BULETINUL INSTITUTULUI POLITEHNIC DIN IAŞI
Publicat de
Universitatea Tehnică „Gheorghe Asachi” din Iaşi
Tomul LVII (LXI), Fasc. 3, 2011
Secţia
ELECTROTEHNICĂ. ENERGETICĂ. ELECTRONICĂ

AN APPLICATION BASED ON FASTICA ALGORITHM FOR SEPARATION OF TWO OR MORE CORRELATED AND UNCORRELATED SIGNALS

BY

ALEXANDRU-MIHNEA SPIRIDONICĂ[*] and ADRIAN-GABRIEL FERARU

“Gheorghe Asachi” Technical University of Iaşi,
Faculty of Electrical Engineering, Energetics
and Applied Informatics

Received, June 12, 2011
Accepted for publication: August 16, 2011

Abstract. A fundamental problem in many research areas is to find a suitable representation of multivariate data. Most times the desired representation is a linear transformation of the initial data. In other words, each component of the linear representation is given by a linear combination of initial variables. Among the well known linear transformation methods we note the Principal Component Analysis (PCA), Factorial Analysis and Projection Pursuit. A method who knew a big growth in recent years in terms of data representation is the Independent Component Analysis (ICA). The goal of this method is a linear representation of data that do not have a Gaussian distribution so that its components to be statistically independent. This paper presents some of the most important concepts of Independent Component Analysis and then, based on the notions presented in the first part of the paper, it realizes a practical application of separation of two or more mixed signals, both correlated and uncorrelated, in order to identify the independent components.

Key words: cocktail-party problem; multivariate data; mixtures of signals; genetic algorithms.

---

[*] Corresponding author: *e-mail*: aspiridonica@ee.tuiasi.ro

1. Introduction

A fundamental problem in many research areas is to find a suitable representation of multivariate data. Most times the desired representation is a linear transformation of the initial data. In other words, each component of the linear representation is given by a linear combination of initial variables. Among the well known linear transformation methods we note the Principal Component Analysis (PCA), Factorial Analysis and the Projection Pursuit. A method who knew a big growth in recent years in terms of data representation is the Independent Component Analysis (ICA). The goal of this method is a linear representation of data that do not have a gaussian distribution so that its components to be statistically independent. Although it is just in the beginning era, this method has become one of the most used statistical signal processing method, being used in numerous important applications. So, ICA may be defined as being a statistical method for separating a multivariate signal into additive components based on statistical independence of non-gaussian signal sources and it is a special case of Blind Source Separation (Hyvarinen & Oja, 2000; Hyvarinen, 1997). According to Lee (2001) ICA is a signal processing method to extract independent sources given only observed data that are mixtures of the unknown sources. Among the areas in which the contributions of this method are remarkable should be noted speech enhancement systems, telecommunications and medical signal processing. So, the ICA algorithm has been successfully applied in biomedical signal processing such the analysis of electroencephalographic data, functional magnetic resonance imaging data (Makeig *et al.*, 1996). Bell and Sejnowski (1997) suggested that independent components of natural scenes are edge filters and those independent image components can be used as features in pattern classification problems such as visual lip-reading and face recognition systems. In 1998 it was developed an overcomplete representation of ICA formulation which includes an additive noise model that can be used to infer more sources than sensors (Hyvarinen, 1997). Feature extraction is also an important application on ICA. A crucial problem in digital signal processing is to find appropiate representations for image, audio or other kind of data for tasks like compression and denoising (Delorme, 2007).

2. Mathematical Model

Most of the applications that use the ICA model are based on the cocktail-party problem. Suppose that in a room unaffected by the additional noise they are two people speaking simultaneously at two microphones placed in different locations. The initials signals that are emitted by these two speakers are noted with $s_1(t)$ and $s_2(t)$. The recorded signals provided by the microphones are denoted by $x_1(t)$ and $x_2(t)$, where x_1 and x_2 represents the amplitudes of the two signals and t is the time index. Each of these signals is a weighted sum of recorded voice signals emitted by these two speakers which have been noted above with $s_1(t)$ and $s_2(t)$. We can read this thing as a linear equation, as it follows:

$$x_1(t) = a_{11}s_1 + a_{12}s_2 \,, \tag{1}$$

$$x_2(t) = a_{21}s_1 + a_{22}s_2 \,; \tag{2}$$

$a_{11}, a_{12}, a_{21}, a_{22}$ are the parameters that depend on the distances of the microphones from the speakers. Ideally it should be able to estimate the two original signals, $s_1(t)$ and $s_2(t)$, based only on recorded signals, $x_1(t)$ and $x_2(t)$. If the parameters a_{ij} would be known the system of equations would be easy to be solved by classical methods, but usually these parameters are not known and the problem becomes more difficult. An approach could be finding some of the information about statistical properties of the signals $s_i(t)$ in order to estimate a_{ij} parameters. It is sufficient to assume that $s_1(t)$ and $s_2(t)$ (the signals emitted by those two speakers) are statistically independent at every moment of time, t. In this situation ICA can be used to estimate the a_{ij} parameters based on the information of their independence. According to this information, it allows separation of the two source signals, $s_1(t)$ and $s_2(t)$ by their mixtures, $x_1(t)$ and $x_2(t)$. As a conclusion, we can speak that the ICA was developed for approaching the problems that are high related with the cocktail-party problem. Due to increase of the interest regarding such analysis, it became clear that this principle has many interesting applications. As a illustration we consider the waveforms from the Figs. 1 and 2. These are not realistic speaking

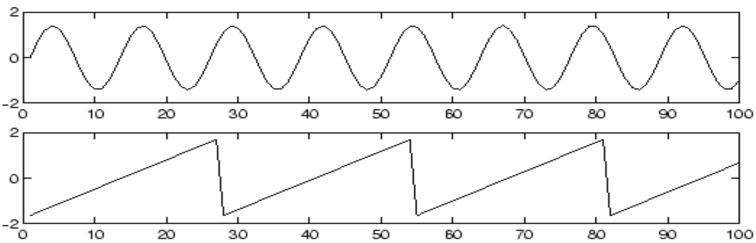

Fig.1 – The initial signals.

signals, but are enough for this illustration. The initial speaking signals are represented in the Fig. 1 and mixed signals, in the Fig. 2. The problem is to determine data from the Fig. 1 based on data from Fig. 2. In the Fig. 3 we have the estimations of the initial source signals, by using only the observed signals from the Fig. 2.

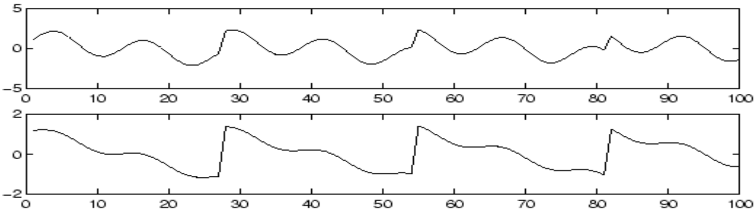

Fig. 2 – The observed mixtures of the initial source signals from Fig. 1.

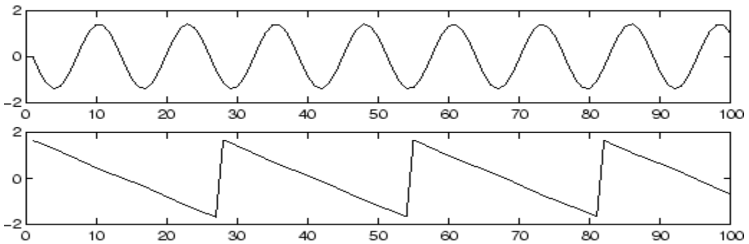

Fig. 3 – The estimations of the initial source signals, by using only the observed signals from the Fig. 2.

3. An Application of Separation of Two or More Mixed Signals by Using Independent Component Analysis Properties

In the second part of this paper was realized an application, using MATLAB language, to compare two or more signals, both independent and dependent with each other. The algorithm underlying this application is known as FASTICA algorithm or Mixedsig algorithm and estimates the independent components from given multidimensional signals. Each line from the mixedsig matrix is an observed signal. It is important to say that FASTICA uses the fixed-point algorithm of Hyvarinen (1997). The output depends on the number of output arguments, namely:

a) [icasig] = FASTICA (mixedsig) – the lines of icasig contain the estimated independent components;

b) [icasig, *A*, *W*] = FASTICA (mixedsig) – provides the separate matrix, *W*, and corresponding mixed matrix, *A*;

c) [*A*,*W*] = FASTICA (mixedsig) – provides only the estimated mixed matrix, *A*, and separate matrix, *W*.

FASTICA may be called with numerous optional arguments. The optional arguments are offered in pairs of parameters, so that the first argument is parameter's name and the next argument is the value for that parameter. The pairs of parameters can be given in any order. Some examples can be very helpful for understanding this problem as it follows:

a) [icasig] = FASTICA (mixedsig,'approach','symm','g','tanh') – done ICA with nonlinearity tanh and in parallel (like the maximum likelihood method for supernormalized data);

b) [icasig] = FASTICA (mixedsig,’lastEig’,10,’numOfIC’,3) – reduces the dimension at 10 and estimates only 3 independent components.

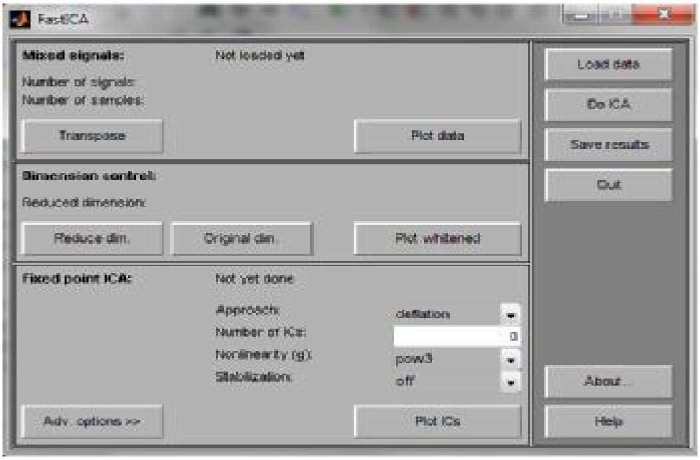

Fig. 4 – The main window of the program FASTICA.

By using the Load Data button we load the necessary data. In the first example we considered a random gaussian signal and a random exponential signal

signal1 = [randn(1,501);exprnd(1,1,501)].

By clicking the „Plot data” button it displayed the graphics of initial mixed signals, and by clicking the button „Plot Whitened” we have a filtering of mixed signals so that these mixed signals to be closer to their original form.

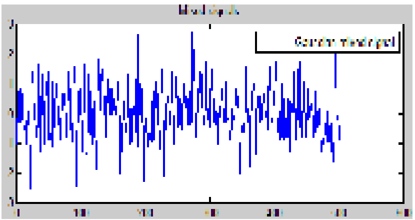
Fig. 5 *a* – Gaussian mixed signal.

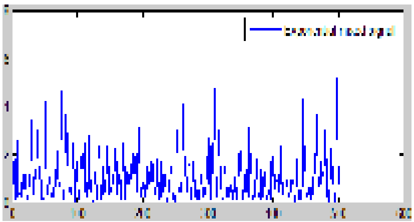
Fig. 5 *b* – Exponential mixed signal.

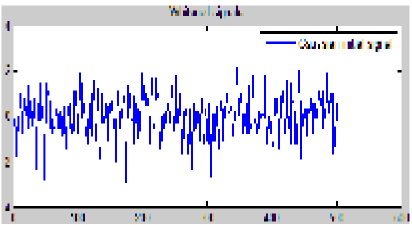
Fig. 6 *a* – Gaussian initial signal.

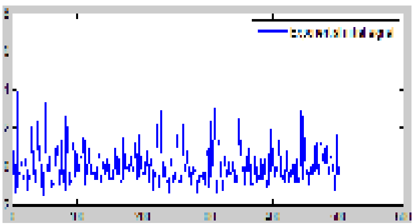
Fig. 6 *b* – Exponential initial signal.

As it can see from the Fig. 5 *a* the random mixed exponential signal has not negative values of the *Oy*-axis, an obvious fact because the exponential distribution has always positive values. In the Figs. 6 ***a*** and 6 ***b*** we can see the initial signals resulting by using the ICA algorithm. For this case we applied the deflation approach, with two independent components, as nonlinearity we used pow3 and the number of necessary steps required for the two signals to return to the initial form was 5 for the first independent component (the random gaussian filtered signal) and 2 for the second independent component (the random exponential filtered signal). Obviously, the number of required steps for the algorithm application is not the same at every rule and this number is dependent upon the moment in which the algorithm finds the best filtering. If the symmetric approach is used instead of deflation, the convergence is realized after 5 steps and the estimated value is near to zero. Also, it can see the statistical indices of each signal (Figs. 7 ***a***, 7 ***b***, 8 ***a***, 8 ***b***).

Check to plot statistics on figure:

	X	Y
min	1	-2.642
max	501	3.093
mean	251	-0.02691
median	251	-0.03985
mode	1	-2.642
std	144.8	1.013
range	500	5.735

Fig. 7 ***a*** – Statistical indices for the random gaussian mixed signal.

Check to plot statistics on figure:

	X	Y
min	1	0.002454
max	501	7.328
mean	251	1.065
median	251	0.6767
mode	1	0.002454
std	144.8	1.121
range	500	7.325

Fig. 7 ***b*** – Statistical indices for the random exponential filtered signal.

Check to plot statistics on figure:

	X	Y
min	1	-2.563
max	501	3.095
mean	251	-2.438e-017
median	251	-0.02385
mode	1	-2.563
std	144.8	1.001
range	500	5.657

Fig. 8 ***a*** – Statistical indices for the random gaussian mixed signal.

Check to plot statistics on figure:

	X	Y
min	1	-5.644
max	501	1.068
mean	251	-4.347e-016
median	251	0.3179
mode	1	-5.644
std	144.8	1.001
range	500	6.713

Fig. 8 ***b*** – Statistical indices for the random exponential filtered signal.

In the second example were considered two strong correlated signals, the second signal being the square of the first signal:

$$x = \text{randn}(1,501);$$
$$x_2 = \text{randn}(1,501).\text{^}2;$$
$$\text{signal} = [x;x2].$$

The graphics are those represented in Figs. 9 and 10.

Because the two signals are highly correlated (the second signal is the square of the first signal), the initial signals are almost identical with the mixed signals. The deflation approach was used and as nonlinearity the „pow3” function, with two independent components. For the filtration of the first signal the number of required steps was 6 until the algorithm found the best fit and for the filtration of the second signal the number of required steps was 2.

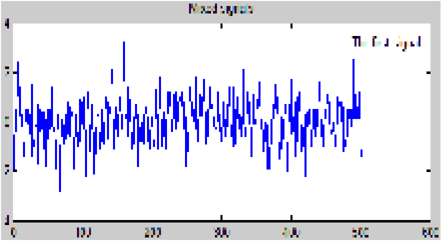

Fig.9 *a* – The first mixed signal.

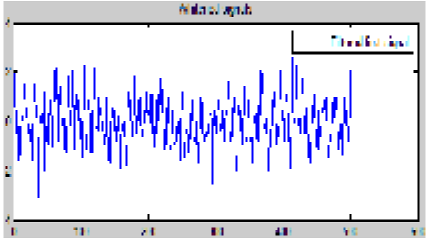

Fig.10 *a* – The first initial signal.

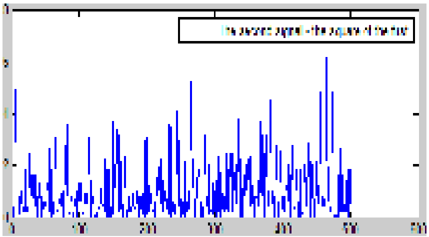

Fig. 9 *b* – The second mixed signal – the square of the first mixed signal.

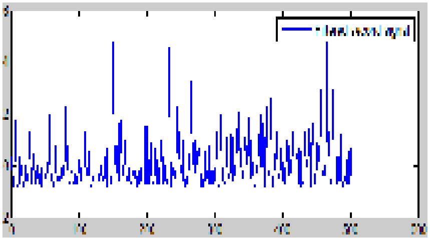

Fig. 10 *b* – The second initial signal.

In the last example from this paper three signals, *a*, *b* and *c* were considered, that depends on vector *x* that contains 100 values random generated, as it follows:

$$x = \text{randn}(1,100);$$
$$a = x.\hat{}2+1;$$
$$b = x.\hat{}3+2*x.\hat{}2+1;$$
$$c = x+1;$$
$$signal = [a;b;c].$$

The graphics are represented in Figs.11 and 12.

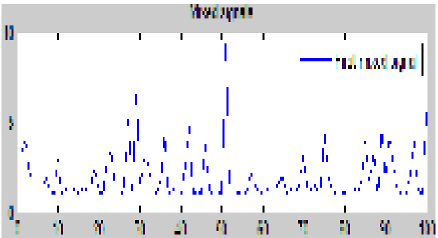

Fig. 11 *a* – The first mixed signal.

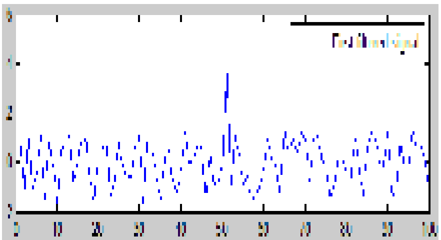

Fig. 12 *a* – The first filtered signal.

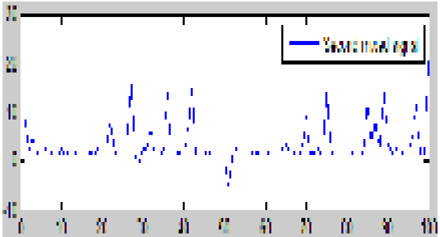

Fig. 11 *b* – The second mixed signal.

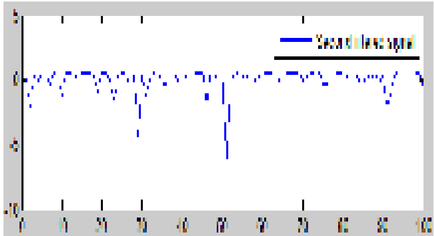

Fig. 12 *b* – The second filtered signal.

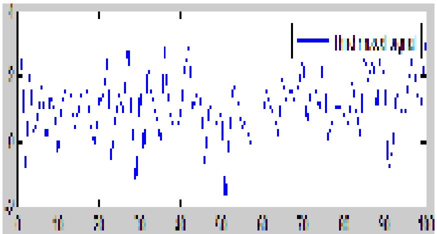

Fig. 11 *c* – The third mixed signal.

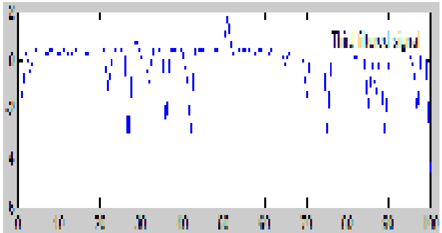

Fig. 12 *c* – The third filtered signal.

Because these three mixed signals are not correlated, we can see that the filtered signals are different one by one. This thing is obvious if we look at the formulas of these signals, formulas that do not reveal any correlation between a signal and another. For the filtering of the first two signals the program used a number of 5 steps and for the last signal the program used a number of only 2 steps. As a remark, it can say that as the mixed signal is given by a more complex formula, so the number of steps required for the ICA algorithm is higher. If as nonlinearity we used not „pow3", but „tanh" (hyperbolic tangent), „gauss" or „skew" then as formula through the signal is given is more complex, the number of steps is higher. In this case, for the filtering of the first signal the program used 32 steps for „tanh" nonlinearity and 40 steps for „gauss" nonlinearity.

4. Conclusions

Independent components analysis is a development of analysis of independent components principles. The algorithm tries to achieve some de-correlations between input signals, which are given in a mixed form, in one by one independent signals. In this work, by using ICA algorithm, we realized an analysis of three types of mixed signals, in order to notice certain similarities or dissimilarities between initial independent components of these. We observed if that mixed signals are highly correlated then exists very small differences from the initial independent components, and if some mixed signals are given by complicated formulas their original form differ more and more. The FASTICA method is a very fast and precise method for estimate the independent components and based on this method can be made a high number of estimations of the initial signals in many areas of technology.

Acknowledgments. This work was developed in the framework of the Project “BRAIN” (PhD scholarships ID 6681, OI POS DRU Ministry of Education, Research and Innovation).

REFERENCES

Bell A.J., Sejnowski T.J., *An Information Maximization Approach to Blind Separation and Blind Deconvolution*. Neural Computation, **7,** *6*, 1129-1159 (1997).

Delorme A., *ICA (Independent Component Analysis) for Dummies*. 2007. http://www. sccn.ucsd.edu/~arno/indexica.html.

Hyvarinen A., *Fast and Robust Method – Point Algorithms for Independent Component Analysis*. Proc. IEEE NNSP Workshop, Amelia Island, Florida, USA, 1997, 388-395.

Hyvarinen A., Oja E., *Independent Component Analysis: Algorithms and Applications, Neural Networks Research Centre*. Univ. of Technology, Helsinki, Finland, 2000.

Hyvarinen A., *One-Unit Contrast Functions for Independent Component Analysis: A Statistical Analysis*. Proc. IEEE NNSP Workshop, Amelia Island, Florida, USA, 1997, 388-397.

Jang G.-J., Lee T.-W., Oh Y.-H., *Learning Statistically Efficient Features for Speaker Recognition*. Proc. of IEEE Internat. Conf. on Acoustics, Speech a. Signal Proc., Salt Lake City, Utah, USA, May 2001.

Makeig S., Bell A.J., Jung T.-P., Sejnowski T.-J., *Independent Component Analysis of Electroencephalographic Data*. Adv. in Neural Inform. Proc. Syst., **8**, MIT Press, Cambridge, MA., 1996.

APLICAŢIE BAZATĂ PE ALGORITMUL FASTICA DE SEPARARE A DOUĂ SAU MAI MULTE SEMNALE CORELATE ŞI NECORELATE

(Rezumat)

Analiza componentelor independente reprezintă o dezvoltare a analizei principiilor componentelor independente. Algoritmul încearcă realizarea unor decorelaţii între semnalele de intrare, care sunt date într-o formă mixată, în unul câte unul semnale independente. În cadrul acestei aplicaţii, utilizând algoritmul ICA, s-a realizat o analiză a trei tipuri de semnale mixate, în scopul de a remarca anumite similarităţi sau diferenţe între componentele independente iniţiale ale acestora. S-a observat că dacă semnalele mixate sunt puternic corelate atunci există diferenţe foarte mici şi faţă de forma iniţială a componentelor independente, iar dacă anumite semnale mixate sunt date de formule complicate atunci forma lor iniţială diferă şi mai mult. Metoda FASTICA este foarte rapidă şi precisă permiţând estimarea componentelor independente şi pe baza acestei metode pot fi realizate diverse estimări ale semnalelor iniţiale din numeroase domenii ale tehnologiei.