UNSUPERVISED SEPARATION OF DISCRETE SOURCES WITH A COMBINED EXTENDED ANTI-HEBBIAN ADAPTATION

Zied Malouche and Odile Macchi
Laboratoire des Signaux et Systèmes, CNRS, Supélec
Plateau de Moulon 91192 Gif-sur-Yvette Cedex FRANCE
Groupement de Recherche TdSI du CNRS
e-mail: malouche@lss.supelec.fr

ABSTRACT

In the classical methods of unsupervised source separation, the a priori hypothesis is independence of sources. In certain applications, there is some additional knowledge on the sources (statistics, distributions, alphabet...). It is the case with discrete sources with known alphabet. Then we can improve separation. Initialization of adaptation is done according to some known algorithm, e.g. thanks to an extended anti-Hebbian algorithm, provided there are not less sensors than sources. As soon as the separation performance index has reached some preassigned level, a second part which involves the output decision error is introduced in the increment. In a noiseless environment, this method allows complete cancellation of steady state adaptation fluctuations and perfect source recovery.

1 Introduction

The unsupervised separation of independent discrete sources is a problem encountered in many applications, e.g. mobile radio telephone, digital television and neural networks for discrimination of mixed biological components. Let $\mathbf{x}$ be an m dimensional observation vector obtained by linearly mixing the p independent, multi-level discrete components a_i of a source vector $\mathbf{a}$. The linear mixture model can be written

$$\mathbf{x} = \mathbf{G}\mathbf{a}, \tag{1}$$

where $\mathbf{G}$ is the $m \times p$ mixture matrix. The two quantities $\mathbf{a}$ and $\mathbf{G}$ are unknown. The aim is to design m vectors $\mathbf{h}_j$ working on $\mathbf{x}$, such that the m outputs

$$y_j = \mathbf{x}^T \mathbf{h}_j \ , \ j = 1, \ldots, m \tag{2}$$

recover the p discrete sources without taking into account neither their order nor their powers, using the a priori knowledge that sources are independent and discrete. Note that m is taken larger or equal to p and it is assumed that

$$rank(\mathbf{G}) = p \tag{3}$$

or equivalently that $\mathbf{G}^T\mathbf{G}$ is regular. Some redundancy is admissible when the same source is restored on several outputs at the same time.

The unsupervised extended anti-Hebbian algorithm introduced in [1][2] allows the output y_j to recover one of the p independent sources contained in $\mathbf{x}$, e.g. a_i, provided that its kurtosis is negative. In particular for binary sources, the convergence performance is perfect [1]. However, for non binary sources, either the convergence speed or the steady state performance are deteriorated (compared to those of the binary case). To overcome this drawback, we introduce a novel optimization criterion which corresponds to a weighting of the two kinds of knowledge about sources: their independence and their discreteness. This idea was originally introduced in [3] and [4] for the case of data transmission equalizers.

This solution is computationally simple: the algorithm increment involves only $O(m)$ multiplications per output. It is modular in the sense that each vector $\mathbf{h}_j$ is updated independently of the others and with the same rule thanks to a multiple-input-single-output (MISO) device. So, if one vector $\mathbf{h}_j$ is poor, this has no degrading influence upon the others. Consequently the recovery of a specific source a_i at the output y_j only depends on the initialization $\mathbf{h}_j(0)$ of the vector $\mathbf{h}_j(n)$ and on the sequence $\mathbf{x}(n)$. Moreover, one does not need to know the number p of sources provided m is large enough ($m \geq p$).

Finally, the weighting coefficient between the two parts of the algorithm can be made adaptive: as separation proceeds with time, the criterion focuses on discreteness which improves the performance still more. All these achievements are confirmed by computer simulations.

2 Source extraction (general case)

Extraction of a source is the recovery on one output y_j of a single source a_i out of the mixture $\mathbf{x}$. In section 2, 3, 4, we shall drop the index j of y_j and $\mathbf{h}_j$. It is irrelevant because the m MISO moduli perform the same way, and we work with only one of them. Extraction of one source a_i can be performed by searching for the vector $\mathbf{h}$ which minimizes the criterion

$$J_s(\mathbf{h}) = E\{F(y)\} \tag{4}$$

where F is a polynomial like

$$F(y) = y^4 - 2y^2. \tag{5}$$

This can be done adaptively thanks to the associated stochastic gradient algorithm

$$\mathbf{h}(n) = \mathbf{h}(n-1) - \mu \mathbf{x}(n) f(y(n)) \tag{6}$$

where f is the derivative of F. As shown in [1][2], this is an extended anti-Hebbian updating. The modified vector

$$\mathbf{c} \triangleq \mathbf{G}^T \mathbf{h}, \tag{7}$$

obeys the algorithm

$$\mathbf{c}(n) = \mathbf{c}(n-1) - \mu \mathbf{G}^T \mathbf{G} \mathbf{a}(n) f(\mathbf{a}(n)^T \mathbf{c}(n-1)) \tag{8}$$

which is shown [1] to be equivalent to (6). The vector $\mathbf{c}$ characterizes the cascade mixture-extraction in the sense that

$$y = \mathbf{a}^T \mathbf{c}. \tag{9}$$

Hence the good equilibria of (8) are those roots $\overline{\mathbf{c}}$ of the equation

$$E\{\mathbf{a} f(\mathbf{a}^T \mathbf{c})\} = 0 \tag{10}$$

which are parallel to an axis. This yields

$$\overline{\mathbf{c}}^{\pm i} = \frac{\pm 1}{\sqrt{A_i}} (0, \ldots, 0, 1, 0, \ldots, 0)^T, \tag{11}$$

where the 1 stands in place i,

$$A_i = E\{a_i^4\}, \tag{12}$$

and we have assumed, without loss of generality, that

$$E\{a_i^2\} = 1. \tag{13}$$

It can be shown that the $\overline{\mathbf{c}}^{\pm i}$ are stable [1] provided the source a_i is sub-Gaussian i.e.

$$A_i < 3. \tag{14}$$

Then according to (9) and (11)

$$\sqrt{A_i} y - (\pm a_i) = 0. \tag{15}$$

So the source a_i is recovered (up to a fixed sign) with a gain that only depends on the fourth order moment of that source, and does not depend on the mixture. This fact will be exploited later to build up the second part of the criterion.

An attraction basin $B^{\pm i}$ is assigned to each good equilibrium $\overline{\mathbf{c}}^{\pm i}$. The specific source that is extracted corresponds to the attraction basin where the initial vector $\mathbf{c}(0) = \mathbf{G}^T \mathbf{h}(0)$ is located.

3 Improvement for discrete sources

The algorithm (6) properly performs the extraction of sub-Gaussian sources. Especially for binary sources, the extraction is perfect as explained in [1]. Nevertheless, for non binary sources, in steady state the algorithm exhibits noticeable residual fluctuations. The following method will improve the extraction quality of a non binary discrete source a_i thanks to the a priori knowledge of its alphabet S_i. A decision function $Q_i(z)$ can be associated in an obvious manner to S_i: $Q_i(z)$ takes the level of S_i that is the closest to z (see Fig.2). Then define the "decision error" as

$$\varepsilon = \sqrt{A_i} y - Q_i(\sqrt{A_i} y). \tag{16}$$

According to (15), ε is the decision error on the output y when the source a_i is to be extracted, because $\mathbf{h}$ is close enough to the corresponding extracting vector. This suggests that, to recover the source a_i in a perfect way, $\mathbf{h}(n)$ could be updated in order to minimize the criterion function

$$J_d(\mathbf{h}) = \frac{1}{2} E\{\varepsilon^2\}, \tag{17}$$

whose associated stochastic gradient algorithm is

$$\mathbf{h}(n) = \mathbf{h}(n-1) - \mu \sqrt{A_i} \mathbf{x}(n) \varepsilon(n). \tag{18}$$

So the modified vector $\mathbf{c}$ obeys the algorithm

$$\mathbf{c}(n) = \mathbf{c}(n-1) - \mu \mathbf{G}^T \mathbf{G} \sqrt{A_i} \mathbf{a}(n) \varepsilon(n). \tag{19}$$

Because $\mathbf{G}^T \mathbf{G}$ is supposed regular, the equilibria of (19) (resp. (18)) are the roots in $\mathbf{c}$ (resp. $\mathbf{h}$) of the equation

$$E\{\mathbf{a}(n) \varepsilon(n)\} = \mathbf{0}, \tag{20}$$

or equivalently

$$E\{\mathbf{a}\, dec(\sqrt{A_i} \mathbf{a}^T \mathbf{c})\} = \sqrt{A_i} \mathbf{c}. \tag{21}$$

It is easy to show that $\overline{\mathbf{c}}^{\pm i}$ defined in (11) is a root of (21). The algorithm (19) is not capable by itself of converging to the desired solution (11) if $\mathbf{c}(0)$ is too bad. This algorithm will always be run when $\mathbf{c}$ is in the vicinity of $\overline{\mathbf{c}}^{\pm i}$. It requires what is called the "open eye" condition in the case of equalization. Example: in the PAM case with $2N_i$ levels, the source a_i is equally distributed and the alphabet

$$S_i = \{\pm(2n-1)\lambda_i,\ n = 1, \ldots, N_i\} \tag{22}$$

is equidistant. Then

$$\lambda_i = \frac{\sqrt{3}(2n-1)}{\sqrt{(2N_i-1)(2N_i+1)}} \tag{23}$$

$$A_i = \frac{3(12N_i^2 - 7)}{5(2N_i-1)(2N_i+1)}. \tag{24}$$

Then the "open eye" condition reads

$$\left| \sqrt{A_i} |y| - |a_i| \right| < \lambda_i. \tag{25}$$

Thus, when (25) holds, the adaptive algorithm (19) (resp. (18)) will eliminate the residual fluctuations of $\mathbf{c}(n)$ (resp. $\mathbf{h}(n)$) in the vicinity of the extracting equilibrium.

4 New combined algorithm

For the sake of simplicity, it is assumed that

$$\begin{cases} S_i = S \\ A_i = A \end{cases} \forall i. \tag{26}$$

To perform high quality extraction, we combine the two criteria (4) and (17) with a weighting coefficient $\alpha \in [0,1]$

$$J(\mathbf{h}) = \alpha J_s(\mathbf{h}) + (1-\alpha) J_d(\mathbf{h}). \tag{27}$$

The corresponding gradient algorithm clearly reads

$$\begin{aligned} \mathbf{h}(n) &= \mathbf{h}(n-1) \\ &\quad -\mu\{\alpha \mathbf{x}(n) f(y(n)) + (1-\alpha)\sqrt{A}\mathbf{x}(n)\varepsilon(n)\}. \end{aligned} \tag{28}$$

The second part of the increment should not be activated while the vector $\mathbf{h}(n)$ is not in the vicinity of an extracting equilibrium. So α should be close to 1 at the beginning, but should approach 0 at the end. One possible strategy is

$$\begin{cases} \alpha(n) \equiv 1 & n < K \\ \alpha(n) \equiv 0 & n \geq K \end{cases}. \tag{29}$$

$\alpha(n)$ can also be made decreasing in a monotonic relationship versus $\varepsilon(n)$ e.g.

$$\alpha(n) = tanh(10(|\varepsilon_1(n)| + |\varepsilon_2(n)|)). \tag{30}$$

5 Source separation

Now to perform source separation, it is necessary to extract the p independent sources $a_i, i = 1, \ldots, p$. For this sake, m differently initialized vectors $\mathbf{h}_j(n)$ are adapted with the algorithm (28). The equilibrium $\overline{\mathbf{c}}^{\pm i}$ has an attraction basin $B^{\pm i}$. At least one vector $\mathbf{c}_j(0) = \mathbf{G}^T \mathbf{h}_j(0)$ must belong to $B^{\pm i}$. The separating system is made up of m similar MISO moduli depicted in (Fig.1). Note the local character of this architecture that allows each modulus to perform independently of the other moduli.

6 Computer simulations

Denote

$$\mathbf{H} \stackrel{\Delta}{=} [\mathbf{h}_1, \ldots, \mathbf{h}_m] \tag{31}$$

the matrix of all moduli and the modified matrix

$$\mathbf{C} \stackrel{\Delta}{=} \mathbf{G}^T \mathbf{H} = [\mathbf{c}_1, \ldots, \mathbf{c}_m]. \tag{32}$$

In the following we take $m = p$. A positive separation index is associated to $\mathbf{C}$. It is zero if and only if separation is achieved by $\mathbf{H}$, that is if

$$\mathbf{C} = \mathbf{\Lambda P}, \tag{33}$$

where $\mathbf{\Lambda}$ is a diagonal matrix and $\mathbf{P}$ a permutation one. Simulations are run for $m = p = 2$, with the mixture matrix

$$\mathbf{G} = \begin{bmatrix} 1 & 0.4 \\ 0.5 & 1 \end{bmatrix}. \tag{34}$$

The two discrete sources are 8-PAM sources. The initial matrix is $\mathbf{H}(0) = \mathbf{I}$. The separation index is plotted in Fig.3 with $\alpha = 1$ (this corresponds to algorithm (6)). The separation is performed but noticeable residual fluctuations remain in steady state.

In Fig.4, as soon as $\alpha(n)$ is switched to 0 (at step $K = 800$), the separation index decreases to reach the null value. Separation is perfect.

With an adaptive α according to (30), Fig.5 clearly shows that the separation index and α are jointly vanishing to yield perfect separation. The contribution of J_d in the criterion (27) allows both perfect and very fast separation.

7 Conclusion

The adaptive combined anti-Hebbian algorithm proposed in this paper performs fast perfect discrete source separation in a noiseless environment. Its implementation is very simple. It has a local character in the sense that the m MISO moduli are adapted independently. This great improvement is realized thanks to the a priori knowledge of the alphabet of sources and also because the sources are recovered with a gain that does not depend on the mixture. To totally cancel the residual fluctuations, the second part added in the algorithm is based on the discreteness of the sources. It is switched on as soon as the separation has reached a given quality. A simpler possibility is to activate this second part in an adaptive way. As a result, separation is perfect.

References

[1] Z. Malouche and O. Macchi, "Adaptive Unsupervised Extraction of One Component with a Single Neuron", *Submitted to IEEE Neural Networks.*

[2] Z. Malouche and O. Macchi, "Extended Anti-Hebbian Adaptation For Unsupervised Source Extraction", *Proc. ICASSP 96, Session DSP 10, #10.*

[3] O. Macchi, C.A.F. Da Rocha and J.M.T. Romano, "Egalisation adaptative autodidacte par rétroprédiction et prédiction", Actes du 14eme colloque GRETSI, Juan-Les-Pins, Septembre 1993, pp. 491-494.

[4] C.A.F. Da Rocha and O. Macchi, "A novel self adaptive recursive equalizer with unique optimum for QAM", Proc. ICASSP, Adelaide, pp III 481-484, 1994.

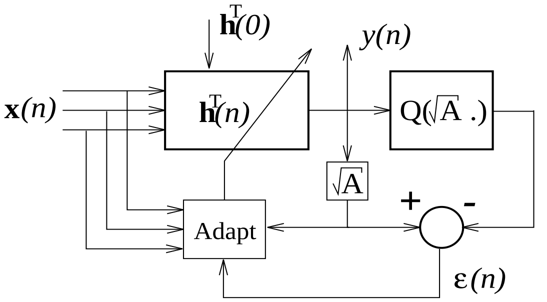

Figure 1: One of the m identical MISO moduli

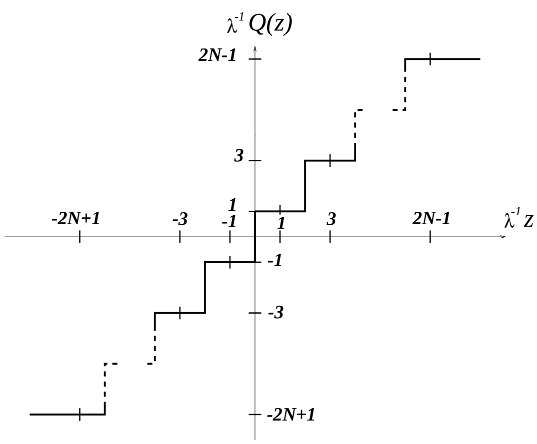

Figure 2: Decision function for $2N$ equidistant levels

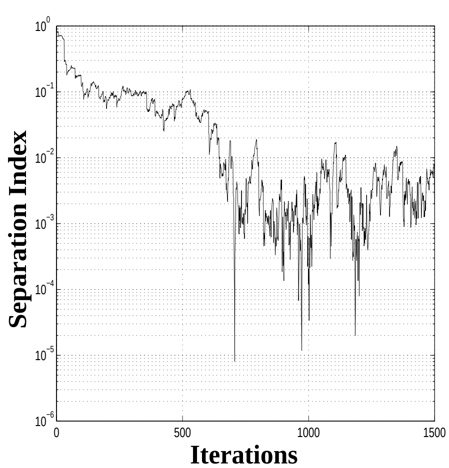

Figure 3: Separation Index for two independent 8-PAM sources: $\alpha(n) \equiv 1$, $\mu = 0.01$.

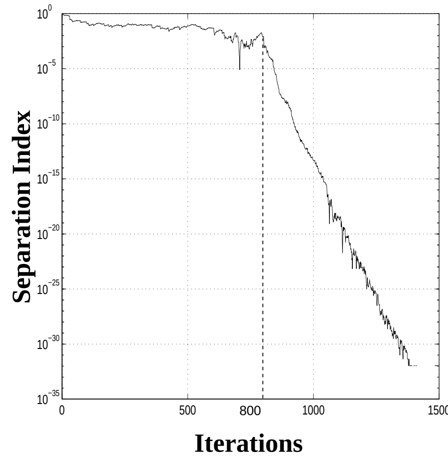

Figure 4: Separation Index for two independent 8-PAM sources: $\alpha(n)$ changes abruptly from 1 to 0 at $n = K = 800$, $\mu = 0.01$.

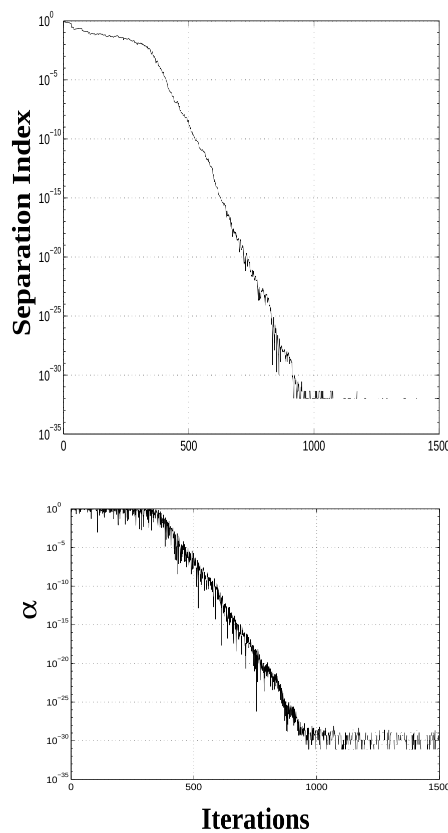

Figure 5: Separation Index for two independent 8-PAM sources: $\alpha(n) = tanh(10(|\varepsilon_1(n)| + |\varepsilon_2(n)|))$, $\mu = 0.01$.