Blind maximum likelihood separation of a linear-quadratic mixture

Shahram Hosseini and Yannick Deville

Université Paul Sabatier, Laboratoire d'Acoustique, Métrologie, Instrumentation
Bat. 3R1B2, 118 route de Narbonne, 31062 Toulouse Cedex, France
hosseini@cict.fr , ydeville@cict.fr

Abstract. We proposed recently a new method for separating linear-quadratic mixtures of independent real sources, based on parametric identification of a recurrent separating structure using an *ad hoc* algorithm. In this paper, we develop a maximum likelihood approach providing an asymptotically efficient estimation of the model parameters. A major advantage of this method is that the explicit form of the inverse of the mixing model is not required to be known. Thus, the method can be easily generalized to more complicated polynomial mixtures.

1 Introduction

Little work has been dedicated to the BSS problem in nonlinear mixtures [1]-[8]. It is well known [1], [2] that the independence hypothesis is not sufficient for separating general nonlinear mixtures because of the very large indeterminacies which make the nonlinear BSS problem ill-posed. A natural idea for reducing the indeterminacies is to constrain the structure of mixing and separating models to belong to a certain set of transformations. This supplementary constraint can be viewed as a regularization of the initially ill-posed problem [5], [8].

In a recent paper [7], we studied a linear-quadratic mixture model which may be considered as the simplest (nonlinear) version of a general polynomial model. Our main aim is to develop an approach which can be easily extended to higher-order polynomial models. Hence, in [7] we proposed a recurrent separating structure whose realization does not require the knowledge of the explicit form of the inverse of the mixing model. A drawback of the proposed approach was that somewhat heuristic criteria had been chosen to identify the model parameters. In the present paper, we develop a rigorous method to identify the parameters of the separating structure in a maximum likelihood framework. Once more, the algorithm is developed so that the inverse of the mixing structure is not required to be known. Thus, it can be extended to more general polynomial mixtures.

2 mixing and separating models

Suppose u_1 and u_2 are two independent random signals. Given the following nonlinear instantaneous mixture model

$$x_i = a_{i1}u_1 + a_{i2}u_2 + b_i u_1 u_2 \quad i = 1, 2 \tag{1}$$

we would like to estimate u_1 and u_2 up to a permutation and a scaling factor (and possibly an additive constant). For simplicity, let's denote $s_1 = a_{11}u_1$ and $s_2 = a_{22}u_2$. s_1 and s_2 will be referred to as the *sources* in the following. (1) can be rewritten as

$$\begin{aligned} x_1 &= s_1 - l_1 s_2 - q_1 s_1 s_2 \\ x_2 &= s_2 - l_2 s_1 - q_2 s_1 s_2 \end{aligned} \tag{2}$$

in which $l_1 = -a_{12}/a_{22}$ and $l_2 = -a_{21}/a_{11}$ represent the linear contributions of the sources in the mixture, and $q_1 = -b_1/(a_{11}a_{22})$ and $q_2 = -b_2/(a_{11}a_{22})$ represent the quadratic contributions. The negative signs are chosen for simplifying the notations of the separating structure.

A more general form of the model (2), containing the additional terms s_1^2 and s_2^2, has been studied by a few authors [9], [10], for the special case of *circular* complex sources, when at least 5 mixtures are available. In the current work, however, we suppose that: 1) the sources are arbitrary real signals, and 2) only two mixtures are available.

The invertibility of the model (2) was briefly discussed in [7]. Here, we present a more rigorous analysis of this subject. Solving the model (2) for s_1 and s_2 leads to the following two pairs of solutions [7], which may be considered as two direct separating structures:

$$\begin{aligned} (\int_1, \int_2)_1 &= ((-b_1 + \sqrt{\Delta_1})/2a_1, (-b_2 + \sqrt{\Delta_2})/2a_2) \\ (\int_1, \int_2)_2 &= ((-b_1 - \sqrt{\Delta_1})/2a_1, (-b_2 - \sqrt{\Delta_2})/2a_2) \end{aligned} \tag{3}$$

where $\Delta_i = b_i^2 - 4a_i c_i$, $a_1 = q_2 + l_2 q_1$, $a_2 = q_1 + l_1 q_2$, $b_1 = q_1 x_2 - q_2 x_1 + l_1 l_2 - 1$, $b_2 = q_2 x_1 - q_1 x_2 + l_1 l_2 - 1$, $c_1 = x_1 + l_1 x_2$ and $c_2 = x_2 + l_2 x_1$. It can be easily verified that $\Delta_1 = \Delta_2 = J^2$, where J is the Jacobian of the mixing model (2) and reads

$$J = 1 - l_1 l_2 - (q_2 + l_2 q_1)s_1 - (q_1 + l_1 q_2)s_2 \tag{4}$$

According to the variation domain of the two sources, three different cases may be considered:

1) $J < 0$ for all the values of s_1 and s_2. In this case (3) becomes:

$$(\int_1, \int_2)_1 = (s_1, s_2) \tag{5}$$

$$(\int_1, \int_2)_2 = (-\frac{q_1 + l_1 q_2}{q_2 + l_2 q_1} s_2 - \frac{l_1 l_2 - 1}{q_2 + l_2 q_1}, -\frac{q_2 + l_2 q_1}{q_1 + l_1 q_2} s_1 - \frac{l_1 l_2 - 1}{q_1 + l_1 q_2}) \tag{6}$$

Thus, the first direct separating structure in (3) leads to the actual sources and the second direct separating structure leads to another solution, equivalent to the first one up to a permutation, a scaling factor, and an additive constant.

2) $J > 0$ for all the values of s_1 and s_2. In this case, the first structure leads to the permuting solution, defined by (6), and the second structure to the actual sources (s_1, s_2). An example is shown in Fig. 1 for the numerical values $l_1 = -0.2$, $l_2 = 0.2$, $q_1 = -0.8$, $q_2 = 0.8$ and $s_i \in [-0.5, 0.5]$.

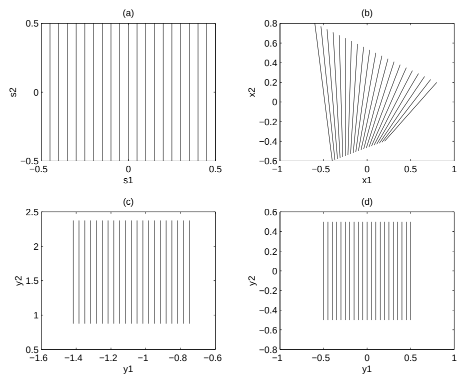

Fig. 1. Case when $J > 0$ for all the source values. Distribution of (a) sources, (b) mixtures, (c) output of the first direct separating structure, (d) output of the second direct separating structure.

3) $J > 0$ for some values of the sources and $J < 0$ for the other values. In this case, each structure leads to the non-permuted sources (5) for some values of the observations and to the permuted sources (6) for the other values. An example is shown in Fig. 2 (with the same coefficients as in the second case, but for $s_i \in [-2, 2]$). The permutation effect is clearly visible in the figure. One may also remark that the straight line $J = 0$ in the source plane is mapped to a conic section in the observation plane (shown by asterisks).

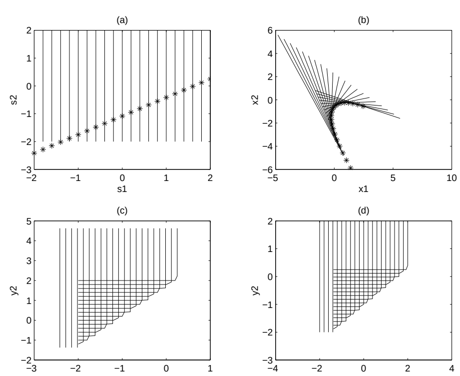

Fig. 2. Case when $J > 0$ for some values of the sources and $J < 0$ for the other values. Distribution of (a) sources, (b) mixtures, (c) output of the first direct separating structure, (d) output of the second direct separating structure.

Thus, it is clear that the direct structures may be used for separating the sources if the Jacobian of the mixing model is always negative or always positive, *i.e.* for all the source values. Otherwise, although the sources are separated *sample by sample*, each retrieved signal contains samples of the two sources. This problem arises because the mixing model (2) is not bijective. This theoretically insoluble problem should not discourage us. In fact, our final objective is to extend the idea developed in the current study to more general polynomial models which will be used to approximate the nonlinear mixtures encountered in the real world. If these real-world nonlinear models are bijective, we can logically suppose that the coefficients of their polynomial approximations take values which make them bijective on the variation domains of the sources. Thus, in the following, we suppose that the sources and the mixture coefficients have numerical values ensuring that the Jacobian J of the mixing model has a constant sign.

The natural idea to separate the sources is to form a direct separating structure using any of the equations in (3), and to identify the parameters l_1, l_2, q_1 and q_2 by optimizing an independence measuring criterion. Although this approach may be used for our special mixing model (2), as soon as a more complicated polynomial model is considered, the solutions $(\int_1, \int_2)$ can no longer be determined so that the generalization of the method to arbitrary polynomial models seems impossible. To avoid this limitation, we propose a recurrent structure. Such structures have been considered since the early work of Hérault and Jutten [11] and then in more complex configurations [12], [13]. We here extend them to linear-quadratic mixtures by introducing the structure shown in Fig. 3. Note that, for $q_1 = q_2 = 0$, this structure is reduced to the basic Hérault-Jutten network. It may be checked easily that, for fixed observations defined by (2), $y_1 = s_1$ and $y_2 = s_2$ corresponds to a steady state for the structure in Figure 3.

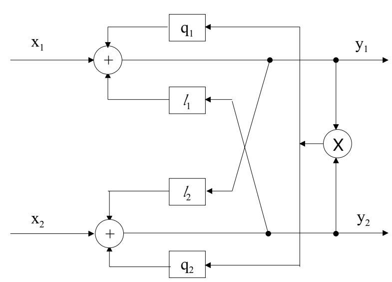

Fig. 3. Recurrent separating structure.

The use of this recurrent structure is more promising because it can be easily generalized to arbitrary polynomial models. However, the main problem with this structure is its stability. In fact, even if the mixing model coefficients are exactly known, the computation of the structure outputs requires the realization of the following recurrent iterative model

$$
\begin{aligned}
y_1(n+1) &= x_1 + l_1 y_2(n) + q_1 y_1(n) y_2(n) \\
y_2(n+1) &= x_2 + l_2 y_1(n) + q_2 y_1(n) y_2(n)
\end{aligned}
\qquad (7)
$$

where a loop on n is performed for each couple of observations (x_1, x_2) until convergence is achieved.

In [7], we have studied the local stability of the model (7) and shown that this model is locally stable at the separating point $(y_1, y_2) = (s_1, s_2)$, if and only if the absolute values of the two eigenvalues of the Jacobian matrix of (7) are smaller than one. In the following, we suppose that this condition is satisfied.

3 Maximum likelihood estimation of the model parameters

Let $f_{S_1,S_2}(s_1, s_2)$ be the joint pdf of the sources, and assume that the mixing model is bijective so that the Jacobian of the mixing model has a constant sign on the variation domain of the sources. The joint pdf of the observations can be written as

$$f_{X_1,X_2}(x_1, x_2) = \frac{f_{S_1,S_2}(s_1, s_2)}{|J(s_1, s_2)|} \tag{8}$$

Taking the logarithm of (8), and considering the independence of the sources, we can write:

$$\log f_{X_1,X_2}(x_1, x_2) = \log f_{S_1}(s_1) + \log f_{S_2}(s_2) - \log |J(s_1, s_2)| \tag{9}$$

Given N samples of the mixtures X_1 and X_2, we want to find the maximum likelihood estimator for the mixture parameters $\mathbf{w} = [l_1, l_2, q_1, q_2]$. This estimator is obtained by maximizing the joint pdf of all the observations (supposing that the parameters in $\mathbf{w}$ are constant), which is equal to

$$E = f_{X_1,X_2}(x_1(1), x_2(1), \cdots, x_1(N), x_2(N)) \tag{10}$$

If $s_1(t)$ and $s_2(t)$ are two i.i.d. sequences, $x_1(t)$ and $x_2(t)$ are also i.i.d. so that $E = \prod_{i=1}^{N} f_{X_1,X_2}(x_1(i), x_2(i))$ and $\log E = \sum_{i=1}^{N} \log f_{X_1,X_2}(x_1(i), x_2(i))$. The cost function to be maximized can be defined as $L = \frac{1}{N} \log E$, which will be denoted using the temporal averaging operator $E_t[.]$ as

$$L = E_t[\log f_{X_1,X_2}(x_1(t), x_2(t))] \tag{11}$$

Using (9):

$$L = E_t[\log f_{S_1}(s_1(t))] + E_t[\log f_{S_2}(s_2(t))] - E_t[\log |J(s_1(t), s_2(t))|] \tag{12}$$

Maximizing this cost function requires that its gradient with respect to the parameter vector $\mathbf{w}$, *i.e.* $\frac{\partial L}{\partial \mathbf{w}}$, vanishes. Defining the score functions of the two sources as

$$\psi_i(u) = -\frac{\partial \log f_{S_i}(u)}{\partial u} \qquad i = 1, 2 \tag{13}$$

and considering that $\frac{\partial \log |J|}{\partial \mathbf{w}} = \frac{1}{J} \frac{\partial J}{\partial \mathbf{w}}$, we can write

$$\frac{\partial L}{\partial \mathbf{w}} = -E_t[\psi_1(s_1) \frac{\partial s_1}{\partial \mathbf{w}}] - E_t[\psi_2(s_2) \frac{\partial s_2}{\partial \mathbf{w}}] - E_t[\frac{1}{J} \frac{\partial J}{\partial \mathbf{w}}] \tag{14}$$

Rewriting (2) in the vector form $\mathbf{x} = \mathbf{f}(\mathbf{s}, \mathbf{w})$ and considering $\mathbf{w}$ as the independent variable and $\mathbf{s}$ as the dependent variable, we can write, using implicit differentiation

$$\mathbf{0} = \frac{\partial \mathbf{f}}{\partial \mathbf{s}} \frac{\partial \mathbf{s}}{\partial \mathbf{w}} + \frac{\partial \mathbf{f}}{\partial \mathbf{w}} \tag{15}$$

which yields

$$\frac{\partial \mathbf{s}}{\partial \mathbf{w}} = -(\frac{\partial \mathbf{f}}{\partial \mathbf{s}})^{-1} \frac{\partial \mathbf{f}}{\partial \mathbf{w}} \tag{16}$$

Note that $\frac{\partial \mathbf{f}}{\partial \mathbf{s}}$ is the Jacobian matrix of the mixing model. Using (14) and (16), the gradient of the cost function L with respect to the parameter vector $\mathbf{w}$ is equal to (see the appendix for the computation details)

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{w}} = -E_t\Big[\Big(&\psi_1(s_1)(1 - q_2 s_1)s_2 + \psi_2(s_2)(l_2 + q_2 s_2)s_2 - (l_2 + q_2 s_2)\Big)/J, \\ &\Big(\psi_1(s_1)(l_1 + q_1 s_1)s_1 + \psi_2(s_2)(1 - q_1 s_2)s_1 - (l_1 + q_1 s_1)\Big)/J, \\ &\Big(\psi_1(s_1)(1 - q_2 s_1)s_1 s_2 + \psi_2(s_2)(l_2 + q_2 s_2)s_1 s_2 - (l_2 s_1 + s_2)\Big)/J, \\ &\Big(\psi_1(s_1)(l_1 + q_1 s_1)s_1 s_2 + \psi_2(s_2)(1 - q_1 s_2)s_1 s_2 - (s_1 + l_1 s_2)\Big)/J\Big] \end{aligned} \tag{17}$$

In practice, the actual sources and their density functions are unknown and will be replaced by the reconstructed sources, *i.e.* by the outputs of the separating structure of Fig 3, y_i, in an iterative algorithm. The score functions of the reconstructed sources can be estimated by any of the existing parametric or non-parametric methods. In our work, we used the kernel estimator proposed in [14] based on third-order cardinal splines. Using (17), the cost function (12) can be maximized by a gradient ascent algorithm which updates the parameters by the rule $\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \frac{\partial L}{\partial \mathbf{w}}$. The learning rate parameter μ must be chosen carefully to avoid the divergence of the algorithm. Note that the algorithm does not require the knowledge of the explicit inverse of the mixing model (direct separating structures (3)). Hence, it can be easily extended to more general polynomial mixing models.

4 Simulation results

The algorithm was tested using different combinations of subgaussian and supergaussian sources, where the subgaussien sources were uniformly distributed on $[-0.5, 0.5]$ and the supergaussian sources were laplacian with pdf $f_S(s) = 5exp(-10|s|)$. The distribution of the mixtures for two uniform sources is like that presented in Fig. 1.b. The distribution of the estimated sources y_1 and y_2 applying our algorithm is shown in Fig. 4. The rectangular shape of this distribution indicates that the independent components are retrieved. Table 1 represents the output Signal to Noise Ratio, defined as $SNR = 0.5 \sum_{i=1}^{2} 10 \log_{10} \frac{E[s_i^2]}{E[(y_i - s_i)^2]}$ achieved by our algorithm for 3 different combinations of the sources. In each case, the experiment was repeated 100 times, corresponding to different seed values of the random variable generator, using 1000 samples of the sources. The results confirm the good performance of the algorithm.

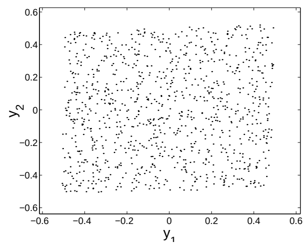

Fig. 4. Distribution of the estimated sources.

	Mean(SNR)	STD(SNR)
s_1 and s_2 uniform	28.0	4.2
s_1 uniform, s_2 laplacian	27.8	3.8
s_1 and s_2 laplacian	26,8	3,1

Table 1. Mean and Standard Deviation of output SNR (in dB) for different combinations of the sources.

5 Conclusion

Nonlinear blind source separation is a difficult, little studied problem. In this work, we investigated one of the simplest structured nonlinear models, *i.e.* the linear-quadratic model. As we aim at generalizing the idea developed in this study to more complicated polynomial models, we proposed a separating structure and an estimation method which do not make use of our knowledge on the explicit form of the inverse of the mixing model. The maximum likelihood approach, developed in this paper, provides an asymptotically efficient estimation of the model parameters and works very well in practice. Some of our objectives for completing this work are: a more precise stability analysis of the recurrent separating network, development of an equivariant estimating method using natural gradient, study of the separability problem, and generalizing the method to more complicated polynomial models and more sources.

Appendix: details of gradient computation

Considering (2), we can write
$\frac{\partial \mathbf{f}}{\partial \mathbf{s}} = \begin{pmatrix} 1 - q_1 s_2 & -l_1 - q_1 s_1 \\ -l_2 - q_2 s_2 & 1 - q_2 s_1 \end{pmatrix}$ and $\frac{\partial \mathbf{f}}{\partial \mathbf{w}} = \begin{pmatrix} -s_2 & 0 & -s_1 s_2 & 0 \\ 0 & -s_1 & 0 & -s_1 s_2 \end{pmatrix}$, which implies, from (16)

$$\frac{\partial \mathbf{s}}{\partial \mathbf{w}} = \frac{-1}{J} \begin{pmatrix} 1 - q_2 s_1 & l_1 + q_1 s_1 \\ l_2 + q_2 s_2 & 1 - q_1 s_2 \end{pmatrix} \cdot \begin{pmatrix} -s_2 & 0 & -s_1 s_2 & 0 \\ 0 & -s_1 & 0 & -s_1 s_2 \end{pmatrix}$$

which yields

$$\frac{\partial s_1}{\partial \mathbf{w}} = \frac{1}{J}\Big[(1-q_2s_1)s_2\ ,\ (l_1+q_1s_1)s_1\ ,(1-q_2s_1)s_1s_2\ ,\ (l_1+q_1s_1)s_1s_2\Big]$$
$$\frac{\partial s_2}{\partial \mathbf{w}} = \frac{1}{J}\Big[(l_2+q_2s_2)s_2\ ,\ (1-q_1s_2)s_1\ ,(l_2+q_2s_2)s_1s_2\ ,\ (1-q_1s_2)s_1s_2\Big] \quad (19)$$

Considering (4)

$$\frac{\partial J}{\partial \mathbf{w}} = -\Big[l_2+q_2s_2, l_1+q_1s_1, l_2s_1+s_2, s_1+l_1s_2\Big] \quad (20)$$

(17) follows directly from (14), (19) and (20).

References

1. A. Hyvarinen and P. Pajunen, Nonlinear independent component analysis: Existence and uniqueness results, *Neural Networks*, 12(3), pp. 429-439, 1999.
2. A. Taleb and C. Jutten, Source separation in post-nonlinear mixtures, *IEEE Trans. on Signal Processing*, 47(10), pp. 2807-2820, 1999.
3. L. Almeida, Linear and nonlinear ICA based on mutual information, In *Proc. IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (AS-SPCC)*, pp. 117-122, Lake Louise, Canada, October 2000.
4. J. Eriksson and V. Koivunen, Blind identifiability of class of nonlinear instantaneous ICA models. In *Proc. of the XI European Signal Proc. Conf. (EUSIPCO 2002)*, volume 2, pp. 7-10, Toulouse, France, September 2002.
5. A. Taleb, A generic framework for blind source separation in structured nonlinear models, *IEEE Trans. on Signal Processing*, 50(8), pp. 1819-1830, August 2002.
6. S. Hosseini and C. Jutten, On the separability of nonlinear mixtures of temporally correlated sources, *IEEE Signal Processing Letters*, 10(2), pp. 43-46, February 2003.
7. S. Hosseini, Y. Deville, Blind separation of linear-quadratic mixtures of real sources using a recurrent structure, in *Proc. IWANN*, vol.2, pp. 241-248, Mao, Menorca, Spain, June 2003.
8. C. Jutten, B. Babaie-Zadeh, S. Hosseini, Three easy ways for separating nonlinear mixtures?, *Signal Processing*, 84(2), pp. 217-229, February 2004.
9. M. Krob and M. Benidir, Blind identification of a linear-quadratic model using higher-order statistics, In *Proc. ICASSP*, vol. 4, pp. 440-443, 1993.
10. K. Abed-Meraim, A. Belouchrani, and Y. Hua, Blind identification of a linear-quadratic mixture of independent components based on joint diagonalization procedure, In *Proc. ICASSP*, pp. 2718-2721, Atlanta, USA, May 1996.
11. C. Jutten and J. Hérault, Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture, *Signal Processing*, 24:1-10, 1991.
12. N. Charkani and Y. Deville, Self -adaptive separation of convolutively mixed signals with a recursive structure. Part I: Stability analysis and optimization of asymptotic behaviour, *Signal Processing*, 73(3)3, pp. 225-254, 1999.
13. N. Charkani and Y. Deville, Self-adaptive separation of convolutively mixed signals with a recursive structure. Part II: Theoretical extensions and application to synthetic and real signals, *Signal Processing*, 75(2), pp. 117-140, 1999.
14. D.-T. Pham, Fast algorithm for estimating mutual information, entropies and score functions, in *Proc. ICA*, pp. 17-22, Nara, Japan, April 2003.