of the use of the equal-time correlation matrix (corrupted by noise). Table 2 shows the mean and the maximun value of the SNR (for the whole data) and the PI (for 5000 points). Based on the results, the use of only fourth order cumulants, which are non sensitive to the Gaussian noises, make the nested Newton's method the most robust under their presence for SNR between 15 and 20dB. For SNR >28 dB the eeA algorithm accuracy is not much better than FP and JADE, and for SNR < 15 dB separation of the Gaussian noise from sources is not achieved. Chosing the mixing matrix to be singular make the algorithms more robust under the presence of small gaussian noises. For medium level Gaussian noises, a whitening method based on the use of the time-delayed correlation matrices (not sensitive to the additive with noise) was used. When the whitening method was applied, at low SNR, the algorithms gives slightly better performance.
***Computational load and stability****:* The FP algorithm shows the better convergence time for all the experiments (previous tuning) in contrast with eeA (which shows the bigger computation time) and JADE. The advantage of JADE is that the use of cross-cumulants do not require gradient descent techniques for optimisation avoiding convergence problems and it has not tuneable parameters because the learning is based in a statistical way rather than a data domain. FP using tanh and ($\exp(u^2/\sigma)$) as nonlinearties shows the better stability under the presence of super-Gaussian sources. The computational load of FP algorithm is smaller than the other algorithms.
***Effectiveness in presence of Gaussian sources****:* For more than two Gaussian sources, the algorithms have very poor performance, especially eeA, because they cannot find the directions of the columns of the mixing matrix in order to locate the edges of the joint density of the observations, when the maximisation of the measure of nongaussianity is done. For achieve Gaussian sources separation, it is necessary to use algorithms based on the information about the time correlation structure of the data. An extended graphic results presentation can be found at: www.geocities.com/guillermoledor2/bss_isfet.html.

**Acknowledgements**
This work is part of the EU. IST programme No. 2000-28084 (SEWING).

## References

[1] A. Cichocki & S-I. Amari. (2002). Adaptive Blind Signal & Image Processing. Wiley.
[2] A. Hyvärinen & E. Oja. (1999). Fast and robust fixed-point algorithms for ICA. *IEEE trans. On Neural Networks*, vol. 10, 626-634.
[3] J.-F. Cardoso & A. Souloumiac. (1993). Blind Beamforming for non-Gaussian signals. IEE *Proceeding part F*, Vol. 140, No. 6. 362-370.
[4] T. Akuzawa, "Nested Newton´s method for ICA and post factor analysis", available from: http://www.mns.brain.riken.go.jp/~akuzawa/
[5] A. Napieralski, W. Wojciak, D. Marcin, M. Szermer (2001). Library of Chemfets for the ELDO environment. *Technical report for the SEWING project.*
[6] S. Martinoia & G. Massobrio (2000). A Behavioural macromodel of the ISFET in SPICE. Sens. Actuat. B 62, 182-189.