CRITERIA FOR COMPLEX SOURCES SEPARATION

Eric MOREAU

MS-GESSY, ISITV, Université de Toulon
Av. G. Pompidou, BP 56
83162 La Valette du Var, France
e-mail: moreau@isitv.univ-tln.fr

ABSTRACT

We consider the problem of sources separation. Two necessary and sufficient conditions involving high-order cumulants are given and proved. Hence, a family of criteria for source separation is obtained. A novel gradient based algorithm is derived in order to optimize the proposed criteria and various computer simulations are presented in order to illustrate the performances of the algorithm.

1 INTRODUCTION

The source separation problem is receiving an increasing interest because of practical applications in diverse fields of engineering and applied sciences like communications and array processing: e.g. interference cancellation in transmission and localization of radiating sources with perturbed arrays. It can be simply formulated as follows: several unknown linear mixtures of certain independent signals called sources are observed. The objective is to recover the original sources without knowing the mixing system. Hence this must be realized from the only knowledge of the observations. This is the reason why this kind of approach is often qualified as "blind" or "unsupervised". It is well known that, in general, conventional methods based on second order statistics (i.e. correlation or power spectrum) are not sufficient to solve the problem.

The first designed adaptive algorithms (available in the real case) include non linearity functions in order to test independence [3]. However these adaptations laws are heuristic and might not converge to a separating state. Recently algorithms based on sources separation criterion were proposed. Some of them require a prewhitening stage prior the proper separation, e.g. [1][2][6] while in [4][5] this preprocessing is not necessary. In this paper we consider the problem with a prewhitening stage.

2 PROBLEM FORMULATION

The classical linear memoryless mixture model is considered. It reads

$$\boldsymbol{x} = \boldsymbol{G}\boldsymbol{a} + \boldsymbol{b} \tag{1}$$

where $\boldsymbol{x}$ is the $(N,1)$ vector of observations, $\boldsymbol{a}$ the $(N,1)$ vector of statistically independent sources, $\boldsymbol{b}$ the $(N,1)$ vector of additive noises and $\boldsymbol{G}$ the (N,N) invertible mixture matrix. The sources separation problem consists in estimating a matrix $\boldsymbol{H}$ such that the vector

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} \tag{2}$$

restores the N input sources a_i. With no noise, this means to identify the inverse $\boldsymbol{G}^{-1}$ of the mixture matrix (up to the product of any invertible diagonal matrix and any permutation matrix). For this task, the key hypothesis is the joint independence of the N sources and the non-zero character of some of their cumulants. Thus Gaussian sources are excluded. Without loss of generality, the sources can be assumed zero-mean with unit variance, i.e. $\mathrm{E}\boldsymbol{a}\boldsymbol{a}^H = \boldsymbol{I}$ where the superscript ^{H} means "conjugate and transpose" and E is the mathematical expectation operator. It is useful to define the matrix $\boldsymbol{S}$ of the global system as

$$\boldsymbol{S} \stackrel{\triangle}{=} \boldsymbol{H}\boldsymbol{G} , \tag{3}$$

hence with no noise and according to (1) and (2)

$$\boldsymbol{y} = \boldsymbol{S}\boldsymbol{a} . \tag{4}$$

Because sources are assumed inobservable, there are some inherent indeterminations in their restitution. That is, in general, we cannot identify the power and the order of each sources. Hence they are said separated if and only if the global matrix reads

$$\boldsymbol{S} = \boldsymbol{D}\boldsymbol{P} \tag{5}$$

where $\boldsymbol{D}$ is an invertible diagonal matrix and $\boldsymbol{P}$ a permutation matrix.

3 SOURCE SEPARATION CRITERIA

Contrast functions as defined in [1] constitute separation criteria in the sense that their maximization solve the source separation problem. We consider white vectors $\boldsymbol{y}$, i.e. vectors $\boldsymbol{y}$ such that

$$\mathrm{E}\boldsymbol{y}\boldsymbol{y}^H = \boldsymbol{I} \tag{6}$$

where $\boldsymbol{I}$ is the (N,N) identity matrix. Let us define the function

$$\mathcal{I}_{(p,q)}(\boldsymbol{y}) \triangleq \sum_{i=1}^{N} |\mathrm{C}_{p,q} y_i| \ , \qquad p,q \in \mathbb{N} \qquad (7)$$

where $\mathrm{C}_{p,q} y_i \triangleq \mathrm{Cum}(\underbrace{y_i,\ldots,y_i}_{p\times},\underbrace{y_i^*,\ldots,y_i^*}_{q\times})$ is the (p,q) order cumulant of random variable y_i. The superscript * stands for the conjugate operation and $p\times$ (resp. $q\times$) means that we have p (resp. q) terms.

Proposition 1 *For any statistically independent random vector* $\boldsymbol{a}$*, any orthonormal matrix* $\boldsymbol{S}$ *and for* $p+q \geq 2$*, we have* $\mathcal{I}_{(p,q)}(\boldsymbol{S}\boldsymbol{a}) \leq \mathcal{I}_{(p,q)}(\boldsymbol{a})$.

Proof: Thanks to the independence of the a_i's, we have

$$\begin{aligned}\mathcal{I}_{(p,q)}(\boldsymbol{S}\boldsymbol{a}) &= \sum_{i=1}^{N} |\sum_{j=1}^{N} s_{ij}^p (s_{ij}^*)^q \mathrm{C}_{p,q} a_j| \\ &\leq \sum_{j=1}^{N} |\mathrm{C}_{p,q} a_j| \sum_{i=1}^{N} |s_{ij}|^{p+q} \ .\end{aligned}$$

Now because $\boldsymbol{S}$ is orthonormal, $\forall j$, $\sum_{i=1}^N |s_{ij}|^2 = 1$. With $p+q \geq 2$ one easily has $\forall j$, $\sum_{i=1}^N |s_{ij}|^{p+q} \leq 1$. Thus

$$\mathcal{I}_{(p,q)}(\boldsymbol{S}\boldsymbol{a}) \leq \sum_{j=1}^{N} |\mathrm{C}_{p,q} a_j| = \mathcal{I}_{(p,q)}(\boldsymbol{a}) \qquad (8)$$

and the proposition 1 is proved. •

Proposition 2 *For any statistically independent random vector* $\boldsymbol{a}$ *having at most one null cumulant of order* (p,q) *and for* $p+q \geq 3$, $\mathcal{I}_{(p,q)}(\boldsymbol{S}\boldsymbol{a}) = \mathcal{I}_{(p,q)}(\boldsymbol{a})$ *if and only if* $\boldsymbol{S} = \boldsymbol{D}\boldsymbol{P}$ *where* $\boldsymbol{P}$ *is a permutation and* $\boldsymbol{D} = \mathrm{diag}(d_1,\ldots,d_N)$ *such that* $\forall i$, $|d_i|^2 = 1$.

Proof: The equality in (8) requires the equalities $\sum_{i=1}^N |s_{ij}|^{p+q} = 1$ for all j such that $|\mathrm{C}_{p,q} a_j| \neq 0$. Since $\forall j$, $\sum_{i=1}^N |s_{ij}|^2 = 1$ and $p+q \geq 3$, this is possible if and only if columns j of $\boldsymbol{S}$ have only one non zero component of modulus 1. That is for at least $N-1$ columns because we assume that $\boldsymbol{a}$ has at most one null cumulant of order (p,q). Because $\boldsymbol{S}$ is orthonormal then $\boldsymbol{S} = \boldsymbol{D}\boldsymbol{P}$ where $\boldsymbol{P}$ is a permutation and $\boldsymbol{D} = \mathrm{diag}(d_1,\ldots,d_N)$ such that $\forall i$, $|d_i|^2 = 1$. •

Theorem 1 *The function* $\mathcal{I}_{(p,q)}(\boldsymbol{y})$, $p+q \geq 3$ *is a contrast over the set of white random vectors having at most one null cumulant of order* (p,q).

Proof: One easily has for all permutation matrix $\boldsymbol{P}$, $\mathcal{I}_{(p,q)}(\boldsymbol{P}\boldsymbol{y}) = \mathcal{I}_{(p,q)}(\boldsymbol{y})$ and for all orthonormal diagonal matrix $\boldsymbol{D}$, $\mathcal{I}_{(p,q)}(\boldsymbol{D}\boldsymbol{y}) = \mathcal{I}_{(p,q)}(\boldsymbol{y})$. Propositions 1 and 2 finish the proof. •

Hence by the theorem, for white random vector $\boldsymbol{y}$ deduced from a linear transformation of an independent vector (cf. (3) with $\boldsymbol{b} = 0$), a necessary and sufficient condition for separation is

$$\sum_{i=1}^{N} |\mathrm{C}_{p,q} y_i| = \sum_{i=1}^{N} |\mathrm{C}_{p,q} a_i| \ . \qquad (9)$$

This leads to the following constrained separation criterion

$$\max \sum_{i=1}^{N} |\mathrm{C}_{p,q} y_i| \quad \text{subject to} \quad \mathrm{E}\boldsymbol{y}\boldsymbol{y}^H = \boldsymbol{I} \ . \qquad (10)$$

Now let us consider the specific case of sources with identical sign ε_p of their (p,p) order cumulants, i.e. $\forall i$, $\mathrm{sgn}(\mathrm{C}_{p,p} a_i) = \varepsilon_p$, we have the following theorem:

Theorem 2 *The function* $\mathcal{J}_{(p,p)}(\boldsymbol{y}) \triangleq \varepsilon_p \sum_{i=1}^N \mathrm{C}_{p,p} y_i$, $p \geq 2$, *is a contrast over the set of white random vectors having non null cumulant of order* (p,p).

Proof: Because $\boldsymbol{y} = \boldsymbol{S}\boldsymbol{a}$, we have $\mathrm{C}_{p,p} y_i = \sum_{i=1}^N |s_{ij}|^{2p} \mathrm{C}_{p,p} a_j$. Because $|s_{ij}|^{2p} \geq 0$ and $\forall i$, $\mathrm{sgn}(\mathrm{C}_{p,p} a_i) = \varepsilon_p$ then $\forall i$, $\mathrm{sgn}(\mathrm{C}_{p,p} y_i) = \varepsilon_p$ and the proof follows directly from Theorem 1. •

As previously we can deduce necessary and sufficient conditions for sources separation and the corresponding constrained maximization criterion reads

$$\max \ \varepsilon_p \sum_{i=1}^{N} \mathrm{C}_{p,p} y_i \quad \text{subject to} \quad \mathrm{E}\boldsymbol{y}\boldsymbol{y}^H = \boldsymbol{I} \ . \qquad (11)$$

4 GRADIENT-BASED ALGORITHM

We consider that a first stage realizes a whitening of the observations. This "classical" stage will not be discussed here, see e.g. [6]. The whiteness of $\boldsymbol{y}$ is then ensured if $\boldsymbol{H}$ is orthonormal. In order to find such an orthonormal matrix $\boldsymbol{H}$ which separates the sources, a gradient-based algorithm is proposed in order to maximize the contrast $\mathcal{I}_{(p,q)}(.)$. For this task one can used a parametrization of $\boldsymbol{H}$ e.g. by planar (Givens) rotations as in [1][2][6]. Here we choose not to parametrize and $\boldsymbol{H}$ is updated thanks to

$$\boldsymbol{H}' = \boldsymbol{H} + \lambda \frac{\partial \mathcal{I}_{(p,q)}}{\partial \boldsymbol{H}} \qquad (12)$$

where $\boldsymbol{H}$ and $\boldsymbol{H}'$ are respectively the separating matrices before and after the iteration, λ a positive constant and $\partial\mathcal{I}_{(p,q)}/\partial\boldsymbol{H}$ the matrix whose (ℓ,m) component is $\partial\mathcal{I}_{(p,q)}/\partial h_{\ell m}$. Recalling that $\boldsymbol{H}$ has to be orthonormal, it is easily seen that the iterative procedure (12) does not keep $\boldsymbol{H}$ orthonormal, i.e. $\boldsymbol{H}'$ is not orthonormal. Hence we need a normalization operation in order to maintain the orthonormal character of $\boldsymbol{H}$ thoughout the iterations. This reads

$$\boldsymbol{H}'' = \mathcal{N}(\boldsymbol{H}') \tag{13}$$

where $\boldsymbol{H}''$ is the separating matrix after the normalization. This operation is described now. Consider the singular value decomposition of $\boldsymbol{H}'$: $\boldsymbol{H}' = \boldsymbol{U}\boldsymbol{W}\boldsymbol{V}^H$ where matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthonormal and $\boldsymbol{W}$ is diagonal, then we impose $\boldsymbol{H}'' = \boldsymbol{U}\boldsymbol{V}^H$.

In order to determine the gradient of the contrast $\mathcal{I}_{(p,q)}(.)$, for the sake of simplicity we consider first the real case where $q = 0$ and we denote $\mathcal{I}_p \equiv \mathcal{I}_{(p,0)}$ and $\mathrm{C}_p \equiv \mathrm{C}_{(p,0)}$. Differentiating $\mathcal{I}_p(.)$ with respect to $h_{\ell m}$, one has

$$\frac{\partial \mathcal{I}_p}{\partial h_{\ell m}} = \sum_{i=1}^{N} \mathrm{sgn}(\mathrm{C}_p y_i) \frac{\partial \mathrm{C}_p y_i}{\partial h_{\ell m}} \tag{14}$$

where $\mathrm{sgn}(\mathrm{C}_p y_i)$ is assumed constant. According to (3) in the real case $\mathrm{C}_p y_i = \sum_{j=1}^{N} s_{ij}^p \mathrm{C}_p a_j$ and we have

$$\begin{aligned} \frac{\partial \mathrm{C}_p y_i}{\partial h_{\ell m}} &= p\ \delta_{i\ell} \sum_{j=1}^{N} g_{mj} s_{ij}^{p-1} \mathrm{C}_p a_j \\ &= p\ \delta_{i\ell} \mathrm{Cum}(\underbrace{y_i, \ldots, y_i}_{(p-1)\times}, x_m) \end{aligned} \tag{15}$$

where $\delta_{i\ell} = 1$ if $\ell = i$ and 0 else. Finally using (15) in (14), one obtains

$$\frac{\partial \mathcal{I}_p}{\partial h_{\ell m}} = p\ \mathrm{sgn}(\mathrm{C}_p y_\ell) \mathrm{Cum}(\underbrace{y_\ell, \ldots, y_\ell}_{(p-1)\times}, x_m)\ . \tag{16}$$

In a similar way for the complex case, we have

$$\frac{\partial \mathcal{I}_{(p,p)}}{\partial h_{\ell m}} = 2p\mathrm{sgn}(\mathrm{C}_{p,p} y_\ell) \mathrm{Cum}(\underbrace{y_\ell, \ldots, y_\ell}_{p\times}, \underbrace{y_\ell^*, \ldots, y_\ell^*}_{(p-1)\times}, x_m^*)\ . \tag{17}$$

In the case $p \neq q$, the derivative of $\mathcal{I}_{(p,q)}(.)$ leads to a more complicated expression and therefore will not be presented here.

Consider now the important special case $p = q = 2$. The algorithm is then based on fourth-order cumulants. We have for zero-mean random variables

$$\mathrm{Cyx} = \mathrm{E}|y_\ell|^2 y_\ell x_m^* - 2\mathrm{E}|y_\ell|^2 \mathrm{E} y_\ell x_m^* - \mathrm{E} y_\ell^2 \mathrm{E} y_\ell^* x_m^* \tag{18}$$

where $\mathrm{Cyx} \triangleq \mathrm{Cum}(y_\ell, y_\ell, y_\ell^*, x_m^*)$. Moreover, in order to simplify, we shall further assume that $\forall i$, $\mathrm{E}a_i^2 = 0$ in the complex case. Note that this assumption is not very restrictive e.g. in digital communication where most signal constellations satisfy it. This implies that $\forall \ell$, $\mathrm{E}y_\ell^2 = 0$ and recalling that $\forall \ell$, $\mathrm{E}|y_\ell|^2 = 1$, we have

$$\mathrm{Cyx} = \mathrm{E}|y_\ell|^2 y_\ell x_m^* - 2\mathrm{E} y_\ell x_m^*\ . \tag{19}$$

In practice the exact cumulants are unknown and can only be approximated by their sample estimates using avalaible data. Hence we replace the mathematical expectation in (19) with sample averages

$$\widehat{\mathrm{Cyx}} = \frac{1}{N_d} \sum_{t=1}^{N_d} y_\ell(t) x_m^*(t)(|y_\ell(t)|^2 - 2) \tag{20}$$

where N_d is the number of avalaible data.

Convergence analysis of the proposed algorithm is beyond the scope of this communication. However computer simulations are presented in order to show that the proposed algorithm works.

5 COMPUTER SIMULATIONS

The performances of the algorithm are associated to an index/measure of performance defined on the global matrix $\boldsymbol{S}$ according to

$$\begin{aligned} \mathrm{ind}_\alpha(\boldsymbol{S}) \ \triangleq\ & \frac{1}{2} \left[\sum_i \left(\sum_j \frac{|s_{ij}|^\alpha}{\max_\ell |s_{i\ell}|^\alpha} - 1 \right) \right. \\ & \left. + \sum_j \left(\sum_i \frac{|s_{ij}|^\alpha}{\max_\ell |s_{\ell j}|^\alpha} - 1 \right) \right] \end{aligned} \tag{21}$$

where $\alpha \geq 1$. This positive index is indeed zero if $\boldsymbol{S}$ satisfies (5) and a small value indicates the proximity to the desired solutions.

Simulations are presented in the case of two sources. The mixing matrix is taken orthonormal such that the prewhitening stage is not necessary

$$\boldsymbol{G} = \begin{pmatrix} \cos\theta & i\sin\theta \\ -i\sin\theta & -\cos\theta \end{pmatrix}, \qquad \theta = 40\frac{\pi}{180}\ . \tag{22}$$

We consider the index ind_2 and its initial value is thus 1.41. The algorithm ($p = q = 2$) is tested for different sources, data-block sizes and noise levels. A white Gaussian noise is considered and the SNR is defined as the ratio of the power of a_i to the power of b_i assumed equal for all i. In each cases, we have use 200 Monte-Carlo runs and the average ind_2 is plotted.

In Fig.1 (resp. Fig.2), the input symbols are independent realizations from two 4-QAM (resp. V27) sources and the data-block size equal 100. In all cases, depending on the SNR, the averaged index decrease monotically. With no noise it achieves the steady-state level of −160dB while with SNR=20dB and SNR=10dB the levels are −37dB and −25dB respectively.

In Fig.3, the input symbols are independent realizations from two 16-QAM sources. For data-block sizes $N_d = 100$, 500 and 2000, the averaged index decrease monotically to reach the steady-state level of −23dB, −30dB and −37dB respectively. Fig.4 illustrates the influence of the noise for $N_d = 2000$. It shows the decrease of the steady-state level of the index: −36dB and −32dB when SNR=20dB and SNR=10dB respectively.

References

[1] P. Comon, "Independent Component Analysis, a New Concept?", *Signal Processing*, Vol. 36, no. 3, pp 287-314, April 1994.

[2] M. Gaeta and J.L. Lacoume, "Source Separation Without a Priori Knowledge: the Maximum Likelihood Solution", In Proc. *EUSIPCO'90*, Barcelona, Spain, pp 621-624, 1990.

[3] C. Jutten and J. Herault, "Blind Separation of Sources, Part I: An Adaptative Algorithm Based on Neuromimetic Architecture", *Signal Processing*, Vol. 24, pp 1-10, 1991.

[4] B. Laheld and J.F. Cardoso, "Adaptive Source Separation Without Prewhitening", In Proc. *EUSIPCO'94*, Edinburgh, Scotland, pp 183-186, Sept. 1994.

[5] E. Moreau and O. Macchi, "Complex Self-Adaptive Algorithm for Source Separation Based on High-Order Contrast", In Proc. *EUSIPCO'94*, Edinburgh, Scotland, pp 1157-1160, Sept. 1994.

[6] E. Moreau and O. Macchi, "High Order Contrasts for Self-Adaptive Source Separation", *International Journal of Adaptive Control and Signal Processing*, Vol. 10, No. 1, pp 19-46, January 1996.

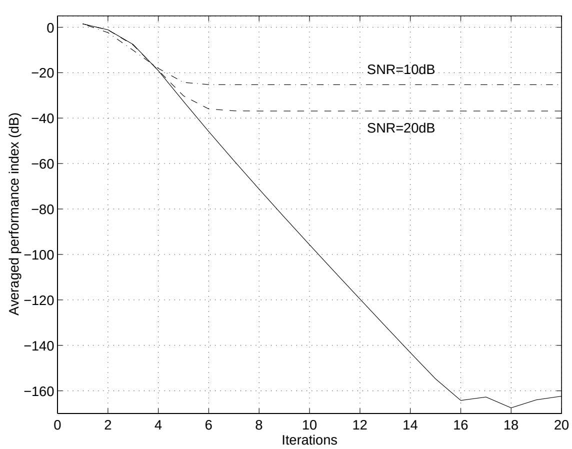

Figure 1: Averaged performance index for two 4-QAM sources, for different SNR and $N_d = 100$.

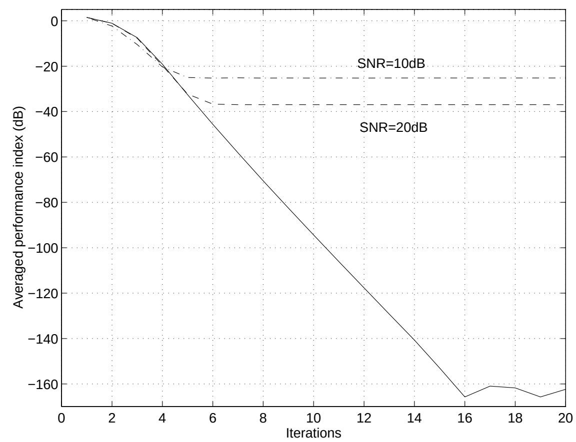

Figure 2: Averaged performance index for two V27 sources, for different SNR and $N_d = 100$.

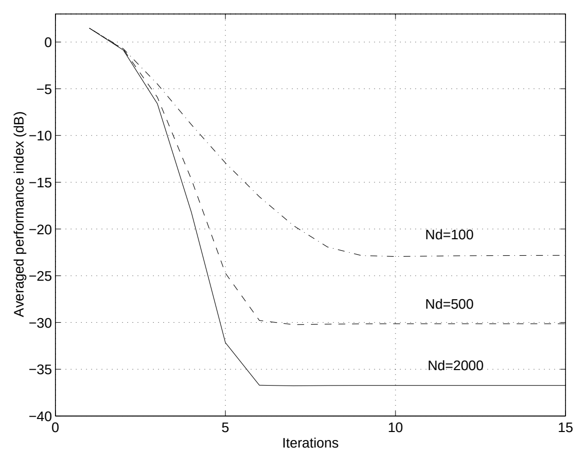

Figure 3: Averaged performance index for two 16-QAM sources, for different N_d and with no noise.

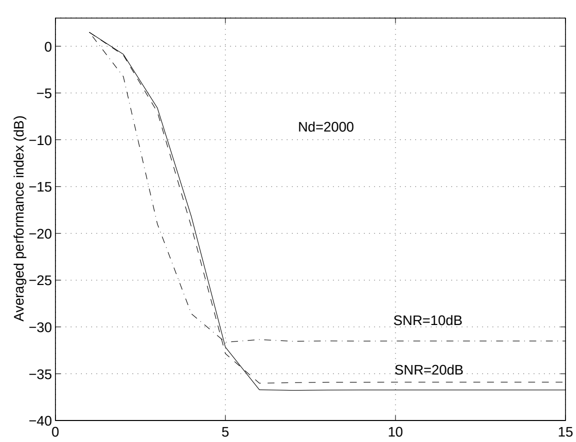

Figure 4: Averaged performance index for two 16-QAM sources, for different SNR and $N_d = 2000$.