Blind Separation of Convolutive Mixtures: A Gauss-Newton Algorithm

Sergio Cruces
TSC Group, ESI Telecomunicación
Universidad de Sevilla
Av. Reina Mercedes, 41012-Sevilla, Spain
sergio@viento.us.es

Luis Castedo
Dpto. Electrónica y Sistemas
Universidad de La Coruña
Campus de Elviña, 15071-La Coruña, Spain
luis@sol.des.fi.udc.es

Abstract

This paper addresses the blind separation of convolutive mixtures of independent and non-Gaussian sources. We present a block-based Gauss-Newton algorithm which is able to obtain a separation solution using only a specific set of output cross-cumulants and the hypothesis of soft mixtures. The order of the cross-cumulants is chosen to obtain a particular form of the Jacobian matrix that ensures convergence and reduces computational burden. The method can be seen as an extension and improvement of the Van-Gerven’s symmetric adaptive decorrelation (SAD) method. Moreover, the convergence analysis presented in the paper provides a theoretical background to derive an improved version of the Nguyen-Jutten algorithm.

1. Introduction

Blind source separation is receiving a growing interest due to its applications in diverse fields such as array processing, multiuser communications, etc ([10]-[8]). While most of the work has been developed in the context of instantaneous mixtures, the more difficult problem of convolutive mixtures has received less attention. Recently, Van Gerven et al. [15] have proposed and analyzed the Symmetric Adaptive Decorrelation (SAD) algorithm that uses second order statistics and exhibits interesting signal separation properties. However, it is well known that methods based on second order statistics are not sufficient to solve the blind source separation problem and higher order statistics are necessary. For this reason, Nguyen and Jutten [12] have proposed an algorithm (NJ) that cancels fourth order cross-cumulants of the output signals to achieve separation. Nevertheless, the convergence of this algorithm has not been analyzed yet.

In this paper we present a method that generalizes and improves the SAD algorithm using higher order cross-cumulants. From the given method, it also can be derived an algorithm which is similar to the NJ algorithm for two sources except for a term that ensures asymptotic stability.

2. Signal model

Let us assume the signal model presented in Figure 1. There are \(N\) independent sources \(s[n] = [s_1[n], \ldots, s_N[n]]^T\) where at most one is Gaussian. The Darmois-Skitovich theorem [3] ensures that the sources can be separated by imposing pairwise independence between them. The sources are convoluted together through a multichannel linear time-invariant system, with memory, to obtain the observations vector \(x[n] = [x_1[n], \ldots, x_M[n]]^T\), where \(M \geq N\). The impulse response of the mixing system is characterized by the sequence of \(M \times N\) matrices \(A[n] = [a_{ij}[n]]\). We will assume that the mixing system is FIR and (possibly) non-causal. As a consequence, \(A[n] \neq 0\) for \(-L_{a1} \leq n \leq L_{a2}\), and the relationship between sources and observations can be written as

\[
x[n] = A[n] * s[n] = \sum_{k=-L_{a1}}^{L_{a2}} A[k]s[n - k]
\]

In order separate the sources, the observations are processed by another multichannel LTI system, with memory, characterized by its \(N \times M\) impulse response sequence matrix \(B[n]\). Again, the separating system will be FIR, (possibly) non causal and non-zero in the interval \(-L_{b1} \leq n \leq L_{b2}\). Therefore, the output vector \(y[n] = [y_1[n], \ldots, y_N[n]]^T\) is given by

\[
y[n] = B[n] * x[n] = \sum_{k=-L_{b1}}^{L_{b2}} B[k]x[n - k]
\]

There are several indeterminacies in the source separation problem [14]. Our approach can handle the sources
ordering indeterminacy, but we must avoid the scaling and delay indeterminacy. Towards this aim, we will suppose that the diagonal terms of the separation sequence matrices $B[k]$ and that of $A[k]$ are equal to the unit impulse $\delta[k]$ ($P$ is an unknown permutation matrix).

Once the signals $y[n]$ are separated, it is necessary to introduce a post-processing system $D(z)$ in order to remove the undesired correlations originated by the separation process. For this task, it can be shown [5] that the $D(z)$ transfer function should be diagonal with elements given by

$$D_{ii}(z) = \frac{\text{Adj}_{ii}(B(z))}{\text{Det}(B[z])}, \quad i = 1, \ldots, N. \quad (1)$$

where $\text{Adj}_{ii}(\cdot)$ is the adjoint operator on the $(i, i)$ matrix element. At the output of the post-processing system we obtain the vector $\hat{s}[n]$ of estimated sources.

The following hypothesis will be needed for the algorithm to perform adequately

1. The mixture is soft in the sense that each observed signal receives a dominant contribution from one of the sources. This typically occurs when each sensor is closer to a different source and these have similar power.

2. There is a subset of cumulants of the sources in which the cumulants of the same order do not have great differences in their magnitude. This is obviously true when the sources are identically distributed.

The soft mixture assumption is appropriate for the separation case of convolutive mixtures, since a sufficient condition for the stability of the post-processing filters also depends on this hypothesis. It can also be shown that the soft mixtures assumption will lead to a $D(z)$ that is not far from the identity matrix, fact that will be important in later demonstrations.

The separation filter length $L = L_{b1} + L_{b2} + 1$ is chosen to match the length of the filters in the numerator of the inverse multichannel transfer function. Therefore $L_{b1} = L_{a1}(N - 1)$ and $L_{b2} = L_{a2}(N - 1)$ if $M = N$, and $L_{b1} = 3L_{a1}(N - 1)$ and $L_{b2} = 3L_{a2}(N - 1)$ when $M > N$. The inverse of a FIR multichannel transfer function has some IIR components, these components will be provided by $D(z)$ in the two sources case. Nevertheless, in the multiple sources case, our approach will be only approximate. We can’t reach exactly the solution but, under the soft mixture hypothesis, the IIR part will be negligible or approximable by setting a bit longer the separation filters length.

3. Statistical dependence measure

We define the cumulant tensor $C_{\alpha_1,\ldots,\alpha_l}(y[n], \ldots, y[n])$ as the tensor where the $(i_1, \ldots, i_l)$ element is $C_{\alpha_1,\ldots,\alpha_l}(y_{i_1}[n], \ldots, y_{i_l}[n]) \triangleq \text{Cum}(y_{i_1}[n], \ldots, y_{i_l}[n], \alpha_1, \ldots, \alpha_l)$ the $(\alpha_1, \ldots, \alpha_l)$-order cross-cumulant between the signals $y_{i_1}[n], \ldots, y_{i_l}[n]$.

We propose that the adjustable elements of the separating matrix $B(z)$ be selected to jointly diagonalize the set of cumulant matrices $C_{\alpha,\beta}(y[n], y[n - k])$ for the different lags $k = -L_{b1}, \ldots, L_{b2}$ and for the set $\Omega$ of $n\Omega$ pairs $(\alpha, \beta)$ of natural numbers. Towards this aim, we propose to minimize the following cost function

$$\Phi_{\Omega} = \sum_{(\alpha,\beta)\in\Omega} w_{\alpha,\beta} \phi_{\alpha,\beta} \quad (2)$$

$$\phi_{\alpha,\beta} = \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \sum_{k=-L_{b1}}^{L_{b2}} (C_{\alpha,\beta}(y_i[n], y_j[n - k]))^2 \quad (3)$$

The $w_{\alpha,\beta}$ are a set of weighting coefficients. Note that $\Phi_{\Omega}$ can be interpreted as a statistical dependence measure of the output vector $y[n]$. Initially, we will consider the set $\Omega$ containing all the possible pairs $(\alpha, \beta)$. However, it will be shown that the minimization of the subset $(1, \beta)$ is sufficient to ensure separation under the soft mixture assumption. Trivial non-separation solutions, where any of the outputs is equal to zero, occur when the mixing or separating matrices are not full row rank. The convergence to these solutions is avoided by the soft mixture hypothesis since, when this hypothesis holds, we found to be on the basin of attraction of the correct separation solution.

4. Minimization algorithm

For simplicity reasons, we will rewrite the dependence measure $\Phi_{\Omega}$ in a vector form. Let us define the vector $z_{(\alpha,\beta)} = \{C_{\alpha,\beta}(y_i[n], y_j[n - k]); i, j, j \neq i = 1, \ldots, N; k = -L_{b1}, \ldots, L_{b2}\}$. If we now define the vector $z = [\sqrt{\phi_{\alpha,\beta}}, \sqrt{\phi_{\alpha,\beta}}^T]_{\alpha,\beta=1}^{n\Omega}$ with $(\alpha, \beta)_i \in \Omega$, the dependence measure can be written as the inner product $\Phi_{\Omega} = z^T z$. It is also convenient to rearrange the separation variables in a vector $b = \{b_{ij}[k]; i = 1, \ldots, N; j, j \neq i = 1, \ldots, N; k = -L_{b1}, \ldots, L_{b2}\}$.

Different algorithms can be used to minimize $\Phi_{\Omega}$. Due to the soft mixture hypothesis we are initially close to the separation solution and the vector $z$ of output cross-cumulants is small (with norm close to zero). In this situation, we can use the Gauss-Newton (GN) method [6] to
find the desired minimum, because the Hessian matrix of the adaptation can be well approximated from the Jacobian matrix. This reason and the quadratic convergence of the GN method reduces the computational burden.

The Gauss-Newton method uses the gradient and the Hessian of \( \Phi_{\Omega} \) with respect to the separation coefficients. The gradient is given by: 
\[
\nabla_b \Phi_{\Omega} = 2Jz
\]
where \( J = \left[ \frac{\partial \phi}{\partial \delta_1}, \ldots, \frac{\partial \phi}{\partial \delta_{NM}} \right]^T \) is the Jacobian matrix. The global Jacobian matrix can also be defined in terms of the sub-Jacobian matrices \( J_{(\alpha, \beta)} = \nabla_b z_{(\alpha, \beta)} \), since \( J = \left[ \sqrt{w_{(\alpha, \beta)}} J_{(\alpha, \beta)}^1, \ldots, \sqrt{w_{(\alpha, \beta)}} J_{(\alpha, \beta)}^N \right] \). Near the solution the Hessian can be approximated by \( H \approx 2JJ^T \). Then, the Gauss-Newton method consists in the following iteration
\[
b_{n+1} = b_n + \mu \Delta, \quad H \cdot \Delta = -2Jz
\]
where \( \mu \in (0, 2) \) is the step-size. Denoting by \( J^T \# \) the pseudo-inverse of \( J^T \), the increment \( \Delta \) is calculated as \( \Delta = -J^T \# z \). or, in an equivalent way, by solving the following system of linear equations
\[
\sum_{i=1}^{n_\alpha} u_i \alpha_\beta, J_{(\alpha, \beta)}^T J_{(\alpha, \beta)} \Delta = -\sum_{i=1}^{n_\alpha} u_i \alpha_\beta, J_{(\alpha, \beta)}^T z_{(\alpha, \beta)}
\]

### 4.1. Jacobian structure

In this section, we will choose an adequate set \( \Omega \) to reduce the computational burden of the algorithm (4). We will exploit the Jacobian structure for this selection, showing that some few elements of the Jacobian matrix has a strong dominance over the rest. This enables to approximate the Jacobian matrix by an strong sparse matrix. As a consequence, the computational burden of the algorithm will be reduced for two reasons: there are less cumulants to estimate and there exist efficient methods to solve linear sparse systems [7].

Each sub-Jacobian \( J_{(\alpha, \beta)} \) is a \( L_b N(N-1) \)-by-\( L_b N(N-1) \) rectangular matrix. Its elements are proportional to
\[
\frac{\partial C_{1, \alpha \beta}(y_i[n], y_j[n-k])}{\partial b_{rs}[m]} = \delta_{ir} \alpha C_{1, \alpha-1, \beta}(x_s[n-m], y_i[n], y_j[n-k]) + \delta_{jr} \beta C_{1, \alpha, \beta-1}(x_s[n-k-m], y_i[n], y_j[n-k])
\]
where \( \delta_{ir} = \{i = r, \ 0 \ \text{else}\} \) and where the index \( (i, j, r, s, m) \) belong to the range \( \{i \neq j, r = 1, \ldots, N; s \neq r = 1, \ldots, M; m, k = -L_{b1}, \ldots, L_{b2}\} \). The resulting expression shows that for \( (\alpha, \beta) \) both greater than 1 and under the assumption of soft mixtures, all the Jacobian elements are close to zero, since the signals \( y_i[n] \) and \( y_j[n] \) have an small and decreasing dependence between each other. Therefore, \( \Phi_{(\alpha, \beta)} \) has a slow variation or sensitivity with respect to the separation coefficients, and this makes this kind of pairs undesirable to use in a gradient based algorithm. When \( \alpha \) or \( \beta \) are equal to 1, the Jacobian is no longer close to zero, therefore avoiding the former problem. We will set \( \alpha = 1 \) since this choice leads to a well behaved structure for the Jacobian matrix while \( \beta = 1 \) does not. Substituting \( \alpha = 1 \) in expression (5) and neglecting the non-dominant terms around the separation solution, we arrive at this first approximation for the Jacobian elements
\[
\frac{\partial C_{1, \alpha \beta}(y_i[n], y_j[n-k])}{\partial b_{rs}[m]} \approx \delta_{ir} C_{1, \beta}(x_s[n-m], y_j[n-k])
\]
The above expression is not true for the case \( \beta = 1 \), but even in that case constitutes a valid approximation for the Jacobian.

The sources can be considered locally stationary with respect to the delay \( L_b \) (e.g. voice signals) or when we work with a block based method of \( n_d \) samples where \( n_d \gg L_b \), the cross-cumulant estimates can be considered stationary. This simplifies the last approximation to
\[
\frac{\partial C_{1, \alpha \beta}(y_i[n], y_j[n-k])}{\partial b_{rs}[m]} \approx \begin{cases} 
\delta_{ir} C_{1 \beta}(x_s[n-m+k], y_j[n]) & \text{if } k \leq m \\
\delta_{ir} C_{1 \beta}(x_s[n], y_j[n-k+m]) & \text{if } k \geq m
\end{cases}
\]

The Jacobian can now be computed from \( n_d 2L_b N(M-1) \) cumulant estimates. We see that the structure of the sub-Jacobians \( J_{(\alpha, \beta)} \) is sparse, they are block toeplitz with blocks of dimension \( N(N-1) \)-by-\( N(N-1) \), and each block is block diagonal formed by sub-blocks of dimension \( (M-1) \)-by-\( (N-1) \). When the sources are white with respect to their cumulants of order \( 1 + \beta \) or when the mixture is instantaneous, the approximation (6) can be further simplified in the following way
\[
\frac{\partial C_{1, \alpha \beta}(y_i[n], y_j[n-k])}{\partial b_{rs}[m]} \approx \begin{cases} 
\delta_{ir} \delta_{km} C_{1 \beta}(x_s[n], y_j[n]) & \text{if } k \leq m \\
\delta_{ir} \delta_{km} C_{1 \beta}(x_s[n], y_j[n-k]) & \text{if } k \geq m
\end{cases}
\]
In this case, the Jacobian precisest only \( n_d N(N-1) \) cumulant computations. The sub-Jacobian matrices \( J_{(\alpha, \beta)} \) are block diagonal with blocks of dimension \( (M-1) \)-by-\( (N-1) \). The proposed algorithm for the two white sources case reduces to the coefficients adaptation:
\[
b_{ij}[k]_{n+1} = b_{ij}[k][n] - \mu \sum_{(i, j) \in \Omega} w_{ij} C_{1 \beta}(x_s[n], y_i[n]) C_{1 \beta}(y_i[n], y_j[n-k]) \sum_{(i, j) \in \Omega} w_{ij} C_{1 \beta}(x_s[n], y_i[n])
\]
Recall that this iteration only requires \( 2n_d (L_b + 1) \) cumulants!
5. Link with another separation methods

For the moment, we have used the feed-forward (FF) separation structure (Figure 1). Another possibility is the feed-backward (FB) filtering structure that can be seen in Figure 2. This structure has the advantage of avoiding post-filtering by a post-processing matrix, but its analysis is not simple due to the feedback. In this structure, the transfer function of the separation filter is \((I + C(z))^{-1}\) where \(C(z)\) is a matrix with the same dimensions as \(A(z)\) and zero diagonal elements. We can establish a connection between our feed-forward algorithm and a similar feed-backward one under the assumption of soft mixtures.

The relation between the feed-forward and the feed-backward separation structure is given by \(D(z)B(z) = (I + C(z))^{-1}\). For the two sources case, the solution of this equation reduces to \(C(z) = I - B(z)\). This establishes the connection between both filtering approaches. For more than two sources and soft mixtures, we may approximate \((I + C(z))^{-1}\) by the linear part of its Taylor series expansion around \(C(z) = 0\) which is \((I + C(z))^{-1} \approx I - C(z)\). Since for the soft mixtures hypothesis we can approximate the post-processing matrix by the identity \(D(z) \approx I\), the connection between both filtering approaches still takes the form \(C(z) \approx I - B(z)\). Having this in mind, we can propose a feed-backward separation algorithm simply updating the coefficients \(c_{ij}[k]\) in the same way the feed-forward algorithm does, but changing the adaptation sign and replacing \(y[n]\) by \(\hat{s}[n]\). The algorithm equivalent to (8) in the feed-backward case is

\[
c_{ij}[k]_{n+1} = c_{ij}[k]_n + \mu \sum_{(1,\beta) \in \Omega} w_{\beta} \frac{C_{1,\beta}(x_j[n], \hat{s}_j[n]) C_{1,\beta}(x_j[n], \hat{s}_j[n-k])}{\sum_{(1,\beta) \in \Omega} w_{\beta} C_{1,\beta}^2(x_j[n], \hat{s}_j[n])} \quad i,j \neq i = 1, \ldots, 2; k = -L_{a1}, \ldots, L_{a2} \tag{9}
\]

This algorithm will also be asymptotically stable for \(\beta > 1\) since it converges to the Gauss-Newton algorithm when approaching to the separation solution. When \(n_{\Omega} = 1\) it simplifies to

\[
c_{ij}[k]_{n+1} = c_{ij}[k]_n + \mu \frac{C_{1,\beta}(\hat{s}_i[n], \hat{s}_j[n-k])}{C_{1,\beta}(x_j[n], \hat{s}_j[n])} \quad i,j \neq i = 1, \ldots, 2; k = -L_{a1}, \ldots, L_{a2}
\]

It can be seen that when \(\beta = 1\) this algorithm reduces to the SAD algorithm [15]. For \(\beta = 3\) it takes a similar, but improved, form of the NJ algorithm [12]. The improvement over this last algorithm is three-fold: the proposed algorithm has a greater speed of convergence, it works with non causal mixtures, and it is asymptotically stable under the given hypothesis while NJ is not stable.

6 Separation example

Before beginning with the separation example we will introduce some additional definitions. The combined multichannel impulse response sequence of mixing and separating systems will be denoted by \(H[n] = B[n] * A[n]\). Let us define the matrix \(H_{E} = \sum_{k=-\infty}^{\infty} H[k]_i j\) whose elements are the energy of \(H[n]\). We normalize the energy matrix in the following way: \(\tilde{H}_{E} = \frac{1}{H_{E}}(D_{x}^{-1}H_{E} + H_{E}D_{x}^{-1})\) where \(D_{x}\) and \(D_{E}\) are diagonal matrices whose diagonal elements are the maxima of the rows and columns of \(H_{E}\), respectively. This performance index is similar to the one used in [11]. The normalized matrix \(\tilde{H}_{E}\) is invariant to scaling by diagonal matrices and time shifts of the transfer functions. Its maxima coefficients by columns or rows are equal to 1, and will be associated with the normalized energy of the direct transfer functions. The rest will be associated to the normalized energy of the interference transfer functions.

Let us suppose we have two independent white source signals \(s[n]\) (of 5000 samples) which are mixed through the channel \(\mathbf{A}(z)\) to obtain a convolutive mixture \(y[n]\). The mixing matrix is formed by five taps FIR filters, and verify the soft mixture condition. The initial filters energy matrix is \(\tilde{H}_{E} = \begin{bmatrix} 1 & 0.21 & 0.98 & 1 \end{bmatrix}\). We run a few iterations of the feed-forward separation algorithm presented in (8) to separate sources. We will use the set of cumulants \(\Omega = \{(1,3)\}\) for the dependence measure. At each iteration we show in Figure 3 and Figure 4 the evolution and performance of the algorithm. In Figure 3 we show how the interference energy is reduced whereas the direct path signals energy is preserved. The final energy matrix is \(\tilde{H}_{E} = \begin{bmatrix} 1 & 0.0009 & 0.0009 & 1 \end{bmatrix}\) which is close to the identity matrix. The algorithm does not converge exactly to the identity matrix when using cross-cumulant estimates due to the finite length of the data. If we have the exact cross-cumulants at each iteration the convergence to the separation solution is achieved with an arbitrary precision. Figure 4 presents the minimization of the dependence measure. It can be seen that the theoretical convergence of the algorithm is quadratic, and therefore, reaches the separation solution in a few iterations.

7 Conclusions

A new signal separation algorithm is presented for the convolutive mixture of independent and non-Gaussian
ACKNOWLEDGMENT

The authors wish to thank J.I. Acha from the University of Seville (Spain) for his cooperation and fruitful discussions on the signal separation topic.

REFERENCES