

# Maximum de vraisemblance exact pour la séparation aveugle d'un mélange instantané de sources autorégressives gaussiennes

Serge DÉGERINE, Abdelhamid ZAIDI

LMC-IMAG, UMR CNRS 5523  
BP 53, F-38041 Grenoble cedex 9, France  
Serge.Degerine@imag.fr, Abdelhamid.Zaidi@imag.fr

**Résumé** – Le présent papier est consacré à l'étude du problème de la séparation aveugle d'un mélange instantané de sources gaussiennes autorégressives, sans bruit additif, par la méthode du maximum de vraisemblance exact. La maximisation de la vraisemblance est décomposée, par relaxation, en deux sous-problèmes d'optimisation, également traités par des techniques de relaxation. Le premier consiste en l'estimation de la matrice de séparation à structure autorégressive des sources fixée. Le second est d'estimer cette structure lorsque la matrice de séparation est fixée. Le premier problème est équivalent à la maximisation du déterminant de la matrice de séparation sous contraintes non linéaires. Nous donnons un algorithme de calcul de la solution de ce problème pour lequel nous précisons les conditions de convergence. Puis nous analysons, par simulation, les performances de l'estimateur ainsi défini et nous montrons l'amélioration qu'il apporte à la procédure de quasi-maximum de vraisemblance.

**Abstract** – This paper deals with the problem of blind separation of an instantaneous mixture of Gaussian autoregressive sources, without additive noise, by the exact maximum likelihood approach. The maximization of the likelihood is divided, using relaxation, into two sub-optimization problems, still solved by relaxation techniques. The first one consists in the estimation of the separation matrix when the autoregressive structure of the sources is fixed. The second one aims to estimate this structure when the separation matrix is fixed. We show that the first problem is equivalent to the determinant maximization of the separation matrix under nonlinear constraints. We propose in this paper an algorithm to compute the solution of such a problem and we look at its convergence properties. Then we study, by computer simulation, the performance of our estimator and show the improvement of its achievements with respect to the quasi-maximum likelihood estimator.

## 1 Introduction

Nous considérons dans ce rapport le problème de la séparation aveugle d'un mélange instantané de sources colorées sans bruit additif par des techniques du second ordre :

$$X(t) = AS(t). \quad (1)$$

Plusieurs méthodes ont été proposées ([2],[4],[6],[8]). La méthode du Quasi-Maximum de Vraisemblance (QMV), introduite par Pham et Garat [6], montre que la maximisation d'une vraisemblance approchée équivaut à la recherche de sources dont des versions filtrées soient décorréélées entre elles. Les filtres optimaux sont liés à la structure temporelle des sources. Celle-ci étant inconnue, ils sont remplacés par des filtres choisis a priori. Par ailleurs Burg [3] a montré que le principe du maximum d'entropie conduit à retenir la modélisation autorégressive dans ce type de situation. Ce modèle a d'ailleurs été utilisé récemment par Amari [1] pour la construction d'un algorithme adaptatif. Nous proposons donc de considérer la vraisemblance exacte associée à un mélange instantané de sources gaussiennes autorégressives. La maximisation est réalisée par des techniques de relaxation. On maximise la vraisemblance de façon alternative par rapport à chacun des deux paramètres, l'autre étant alors fixé, que constituent la matrice de séparation d'une part et la structure autorégressive des sources d'autre part. Chacune des

deux étapes de maximisation est également réalisée par relaxation. La seconde est une simple adaptation de la méthode du Maximum de Vraisemblance Exact (MVE) pour le modèle autorégressif de Pham [7]. Par contre, la première équivaut au problème géométrique  $\mathcal{P}$  suivant : Étant donnés  $K$  ellipsoïdes de  $\mathbb{R}^K$  cocentriques non homothétiques deux à deux,  $\mathcal{E}_1, \dots, \mathcal{E}_K$ , trouver  $K$  vecteurs  $V_1 \in \mathcal{E}_1, \dots, V_K \in \mathcal{E}_K$  qui maximisent le volume du parallépipède que forment ces vecteurs. Ce problème apparemment simple, n'entre pas dans le cadre de l'optimisation convexe classique [9]. Nous donnons un algorithme de calcul de la solution pour lequel nous précisons les propriétés de convergence. Nous étudions également le problème de l'unicité et celui de l'existence de maxima locaux dans le cas particulier où les axes principaux des ellipsoïdes sont colinéaires aux vecteurs de base de  $\mathbb{R}^K$ . Cette situation correspond au comportement asymptotique de l'estimateur de la structure des sources.

## 2 Présentation de la méthode MVE

### 2.1 Hypothèses

- La matrice  $A$  est carrée (autant de sources que de capteurs) et inversible

- Les sources  $S_k$  et  $S_j$  sont indépendantes  $\forall k \neq j = 1, \dots, K$
- $S_k(\cdot)$  est gaussienne autorégressive d'ordre  $p_k : AR(p_k)$

où  $R_k^T$  est la matrice de covariance de  $(S_k(1), \dots, S_k(T))^t$ , définie par  $\beta_k(j)$ ,  $j = 1, \dots, p_k$ ,  $\Lambda_k^T = \frac{1}{T} X_T (R_k^T)^{-1} X_T^t$  et  $\{e_k, k = 1, \dots, K\}$  est la base canonique de  $\mathbb{R}^K$ .

## 2.2 Notations

- $B = A^{-1}$  est la matrice de séparation
- $\sigma_k^2$  est la variance du processus d'innovation de  $S_k(\cdot)$
- $S_T = (S(1), \dots, S(T))$  est la matrice des sources (de taille  $(K, T)$ )
- $X_T = (X(1), \dots, X(T))$  est la matrice des observations (de taille  $(K, T)$ )

## 2.3 Indéterminations

La représentation (1) n'est pas unique au sens où nous pouvons toujours écrire :

$$X(\cdot) = AS(\cdot) = (AD\mathcal{E}\Pi) (\Pi^t \mathcal{E} D^{-1} S(\cdot)) = \tilde{A} \tilde{S}(\cdot),$$

où  $D$  est une matrice diagonale à éléments strictement positifs,  $\mathcal{E}$  est une matrice signe (diagonale à éléments égaux à  $\pm 1$ ) et  $\Pi$  est une matrice de permutation. Afin de fixer la paramétrisation du modèle, nous adoptons les conventions suivantes : (i) la variance d'innovation de chaque source  $S_k(\cdot)$  est fixée à 1 ( $D$ ), (ii) la première composante non nulle de chaque ligne de  $B$  est positive ( $\mathcal{E}$ ), (iii) les lignes de  $B$  sont rangées par ordre décroissant de leur première composante non nulle ( $\Pi$ ).

## 2.4 Fonction de vraisemblance

### 2.4.1 Choix du paramètre global

Les sources  $S_k(\cdot)$ , composantes de  $S(\cdot)$ , sont gaussiennes autorégressives d'ordre  $p_k$ , mutuellement indépendantes. Chaque modèle  $AR(p_k)$  est paramétré par ses autocorrélations partielles  $\beta_k(i)$ ,  $i = 1, \dots, p_k$ , car la variance de l'innovation est fixée égale à 1. Ce choix de normalisation laisse libre le paramètre  $A$ . On note  $\beta$  le paramètre de la structure des sources constitué par l'ensemble  $\{\beta_k(i), i = 1, \dots, p_k, k = 1, \dots, K\}$ . Rappelons que la seule contrainte sur  $\beta$  est que chaque  $\beta_k(i)$  est en module strictement inférieur à 1 et que  $\beta_k(\cdot) \neq \beta_j(\cdot)$  pour  $k \neq j$ , c'est-à-dire que les modèles  $AR(p_k)$  sont supposés distincts pour garantir l'identifiabilité de  $A$ . Chaque ordre  $p_k$  est fixé a priori ou estimé selon la procédure préconisée par De Luna [5]. Le paramètre global du problème de la séparation sera constitué par le couple  $(B, \beta)$ .

### 2.4.2 Fonction log-vraisemblance

L'expression de la fonction log-vraisemblance du problème est donnée, à une constante additive près, par :

$$L_T(B, \beta; X_T) = T \left( \log |\det B| - \frac{1}{2} \sum_{k=1}^K e_k^t B \Lambda_k^T B^t e_k \right) + \frac{1}{2} \sum_{k=1}^K \log |\det (R_k^T)^{-1}|$$

## 3 Algorithme de relaxation

On maximise la vraisemblance de façon alternative par rapport à chacun des deux paramètres, l'autre étant alors fixé, que constituent  $B$  et  $\beta$ . Chacune des deux étapes de maximisation est réalisée par relaxation.

### 3.1 Maximisation de $L_T$ par rapport à $\beta$ , à $B$ fixée

La maximisation de  $L_T$  par rapport à  $\beta$  est équivalente à la minimisation, de façon indépendante, du critère :

$$(S_T)_k. (R_k^T(\beta_k))^{-1} (S_T)_k. + \log(\det(R_k^T(\beta_k))), \quad (2)$$

associé à chaque source. L'estimation du paramètre  $\beta_k(\cdot)$  d'un processus  $AR$  par la méthode MVE a été étudiée par Pham [7]. Elle est réalisée par relaxation sur les  $\beta_k(j)$ ,  $j = 1, \dots, p_k$ . Nous proposons une version adaptée de cette méthode pour tenir compte du fait que  $\sigma_k^2 = 1$ . On démontre que  $\forall k = 1, \dots, K, \forall j = 1, \dots, p_k$ , à  $\beta_k(i)$   $i \neq j$  fixés, le critère (2) admet un seul minimum dans  $] -1, 1[$ . Ce minimum est déterminé par la résolution d'une équation du troisième degré.

### 3.2 Maximisation de $L_T$ par rapport à $B$ , à $\beta$ fixé

On démontre que la maximisation de  $L_T$  par rapport à  $B$ , à  $\beta$  fixé, est équivalente aux deux problèmes d'optimisation suivants :

- $\mathcal{P}$  : Soit  $K$  ellipsoïdes de l'espace  $\mathbb{R}^K$   $\mathcal{E}_1, \dots, \mathcal{E}_K$  concentriques et 2 à 2 non homothétiques, trouver  $K$  vecteurs  $B_1 \in \mathcal{E}_1, \dots, B_K \in \mathcal{E}_K$  qui maximisent le volume du parallélépipède que forment ces vecteurs.
- $\mathcal{P}^*$  : Maximiser le déterminant de  $B$  sous la contrainte  $\{\|B_k\|_{\Lambda_k^T} = 1 \quad k = 1, \dots, K\}$ .

Ce problème apparemment simple, n'entre pas dans le cadre de l'optimisation convexe classique [9].

#### 3.2.1 Principe de l'algorithme

La solution est recherchée par relaxation sur les lignes de la matrice  $B$ . En effet, les expressions du gradient et de la hessienne de la fonction  $L_T$  par rapport à la  $k^{\text{ème}}$  ligne de  $B$  sont données par :

$$\frac{\partial L_T}{\partial B_k} = T A (e_k - B \Lambda_k^T B^t e_k),$$

$$\frac{\partial^2 L_T}{\partial B_k^2} = - \left( \frac{1}{\det(B)^2} U U^t + \Lambda_k^T \right).$$

où  $U = (|B_{k1}|, \dots, |B_{kK}|)^t$  est le vecteur formé par les cofacteurs d'ordre  $(k, j)_{j=1 \dots K}$  de la matrice  $B$ . Ainsi, le gra-

dient de  $L_T$  s'annule en un seul point (à un facteur signe près) et la hessienne est définie négative.

**Théorème 3.1** *À structure autorégressive des sources fixée, l'algorithme de construction de la matrice de séparation converge sinon l'ensemble des points critiques de la fonction log-vraisemblance a la puissance du continu.*

### 3.2.2 Cas bidimensionnel: $K = 2$

Dans ce cas particulier, la solution est explicite et est obtenue à partir de la diagonalisation conjointe des matrices  $\Lambda_i^T, i = 1, 2$ . On démontre que les équations normales admettent deux solutions (à un facteur signe près). L'une correspond à un maximum  $\hat{B}$ , l'autre correspond à un point selle  $\tilde{B}$ .

Soient  $\alpha_1^2, \alpha_2^2$  les valeurs propres de  $\Lambda_1$  relativement à  $\Lambda_2$ , avec  $0 < \alpha_1 < \alpha_2$  (si  $\alpha_1 = \alpha_2$ ,  $\Lambda_1$  et  $\Lambda_2$  sont proportionnelles),  $W$  la matrice définie de façon unique (à un facteur de signe près) par :

$$W\Lambda_1 W^t = \begin{pmatrix} \alpha_1^2 & 0 \\ 0 & \alpha_2^2 \end{pmatrix}, \quad W\Lambda_2 W^t = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

alors :  $\hat{B} = \begin{pmatrix} \frac{1}{\alpha_1} & 0 \\ 0 & 1 \end{pmatrix} W, \tilde{B} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \alpha_1 & 0 \\ 0 & \frac{1}{\alpha_2} \end{pmatrix} \hat{B}$

### 3.2.3 Situation asymptotique

**Lemme 3.1** *Lorsque la structure autorégressive des sources est connue,*

$$\Lambda_k^T \xrightarrow{T \rightarrow +\infty} AD_k A^t \quad p.s.$$

où  $D_k$  est une matrice diagonale telle que  $D_k(k, k) = 1 < D_k(j, j) \quad \forall j \neq k = 1, \dots, K$ .

Dans ce cas  $B = A^{-1}$  est l'unique solution du problème  $\mathcal{P}$ . Notons que toute matrice obtenue par permutation des lignes de  $B$  est également solution des équations normales et que certaines peuvent correspondre à des maxima locaux. Ceci souligne la difficulté du problème global. Cependant l'étape d'estimation de la structure des sources permet d'éviter les maxima locaux. Par ailleurs, on démontre que lorsque  $T$  tend vers  $+\infty$ , le critère (2) admet un minimum unique atteint en  $\beta_k$ .

## 3.3 Comparaison avec la procédure QMV

On montre que la vraisemblance approchée de la méthode QMV, lorsque la structure des sources est fixée, est celle de la vraisemblance exacte avec préfenêtrage ( $X_i(t) = 0$  pour  $-p_i < t \leq 0$ ). Elle peut donc être résolue par relaxation plutôt que par la méthode de Newton, mal adaptée à la présence de solutions multiples aux équations normales. Les résultats de simulations montrent en effet que la méthode MVE améliore de façon très sensible les performances de la procédure QMV. Ceci est dû bien sûr à l'estimation de la structure des sources, mais aussi au fait d'éviter le piège de solutions parasites grâce à la méthode de relaxation. Il faut d'ailleurs souligner que la méthode MVE s'accommode très bien d'une estimation grossière de la structure des sources, par exemple en utilisant la méthode de Yule-Walker, ce qui lui conserve tout son intérêt sur le plan pratique.

## 4 Résultats de simulation

Les simulations présentées dans cette partie concernent l'étude des qualités de l'estimateur MVE et l'analyse de la part de l'amélioration qu'il apporte à la procédure QMV due à chacune des approximations que constituent le choix des filtres, l'utilisation du préfenêtrage et l'algorithme de résolution des équations normales.

### 4.1 Modèles

Les modèles suivants seront utilisés lors des simulations. Les modèles  $M_1, M_2$  et  $M_4$  sont autorégressifs définis par les autocorrélations partielles  $\beta$ . Le modèle  $M_3$  est une moyenne mobile définie par ses paramètres habituels  $b$ .

$M_1$ : 2 sources  $AR(2), \beta = \{0.4, 0.25; 0.6, 0.8\}$

$M_2$ : 2 sources  $AR(3), \beta = \{0.3, 0.2, 0.1; 0.4, 0.6, 0.8\}$

$M_3$ : 2 sources  $MA(2), b = \{-0.3, -0.25; -0.12, -0.8\}$

$M_4$ : 3 sources  $AR(2), AR(3), AR(4)$ , où

$$\beta = \{0.4, 0.25; 0.6, 0.8, -0.5; 0.5, 0.7, -0.5, 0.4\}$$

$$A = \begin{pmatrix} 1.0 & 0.2 \\ -0.4 & 0.8 \end{pmatrix}, \quad A = \begin{pmatrix} 1.2 & 1 & 0.6 \\ -0.45 & 0.8 & 1 \\ 0.78 & -1.54 & 1.68 \end{pmatrix}$$

### 4.2 Mesure de la qualité des estimateurs

Afin de comparer des estimateurs obtenus à partir de procédures de séparation utilisant des règles de normalisation différentes, nous avons retenu deux critères de comparaison (synthèse du biais et synthèse de la racine de l'erreur quadratique moyenne) définis à partir de versions normalisées des matrices  $A$  et  $\hat{A}$  selon les règles suivantes : (i) les colonnes de  $A$  sont de norme 1, (ii) la première composante non nulle de chaque colonne de  $A$  est positive, (iii) les colonnes de  $A$  sont rangées par ordre décroissant de leur première composante non nulle (ou des suivantes en cas d'égalité). Soient  $T$  la longueur des séquences et  $r$  le nombre de répétitions alors :

$$\sqrt{EQM_{ij}} = \sqrt{\frac{1}{r} \sum_{n=1}^r (\hat{A}_{ij}^n - A_{ij})^2}; \quad b_{ij} = \left( \frac{1}{r} \sum_{n=1}^r \hat{A}_{ij}^n \right) - A_{ij}$$

$$SYN_{EQM} = \sqrt{\frac{\sum_{i=1}^K \sum_{j \neq i=1}^K EQM_{ij}}{K(K-1)}}; \quad SYN_b = \sqrt{\frac{\sum_{i=1}^K \sum_{j \neq i=1}^K b_{ij}^2}{K(K-1)}}$$

### 4.3 Simulations

#### 4.3.1 Comparaison avec la procédure QMV

Les filtres arbitraires ou optimaux  $\Psi_p$  utilisés par la procédure QMV sont obtenus à partir des filtres  $F_p$  et de la relation (3).

$$\Psi_p(j) = \sum_{i=\max\{0, -j\}}^{\min\{p, p-j\}} F_p(i) F_p(i+j) \quad j = -p, \dots, p. \quad (3)$$

L'indice  $p$  dans  $F_p$  correspond à l'ordre d'approximation du modèle  $AR$ .

$F_1 = \{1.0, 1.0; -1.0, 1.0; -0.5, 1.0\}$

$F_2 = \{1.0, 1.0, 1.0; 1.0, -1.0, 1.0; -0.5, 1.0, 0.5\}$

$F_3 = \{1.0, 1.0, 1.0, 1.0; 1.0, -1.0, -1.0, 1.0; 1.0, -0.5, 0.5, 1.0\}$

$F_4 = \{1.0, 1.0, 1.0, 1.0, 1.0; 1.0, -1.0, -1.0, -1.0, 1.0; 1.0, 0.5, -1.0, 0.5, 1.0\}$ .

Les filtres optimaux  $\Psi_p$  sont ceux associés aux filtres  $\Phi_p$  des modèles autorégressifs des sources par l'algorithme de Levinson-Durbin.

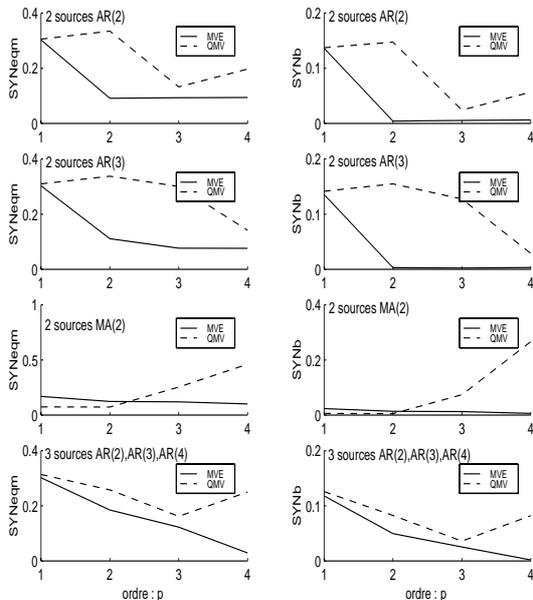


FIG. 1: Comparaison des performances des procédures MVE et QMV avec des filtres arbitraires

TAB. 1: Effet de l'algorithme de résolution des équations normales: QMV avec filtres optimaux; MVE avec préfenêtrage (MVEP) et filtres optimaux;  $T = 50$ ;  $r = 500$

	$M_1$		$M_2$		$M_4$	
	MVEP	QMV	MVEP	QMV	MVEP	QMV
$SYN_b$	0.0572	0.0728	0.0345	0.0482	0.0431	0.0702
$SYNEQM$	0.1891	0.2041	0.1421	0.1686	0.1719	0.2313

### 4.3.2 Robustesse de l'estimateur MVE

Lorsqu'il n'est pas estimé on utilise, dans les tableaux ci-dessous, l'ordre autorégressif exact des sources.

TAB. 2: Substitution dans MVE de l'estimateur de la structure des sources par celui de Yule-Walker (MVEY);  $T = 50$ ;  $r = 500$

	$M_1$		$M_2$		$M_4$	
	MVE	MVEY	MVE	MVEY	MVE	MVEY
$SYN_b$	0.0650	0.0687	0.0290	0.0310	0.0446	0.0535
$SYNEQM$	0.1921	0.2047	0.1485	0.1510	0.1729	0.1909

TAB. 3: Inclusion dans la procédure MVE de l'estimation de l'ordre autorégressif (MVEO);  $T = 250$ ;  $r = 200$

	$M_1$		$M_2$		$M_4$	
	MVE	MVEO	MVE	MVEO	MVE	MVEO
$SYN_b$	0.0043	0.0050	0.0023	0.0031	0.0031	0.0058
$SYNEQM$	0.0913	0.0947	0.0767	0.0783	0.0283	0.0512

### 4.3.3 Commentaires

La figure 1 montre bien l'amélioration qu'apporte l'estimation de la structure des sources. Nous remarquons que l'EQM varie peu lorsque l'ordre d'approximation autorégressive des sources dépasse l'ordre exact. Cette propriété n'est pas vérifiée par l'estimateur QMV. Nous constatons aussi que l'EQM atteint son minimum lorsque cet ordre coïncide avec l'ordre exact. Celui-ci, étant inconnu dans la pratique, pourrait être estimé. Le tableau 3 illustre la robustesse de l'estimateur MVE vis à vis de cette estimation (selon la procédure préconisée par De Luna [5]). Par ailleurs, l'estimation de la structure autorégressive des sources par la méthode MVE est assez onéreuse. Le tableau 2 montre que la méthode MVE s'accommode très bien d'une estimation grossière (Yule-Walker) de la structure des sources. Le tableau 1, illustre la perte de la qualité due à l'algorithme de résolution des équations normales de la procédure QMV.

## Références

- [1] S-I. Amari. *ICA of temporally correlated signals-learning algorithm*. First International Workshop on Independent Component Analysis and Signal Separation, Aussois, France, pp. 13-18, Jan 1999.
- [2] A. Belouchrani, K. Abed-Meriam, J-F. Cardoso et E. Moulines. *A blind Source Separation Technique using second-order statistics*. IEEE Trans. Signal Processing, vol.45, No. 2, pp. 434-444, feb. 1997.
- [3] J-P. Burg. *Maximum entropy spectral analysis*. 37<sup>th</sup> Annual International SEG Meeting, Oklahoma City, Oklahoma, 1967.
- [4] S. Dégerine et R. Malki. *Innovations partielles canoniques d'un mélange instantané*. in Proc. XVI Colloque GRETSI, Grenoble, France, pp. 7-10, sept. 1997.
- [5] X. De Luna. *An improvement of Akaike's PFE criterion to reduce its variability*. Journal of Time Series Analysis, 19(4), pp. 457-471, 1995.
- [6] D-T. Pham et P. Garat. *Blind separation of mixture of independent sources through a quasi-maximum likelihood approach*. IEEE, Trans. on Signal Processing, vol. 45, N° 7, pp. 1712-1725, july 1997.
- [7] D-T. Pham. *Maximum likelihood estimation of autoregressive model by relaxation on the reflection coefficients*. IEEE, Trans. on Acoustics, Speech, and Signal Processing, vol. ASSP 36, N° 1, pp. 175-177, Jan 1988.
- [8] L. Tong, V-C. Soon, Y-F. Huang et R. Liu. *AMUSE: an algorithm for multiple unknown signal extraction*. in Proc. IEEE ISCAS, New Orleans, pp. 1365-1368, 1990.
- [9] L. Vandenberghe, S. Boyd et S-P. Wu. *Determinant maximization with linear matrix inequality constraints*. SIAM J. Matrix Anal. Appl. 19, N° 2, pp. 499-553, 1997.