A BATCH SUBSPACE ICA ALGORITHM.

Ali MANSOUR and Noboru OHNISHI

Bio-Mimetic Control Research Center (RIKEN),
2271-130, Anagahora, Shimoshidami, Moriyama-ku, Nagoya 463 (JAPAN)
email:mansour@nagoya.riken.go.jp and ohnishi@ohnishi.nuie.nagoya-u.ac.jp
http://www.bmc.riken.go.jp

ABSTRACT

For the blind separation of sources (BSS) problem (or the independent component analysis (ICA)), it has been shown in many situations, that the adaptive subspace algorithms are very slow and need an important computation efforts. In a previous publication, we proposed a modified subspace algorithm for stationary signals. But that algorithm was limited to stationary signals and its convergence was not fast enough.
Here, we propose a batch subspace algorithm. The experimental study proves that this algorithm is very fast but its performance are not enough to completely achieve the separation of the independent component of the signals. In the other hand, this algorithm can be used as a pre-processing algorithm to initialized other adaptive subspace algorithms.
Keywords: blind separation of sources, ICA, subspace methods, Lagrange method, Cholesky decomposition.

1. INTRODUCTION

The blind separation of sources (BSS) problem [1] (or the Independent Component Analysis "ICA" problem [2]) is a recent and important problem in signal processing. According to this problem, one should estimate, using the output signals of an unknown channel (i.e. the observed signals or the mixing signals), the unknown input signals of that channel (i.e. sources). The sources are assumed to be statistically independent from each other.

At first the BSS was proposed in a biological context [3]. Actually, one can find this problem in many different situations: speech enhancement [4], separation of seismic signals [5], sources separation method applied to nuclear reactor monitoring [6], airport surveillance [7], noise removal from biomedical signals [8], *etc.*

Since 1985, many researchers have been interested in BSS [9, 10, 11, 12]. Most of the algorithms deal with a linear channel model: The instantaneous mixtures (i.e. memoryless channel) or the convolutive mixtures (i.e. the channel effect can be considered as a linear filter). The criteria of those algorithms were generally based on high order statistics [13, 14, 15]. Recently, by using only second order statistics, some subspace methods have been explored to separate blindly the sources in the case of convolutive mixtures [16, 17].

In previous works, we proposed two subspace approaches using LMS [18, 17] or a conjugate gradient algorithm [19] to minimize subspace criteria. Those criteria were been derived from the generalization of the method proposed by Gesbert *et al.* [20] for blind identification[1]. To improve the convergence speed of our algorithms, we proposed a modified subspace algorithm for stationary signals [21]. But that algorithm was limited to stationary signals and its convergence was not fast enough. Here, we propose a new subspace algorithm, which improves the performance of our previous methods.

2. MODEL, ASSUMPTIONS & CRITERION

Let $Y(n)$ denotes the $q \times 1$ mixing vector obtained from p unknown and statistically independent sources $S(n)$ and let the $q \times p$ polynomial matrix $\mathcal{H}(z) = (h_{ij}(z))$ denotes the channel effect (see fig. 1). In this paper, we assume that the filters $h_{ij}(z)$ are causal and finite impulse response (FIR) filters. Let us denote by M the highest degree[2] of the filters $h_{ij}(z)$. In this case, $Y(n)$ can be written as:

$$Y(n) = \sum_{i=0}^{M} \mathbf{H}(i) S(n-i), \tag{1}$$

where $S(n-i)$ is the $p \times 1$ source vector at the time $(n-i)$ and $\mathbf{H}(i)$ is the real $q \times p$ matrix corresponding to the filter matrix $\mathcal{H}(z)$ at time i.

Let $Y_N(n)$ (resp. $S_{M+N}(n)$) denotes the $q(N+1) \times 1$ (resp. $(M+N+1)p \times 1$) vector given by:

$$Y_N(n) = \begin{pmatrix} Y(n) \\ \vdots \\ Y(n-N) \end{pmatrix},$$

$$S_{M+N}(n) = \begin{pmatrix} S(n) \\ \vdots \\ S(n-M-N) \end{pmatrix}.$$

[1] In the identification problem, the authors generally assume that they have one source and that the source is an iid signal.

[2] M is called the degree of the filter matrix $\mathcal{H}(z)$.

0-7803-5988-7/00/$10.00 © 2000 IEEE

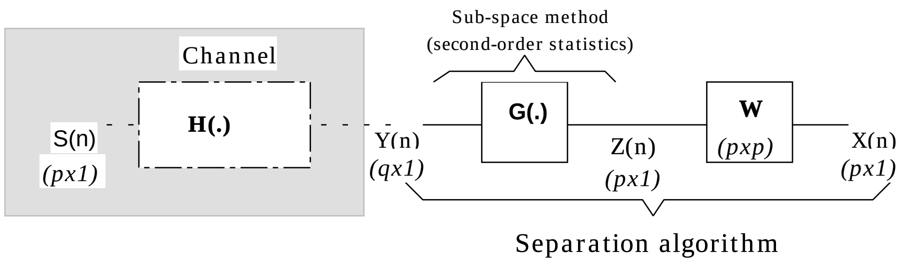


Figure 1: General Structure.

By using $N > q$ observations of the mixture vector, we can formulate the model (1) in another form:

$$Y_N(n) = \mathbf{T}_N(\mathbf{H}) S_{M+N}(n), \tag{2}$$

where $\mathbf{T}_N(\mathbf{H})$ is the Sylvester matrix corresponding to $\mathcal{H}(z)$. The $q(N+1) \times p(M+N+1)$ matrix $\mathbf{T}_N(\mathbf{H})$ is given by [22] as:

$$\begin{bmatrix} \mathbf{H}(0) & \mathbf{H}(1) & \dots & \mathbf{H}(M) & 0 & \dots & 0 \\ 0 & \mathbf{H}(0) & \dots & \mathbf{H}(M-1) & \mathbf{H}(M) & 0 & \dots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \mathbf{H}(0) & \mathbf{H}(1) & \dots & \mathbf{H}(M) \end{bmatrix}.$$

It was proved in [23] that the rank of Sylvester matrix $\mathbf{T}_N(\mathbf{H}) = p(N+1) + \sum_{i=1}^{p} M_i$, where M_i is the degree of the ith column[3] of $\mathcal{H}(z)$. Now, it is easy to prove that the Sylvester matrix has a full rank and it is left invertible if each column of the polynomial matrix $\mathcal{H}(z)$ has the same degree and $N > Mp$ (see [24] for more details). From equation (2), one can conclude that the separation of the sources can be achieved by estimating a $(M+N+1)p \times q(N+1)$ left inverse matrix $\mathbf{G}$ of the Sylvester matrix. To estimate G, one can use criterion proposed in [17] obtained from the generalization of the criterion in [20]:

$$\min C(\mathbf{G}) = \ \mathrm{E}\ \|(\mathbf{I}\ \ \mathbf{0})\mathbf{G}Y_N(n) - (\mathbf{0}\ \ \mathbf{I})\mathbf{G}Y_N(n+1)\|^2, \tag{3}$$

here E stands for the expectation, $\mathbf{I}$ is the identity matrix and $\mathbf{0}$ is a zero matrix of appropriate dimensions. It has been shown in [17] that the above minimization lead us to a matrix $\mathbf{G}^\star$ such:

$$\mathbf{Perf} = \mathbf{G}^\star \mathbf{T}_N(\mathbf{H}) = \mathrm{diag}(\mathbf{M}, \cdots, \mathbf{M}), \tag{4}$$

where $\mathbf{M}$ is any $p \times p$ matrix. Using the last equation, it becomes clear that the separation is reduced to the separation of an instantaneous mixture with a mixing matrix $\mathbf{M}$. In other words, this algorithm can be decomposed into two steps: First step, by using only second-order statistics, we reduce the convolutive mixture problem to an instantaneous mixture (deconvolution step); then in the second step, we must only separate sources consisting of a simple instantaneous mixture (typically, most of the instantaneous mixture algorithms are based on fourth-order statistics).

[3] The degree of a column is defined as the highest degree of the filters in this column.

Finally, to avoid the spurious solutions (i.e. a singular matrix $\mathbf{M}$), one must minimize that criterion subject to a constraint [17]:

$$\text{Subject to } \mathbf{G}_0 \mathbf{R}_N(n) \mathbf{G}_0^T = \mathbf{I}, \tag{5}$$

here $\mathbf{R}_N(n) = \ \mathrm{E}\ Y_N(n) Y_N^T(n)$, and the $p \times q(N+1)$ matrix $\mathbf{G}_0$ stands for the first bloc line of $\mathbf{G} = (\mathbf{G}_0^T \ \ \cdots \ \ \mathbf{G}_{(M+N)}^T)^T$. The minimization using a LMS algorithm of the above criterion with respect to a constraint was discuss in our previous work [17]. In addition, the minimization of a modified version of the above criterion was done using a conjugate gradient algorithm [19].

3. ALGORITHM

From the previous section, it is clear that the minimization of the criterion (3) should be done subject to a p^2 constraints[4]. Let *const* denotes the constraint vector (i.e. $const = \mathrm{Vec}\ (\mathbf{G}_0 \mathbf{R}_N(n) \mathbf{G}_0^T - \mathbf{I})$, here Vec is the operator that corresponds to a $p \times q$ matrix a pq vector). The minimization of the criterion (3) subject to the constraints (5) can be formulated using the Lagrange method as:

$$\mathcal{L}(\mathbf{G}, \lambda) = C(\mathbf{G}) - \lambda\ const \tag{6}$$

here λ is a line vector, stands for the Lagrange parameters. The minimization of the above equation with respect to λ leads us to the constraint equation (5). Using the derivative $\partial C(\mathbf{G}) / \partial \mathbf{G}$ given in [17], the equation (5) and (6), one can write:

$$\begin{aligned} \frac{\partial \mathcal{L}(\mathbf{G}, \lambda)}{\partial \mathbf{G}} &= \begin{pmatrix} \mathbf{I}_p & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 2\mathbf{I}_{(M+N-1)p} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_p \end{pmatrix} \mathbf{G}\mathbf{R}_N(n) \\ &- \begin{pmatrix} \mathbf{0} & \mathbf{I}_{(M+N)p} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{G}\mathbf{R}_N^T(n+1) \\ &- \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{I}_{(M+N)p} & \mathbf{0} \end{pmatrix} \mathbf{G}\mathbf{R}_N(n+1) - \begin{pmatrix} 2\mathbf{\Gamma}\ \mathbf{G}_0 \mathbf{R}_N(n) \\ \mathbf{0} \end{pmatrix}, \end{aligned}$$

where $\mathbf{R}_N(n+1) = \ \mathrm{E}\ Y_N(n) Y_N^T(n+1)$ and $\mathbf{I}_l$ is the $l \times l$ identity matrix. By canceling the above equation and after some algebraic operations, one can find that the bloc lines

[4] Using the symmetrical form of the equation (5), one can decrease the constraint number to $p(p+1)/2$.

of the optimal $\mathbf{G}^\star$ should satisfy:

$$\begin{aligned} \mathbf{G}_0\mathbf{R}_N(n)\mathbf{G}_0^T &= \mathbf{I}, & (7)\\ 2\mathbf{G}_i\mathbf{R}_N(n) &= \mathbf{G}_{(i+1)}\mathbf{R}_N^T(n+1) + \\ & \quad \mathbf{G}_{(i-1)}\mathbf{R}_N(n+1), & (8)\\ \mathbf{G}_{(M+N)}\mathbf{R}_N &= \mathbf{G}_{(M+N-1)}\mathbf{R}_N(n+1), & (9)\end{aligned}$$

here $1 \leq i \leq M+N-1$. Let $\mathbf{A} = \mathbf{R}_N^T(n+1)\mathbf{R}_N^{-1}(n)$ and $\mathbf{B} = \mathbf{R}_N^T(n+1)\mathbf{R}_N^{-1}(n)$, we should mention that $\mathbf{A}$ and $\mathbf{B}$ exist if and only if (iff) $\mathbf{R}_N(n)$ is full rank[5]. Finally, using some algebraic operations, we can prove that the previous matrix equation system can be solved by a recursion formula:

$$\mathbf{G}_{(M+N-i-1)} = \mathbf{G}_{(M+N-i-2)}\mathbf{D}_i \qquad (10)$$

her $0 \leq i \leq M+N-1$ and the $\mathbf{G}_0$ can be obtained from the first equation (7), using a simple Cholesky decomposition. In addition, the matrices $\mathbf{D}_i$ can also be obtained by:

$$\mathbf{D}_{(i+1)} = \mathbf{B}(2\mathbf{I} - \mathbf{D}_i\mathbf{A})^{-1} \qquad (11)$$

here $0 \leq i \leq M+N-1$ and $\mathbf{D}_0 = \mathbf{B}$. Even if relationships (10) and (11) looks complicated, but the time needed to obtain the matrix $\mathbf{G}$ still very comparable[6] to the time needed for the convergence of LMS version [17] or even the Conjugate Gradient version [21, 19].

4. EXPERIMENTAL RESULTS

The experiments discussed here are conducted using two sources ($p = 2$) with uniform probability density function (pdf) and four sensors ($q = 4$), and the degree of $\mathcal{H}(z)$ is chosen as ($M = 4$).

To show the performances of the subspace criterion, the matrix $\mathbf{Perf} = \mathbf{G}^\star\mathbf{T}_N(\mathbf{H})$ is plotted. In the other hand, we know that the deconvolution is achieved iff the matrix **Perf** is a bloc diagonal matrix as shown in equation (4). Figure 2 shows the performances of the batch subspace algorithm discussed in this paper. It is clear from that figure 2 that the first step of the algorithm (the deconvolution) was not satisfactory achieved (**Perf** is not a bloc diagonal as in equation (4). This problem was obtained because the criterion (3) is a flat function around its minima (see figure (2)).

Figure 3 shows us the performance results and the criterion convergence of the LMS algorithm (first column), and the performance results and the criterion convergence of the same LMS algorithm but the matrix $\mathbf{G}$ is initialized using the result of the batch algorithm (second column). We should mention that the time needed to obtain the minima by the initialized version was almost half the time needed by the non initialized version. Figures 3 (c) and (d) show the criterion convergence (the stop condition was the limit of the sample number, i.e. 10000). The experimental studies show that the Conjugate Gradient version of the subspace algorithm can converge faster and lead us to better performances if that algorithm has been initialized using the batch proposed algorithm (these results will be omitted in this short paper).

The second step of the algorithm consists on the separation of a residual instantaneous mixture (corresponding to $\mathbf{M}$, see equation (4)). This separation can be processed using any source separation algorithm applicable to instantaneous mixtures. Here, we chose the minimization of a cross-cumulant criterion using Levenberg-Marquardt method [25]. Figure (4) shows us the different signals (see figure (1)). It is clear that the sources X and the estimated signals S are independent signals and the vector Z, output of the subspace criterion, corresponds to an instantaneous mixture, and the observed vector Y corresponds to a convolutive mixture (see [26, 27]).

Finally, the estimation of the second and the high order statistics was done according to the method described in [28].

5. CONCLUSION

In this paper, we propose a batch algorithm for source separation in convolutive mixtures based on a subspace approach. This new algorithm requires, as same as the other subspace methods, that the number of sensors is larger than the number of sources. In addition, 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, should be conducted to achieve the separation.

The experimental study shows that the the present algorithm can be used for initialized an adaptive subspace algorithm. The initialized algorithms need less time to converge. These results were discussed in the case of two subspace algorithms which are based on LMS or on a conjugate gradient method. Finally, the subspace LMS criterion and the Conjugate gradient criterion will become more stable and faster if they are initialized using the present algorithm.

REFERENCES

[1] C. Jutten and J. Hérault, "Blind separation of sources, Part I: An adaptive algorithm based on a neuromimetic architecture," *Signal Processing*, vol. 24, no. 1, pp. 1–10, 1991.

[2] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, no. 3, pp. 287–314, April 1994.

[5] It is easy to prove that $\mathbf{R}_N(n)$ is full rank iff one add some additive independent noise to the observed signals, because one of the subspace assumption $q > p$. In the other hand and by using the criterion (3), one can prove the existence of some spurious minima, if the model have some additive noise (the demonstration will be omitted here because the limit of the sheet number). However, the experimental study shows that one still obtain good results for a 20 dB ratio of signal to noise (RSN). In our simulation, we added a Gaussian noise with RSN $\geq$ 20dB.

[6] Indeed, using C code program and an ultra 30 creator sun station, it needs few minutes (less than 5) to obtained the matrix $\mathbf{G}$. But the convergence of the conjugate gradient needs from 40 to 100 minutes and the LMS algorithm needs few hours to converge.

[3] B. Ans, J. C. Gilhodes, and J. Hérault, "Simulation de réseaux neuronaux (sirene). II. hypothèse de décodage du message de mouvement porté par les afférences fusoriales IA et II par un mécanisme de plasticité synaptique," *C. R. Acad; Sci. Paris*, vol. série III, pp. 419–422, 1983.

[4] L. Nguyen Thi and C. Jutten, "Blind sources separation for convolutive mixtures," *Signal Processing*, vol. 45, no. 2, pp. 209–229, 1995.

[5] N. Thirion, J. MARS, and J. L. BOELLE, "Separation of seismic signals: A new concept based on a blind algorithm," in *Signal Processing VIII, Theories and Applications*, Triest, Italy, September 1996, pp. 85–88, Elsevier.

[6] G. D'urso and L. Cai, "Sources separation method applied to reactor monitoring," in *Proc. Workshop Athos working group*, Girona, Spain, June 1995.

[7] E. Chaumette, P. Common, and D. Muller, "Application of ica to airport surveillance," in *HOS 93*, South Lake Tahoe-California, 7-9 June 1993, pp. 210–214.

[8] A. Kardec Barros, A. Mansour, and N. Ohnishi, "Removing artifacts from ecg signals using independent components analysis," *NeuroComputing*, vol. 22, pp. 173–186, 1999.

[9] J. F. Cardoso and P. Comon, "Tensor-based independent component analysis," in *Signal Processing V, Theories and Applications*, L. Torres, E. Masgrau, and M. A. Lagunas, Eds., Barcelona, Espain, 1990, pp. 673–676, Elsevier.

[10] S. I. Amari, A. Cichoki, and H. H. Yang, "A new learning algorithm for blind signal separation," in *Neural Information Processing System 8*, Eds. D.S. Toureyzky et. al., 1995, pp. 757–763.

[11] O. Macchi and E. Moreau, "Self-adaptive source separation using correlated signals and cross-cumulants," in *Proc. Workshop Athos working group*, Girona, Spain, June 1995.

[12] A. Mansour and C. Jutten, "A direct solution for blind separation of sources," *IEEE Trans. on Signal Processing*, vol. 44, no. 3, pp. 746–748, March 1996.

[13] M. Gaeta and J. L. Lacoume, "Sources separation without a priori knowledge: the maximum likelihood solution," in *Signal Processing V, Theories and Applications*, L. Torres, E. Masgrau, and M. A. Lagunas, Eds., Barcelona, Espain, 1994, pp. 621–624, Elsevier.

[14] N. Delfosse and P. Loubaton, "Adaptive blind separation of independent sources: A deflation approach," *Signal Processing*, vol. 45, no. 1, pp. 59–83, July 1995.

[15] A. Mansour and C. Jutten, "Fourth order criteria for blind separation of sources," *IEEE Trans. on Signal Processing*, vol. 43, no. 8, pp. 2022–2025, August 1995.

[16] A. Gorokhov and P. Loubaton, "Subspace based techniques for second order blind separation of convolutive mixtures with temporally correlated sources," *IEEE Trans. on Circuits and Systems*, vol. 44, pp. 813–820, September 1997.

[17] A. Mansour, C. Jutten, and P. Loubaton, "An adaptive subspace algorithm for blind separation of independent sources in convolutive mixture," *IEEE Trans. on Signal Processing*, vol. 48, no. 2, pp. 583–586, February 2000.

[18] A. Mansour, C. Jutten, and P. Loubaton, "Subspace method for blind separation of sources and for a convolutive mixture model," in *Signal Processing VIII, Theories and Applications*, Triest, Italy, September 1996, pp. 2081–2084, Elsevier.

[19] A. Mansour, A. Kardec Barros, and N. Ohnishi, "Subspace adaptive algorithm for blind separation of convolutive mixtures by conjugate gradient method," in *The First International Conference and Exhibition Digital Signal Processing (DSP'98)*, Moscow, Russia, June 30-July 3 1998, pp. I–252–I–260.

[20] D. Gesbert, P. Duhamel, and S. Mayrargue, "Subspace-based adaptive algorithms for the blind equalization of multichannel fir filters," in *Signal Processing VII, Theories and Applications*, M.J.J. Holt, C.F.N. Cowan, P.M. Grant, and W.A. Sandham, Eds., Edinburgh, Scotland, September 1994, pp. 712–715, Elsevier.

[21] A. Mansour and N. Ohnishi, "A blind separation algorithm based on subspace approach," in *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing (NSIP'99)*, Antalya, Turkey, June 20-23 1999, pp. 268–272.

[22] T. Kailath, *Linear systems*, Prentice Hall, 1980.

[23] 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*, vol. 23, no. 6, pp. 1043–1047, December 1978.

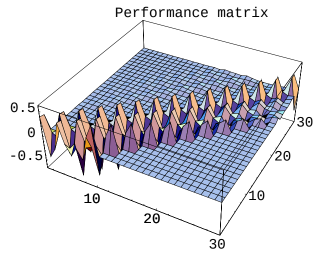
[24] A. Mansour, C. Jutten, and P. Loubaton, "Robustesse des hypothèses dans une méthode sous-espace pour la séparation de sources," in *Actes du XVIème colloque GRETSI*, Grenoble, France, September 1997, pp. 111–114.

[25] A. Mansour and N. Ohnishi, "Multichannel blind separation of sources algorithm based on cross-cumulant and the levenberg-marquardt method.," *IEEE Trans. on Signal Processing*, vol. 47, no. 11, pp. 3172–3175, November 1999.

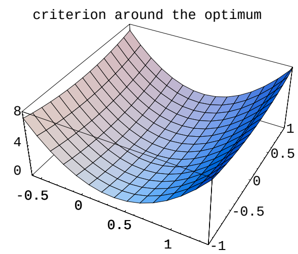
[26] G. Puntonet, C., A. Mansour, and C. Jutten, "Geometrical algorithm for blind separation of sources," in *Actes du XVème colloque GRETSI*, Juan-Les-Pins, France, 18-21 September 1995, pp. 273–276.

[27] A. Prieto, C. G. Puntonet, and B. Prieto, "A neural algorithm for blind separation of sources based on geometric prperties.," *Signal Processing*, vol. 64, no. 3, pp. 315–331, 1998.

[28] A. Mansour, A. Kardec Barros, and N. Ohnishi, "Comparison among three estimators for high order statistics.," in *Fifth International Conference on Neural Information Processing (ICONIP'98)*, Kitakyushu, Japan, 21-23 October 1998, pp. 899–902.

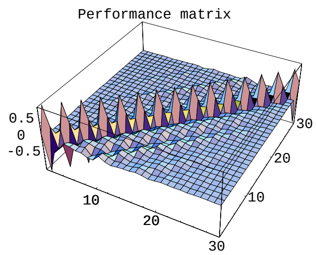


(a) Performance matrix **Perf**

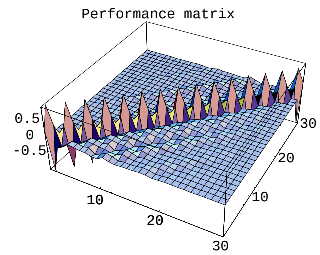


(b) The criterion is flat around its minima.

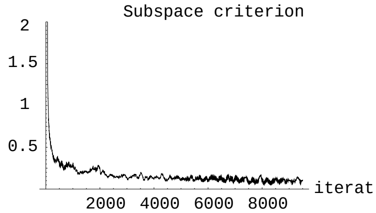
Figure 2: Performances and Properties.



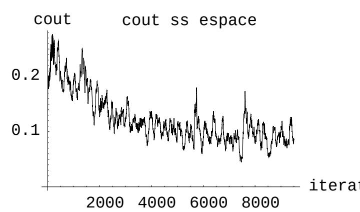
(a) Performance matrix **Perf**, by only using LMS



(b) The Performance matrix, using initialized LMS.

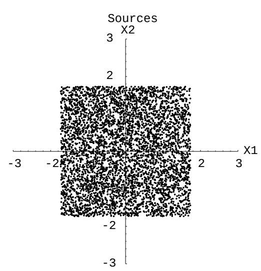


(c) Criterion convergence of the LMS version.

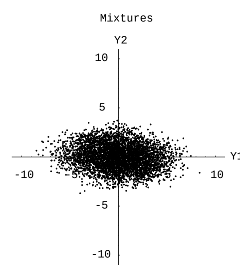


(d) Criterion convergence of the initialized LMS version.

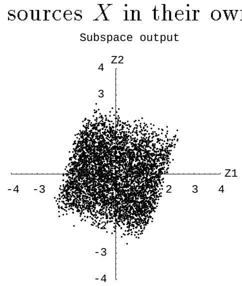
Figure 3: Performances and convergence.



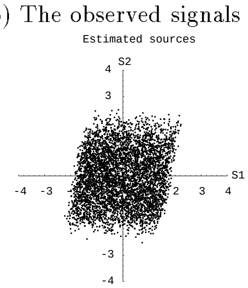
(a) The sources X in their own plane.



(b) The observed signals Y.



(c) The first step output signals Z.



(d) the output signals S.

Figure 4: Different signals.