

# Séparation de sources

## Application à la séparation de signaux et de brouilleurs dans un satellite de télécommunications

Svetlana Bozinoski<sup>(1)</sup>, Anisse Taleb<sup>(1)</sup>, Jean-Claude Guizzo<sup>(2)</sup> et Christian Jutten<sup>(1)</sup>

<sup>(1)</sup>INPG-TIRF, 46, avenue Félix Viallet, 38031 Grenoble Cedex, France

<sup>(2)</sup>ALCATEL ESPACE Toulouse

### RÉSUMÉ

Dans ce papier nous appliquons les récents développements en séparation de sources dans un contexte réaliste de communications spatiales. Un nouvel algorithme est dérivé. Des résultats expérimentaux sont exposés et valident l'approche.

### ABSTRACT

In this paper we apply the recent source separation algorithms in a realistic context of spatial communications. A new algorithm is derived. Some computer simulations are brought at the end of this paper.

## 1 Introduction

Nous avons appliqué les récents algorithmes adaptatifs de séparation de sources à la séparation d'émetteurs dans un système de communications spatiales. La nature adaptative présente l'avantage d'être facile à implanter, de traiter les données en ligne et de rendre possible la poursuite des modèles.

Bien que remplissant déjà les objectifs principaux qui leur sont assignés, les architectures à bord des satellites peuvent être améliorées à la fois du point de vue des performances en communication (meilleure résistance au brouillage), et du point de vue de la séparation des sources d'émissions indésirables (brouilleurs).

Les signaux utilisés sont des signaux de télécommunications (sources utiles) dont les bits d'information sont modulés en BPSK, puis en QPSK par une séquence d'étalement. Les signaux reçus sont tels qu'ils seraient reçus par l'antenne d'un satellite de télécommunications. Dans le scénario de communication, nous avons considéré un modèle de mélange linéaire instantané à valeurs complexes. En effet, les signaux sont modulés et donc complexes, l'influence de l'antenne est modélisée par une multiplication par scalaires complexes. On suppose également que les signaux sources sont circulaires. Le mélange s'écrit :

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \quad (1)$$

où  $\mathbf{s}(t)$  est le vecteur des sources inconnues et supposées statistiquement indépendantes,  $\mathbf{x}(t)$  est le vecteur des signaux observés (mélanges) et  $\mathbf{A}$  est la matrice de mélange inconnue à valeurs complexes. Les sources au sol sont proches : elles sont situées dans le lobe principal de directivité de l'antenne du satellite. Les capteurs de réception sur le satellite étant très

proches également, le mélange peut être très mal conditionné. Les brouilleurs considérés sont des sources dont les propriétés statistiques sont inconnues. Le brouilleur est un signal pulsé paramétré de telle sorte qu'il brouille l'ensemble du canal.

## 2 Algorithmes adaptatifs de séparation

L'objectif de la séparation aveugle de sources est de retrouver les sources ne connaissant que les vecteurs d'observation  $\mathbf{x}(t)$  et le nombre  $n$  de sources. Généralement, on cherche à estimer une matrice  $\mathbf{B}$  telle que  $\mathbf{B}\mathbf{A} = \mathbf{\Lambda}\mathbf{P}$  où  $\mathbf{\Lambda}$  est une matrice diagonale à coefficients complexes et  $\mathbf{P}$  est une matrice de permutation.

En séparation adaptative de sources, on met à jour une matrice  $\mathbf{B}$  de dimension  $(n \times n)$  de sorte que le vecteur de dimension  $n$  :  $\mathbf{y}(t) = \mathbf{B}\mathbf{x}(t)$  soit à composantes indépendantes.

Après l'étude des algorithmes de Héroult-Jutten [1], Laheld - Cardoso [2], Cichocki-Amari [3], Bell-Sejnowski [7], Wang-Karhunen (bigradient) [12], nous avons choisi les algorithmes de la forme :

$$\mathbf{B}(t+1) = \mathbf{B}(t) + \mu_t \mathbf{H}\mathbf{B}(t) = (\mathbf{I} + \mu_t \mathbf{H})\mathbf{B}(t) \quad (2)$$

où  $\mu$  est le pas d'adaptation et  $\mathbf{H}$  une matrice qui ne dépend que des sorties du système de séparation. Ce type d'algorithmes, introduit sous le nom de EASI par Cardoso *et al.* [2], bénéficie de la propriété d'équivariance qui est remarquable dans le cas de mélanges mal conditionnés.

La mise à jour de la matrice séparante  $\mathbf{B}$  dans l'algorithme EASI est faite selon (2), avec  $\mathbf{H} = \mathbf{H}(\mathbf{y}(t))$  tel que :

$$\mathbf{H}(\mathbf{y}) = \frac{\mathbf{y}\mathbf{y}^\dagger - \mathbf{I}}{1 + \mu_t \mathbf{y}^\dagger \mathbf{y}} + \frac{\mathbf{f}(\mathbf{y})\mathbf{y}^\dagger - \mathbf{y}\mathbf{f}(\mathbf{y})^\dagger}{1 + \mu_t \|\mathbf{y}^\dagger \mathbf{f}(\mathbf{y})\|} \quad (3)$$

où  $\mathbf{y}^\dagger$  est le transposé conjugué de  $\mathbf{y}$ , et  $\mathbf{f}(\cdot)$  est le vecteur  $\mathbf{f} = [f_1(\cdot) \dots f_N(\cdot)]^T$ .

Ch. Jutten est professeur à l'ISTG de l'Université Joseph Fourier de Grenoble.

Ce travail a été financé par le contrat DRET  $n^o$  95-492

Dans le cadre des communications, on applique l'algorithme (3) avec des fonctions nonlinéaires cubiques :  $f(y) = |y|^2 y$ . Dans le cas de mélanges de sources circulaires, la convergence est assurée en quelques milliers d'itérations et les performances obtenues sont d'excellente qualité :  $-20$  à  $-30$  db de diaphonie résiduelle avec des sources de même puissance distantes de  $0.1$  degré. Cependant avec ces fonctions cubiques, Cardoso et Laheld ont montré que l'algorithme est asymptotiquement stable à la condition que la somme des kurtosis des sources soit inférieure à  $0$ . Si cette condition est vérifiée avec des signaux de communications, ce n'est pas forcément le cas avec des brouilleurs quelconques : l'algorithme peut alors converger vers des valeurs erronées. Une solution heuristique à ce problème consiste à changer le signe de la fonction nonlinéaire qui s'applique au brouilleur  $y_b$  :  $f(y_b) = -|y_b|^2 y_b$ . Cette approche est cependant peu satisfaisante et peu réaliste quand le nombre de sources est supérieur à  $2$ . Enfin, le choix des fonctions découle d'une fonction de contraste et suppose que les sources sont circulaires, condition que ne vérifient pas forcément les brouilleurs.

### 3 Approche basée sur le minimum d'information mutuelle

La majorité des algorithmes de séparation de sources, basés sur l'hypothèse de l'indépendance statistique des sources, utilisent des fonctions nonlinéaires choisies *a priori* ou dérivées d'une fonction de contraste.

Dans le cas de mélanges linéaires instantanés réels, il a été prouvé que les fonctions nonlinéaires réalisant l'optimalité d'un point de vue taux de réjection ou variance de l'estimation [8, 11, 2], sont proportionnelles à la dérivée logarithmique des densités de probabilité marginales des sources  $\psi_i(u) = p'_{s_i}/p_{s_i}(u)$ .

Dans le cas complexe, un cadre théorique autorisant la manipulation directe de variables aléatoires à valeurs complexes a été introduit par Amblard *et al.* [6]. En utilisant ces outils, nous allons dériver le critère d'indépendance et en déduire un algorithme pour la séparation de sources dans le cas complexe. L'indépendance statistique des signaux de sortie peut être formulée dans le cas complexe de la manière suivante :

$$p_{\mathbf{Y}, \mathbf{Y}^*}(\mathbf{y}, \mathbf{y}^*) = \prod_{i=1}^n p_{Y_i, Y_i^*}(y_i, y_i^*) \quad (4)$$

Une manière classique pour mesurer l'indépendance est l'utilisation de la divergence de Kullback-Liebler entre les deux termes de (4). Ceci nous donne :

$$\mathcal{J}(\mathbf{Y}, \mathbf{Y}^*) = \sum_{i=1}^n H(Y_i, Y_i^*) - H(\mathbf{Y}, \mathbf{Y}^*) \quad (5)$$

Cette quantité n'est rien d'autre que l'information mutuelle entre les couples  $(Y_i, Y_i^*)$ .  $H(\cdot)$  désigne l'entropie différentielle.  $I(\mathbf{Y})$  est positive, et s'annule quand les sorties sont statistiquement indépendantes. Cette quantité est donc à minimiser par rapport à la matrice  $\mathbf{B}$ .

En notant  $\mathbf{Y} = \mathbf{B}\mathbf{X}$  et  $\mathbf{Y}^* = \mathbf{B}^*\mathbf{X}^*$ , l'entropie des sorties peut être écrite sous la forme suivante :

$$H(\mathbf{Y}, \mathbf{Y}^*) = H(\mathbf{X}, \mathbf{X}^*) + \log |\det(\mathbf{B}\mathbf{B}^*)| \quad (6)$$

Le premier terme correspond à l'entropie conjointe des entrées (mélange) et ne dépend pas de  $\mathbf{B}$ . Le gradient par rapport à  $\mathbf{B}$  de  $H(\mathbf{Y}, \mathbf{Y}^*)$  s'écrit donc :

$$\frac{\partial}{\partial \mathbf{B}} H(\mathbf{Y}, \mathbf{Y}^*) = \frac{\partial}{\partial \mathbf{B}} \log |\det(\mathbf{B}\mathbf{B}^*)| = 2\mathbf{B}^{-\dagger} \quad (7)$$

Le gradient des autres termes est facile à calculer, pourvu qu'on prenne des précautions lors de la dérivation par rapport aux variables complexes, et donne finalement pour  $\mathcal{J}(\mathbf{Y}, \mathbf{Y}^*)$  :

$$\frac{\partial \mathcal{J}}{\partial \mathbf{B}} = -2E[\mathbf{I} + \Psi(\mathbf{y}, \mathbf{y}^*)\mathbf{y}^\dagger]\mathbf{B}^{-\dagger}, \quad (8)$$

où  $\Psi(\mathbf{y}, \mathbf{y}^*)$  est un vecteur  $n \times 1$ , dont la composante  $i$  est égale à :

$$\psi_{Y_i, Y_i^*}(y_i, y_i^*) = \frac{\partial}{\partial y_i^*} \log p_{Y_i, Y_i^*}(y_i, y_i^*)$$

L'algorithme stochastique *équivalent* [2] qui résulte du critère d'information mutuelle minimale s'écrit sous la forme (2) avec :

$$h_{ij}(\mathbf{y}) = \begin{cases} 0 & \text{si } i = j, \\ \psi_{Y_i, Y_i^*}(y_i, y_i^*)y_j^* & \text{sinon.} \end{cases} \quad (9)$$

En fait, le terme  $h_{ii}(\mathbf{y})$  devrait être égal dans (9) à  $1 + \psi_{Y_i, Y_i^*}(y_i, y_i^*)y_i^*$ . Mais, ce terme est nul en moyenne et donc ne sera pas pris en compte dans la version stochastique. En effet :

$$\begin{aligned} E[1 + \psi_{Y_i, Y_i^*}(y_i, y_i^*)y_i^*] &= 1 + \int p_{Y_i, Y_i^*}(y_i, y_i^*)y_i^* \frac{\partial}{\partial y_i^*} \\ &\quad \log p_{Y_i, Y_i^*}(y_i, y_i^*) dy_i dy_i^* \\ &= 1 + \int y_i^* \frac{\partial}{\partial y_i^*} p_{Y_i, Y_i^*}(y_i, y_i^*) dy_i dy_i^* \end{aligned}$$

et en intégrant par parties le second membre, on trouve  $-1$ .

Le développement précédent a l'avantage de ne pas supposer que la densité de probabilité des sources est circulaire. Dans [10], A. Taleb et C. Jutten présentent une méthode d'estimation des fonctions  $\psi_i$  dans le cas réel multivariable. L'extension de cette méthode au cas complexe est simple, et peut être considérée comme un cas réel à 2 variables. Nous utilisons pour cela un réseau de type perceptron multicouches pour chaque fonction  $\psi_{Y_i, Y_i^*}(y_i, y_i^*)$  à estimer. Chaque réseau de neurones reçoit deux entrées, correspondant à  $\Re(y_i)$  et  $\Im(y_i)$  et fournit en sortie des estimations  $h^{Re}(\mathbf{w}_i, \Re(y_i), \Im(y_i))$  et  $h^{Im}(\mathbf{w}_i, \Re(y_i), \Im(y_i))$  de  $\Re(\psi_{Y_i, Y_i^*}(y_i, y_i^*))$  et  $\Im(\psi_{Y_i, Y_i^*}(y_i, y_i^*))$ . Les paramètres du réseau  $\mathbf{w}_i$  sont estimés de façon à minimiser l'erreur quadratique :

$$\begin{aligned} \mathcal{E}_i &= E[(h^{Re}(\mathbf{w}_i, \Re(y_i), \Im(y_i)) - \Re(\psi_{Y_i, Y_i^*}(y_i, y_i^*)))^2 \\ &\quad + (h^{Im}(\mathbf{w}_i, \Re(y_i), \Im(y_i)) - \Im(\psi_{Y_i, Y_i^*}(y_i, y_i^*)))^2] \end{aligned}$$

En calculant le gradient de  $\mathcal{E}_i$ , on en déduit l'algorithme de gradient correspondant qui en fait ne nécessite pas la connaissance de la fonction désirée  $\psi_{Y_i, Y_i^*}(y_i, y_i^*)$ .

En raison de l'indétermination sur le module des signaux de sorties, une modification doit être apportée à l'algorithme (9). Celle-ci consiste à rajouter des contraintes sur la variance

des sorties. Nous proposons la contrainte sur les éléments diagonaux de  $\mathbf{H}$  :

$$h_{ii}(\mathbf{y}) = 1 - y_i y_i^*, \psi \quad (10)$$

qui impose à la convergence que la variance des sources estimées soit égale à l'unité.

## 4 Validation expérimentale

Nous considérons 2 sources de kurtosis différents (Fig. 1) et 2 capteurs. Le mélange (Fig. 2) est pollué par un bruit additif  $\mathbf{b}(t)$ , centré blanc gaussien, de matrice de covariance  $E[\mathbf{b}\mathbf{b}^\dagger] = \sigma^2 \mathbf{I}$ , et s'écrit :

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{b}(t). \psi \quad (11)$$

Dans cet exemple  $\sigma^2 = 0.1$ .

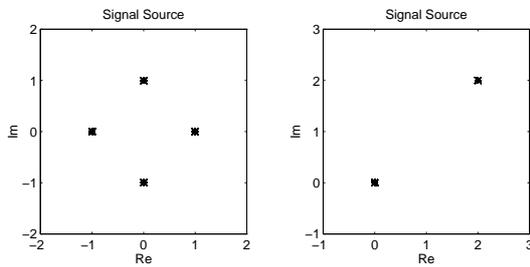


FIG. 1 — Source et brouilleur

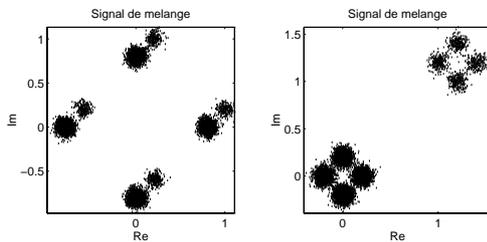


FIG. 2 — Signaux de mélange

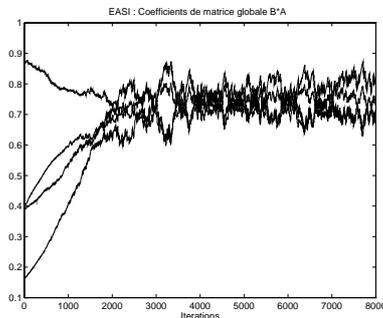


FIG. 3 — Evolution temporelle du module de  $\mathbf{BA}$  pour l'algorithme EASI.

Nous pouvons constater sur les figures 3 et 4 que l'algorithme EASI converge mal (la matrice  $\mathbf{AB}$  n'est pas diagonale !). Ceci est dû au fait que, le kurtosis du brouilleur valant 400, la condition de stabilité de l'algorithme EASI n'est plus

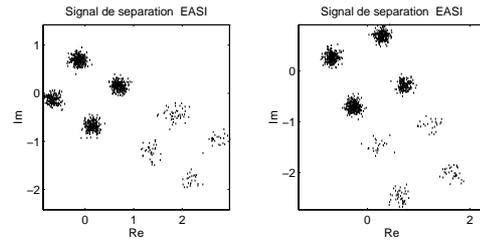


FIG. 4 — Constellation des sources estimées par l'algorithme EASI

vérifiée. Pour y remédier, on peut utiliser les fonctions non linéaires cubiques avec des signes alternés :

$$\mathbf{f} = \begin{bmatrix} |y_1|^2 y_1 & -|y_2|^2 y_2 \end{bmatrix}^T \quad (12)$$

Avec cette modification, nous constatons la convergence de l'algorithme sur les figures 5 (la matrice  $\mathbf{BA}$  devient approximativement diagonale). Nous retrouvons sur la figure 6 une estimation des sources (bruitées).

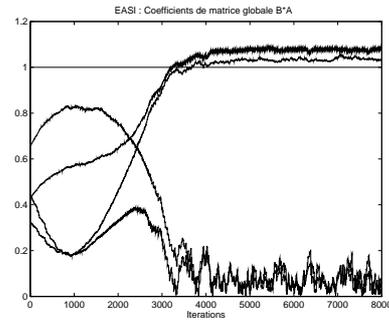


FIG. 5 — Evolution temporelle du module de  $\mathbf{BA}$  pour l'algorithme EASI modifié par (12).

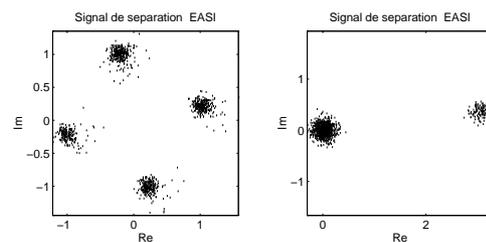


FIG. 6 — Constellation des sources estimées par l'algorithme EASI modifié par (12).

Les figures 7 et 8 montrent les résultats de séparation fournis par la méthode développée au paragraphe 3, qui s'affranchit de la connaissance des ddp des sources.

Cet exemple simple montre l'intérêt d'une estimation directe des fonctions non linéaires optimales pour le critère entropique (5). Ce nouvel algorithme permet de s'affranchir des contraintes statistiques sur les sources : circularité, signe du kurtosis et d'éviter un choix parfois difficile de fonctions non linéaires simples et robustes.

Les inconvénients résident dans une complexité accrue de l'algorithme, une convergence parfois lente des réseaux

de neurones qui estiment les  $\psi_i$  (à moins d'utiliser des méthodes du second ordre), une étude complexe de la stabilité et des performances de l'ensemble, qui contient 2 systèmes adaptatifs rebouclés.

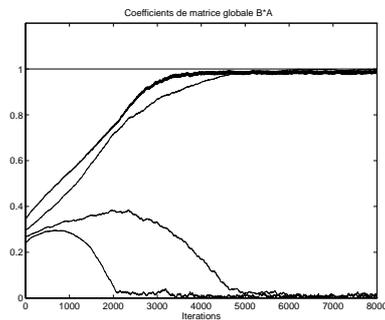


FIG. 7 — Evolution temporelle du module de **BA** pour l'algorithme proposé

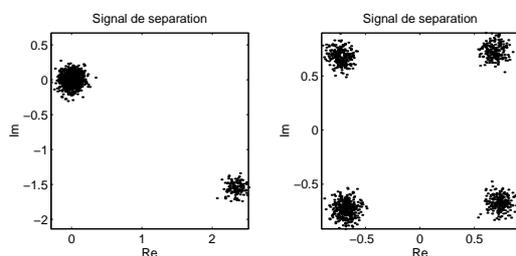


FIG. 8 — Constellation des sources estimées par l'algorithme proposé

## 5 Conclusion

La nature adaptative de l'algorithme EASI présente l'avantage d'être facile à implanter dans un processeur de signal, de traiter les données en ligne, d'effectuer la poursuite des modèles et d'assurer la séparation dans le contexte difficile de mélange presque singulier.

Cependant, l'algorithme EASI nécessite le choix de fonctions non-linéaires intervenant dans la mise à jour de la matrice séparante. La fonction optimale dépend des distributions des sources, bien sûr inconnues. Les performances en terme de convergence et de séparabilité peuvent être améliorées si l'on choisit de façon judicieuse les fonctions non linéaires qui interviennent dans la phase d'adaptation.

Pour cela, nous avons proposé une approche fondée sur la minimisation de l'information mutuelle. Les fonctions non-linéaires optimales intervenant dans l'algorithme de séparation sont estimées de manière adaptative. Les résultats expérimentaux montrent l'intérêt de cette approche et les performances que l'on peut en attendre.

## Références

- [1] C. Jutten, J. Héroult Blind separation of sources, Part I : An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, vol. 24, pp1-20, 1991.
- [2] J. F. Cardoso, B. Laheld. Equivariant adaptive source separation *IEEE trans. on Signal Processing*, 44(12), pp 3017-3030, 1996.
- [3] A. Cichocki, R. Unbehauen. *Neural networks for optimization and signal processing*, 2nd ed., John Wiley, New-York, 1994.
- [4] S. Amari, A. Cichocki, H. H. Yang. Recurrent neural networks for blind separation of sources *Proc. 1995 International Symposium on Nonlinear Theory and Applications*, pp. 37-42, 1995.
- [5] S. Amari. Natural gradient works efficiently in learning *Nural Computation*, To appear 1997.
- [6] P. O. Amblard, M. Gaeta, J. L. Lacoume. Statistics for complex variables and signals - Part I : Variables. *Signal Processing*, 53(1996), pp 1-13, 1996.
- [7] A. Bell, T. Sejnowski. An information-maximisation approach to blind separation and blind deconvolution *Neural Computation*, vol. 7, pp. 1129-1159.
- [8] D. T. Pham, Ph. Garat, C. Jutten. Separation of a mixture of independant sources through a maximum likelihood approach, *Proc. European Signal Processing Conf. EUSIPCO*, pp. 771-774, Brussels, 1992.
- [9] A. Taleb, C. Jutten. Séparation de sources dans des mélanges non linéaires. *Seizième Colloque GRETSI*, 1997.
- [10] A. Taleb, C. Jutten. Entropy Optimization : Application to blind source separation, Submitted To *ICANN*, 1997.
- [11] D. T. Pham. Séparation aveugle de sources via une analyse en composantes indépendantes. *Quinzième Colloque GRETSI*, pp. 289-292, 1995.
- [12] L. Wang, J. Karhunen, E. Oja, R. Vigarío. Blind separation of sources using nonlinear PCA type learning algorithms. *Proc. Int. Conf. on Neural Networks and Signal Processing*, China, pp. 847-850, 1995.