COMPARATIVE ANALYSIS OF THE ICA ALGORITHMS APPLIED ON A 2D SIGNAL

Nikola Bešić[•], *Gabriel Vasile*[†], *Budimir Lutovac, Srdjan Stanković and Dragan Filipović*[‡]

Keywords: *image, FastICA, jade, fobi, cdf*

Abstract: Blind source separation is one of the major areas of research in signal and image processing today. Being a broad area, it actually comprises various distinctive techniques. The revolutionary one, considered here, is Independent Component Analysis (ICA). Nowadays, the FastICA method, based on the fixed-point iterative algorithm, is probably the most widespread method in estimating independent components. Aside from it, there are numerous methods based on tensorial decomposition, among which FOBI and JADE are regarded as the most representative. In this paper we introduce these ICA methods and analyze their performances, illustrated through the application on a synthetic 2D signal (image).

1. INTRODUCTION

The very purpose of the blind source separation is recovering source signals without having detailed knowledge of the mixing process. If we assume having the vector **X** consisted of *n* either time or space dependent observations:

$$\begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ \vdots \\ x_n(t) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} s_1(t) \\ s_2(t) \\ s_3(t) \\ \vdots \\ s_n(t) \end{bmatrix}, \tag{1}$$

we are supposed to relate it to the vector **s** consisted of *n* sources being functions of the same variable (time or space) as the corresponding observations [1]. The matrix **A** is called a mixing matrix, and it provides information concerning the "involvement" of each of the

• N. Bešić is with the GIPSA-lab, Grenoble INP, Grenoble, FR and the University of Montenegro, Podgorica, ME.
Corresponding author: nikola.besic@gipsa-lab.grenoble-inp.fr
† G. Vasile is with the GIPSA-lab, CNRS, Grenoble, FR.
‡ B. Lutovac, S. Stanković and D. Filipović are with the University of Montenegro, Podgorica, ME.

sources for each of the observations.

If we knew coefficients of the mixing matrix (a_{ij}), the problem would be quite simple – reduced to solving a system of linear equations, but in fact, we do not have any information on the mixing process. That is why this kind of a separation is called blind. Therefore, the only thing we can do it to assume the statistical relation of the source signals and then to try to estimate these signals. Exactly this kind of the statistical relation is the base for the classification of blind source separation methods, where the most distinguished ones are the Principal Component Analysis (PCA) and the Independent Component Analysis (ICA).

The PCA assumes that the sources are statistically uncorrelated. This method is limited to the analysis of the second order statistics – covariance matrix of the observations' vector (**x**). It can be demonstrated that the resulting mixing matrix after the normalization corresponds to the unitary matrix of covariance matrix eigenvectors. The one-dimensional sources are taking role of the eigenvalues, providing the "power" of the corresponding mixing matrix column.

On the other side, the ICA is rather based on the assumption of the statistical independence of the sources. In this case, the higher order statistics is considered as well. This fact becomes essential once we deal with the non-Gaussian observations.

There are several different approaches in estimating independent components, among which the most representative are the ones based on emphasizing the non-Gaussianity of the sources (FastICA) [2] and the ones exploiting higher-order statistics through tensorial decompositions (FOBI, JADE). Even though the latter appear to be equally attractive in the context of ICA, the method emphasized in this this paper is rather the FastICA algorithm, given its performances in case of high-dimensional data and the availability. Therefore, starting by describing the idea behind, we introduce the algorithm itself with a strong emphasis on the role of different nonlinearities which are necessarily used in estimating independent components. Finally, by applying the algorithm on two synthesized "mixed" images (2D signals), we compare the statistics of the estimated sources with the simulated ones (used in "mixing"). The comparison is made with respect to the chosen nonlinearity. The presented comparative analysis is additionally enhanced by introducing the results achieved with Forth Order Blind Identification (FOBI) [3] and Joint Approximate Diagonalization of Eigenmatrices (JADE) [4].

2. THE IDEA BEHIND THE FASTICA

The most important restriction concerning ICA (and therefore FastICA) is the statistical distribution of the independent components. It cannot be Gaussian, for if the mixing matrix actually represents orthogonal transform, the observations will also have the joint Gaussian pdf, leading to the unidentifiability of mixing matrix columns directions (for the elements random vector, the joint pdf will have the shape of a circle).

This restriction represents the founding principle of the concerned approach in the IC estimation – FastICA [4]. Namely, after the Central Limit Theorem, the sum of the independent random variables always tends to follow normal distribution. If we assume that there is a vector **y** which is related to the observation vector by means of relation:

$$\mathbf{y} = \mathbf{wx}, \tag{2}$$

the goal is to find vector **w** so that it corresponds to one of the rows of the inverse mixing matrix $\mathbf{A}^{-1}$. As we don't know the mixing matrix, the established criterion is minimizing the Gaussianity of vector **y**. By introducing the relation **z** = **wA** we could rewrite the previous equation as:

$$\mathbf{y} = \mathbf{wx} = \mathbf{wAs} = \mathbf{zs}\,, \tag{3}$$

with the vector **z** indicating the mismatch between the estimated vector **w** and the corresponding row of the inverse mixing matrix $\mathbf{A}^{-1}$.

As the central theorem states the product will be the least Gaussian if we have just one member of the sum, saying if the **z** vector has just one non-zero element, which implies that $y_i = s_i$. So we need to determine vector **w** which maximizes the non-Gaussianity of **y**. In order to do that, the measure of the Gaussianity must be introduced.

The four measures of non-Gaussianity considered in the FastICA approach are the following:

- **Kurtosis** – defined in the excess form as:

$$\text{kurt}\,(y) = \frac{E\left[y^4\right]}{\left(E\left[y^2\right]\right)^2} - 3.$$

The normalized fourth moment of the Gaussian variable equals 3, meaning that the excess kurtosis different from zero points out non-Gaussianity. Generally, if the value is positive, we have platykurtic distribution (spiky pdf) while the negative value characterize leptokurtic distributions (flat pdf).

- **Approximated negentropy** – defined using nonlinear function G_i [6], Gaussian variable v and positive coefficient k_i:

$$J(y) \approx \sum k_i \left[E\{G_i(y)\} - E\{G_i(v)\}\right]^2.$$

The Gaussian variable has the largest entropy among all the random variable with the same variance. Therefore, the negentropy is the quantity having maximal value for non-Gaussian variable.

- **Mutual information**, defined in the context of the information theory.
- **Maximum likelihood estimation**, based on the log-likelihood function L [5],

where we ought to presume the statistics of the sources:

$$L = \sum_{t=1}^{T}\sum_{i=1}^{n}\log f_i\left(\mathbf{w}_i^T\mathbf{x}(t)\right) + T\log\left|\det\mathbf{W}\right| ,$$

with f_i being the density function of s_i. **W** is the estimated demixing (inverse mixing) matrix, while *T* represents the number of observations (length of vector **x**).

3. THE FASTICA ALGORITHM

The algorithm used in this analysis [7] is based on maximizing the negentropy as the measure of non-Gaussianity. In this case, the selection of the nonlinear function appears to be critical with respect to the nature of the observation data. Therefore, having 2D signals (images) as observations, in this article we compare three mostly used functions:

$$\begin{aligned} &\text{pow3:} \quad G_1(y) = \frac{y^4}{4} \\ &\text{tanh:} \quad G_2(y) = \frac{1}{a}\log\left(\cosh\left(ay\right)\right). \\ &\text{skew:} \quad G_3(y) = \frac{y^3}{3} \end{aligned} \tag{4}$$

The aim of the iterative routine is to maximize the function $F(\mathbf{w})$ given the constraint:

$$\begin{aligned} &\max F(\mathbf{w}) = \max\left[E\left\{G\left(\mathbf{w}^T\mathbf{x}\right)\right\} - E\left\{G\left(v\right)\right\}\right] \\ &\left\|\mathbf{w}\right\|^2 = 1 \end{aligned} . \tag{5}$$

The important remark here would be that in the frame of the formalism, introduced in this paragraph, **W** is rather associated to the mixing than to the demixing matrix, as it was the case in the preceding section.

After applying the Lagrange multiplier, the function $F(\mathbf{w})$ can be rewritten in terms of the first derivatives of the function given in the (4) (g_i) and the optimal value of **w** ($\mathbf{w_0}$):

$$F^*(\mathbf{w}) = E\left\{\mathbf{x}g(\mathbf{w}^T\mathbf{x})\right\} - E\left\{\mathbf{w}_0^T\mathbf{x}g(\mathbf{w}_0^T\mathbf{x})\right\}\mathbf{w}. \tag{6}$$

This way, the iteration is reduced to the Newton method used in order to find a vector **w** leading to the maximal negentropy. Given that we are actually dealing with the nonlinear system of equation, this has to be done using Jacobian matrix $J(\mathbf{w})$:

$$J(\mathbf{w}) = E\left\{g'(\mathbf{w}^T\mathbf{x})\right\}\mathbf{I} - E\left\{\mathbf{w}_0^T\mathbf{x}g(\mathbf{w}_0^T\mathbf{x})\right\}\mathbf{I}. \tag{7}$$

Finally, the principal step of the algorithm is the following:

$$\mathbf{w}_{n+1} = \mathbf{w}_n - J^{-1}(\mathbf{w}_n)F^*(\mathbf{w}_n) \ , \tag{8}$$

with $\mathbf{w}_{n+1}$ being normalized after each iteration. The convergence of the algorithm is verified by calculating a dot product of $\mathbf{w}_{n+1}$ and $\mathbf{w}_n$, which ought to be zero.

4. TENSORIAL DECOMPOSITIONS

Two most representative tensorial ICA methods, which are rather based on tensorial decompositions and do not have the character of iterative algorithm as the FastICA, are Forth-Order Blind Identification (FOBI) and Joint Approximate Diagonalization of Eigenmatrices (JADE).

FOBI is one of the first and most simple ICA methods [3]. The independent components are simply derived as eigenvectors of the kurtosis matrix, estimated using whitened set of observations (**x**):

$$K_I(\mathbf{x}) = E\left[\left(\mathbf{x}^T\mathbf{I}\mathbf{x}\right)\mathbf{x}\mathbf{x}\right]^T - 2\mathbf{I} - tr(\mathbf{I})\mathbf{I} = E\left[\left(\mathbf{x}\right)^2\mathbf{x}\mathbf{x}^T\right] - (n+2)\mathbf{I}. \tag{9}$$

The most notable drawback of this method would be the condition that all the sources must have quite distant kurtosis values, implicating the failure in case of having several mechanisms characterized with the same distribution.

JADE is a generalization of FOBI [4]. By considering covariance matrix to be a second order cumulant tensor, the kurtosis matrix (9) can be considered as a fourth order cumulant tensor of the identity matrix ($\mathbf{K}_I = \mathbf{F}(I)$). Replacing the identity matrix with a set of tuning matrices (eigenmatrices of the cumulant tensor: {M_1, ...M_p,}) results in a set of cumulants {K_{M1}, ...K_{Mp}}. The whitened de-mixing matrix **D** is estimated by jointly diagonalizing these matrices, which reduces to the maximization problem:

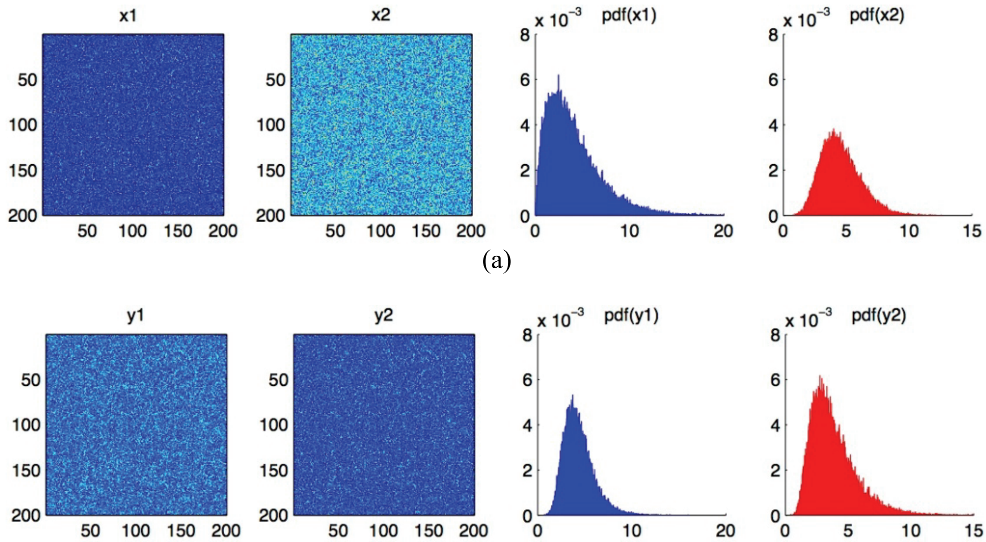
$$max\, J(\mathbf{D}) = max \sum\left(\left(\mathrm{diag}\left(\mathbf{D}\mathbf{K}_{Mp}\mathbf{D}\right)\right)\right)^2 \ , \tag{10}$$

where $\left\|diag(.)\right\|^2$ is the squared l_2 norm of the diagonal. Given that the maximization of the diagonal elements is equivalent to the minimization of the off-diagonal ones, the resulting de-mixing matrix **D** jointly diagonalize the set of cumulants. This algorithm overcomes the mentioned drawback of FOBI, but stays limited to low-dimensional problems.

5. THE COMPARATIVE ANALYSIS

In order to compare the performances of three introduced nonlinear functions in case of 2D observations, we have simulated two images characterized with the Gamma distribution:

$$\mathrm{pdf}\,(x\,|\,k,\theta)=\frac{x^{k-1}e^{-\frac{x}{\theta}}}{\theta^{k}\Gamma\left(k\right)}. \tag{11}$$



(a)

(b)

Fig. 1. The simulated sources (a) and their linear mixture (b): images and their statistics (pdf).

The defining shape k and scale θ parameters are, however, different (Fig. 1a). In case of the first one $k = \theta = 2$, while for the second one $k = 9$, $\theta = 0.5$.

Following step was mixing these two images in order to synthesize the observation data (Fig. 1b). The mixing is both linear (1) and stationary (meaning that the coefficients a_{ij} don't represent a function of space).

Finally, the introduced algorithm using all three introduced nonlinear functions (pow3, tanh and skew) is applied on the synthesized mixture (Fig. 2a, 2b, 2c). Probability density functions (pdf) of the sources are estimated by normalizing their histograms.

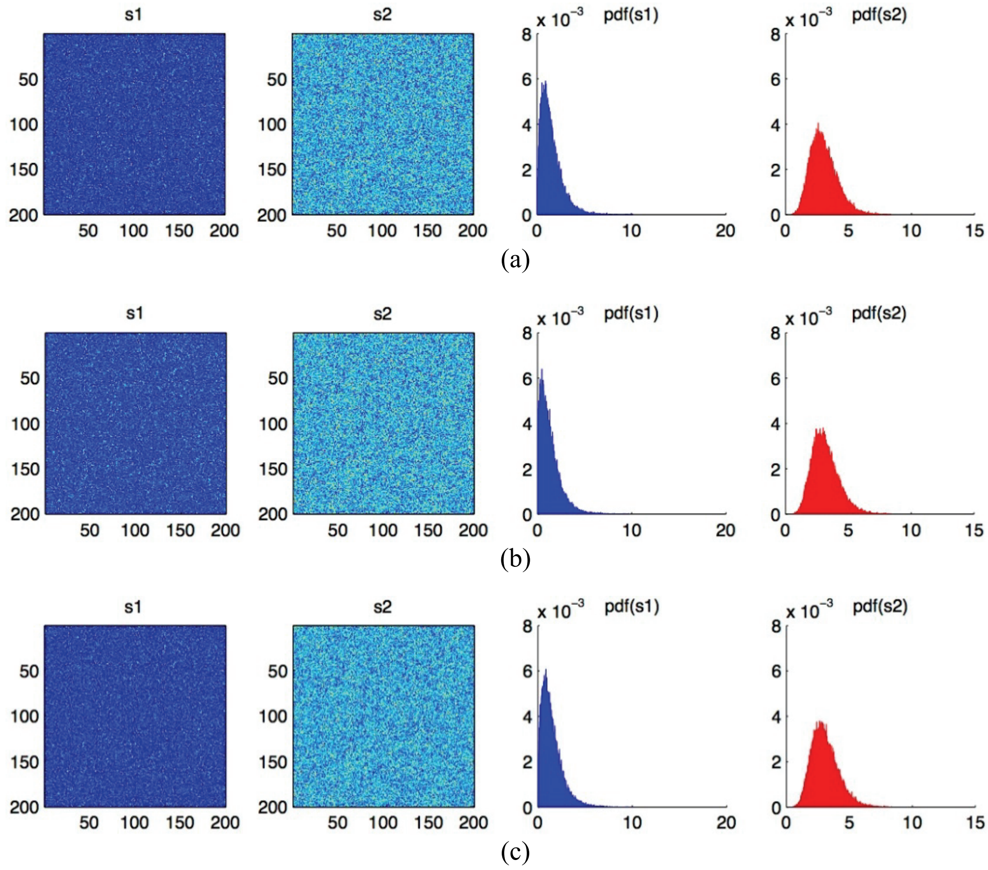


Fig. 2. The comparison of the sources estimated using different non-linearities of the FastICA algorithm, in terms of images and their statistics: (a) first criterion – *pow3*, (b) second criterion – *tanh*, (c) third criterion – skew.

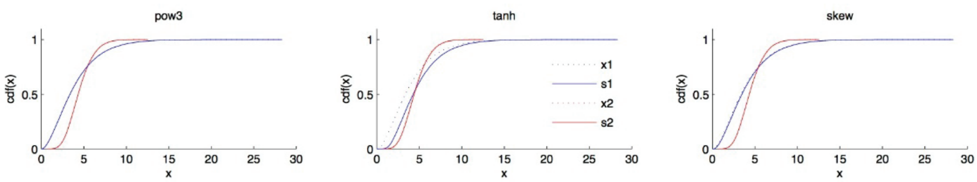


Fig. 3. The comparison between the simulated and the scaled estimated sources (FastICA) in terms of their statistics (cdf).

The evaluation of the different criteria is performed by deriving corresponding cumulative distribution functions (cdf) (Fig. 3) and calculating the mean square error (MSE) between the simulated (original) images and the estimated sources. Although the more conventional statistical tool would have been the estimation of the shape and scale parameters of the estimated sources, given the purpose (comparative analysis) and the simplicity of this method, it appears as a more suitable tool.

The obvious remark would be that the algorithm (in case of all three functions) fails to estimate the magnitude of a signal. This is, however, justified by the fact that the preprocessing and postprocessing of the observation data assume both (de)whitening and (de)centering. Therefore, before deriving the cdf, each pdf is linearly scaled in order to fit a range of values of original 2D signals. The results of the MSE comparison, provided in Table 1, unambiguously indicate the superiority of the *pow3* criterion.

Table I
MSE between the cdf of the simulated the scaled estimated sources (FastICA).

	pow3	*tanh*	*skew*
I	1.1672e-06	0.0037	5.6857e-05
II	1.6364e-08	5.4661e-05	1.5862e-07

The results obtained by means of tensorial decompositions are, as it can be seen in Fig. 4 and Table 2, not as nearly good as the ones obtained by FastICA (*pow3*) method. Such a big difference in terms of chosen quantitative evaluation can be considered as surprising, given the very low-dimensionality of the problem. Aside from that, the kurtosis value of the simulated sources are quite different.

Table II
MSE between the cdf of the simulated the scaled estimated sources (tensorial decomposition).

	FOBI	**JADE**
I	0.0537	0.0506
II	0.0231	0.0379

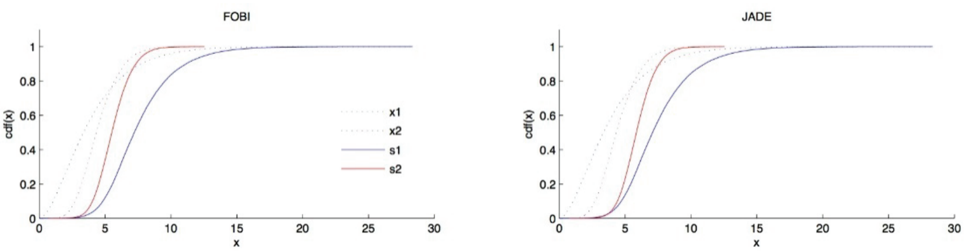


Fig. 4. The comparison between the simulated and the scaled estimated sources (tensorial decomposition) in terms of their statistics (cdf).

6. CONCLUSIONS

The FastICA algorithm is the prevalent ICA method, allowing a blind estimation of the independent sources. It is the iterative fixed-point algorithm, adopting a non-Gaussianity of the sources as an indicator of the independence. Depending on a measure of non-Gaussianity, there are several realizations of the algorithm. The one considered as most reliable and therefore used in this article – the approximated negentropy, depends strongly on a choice of a nonlinearity allowing the approximation. Therefore, we have analyzed the performances of the FastICA algorithm depending on a choice of the nonlinear function, illustrating the analysis using a 2D signal (image) as the observation. The comparative analysis was performed by considering the MSE between the cumulative distribution functions of the original (simulated) images and the scaled estimated sources. Eventually, we conclude that the quartic function (criterion *pow3*) appears to be the most suitable. The comparative analysis was additionally enhanced by introducing the most representative tensorial decompositions: FOBI and JADE. However, with respect to the same criteria, their performances were not comparable with the FastICA method in case of the considered observations. The important remark would be that the derived conclusions are strongly conditioned by the choice of the statistical parameters of the simulated images.

REFERENCES

[1] P. Comon and C. Juttern, *Handbook of Blind Source Separation: Independent Component Analysis and Applications,* Academic Press, 2010.

[2] A. Hyvarinen and E. Oja, "Independent Component Analysis: Algorithms and Applications, " *Neural Networks*, vol. 13, pp. 411-430, 2000.

[3] J. -F. Cardoso, “Source separation using higher-order moments,” *in Proc. ICASSP*, Glasgow, May1989, pp.2109–2112.

[4] J. -F. Cardoso and A. Souloumiac, “Blind beam forming for non-Gaussian signals,” IEE Proceedings-F, vol.140, no.6, pp.362–370, Dec. 1993.

[5] A. Hyvarinen, J. Karhunen and E. Oja, *Independent Component Analysis*, John Wiley and Sons, Inc., New York, NY, USA, 2001.

[6] D.-T. Pham, P. Garrat, and C. Jutten, “Separation of a mixture of independent sources through a maximum likelihood approach,” In *Proc. EUSIPCO,* pp. 771–774, 1992.

[7] A. Hyvärinen, “New approximations of differential entropy for independent component analysis and projection pursuit,” In *Advances in Neural Information Processing Systems,* volume 10, pp. 273–279. MIT Press, 1998.

[8] A. Hyvärinen, “Fast and robust fixed-point algorithms for independent component analysis,” *IEEE Transactions on Neural Networks,* 10(3), 626–634. 1999.