

Méthodes de séparation de sources dans le cas sous-déterminé

Olivier BERMOND, Jean-François CARDOSO

ENST, département TSI
46, rue Barrault 75634 Paris Cedex 13
bermond,cardoso@tsi.enst.fr

Résumé – Dans cette contribution, le problème de séparation de sources dans le cas où l'on dispose de moins de capteurs que de sources est étudié. Des résultats d'identifiabilité sont rappelés, puis plusieurs méthodes d'identification du mélange, d'une part, et d'extraction des sources, d'autre part, sont présentées. Finalement les algorithmes d'identification sont illustrés par des simulations.

Abstract – In this contribution, the underdetermined blind source separation problem is addressed. We recall some known identifiability results, and present various methods for the identification of the mixture matrix, and the extraction of the sources. Finally computer simulations illustrate the identification algorithms.

1 Introduction

Le problème de séparation de sources ou d'analyse en composantes indépendantes (ACI) a reçu une attention croissante au cours des dernières années, et l'étude du modèle le plus simple (cas d'un mélange instantané carré ou sur-déterminé, *i.e.* plus de capteurs que de sources) est arrivée à une certaine maturité. En revanche, le problème sous déterminé n'a que peu été étudié. Le but de cet article est de présenter quelques méthodes pour résoudre ce problème, dans le cas particulier de sources réelles et à densité continue. Dans une première partie, le problème sera posé et des résultats sur l'identifiabilité du modèle seront rappelés, puis la résolution des deux parties du problème, l'identification du mélange et la reconstruction des sources, sera étudiée. Finalement des simulations illustreront les résultats théoriques.

2 Modèle et identifiabilité

Le modèle considéré est le modèle classique de séparation de sources; on dispose d'une série d'observations réelles :

$$\mathbf{x}_t = A\mathbf{s}_t + \varepsilon_t \quad 1 \leq t \leq T \quad (1)$$

où \mathbf{s}_t est le vecteur des sources ($n \times 1$), ε_t ($m \times 1$) est un bruit additif d'observation, et A ($m \times n$) est la matrice (réelle) de mélange. On suppose les processus $(\mathbf{s}_t)_t$ et $(\varepsilon_t)_t$ i.i.d (indépendants identiquement distribués) et indépendants entre eux. En outre, le vecteur source \mathbf{s} est supposé à composantes indépendantes, de densité de probabilité inconnue $r(\mathbf{s}) = \prod_i r_i(s_i)$, et le bruit d'observation est supposé gaussien $\varepsilon \sim \mathcal{N}(0, R_\varepsilon)$.

Nous supposons que le nombre de sources excède le nombre de capteurs: $m < n$.

Dans ce cas les méthodes classiques ne s'appliquent pas, et la matrice de mélange n'étant pas inversible, identifier

cette matrice ne suffit pas pour reconstruire les sources. L'identification et la reconstruction sont donc deux problèmes distincts dans le cas sous-déterminé.

Taleb *et al* [1] ont rappelé l'identifiabilité du mélange, démontrée par Kagan *et al* [2]. Ce résultat dit, sous réserve que les sources ne soient pas gaussiennes, que l'on peut identifier A aux indéterminations classiques près (multiplication à droite par une matrice diagonale et une matrice de permutation). En revanche, si \mathbf{s} et \mathbf{y} sont des v.a. à composantes indépendantes qui vérifient

$$\mathbf{x} = A\mathbf{s} = A\mathbf{y}$$

elles ne sont pas nécessairement égales, et le résultat de Kagan *et al* permet seulement d'affirmer que leurs secondes fonctions caractéristiques diffèrent d'un polynôme¹.

3 Estimation du mélange

3.1 fonction caractéristique

Le résultat d'identifiabilité précédemment cité est basé sur l'utilisation des fonctions caractéristiques. Une idée simple consiste à les utiliser pour l'estimation de la matrice de mélange. En effet, la seconde fonction caractéristique $\phi_{\mathbf{x}}$ de \mathbf{x} s'écrit simplement en fonction de celle de \mathbf{s} et des colonnes de A :

$$\begin{aligned} \phi_{\mathbf{x}}(\mathbf{u}) &\stackrel{def}{=} \log E \left[e^{i\mathbf{u}^T \mathbf{x}} \right] \\ &= \sum_{k=1}^n \phi_{s_k}(\mathbf{a}_k^T \mathbf{u}) + \phi_{\varepsilon}(\mathbf{u}) \end{aligned}$$

\mathbf{a}_k étant la k -ième colonne de A . On dispose d'une estimée de $\phi_{\mathbf{x}}$:

$$\hat{\phi}_{\mathbf{x}}(\mathbf{u}) = \log(\hat{E} e^{i\mathbf{u}^T \mathbf{x}})$$

¹Taleb *et al* ont démontré que dans le cas de sources discrètes $\mathbf{s} = \mathbf{y} + \mathbf{z}$ où \mathbf{z} est déterministe.

où \hat{E} est l'opérateur d'espérance empirique :

$$\hat{E}[f(\mathbf{x})] = \frac{1}{T} \sum_t f(\mathbf{x}_t).$$

Si l'on connaît la distribution des sources et celle du bruit (et donc $\phi_{\mathbf{s}}$ et ϕ_{ε}), une estimée de la matrice de mélange est fournie par la minimisation du critère des moindres carrés entre la fonction caractéristique empirique et théorique:

$$C_{\phi} = \sum_{p=1}^N |\phi_{\mathbf{x}}(\mathbf{u}_p; A) - \hat{\phi}_{\mathbf{x}}(\mathbf{u}_p)|^2. \quad (2)$$

Ici, N représente le nombre de points où l'on calcule les fonctions caractéristiques, et donc le nombre de contraintes du problème à résoudre.

Le problème est que l'on ne connaît pas (en général) la distribution des sources, et l'on ne dispose donc pas de la fonction caractéristique théorique; dans ce cas, il est possible de l'approcher en remplaçant les distributions des sources inconnues par un modèle paramétrique, et en estimant les paramètres correspondants en même temps que la matrice. Il faut pour cela que le modèle paramétrique choisi soit capable d'approcher suffisamment la vraie fonction caractéristique des sources.

3.2 Statistiques d'ordre supérieur

Une alternative à l'utilisation de la fonction caractéristique est celle de statistiques plus simples, qui sont les cumulants. Comon *et al* [3] ont présenté le lien entre les cumulants et les polynômes de plusieurs variables, et notamment le fait que l'estimation de la matrice de mélange revient à trouver une décomposition d'un polynôme en somme de puissances de formes linéaires. Les cumulants considérés sont typiquement d'ordre 4. Un algorithme d'estimation dans le cas d'un mélange complexe 2 capteurs - 3 sources a été présenté dans [4].

Cardoso [5] a présenté une méthode algébrique basée sur le même type de décomposition des cumulants d'ordre 4. Cette méthode s'applique pour un nombre de sources limitée, mais plus grand que le nombre de capteurs. Un algorithme d'identification est présenté pour des sources ayant toutes un kurtosis de même signe; FOOBI (Fourth Order Only Blind Identification).

Il est à noter que les algorithmes de Comon et de Cardoso ne nécessitent aucune information a priori sur la distribution des sources.

L'utilisation des cumulants peut se faire sans exploiter leur structure algébrique particulière, mais comme outils d'ajustement du modèle. En effet, comme dans le chapitre précédent, il est possible de résoudre un critère des moindres carrés entre les cumulants empiriques, et les cumulants théoriques supposés². Le critère d'ajustement pour les cumulants d'ordre 4 fait intervenir les colonnes de A ainsi que les kurtosis $\kappa_1, \dots, \kappa_n$ des sources :

$$C_4(A, \kappa_1, \dots, \kappa_n) =$$

$$\sum_{i,j,k,l} |\hat{c}um(x_i, x_j, x_k, x_l) - \sum_{p=1}^n \kappa_p a_{ip} a_{jp} a_{kp} a_{lp}|^2$$

et le critère de contrainte du second ordre est:

$$C_2(A) = \sum_{i,j} |\hat{E}[x_i x_j] - \sum_{p=1}^n a_{ip} a_{jp}|^2. \quad (3)$$

Un critère d'ajustement utilisant l'ordre 2 et l'ordre 4 s'obtient par la combinaison linéaire :

$$C(A, \kappa_1, \dots, \kappa_n) = C_2(A) + \frac{1}{12} C_4(A, \kappa_1, \dots, \kappa_n) \quad (4)$$

où le coefficient 1/12 est suggéré par un développement de la divergence de Kullback en termes de cumulants.

Contrairement à l'ajustement des fonctions caractéristiques, les nombre d'inconnues et de contraintes indépendantes sont ici fixés et valent respectivement

$$\begin{aligned} N_{inc} &= (m+1)n \\ N_{cont} &= \frac{m(m+1)}{2} + \frac{m(m+1)(m+2)(m+3)}{24} \end{aligned}$$

Le premier terme de N_{cont} est le nombre de contraintes au second ordre, et le deuxième celui au quatrième (dimension de l'espace des tenseurs supersymétriques). Une condition nécessaire d'identifiabilité par cette méthode est que le nombre d'inconnues n'excède pas le nombre de contraintes; en d'autres termes le nombre de sources identifiables par cet ajustement de cumulants est majoré:

$$n \leq \frac{m}{2} + \frac{m(m+2)(m+3)}{24} \quad (5)$$

Cette méthode ne nécessite pas non plus de connaissances a priori sur la distribution des sources, puisque les kurtosis correspondants sont estimés en même temps que la matrice de mélange.

Limite d'identifiabilité : Pour l'algorithme FOOBI, il existe une borne du nombre de sources identifiables; pour appliquer la méthode, il faut nécessairement que $n(n-1) \leq m^2(m-1)^2/2$. La borne (5) représente un majorant du nombre maximum de sources identifiables par l'ajustement de cumulant. Le tableau ci-dessous rassemble ces majorants du nombre de sources pour FOOBI et pour l'ajustement de cumulant (ligne "CUM").

Il est intéressant d'avoir une idée du nombre de sources réellement identifiable; pour ce faire, nous avons cherché, pour chaque dimension, les solutions $(A, \kappa_1, \dots, \kappa_n)$ annulant le critère (4), en remplaçant les cumulants empiriques par leur vraies valeurs (exprimées en fonction de la vraie matrice et des kurtosis). Nous avons observé, pour certaines valeurs de m et n , plusieurs solutions du critère sensiblement différentes. Cet étude a donné un majorant plus précis du nombre de sources identifiables par l'ajustement de cumulants, il est indiqué à la ligne "OBS" du tableau.

m	2	3	4	5	6	7	8
FOOBI	2	4	9	14	21	30	40
CUM	2	5	9	14	21	29	40
OBS	2	4	8	13	20	28	36

Robustesse au bruit : L'algorithme FOOBI n'utilisant que les cumulants d'ordre 4, il est robuste au bruit gaussien. En revanche le critère (4) doit être modifié à l'ordre 2

²cette technique est connue sous le nom de "cumulant matching", déjà largement utilisée dans l'estimation de modèles ARMA par exemple.

pour s'adapter à la présence d'un bruit gaussien (en incluant la covariance du bruit dans les paramètres à estimer³).

3.3 Algorithme EM

Une alternative aux méthodes (algébriques ou d'ajustement) précédemment citées, est l'estimation par maximum de vraisemblance. En général, la forme intégrale de la vraisemblance ne permet pas le calcul de son maximum, mais il est possible d'avoir recours à l'algorithme EM [6]. Il est alors nécessaire d'estimer la distribution des sources conjointement à la matrice de mélange. Une méthode paramétrique pour laquelle la distribution des sources est modélisée par un mélange de gaussiennes a été présenté dans [7]. L'algorithme EM est itératif, et la réestimation de la matrice de mélange à la $(k+1)$ -ième itération s'écrit de manière simple dans le cas d'un bruit gaussien:

$$A^{(k+1)} = R_{xs}^{(k)} R_{ss}^{(k)-1} \quad (6)$$

$$R_{\varepsilon}^{(k+1)} = R_{xx} - R_{xs}^{(k)} R_{ss}^{(k)-1} R_{xs}^{(k)T} \quad (7)$$

où l'on a défini les moments conditionnels empiriques suivants :

$$R_{xs}^{(k)} = \hat{E} [\mathbf{x} E[\mathbf{s}^T | \mathbf{x}, \theta^k]] \quad (8)$$

$$R_{ss}^{(k)} = \hat{E} [E[\mathbf{s}\mathbf{s}^T | \mathbf{x}, \theta^k]] \quad (9)$$

θ représentant ici le paramètre global à estimer; c'est à dire la matrice A , la covariance du bruit R_{ε} et les paramètres du mélange de gaussiennes.

Avec l'algorithme EM, le nombre de sources identifiables n'est pas limité, et l'identification est robuste au bruit additif Gaussien. En revanche, il s'agit d'une méthode lourde, et dépendant fortement de son initialisation (à cause des minima locaux de la fonctionnelle EM et/ou de la vraisemblance).

4 Extraction des sources

Comme souligné précédemment, le problème de séparation de sources n'est pas équivalent à celui de l'estimation de la matrice de mélange dans le cas sous-déterminé. En effet, même si l'on connaît la matrice A , il existe un espace affine de solutions à l'équation

$$\mathbf{x} = A\mathbf{s} + \varepsilon. \quad (10)$$

Comon et Grellier [8] proposent une méthode d'extraction de sources dans le cas de sources complexes et discrètes, en exploitant leurs distributions particulières.

Il est clair que l'on doit utiliser une information a priori sur les sources pour résoudre le problème de leur extraction.

4.1 Minimum de l'EQM

Supposons la matrice de mélange A connue. On peut définir parmi les solutions de (10) celle qui minimise l'erreur

³Dans ce cas, le nombre de sources identifiables est plus faible, puisqu'il y a plus d'inconnues à identifier

quadratique moyenne (EQM). On sait que cette solution (MEQM) est donnée par l'espérance conditionnelle :

$$\hat{\mathbf{s}}_{meqm} = \frac{\int \mathbf{s} r(\mathbf{s}) e^{-\frac{1}{2}(\mathbf{x}-A\mathbf{s})^T R_{\varepsilon}^{-1}(\mathbf{x}-A\mathbf{s})} d\mathbf{s}}{\int r(\mathbf{s}) e^{-\frac{1}{2}(\mathbf{x}-A\mathbf{s})^T R_{\varepsilon}^{-1}(\mathbf{x}-A\mathbf{s})} d\mathbf{s}} \quad (11)$$

qui nécessite donc la connaissance de la densité r des sources et de la covariance du bruit R_{ε} , ou du moins de leurs estimées. L'algorithme EM précédemment cité permet de réestimer conjointement A , R_{ε} et les paramètres du mélange de gaussiennes modélisant la distribution des sources; et dans ce cas, on peut calculer (de manière exacte) l'espérance conditionnelle correspondant à ces estimés.

4.2 Maximum a Posteriori

Pour résoudre (10) on peut aussi considérer la solution donnée par le maximum de la loi a posteriori (MAP), défini par:

$$\begin{aligned} \hat{\mathbf{s}}_{map} &= \underset{\mathbf{s}}{\text{Argmax}} p(\mathbf{s} | \mathbf{x}) \\ &= \underset{\mathbf{s}}{\text{Argmax}} \left\{ \log r(\mathbf{s}) - \frac{1}{2}(\mathbf{x} - A\mathbf{s})^T R_{\varepsilon}^{-1}(\mathbf{x} - A\mathbf{s}) \right\} \end{aligned} \quad (12)$$

On peut montrer [9] qu'une approximation de la loi a posteriori par une loi gaussienne dans l'expression du MEQM (11) donne précisément le MAP. C'est à dire que ce dernier correspond à une approximation du MQEM.

Dans l'hypothèse d'un *bruit négligeable*, le MAP est particulièrement facile à calculer pour certaines lois a priori; les gaussiennes généralisées. Dans ce cas, on a

$$\log r(\mathbf{s}) = c - \lambda |\mathbf{s}|_{\alpha}$$

où c et $\lambda > 0$ sont des constantes, et $|\mathbf{s}|_{\alpha} = \sum |s_i|^{\alpha}$ est la puissance α -ième de la norme α de \mathbf{s} ($\alpha > 0$).

Pour de tels a priori, le MAP correspond au minimum de la norme α sur l'espace affine des solutions $\{\mathbf{x} = A\mathbf{s}\}$. En particulier, pour $\alpha = 2$, *i.e.* pour un a priori gaussien, le MAP est donné par la formule de pseudo-inversion de la matrice A :

$$\hat{\mathbf{s}}_2 = A^T (A A^T)^{-1} \mathbf{x}. \quad (13)$$

Pour $\alpha = 1$ (a priori Laplacien ou double exponentiel), le minimum de la norme 1 sur l'espace des solutions $\{\mathbf{x} = A\mathbf{s}\}$ est atteint en un sommet de cet espace, et on peut le calculer par des algorithmes de programmation linéaire (voir par exemple [10]). Il s'agit dans ce cas d'un vecteur solution ayant le maximum de composantes nulles (précisément $n - m$ si A est de rang m); on parle alors de représentations "parsimonieuses".

Pour $\alpha < 1$, le MAP correspond aussi à un sommet de l'espace des solutions et Kreutz-Delgado *et al* [11] ont proposé un algorithme pour le calcul de cet estimateur.

Quand le bruit n'est plus négligeable (et pour n'importe quelle loi a priori) l'estimateur du MAP peut se calculer par un algorithme de gradient classique, sauf dans le cas d'un a priori gaussien où il a une expression simple

$$\hat{\mathbf{s}}_{2b} = R_{\varepsilon} A^T (A R_{\varepsilon} A^T)^{-1} \mathbf{x} \quad (14)$$

5 Simulations

Fonction caractéristique :

Des expériences ont été menées en utilisant des mélanges de gaussiennes pour la distribution des sources, mais les résultats ne sont pas très concluants dès que la vraie distribution des sources (synthétiques) utilisées n'est pas un mélange de gaussiennes.

D'autres expériences, utilisant une modélisation par une densité de Laplace ont montré une meilleure efficacité pour des sources symétriques et fortement kurtiques, pour une moindre complexité.

Malgré l'intérêt de la méthode par rapport à l'identifiabilité théorique de n'importe quelle taille de matrice de mélange, les résultats obtenus montre qu'elle n'est pas très robuste à une mauvaise paramétrisation de la distribution des sources.

Ajustement de cumulants :

Des expériences ont été menées pour comparer les algorithmes FOABI et celui d'ajustement de cumulants par la minimisation de (4), les résultats ont été comparés à ceux de l'algorithme EM :

Nous avons mené 100 expériences d'identification de la matrice

$$A = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{pmatrix}$$

en utilisant un échantillon (de taille 5000) de signaux synthétiques, simulés suivant une loi de Laplace. Cette série d'expériences a été conduite pour plusieurs rapports signal à bruit (RSB) (exprimés en dB), et pour trois algorithmes; l'ajustement de cumulants par minimisation du critère des moindres carrés (4) (colonne "CUM"), l'algorithme FOABI, ainsi que l'algorithme EM utilisant les mélanges de gaussiennes. L'erreur d'estimation est l'erreur relative ($\|\cdot\|_F$ désignant la norme de Froebenius) :

$$err = \frac{\|A - \hat{A}\|_F}{\|A\|_F}$$

où \hat{A} représente la matrice estimée après avoir fixé les indéterminations d'échelle et de permutation. μ est l'erreur moyenne, et σ son écart type.

RSB		CUM	FOABI	EM
40	μ	0.0561	0.0601	0.0391
	σ	0.0428	0.0456	0.0297
20	μ	0.0559	0.0587	0.0401
	σ	0.0445	0.0471	0.0307
5	μ	0.0743	0.0783	0.0696
	σ	0.0578	0.0606	0.0525

6 Conclusion

Nous avons vu que l'identification de la matrice de mélange dans le cas sous-déterminé était possible, soit par des méthodes utilisant les cumulants, soit par l'algorithme EM. Ce dernier se révèle beaucoup plus lourd à mettre en oeuvre que les méthodes algébriques ou d'ajustement, mais offre en revanche un meilleur estimé, et est valable

(théoriquement) pour n'importe quelle dimension de la matrice de mélange.

Il est possible d'utiliser des cumulants d'ordre supérieur à 4 pour faire de l'ajustement, ce qui permettrait d'augmenter le nombre de sources identifiables par cette méthode, au dépens d'une complexité algorithmique croissante.

D'autre part, le problème d'extraction des sources (dans le cas de sources continues) est plus complexe, et sa résolution dépend fortement du contexte d'utilisation (par exemple, les solutions "parsimonieuses" seront préférées dans le cadre du codage d'image). Des expériences sont à mener, notamment sur des signaux réels, afin de comparer les différentes solutions proposées. Certains résultats seront présentés à la conférence.

References

- [1] A. Taleb and C. Jutten. On underdetermined source separation. In *ICASSP'99*, Phoenix, may 1999.
- [2] A.M. Kagan, Y.V. Linnik, and C.R. Rao. *Characterization Problems in Mathematical Statistics*. Wiley, 1973.
- [3] P. Comon and B. Mourrain. Decomposition of quantities in sums of powers of linear forms. *Signal Processing, Elsevier*, (53):93–107, September 1996.
- [4] P. Comon. Blind channel identification and extraction of more sources than sensors. In *SPIE Conf. Adv. Sig. Proc.*, VIII, San Diego, July 1998.
- [5] J. F. Cardoso. Super-symmetric decomposition of the fourth-order cumulant tensor. Blind identification of more sources than sensors. In *ICASSP*, 1991.
- [6] A. P. Dempster, N. M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc. B*, 39:1–38, 1977.
- [7] O. Bermond, E. Moulines, and J.F. Cardoso. Séparation et déconvolution aveugle de signaux bruités: modélisation par mélanges de gaussiennes. In *GRETSI'97*, Grenoble, September 1997.
- [8] P. Comon and O. Grellier. Nonlinear inversion of underdetermined mixtures. In *ICA'99*, pages 461–465, Aussois (France), January 1999.
- [9] O. Bermond and J.F. Cardoso. Approximate likelihood for noisy mixtures. In *ICA'99*, pages 325–330, Aussois (France), January 1999.
- [10] P.G. Ciarlet. *Introduction à l'analyse numérique matricielle et à l'optimisation*. Masson, Paris, 1990.
- [11] Kenneth Kreutz-Delgado and Bhaskar D. Rao. A general approach to sparse basis selection: majorization, concavity, and affine scaling. Technical Report UCSD-CIE-97-7-1, University of California, San Diego, July 1997.