14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

RISK-AVERTING COST FUNCTION FOR INDEPENDENT COMPONENT ANALYSIS IN SIGNALS WITH MULTIPLICATIVE NOISE

D. Blanco, D.P. Ruiz, M.C. Carrion

University of Granada
Department of Applied Physics
Science Faculty, Av. Severo Ochoa, s/n,
Granada, 18071, Spain

B. Mulgrew

University of Edinburgh
Institute for Digital Communications
Kings Buildings, Mayfield Road,
Edinburgh, EH9 3JL, United Kingdom

ABSTRACT

The FMICA is a method to extract the mixture of independent sources when they are contaminated with multiplicative noise, and notably improves the standard ICA methods in the presence of this kind of noise, although its results worsen when the level of noise increases. In this paper, whether this worsening is due to the existence of local minima or problems in the convergence of the statistical functions used is studied by a modification in the cost function that appears in FMICA. This new cost function has the property that, asymptotically, it does not present local minima, so it provides insights on the global convergence of the original cost function and it leads to the improvement of the behaviour of the FMICA for high noise levels, increasing the applicability of the method.

1. INTRODUCTION

Multiplicative noise appears in many situations, the most common being coherent images, such as synthetic aperture radar, laser or ultrasound images. In these signals, the information in a pixel is the result of the coherent sum of different scattered waves with fluctuating phases, which can be modelled as multiplicative noise [1]. The Independent Component Analysis (ICA) has been widely applied to almost all kinds of images (optical, stereo, video, multispectral, hyperspectral,...) in the solution of many problems (unsupervised classification, target detection, formation of thematic maps, denoising,...). However, there are quite few applications of ICA to coherent images, mainly because the model of linear mixture of independent sources that is assumed by the ICA methods is not satisfied in this case, due to the presence of this multiplicative noise. In [2], [3] and [4], ICA is applied to SAR images, and in [5] and [6] to ultrasound images. Although in [3] the noise is taken into account in the pre-processing step of PCA, in all these works the ICA methods are used as if there were no multiplicative noise, a fact that in theory and in practice reduces their applicability.

On the other hand, the Fourth-order Multiplicative ICA [7] (FMICA) is a method that tries to overcome the problem of the multiplicative noise, in order to extend the ICA approach to signals with this kind of noise, as coherent images.

As shown in [7] and [8], the FMICA presents very good results from low to medium noise levels, notably improving the standard ICA methods, and these results generally worsen fastly as the noise level rises above a certain point. Therefore, although the results of the FMICA are promising, to use it in a real problem it is necessary for the noise level in that situation to be within the application range of the FMICA. If the noise level is higher, the method cannot be used. Thus, it would be interesting to have an applicability range as broad as possible in the noise. The goal of this paper is to extend the application range of FMICA to higher noise levels, increasing the applicability of the method. This is done studying the global convergence of the FMICA method.

To overcome the limitations of the standard ICA methods in the multiplicative noise environment, firstly the Multiplicative ICA (MICA) method was designed in [9], using the second- and third-order statistics of the noisy signals. Although this method provides promising results, it also has serious limitations that make its application to real problems difficult. Specifically, it needs more than seven sources with up to one symmetrical (symmetric power density function (PDF) with respect to its mean). Due to the abundance of the symmetrical signals in the nature, this limitation is difficult to accept for an applicable method. To overcome this drawback, a new method, the FMICA method, was designed in [7] and [8]. This method adds fourth-order statistics to the available information of the signals, so that the method can find the mixture for any sources, symmetrical or non-symmetrical, and for any multiplicative noise. Also, the minimum number of sources necessary for the method to converge decreases from eight in MICA to just three in FMICA. Basically, FMICA builds a non-linear cost function that depends on the inverse of the mixture, using the signal statistics up to fourth-order, in such a way that the cost function presents its minimum value at the correct solution. The local convergence of FMICA is also studied, and it notably improves the standard ICA methods in this situation.

The results of FMICA worsen when the noise level increases, as is expected, and this worsening may be due to a lack of convergence in the involved statistical functions or to the existence of a local minimum, where the non-linear minimization method stops before it can reach the correct solution. If the cause is a lack of convergence, the behaviour can be improved only by increasing the number of data, but if the cause is the existence of local minima, it is possible to modify the FMICA to avoid these minima and thereby achive better results. This is what is proposed in this paper: to study the

This work was partially supported by the “Ministerio de Ciencia y Tecnología” of Spain under Project TEC2004-06096-C03-02/TCM.

global convergence of the FMICA method to improve its results. The paper is organized as follows. In Section II the model followed by the signals and their statistics up to the fourth order, are presented. In Section III the FMICA method is briefly outlined. In Section IV a risk-averting modification of the cost function of the FMICA is introduced to study qualitatively the global convergence of the method. This will provide a method, the risk-FMICA method, which can improve the results of the FMICA, for high noise levels. This is corroborated in Section V, where this improvement is shown by simulations. The paper finishes with the main conclusions.

2. MULTIPLICATIVE ICA MODEL

The MICA model [9] assumes that the recorded signals are a linear mixture of independent sources contaminated by multiplicative noise. This can be expressed as:

$$z_i = v_i x_i,\ i = 1,\ldots,N \text{ with } \mathbf{x} = \mathbf{A}\mathbf{s} \tag{1}$$

where $\mathbf{s} = [s_i,\ldots,s_N]^T$ is the vector of independent sources, with unit variance to eliminate the arbitrary scaling factor associated to all the ICA problems; $\mathbf{v} = [v_i,\ldots,v_N]^T$ is the multiplicative noise vector, which is formed by mutually independent random variables with mean one, and the noise-free data $\mathbf{x}$ and this noise are also independent [1]. $\mathbf{A}$ is the $N\times N$ mixing matrix and N is the number of signals. For simplicity, real signals and the same number of sources and signals are assumed in the paper, but these assumptions can be relaxed without loss of generality. Independence of the elements of $\mathbf{s}$ and $\mathbf{v}$, and mean one in these last elements are the only needed statistical conditions in the model.

The ICA methods use the statistics of the outputs of a linear transformation $\mathbf{u} = \mathbf{W}\mathbf{x}$ to find the inverse of the mixing matrix, called the *unmixing matrix,* since in the case $\mathbf{W} = \mathbf{A}^{-1}$ the outputs $\mathbf{u}$ are independent. As in the ICA case, we are interested in the statistical properties of $\mathbf{y} = \mathbf{B}\mathbf{z}$, when $\mathbf{B} = \mathbf{A}^{-1}$. In order to study them, we define the covariance, third- and fourth-order cumulants of the components of $\mathbf{y}$ (noted σ^y_{ijk}, γ^y_{ij} and κ^y_{ijkl} respectively) as:

$$\begin{aligned}
\sigma^y_{ij} =& \mathscr{E}\{(y_i-\mu^y_i)(y_j-\mu^y_j)\}\\
\gamma^y_{ijk} =& \mathscr{E}\{(y_i-\mu^y_i)(y_j-\mu^y_j)(y_k-\mu^y_k)\}\\
\kappa^y_{ijkl} =& \mathscr{E}\{(y_i-\mu^y_i)(y_j-\mu^y_j)(y_k-\mu^y_k)(y_l-\mu^y_l)\}\\
&-\sigma^y_{ij}\sigma^y_{kl}-\sigma^y_{ik}\sigma^y_{jl}-\sigma^y_{il}\sigma^y_{jk}
\end{aligned} \tag{2}$$

where μ^y_i is the mean of the component y_i and $\mathscr{E}\{\cdot\}$ is the expectation operator, and all the indices go from 1 to N, as all the indices will do in the rest of the paper, unless otherwise stated. These functions can be expressed as function of the mixing matrix and of the statistical properties of the noise and of the sources, taking into account that it is assumed that $\mathbf{BA} = \mathbf{I}$. To do so, we define the following parameters:

$$\{\gamma^s_i, \kappa^s_i, \eta_i, \omega_{ij}, \rho_i, \phi_i\} \tag{3}$$

where γ^s_i and κ^s_i are the skewness and kurtosis of s_i and the rest of parameters are

$$\begin{aligned}
\eta_i = \sqrt{\sigma^v_i}\mu^x_i;\ \ \omega_{ij} = \sqrt{\sigma^v_i}A_{ij};\ \ \rho_i = \gamma^v_i/(\sigma^v_i)^{3/2};\\
\phi_i = \kappa^v_i(\kappa^x_{iiii}+4\mu^x_i\gamma^x_{iii}+6(\mu^x_i)^2\sigma^x_{ii}+3(\sigma^x_{ii})^2+(\mu^x_i)^4)
\end{aligned} \tag{4}$$

with σ^v_i, γ^v_i and κ^v_i the variance, skewness and kurtosis of v_i; and μ^x_i, σ^x_{ii}, γ^x_{iii} and κ^x_{iiii} the mean, variance, skewness and kurtosis of x_i. With these parameters, γ^y_{ij}, σ^y_{ijk} and κ^y_{ijkl} result:

$$\begin{aligned}
\sigma^y_{ij} =& \delta_{ij} + \sum_r B_{ir}B_{jr}(\sum_s \omega^2_{rs} + \eta^2_r)\\
\gamma^y_{ijk} =& \gamma^s_i\delta_{ijk} + \lfloor \sum_r B_{ir}B_{jr}(B_{kr}\beta_r/3 + \omega^2_{ri}\gamma^s_r + 2\omega_{ri}\eta_r)\rfloor_{ijk}\\
\kappa^y_{ijkl} =& \kappa^s_i\delta_{ijkl} + \lfloor \sum_r B_{ir}B_{jr}B_{kr}(B_{lr}\phi_r/4 + \psi_{ri})\rfloor_{ijkl}\\
&+\lceil \sum_r B_{ir}B_{jr}(\chi_{rkl} + \sum_s B_{ks}B_{ls}\xi_{rs}/2)\rceil_{ijkl}
\end{aligned} \tag{5}$$

with

$$\begin{aligned}
\beta_i =& \rho_i(\sum_r \omega^2_{ir}(\omega_{ir}\gamma^s_r + 3\eta_i) + \eta^3_i)\\
\psi_{ij} =& \rho_i\omega_{ij}(\kappa^s_j\omega^2_{ij} + 3(\eta_i\gamma^s_j\omega_{ij} + \eta^2_i + \sum_r \omega^2_{ir}))\\
\chi_{ijk} =& \omega_{ij}(\delta_{jk}(\omega_{ij}\kappa^s_j + 2\eta_i\gamma^s_j) + 2\omega_{ik})\\
\xi_{ij} =& \sum_r \omega_{ir}\omega_{jr}\big(\omega_{ir}\omega_{jr}\kappa^s_r + 2(\gamma^2_r(\eta_i\omega_{jr} + \eta_j\omega_{ir})\\
&+ 2\eta_i\eta_j + \sum_s \omega_{is}\omega_{js})\big)
\end{aligned} \tag{6}$$

all the sums go from 1 to N, as all will do in the rest of the paper unless otherwise stated; and for any arbitrary functions f_{ijk}, dependent on the indices i, j and k, and g_{ijkl}, dependent on the indices i, j, k and l, it is defined

$$\begin{aligned}
\lfloor f_{ijk}\rfloor_{ijk} \triangleq& f_{ijk} + f_{jki} + f_{kij}\\
\lfloor g_{ijkl}\rfloor_{ijkl} \triangleq& g_{ijkl} + g_{jkli} + g_{klij} + g_{lijk}\\
\lceil g_{ijkl}\rceil_{ijkl} \triangleq& g_{ijkl} + g_{ikjl} + g_{iljk} + g_{jkil} + g_{jlik} + g_{klij}
\end{aligned} \tag{7}$$

The last definitions have been made to avoid writing too many terms in (5), which appear because the permutation symmetries of the cumulants. A more detailed deduction of the previous expressions can be found in [7].

It is important to point out that, as the problem is blind, the parameters $\{\eta_i, \omega_{ij}, \rho_i, \phi_i, \gamma^s_i, \kappa^s_i\}$ are unknowns of the problem.

3. FOURTH-ORDER MULTIPLICATIVE ICA METHOD

ICA searches for the linear transformation for which the outputs are as independent as possible. If the data satisfy the ICA model, the solution is the inverse of the mixing matrix. In the case of MICA model, the outputs of the inverse of the mixing matrix are not independent but they possess a specific statistical structure that can be used to find this matrix. This is exactly how the FMICA finds the solution. This structure is explicitly shown in (5), which will be satisfied if the unmixing matrix $\mathbf{B}$ is the inverse of the mixing matrix, and the rest of the parameters in (3) take their theoretical values.

On the other hand, the covariance, third- and fourth-order cumulants of the output $\mathbf{y}$ can be estimated from the noisy data, for any matrix $\mathbf{B}$. If these three estimated functions are noted as $\hat{\sigma}^y_{ij}$, $\hat{\gamma}^y_{ijk}$ and $\hat{\kappa}^y_{ijkl}$, they can be obtained from the covariance, third- and fourth-order cumulants of the noisy data $\mathbf{z}$, which are noted as $\hat{\sigma}^z_{ij}$, $\hat{\gamma}^z_{ijk}$ and $\hat{\kappa}^z_{ijkl}$, respectively. The explicit relation is straightforward to obtain, taking into account the relation $\mathbf{y} = \mathbf{Bz}$, and it is:

$$\hat{\sigma}^y_{ij} = \sum_{mn} B_{im}B_{jn}\hat{\sigma}^z_{mn}\ ;\ \hat{\gamma}^y_{ijk} = \sum_{mnp} B_{im}B_{jn}B_{kp}\hat{\gamma}^z_{mnp}$$
$$\hat{\kappa}^y_{ijkl} = \sum_{mnpq} B_{im}B_{jn}B_{kp}B_{lq}\hat{\kappa}^z_{mnpq} \tag{8}$$

It can be seen that the functions $\hat{\sigma}^y_{ij}$, $\hat{\gamma}^y_{ijk}$ and $\hat{\kappa}^y_{ijkl}$ depend only on the unmixing matrix $\mathbf{B}$, while the functions σ^y_{ij}, γ^y_{ijk} and κ^y_{ijkl} in (5) depend on the unmixing matrix and also on the set of parameters in (3), which, as the problem is blind, are unknown.

Hence, the estimated functions (8) will be equal to the functions (5) when $\mathbf{B} = \mathbf{A}^{-1}$ and the rest of the parameters in (3), take their theoretical values, which will be called the correct solution. To measure how well the structure is reproduced for a specific matrix $\mathbf{B}$ and a specific set of parameters (3), a cost function $J = J(B_{ij}, \eta_i, \omega_{ij}, \rho_i, \phi_i, \gamma^s_i, \kappa^s_i)$ can be built as:

$$J = \sum_{ij}(\mu^x_i \sum_k \omega_{ik}B_{kj} - \eta_i\delta_{ij})^2 + \sum_{i\geq j}\left(\sigma^y_{ij} - \hat{\sigma}^y_{ij}\right)^2$$
$$+ \sum_{i\geq j\geq k}\left(\gamma^y_{ijk} - \hat{\gamma}^y_{ijk}\right)^2 + \sum_{i\geq j\geq k\geq l}\left(\kappa^y_{ijkl} - \hat{\kappa}^y_{ijkl}\right)^2 \tag{9}$$

with the definitions in (5) and (8). The first sum in the cost function is included to take into account the theoretical relation between the parameters B_{ij}, ω_{ij} and η_i. This function is formed by $N_2 = N(N+1)/2((N+2)/3(1+(N+3)/4)) + N^2$ terms, is function on the $N_1 = N(2N+5)$ parameters, and will be zero at the correct solution.

Thus, the problem is reduced to find the value of the parameters $\{\phi_i, \eta_i, \omega_{ij}, \rho_i, \gamma^s_i, \kappa^s_i, B_{ij}\}_{i,j=1,\ldots,N}$ that minimizes the cost function (9), which means a problem of non-linear minimization of J. Although the non-linear minimization method that is most used in the ICA literature is the steepest descendant method using the natural gradient, it is necessary to resort to another minimization method here. The natural gradient of the cost function is not easy to establish, since the set of parameters is not a multiplicative group, and the standard steepest descendant method is too slow. In the FMICA method, the minimization is accomplished using the quasi-Newton method called BFGS (Broyden-Fletcher-Goldfard-Shanno). In this method the set of parameters, which are grouped in a $N_1 \times 1$ vector $\mathbf{b}$, is updated in the step k as:

$$\mathbf{b}(k+1) = \mathbf{b}(k) - \mu(k)\mathbf{H}(k)\nabla_k J \tag{10}$$

where $\mu(k)$ is the learning ratio in the step k, $\nabla_k J$ is the gradient of J in the step k, and the matrix $\mathbf{H}(k)$ is an estimation of the inverse of the Hessian in the step k, which is forced to be positive definite and symmetrical, and is obtained using the value of the parameters and the gradient of J in the steps k and $k-1$ as

$$\mathbf{H}(k) = (\mathbf{I} - u_k\mathbf{p}_k\mathbf{q}^T_k)\mathbf{H}(k-1)(\mathbf{I} - u_k\mathbf{q}_k\mathbf{p}^T_k) + u_k\mathbf{p}_k\mathbf{p}^T_k \tag{11}$$

with $\mathbf{p}_k = \mathbf{b}(k) - \mathbf{b}(k-1)$, $\mathbf{q}_k = \nabla_k J - \nabla_{k-1} J$ and $u_k = 1/(\mathbf{q}^T_k\mathbf{p}_k)$.

The BFGS method keeps a great part of the speed of the Newton method, but $\mathbf{H}$ is never singular and it is not necessary to compute second derivates, as in the Newton method. The readers are referred to the literature in non-linear optimization for details about BFGS method, for example [10].

Only the gradient and the starting point are necessary for the FMICA method to be completed. The first is omitted due to lack of space, but can be found in [8]. As in most non-linear optimization methods, adequate starting points are necessary for the method to converge. Although the initial value for most of the parameters can be determined by assuming a not too high noise level, this is not possible for the initial values of the unmixing matrix, $\mathbf{B}(0)$. To do so, it is necessary to resort to a standard ICA method (the FastICA in this paper), so its solution is taken as the initial value for $\mathbf{B}$ in the FMICA method. Once this is done, $\tilde{\mathbf{y}} = \mathbf{B}(0)\mathbf{z}$ it is defined, and with it the starting point results:

$$\begin{aligned} \mathbf{B}(0) =& \text{solution of FastICA method;} \\ \phi_i(0) =& \rho_i(0) = 0;\ \ \gamma^s_i(0) = \hat{\gamma}^{\tilde{y}}_{iii};\ \ \kappa^s_i(0) = \hat{\kappa}^{\tilde{y}}_{iiii}; \\ \eta_i(0) =& \sqrt{\hat{\sigma}^v_i}\hat{\mu}^z_i \ \text{ and } \ \omega_{ij}(0) = \sqrt{\hat{\sigma}^v_i}[\mathbf{B}(0)^{-1}]_{ij} \end{aligned} \tag{12}$$

where $\hat{\sigma}^v_i$ is obtained as $\hat{\sigma}^v_i = \frac{\hat{\sigma}^z_{ii} - \sum_k [\mathbf{B}(0)^{-1}]_{ik}}{\sum_k [\mathbf{B}(0)^{-1}]^2_{ik} + (\hat{\mu}^z_i)^2}$. These last estimatated values are inexact, due to the errors in $\mathbf{B}(0)$, but this is not a problem since the values of $\hat{\sigma}^v_i$ are inside other parameters (η_i and ω_{ij}) that are updated in the minimization process. In fact, a initial value of zero for all the $\hat{\sigma}^v_i$ provides, in most of the cases, a good initialization, such as the method converges, but the convergence is slower. The FMICA method consists in the minimization of (9) with the update formula (10), the gradient of J and the initialization (12).

4. RISK-AVERTING COST FUNCTION

Although the FMICA method provides much better results than does the standard ICA methods, these results worsen as the noise level increases. This is expected, because the correct solution is a minimum of the cost function only asymptotically, so for a finite number of data the convergence will not be perfect and the correct solution is not exactly a minimum. The higher the noise level the worse the approximation, and, as the noise is multiplicative, the worsening increases faster with the noise level than in the case of additive noise. The performance of a method is characterized by the distance of the global transformation $\mathbf{C} = \mathbf{BA}$ to the identity or any permutated sign-shifted version of it. A bad result, i.e. a large distance between the global transformation and the identity, may also be due to a local minimum where the method has stopped. If the first of the two causes is responsible for a bad result, the only way to improve is to increase the number of data; but if the existence of a local minimum is the cause of a bad result, it is possible to design a way to avoid the such a minimum.

To distinguish between the two causes requires an analytical study of when and where local minima exit in the cost function (9). This means to study the global convergence of the cost function, which is a very complex mathematical problem. A non-exhaustive alternative to this analytical study is to modify the cost function (9) in order to ensure that the new cost function does not present local minima. This is possible, at least asymptotically, through the introduction of a risk-averting cost function that tends to a convex cost function when a risk-sensitivity index increases to infinity. This is always possible for any cost function [11], and in this case the risk-averting cost function is:

$$G = \sum_{ij} \exp\left(\tau(\mu_i^x \sum_k \omega_{ik} B_{kj} - \eta_i \delta_{ij})^2\right) + \sum_{i\geq j} \exp\left(\tau(\sigma_{ij}^y - \hat{\sigma}_{ij}^y)^2\right) + \sum_{i\geq j\geq k} \exp\left(\tau(\gamma_{ijk}^y - \hat{\gamma}_{ijk}^y)^2\right) + \sum_{i\geq j\geq k\geq l} \exp\left(\tau(\kappa_{ijkl}^y - \hat{\kappa}_{ijkl}^y)^2\right) \quad (13)$$

where the index τ is called the risk-sensitivity index. It can be proven that for a cost function as (13), when the risk-sensitivity index τ increases to infinity, the region where G is convex expands monotonically and tends to the whole space, except for the possible intersection of a finite number of manifolds of dimension smaller than the dimension of the parameter space [11]. So, except for the rare case where the manifolds all intersect at the same point (more difficult as the number N_2 increases), it will always be possible, at least asymptotically, to eliminate any local minimum. Of course, in practice it is not possible to increase τ as much as desired, since the minimization methods diverge if τ is increased above a certain limit. Thus, in practice, it would be possible to eliminate some local minima, but not others, and the practical utility of the G can be tested only by simulations.

It is straightforward to see that the correct solution is a minimum of G, and therefore it is possible to repeat the same steps as for the FMICA method. The BFGS is a non-linear minimization method that can be applied to any cost function, so it can be used to find the minimum in G, with the same initialization as in (12). To complete the method, only the gradient of G is needed. This can be easily derived from the gradient of J. To do so, we express the cost function J as $J = \sum_{p=1}^{N_2} (f_p)^2$, where the explicit expression of f_p for $p = 1, \ldots, N_2$ can be determined comparing the previous expression with (9). After this, the risk-averting cost function can be written as $G = \sum_{p=1}^{N_2} \exp\left(\tau(f_p)^2\right)$. Then, the components of the gradient of both cost functions can be related, taking into account that:

$$\frac{\partial J}{\partial b_i} = 2\sum_{p=1}^{N_2} f_p \frac{\partial f_p}{\partial b_i} \text{ and } \frac{\partial G}{\partial b_i} = 2\tau \sum_{p=1}^{N_2} \exp\left(\tau(f_p)^2\right) f_p \frac{\partial f_p}{\partial b_i} \quad (14)$$

The Risk-FMICA method will consist of the minimization of (13), using the update formula (10), with the gradient of G (which can be obtained from the gradient of J using (14)) and the initialization (12). If the Risk-FMICA method obtains better results than the FMICA, it means that in the second case the method stopped at a local minimum, and the method using G is preferred. If there is no difference, it would be not possible to distinguish wheter the methods stopped at a local minimum if the distance of the global transformation to the identity is due to a poor convergence in the statistical functions. In the next chapter, the utility of the risk-averting cost function will be tested by simulations.

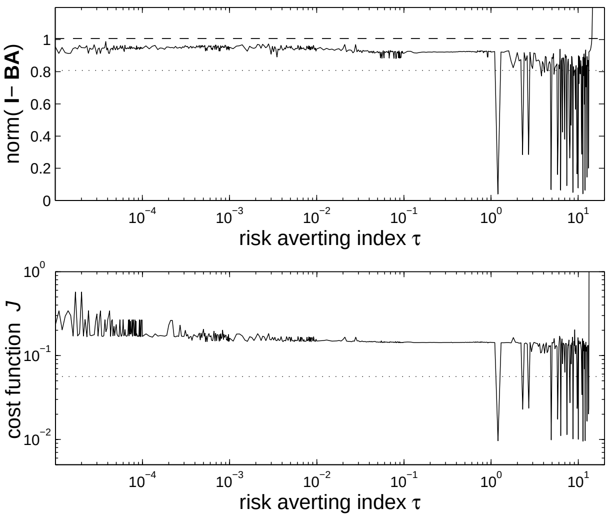

Fig. 1. Distance of the global transformation to the identity (upper graph) and value of the cost function J (lower graph) as function of τ, for four signals with a noise level of $st = 0.09$. Solid line = Risk-FMICA, dotted line = FMICA and dashed line = FastICA.

5. RESULTS

In this section, the behaviour of the Risk-FMICA method is compared with the FMICA method in [7] and with the FastICA method [12].

In the simulations, the mixture of four sources contaminated with Rayleigh multiplicative noise is studied. The sources are obtained through the exponentiation of uniform signals, and therefore their PDFs are truncated logarithmic. We have selected this kind of sources and noise because both are non-symmetrical and it has been shown in [8] that in this case the FMICA presents its worst results. The number of data per signal is 100,000.

The sources are mixed with a 4×4 matrix, generated randomly, and the mixed signals are contaminated with multiplicative Rayleigh noise of standard deviation st. The cost function J is built and then minimized using the BFGS algorithm. When this is done, the results of FMICA are much better than the results of the FastICA method up to a noise level of $st = 0.09$, as it can be seen in Figure 2. This is the typical behaviour of the FMICA method, i.e. the method notably improves the results of the FastICA method until a noise level where the improvement is reduced. There is always a rather wide region in st where the FMICA results are much better than those of the FastICA, which will be the applicability region, and in this example corresponds with the region $st \in [0, 0.8]$. The goal of this paper is to extend this applica-

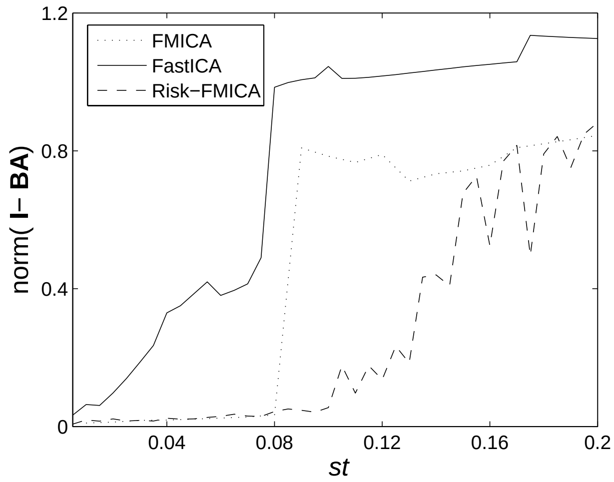

Fig. 2. Distance of the global transformation to the identity as function of the *st* for four signals. Solid line = Risk-FMICA, dotted line = FMICA and dashed line = FastICA.

bility region to higher *st* values.

For the same signals, keeping $st = 0.09$, the risk-averting cost function G is built for a different risk-sensitivity index τ. That cost function is minimized with the BFGS method and the distance of the resulting global transformation to the identity is plotted in the upper graph in Figure 1, for different values of τ (solid line). The dotted line corresponds with the value of the distance for the FMICA method and the dashed line corresponds to the result for the FastICA method (both independent of τ). Also, the value of the cost function J for the parameters derived as the result of the minimization of G are plotted in the lower graph (solid line), where the dotted line represents with the value of J for the solution of the FMICA method. This figure shows how it is possible to attain a global transformation very close to the identity for some τ values, although this does not happen for a uniform region in τ, but just for specific values. Fortunately, the τ values with better results are also the ones with smaller values in the cost function J (the cost function built with the parameters resulting from the minimization on G). Hence, it is possible to find the τ values that improve the results of the BFGS method by finding the τ values that produces smaller values in J.

This procedure has been repeated for different sources, noises and mixtures, with the result being that only between τ equal to one and τ near ten, does Risk-FMICA method provide better results than the FMICA method. Values of τ much greater than 10 cause the BFGS to diverge. Thus, in practice, the search for the optimum value of τ is limited to the previous region. Although this could be considered as a time-consuming approach, the greater the value of τ the faster the BFGS method converges. For the results presented in Figure 2, the search has been made for values of $\tau_i = 1 + (i-1)0.04$ for $i = 1, \ldots, 125$, which corresponds to a grid of the region $\tau \in [1, 5]$. In practice, it is faster to obtain the 125 minimizing G than only one minimization of J, for high noise levels and τ in the previous region (it takes just few minutes in a standard PC). The results of the comparison of FastICA, FMICA and Risk-FMICA are shown in Figure 2.

It can be shown how it is possible to improve the results of the FMICA method through the introduction of a risk-averting cost function, extending in this way the applicability region of the methods.

This study has been repeated for different PDFs for the sources and the noise, different number of sources and number of data per signal, and the behaviour is similar to that shown here, although they are omited due to lack of space.

6. CONCLUSIONS

The introduction of a risk-averting cost function has facilitated the study of global convergence of the FMICA. Although the study is not exhaustive, through an easy analysis it is possible to design a method to improve the FMICA method using a risk-averting cost function. The method so designed, the Risk-FMICA, improves the result of the FMICA method in the region where this fails, extending the applicability region that the FMICA possessed.

7. REFERENCES

[1] S. Chitroub and B. Sansal, "Statistical characterisation and modelling of SAR images," *Signal Processing*, vol. 82, pp. 69–92, 2002.

[2] X. Zhang and C. H. Chen, "A new independent component analyis (ICA) method and its application to SAR images," in *Proc. of NNSP*, 2001, pp. 283–292.

[3] S. Chitroub and B. Sansal, "Unsupervised learning rules for POLISAR images analysis," in *Proc. of NNSP*, 2002, pp. 567–576.

[4] J. Karvonen and M. Similä, "Independent component analysis for ice SAR image classification," in *Proc. of IGARSS*, 2001, pp. 1255–1257.

[5] H. Neemuchwala, A. Hero, and P. Carson, "Feature coincidence trees for regisration of ultrasound breast images," in *Proc. of IEEE Int. Conf. on Image Proc.*, 2001.

[6] H. Neemuchwala, A. Hero, and P. Carson, "Image registration using entropic graph-mathching criteria," in *Proc. of Asilomar Conf. on Signal and Sistem*, 2002.

[7] D. Blanco, B. Mulgrew, D. P. Ruiz, and M. C. Carrion, "ICA in signals with multiplicative noise using fourth-order statistic," in *Proc. of EUSIPCO*, 2005.

[8] D. Blanco, B. Mulgrew, D. P. Ruiz, and M. C. Carrion, "A fourth-order statistic ICA method in signals with multiplicative noise," *Sent to IEEE Trans. Signal Processing*.

[9] D. Blanco and B. Mulgrew, "ICA in signals with multiplicative noise," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 2648–2657, 2005.

[10] J. Nocedal and S. J. Wright, *Numerical optimization*, Springer, 2000.

[11] J. T. Lo, "Minimization through convexitization in training neural networks," in *Proc. of Internation Joir Conference on Neural Networks*, 2002.

[12] A. Hayvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation*, vol. 9, pp. 1483–1492, 1997.