



## SÉPARATION ADAPTATIVE DE SOURCES EN AVEUGLE. IMPLANTATION COMPLEXE SANS CONTRAINTES.

Beate Christina Hvam Laheld, Jean-François Cardoso  
Télécom Paris, Dept. Signal, 46 rue Barrault, 75634 Paris Cedex 13  
email: laheld@sig.enst.fr, cardoso@sig.enst.fr

### Résumé

Nous présentons un nouvel algorithme adaptatif pour la séparation de sources, dérivant classiquement d'un critère à cumulants. L'approche classique est en deux étapes : blanchiment puis optimisation sous contrainte unitaire. Nous montrons comment l'intégration des deux étages en un seul fournit un algorithme sans paramétrisation, à la fois simple et efficace, valable dans le cas réel comme dans le cas complexe. Les simulations avec des signaux de communication sont très satisfaisantes.

### 1 Introduction

Le problème de *séparation de sources* considère  $m$  combinaisons linéaires, linéairement indépendantes, de  $n$  signaux statistiquement indépendants (les signaux 'sources'), et consiste à estimer les signaux sources à partir des seules observations, c'est-à-dire en l'absence de toute information sur les coefficients du mélange.

Des techniques de traitement par blocs utilisant des cumulants d'ordre 4 sont décrites dans [4, 7, 3] (voir [13] pour une analyse des performances). Dans cet article, nous considérons une approche *adaptative*. Les techniques auto-adaptatives pour la séparation des sources utilisent des non-linéarités du signal de sortie du séparateur dans la règle d'adaptation [6, 8, 11]. Lorsque ces techniques dérivent d'un critère du quatrième ordre, on trouve des non-linéarités cubiques [10, 2, 9, 5]. Toutefois, les critères à base de cumulants sont définis à partir des signaux blanchis et le critère sert alors à adapter une matrice unitaire qui fait suite à la matrice de blanchiment. Autrement dit, il faut adapter deux étages pour obtenir la séparation, le second étant contraint à l'unitarité. Le but de cette communication est de présenter un algorithme où les deux étages sont intégrés en un seul. On obtient un algorithme sans contrainte, de forme simple et qui s'applique indifféremment aux signaux complexes comme aux signaux réels.

### 2 Modèle et hypothèses

Nous considérons le modèle de données suivant :

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

où les sorties de  $m$  capteurs forment à l'instant  $t$  un vecteur  $\mathbf{x}(t)$  modélisé comme le mélange instantané par la matrice  $A$  (de taille  $m \times n$ ) de  $n$  signaux  $u_i(t)$ ,  $i = 1, \dots, n$  statistiquement indépendants appelés 'signaux sources' regroupés dans le vecteur  $\mathbf{u}(t)$ . À l'inverse des méthodes par blocs, les algorithmes adaptatifs de séparation de sources

### Abstract

A new adaptive blind algorithm for the separation of independent complex (or real) sources is presented. The algorithm stems from a classical cumulant-based separation criterion, but is not implemented by chaining a whitening stage and a unitary stage. We rather show that their integration into a single block leads to a parameter free implementation which is both simple and efficient and works indifferently for both real and complex signals.

ne peuvent pas traiter facilement le problème du bruit additif. Aussi notre discussion se limite-t-elle au cas où le bruit peut être négligé. Les hypothèses sont

- $n \leq m$  signaux sources mutuellement indépendants.
- Les distributions des sources sont non-gaussiennes et centrées. Dans le cas complexe, on suppose que  $E u_i(t)^2 = 0, \forall i = 1, n$ . Le kurtosis de la  $i$ -ème source est alors  $k_i = E |u_i(t)|^4 - 2E^2 |u_i(t)|^2$ . Dans le cas réel, il est  $k_i = E u_i(t)^4 - 3E^2 u_i(t)^2$ . Dans tous les cas, on suppose les kurtosis non nuls et de même signe.
- La matrice de mélange  $A$ , de taille  $m \times n$  est fixe et de rang  $n$ .
- Les puissances des sources sont intégrées dans la matrice de mélange de sorte que  $E(\mathbf{u}(t)\mathbf{u}(t)^*) = I$ .

Ici comme plus bas, le transconjugé d'un vecteur  $\mathbf{x}$  est noté  $\mathbf{x}^*$  et  $I$  est la matrice identité.

La séparation de sources est obtenue par toute matrice  $B$  de taille  $n \times m$  qui, agissant sur  $\mathbf{x}$ , fournit un vecteur  $\mathbf{z} = B\mathbf{x}$ , dont les composantes sont les signaux sources, éventuellement permutés et déphasés (ces indéterminations sont inhérentes sous nos hypothèses).

### 3 Les algorithmes.

Les deux algorithmes que nous présentons ci-dessous se basent sur le fait qu'une matrice de séparation  $B$  pour des sources de kurtosis négatifs (resp. positifs) peut se factoriser en  $B = UW$  où  $W$  est une matrice  $n \times m$  de décorrélation spatiale (blanchiment) et  $U$  est une matrice unitaire telle que la somme des kurtosis de ses sorties  $\mathbf{z} = B\mathbf{x}$  soit minimale (resp. maximale).

Une implantation de cette idée conduit à faire suivre un blanchisseur adaptatif (adapté par le second-ordre) d'une matrice unitaire (adaptée par le quatrième ordre). Une telle approche maintient donc deux étages distincts. Nous avons



proposé dans [2] une implantation originale pour réaliser la contrainte unitaire sans paramétrisation explicite. Toutefois cette implantation ne conservait qu'approximativement l'unitarité et de plus, l'algorithme global présentait des attracteurs indésirables. Nous proposons ici une approche beaucoup plus satisfaisante.

### 3.1 Algorithme PFS

Le but de cet article est de montrer comment les deux étages peuvent être combinés en une seule matrice de séparation dont l'adaptation se fait par une règle unique, *faisant intervenir uniquement les signaux de sorties*, conformément au schéma (1).

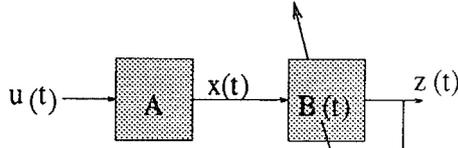


Figure 1: Un seul étage adapté uniquement sur sa sortie

La minimisation de la somme des kurtosis des sorties sous contrainte de blancheur amène après diverses simplifications (voir appendice), à l'algorithme **PFS** (Parameter Free Separation) suivant d'adaptation d'une matrice  $B$  de taille  $n \times m$ :

#### Algorithme PFS

$$e_i = |z_i(t)|^2 z_i(t) \quad \forall i = 1, n \quad (2)$$

$$H = I - z(t)z(t)^* - b(ez(t)^* - z(t)e^*) \quad (3)$$

$$B(t) = (I + \lambda H) B(t-1) \quad (4)$$

où  $I$  désigne la matrice identité,  $\lambda$  et  $b$  sont deux constantes réelles positives (on doit changer le signe de  $b$  pour traiter des signaux de kurtosis positifs). La première ligne (2) définit le vecteur  $e$  comme étant une distortion cubique, composante par composante de la sortie  $z$ . Cette distortion cubique a pour origine le critère sous-jacent construit sur les cumulants d'ordre 4. La seconde ligne (3) construit une matrice  $H$  de taille  $n \times n$ . Il est facile de voir que cette matrice est nulle en moyenne si  $B$  est maintenue constante à une valeur telle que  $BA = I$ . En effet, le terme  $I - zz^*$  est de moyenne nulle lorsque la sortie  $z$  est blanche. D'autre part, la moyenne de  $ze^* - ez^*$  est aussi nulle lorsque les composantes de  $z$  sont indépendantes. Le scalaire  $b$  permet de régler le poids relatif du terme de blanchiment et du terme d'ordre supérieur. Finalement, la dernière ligne (4) donne la mise à jour de la matrice de séparation.

Il est à noter que la mise à jour se fait par multiplication à gauche (et donc par une action à droite dans le schéma de la figure 1) du séparateur  $B$  par une matrice  $H$  ne dépendant que des *sorties* du séparateur. Autrement dit, tout se passe du côté des sorties : l'algorithme 'ne voit pas' l'entrée  $x$  de telle sorte que son comportement est, à bruit nul, indépendant de la matrice de mélange  $A$ .

Le code Matlab de la mise à jour **PFS** peut s'écrire

```
zz = z*z';
ez = ((diag(zz)).*z)*z';
H = id - zz - b*(ez-ez');
B = (id + lambda*H) * B;
```

et souligne la simplicité de l'implantation.

### 3.2 Algorithme SPFS

La simplicité de l'algorithme PFS le laisse toutefois vulnérable à une initialisation trop forte du séparateur  $B$  ou à la présence d'outliers lorsque  $\lambda$  n'est pas 'petit'. Ceci est la simple conséquence du fait que la matrice  $H$  dépend d'une puissance quatrième des sorties qui favorise l'instabilité. Nous utilisons une version stabilisée de l'algorithme, dite **SPFS** (Stabilized PFS), pour y remédier. Parmi les différentes techniques possibles de stabilisation, nous avons considéré la suivante :

#### Algorithme SPFS

$$e_i = |z_i(t)|^2 z_i(t) \quad \forall i = 1, n \quad (5)$$

$$H = \frac{I - z(t)z(t)^*}{\|z(t)\|^2} - b \frac{z(t)e^* - ez(t)^*}{\|z(t)\|^4} \quad (6)$$

$$B(t) = (I + \lambda H) B(t-1) \quad (7)$$

dont nous illustrons les performances dans la section consacrée aux simulations. L'algorithme SPFS ne diffère du PFS que par la présence de facteurs stabilisants dans  $H(t)$ . Cette stabilisation nous a donné satisfaction mais rend plus difficile l'analyse des performances.

### 3.3 Signaux réels

Nous avons traité jusqu'ici le cas général de signaux complexes car nous pensons que la séparation de sources trouve un domaine d'application privilégié dans le traitement d'antenne à bande étroite où le modèle de mélange instantané se vérifie effectivement, *lorsque l'on considère des signaux complexes*. Soulignons que notre approche s'applique sans modification au cas de signaux réels où l'opération de transposition prend simplement la place de la transconjugaison. De façon frappante, le code Matlab donné ci-dessus s'applique indifféremment au traitement réel comme au traitement complexe.

## 4 Points stationnaires de PFS

La méthode de l'ODE consiste à ramener l'étude d'un algorithme adaptatif stochastique à celle d'une équation différentielle déterministe du premier ordre [1]. L'ODE de l'algorithme PFS s'écrit :

$$\frac{\partial B}{\partial t} = -E_B(HB) \quad (8)$$

où la notation  $E_B$  indique que l'espérance est prise à  $B$  fixée. Si l'on exclut les matrices  $B$  singulières, il apparaît donc que les points stationnaires de l'ODE sont ceux pour lesquels  $E_B(H) = 0$ . La nullité de  $E_B(H)$  implique simultanément celle des parties hermitiennes et anti-hermitiennes. Les points stationnaires non singuliers vérifient donc :

$$E_B(I - zz^*) = 0 \quad (9)$$

$$E_B(ez^* - ze^*) = 0 \quad (10)$$



La première condition implique que les sorties sont blanches et donc que tout point stationnaire non singulier vérifie  $BA = U$  où  $U$  est une matrice unitaire.

La seconde condition, lorsqu'on y reporte la première, donne une équation matricielle du quatrième ordre en une matrice unitaire. La forme en est simple mais la solution générale, qui permettrait de caractériser l'ensemble des points stationnaires, nous échappe. Dans le cas à deux sources on peut montrer que les seuls points stationnaires stables sont des matrices de séparation et que ces points stables. Les points stationnaires instables correspondent au mélange maximum, c'est à dire que les coefficients de  $BA$  ont tous le même module.

## 5 Simulations

Nous illustrons différents points par les figures suivantes qui montrent l'évolution de la norme des coefficients de la matrice globale  $B(t)A$  en fonction du nombre d'itérations. La convergence vers une solution correcte se vérifie lorsque  $n$  trajectoires tendent vers 1 tandis que les autres tendent vers 0. Les figures indiquent le type d'algorithme utilisé, la valeur des paramètres  $\lambda$  et  $b$ , le nombre de sources et de capteurs, la distribution des sources (qui sont par ailleurs i.i.d.) et le conditionnement de la matrice de mélange. Les simulations montrent un bon comportement de l'algorithme pour des signaux complexes ou réels et semblent indiquer l'absence de minima locaux parasites. Sauf mention du contraire, le nombre de sources est supposé connu.

- La figure 2 montre une convergence rapide dans un cas favorable : sources à module constant, pas de bruit, bon conditionnement du mélange, nombre de sources supposé connu.

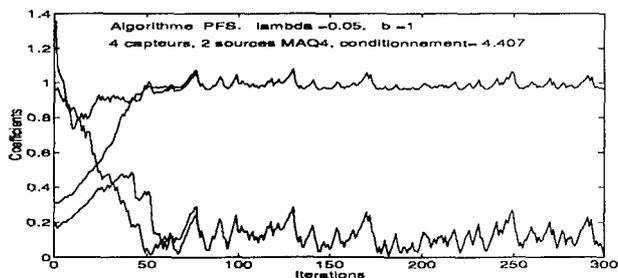


Figure 2: Un cas favorable, convergence rapide.

- La figure 3 montre une dégradation du comportement lorsque les sources sont modulées en MAQ16. On sait (??) que dans les techniques par blocs, les signaux de module constant sont particulièrement favorables.

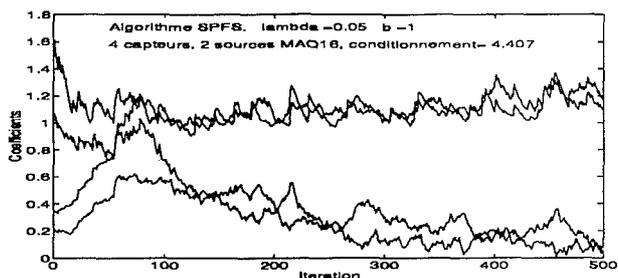


Figure 3: Variance accrue en MAQ16

- La présence d'un bruit gaussien, blanc temporellement et spatialement (i.e. de covariance  $\sigma I$ ), ajouté à  $x(t)$  est

toléré à faible niveau (figure 4). Nous avons ici un SNR de -20 dB, défini comme le rapport  $\|A\|^2/\sigma$  (en norme de Frobenius). Nous observons que la variance et le temps

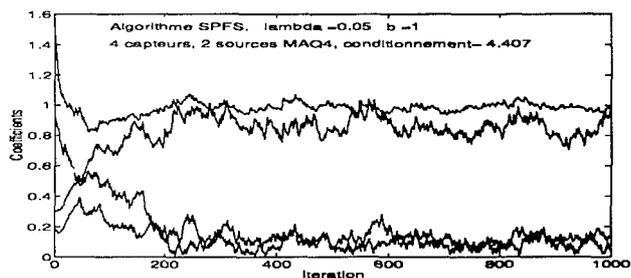


Figure 4: Tolérance à un bruit faible. RSB=20 dB.

de convergence augmentent, mais la séparation reste satisfaisante. Il y a peu d'espoir de tolérer des bruits significativement plus forts tout en conservant la simplicité de l'algorithme et une convergence rapide.

- Nous illustrons une propriété frappante de l'algorithme : son insensibilité après convergence à la sévérité du mélange. Dans la figure 5, le conditionnement de  $A$  est en effet supérieur à 3000. Cette robustesse est due au fait que le séparateur s'adapte seulement sur sa sortie. Il est malheureusement clair que cette robustesse sera détruite par un bruit additif sur  $x$  et ce d'autant plus vite que le conditionnement est défavorable.

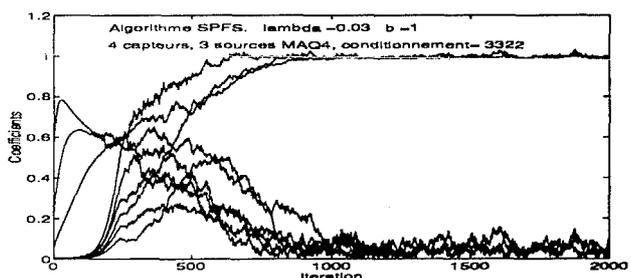


Figure 5: Mélange sévère.

- Lorsque le nombre de sources est surestimé, il est difficile de prédire le comportement de l'algorithme. La figure 6 montre un des comportements observés. Ici  $m = 3$  et  $n = 2$ , mais on adapte une matrice de taille  $3 \times 3$ . Chaque sous-graphe concerne une sortie et montre la norme des coefficients de chaque source participant à cette sortie. La troisième ne contient rapidement ( $\sim 100$  échantillons) plus qu'un seul signal, tandis que le second se partage plus longtemps sur les deux autres sorties avant de finalement ( $\sim 800$  échantillons) choisir la seconde. De façon apparemment paradoxale, le faible bruit présent (RSB=20dB) contribue à la stabilité de la solution (simulation non incluse). Ceci se comprend si l'on songe que, quand  $n$  est surestimé, l'algorithme tente de fabriquer un signal inexistant.

## 6 Conclusion

Nous avons introduit un nouveau type d'algorithme adaptatif pour la séparation de sources complexes ou réelles, dérivant d'algorithmes de minimisation sous contrainte unitaire après blanchiment. Toutefois, en intégrant l'étage de blanchiment et l'étage unitaire, nous aboutissons à un séparateur global dont l'adaptation se fait sans paramétrisation ou contraintes.

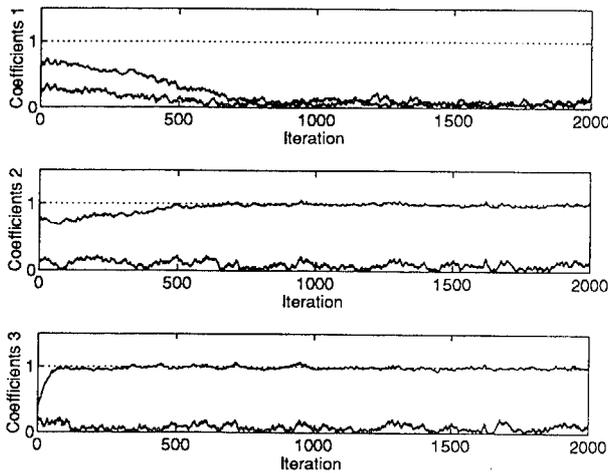


Figure 6: Surestimation du nombre de sources

L'algorithme le plus simple (FPS) se prête à une étude théorique : son ODE associée peut être facilement établie. Cet algorithme doit toutefois être stabilisé : nous suggérons la version SPFS qui est ici illustrée de plusieurs simulations. Celles-ci montrent une convergence très satisfaisante et indiquent que l'algorithme supporte dans une certaine mesure la surestimation du nombre de sources. Par ailleurs, la robustesse de l'algorithme par rapport au conditionnement de  $A$  est excellente à bruit nul. L'analyse des performances asymptotiques de FPS sera publiée prochainement. Nous explorons aussi la possibilité de remplacer la distortion cubique (2) par d'autres non-linéarités. La prise en compte explicite d'un bruit additif est plus délicate dans un traitement adaptatif que dans une approche bloc : des solutions préservant la simplicité d'implantation restent à trouver.

## Appendice

Nous expliquons brièvement comment l'algorithme PFS dérive d'une minimisation d'un critère à cumulants optimisé sous contrainte unitaire après blanchiment. Nous partons de la structure classique en deux étages.

**Blanchiment.** Si l'on adapte une matrice  $W$  de taille  $n \times m$  par la loi  $W(t) = (I + \lambda(I - \mathbf{y}(t)\mathbf{y}(t)^*)) W(t-1)$  où  $\mathbf{y}(t) \stackrel{\text{def}}{=} W(t)\mathbf{y}(t)$ , on obtient un algorithme stochastique dont le champ moyen s'annule clairement lorsque  $E\mathbf{y}(t)\mathbf{y}(t)^* = I$  c'est-à-dire lorsque la sortie de  $W$  est blanche. Sous réserve de stabilité, on dispose donc d'un blanchisseur adaptatif.

**Partie unitaire.** Soit  $U$  une matrice unitaire agissant sur  $\mathbf{y}(t)$ , une version blanchie de  $\mathbf{x}(t)$ . On note  $\mathbf{z}(t) = U\mathbf{y}(t)$  et on montre facilement que la somme des kurtosis de  $\mathbf{z}(t)$  est minimale (resp. maximale) lorsque  $U$  est une matrice de séparation si tous les signaux sources sont de kurtosis négatifs (resp. positifs). Le gradient stochastique de la somme des kurtosis des sorties par rapport à la matrice  $U$  conduit à la règle d'adaptation  $U(t) = (I + \beta G)U(t-1)$  où  $\beta$  est un petit scalaire et  $G = \mathbf{z}(t)\mathbf{e}(t)^* - \mathbf{e}(t)\mathbf{z}(t)^*$  avec  $\mathbf{e}$  défini comme dans (2) [2]. Un gradient sous contrainte unitaire est défini indépendamment de toute paramétrisation unitaire explicite, mais la règle d'adaptation ci-dessus ne préserve

l'unitarité de  $U$  qu'au second ordre près en  $\beta$ . Ce problème se résout en combinant blanchisseur et partie unitaire.

**Une unique règle d'adaptation.** Les mises à jour de  $W$  et de  $U$  indiquées plus haut sont toutes deux multiplicatives à gauche. En faisant le produit de ces facteurs, chacun voisin de l'identité (pour des pas faibles), on obtient un nouveau facteur multiplicatif voisin de l'identité. En n'en gardant que le premier ordre en  $\lambda$  et  $\beta$ , on obtient l'algorithme PFS. Ceci résout le problème de l'écart à l'unitarité mentionné plus haut, puisque ces écarts sont ramenés à zéro par le facteur de blanchiment. Ces explications ne prétendent pas à la rigueur : elles expliquent la construction de PFS dont les propriétés peuvent s'étudier indépendamment de son origine.

## References

- [1] A. Benveniste, M. Métivier, and P. Priouret. Algorithmes adaptatifs et approximations stochastiques. In *Masson*, 1987.
- [2] J.-F. Cardoso and B. Laheld. Adaptive blind source separation for channel spatial equalization. In *Proc. of COST 229 workshop on adaptive signal processing*, pages 19–26, 1992.
- [3] J.-F. Cardoso and A. Souloumiac. An efficient technique for blind separation of complex sources. In *Proc. IEEE SP Workshop on Higher-Order Stat., Lake Tahoe, USA*, pages 275–279, 1993.
- [4] P. Comon. Independent component analysis. In *Proc. Int. Workshop on Higher-Order Stat., Chamrousse, France*, pages 111–120, 1991.
- [5] N. Delfosse and P. Loubaton. Séparation adaptative de sources indépendantes par une approche de déflation. In *Proc. GRETSI*, 1993.
- [6] L. Féty. Méthodes de traitement d'antenne adaptées aux radio-communications. *Thèse de doctorat. Télécom Paris*, June 1988.
- [7] M. Gaeta and J.-L. Lacoume. Source separation without a priori knowledge: the maximum likelihood solution. In *Proc. EUSIPCO*, pages 621–624, 1990.
- [8] C. Jutten and J. Héroult. Blind separation of sources: an adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24:1–10, 1991.
- [9] P. Loubaton and P. A. Regalia. Blind deconvolution of multivariate signals: a deflation approach. In *Proc. ICC*, page To appear, 1993.
- [10] E. Moreau and O. Macchi. New self-adaptive algorithms for source separation based on contrast functions. In *Proc. IEEE SP Workshop on Higher-Order Stat., Lake Tahoe, USA*, pages 215–219, 1993.
- [11] D. T. Pham, P. Garrat, and C. Jutten. Separation of a mixture of independent sources through a maximum likelihood approach. In *Proc. EUSIPCO*, pages 771–774, 1992.
- [12] V. C. Soon, L. Tong, Y. F. Huang, and R. Liu. An extended fourth order blind identification algorithm in spatially correlated noise. In *Proc. ICASSP*, pages 1365–1368, 1990.
- [13] A. Souloumiac and J.-F. Cardoso. Performances en séparation de sources. In *Proc. GRETSI, Juan les Pins, France*, 1993.