A GAUSS-NEWTON METHOD FOR BLIND SOURCE SEPARATION OF CONVOLUTIVE MIXTURES

Sergio Cruces[†]

TSC Group, ESI Telecomunicación, Univ. Sevilla

Camino de Descubrimientos, 41092-Sevilla, Spain

sergio@viento.us.es

Luis Castedo[‡]

Dpto. Electrónica y Sistemas, Univ. A Coruña

Campus de Elviña, 15071-A Coruña, Spain

luis@sol.des.fi.udc.es

ABSTRACT

In this paper we present several Gauss-Newton algorithms for blind source separation of convolutive mixtures. The algorithms can be interpreted as generalizations of two previous approaches due to Gerven-Compernolle [7] and Nguyen-Jutten [5]. Since they are of the Gauss-Newton type, they exhibit a fast rate of convergence. Also, we present a stability analysis for two sources and instantaneous mixtures where we show that the algorithms cannot converge to non-separating solutions.

1. INTRODUCTION

Blind Source Separation (BSS) is a fundamental problem in signal processing with a large number of applications in speech processing, array signal processing, multiuser communications, etc. [2]. If we assume that several statistically independent non-Gaussian sources are transmitted through a Multiple Input Multiple Output (MIMO) Linear and Time Invariant (LTI) system, therefore arriving at an array of sensors in the form of a convolutive mixture, the BSS problem consists in recovering the original sources from the observations only, without knowing the transmission channel and the sources.

Most of the existing approaches to BSS have been developed for instantaneous mixtures. In practical situations, however, this hypothesis is not true and it is more common to find convolutive mixtures. Different algorithms [5][7] have been proposed to separate convolutive mixtures of sources in a blind way. In this paper we present a new approach based on the minimization of a statistical dependence measure by means of Gauss-Newton algorithms.

This paper is organized as follows. The signal model is presented in Section 2. Section 3 develops the proposed method and derives two simplified versions. It also revisits existing algorithms [5][7] and interprets them as particular cases of the proposed one. A convergence analysis for two sources and instantaneous mixtures will be presented in Section 4. Section 5 shows the results of computer simulations and Section 6 is devoted to the conclusions.

2. SIGNAL MODEL

Let us consider N non-Gaussian and statistically independent sources $\mathbf{s}[n]=[s_1[n],\ldots,s_N[n]]^T$ that are mixed through a MIMO LTI system with memory. The elements of the vector of $M \geq N$ observed signals $\mathbf{x}[n]=[x_1[n],\ldots,x_M[n]]^T$ are given by

$$x_i[n]=\sum_{j=1}^{N}\sum_{k=-L_{a1}}^{L_{a2}} a_{ij}[k]s_j[n-k].$$

This mixture is characterized by the system's transfer matrix

$$\mathbf{A}(z)=\left[\sum_{k=-L_{a1}}^{L_{a2}} a_{ij}[k]z^{-k}\right]_{ij}.$$

In order to separate the sources, we introduce a MIMO LTI system with memory $\mathbf{B}(z)=\left[\sum_{k=-L_{b1}}^{L_{b2}} b_{ij}[k]z^{-k}\right]_{ij}$ to produce the vector of separated signals $\mathbf{y}[n]=[y_1[n],\ldots,y_N[n]]^T$ where

$$y_i[n]=\sum_{j=1}^{M}\sum_{k=-L_{b1}}^{L_{b2}} b_{ij}[k]x_j[n-k].$$

We will also make the following assumptions throughout this paper:

AS1. The mixture is *soft* in the sense that each observed signal has a dominant contribution from a different source. This hypothesis is valid when each sensor is closer to one source than to the others.

AS2. Sources are *almost* identically distributed in the sense that cumulants of the same order do not exhibit large differences in magnitude.

As pointed out by several authors [2][6], the BSS problem has multiple valid solutions since sources can be separated with arbitrary orders, delays and scale factors. To avoid the last two indeterminacies we will assume $diag(\mathbf{B}(z))=diag(\mathbf{I})$ and $diag(\mathbf{A}(z)\mathbf{P})=diag(\mathbf{I})$, where $\mathbf{I}$ and $\mathbf{P}$ are the identity and an unknown permutation matrix, respectively.

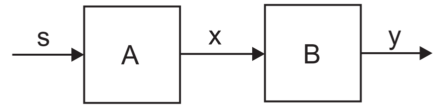

Figure 1. Feed-forward separation model.

3. SEPARATION ALGORITHM

The BSS problem can be solved relying only on the non-Gaussianity and statistical independence of the sources because, under these hypotheses, it has been shown in [2] that the sources are separated if and only if the components of $\mathbf{y}[n]$ become pair-wise independent. When this occurs, all the cross-cumulants between different outputs vanish. This idea suggests that the separating system be selected to minimize a weighted sum of output cross-cumulants.

Let us denote $C^{x_{i_1},y_{i_2},\ldots,z_{i_q}}_{\alpha_1,\alpha_2,\ldots,\alpha_q}(k_1,\ldots,k_{(q-1)})$ = $Cum(x_{i_1}[n]\times\alpha_1, y_{i_2}[n-k_1]\times\alpha_2,\ldots,z_{i_q}[n-k_{(q-1)}]\times\alpha_q)$ the $(i_1,\ldots,i_q)$-order cross-cumulant of $x_{i_1}, y_{i_2},\ldots,z_{i_q}$. Similarly, let us define $C^{\mathbf{x},\mathbf{y},\ldots,\mathbf{z}}_{\alpha_1,\alpha_2,\ldots,\alpha_q}(k_1,\ldots,k_{(q-1)})$ as the tensor whose elements at the $(i_1,\ldots,i_q)$ position are $C^{x_{i_1},y_{i_2},\ldots,z_{i_q}}_{\alpha_1,\alpha_2,\ldots,\alpha_q}(k_1,\ldots,k_{(q-1)})$.

Let us consider now the following set of cross-cumulant matrices of the output vector

$$\Pi = \left\{\mathbf{C}^{\mathbf{y},\mathbf{y}}_{\alpha,\beta}[k] : \ k=-L_{b1},\ldots,L_{b2}\ ;\ (\alpha,\beta)\in\Omega\right\}$$

(Ω is a set of n_Ω pairs of natural numbers). It is apparent that these matrices become diagonal when $\mathbf{y}[n]$ is a vector of statistically independent signals. Therefore, we propose as a criterion to achieve source separation the joint diagonalization of Π. This diagonalization is equivalent to minimizing the following dependence measure

$$\Phi_\Omega = \sum\nolimits_{(\alpha,\beta)\in\Omega} w_{(\alpha,\beta)}\Phi_{(\alpha,\beta)} \qquad (1)$$

$$\Phi_{(\alpha,\beta)} = \sum\nolimits_k \left\| \mathbf{C}^{\mathbf{y},\mathbf{y}}_{\alpha,\beta}[k] - \Lambda^{\mathbf{y},\mathbf{y}}_{\alpha,\beta}[k] \right\|^2_F \qquad (2)$$

being $w_{(\alpha,\beta)}$ a set of weighting coefficients, $\|\cdot\|_F$ the Frobenius norm and $\Lambda^{\mathbf{y},\mathbf{y}}_{\alpha,\beta}[k]$ a temporal sequence of diagonal matrices defined as $\Lambda^{\mathbf{y},\mathbf{y}}_{\alpha,\beta}[k] = diag\left(diag\left(\mathbf{C}^{\mathbf{y},\mathbf{y}}_{\alpha,\beta}[k]\right)\right)$. The dependence measure (1) is the weighted square distance between the set of cross-cumulant matrices Π and its diagonal counterpart $\Pi_{diag} = \left\{\Lambda^{\mathbf{y},\mathbf{y}}_{\alpha,\beta}[k] : \ k=-L_{b1},\ldots,L_{b2}\ ;\ (\alpha,\beta)\in\Omega\right\}$.

3.1 Minimization algorithm

Many different algorithms can be used to minimize Φ_Ω. Among them we choose a Gauss-Newton (GN) one since it will never converge to a local minimum and exhibits quadratic convergence. In order to derive the GN iteration, let us rearrange the variables in vector notation. If we denote $vec(\cdot)$ as the correspondence that assigns non-diagonal elements of a $N\times Q$ matrix (with $N\le Q$) to a vector $(N-1)Q\times 1$ indexing first by columns and then by rows, we can define the following vectors

$$\mathbf{z}_{(\alpha,\beta)} = \left[vec\left(\mathbf{C}^{\mathbf{y},\mathbf{y}}_{\alpha,\beta}[-L_{b1}]\right)^T,\ldots,vec\left(\mathbf{C}^{\mathbf{y},\mathbf{y}}_{\alpha,\beta}[L_{b2}]\right)^T \right]^T \qquad (3)$$

$$\mathbf{z} = \left[\sqrt{w_{(\alpha,\beta)_1}}\,\mathbf{z}^T_{(\alpha,\beta)_1},\ldots,\sqrt{w_{(\alpha,\beta)_{n_\Omega}}}\,\mathbf{z}^T_{(\alpha,\beta)_{n_\Omega}} \right]^T \qquad (4)$$

$$\mathbf{b} = \left[vec\left(\mathbf{B}[-L_{b1}]\right)^T,\ldots,vec\left(\mathbf{B}[L_{b2}]\right)^T \right]^T \qquad (5)$$

Substituting (3),(4) and (5) in (1), the dependence measure can be written as the inner product $\Phi_\Omega = \mathbf{z}^T\mathbf{z}$. Denoting the Jacobian matrix as $\mathbf{J} = \nabla_{\mathbf{b}}\mathbf{z}$, the GN iteration will consist in updating the separation coefficients vector as

$$\mathbf{b}^{(n+1)} = \mathbf{b}^{(n)} + \mu\,\Delta \qquad (6)$$

where $0<\mu<2$ is the adaptation step and Δ is the result of solving the following system of linear equations

$$\mathbf{J}\mathbf{J}^T\ \Delta = -\mathbf{J}\,\mathbf{z} \qquad (7)$$

3.2 Simplifications based on the Jacobian matrix structure

The main limitation of the above algorithm is the computational complexity required by the evaluation of the Jacobian matrix for arbitrary pairs (α,β), i.e., the order of the cross-cumulant involved in the dependence measure. Initially, every pair of natural numbers (α,β) could belong to Ω. However, the algorithm complexity can be considerably reduced if we choose $\alpha=1$ and $\beta>1$. In this case, the dependence measure is approximately quadratic in the vicinity of the separation solution.

Under the assumptions AS1 and AS2, for $\alpha=1$, $\beta>1$ and stationary sources which are white sequences (or equivalently, when there are instantaneous mixtures) it is shown in [3] that the elements of the Jacobian matrix can be approximated by

$$\frac{\partial C^{y_i,y_j}_{1,\beta}[k]}{\partial b_{rs}[m]} \approx \delta_{ir}\delta_{km} C^{x_s,y_j}_{1,\beta}[0] \qquad (8)$$

where δ_{ij} is the Kronecker operator ($\delta_{ij}=1$ if $i=j$, $\delta_{ij}=0$ otherwise). Taking into account these approximations, the proposed algorithm in (6) for two white sources and two sensors reduces to the following adaptation rule

$$b^{(n+1)}_{i,j}[k] = b^{(n)}_{i,j}[k] - \mu \frac{\sum\limits_{(1,\beta)\in\Omega} w_{(1,\beta)} C^{x_j,y_j}_{1,\beta}[0]\, C^{y_i,y_j}_{1,\beta}[k]}{\sum\limits_{(1,\beta)\in\Omega} w_{(1,\beta)} (C^{x_j,y_j}_{1,\beta}[0])^2} \qquad (9)$$

with $i,j\big|_{j\neq i} = 1,2;$ and $k=-L_{b1},\ldots,L_{b2}$. Recall that iteration (9) only requires $2n_\Omega(L_{b1}+L_{b2}+1)$ cumulants!

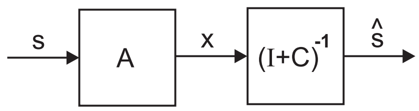

Figure 2. Feed-backward separation model.

For more sources and sensors, we observe from (8) that the Jacobian $\mathbf{J}$ is a sparse matrix and that there exists efficient methods [4] to solve linear sparse systems such as (7).

3.3 Feed-backward algorithm

The algorithm (6) was obtained for the feed-forward separation system represented in Figure 1. A different version of this algorithm can be obtained for the feed-backward separation structure plotted in Figure 2. It has been shown in [3] that changing the sign of the adaptation coefficients $\mathbf{b}$ and replacing the feed-forward outputs $\mathbf{y}[n]$ by the feed-backward ones $\hat{\mathbf{s}}[n]$ the simplified adaptation rule (9) is converted into the following one

$$c_{i,j}^{(n+1)}[k]=c_{i,j}^{(n)}[k]+\mu\frac{\sum_{(1,\beta)\in\Omega} w_{(1,\beta)} C_{1,\beta}^{x_j,\hat{s}_j}[0]\, C_{1,\beta}^{\hat{s}_i,\hat{s}_j}[k]}{\sum_{(1,\beta)\in\Omega} w_{(1,\beta)} \left(C_{1,\beta}^{x_j,\hat{s}_j}[0]\right)^2} \qquad (10)$$

where $i,j\big|_{j\neq i}=1,2;$ and $k=-L_{b1},\ldots,L_{b2}$. When $n_\Omega=1$, this expression reduces to

$$c_{i,j}^{(n+1)}[k]=c_{i,j}^{(n)}[k]+\mu\frac{C_{1,\beta}^{\hat{s}_i,\hat{s}_j}[k]}{C_{1,\beta}^{x_j,\hat{s}_j}[0]}; \qquad \begin{array}{l} i,j\big|_{j\neq i}=1,2; \\ k=-L_{b1},\ldots,L_{b2}; \end{array} \qquad (11)$$

This feed-backward simplified algorithm allows us to establish a link with existing BSS methods. When $\Omega=\{(1,1)\}$, i.e. $\beta=1$, the algorithm (11) reduces to the Symmetric Adaptive Decorrelation (SAD) algorithm proposed in [7]. On the other hand, for $\beta=3$ the algorithm (11) is very similar to the Nguyen-Jutten (NJ**)** method proposed in [5]. Moreover, our approach presents several advantages with respect to the NJ algorithm: it exhibits greater speed of convergence since it corresponds to a Gauss-Newton method and it is asymptotically stable (as shown in the following section). Finally, it is interesting to note that, in contrast to the existing approaches [5][7], the new algorithm performs adequately for non-causal mixtures.

4. CONVERGENCE ANALYSIS

The dependence measure Φ_Ω to be minimized in our method is not a quadratic form of the separating system coefficients and thus may contain undesired minima that will impair the convergence of the adaptive algorithms. In the sequel we present a convergence analysis to show that, even though these minima may exist (depending on the chosen set Ω), they are not stable points. Since a general proof is rather involved, we will consider the simpler case of two sources, two sensors and instantaneous mixtures. In this case, the feed-forward algorithm (9) reduces to

$$b_{i,j}^{(n+1)}=b_{i,j}^{(n)}-\mu\frac{C_{1,\beta}^{y_i,y_j}}{C_{1,\beta}^{x_j,y_j}}; \qquad i,j\big|_{j\neq i}=1,2; \qquad (12)$$

In the sequel we will demonstrate that when β is an odd number greater than one, the adaptation rule (12) may exhibit several non-separating stationary points but the only stable stationary points correspond to a separating solution.

Let $\mathbf{H}=\mathbf{BA}$ be the global system transfer matrix ($\mathbf{y}=\mathbf{H}\,\mathbf{s}$). The stationary points of Φ_Ω occur when the updating term in (12) vanishes, i.e., when

$$C_{1,\beta}^{y_i,y_j}=h_{ii}h_{ji}^{\beta}C_{\beta+1}^{s_i}+h_{ij}h_{jj}^{\beta}C_{\beta+1}^{s_j}=0; \quad i,j\big|_{j\neq i}=1,2;$$

Assuming that β is chosen so that the cumulants of the sources $C_{\beta+1}^{s_i}$ and $C_{\beta+1}^{s_j}$ are non-zero, it is straightforward to show that the stationary points in (12) can be classified into the following groups:

1. $\mathbf{H}$ is a (possibly scaled) permutation matrix.
2. $\mathbf{H}$ is singular with at least one null row.
3. $h_{ij}=\pm\eta\cdot h_{ii}$ and $h_{jj}=\eta\cdot h_{ji}$ with $\eta^{\beta+1}=\pm\dfrac{C_{\beta+1}^{s_i}}{C_{\beta+1}^{s_j}}$.

The first case corresponds to the separation solution. These stationary points are asymptotically stable (if $0<\mu<2$) because the approximation (8) for the Jacobian matrix becomes exact and the algorithm coincides with a Gauss-Newton minimization method. This circumstance does not hold for the non-separating stationary points. Note that in the second case, one of the outputs is zero, the denominator of the updating term in (12) vanishes and the algorithm becomes unstable. For the third case, the stability ODE analysis has been done leading to the following characteristic polynomial in λ

$$\left|\mathbf{I}-\mu\cdot\mathbf{J}\right|=\lambda^2-2(1-\mu)\lambda+(1-\mu^2)-\mu^2\beta^2\frac{C_{1,1,\beta-1}^{x_i,y_i,y_j}C_{1,1,\beta-1}^{x_j,y_j,y_i}}{C_{1,\beta}^{x_j,y_j}C_{1,\beta}^{x_i,y_i}}$$

However, since β was chosen to be odd, this expression is simplified to

$$\left|\mathbf{I}-\mu\cdot\mathbf{J}\right|=\lambda^2-2(1-\mu)\lambda+(1-\mu^2)-\mu^2\beta^2$$

Now, we can use the Anderson-Jury inequalities [1] to check if the roots are within the stability region. These are $\beta^2<1$ and $0<\mu<\dfrac{2}{1+\beta}$. Then, since by hypothesis $\beta>1$, we have reached a contradiction and the algorithm will also be unstable in the third case.

5. COMPUTER SIMULATIONS

In this section we present the results of several computer experiments carried out to illustrate the performance of the proposed algorithms.

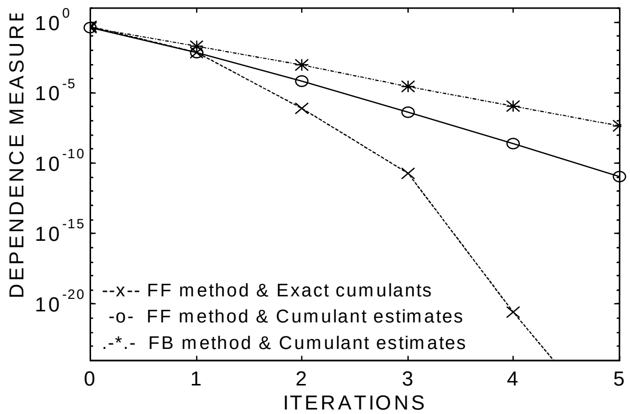

Figure 3. Dependence Measure Φ_Ω versus Iterations.

We have considered two independent uniformly distributed white sources which are mixed through the convolutive channel

$$\mathbf{A}(z) = \begin{pmatrix} 1 & -0.19 + 0.05z^{-1} + 0.00z^{-2} + 0.20z^{-3} + 0.69z^{-4} \\ -0.52 + 0.38z^{-1} - 0.32z^{-2} - 0.04z^{-3} - 0.33z^{-4} & 1 \end{pmatrix}$$

It is apparent that the mixing matrix satisfies the soft mixture condition. We applied the GN feed-forward algorithm (9) with $\Omega = \{(1,3)\}$. Cumulants were estimated from 5000 samples of the observations. After 2 iterations we arrived at the separating matrix

$$\mathbf{B}(z) = \begin{pmatrix} 1 & 0.190 - 0.043z^{-1} - 0.011z^{-2} - 0.189z^{-3} - 0.706z^{-4} \\ 0.512 - 0.375z^{-1} + 0.323z^{-2} + 0.028z^{-3} + 0.347z^{-4} & 1 \end{pmatrix}$$

which is close to the optimum separation matrix

$$\mathbf{B}(z) = \begin{pmatrix} 1 & -A_{12}(z) \\ -A_{21}(z) & 1 \end{pmatrix}.$$

To measure the performance of the proposed method we use $\mathbf{H}_E = \left[\sum_k \left|h_{ij}[k]\right|^2\right]_{ij}$, the energy matrix of the overall transfer matrix $\mathbf{H}$. This matrix is normalized as follows in order that the diagonal elements are equal to one

$$\overline{\mathbf{H}}_E = \tfrac{1}{2}\left(\mathbf{D}_r^{-1}\mathbf{H}_E + \mathbf{H}_E\mathbf{D}_c^{-1}\right)$$

$$\mathbf{D}_r = diag\left(\max_{rows}\{\mathbf{H}_E\}\right), \qquad \mathbf{D}_c = diag\left(\max_{colums}\{\mathbf{H}_E\}\right)$$

Perfect separation of the sources is achieved when $\overline{\mathbf{H}}_E$ is the identity matrix. In our computer experiment the initial and final normalized energy matrices of the overall transfer function were

- *Initial Energy Matrix:* $\overline{\mathbf{H}}_E = \begin{pmatrix} 1 & 0.6277 \\ 0.5547 & 1 \end{pmatrix}$
- *Final Energy Matrix:* $\overline{\mathbf{H}}_E = \begin{pmatrix} 1 & 0.0005 \\ 0.0005 & 1 \end{pmatrix}$

It is clearly seen that an exact separation is almost reached since the final energy matrix is close to the identity matrix. A residual value appears in the non-diagonal elements because we are using estimates of the cumulants instead of the true values.

Figure 3 illustrates the algorithm rate of convergence by presenting the dependence measure Φ_Ω versus iterations. The dotted line represents the case of using true cross-cumulant values (which corresponds to a quadratic convergence) whereas the solid line represents the case of using cumulant estimates (which corresponds to a linear convergence). Finally, the dotted-pointed line represents the case of applying the feed-backward algorithm (11) using cumulant estimates. Observe that source separation was achieved in all cases.

6. CONCLUSIONS

A new blind source separation algorithm for convolutive soft mixtures has been presented. It is based on a Gauss-Newton algorithm that minimizes a dependence measure that involves cross-cumulants of the outputs. Exploiting the Jacobian matrix structure, we found simplified versions of the algorithm with lesser computational complexity. We also showed that existing methods [5][7] can be interpreted as particular cases of our approach. The proposed algorithms exhibit a fast convergence (typically three iterations) because they are of the Gauss-Newton type. A convergence analysis was also carried out in the simplified case of instantaneous mixtures of two sources and it is shown that the unique stable stationary point of the feed-forward algorithm corresponds to the separation solution.

7. REFERENCES

[1] Ackerman J.,"Sampled-Data Control Systems: Analysis and Synthesis, Robust Systems Design". Springer-Verlag, 1995.

[2] Comon P., "Independent component analysis, a new concept? ". *Signal Processing*, **36**, 287-314 (1994).

[3] Cruces S. and Castedo L. "Blind separation of convolutive mixtures: A Gauss-Newton algorithm" *Proc. of the IEEE SP Workshop HOS'97*, pp. 326-330,Canada, July 1997.

[4] Golub G.H., Van Loan C.F., "Matrix computations", *Johns Hopkins University Press*, third edition, 1996.

[5] Nguyen-Thi H.L. and Jutten C., "Blind source separation for convolutive mixtures" *Signal Processing.* **45**, pp.209-229, 1995.

[6] Tong L., Liu R.W., Soon V.C., Huang Y.-F. "Indeterminacy and identifiability of blind identification" *IEEE Trans. on Circuits and Systems*, **38**, 5, pp. 499-509, 1991.

[7] Van-Gerven S. and Van-Compernolle D. "Signal separation by symmetric adaptive decorrelation: Stability, convergence and uniqueness" IEEE Trans. on Signal Processing, **43**, 7, pp.1602-1612, 1995.

† This work has been supported by the Spanish National Research Plan, CICYT (Grant No. TIC96-0500-C10-08).

‡ This work has been supported by the Grants: CICYT (TIC96-0500-C10-02) and Xunta de Galicia (XUGA 10502A96).