Cyclic maximization of non-Gaussianity for blind signal extraction of complex-valued sources

Iván Durán-Díaz *, Sergio Cruces, María Auxiliadora Sarmiento-Vega, Pablo Aguilera-Bonet

Escuela Técnica Superior de Ingenieros, Departamento de Teorı́a de la Senal y Comunicaciones, University of Seville, Camino de los descubrimientos s/n, 41092 Sevilla, Spain

A R T I C L E   I N F O

Article history:
Received 17 September 2010
Received in revised form
19 January 2011
Accepted 24 March 2011
Communicated by E.W. Lang
Available online 18 May 2011

Keywords:
Independent component analysis
Blind signal extraction
Negentropy criterion

A B S T R A C T

This article presents a new algorithm for the blind extraction of communications sources (complex-valued sources) through the maximization of negentropy approximations based on nonlinearities. A criterion based on the square modulus of a nonlinearity of the output is used. We decouple the arguments of the criterion so that the algorithm maximizes it cyclically with respect to each argument by means of the Cauchy–Schwarz inequality. A proof of the ascent of the objective function after each iteration is also provided. Numerical simulations corroborate the good performance of the proposed algorithm in comparison with the existing methods.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

During the last years several algorithms have been proposed for blind signal extraction and blind signal separation of complex sources [1–8]. Most of the higher-order-cumulant-based contrast functions have been easily extended from the real to the complex domain [1,4–9,11]. However, in general, these contrast functions are very sensitive to outliers [6]. Alternatively, nonlinear functions that try to approximate either the probability density function of the sources (like in the ML and Infomax criteria) or the output’s negentropy can be used as contrast functions. However, as noted by Adali et al. [3], while contrast-based algorithms can be easily extended from the real to the complex case, the extension for nonlinearities-based contrasts is not straightforward, so this case is not so widely discussed as that of cumulant-based contrast for complex sources. Nevertheless a few contrast functions and algorithms have been proposed in the context of nonlinearities of the outputs. For example, the extension of the well-known negentropy-based FastICA to the complex case by Bingham and Hyvärinen [2], a complex INFOMAX ICA algorithm for frequency domain by Anemuller [12], or several works by Adali et al. [6,13,14]. In [14] an extension of the complex-FastICA algorithm was proposed in order to deal with noncircular sources by means of the insertion of a term given by the pseudo-covariance matrix of the data into the fixed-point algorithm, but still using the contrast proposed by Bingham and Hyvärinen for FastICA, which only used the modulus information. In [6], Novey and Adali proposed a new negentropy-based contrast exploiting the information from both the modulus and phase of the output, and also derived a quasi-Newton algorithm able to work with circular and noncircular sources, generalizing thus their work in [14]. Adali et al. summarized in [13] three nonlinearities-based algorithms in the context of non-Gaussianity or negentropy (the one proposed in [6]), maximum likelihood (the one proposed in [15]) and minimization of mutual information, respectively.

One of the applications of BSS techniques for complex-valued sources is the blind detection of communications signals. In digital communications systems two kinds of interferences appear: co-channel interference (CCI) and inter-symbol interference (ISI). The first one is due to the fact that users share the same time-frequency-code slots while the second one is caused by multipath channels [16]. Traditionally interference has been removed by means of the supervised identification and equalization of the channel by using pilot symbols (training sequences). Blind identification and equalization of the channel avoid training sequences, which implies a more efficient use of the channel bandwidth [17–20]. The problem of removing the interference at the receiver can be seen as a problem of BSS in which the mixing system depends on the impulsive response of the channels [21,22]. Previous works that have applied the concepts of BSS to the blind detection of users in digital communications systems can be found in [5,23–25].

In this work we propose a new algorithm for the blind extraction of communications signals. Starting from a contrast
based on the kind of nonlinearities proposed in [6], we define a new criterion based on three independent candidates for the extraction vector. We suggested a new algorithm for the cyclic maximization of the output’s non-Gaussianity by means of fast and simple iterations based on the Cauchy–Schwarz inequality. In comparison with some of the existing approaches, like the one proposed in [6] and complex-FastICA [2], the proposed algorithm exhibits an improved performance for communications signals, QPSK or 8-PSK. Unlike in [2,6], we also give a proof of the monotonic ascent of the contrast function for the proposed algorithm.

2. Signal model and notation

We consider the existence of N independent complex sources \( s(k) = [s_1(k), \ldots, s_N(k)] \) of zero mean and unit variance that are mixed by a full-column rank mixing matrix \( A \in \mathbb{C}^{N \times N} \) to yield the vector of N observations

\[
x(k) = As(k).
\]

(1)

The prewhitening of the observations by the matrix \( W \) yields the new observations vector, \( z(k) = Wx(k) \), whose covariance, \( E\{zz^H\} \), is the identity matrix of dimension \( N \times N \). The output signal \( y(k) \) that estimates one of the sources is obtained as \( y(k) = u^H(z(k)) \), where \( u \) is the unit norm extraction vector (enforcing the output to have unit variance). The overall transfer vector from the sources to the output is defined by the row vector \( r^i = u^HWA \).

In this article the superscripts \((i^*, j), (i) \) denote conjugate, transpose and conjugate transpose, respectively.

3. A cyclic method for blind extraction

Several criteria have been proposed in the context of complex maximization of the non-Gaussianity and in the context of the maximum likelihood principle for independent component analysis. These criteria use nonlinear functions of either the output (e.g., [6]) or the output’s modulus (e.g., [2]). In this last case the sources are assumed to have circular symmetry since the phase information is lost (the phase information is relevant for the approximation of nonsymmetric probability density functions and, thus, for the extraction of noncircular sources).

The criterion proposed in [6] consists in the search for the extrema of the function \( J_{\text{neg}}(y) = E\{G(y)^2\} \), which minimizes the entropy or non-Gaussianity of the output. In this work we start from a similar criterion, approximating the negentropy of the output (of unit variance) by

\[
J_{\text{neg}}(y) = h_N - vE\{G(y)^2\},
\]

where \( h_N \) is the entropy of a Gaussian variable with unit variance and \( G(y) \) is a selected nonlinearity. Actually we will show in Section 4 that any value selected for \( h_N \) larger than a certain threshold guarantees the ascent of \( J_{\text{neg}}(y) \) through iterations. The parameter \( v \) takes values within the set \( \{-1, 1\} \), depending on whether the maximization or minimization of the function \( E\{G(y)^2\} \) is appropriate. This value is determined by the selected function, \( G(y) \) (see [6]).

We propose to define a function \( F(y) \) such that \( G(y) = |F(y)| \). This will be useful for decoupling the problem, as it will be shown below. Taking into account that \( E\{|F(y)|^2\} = 1 \), we can rewrite (2) as

\[
J_{\text{neg}}(y) = E|F(y)|^2h_N - vE\{|F(y)|^2y^*\}
\]

and we will maximize (3) subject to the unit norm constraint \( u^H u = 1 \).

Since Gaussian variable has the largest entropy for a fixed variance, when function \( G(y) \) is correctly selected we have \( h_N \geq E\{|F(y)|^2y^*\} \), so expression (3) can be rewritten as

\[
J_{\text{neg}}(y) = E|y|^2h_N - vE\{|F(y)|^2y^*\} = u^H u h_N - vE\{u^H F(y^*) z^H\} u,
\]

(4)

where the second equality is given by the fact that \( y(k) = u^H(z(k)) \) and \( E\{u^H(z(k))\} = I \).

In order to simplify the maximization of (4) we propose to decouple the dependence between the involved output variables by means of the introduction of three independent candidates for the extraction vector, \( u^{(1)}, u^{(2)}, u^{(3)} \), whose respective outputs, \( y^{(1)}(k), y^{(2)}(k), y^{(3)}(k) \), are obtained from \( y(k) = u^{(i)}(z(k)) \) for \( j = 1, \ldots, 3 \). Then, after decoupling the results, the resulting function is

\[
J_{\text{neg}}(u^{(1)}, u^{(2)}, u^{(3)}) = \|u^{(1)} y^{(2)} u^{(3)} h_N - v(u^{(1)} y^{(2)} y^{(3)}) E\{z F(u^{(1)} y^{(2)} z^H)^2 z^H\} u^{(3)}\|.
\]

(5)

The function given by (5) will be cyclically maximized w.r.t. each one of the unit norm vectors \( u^{(i)} \) for \( j = 1, \ldots, 3 \), while keeping fixed the remaining variables, i.e., the maximization is carried out by means of a cyclical process which consists of three steps: maximization with respect to \( u^{(1)} \), with respect to \( u^{(2)} \) and with respect to \( u^{(3)} \) (and then back to the first step). Let us denote the extraction vector optimized at the ith iteration as \( u^{i}\). This vector is given by \( u^{(i)} = u^{(mod i - 1) + 1} \). Note the difference between the superscripts \( i \) and \( (i) \): while the first one refers to the ith candidate, the second one denotes the ith iteration of the algorithm. The introduction of three variables allows us to isolate the term \( |F(u^{(i)} z)|^2 \), the only one for which we cannot obtain a quadratic dependence. As we will show below, it is possible to simplify the maximization w.r.t. the third argument, \( u^{(3)} \) so as it coincides with the maximization of \( f(y) \) w.r.t. \( u^{(1)} \) and \( u^{(2)} \). The use of three variables (and not two) allows us to get a quadratic dependence of the function (5) w.r.t. the arguments \( u^{(1)} \) and \( u^{(2)} \), what makes possible the use of the Cauchy–Schwarz inequality to solve the maximization w.r.t. one of these arguments (when keeping fixed the others).

For a properly chosen \( F(y) \) (see Section 4), at the maxima of \( J_{\text{neg}} \) we should get \( y^{(i)}(k) = y^{(2)}(k) = y^{(3)}(k) = \hat{s}(k) \), which estimates one of the sources. So, in order to accelerate the convergence speed of the algorithm, after each iteration we choose to project the result onto the manifold \( y^{(1)}(k) = y^{(2)}(k) = y^{(3)}(k) \) that contains the desired solution of the problem. This projection is done by simply saving the values of \( y^{(i)}(k) = (u^{(i)} y^{(2)} u^{(3)} z^H k^H) / y^{(3)}(k) \) after each iteration. Then, after the projection, every output candidate will share the same common value, \( u^{(1)} = u^{(2)} = u^{(3)} = u^{i}\).

The first step of the cyclic maximization of (5) is the maximization w.r.t. \( u^{(3)} \) while keeping \( u^{(1)} \) and \( u^{(2)} \) fixed to \( u^{(i-1)} \). So, for \( i = 1, 4, 7, \ldots \) (i.e., \( u^{(0)} = u^{(1)} \)) the optimum of the extraction vector is given by

\[
u^{(i)} = \arg \max_{u^{(i)} = u^{(i-1)}} \left| |u^{(1)}|^2 u^{(i-1)} h_N - v|u^{(1)}|^2 E\{z F(y^{(i-1)})^2 (y^{(i-1)})^*\}\right|
\]

(6)

An analogous reasoning leads us to the optimum of the extraction vector for \( i = 2, 5, 8, \ldots \) (the second step of the cyclic maximization, i.e., the maximization w.r.t. \( u^{(2)} \) while keeping \( u^{(1)} \) and \( u^{(3)} \) fixed to \( u^{(i-1)} \))

\[
u^{(i)} = \arg \max_{u^{(i)} = u^{(i-1)}} \left| |u^{(2)}|^2 u^{(i-1)} h_N - v|u^{(2)}|^2 E\{z F(y^{(i-1)})^2 (y^{(i-1)})^*\}\right|
\]

(7)

and for \( i = 3, 6, 9, \ldots \) (the third step of the cyclic maximization, i.e., the maximization w.r.t. \( u^{(3)} \) while keeping \( u^{(1)} \) and \( u^{(2)} \) fixed to \( u^{(i-1)} \)).
fixed to $u^{(e-1)}$)

$$u^{(0)} = \arg \max_{u \in \mathbb{U}} \left[ u^{(e-1)}h_N - \mathbb{V} \left[ |f(u^{(e-1)}z)|^2 \right] \right].$$

(8)

Therefore the first and second steps (the optimization of $u^0$ while keeping fixed $u^{(1)}$ and $u^{(2)}$ and the optimization of $u^{(2)}$ while keeping fixed $u^{(1)}$ and $u^{(3)}$ are equal.

From Eqs. (6)-(8) it can be seen that we no longer need to maintain the notation for all the extraction candidates (the superscript $[,]$, but only for the extraction vector at the $i$th iteration, $u^i$, and at the previous iteration, $u^{(i-1)}$, so that the cyclical maximization of (5) is done through the iterations until the convergence according to the procedure described in Table 1.

We consider that the algorithm converges when $|u^{(0)} - u^{(i-1)}| < \varepsilon$.

According to the Cauchy–Schwarz inequality, the solution of the constrained maximization problem in step (I) is

$$u^{(0)} = u^{(e-1)}h_N - \mathbb{V} \left[ |f(y^{(e-1)}z)|^2 \right].$$

(9)

On the other hand, the constrained maximization in step (III) can be done by means of a gradient algorithm taking $u_0 = u^{(e-1)}$ as initial value. In Appendix A.1 we show that the sub-iterations of this algorithm are

$$u_i = u_{i-1} - \gamma \mathbb{V} \left[ |f(y^{(e-1)}z)|^2 \right].$$

(10)

provided that the function $f(y)$ satisfies

$$\frac{\partial f(y)}{\partial y} = (p-1)f(y)^2.$$

(11)

For functions that satisfy (11) and for small values of $p-1$ the expression to be maximized in step (III) does not change significantly with $u$, hence we can only take one sub-iteration of the gradient algorithm. Thus, for functions that satisfy (11), the constrained maximization in step (III) is simplified to the following iteration:

$$u^{(i)} = u^{(e-1)} - \gamma \mathbb{V} \left[ |f(y^{(e-1)}z)|^2 \right].$$

(12)

Taking $\gamma = 1/h_N$, the iterations for steps (I)-(III) coincide, so that the proposed algorithm is reduced to the iteration (9) until convergence, as shown in Table 2.

The only family of functions that holds Eq. (11) is $f(y) = r \cdot y^{(p-1)}$, i.e., $G(y) = y^{p-1}$ (although in simulations section we also tested the function $G = (0.1 + |y|^2)^{1/4}$, which does not hold the condition, with good results). So, $p$ is the exponent of a polynomial function used for approximating negentropy. In general is a real-valued constant that can be noninteger. Since for simplifying the iteration in step (III) we use small values of $p-1$ (values of $p$ close to 1) this provides a contrast function robust against outliers since it grows slower than quadratic with respect to its argument [2].

4. Monotonic ascent

In this section we derive a condition for the iteration (9) that guarantees the ascent of the function $J_{\neg \neg \neg \neg} (u^1z)$ given by (3) for nonlinearities that satisfy (11). For this purpose we define the vectors $u_k, u_r \in \mathbb{R}$ so that $u = u_k + ju_r$ and the output is given by $y = u|z - ju|^2z$.

Theorem 4.1. If the function $\phi(u_k, u_r)^2$ is convex in the convex set $\{|u| \leq 1\}$ then iteration (9) guarantees a monotonic ascent in the function $J_{\neg \neg \neg \neg} (u^1z)$ given by Eq. (3).

Proof. The proof of this theorem is given in Appendix A.2.

A well-known result is that the function $\phi(u_k, u_r)^2$ is convex on a convex set if and only if its Hessian matrix is positive definite [26]. By using this result we derived the following lemma, whose proof can be found in Appendix A.3.

Lemma 4.2. If we denote the real and imaginary parts of a complex value as $\Re (\cdot)$ and $\Im (\cdot)$ respectively, for a nonlinearity $f(y)$ satisfying (11), if the following condition

$$h_N > \left| (p-1) \Re \left( \mathbb{E} \left( f(y)^2 \frac{\partial f(y)}{\partial y} z_i \right) \right) \right| + \left| v \mathbb{R} \left( \mathbb{E} (f(y)^2 z_i) \right) \right|$$

$$+ \left| \sum_{j=1}^N \mathbb{E} \left( f(y)^2 \frac{\partial f(y)}{\partial y} z_j \right) \right| + \left| v \mathbb{R} \left( \mathbb{E} (f(y)^2 z_j) \right) \right|$$

$$+ \left| \sum_{j=1}^N \mathbb{E} \left( f(y)^2 \frac{\partial f(y)}{\partial y} z_j \right) \right| + \left| v \mathbb{R} \left( \mathbb{E} (f(y)^2 z_j) \right) \right|$$

(13)

is satisfied, then iteration (9) guarantees a monotonic ascent of $J_{\neg \neg \neg \neg} (u^1z)$.

4.1. A more practical condition

Now we will propose a practical condition based on Lemma 4.2. We will distinguish two different cases:

(a) For $1 < p < 2$ we set $v = 1$ (for communications signals like QPSK or 8-PSK this fact is proved in the Appendix A.4), i.e., $E(f(y)^2) = 1$. This will be minimized. As $0 < p-1 < 1$, the maximum value of $E(f(y)^2)$ is 1 and it is achieved at the extraction. For sources with circular symmetry, e.g., communications signals like QPSK or 8-PSK) $\mathbb{E} |s|^2 = 1$ and thus $\mathbb{E} (zz^T) = 0$. On the other hand let us recall that $\mathbb{E} |z|^2 = 1$. Therefore, for circular sources terms in (13) including $z_i z_j$ (with $i \neq j$) or $z_i^2$ (with $i = j$) are negligible with respect to $v \mathbb{R} (f(y)^2 z_i)$ and $\mathbb{E} (f(y)^2)$. Thus the condition (13) can be approximated by

$$h_N > p \Re \left( f(y)^2 \right)$$

(14)

If $p-1$ is small, $f(y)^2$ has few variations and is weakly dependent with $|z|^2$, so that $E(f(y)^2 |z|^2) = E(f(y)^2) = E(f(y)^2)$. Since the maximum value of $E(f(y)^2)$ is 1 we can
reasonably approximate (14) by
\[ h_n > p. \]  
(15)

(b) For \( 0 < p < 1 \) we set \( v = -1 \). Since dominant term in right hand of Eq. (13) is \( v p E[F(y)|^2 |z|^2] \) and is negative, for satisfying (13) it is sufficient to define
\[ h_n > 0. \]  
(16)

5. Extraction of all the sources

The extraction of all the \( N \) sources can be achieved by running the algorithm the same number of times, obtaining the vector of outputs \( y(k) = [y_1(k), \ldots, y_N(k)]^T \), where \( y_j(k) = U^j k z(k) \). If we define the extraction matrix as \( U = [u_1, \ldots, u_N] \), then \( y(k) = U^j k z(k) \). The output can be also expressed in terms of the sources vector as \( y(k) = G s(k) \), where \( G = U^H W A \) is the overall transfer matrix. Among the methods that prevent the outputs from converging to the same source, we consider the symmetric orthogonalization and the deflation approach.

5.1. Symmetric orthogonalization

This method consists in the execution of the algorithm \( N \) times simultaneously and the symmetric decorrelation of the extraction vectors. This decorrelation is carried out after each iteration by means of
\[ U^{i+1} = U^{i} U^{i} H U^{i} J^{-1/2}, \]  
(17)
where \( U^{i} \) is the extraction matrix at \( i \)th iteration [2]. To obtain more accurate results, after convergence the algorithm is run a second time for each output without constraints [27].

5.2. Deflation method

In this case the estimation of the sources is done sequentially, removing each source from the data after its estimation [28]. Let \( u_1, \ldots, u_{j-1} \) be the unit norm extraction vectors for the \( j-1 \) first extracted sources. The observations vector for the \( j \)th extraction (\( j > 1 \)), once the \( j-1 \) previous extracted sources have been removed, is
\[ x_j \left( 1 - \sum_{q=1}^{j-1} u_q u_q^H \right) z. \]  
(18)

If \( W \) is the prewhitening matrix of dimension \( (N-j+1) \times N \) for \( x_j \) and \( v_j \) is the unitary vector of dimension \( (N-j+1) \times 1 \) that achieves the extraction for the prewhitened data \( W x_j \), then
\[ u_j = \left( 1 - \sum_{q=1}^{j-1} u_q u_q^H \right) W v_j \]  
(19)
is the \( N \times 1 \) extraction vector applied to \( z(k) \). Again, after convergence, we can run the algorithm for the original observations vector, \( z(k) \) (i.e., for \( z(k) \)), to obtain more accurate results.

While symmetric orthogonalization is faster, the deflation approach has the advantage that proof of monotonic ascent still remains valid. This is due to the fact that the iterations of the extraction algorithm are not modified.

6. Simulations

In order to illustrate the behavior of the proposed algorithm we performed a set of computer experiments. In these simulations we mixed eight sources and varied the constellation of the sources: QPSK and 8-PSK, depending on the experiment. We compared the proposed algorithm with FastICA [2] and N-CMN [6] algorithms using them for the extraction of all sources. For all the algorithms we carried out two executions: in the first one we enforced the orthogonality of the extraction vectors, while in the second one we took the result of the first stage as starting point and the algorithms were executed without this orthogonality constraint. Although the only family of functions that satisfy (11) is \( F(y) = ry^{p-1} \) \( \forall \gamma, p \in \mathbb{R} \), this algorithm was tested with the nonlinearities \( G(y) = y^{1 + 25} \) (i.e., \( F(y) = y^{0.25} \)) and \( G(y) = (0.1 + |y|^2)^{1/4} \) (i.e., \( F(y) = (0.1 + |y|^2)^{1/4}/y \)). These functions were proposed in [6,2] respectively in the context of complex maximization of non-Gaussianity. For N-CMN algorithm we chose the same nonlinearities. Finally the nonlinearities chosen for the FastICA algorithm were the already mentioned and another one proposed in [2] for this algorithm: \( \log(0.1 + |y|^2) \).

In a first set of simulations we compared the algorithms in the absence of noise in terms of Amari’s index, which is a quality index commonly accepted in the context of BSS [6,29] (normalized to take values between 0 and 1), which measures the quality of the extraction of the sources
\[ I_A = \frac{1}{2N} \sum_{i=1}^{N} \left( \sum_{j=1}^{N} \left( \frac{|g_j|}{\max(1, |g_{j1}, \ldots, g_{jN}|)} \right)^{-1} \right) \]

Fig. 1. Amari’s index versus data length for the extraction of all the sources in noiseless mixtures. We carried out two executions for all the algorithms: the first one enforced the orthogonality of the extraction vectors, while in the second one the result of the first stage was taken as starting point and the algorithms were executed without this orthogonality constraint.
where \( g_0 \) are the elements of the overall transfer matrix, \( G \), defined above. The value of \( I_A \) decreases as the quality of the extraction increases. We averaged the values of this measure by performing 100 Monte Carlo runs for each case. At each run the sources, the mixing matrix and the initial values for the extraction vectors were randomly generated.

The data length was varied from 200 to 1000 samples. Results are shown in Fig. 1, where Amari’s index is plotted versus data length. We observe that results are similar for QPSK and 8-PSK sources. The proposed algorithm performs with better quality than the others, being its results better for the nonlinearity that exactly satisfies (11). The proposed algorithm shows satisfactory results even for a small set of data (200 samples), whereas the quality of separation of the N-CMN and FastICA algorithms decreases considerably in this case. In all cases the proposed algorithm shows an Amari’s index between 25 and 30 dB (for the nonlinearity \( G(y) = y^{1.25} \)) and between 15 and 20 dB (for \( G(y) = 0.1 + |y|^2 \)) lower than those shown for N-CMN and FastICA algorithms. We can see that even for a nonlinearity that does not satisfy (11) the performance of our method is better than the others.

In a second experiment we compared the behavior of the algorithms in the presence of additive Gaussian noise for a fixed data length of only 200 samples. We mixed eight sources into twelve observations and varied the SNR, taking values from 0 to 40 dB. The SNR was defined as

\[
SNR = \frac{E[|\sum_{j=1}^{N} a_j s_j|^2]}{E[|\hat{s}_l|^2]} \quad \forall i
\]

being \( a_j \) the \((i,j)\)th entry of \( A \). After dimension reduction by means of PCA, at each run, all the eight sources were estimated. Results are shown in Fig. 2, where Amari’s index is plotted versus SNR. Results are similar again for QPSK and 8-PSK sources, and the proposed algorithm shows the best performance. For our method \( I_A \) curve decreases almost linearly with the value of SNR in dB whereas for N-CMN and FastICA the value of Amari’s index seems to tend to a limit.

Finally in a third experiment we performed 1000 Monte Carlo runs to compute the CPU time per run used by each algorithm. Although the CPU time is quite dependent on the implementation done for the algorithms we used this comparison in order to illustrate that the algorithm has a similar execution time than the others although its quality of separation is better. Eight QPSK sources (200 samples each) were mixed into eight observations in the absence of noise, providing similar execution times for all the algorithms, being the proposed algorithm with the nonlinearity \( y^{1.25} \) the fastest. That is, the improvement of the quality in the proposed algorithm is not achieved by increasing the complexity.

7. Conclusions

We have proposed an algorithm for the extraction of complex sources through the maximization of approximations of negentropy based on nonlinearities. The procedure is based on the decoupling of the original variables of the criterion and on the cyclical maximization of the contrast with respect to each of the variables. We have shown through computer simulations that the proposed method offers an improved performance in terms of quality of extraction with respect to other well-known approaches without an increase of the computational effort. We have also provided a proof of the monotonous ascent of the contrast function for the proposed algorithm.

Acknowledgments

This work has been supported by the MCINN project of the Government of Spain (Grant TEC2008-06259).

Appendix A

A.1. Maximization in step III

Let us define \( \psi(\mathbf{u}) \) as the function to be maximized in step (III).

\[
\psi(\mathbf{u}) = [E[|y^{(l-1)}|^2]] - \nu E\left(|F(\mathbf{u}^l(\mathbf{z}))|^2(y^{(l-1)})^*\right].
\]

Now we will consider the maximization of \( \psi(\mathbf{u}) \) w.r.t. \( \mathbf{u} \), s.t. \( \|\mathbf{u}\| = 1 \) by means of a gradient algorithm. Taking into account that \( \nabla_\mathbf{u}^l y = \mathbf{z} \) and \( \nabla_\mathbf{u}^l y^* = 0 \), the gradient of \( \psi(\mathbf{u}) \) is given by \( \nabla_\mathbf{u} \psi(\mathbf{u}) = \mathbf{z}(\partial \psi(\mathbf{u}) / \partial y) \). The sub-iteration of the gradient algorithm is given by

\[
\mathbf{u}_{l-1} - \nu E\left[\mathbf{z}_{y^{(l-1)}} \frac{\partial F(y)^2}{\partial y}(y^{(l-1)})^*\right]
\]

\[
\mathbf{u}_l = \mathbf{u}_{l-1} - \nu E\left[\mathbf{z}_{y^{(l-1)}} \frac{\partial F(y)^2}{\partial y}(y^{(l-1)})^*\right].
\]
For functions that satisfy Eq. (11) and defining $z = \beta(p-1)$, if we substitute (11) into (23) we have (10). □

A.2. Proof of Theorem 4.1

The function $\phi(u_i, u_j)^\top \in \mathbb{R}$ is convex in $S = \{u : ||u|| < 1\}$ if and only if it satisfies the gradient inequality [26], i.e., for any vectors $u, v \in S$

$$\phi(u_i, u_j)^\top - \phi(v_i, v_j)^\top \geq [u_i - v_i, u_j - v_j] \nabla \phi \frac{v_i, v_j]}{||v_i, v_j||} = g i \{u - v\nabla \phi \}^\top \{u - v\nabla \phi \},$$

where $v_i, v_j \in \mathbb{R}$ are defined so that $v = v_i + v_j$ and $\phi(u) \equiv \phi(u_i^\top, u_j^\top)$. In particular, gradient inequality holds for $u = u_i$ and $v = u_{i-1}$

$$\phi(u_i) - \phi(u_{i-1}) \geq g i \{u_i, u_i\nabla \phi \}^\top \{u_i, u_i\nabla \phi \},$$

and $\phi(u_i) - \phi(u_{i-1}) \geq g i \{u_i, u_i\nabla \phi \}^\top \{u_i, u_i\nabla \phi \} \geq 0$. Taking into account that $E(y)^2 - 1$ the ascent of $\phi$ implies the ascent of (3). A similar proof is provided in [30] for the real case. □

A.3. Proof of the Lemma 4.2

The Hessian matrix of the function $\phi(u_i, u_j)^\top$ is positive definite if it is a diagonal dominant matrix with real positive diagonal entries [31]. For function $Fl_y$ satisfying (11), the condition of real positive diagonal entries results to

$$h_N > |v(p-1)R\{E(F(y)^2 Y^2)\}| + v^2 E[F(y)|y|z^2].$$

Besides, the condition of diagonal dominant matrix results in two conditions:

$$h_N > |v(p-1)R\{E(F(y)^2 Y^2)\}| + v^2 E[F(y)|y|z^2].$$

and

$$h_N > -|v(p-1)R\{E(F(y)^2 Y^2)\}| + v^2 E[F(y)|y|z^2].$$

for $1 \leq i \leq N$. Condition (13) includes conditions (27) and (28) and leads to a convex function $\phi(u_i, u_j)^\top$. Due to Theorem 4.1 this implies that, for the iteration (9), $\phi(u_i) - \phi(u_{i-1}) \geq 0$. □

A.4. Choice of $v$ for QPSK and 8-PSK signals

From [6] we know that for circular sources and a nonlinearity $G(y) = y^p$, the function $E[F(y)^2]$ must be minimized (respectively, maximized) when

$$E[|s_1|^p] < \mu E[|s_1|^p - 1]$$

being $s$, the source to be extracted. For signals like QPSK or 8-PSK, that have the property $|s_1| = 1$ this condition is equivalent to $p > 1$ (resp., $< 1$).

Thus, for $1 < p < 2$ the function $E(F(y)^2)$ must be minimized (i.e., $v = 1$) while for $0 < p < 1$ this function must be maximized (i.e., $v = -1$). □

References


Iván Durañ-Díaz was born in Seville, Spain, in 1975. He received the Telecommunication Engineer degree from the University of Seville in 2001, and the Ph.D. degree from the University of Seville in 2009. Since 2002 he has been with the Radiocommunications Systems Group at the Signal Theory and Communications, Department of the University of Seville, where he is currently an Assistant Professor. He teaches undergraduate courses on communications theory, digital transmission systems and radiocommunications. His current research interests are in the area of statistical signal processing, spread spectrum and wireless communications systems.

Sergio Cruces was born in Spain. He received the Telecommunication Engineering degree in 1994 and the Ph.D. degree in 1999, both at the University of Vigo (Spain). From 1994 to 1995 he worked as project engineer for the Department of Signal Theory and Communications of the same university. In 1995 he joined the Signal Theory and Communications group of the University of Seville, where he is currently an Associate Professor. He teaches undergraduate and postgraduate courses on digital signal processing and mathematical methods for communication. His current research interests include statistical signal processing and information theoretic methods in neuroscience and communications.

María Auxiliadora Sarmiento-Vega was born in Seville, Spain, in May 1978. She received the Telecommunication Engineering from the University of Seville, Seville, Spain in 2002. From 2003 to 2005, she was working as a project engineer for the Department of Medical Physiology and Biophysics of the same university. In 2005 she joined the Radiocommunications System group in the Department of Signal Theory and Communications in the same university where she is currently an Assistant Professor. Her research interests focus on the area of blind signal separation especially in speech mixtures.

Pablo Aguilera-Bonet was born in Jerez, Spain, in 1984. He received the Telecommunication Engineering and M.Sc. in Electronics, Signal Processing and Communications degrees from the University of Seville, Spain, in 2007 and 2010, respectively. From 2007 to 2009, he worked with a research scholarship at the Department of Signal Theory and Communications, University of Seville. He is currently a Ph.D. student at the University of Seville, Spain. He has been a Visiting Student at the Communications and Signal Processing group at the Imperial College, London. His research interests include statistical signal processing, signal separation and latent component analysis.