A Version of the FastICA Algorithm Based on the Secant Method Combined with Simple Iterations Method

Doru Constantin and Luminita State

University of Pitesti
Department of Computer Science
Address Str. Tg. din Vale, No.1, Romania
cdomanid@yahoo.com, radus@sunu.rnc.ro

Abstract. The work proposes a new algorithm for the estimation of the ICA model, an algorithm based on secant method and successive approximations. The first sections briefly present the standard FastICA algorithm based on the Newton method and a new version of the FastICA algorithm. The proposed algorithm to estimate the independent components combines the secant iterations with successive approximations technique. The final section presents the results of a comparative analysis experimentally derived conclusions concerning the performance of the proposed method. The tests were performed of several samples of signal files.

Keywords: Independent Component Analysis, Blind Source Separation, Numerical Method.

1 Introduction

An important problem arising in signal processing, mathematical statistical and neural networks is represented by the need of getting adequate representations of multidimensional data. The problem can be stated in terms of finding a function f such that the m dimensional transform defined by $s = f(x)$ possesses some desired properties, where x is a $m-$ dimensional random vector. Being given its computational simplicity, frequently the linear approach is attractive, that is the transform is

$$s = Wx \tag{1}$$

where W is a matrix to be optimally determined from the point of view of a pre-established criterion.

There are a long series of methods and principles already proposed in the literature for solving the problem of fitting an adequate linear transform for multidimensional data [1,4], as for instance, Principal Component Analysis (PCA), factor analysis, projection methods and Independent Component Analysis (ICA).

The aim of Independent Component Analysis is to determine a transform such that the components $s_i, i = \overline{1..n}$ becomes statistically independent, or at

A. Elmoataz et al. (Eds.): ICISP 2008, LNCS 5099, pp. 312–320, 2008.

least almost statistically independent. In order to find a suitable linear transform to assure that (1) $s_i, i = \overline{1..n}$ become 'nearly' statistically independent several methods have been developed so far. Some of them, as for instance Principal Component Analysis and factor analysis are second order approaches, that is they use exclusively the information contained by the covariance matrix of the random vector x, some of them, as for instance the projection methods and blind deconvolution are higher order methods that use an additional information to reduce the redundancies in the data. Independent Component Analysis has became one of the most promising approaches in this respect and, consists in the estimation of the generative model of the form $x = As$, where the $s = (s_1 s_2, \dots s_n)^T$ are supposed to be independent, and A is the mixing matrix $m \times n-$ dimensional of the model. The data model estimation in the framework of independent component analysis is stated in terms of a variational problem formulated on a given objective function.

The aim of the research reported in this paper is to introduce a new version of the FastICA algorithm; an algorithm that is based on secant iterations combined with successive approximations and to analyze the performances of the algorithm in signal applications.

2 Fixed-Point ICA Based on Iterations Method and Secant Method

2.1 The Standard FastICA Algorithm

In this section the ICA model and the standard FastICA algorithm are briefly exposed. The ICA model is state as $x = As$, where x is the observations vector and A is the mixing matrix of the original sources, s. The aim is to determine the sources, on the basis of x. One of the basic working assumption in estimation the ICA model is that the sources s are statistically independent and they have nongaussian distributions. This way the problem becomes to find the weighting matrix W (the demixing matrix), such that the transform $y = Wx$ gives suitable approximations of the independent sources.

In the following, the numerical estimation of the independent components is going to be obtained using the secant method combined with the successive approximation approaches, the variational problem being imposed on the negentropy taken as criterion function.

The negentropy is defined by:

$$I(y) = H(y_{gauss}) - H(y) \tag{2}$$

where $H(y) = -\int p_y(\eta) \log p_y(\eta) d\eta$ is the differential entropy of the random vector y.

Being given that the Gaussian repartition is of largest differential entropy in the class of the repartitions having the same covariance matrix, the maximization of the negentropy (2) gives the best estimation of the ICA model. Although this approaches is well founded from information point of view the direct use of the expression (2) is not computationally tractable, and some approximations are needed instead. We use the approximation introduced in (Hyvarinen, 98):

$$I(y) = [E\{G(y)\} - E\{G(\nu)\}]^2 \tag{3}$$

where G is an nonquadratic function, ν and y are Gaussian variables of zero mean and unit variance. Some of the most frequently used expressions of G are,

$$G_1(y) = \frac{1}{a_1}\log cosh(a_1 y);\ \ 1 \leq a_1 \leq 2,\ G_2(y) = -\exp(-\frac{y^2}{2});\ G_3(y) = \frac{y^4}{4}$$

Note that the expressions of their first order derivatives are given by: $g_1(y) = \tanh(a_1 y);\ g_2(y) = y\exp(-\frac{y^2}{2});\ g_3(y) = y^3$, respectively.

The variational problem can be formulated as a constraint optimization problem as follows,

$$\max F(w),\ \ \|w\|^2 = 1 \tag{4}$$

that is the objective function $F(w)$ has to be maximized on the unit sphere. In case the negentropy is taken as the objective function, we get,

$$F(w) = [E\{G(y)\} - E\{G(\nu)\}]^2 \tag{5}$$

where $y = w^T z$.

To solve the optimization problem from the (4) relation we write the Lagrange function using the Lagrange multiplications method:

$$L(w) = F(w) - \lambda(\|w\|^2 - 1) \tag{6}$$

The necessary conditions for the critical points are:

$$\frac{\partial L(w,\lambda)}{\partial w} = 0\ \ and\ \ \frac{\partial L(w,\lambda)}{\partial \lambda} = 0 \tag{7}$$

Applying (7) in the (6) relation we have:

$$\frac{\partial L(w,\lambda)}{\partial w} = \frac{\partial F(w)}{\partial w} - 2\lambda w = 0 \tag{8}$$

The expression of the gradient $\frac{\partial F(w)}{\partial w}$ is calculated like this:

$$\begin{aligned} \frac{\partial F(w)}{\partial w} &= \frac{\partial\,[E\{G(y)\} - E\{G(\nu)\}]^2}{\partial w} = \\ &= 2 * \left[E\left\{G(w^T z)\right\} - E\{G(\nu)\}\right] * \left[E\left\{z g(w^T z)\right\}\right] \end{aligned} \tag{9}$$

and $\gamma = \left[E\left\{G(w^T z)\right\} - E\left\{G(\nu)\right\}\right]$ is a constant, because ν is a Gaussian random variable. Thus we obtain:

$$\frac{\partial F(w)}{\partial w} = \gamma \left[E\left\{zg(w^T z)\right\}\right] \tag{10}$$

and it presents the gradient of the negentropy.

Replacing the (10) relation in (8) we obtain:

$$F^*(w) = E\left\{zg(w^T z)\right\} - \beta w = 0 \tag{11}$$

where β is a real constant, $\beta = E\left\{w_0^T zg(w_0^T z)\right\}$, where w_0 is the critical value of w.

The Newton method applied to (11) gives

$$w \leftarrow E\{zg\left(w^T z\right)\} - E\{g^{'}\left(w^T z\right)\}w \tag{12}$$

The weighting vectors being normalized, we arrive at the following approximation scheme,

1. $w \leftarrow E\{zg\left(w^T z\right)\} - E\{g^{'}\left(w^T z\right)\}w$
2. $w \leftarrow w/\left\|w\right\|$.

2.2 The FastICA Algorithm Based on the Simple Iterations Method and Secant Method

In this sections, we proposed a version of the FastICA algorithm based on the secant method combined with successive approximations. Experimentally supported conclusions concerning the performance of these methods are reported in the next section.

A secant method [3] in solving the equation (11) yields to the following iterative scheme:

1. select the initial approximation w_0 and a randomly generated value a.
2. apply the updating rule: $\Delta w = -\frac{F^*(w)}{F^*(w) - F(a)}(w - a)$.
3. if the convergence criterion does not hold then go to the step 2, else take the last value of the w^k as the approximative solution of (11).

The convergence criterion is $\left\|w^{k+1} - w^k\right\| < \varepsilon$, where $\varepsilon = 10^{-N}$ is a small real constant, $N \in N^*$ given.

Using the successive approximation method, the approximations sequence can be written as $w = \varphi(w)$ where $\varphi(w) = w - {}^1\!/_M * F^*(w)$ and M is the maximum value of $F^{*'}$.

The updating rule becomes: $w \leftarrow w - {}^1\!/_M * [E\left\{zg(w^T z)\right\} - \beta w]$, where g is previously defined.

The deducing of the ICA algorithm based on secant iterations combined with successive approximations (SSAM Algorithm). In this section we are developing the SSAM algoritm (Secant Method and Successive Approximations Method- based algorithm) that combines two numerical methods.

Aiming to obtain a better data representation and a better convergence order, we introduced the independent components analysis algorithm (SSAM algoritm) operating by: two iterations are performed by simple iterations method and the next iteration is calculated by secant method. The secant method is applying to the function F defined by $F(w) = w - F^*(w)$ using that two points previously obtained by iterations method. By example, let w_0 the initial iteration. Taken $f(w) = F^*(w)$, $w_1 = f(w_0)$ and $w_2 = f(w_1)$ then w_3 is given by:

$$w_3 = \frac{w_0 F(w_1) - w_1 F(w_0)}{F(w_1) - F(w_0)} \tag{13}$$

Substituting the values of the function F we obtain:

$$w_3 = \frac{w_0 w_2 - {w_1}^2}{w_0 - 2w_1 + w_2} \tag{14}$$

Starting with w_3, another two iterations are generated with successive approximations method and the next value is obtained with secant method and then we repeat this procedure. We note that in the sequence $w_0, w_1, w_2, w_3, w_4, w_5, w_6, w_7...$ every element $w_{3i},\ i = 0, 1, ...,$ may be written in terms of the w_{3i-3}, where $w_k = f(f(w_{k-2}))$. Denote this sequence by $w_0, w_1, w_2, ...,$ we obtain the following iterative scheme:

$$w_{k+3} = \frac{w_k f(f(w_k)) - f(w_k)^2}{w_k - 2f(w_k) + f(f(w_k))} \tag{15}$$

As compared to the Newton method, the convergence rate of the iterative method (16) is at least of order two. Thus we get an iterative scheme yielding to an improved convergence rate in estimating the independent components.

SSAM Algoritm - version of the standard algorithm for estimating the independent components

Step 1 : Center the data to mean.
Step 2 : Apply the whitening transform to data.
Step 3 : Select the number of independent components n and set counter $r \leftarrow 1$.
Step 4 : Select the initial guess of unit norm for w_r.
Step 5 : Apply the updating rules:

$$w_r^k = \frac{w_r^{k-3} f(f(w_r^{k-3})) - f(w_r^{k-3})^2}{w_r^{k-3} - 2f(w_r^{k-3}) + f(f(w_r^{k-3}))} \tag{16}$$

where w_r^0 is a initial value random generated and $f(w_r^k) = F^*(w_r^k)$, $w_r^{k-1} = f(w_r^{k-2}), w_r^{k-2} = w_r^{k-3}$.

Step 6 : Apply the orthogonalization transform: $w_r \leftarrow w_r - \sum_{j=1}^{r-1} (w_r^T w_j) w_j$

Step 7 : $w_r \leftarrow w_r / \|w_r\|$.

Step 8 : If w_r has not converged ($\left\| {w_r}^{k+1} - {w_r}^{k} \right\| > \varepsilon$, where ε is a small real constant), go back to step 5.

Step 9 : Set $r \leftarrow r + 1$. If $r \leq n$ then go to step 4.

3 Experimental Analysis

The assessment of the performances of the proposed algorithm for determining of the independent components is achieved in problems of signals recognition. We define an absolute mean sum error (AbsMSE) for comparing the accuracy of matching between original signals and restored signals. Then AbsMSE is defined as follows:

$$AbsMSE = \sum_{i=1}^{N} |s_i - s\_estimated_i| / N \tag{17}$$

where s_i and $s\_estimated_i$ represent the i-th pixel values for original and restored signals, respectively, and N is the total number of pixels.

All the presented tests comprise the recognition performances of the independent components using as an objective function the negentropy for which they used one at a time in the approximation the three functions adopted in the field [1].

In a comparative study the proposed method based on successive approximations combined with secant iterations has recognition performances of the original signals which are similar with the implemented methods, such as FastICA based on the secant method [2] or standard method.

3.1 Experimentally Derived Conclusions on the Performance of the Algorithm in the Case of the Mixtures of the Signals

Test I. We consider as observation data two signals which are mixed and recorded based on two independent components. In this first test, the original

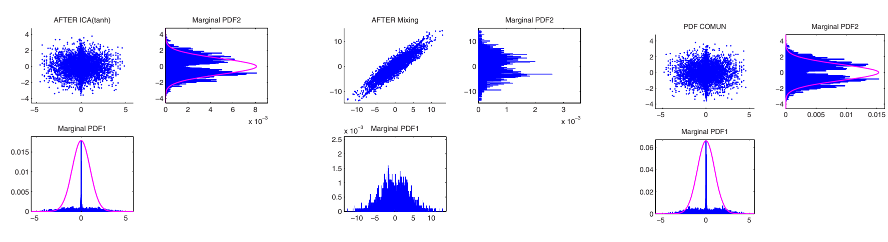

Fig. 1. Source Signals Discovered by the Algorithm (Left: 3 images), The Mixed Signals (Middle: 3 images) and Original Signals (Right: 3 images)

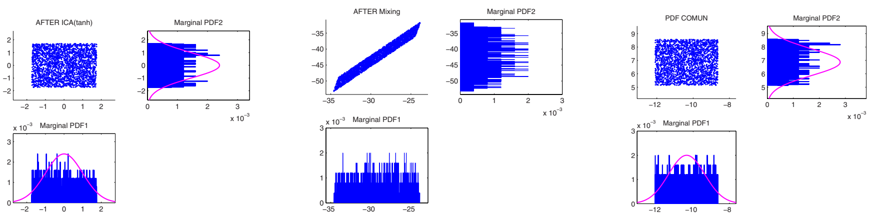

Fig. 2. Source Signals (uniform) Discovered by the Algorithm (Left: 3 images), The Mixed Signals (uniform) (Middle: 3 images) and Original Signals (uniform) (Right: 3 images)

sources are signals generated using the Matlab functions, and the results obtained after applying the SSAM algorithm based on successive approximations and secant iterations show a recognition performance similar to the standard FastICA method based on the Newton and to the method FastICA method based on the secant method [2].

The source signals discovered by the algorithm, the mixed signals and the source signals generated by Matlab subjected to the analysis procedure in independent components are represented in figure 1. In the respective figure we can notice the marginal densities corresponding to the two signals as well as the joint density which is common to the mixtures for the source signals discovered by the algorithm, for the mixed signals and for the source signals, respectively.

The results of the test regarding the appliance of the SSAM algorithm are given in table 1.

Test II. This test resembles the previously test with the difference that it uses, as original signals, the uniform distribution signals.

The figure 2 comprise the original source signals, the mixed signals and the source signals discovered by the algorithm for the uniform signals case.

The results obtained after the comparative study regarding the proposed method and other methods used in the estimation of the ICA model, are similar to the ones from the first test conform with table 1.

Table 1. AbsMSE of versions of the FastICA Algorithm for experimental tests

FastICA	Test I	Test I	Test I	Test II	Test II	Test II	TestIII	TestIII	TestIII
Basic Method	tanh (g_1)	exp (g_2)	kurt (g_3)	tanh (g_1)	exp (g_2)	kurt (g_3)	tanh (g_1)	exp (g_2)	kurt (g_3)
Newton	$1.0702*10^{-2}$	$1.0706*10^{-2}$	$1.0697*10^{-2}$	$2.3617*10^{-2}$	$2.3166*10^{-2}$	$2.4172*10^{-2}$	$4.493*10^{-3}$	$4.491*10^{-3}$	$5.813*10^{-3}$
Secant	$1.0716*10^{-2}$	$1.0682*10^{-2}$	$1.2793*10^{-2}$	$2.3271*10^{-2}$	$2.2754*10^{-2}$	$2.3181*10^{-2}$	$5.792*10^{-3}$	$5.741*10^{-3}$	$5.549*10^{-3}$
SSAM	$1.0706*10^{-2}$	$1.0695*10^{-2}$	$1.0699*10^{-2}$	$2.3295*10^{-2}$	$2.2981*10^{-2}$	$2.2996*10^{-2}$	$5.284*10^{-3}$	$5.352*10^{-3}$	$5.382*10^{-3}$

3.2 Experimentally Derived Conclusions on the Performance of the Algorithm in the Case of Image Signal

Test III. The achieved test refers to the capacity of the proposed algorithm of recognizing independent image faces from images of the mixed faces which can represent joint and superimposed faces as well as deteriorated images subjected to restoration.

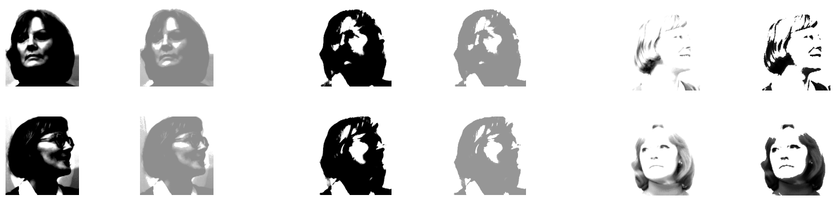

Fig. 3. Original Faces Discovered by the Algorithm (Left: 2 images), The Mixed Faces for Face Recognition (Middle: 2 images) and Original Faces for Face Recognition (Right: 2 images)

In this test we considered again the bidimensional case with two mixed images over which we apply the deterioration algorithm of the original faces. The original image faces, the mixed image faces and the image faces discovered by the proposed algorithm are in figure 3. Just as the previously examples, the obtained results offer good recognition performances of the original faces, showing also in this case a qualitative superiority with respect to convergence speed in the recognition (the results are presented in table 1) compared with other used methods and anterior specified.

4 Summary and Conclusions

In this article we developed an algorithm for estimating the independent components based on a iterative scheme that combines the successive approximations method to the secant method. We derived a suitable algorithm and supplied a comparative analysis of its recognition capacities against the previously developed algorithm. In order to derive conclusive remarks the tests were performed on different signal samples.

References

1. Hyvarinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. John Wiley & Sons, Chichester (2001)
2. Cho, Y.H., Hong, S.J., Park, S.M.: Face Recognition by Using Combined Method of Centroid Shift and Fixed-Point ICA Based on Secant Method. In: Proceedings of the Sixth International Conference on Intelligent Systems Design and Applications (ISDA 2006), IEEE Computer Society, Los Alamitos (2006)

3. State, L., Cocianu, C., Panayiotis, V., Constantin, D.: PRINCIPAL DIRECTIONS - BASED ALGORTHM FOR CLASSIFICATION TASKS. In: Proceedings of the Ninth International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC 2007), Timisoara, September 26-29, 2007, IEEE Computer Society, Los Alamitos (2007)
4. Stone, J.V.: Independent Component Analysis. The MIT Press, Cambridge (2004)
5. Bartlett, M.S., Movellan, J.R., Sejnowski, T.J.: Face Recognition by Independent Component Analysis. IEEE Transactions on Neural Networks 13 (2002)