

An Adaptive Subspace Algorithm for Blind Separation of Independent Sources in Convolutional Mixture

Ali MANSOUR¹, Christian JUTTEN^{2,4}, Philippe LOUBATON^{3,4}.

¹ BMC Research Center (RIKEN), 2271-130 Anagahora, Moriyama-ku, Nagoya 463 (JAPAN)

Tel: (+ 81) 52 - 736 - 5867/ Fax: (+ 81) 52 - 736 - 5868.

² INPG-TIRF, 46 avenue Félix Viallet, 38031 Grenoble Cedex (France).

³ Univ. de Marne la Vallée, 2 rue de la Butte Verte, 93166 Noisy-Le-Grand Cedex (France).

⁴ PRC-GdR-ISIS, CNRS.

email: mansour@nagoya.riken.go.jp, chris@tirf.inpg.fr, loubaton@pekin.univ-mlv.fr.

EDICS SP 4.1.4.

IEEE Trans. on Signal Processing VOL 48, N 2, February 2000

Abstract

We propose an algorithm for blind separation of sources in convolutional mixtures based on a subspace approach. The advantage of this algorithm is that it reduces a convolutional mixture to an instantaneous mixture by using only second-order statistics (but more sensors than sources). Furthermore, the sources can be separated by using any algorithm for an instantaneous mixture (based generally on fourth-order statistics). Otherwise, the classical assumptions for blind separation of sources (at most one source can be a Gaussian signal and the sources are statistically independent) and some new subspace assumptions are considered. The assumptions concern the subspace model and their properties are emphasized. Finally, an experimental study is conducted and results are discussed.

1 Introduction

The blind separation of sources problem, which was introduced in 1984 by Héroult and Ans [10] for modeling motion decoding in neurobiology, has been extensively studied during the last ten years, especially by researchers in the signal processing field, and more recently by those in the neural network field [3, 2].

Since 1985, many authors [7, 6, 17] have addressed the problem of source separation in memoryless linear mixtures (also called instantaneous mixtures).

In the last five years, only a few methods have been proposed for separating sources from linear convolutive mixture [23, 19].

In both cases, separation methods are based on the statistical independence of the sources, which is the main assumption in the source separation problem. This assumption is usually exploited using, in most source separation algorithms, higher order statistics (typically fourth-order statistics).

However, separation algorithms based on second-order statistics are very efficient if specific assumptions on sources or on mixtures are satisfied. For instance, in memoryless mixtures, if the sources are not independent and identically distributed (iid) and have different spectra, second-order separation is possible [20, 4]. Van Gerven *et al.* proposed a second-order separation algorithm for instantaneous mixtures [21], or for convolutive mixtures [22] if the mixture involves **strictly causal finite impulse response (FIR) filters**.

In the general case, the blind separation of sources is based on two basic assumptions:

1. The source signals are statistically independent [15, 12, 16].
2. At most one source is a Gaussian signal [17, 14].

In the following, we assume that these two assumptions are true.

2 Mathematical model

The blind source separation problem is characterized by estimating the source signals from q sensors receiving linear mixtures of the sources. Let us denote by $X(n) = (x_1(n), x_2(n), \dots, x_q(n))^T$ a set of q observed mixtures of p unobservable independent sources $S(n) = (s_1(n), s_2(n), \dots, s_p(n))^T$. The channel effect is considered as a convolutive mixture:

$$X(n) = \sum_{i=0}^M A(i)S(n-i) \iff X(z) = A(z)S(z), \quad (1)$$

where $A(i)$ is a $q \times p$ scalar matrix such that the channel effect $A(z) = (a_{ij}(z)) = \sum_{i=0}^M A(i)z^{-i}$, and M is the degree¹ of $A(z)$. In addition, let $X_N(n)$ and $S_{M+N}(n)$ be random vectors defined by:

$$X_N(n) = \begin{pmatrix} X(n) \\ \dots \\ X(n-N) \end{pmatrix} \text{ and } S_{M+N}(n) = \begin{pmatrix} S(n) \\ \dots \\ S(n-M-N) \end{pmatrix}. \quad (2)$$

Using (1) and (2), we can easily prove that:

$$X_N(n) = T_N(A)S_{M+N}(n). \quad (3)$$

¹The degree of a polynomial matrix $A(z)$ corresponds to the highest value of the degree among the degrees of all the elements $a_{ij}(z)$.

In (3), $T_N(A)$ is the $q(N+1) \times (M+N+1)p$ Sylvester matrix associated with $A(z)$:

$$T_N(A) = \begin{bmatrix} A(0) & A(1) & A(2) & \dots & A(M) & 0 & 0 & \dots & 0 \\ 0 & A(0) & A(1) & \dots & A(M-1) & A(M) & 0 & \dots & 0 \\ \vdots & \vdots & & & & & & & \\ 0 & \dots & \dots & & 0 & A(0) & A(1) & \dots & A(M) \end{bmatrix}. \quad (4)$$

The Sylvester matrices are well known in system theory [13] for their attractive properties, as long as the following three assumptions are satisfied²:

- **H1**: The number of sensors is larger than the number of sources, $p < q$.
- **H2**: $A(z)$ is irreducible ($\text{Rank}(A(z)) = p, \forall z$ excluding $z = 0$ but including $z = \infty$).
- **H3**: $A(z)$ is column reduced:
 $A(z)$ can be written as:

$$A(z) = A_{cc} \text{diag}\{z^{-M_1}, \dots, z^{-M_p}\} + A_1(z), \quad (5)$$

where M_j denotes the degree³ of the j -th column of $A(z)$, A_{cc} is a non polynomial matrix, and $A_1(z)$ is a polynomial matrix, whose degree of the j -th column is less than M_j . By definition, $A(z)$ is reduced by column if and only if A_{cc} is a full-rank matrix.

As long as $p < q$, these assumptions have been shown in [9] to be realistic (it is easy to verify that if $A(z)$ is a square matrix, then the rank of $A(z)$ will be less than p , at least for some z_i such that $\det(A(z_i)) = 0$). It can be shown [5] that under the assumptions **H2** and **H3**:

$$\text{Rank}(T_N(A)) = p(N+1) + \sum_{i=1}^p M_i, \quad (6)$$

as long as $N \geq \sum_{i=1}^p M_i$, where M_i is the degree of the i -th column $A_i(z)$ of $A(z)$.

One should note that $p(N+1) + \sum_{i=1}^p M_i$ is precisely the number of nonzero columns of $T_N(A)$. In particular, if all the degrees $(M_i)_{i=1,p}$ coincide with M , then, $T_N(A)$ is full column rank if $N \geq pM$. Therefore, $T_N(A)$ admits a left inverse.

²In this case, the filter $A(z)$ can be completely determined using the Sylvester matrix [13, 9, 18].

³The degree of the j -th column of $A(z)$ equals the maximum degree of the j -th column component $a_{ij}(z)$, $\forall i = 1, q$.

3 Criterion and constraint

Let us assume that the degrees M_i are equal⁴ to M :

$$M_i = M \quad i = 1 \dots p. \quad (7)$$

Generalizing the method proposed by Gesbert *et al.* [8] for identification (in the identification problem, the authors assume that they have one source, $p = 1$, and that the source is an iid signal), we propose the estimation of a left inverse matrix of the Sylvester matrix $T_N(H)$ by adaptatively minimizing a cost function.

Let us choose $N \geq pM$. Then, as mentioned in the previous section, $T_N(A)$ is left invertible, so that there exists a $p(M + N + 1) \times q(N + 1)$ matrix B for which:

$$BT_N(A) = I_{(M+N+1)p}, \quad (8)$$

which implies:

$$BX_N(n) = S_{M+N}(n). \quad (9)$$

Using definitions (2), it is obvious that the first $(M + N)p$ rows of $BX_N(n)$ are equal to the last $(M + N)p$ rows of $BX_N(n + 1)$. More precisely, if B is a $p(M + N + 1) \times q(N + 1)$ matrix for which

$$(I_{(M+N)p} \ 0_p)BX_N(n) = (0_p \ I_{(M+N)p})BX_N(n + 1) \quad (10)$$

then for each n it can be proved that (see Appendix A):

$$BT_N(A) = \mathcal{H} = \begin{pmatrix} H & 0 & 0 & \dots \\ 0 & H & 0 & \dots \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \dots & H \end{pmatrix} = \text{diag}(H, H, \dots, H), \quad (11)$$

for some $p \times p$ scalar matrix H . Therefore, it is possible to identify a left inverse of $T_N(A)$ by minimizing the quadratic cost function, derived from (10) by averaging:

$$\mathcal{J}(B) = E\|(I_{(M+N)p} \ 0_p)BX_N(n) - (0_p \ I_{(M+N)p})BX_N(n + 1)\|^2. \quad (12)$$

The minimization of (12) should be done under a constraint ensuring that the matrix H in (11) (corresponding to the minimum of (12)) is invertible⁵. We note that the first p -components of $BX_N(n)$ are equal to $HS(n)$ and it can be corresponded as a memoryless linear mixture (with a mixing matrix H) of the sources $S(n)$. Finally the sources $S(n)$ can be retrieved by using a source

⁴If this assumption is not satisfied, then, by adopting another parameterization also based on the Sylvester matrix, it is possible to separate the sources [18].

⁵If the matrix H , corresponding to the minimum of (12), is not a regular matrix, then the second step of our algorithm, the separation of the instantaneous residual mixture, cannot be achieved.

separation algorithm for instantaneous mixtures⁶.

As a constraint for the minimization of $\mathcal{J}(B)$, we propose:

$$B_0 R_X B_0^T = I_p, \quad (13)$$

where B_0 is the first block row ($p \times q(N+1)$) of B , and $R_X = EX_N(n)X_N(n)^T$ is the covariance matrix of $X_N(n)$. This constraint is sufficient to ensure the invertibility of H . In fact:

$$\begin{aligned} I_p = B_0 R_X B_0^T &= EB_0 X_N(n) X_N^T(n) B_0^T \\ &= EB_0 T_N(A) S_{(M+N)} S_{(M+N)}^T T_N(A)^T B_0^T \\ &= B_0 T_N(A) E S_{M+N} S_{(M+N)}^T (B_0 T_N(A))^T \\ &= H E S(n) S(n)^T H^T \\ &= H R_S H^T, \end{aligned} \quad (14)$$

where $R_S = ES(n)S(n)^T$. Clearly, R_S is a diagonal full-rank matrix because the sources are independent signals, and the constraint (13) ensures the invertibility of H .

In practice, an adaptive minimization algorithm can be derived by a classical LMS algorithm, involving only second-order statistics. After some calculations, we can find that the weight matrix B should be adapted for minimizing the cost function (12), by:

$$\begin{aligned} \Delta B = \frac{\partial \mathcal{J}(B)}{\partial B} &= \begin{pmatrix} I_p & 0 & \dots & 0 & 0 \\ 0 & 2I_p & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \dots & 0 \\ 0 & \dots & \dots & 2I_p & 0 \\ 0 & \dots & \dots & 0 & I_p \end{pmatrix} B R_X \\ &\quad - \begin{pmatrix} 0 & I_{(M+N)p} \\ 0 & 0 \end{pmatrix} B R_{X1}^T - \begin{pmatrix} 0 & 0 \\ I_{(M+N)p} & 0 \end{pmatrix} B R_{X1} \end{aligned}$$

where I_p is a $p \times p$ identity matrix and $R_{X1} = EX_N(n)X_N(n+1)^T$. After each step, the current estimate $B(n)$ of B must be constrained to satisfy (13): this is done by normalizing B_0 as shown in the following equation:

$$B_0 = (B_0 \hat{R}_X(n) B_0^T)^{-1/2} B_0 \quad (15)$$

where $\hat{R}_X(n)$ is an estimate of R_X and $(B_0 \hat{R}_X(n) B_0^T)^{-1/2}$ is the square root of $B_0 \hat{R}_X(n) B_0^T$, which can be calculated using Cholesky decomposition. Finally, the residual instantaneous mixture (corresponding to H) is processed using any source separation algorithm for instantaneous mixtures (for example, we used a J-H algorithm [11]).

⁶We can say that the majority of the filter coefficients $A(z)$ (the total number of coefficients is $q \times p \times (M+1)$) can be estimated using second-order statistics, but the $p \times p$ coefficients of the matrix H must be estimated by using an instantaneous mixture algorithm and fourth-order statistics.

4 Experimental results

The experiments discussed here are conducted using two sources ($p = 2$), three or four sensors ($q = 3$ or 4) and the degree of $A(z)$ is chosen between 1 and 5 ($1 \leq M \leq 5$). Convergence is obtained after about 7000 iterations with a LMS algorithm (see section 3), providing good separation: with stationary signals, the crosstalk is about -21dB .

The matrix $B_{\min}T_N(A)$ (where B_{\min} is the estimated matrix minimizing (12) using only second-order statistics) is illustrated in Fig. 2. This matrix is block diagonal with diagonal blocks of size 2×2 (because there are two sources) corresponding to H in equation (11).

The second step of the separation algorithm involves separating the instantaneous mixture H using higher order statistics (basically fourth order).

First, we will illustrate the performance of our algorithm when the sources are uniform iid signals and the mixtures consist of low-pass filters in Fig. 3. The algorithm developed in this study assumes indirectly that M is perfectly known (see assumption (7)). We show in section 4.1 the importance of this assumption from a practical point of view.

4.1 Filter degree

Theoretically, the algorithm is derived under the condition that the columns of $A(z)$ have the same degree ($M_i = M$). This also means that the degree M must be perfectly known⁷. The influence of a wrong estimation of M on the algorithm robustness is of considerable interest. We studied two possible cases: under estimation and over estimation of M .

4.1.1 Under estimation of M

We first test the influence of an under estimation of M . A preliminary experimental study showed that the separation is possible even if the estimated degree \hat{M} of $A(z)$ is less than the actual degree M . Then, we addressed the question: what are the conditions to achieve a good separation with an underestimated value of M ?

We experimentally tested the following conjecture: " The separation is possible if the estimated degree \hat{M} is sufficient to provide good fundamental modeling of filters $A_{ij}(z)$ ". Experimental study confirms this conjecture, and some of the experiments are as follows:

- We verify that if filters have a degree equal to $M = 2n + 1$ with n resonances ($n = 1, 2, 3, \dots$), the underestimation ($\hat{M} = 2n$) leads to good separation. For example, Fig. 4 illustrates the total matrix $B.T_N(A)$ when $\hat{M} = 2$ while the actual degree is $M = 3$ (most of the filters $A_{ij}(z)$

⁷An estimator of M is proposed by Abed Meraim *et al.* [1].

have one resonance): the matrix $B.T_N(A)$ is block diagonal, with constant blocks corresponding to the remaining instantaneous mixture H .

- In contrast, an underestimation such that the order \hat{M} is not sufficient to model all the resonances of the mixing filters, does not provide good separation. For instance, in the case of filters with two resonances ($M = 5$), $\hat{M} = 2$ or $\hat{M} = 3$ cannot lead to good results. For example, Fig. 5 illustrates the case when $\hat{M} = 2$ and the degree $M = 5$: the matrix $B.T_N(A)$ is no longer block diagonal.

4.1.2 Over estimation

Theoretically, an over estimation of the filter order implies that the span⁸ of $T_N(A)$ allows us to identify the filter $A(z)$ up to a scalar filter [1], i.e, we can only estimate $A'(z) = \alpha(z)A(z)$. This property can be observed experimentally. For example, consider $A(z)$ with degree $M = 3$ and assume that $\hat{M} = 4$. Results obtained with this wrong estimation are shown in Fig. 6, where clearly $B.T_N(A)$ is not a block diagonal matrix as shown in Fig. 6, the diagonal blocks are no longer constant.

The estimated sources (not shown here) are strongly distorted by filters $\alpha(z)$ and the separation can never be achieved. This experiment proves that the algorithm is not robust to an overestimation of the filter degree M .

5 CONCLUSION

In this paper, we have proposed an adaptative algorithm for source separation in convolutive mixtures based on a subspace approach. This algorithm applies if the number of sensors is larger than the number of sources, and it allows the separation of convolutive mixtures of independent sources using mainly second-order statistics. A simple instantaneous mixture, the separation of which generally needs high-order statistics, appears only if filters have the same order. Most of the parameters can be estimated using a simple LMS algorithm.

Theoretically, the algorithm requires knowledge of the filter degree M . However, we showed experimentally that an underestimation of M ($\hat{M} < M$) is acceptable as long as it allows a correct modeling of the filter behavior. The algorithm, based on LMS method, converge slowly and we currently study algorithm based on Gradient Conjugate in order to improve convergence speed and to be able to process stationary signals as well as non-stationary ones.

A Type of solutions

Let us denote by $\mathcal{H} = B.T_N(A)$. Using (10), we remark for each n that \mathcal{H} satisfies:

$$[I_{(M+N)p} \ 0_p] \mathcal{H} S_{(M+N)}(n) = [0_p \ I_{(M+N)p}] \mathcal{H} S_{(M+N)}(n+1) \quad (16)$$

⁸The space spanned by the column of $T_N(H)$.

$$\iff [0_p \ (I_{(M+N)p} \ 0_p)\mathcal{H}]S_{(M+N+1)}(n+1) = [(0_p \ I_{(M+N)p})\mathcal{H} \ 0_p]S_{(M+N+1)}(n+1),$$

where 0_p is a $(M+N)p \times p$ zero matrix. Let us assume that the input signal is persistently exciting then we can write:

$$[0_p \ (I_{(M+N)p} \ 0_p)\mathcal{H}] = [(0_p \ I_{(M+N)p})\mathcal{H} \ 0_p].$$

Let us assume that:

$$\mathcal{H} = (H_{ij}) = \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} c \\ d \end{bmatrix}.$$

where H_{ij} is a $p \times p$ matrix, a and d are $p \times (M+N+1)p$ matrix, and b and c are $(M+N)p \times (M+N+1)p$ matrix. From the equation (17), we can deduced that:

$$[b \ 0_p] = [0_p \ c]. \quad (17)$$

Last equation implies:

- First block column (i.e the first p columns) of b is equal to zero:

$$H_{i1} = 0, \quad \forall 1 < i \leq M+N+1, \quad (18)$$

- Last column of c is equal to zero:

$$H_{i(M+N+1)} = 0, \quad \forall 1 \leq i < m+N+1, \quad (19)$$

- And we have the follow relationship $H_{ij} = H_{(i-1)(j-1)}$

Finally, we can write that:

$$\mathcal{H} = \begin{pmatrix} H & 0 & 0 & \dots \\ 0 & H & 0 & \dots \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \dots & H \end{pmatrix} = \text{diag}(H, H, \dots, H). \quad (20)$$

Finally, \mathcal{H} is block diagonal matrix, as in (11).

References

- [1] K. Abed Meraim and Y. Hua. Blind identification of multi-input multi-output system using minimum noise subspace. *IEEE Trans. on Signal Processing*, 45(1):254–258, January 1997.
- [2] S. I. Amari, A. Cichocki, and H. H. Yang. A new learning algorithm for blind signal separation. In *Neural Information Processing System 8*, pages 757–763, Eds. D.S. Toureyzky et. al., 1995.

- [3] A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159, November 1995.
- [4] A. Belouchrani, K. Abed-Meraim, J. F. Cardoso, and E. Moulines. Second-order blind separation of correlated sources. In *Int. Conf. on Digital Sig.*, pages 346–351, Nicosia, Cyprus, July 1993.
- [5] R. Bitmead, S. Kung, B. D. O. Anderson, and T. Kailath. Greatest common division via generalized Sylvester and Bezout matrices. *IEEE Trans. on Automatic Control*, 23(6):1043–1047, December 1978.
- [6] J. F. Cardoso, A. Belouchrani, and B. Laheld. A new composite criterion for adaptive and iterative blind source separation. In *Proceedings of International Conference on Speech and Signal Processing 1994, ICASSP 1994*, pages 273–276, Adelaide, Australia, April 18-22 1994.
- [7] P. Comon, C. Jutten, and J. Héroult. Blind separation of sources, Part II: Statement problem. *Signal Processing*, 24(1):11–20, November 1991.
- [8] D. Gesbert, P. Duhamel, and S. Mayrargue. Subspace-based adaptive algorithms for the blind equalization of multichannel fir filters. In M.J.J. Holt, C.F.N. Cowan, P.M. Grant, and W.A. Sandham, editors, *Signal Processing VII, Theories and Applications (EUSIPCO'94)*, pages 712–715, Edinburgh, Scotland, September 1994. Elsevier.
- [9] A. Gorokhov and Ph. Loubaton. Subspace based techniques for second order blind separation of convolutive mixtures with temporally correlated sources. *IEEE Trans. on Circuits and Systems*, 44:813–820, September 1997.
- [10] J. Héroult and B. Ans. Réseaux de neurones à synapses modifiables: Décodage de messages sensoriels composites par un apprentissage non supervisé et permanent. *C. R. Acad. Sci. Paris, série III*:525–528, 1984.
- [11] C. Jutten and J. Héroult. Blind separation of sources, Part I: An adaptive algorithm based on a neuromimetic architecture. *Signal Processing*, 24(1):1–10, 1991.
- [12] C. Jutten, L. Nguyen Thi, E. Dijkstra, E. Vittoz, and Caelen J. Blind separation of sources: An algorithm for separation of convolutive mixtures. In *International Signal Processing Workshop on Higher Order Statistics*, pages 273–276, Chamrousse, France, July 1991.
- [13] T. Kailath. *Linear systems*. Prentice Hall, 1980.
- [14] J. L. Lacoume and M. Gaeta. The general source separation problem. In *Fifth ASSP Workshop on estimation and modeling*, pages 154–158, Rochester New-York, October 1990.

- [15] B. Laheld and J. F. Cardoso. Adaptive source separation with uniform performance. In M.J.J. Holt, C.F.N. Cowan, P.M. Grant, and W.A. Sandham, editors, *Signal Processing VII, Theories and Applications (EUSIPCO'94)*, pages 1–4, Edinburgh, Scotland, September 1994. Elsevier.
- [16] A. Mansour and C. Jutten. Fourth order criteria for blind separation of sources. *IEEE Trans. on Signal Processing*, 43(8):2022–2025, August 1995.
- [17] A. Mansour and C. Jutten. A direct solution for blind separation of sources. *IEEE Trans. on Signal Processing*, 44(3):746–748, March 1996.
- [18] A. Mansour, C. Jutten, and Ph. Loubaton. Subspace method for blind separation of sources and for a convolutive mixture model. In *Signal Processing VIII, Theories and Applications (EUSIPCO'96)*, pages 2081–2084, Triest, Italy, September 1996. Elsevier.
- [19] L. Nguyen Thi and C. Jutten. Blind sources separation for convolutive mixtures. *Signal Processing*, 45(2):209–229, 1995.
- [20] R. L. L. Tong and Y. H. V. C. Soon. Indeterminacy and identifiability of blind identification. *IEEE Trans. on CAS*, 38:499–509, May 1991.
- [21] S. Van Gerven and D. Van Compernelle. Signal separation by symmetric adaptive decorrelation: Stability, convergence, and uniqueness. *IEEE Trans. on Signal Processing*, 43(7):1602–1612, July 1995.
- [22] S. Van Gerven, D. Van Compernelle, L. Nguyen Thi, and C. Jutten. Blind separation of sources: A comparative study of a 2nd and a 4th order solution. In M.J.J. Holt, C.F.N. Cowan, P.M. Grant, and W.A. Sandham, editors, *Signal Processing VII, Theories and Applications (EUSIPCO'94)*, pages 1153–1156, Edinburgh, Scotland, September 1994. Elsevier.
- [23] D. Yellin and E. Weinstein. Criteria for multichannel signal separation. *IEEE Trans. on Signal Processing*, 42(8):2158–2167, August 1994.

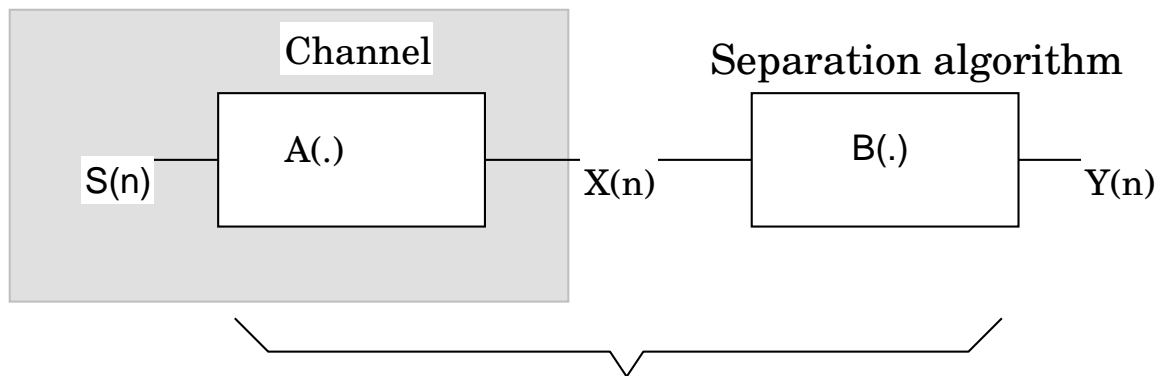


Figure 1: General structure.

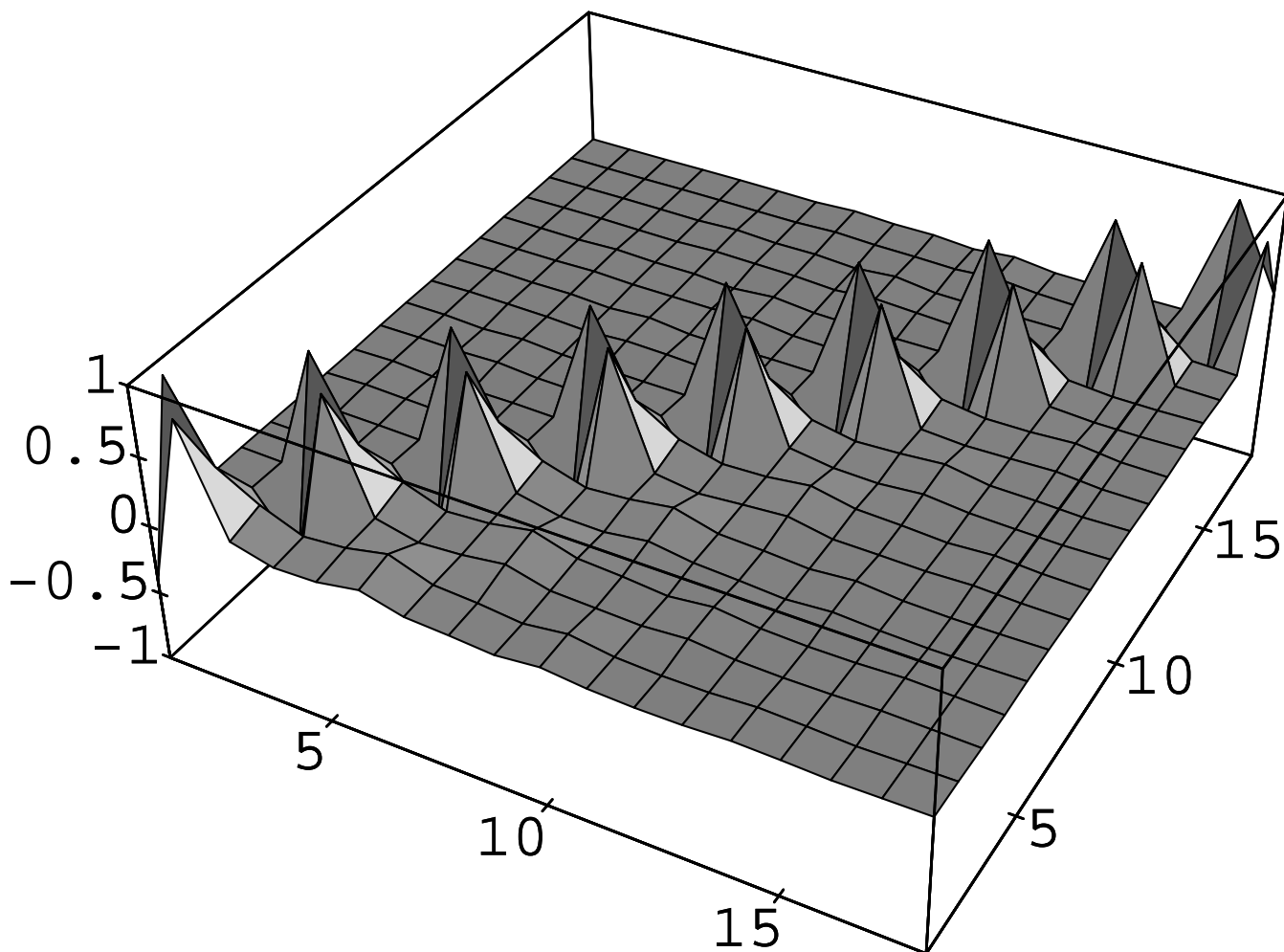
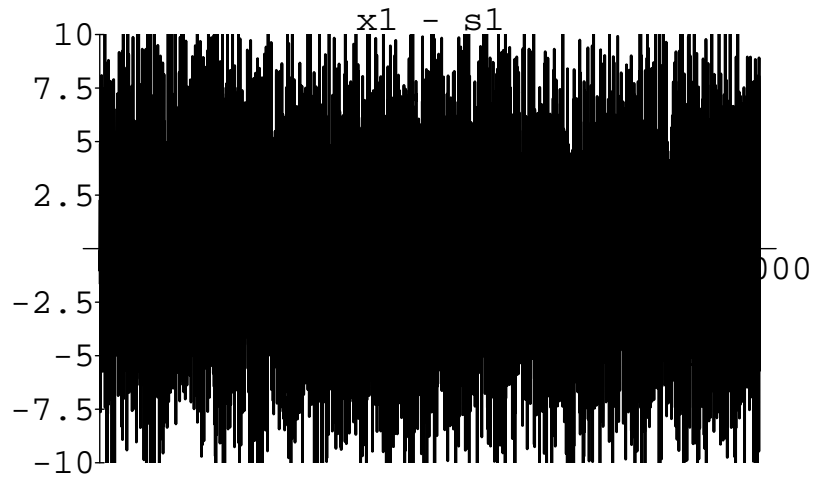
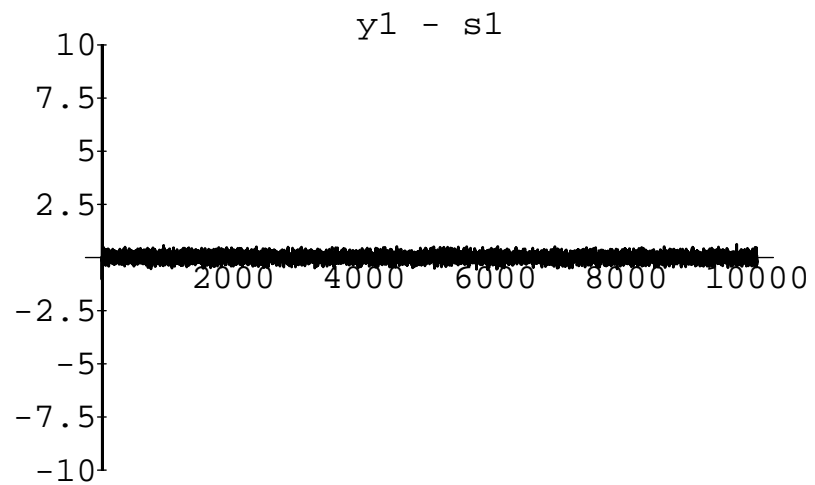


Figure 2: Representation of the matrix $BT_N(A)$ at the convergence. $M = 3, q = 4, p = 2$, the convergence is obtained after 7000 samples, the sources are two uniform iid signals



(a) Difference between the source s_1 and the mixture signal x_1 .



(b) Estimation Error: difference between the source s_1 and the estimated signal y_1 .

Figure 3: Algorithm performance: after the convergence of the subspace identification and the separation of the residual instantaneous mixture.

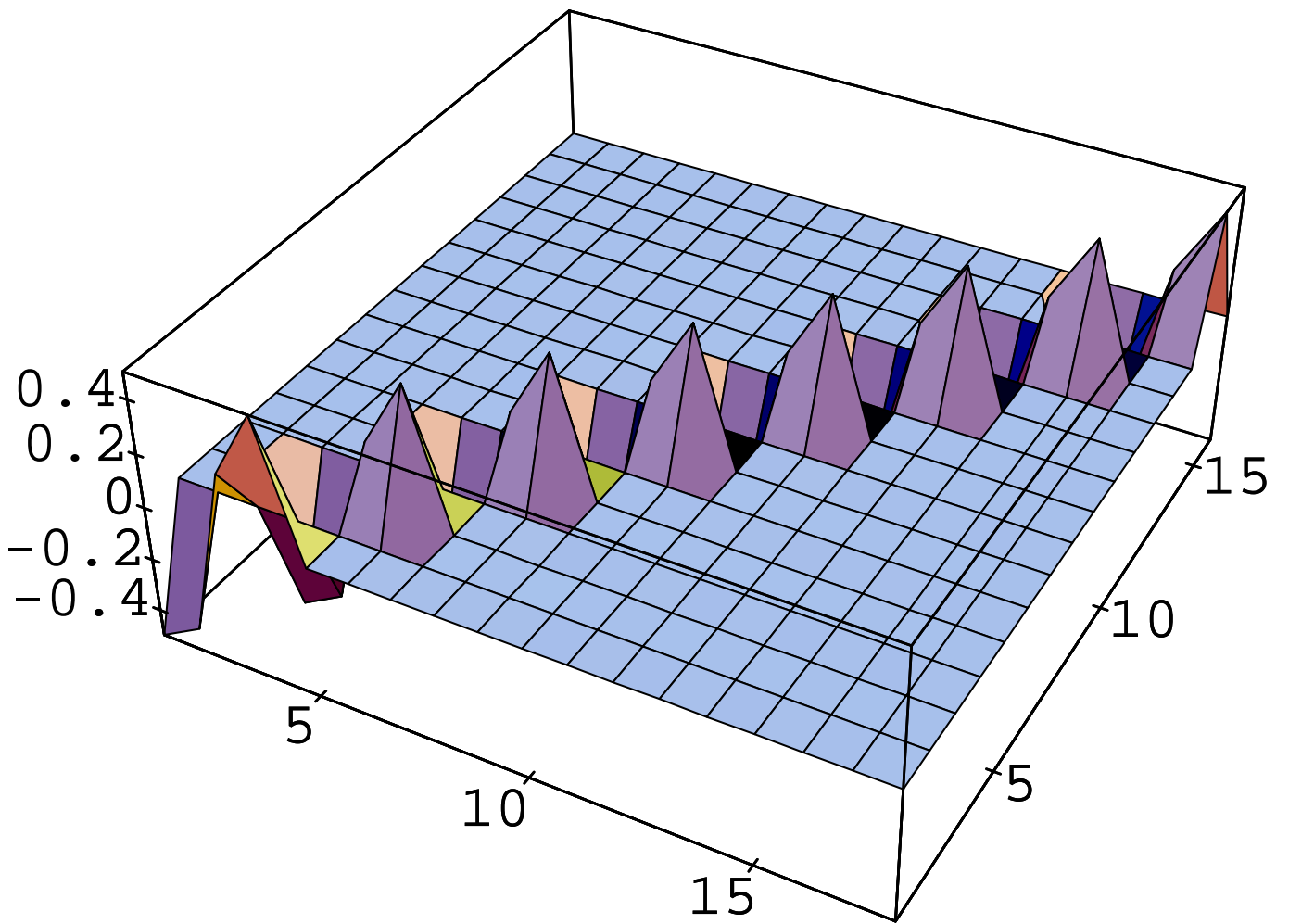


Figure 4: Total matrix $BT_N(A)$ in the case of an acceptable under-estimation of $(\hat{M} = 2, M = 3)$. The matrix remains block diagonal.

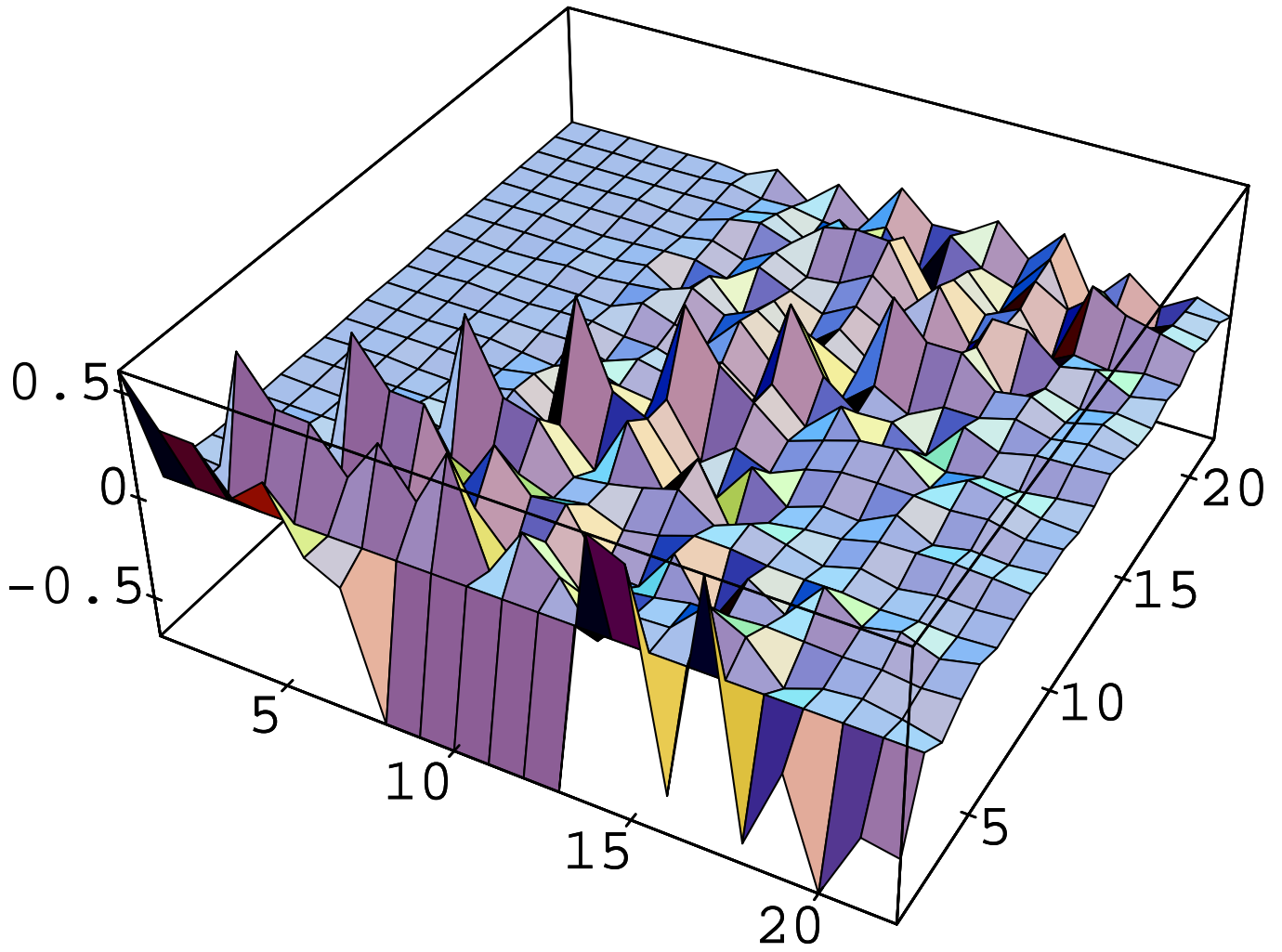


Figure 5: Total matrix $BT_N(A)$ in the case of a too strong under-estimation of M ($\hat{M} = 2, M = 5$). The matrix is no longer block diagonal: the remaining mixture is not an instantaneous mixture, and separation can not be achieved.

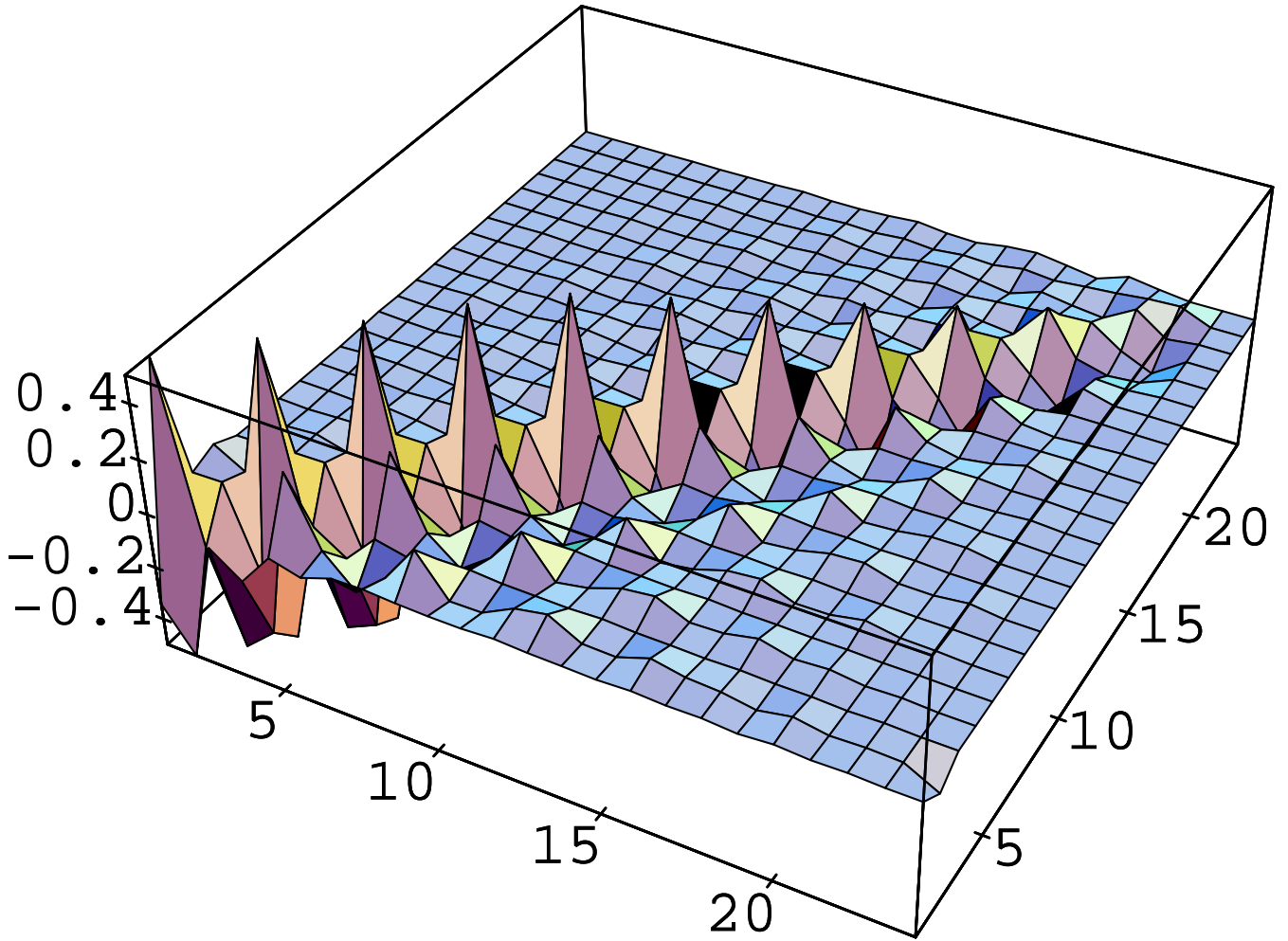


Figure 6: Total matrix $BT_N(A)$ with an estimation of the filter degree ($\hat{M} = 4, M = 3$). The matrix is no longer block diagonal: the remaining mixture is not an instantaneous mixture, and separation can not be achieved.