

Critère de séparation de sources basé sur les statistiques d'ordre

Jani EVEN, Eric MOISAN

Laboratoire des Images et des Signaux, groupe Non-Linéaire (UMR 5083)
ENSIEG, Domaine universitaire, BP 46, 38405 Saint Martin d'Hères Cedex, France
jani.even@lis.inpg.fr, eric.moisan@lis.inpg.fr

Résumé – Cette article présente l'utilisation des statistiques d'ordre pour la séparation aveugle de mélanges instantanés de sources réelles. Un critère basé sur ces statistiques d'ordre est introduit, puis une méthode itérative de séparation est exposée. Enfin, les avantages du critère sont soulignés par des résultats de simulation.

Abstract – This paper shows the possibility to blindly separate instantaneous mixtures of real signals by means of an iterative method based on order statistics. First a new criterion using order statistics is explained, and then significant results underline the advantages of this new method.

1 Présentation du problème

En séparation aveugle de sources, il faut estimer des signaux (les sources) en observant uniquement des mélanges de ceux-ci (les observations). Un premier groupe de méthodes suppose l'indépendance statistique des sources et exploite alors les statistiques d'ordre supérieur à deux de ces signaux pour les séparer (à condition qu'ils ne suivent pas des lois gaussiennes) [6]. Un second ensemble de méthodes se restreint à l'ordre deux mais suppose que les sources présentent des autocorrélations distinctes [1]. Les statistiques d'ordre, construites sur des observations de courte durée, permettent de prendre en compte de manière conjointe les propriétés liées à la corrélation temporelle des sources et celles liées aux lois de ces sources [7]. Un critère basé sur ces statistiques particulières s'avère performant pour la séparation de sources. Nous présentons ici une version adaptée au cas des mélanges instantanés sans bruit additif qui donne effectivement de bons résultats. On suppose que l'on observe N mélanges linéaires instantanés de N sources. Les notations sont les suivantes :

$$\begin{bmatrix} s_1(k) \\ \vdots \\ s_N(k) \end{bmatrix} \Rightarrow \boxed{A} \Rightarrow \begin{bmatrix} v_1(k) \\ \vdots \\ v_N(k) \end{bmatrix} \Rightarrow \boxed{M} \Rightarrow \begin{bmatrix} \hat{s}_1(k) \\ \vdots \\ \hat{s}_N(k) \end{bmatrix}$$

Les sources $\{s_i\}_{i \in [1, N]}$ sont mélangées par la matrice constante A de taille $N \times N$ et les observations $\{v_i\}_{i \in [1, N]}$ sont traitées par la matrice de séparation M de taille $N \times N$ afin d'obtenir des estimées $\{\hat{s}_i\}_{i \in [1, N]}$ des sources. La méthode de séparation que nous proposons ici repose sur l'adaptation de la matrice M , de façon à minimiser un critère \mathcal{J}_1 appliqué aux sources estimées \hat{s}_i . Ce critère s'appuie sur les statistiques d'ordre construites sur chaque source estimée.

2 Le critère

Les statistiques d'ordre pour un signal x sont obtenues à l'aide de fenêtres glissantes de taille T ($T \ll$ nombre de points du signal) qui parcourent le signal. L'ordonnement des valeurs dans ces fenêtres permet d'avoir des

réalisations des statistiques d'ordre. A l'instant t on dispose des T dernières valeurs du signal : $[x_t, \dots, x_{t-T+1}]$. On ordonne ces valeurs pour avoir une réalisation des T statistiques d'ordre : $x_{(1)}^{(t)} \leq x_{(2)}^{(t)} \leq \dots \leq x_{(T)}^{(t)}$. On s'intéresse enfin à la quantité $\delta_x^{(t)} = x_{(T)}^{(t)} - x_{(1)}^{(t)}$, dont on approche l'espérance mathématique par la moyenne empirique des réalisations (éq. 1).

$$\mathcal{J}_1(x) = \mathcal{E}\{\delta_x\} \approx \frac{1}{K} \sum_{t=1}^K \delta_x^{(t)} \quad (1)$$

où K est le nombre de fenêtres de taille T recouvrant l'ensemble des mesures de x .

L'expérience montre que, sous la contrainte d'une puissance constante, la quantité précédente, appliquée à une combinaison linéaire de N signaux indépendants, présente N minima locaux. Chaque minimum correspond au cas où la combinaison linéaire coïncide en fait avec un seul des signaux indépendants.

En particulier, il existe un minimum global correspondant à l'une des N sources (éq. 2).

$$\begin{aligned} \exists j \in [1, N] \text{ tel que } \forall [\alpha_1 \dots \alpha_N] \in \mathbb{R}^N \\ \setminus \mathcal{E}\left\{\sum_i \alpha_i^2 s_i^2\right\} = 1 \\ \mathcal{J}_1([\alpha_1 \dots \alpha_N] \times \begin{bmatrix} s_1 \\ \vdots \\ s_N \end{bmatrix}) \geq \mathcal{J}_1(s_j) \end{aligned} \quad (2)$$

Par exemple dans le cas simple d'un mélange instantané de 2 sources s_1 et s_2 sans bruit additif, une estimée \hat{s} est une combinaison linéaire des observations v_1 et v_2 (éq. 3). La puissance unitaire de l'estimée est assurée grâce au

paramètre λ .

$$\begin{aligned}\hat{s} &= \lambda \begin{bmatrix} 1 & b_{12} \end{bmatrix} \times \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \\ &= \lambda \begin{bmatrix} 1 & b_{12} \end{bmatrix} \times \begin{bmatrix} 1 & a_{12} \\ a_{21} & 1 \end{bmatrix} \times \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \quad (3)\end{aligned}$$

Si le paramètre b_{12} prend les valeurs $-a_{12}$ ou $-1/a_{21}$, on estime respectivement s_1 ou s_2 . Sur la courbe (1), on a représenté les valeurs du critère en fonction de b_{12} pour deux sources uniformes blanches, avec des paramètres de mélange $a_{12} = -0.3$ et $a_{21} = 0.6$. Le critère présente bien deux minima locaux en 0.3 et $-1/0.6$. La détection de ces minima permet de séparer le mélange. La preuve théorique

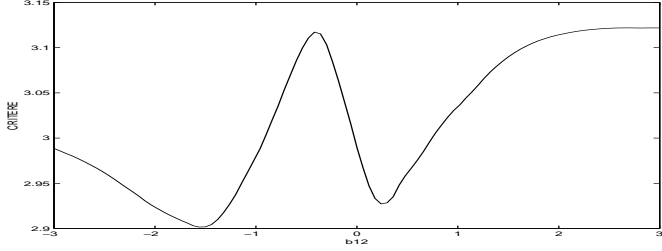


FIG. 1 – Critère en fonction de b_{12} pour le mélange instantané de deux bruits blancs uniformes avec $a_{12} = -0.3$ et $a_{21} = 0.6$.

que ce critère constitue bien un contraste [6] n'est pas encore établie, mais toutes les simulations effectuées ont montré son efficacité.

Dans le cas d'un mélange de N sources indépendantes, les N minima accessibles sur le critère ne sont pas équivalents. Comment obtenir alors les N combinaisons linéaires correspondantes afin d'estimer les N sources ?

3 Cas particulier de deux sources

Pour l'exploitation du critère \mathcal{J}_1 dans le cas d'un mélange instantané de deux sources réelles, la matrice de séparation M ne dépend que d'un seul paramètre grâce aux contraintes de puissance unitaire et d'orthogonalité des estimées.

Le déterminant de la matrice de mélange A , noté Δ , est fonction des moments portant sur les mesures v_1 et v_2 . En notant $P_1 = \mathcal{E}\{v_1^2\}$, $P_2 = \mathcal{E}\{v_2^2\}$ et $P_{12} = \mathcal{E}\{v_1 v_2\}$ les puissances des mesures on a :

$$\Delta^2 = P_1 P_2 - P_{12}^2$$

A est donc inversible si $\mathcal{E}\{v_1^2\}\mathcal{E}\{v_2^2\} \neq \mathcal{E}^2\{v_1 v_2\}$ c'est à dire si v_1 et v_2 ne sont pas colinéaires.

On va chercher la matrice M sous la forme :

$$M = \Delta^{-1} \begin{bmatrix} c & 0 \\ 0 & f \end{bmatrix} \begin{bmatrix} 1 & d \\ e & 1 \end{bmatrix}$$

L'orthogonalité des sources estimées impose (voir [8]) :

$$\mathcal{E}\{\hat{s}_1 \hat{s}_2\} = 0 \Rightarrow d(e) = -\frac{P_{12} + eP_1}{P_2 + eP_{12}}$$

La contrainte de puissance unitaire des sources estimées s'écrit :

$$\mathcal{E}\{\hat{s}_1^2\} = 1 \Rightarrow f(e) = \frac{\Delta}{\sqrt{P_2 + e(2P_{12} + eP_1)}}$$

$$\mathcal{E}\{\hat{s}_2^2\} = 1 \Rightarrow c(e) = \frac{\Delta}{\sqrt{P_1 + d(e)(2P_{12} + d(e)P_2)}}$$

Tous les coefficients de la matrice M sont ainsi fonction du seul paramètre e .

Pour estimer conjointement les deux sources, sous la contrainte qu'elles soient orthogonales, on utilise un critère portant sur les deux estimées. Ce critère est la somme des espérances des accroissements des estimées.

$$\mathcal{J}_2 = \mathcal{E}\{\delta_{\hat{s}_1}\} + \mathcal{E}\{\delta_{\hat{s}_2}\}$$

Comme la matrice de séparation ne dépend que du paramètre e , le critère \mathcal{J}_2 est lui aussi une fonction de e seulement.

Pour séparer le mélange, il faut déterminer le minimum du critère par rapport à e . La recherche de ce minimum est effectuée à l'aide d'un algorithme du gradient du critère par rapport au paramètre, noté : $\nabla_e \mathcal{J}_2$ (Comme son expression littérale est simple à obtenir mais un peu lourde elle n'est pas donnée ici, voir [8]). Une première implémentation a été faite hors ligne, c'est à dire que l'on estime le gradient du critère en un point donné en utilisant toutes les données disponibles.

Les grandes lignes d'un algorithme itératif de séparation sont données ci dessous :

- $k = 0$: Initialisation du séparateur e_0 , choix de la taille T des fenêtres et du pas $\mu > 0$.
- Pour $k \geq 0$: Répéter jusqu'à convergence :
 1. Application de $M(e_k)$ aux observations V pour avoir les estimées \hat{S}_k
 2. Calcul du gradient $\nabla_e \mathcal{J}_2(e_k)$ sur les estimées \hat{S}_k
 3. Evolution du paramètre e : $e_{k+1} = e_k - \mu \nabla_e \mathcal{J}_2(e_k)$

Après convergence, la matrice $M(e_{final})$ permet de retrouver les deux sources à partir des observations.

Une version adaptative de cet algorithme existe : Le gradient est remplacé par un gradient stochastique et l'adaptation de e est effectuée à chaque fenêtre temporelle.

4 Plus de deux sources

4.1 Méthode de déflation

Quand on a plus de deux sources, l'expression des contraintes d'orthogonalité des sources deux à deux n'est plus exploitable facilement.

Dans une première approche, nous avons procédé par déflation. On estime une seule source à l'aide du critère \mathcal{J}_1 portant sur une combinaison linéaire de puissance unitaire des N observations (éq. 4).

$$\begin{aligned}\hat{s} &= \lambda \begin{bmatrix} m_1 & \dots & m_N \end{bmatrix} \times \begin{bmatrix} v_1 \\ \vdots \\ v_N \end{bmatrix} \\ &= \lambda \begin{bmatrix} \alpha_1 & \dots & \alpha_N \end{bmatrix} \times \begin{bmatrix} s_1 \\ \vdots \\ s_N \end{bmatrix} \quad (4)\end{aligned}$$

Un algorithme du gradient permet de trouver le minimum le plus important de $\mathcal{J}_1(\hat{s})$ et d'estimer la source qui lui correspond. On peut alors modifier les observations pour

leur retrancher les contributions de cette source. Ensuite on applique l'algorithme sur les observations modifiées et on itère la méthode jusqu'à obtention de la dernière source. Cette méthode est assez lourde et donne des résultats assez médiocres quant à l'estimation des dernières sources. Il est préférable d'estimer conjointement toutes les sources.

4.2 Orthogonalisation conjointe

Finalement la solution retenue pour séparer des mélanges de taille supérieure à deux est l'utilisation d'un blanchiment spatial conjointement à la séparation par statistiques d'ordre. Dans ce cas le critère prend en compte autant d'estimées qu'il y a de sources, car chacune des estimées doit approcher une des sources.

Le critère utilisé pour séparer les sources est la somme des espérances des accroissements des estimées (éq. 5).

$$\mathcal{J}_N = \sum_{i=1}^N \mathcal{J}_1(\hat{s}_i) = \sum_{i=1}^N \mathcal{E}\{\delta_{\hat{s}_i}\} \quad (5)$$

Une première approche consiste à blanchir spatialement les observations puis à appliquer l'algorithme de séparation. La première version de cet algorithme cherche la matrice de séparation sous la forme d'un produit de matrices de rotation de Givens afin de conserver l'orthogonalité des estimées. On procède à la séparation en adaptant successivement les matrices de Givens, comme dans [6]. Cette approche impose une structure pour la matrice de séparation, ce qui fait que l'algorithme n'est pas équivariant [4]. C'est pour cela que nous lui avons préféré une méthode qui ne paramétrise pas la matrice de séparation.

On rajoute un terme \mathcal{J}_\perp au critère \mathcal{J}_N pour garantir l'orthogonalité des sources estimées conjointement à la séparation (éq. 6).

$$\mathcal{J} = \sum_{k=1}^N \mathcal{E}\{\delta_{\hat{s}_k}\} + \frac{1}{2}[\text{trace}(R_{\hat{S}}) - \log(\det(R_{\hat{S}})) - N] \quad (6)$$

En fait \mathcal{J}_\perp mesure la différence entre la covariance des estimées ($R_{\hat{S}} = M R_V M^T$) et la matrice identité [4].

Pour trouver le minimum de ce critère on utilise le gradient relatif par rapport à la matrice de séparation M . Afin d'assurer l'orthogonalité des estimées, il est nécessaire de projeter le gradient relatif de \mathcal{J}_N sur l'espace des matrices antisymétriques (éq.7).

$$\nabla \mathcal{J}_N(M)_{antisym} = \frac{\partial \mathcal{J}_N}{\partial M} M^T - M \frac{\partial \mathcal{J}_N}{\partial M}^T \quad (7)$$

Car pour une matrice orthogonale U , la relation $U U^T = I$ est modifiée pour une variation relative en $U + \epsilon U$ et devient :

$$(U + \epsilon U)(U + \epsilon U)^T = I + \epsilon + \epsilon^T + \epsilon \epsilon^T$$

L'orthogonalité est assurée au premier ordre en ϵ :

$$(U + \epsilon U)(U + \epsilon U)^T = I + o(\epsilon)$$

si $\epsilon^T = -\epsilon$, c'est à dire si ϵ est antisymétrique [4] [3].

Par contre le gradient relatif de \mathcal{J}_\perp (son calcul est détaillé dans [4]) évolue sur l'espace des matrices symétriques (éq.8).

$$\nabla \mathcal{J}_\perp(M) = R_{\hat{S}} - I_N \quad (8)$$

Le gradient relatif du critère \mathcal{J} par rapport à M (éq.9) permet d'obtenir une équation d'évolution de M (éq. 10) garantissant la séparation du mélange.

$$\nabla \mathcal{J}(M) = R_{\hat{S}} - I_N + \frac{\partial \mathcal{J}_N}{\partial M} M^T - M \frac{\partial \mathcal{J}_N}{\partial M}^T \quad (9)$$

$$M_{k+1} = [I_N - \mu \nabla \mathcal{J}(M_k)] \times M_k \quad (10)$$

Les grandes lignes d'un algorithme itératif de séparation sont :

- $k = 0$: Initialisation du séparateur à M_0 , choix de la taille T des fenêtres et du pas $\mu > 0$.
- $k \geq 0$: Répéter jusqu'à convergence :
 1. Application de M_k aux observations V pour avoir les estimées \hat{S}_k
 2. Calcul du gradient $\nabla \mathcal{J}(M_k)$ sur les estimées \hat{S}_k
 3. Evolution de M : $M_{k+1} = [I_N - \mu \nabla \mathcal{J}(M_k)] \times M_k$

La matrice M_{finale} permet de séparer les sources. Le produit $M_{finale} \times A$ est une matrice de permutation (un seul coefficient non nul égal à ± 1 par ligne et par colonne).

Dans la version adaptative de cet algorithme, le gradient du critère est remplacé par un gradient stochastique et M évolue à chaque fenêtre d'ordonnancement.

5 Simulations

5.1 Paramètres des simulations

Les algorithmes développés ont été testés en simulant des mélanges instantanés.

Les sources sont des bruits de lois et de corrélations diverses.

Les résultats proposés dans cette partie sont ceux de l'algorithme itératif du 4.2. Les deux principaux paramètres de la méthode sont μ le pas d'adaptation et T l'horizon utilisé pour les statistiques d'ordre. La matrice de séparation est initialisée à $M_0 = \Lambda \times I_N$ où Λ est une matrice diagonale de normalisation ($\lambda_{ii} = \frac{1}{\sqrt{\mathcal{E}\{v_i^2\}}}$).

Comme pour tout algorithme itératif, le pas doit être choisi assez faible pour ne pas faire diverger l'algorithme et il y a un compromis à réaliser entre la vitesse de convergence et la variance de l'estimée après convergence.

Par contre, le choix de la taille T des fenêtres utilisées pour le calcul des statistiques d'ordre est lié aux corrélations des sources. Effectivement pour de très grandes valeurs de T les valeurs des signaux en début et en fin de fenêtre sont décorréelées, tandis que pour de faibles valeurs, tous les échantillons sont corrélés. En fait, l'expérience montre que le choix de la taille T est surtout primordial quand l'information nécessaire à la séparation est contenue dans les corrélations des sources.

Le problème est que l'on se place dans le cadre de la séparation aveugle de sources, donc on ne dispose pas d'information *a priori* sur les corrélations des sources.

De plus la justification théorique permettant de motiver un choix optimal de T reste à établir et ce sont les simulations qui ont permis d'étudier l'influence de T .

5.2 Exemple de simulation

Les trois sources, statistiquement indépendantes, sont des signaux centrés de 300 points.

- Source s_1 : Uniforme blanche
- Source s_2 : Gaussienne blanche
- Source s_3 : Gaussienne filtrée AR1 : $H(z) = \frac{0.1}{1-0.9z^{-1}}$

Matrice de mélange :

$$A = \begin{bmatrix} 0.323 & 0.226 & 0.06 \\ 0.279 & -0.222 & -0.515 \\ 0.604 & -0.151 & -0.262 \end{bmatrix}$$

Méthode à l'ordre deux : SOBI [2]

$$M \times A = \begin{bmatrix} 0.75 & 0.663 & 0.027 \\ 0.011 & 0.002 & -1.002 \\ 0.664 & -0.752 & -0.044 \end{bmatrix}$$

Chaque ligne de la matrice du système complet (Matrice de séparation \times matrice de mélange) correspond à une des estimées. Alors pour une bonne séparation, cette matrice doit être une matrice de permutation (un seul coefficient non nul égal à ± 1 par ligne et par colonne). On peut mesurer la qualité de la séparation en calculant pour chaque estimée le rapport entre la somme des puissances parasites (les résidus des autres sources) et la puissance de l'estimée. Les sources étant normées, les puissances résiduelles peuvent être directement calculées à partir des coefficients de la matrice du système complet.

- Puissance résiduelle estimée 1 : $-2.1dB$
- Puissance résiduelle estimée 2 : $-78dB$
- Puissance résiduelle estimée 3 : $-2.1dB$

Comme on s'y attendait, cette méthode ne permet pas de séparer les deux premières sources qui ont la même autocorrélation (s_1 et s_2). On obtient deux combinaisons linéaires orthogonales des ces sources. Par contre la source s_3 , d'autocorrélation différente, est bien estimée.

Méthode à l'ordre supérieur : JADE [5]

$$M \times A = \begin{bmatrix} -0.134 & 0.846 & -0.478 \\ -0.086 & -0.535 & -0.867 \\ 0.989 & 0.066 & -0.161 \end{bmatrix}$$

- Puissance résiduelle estimée 1 : $-9.2dB$
- Puissance résiduelle estimée 2 : $-8.2dB$
- Puissance résiduelle estimée 3 : $-30.2dB$

Avec cette méthode les deux sources gaussiennes (s_2 et s_3) ne sont pas séparées. On en conserve deux combinaisons linéaires orthogonales. La source uniforme est bien estimée.

Méthode utilisant les statistiques d'ordre : Fenêtres de taille 30

$$M \times A = \begin{bmatrix} -0.069 & 1 & 0.022 \\ 0.001 & -0.027 & -1.003 \\ 0.999 & 0.068 & -0.018 \end{bmatrix}$$

- Puissance résiduelle estimée 1 : $-45.6dB$
- Puissance résiduelle estimée 2 : $-62.9dB$
- Puissance résiduelle estimée 3 : $-46.2dB$

Comme le laissait envisager les propriétés des statistiques d'ordre qui cumulent les avantages des deux procédures précédentes, la méthode permet bien d'estimer les trois sources.

6 Conclusion

En séparation de mélanges instantanés, le critère de séparation fondé sur les statistiques d'ordre donne de bons résultats, qu'il soit implémenté sous forme itérative ou adaptative. Il est surtout important de noter, que pour le problème de séparation aveugle de sources, ce critère combine les avantages des deux grandes classes de méthodes. Effectivement il est possible de séparer des gaussiennes d'autocorrélations différentes mais aussi des sources de même autocorrélation mais de lois différentes. Il faut cependant encore fournir une preuve théorique de son efficacité...

Références

- [1] A. Belouchrani. *Séparation autodidacte de sources : algorithmes, performances et applications à des signaux expérimentaux*. PhD thesis, ENST, 1995.
- [2] A. Belouchrani, K. Abed-Meraim, and J.F. Cardoso. A blind source separation technique using second order statistics. *IEEE Trans. on SP*, 45 :434-444, February 1997.
- [3] J.F. Cardoso. Blind signal separation : statistical principles. In *Proceeding of the IEEE, vol. 9., no 10, pp. 2009-2025*, 1998.
- [4] J.F. Cardoso and B. Laheld. Equivariant adaptive source separation. In *IEEE Trans. on Sig. Proc., 44(12) :3017-3030*, 1996.
- [5] J.F. Cardoso and A. Souloumiac. Blind beamforming for non Gaussian signals. *IEE Proceedings-F*, 140(6) :362-370, December 1993.
- [6] P. Comon. Independent component analysis, a new concept ? *Signal Processing*, (36), 1994.
- [7] H.-A. David. *Order Statistics, second edition*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., 1981.
- [8] J. Even. Séparation adaptative de sources à l'aide de statistiques d'ordre. Rapport de DEA, INPG, 2000.