

# SÉPARATION AVEUGLE AU SECOND ORDRE DE SOURCES CORRÉLÉES.

A. Belouchrani, K.Abed-Meraim

Télécom Paris / CNRS URA 820 / GdR TdSI, 46 rue Barrault, 75634 PARIS

email: adel@sig.enst.fr

## Résumé

La séparation aveugle de sources consiste à retrouver un ensemble de signaux statistiquement indépendants à partir de mélanges observés. Cet article introduit une nouvelle technique de séparation de sources exploitant la cohérence temporelle des signaux sources. A la différence des techniques classiques de séparation, cette approche utilisant seulement les statistiques du second-ordre est basée sur “la diagonalisation conjointe” des matrices de corrélation temporelle. Des simulations numériques illustrent la validité de la méthode dans des contextes difficiles.

## 1 Introduction

L’approche dite de séparation aveugle de sources consiste à estimer les signaux sources sans information *a priori* sur la forme du front d’onde ou la géométrie de l’antenne.

Ce problème a déjà fait l’objet de nombreux travaux qui n’exploitent que la structure spatiale instantanée du processus en utilisant les statistiques d’ordre supérieur à deux. Lacoume [1], Comon [2], et Cardoso [3] ont traité ce problème de séparation par des approches exploitant les cumulants d’ordre quatre alors que Jutten[4], Pham [5] et Fety [6] proposent des méthodes adaptatives utilisant des transformations non linéaires.

Dans le cas où les signaux sources possèdent une cohérence temporelle, Fety [6] et Tong et al. [7] montrent qu’il est possible de séparer les différents signaux en utilisant les matrices de corrélation des signaux capteurs. Ces matrices (voir plus loin) possèdent une structure simple qui permet de résoudre le problème de l’indentification aveugle par des procédures de décomposition propres.

Dans cet article, on propose une technique d’identification aveugle basée sur la diagonalisation conjointe d’un ensemble de matrices de corrélation. La robustesse est significativement accrue, en utilisant un ensemble de matrices plutôt qu’une seule matrice comme dans [6, 7]. D’autre part, un algorithme efficace est décrit.

L’article est organisé comme suit. Le modèle des données et les statistiques utilisées sont présentés dans la première section. La section suivante donne les conditions d’identifiabilité et décrit la solution proposée. La section 3 décrit une implantation de notre approche, suivie d’une étude numérique des performances.

## Abstract

Blind separation of sources consists in recovering a set of statistically independent signals whose only instantaneous mixtures are observed. Such mixtures occur in narrow band array data which can then be processed without knowing the array manifold (blindness). This paper introduces a new source separation technique exploiting the possible time coherence of the source signals. In contrast to other previously reported techniques, the proposed approach relies only on second-order statistics, being based on a ‘joint diagonalization’ of correlation matrices. The effectiveness of the method in difficult contexts is illustrated by numerical simulations.

## 2 Modèle et statistiques

On considère une antenne à  $n$  capteurs recevant des signaux émis par  $m$  sources à bande étroite. Le signal reçu à l’instant  $t$  noté  $x(t)$  est un vecteur aléatoire complexe ( $n \times 1$ ) qui s’écrit classiquement :

$$x(t) = y(t) + n(t) = As(t) + n(t) \quad (1)$$

où  $s(t)$  est un vecteur ( $m \times 1$ ) complexe dont la  $p$ -ème composante  $s_p(t)$  est le signal émis par la  $p$ -ème source.  $A$  est la matrice complexe  $n \times m$  de mélange supposée de rang plein mais toutefois inconnue.

Les signaux sources sont stationnaires au second ordre et mutuellement décorrélés :

$$E(s_i(t)s_j^*(t')) = \rho_i(t-t') \delta_{ij} \quad \forall t, t' \quad (2)$$

Le bruit additif  $n(t)$  est supposé spatialement et temporellement blanc, de variance  $\sigma$  et décorrélé des signaux sources.

Sous ces hypothèses, les matrices de corrélation ont la structure simple suivante :

$$R(\tau) \stackrel{\text{def}}{=} E(x(t+\tau)x^*(t)) = AS(\tau)A^H + \delta(\tau)\sigma I \quad (3)$$

où  $S(\tau) \stackrel{\text{def}}{=} E(s(t+\tau)s^*(t))$  est la matrice de corrélation des signaux sources, diagonale grâce à (2), ayant pour éléments diagonaux les  $\rho_i(\tau)$  pour  $1 \leq i \leq m$ .

### 3 Indéterminations et principe d'identification

Le but recherché est d'identifier la matrice de mélange  $A$  à partir de (3) sans connaître les  $S(\tau)$ . Le problème tel qu'il est posé ne permet pas une identification complète de  $A$ , car l'échange d'un facteur scalaire entre le signal source et le vecteur colonne de  $A$  correspondant n'affecte pas les signaux observés. Ainsi, on peut supposer sans aucune restriction que les signaux sources sont de variance unité, i.e.  $S(0) = I$ . Toutefois, on notera que cette normalisation laisse la phase de chaque colonne de  $A$  indéterminée. De même, on remarquera que la numérotation des signaux sources est arbitraire. Il en est donc de même pour les vecteurs colonnes de  $A$ . Ainsi,  $A$  est identifiable au mieux à une permutation et à des termes de phases près.

Soit  $W$  une matrice complexe  $m \times n$  telle que  $Wy(t)$  soit spatialement blanc (une telle matrice est dite matrice de blanchiment). En même temps que le blanchiment, la matrice  $W$  transforme le vecteur des signaux capteurs en un vecteur de dimension réduite  $m \times 1$ . Par ailleurs, on a par définition :

$$\begin{aligned} I &= E((Wy(t))(Wy(t))^*) = WAS(0)A^H W^H \\ &= (WA)(WA)^H \end{aligned} \quad (4)$$

montrant que si  $W$  est une matrice de blanchiment,  $WA$  est alors une matrice unitaire  $m \times m$ . Il s'en suit que pour toute matrice de blanchiment  $W$ , il existe une matrice unitaire ( $m \times m$ )  $U$  telle que :

$$WA = U \quad \text{ou} \quad A = W\#U \quad (5)$$

où  $\#$  désigne la pseudo-inverse.

Comme  $AS(0)A^H = R(0) - \sigma I$ , l'équation (4) montre qu'une telle matrice  $W$  peut être déterminée à partir de la partie signal de la covariance des signaux capteurs  $R(0)$ . D'autres déterminations possibles peuvent être adoptées telle que celle suggérée dans [8].

Les corrélations du signal blanchi  $\underline{x}(t) = Wx(t)$  sont données par :

$$\forall \tau \neq 0 \quad \underline{R}(\tau) = WR(\tau)W^H = US(\tau)U^H. \quad (6)$$

Comme  $U$  est unitaire et  $S(\tau)$  diagonale, la relation (6) signifie que les matrices de corrélation blanchies se diagonalisent par la transformation unitaire  $U$ .

Un principe simple d'identification aveugle consiste alors en une diagonalisation conjointe des matrices de corrélation blanchies.

On trouve cette idée sous différentes formes dans [6] et dans [7].

Rappelons que si les valeurs propres d'une matrice sont uniques, il n'en est pas de même pour les vecteurs propres. Pour des valeurs propres distinctes, les vecteurs propres normés sont déterminés à une phase et un ordre près. Comme les vecteurs propres des matrices de corrélation blanchies sont les colonnes de la matrice  $U$ , on trouve les mêmes indéterminations que dans le problème de la séparation de sources.

La diagonalisation d'une seule matrice  $\underline{R}(\tau)$  résout le problème de la séparation de sources si ses valeurs propres sont distinctes : les indéterminations apparaissent dans le cas de valeurs propres dégénérées. *A priori*, il n'est pas possible de déterminer un retard  $\tau$  pour lequel les valeurs propres de  $\underline{R}(\tau)$  soient distinctes. Si les sources ont des spectres de formes différentes, la dégénérescence des valeurs propres est peu probable. Mais ce problème n'est pas purement académique car il est évident que quand les valeurs propres de  $\underline{R}(\tau)$  sont proches de la dégénérescence, la robustesse de la détermination de  $U$  à partir d'une décomposition propre est sérieusement affectée. La situation est plus favorable si on considère la diagonalisation *simultanée* d'un ensemble de matrice  $\{\underline{R}(\tau_i) | i = 1, \dots, K\}$  de  $K$  matrices de corrélation blanchies. Cet ensemble est conjointement diagonalisé (au sens de la section 3.1) par  $U$  comme dans (6).

La matrice  $U$  est essentiellement unique (à une phase et une permutation près) si et seulement si pour toute paire  $(i, j)$  de sources, il existe au moins un retard  $\tau_i$  dans  $(\tau_1, \dots, \tau_p)$  tel que  $\rho_i(\tau_i) \neq \rho_j(\tau_i)$ . Il est clair que des sources, ayant des spectres normalisés identiques, ne peuvent pas être séparées.

#### 3.1 Critère de diagonalisation conjointe

On vient de voir qu'avec des statistiques exactes les matrices de corrélation blanchies se diagonalisent toutes sous une même transformation unitaire  $U$ .

En pratique, on ne dispose que de statistiques empiriques. Une diagonalisation conjointe exacte n'est en général pas possible à cause des erreurs d'estimation. Il faut donc utiliser une diagonalisation conjointe approchée [9]. Pour cela, on utilisera un critère de "proximité" à la diagonalisation parfaite. Diagonaliser une matrice normale  $M$  de taille  $m \times m$  revient à trouver une base orthonormée  $V = \{v_i | i = 1..m\}$  telle que, dans cette base, la somme des modules carrés des termes diagonaux soit maximale, ce qui est équivalent à trouver la base  $V$  maximisant le critère :

$$C(M, V) \stackrel{\text{def}}{=} \sum_i |v_i^* M v_i|^2 \quad (7)$$

sur l'ensemble des matrices unitaires  $V = [v_1, \dots, v_n]$ .

Pour un ensemble de matrices, on étend naturellement ce critère en définissant la diagonalisation conjointe approchée d'une famille  $\mathbf{M} = \{M_k | k = 1..K\}$  de  $K$  matrices normales comme la maximisation du critère :

$$C(V) \stackrel{\text{def}}{=} \sum_k C(M_k, V) = \sum_{k,i} |v_i^* M_k v_i|^2 \quad (8)$$

sous les mêmes contraintes unitaires.

L'algorithme de diagonalisation conjointe utilisé est donné en annexe : cet algorithme est basé sur une généralisation de la technique de Jacobi [10] de diagonalisation d'une seule matrice hermitienne. Le coût de calcul d'un tel algorithme est essentiellement équivalent à  $K$  fois celui de la diagonalisation d'une seule matrice hermitienne par la technique de Jacobi.

### 3.2 Implantation

L'implantation de l'algorithme de séparation s'effectue comme suit. D'abord, une estimation de la covariance  $\hat{R}(0)$  est calculée à partir des  $T$  échantillons reçus, puis celle-ci est diagonalisée :

$$\hat{R}(0) = L\Delta L^H \quad (9)$$

où  $L = [l_1, \dots, l_n]$  et  $\Delta = \text{diag}(\lambda_1, \dots, \lambda_n)$  avec  $\lambda_i \geq \lambda_j$  pour  $i > j$ . Le nombre de sources peut alors être estimé à partir du spectre  $\Delta$ ; toutefois ce problème d'estimation n'est pas considéré par notre travail. Une estimée  $\hat{\sigma}$  de la puissance du bruit est formée en moyennant les  $n - m$  plus petites valeurs propres de  $\Delta$ . Une matrice de blanchiment  $\hat{W}$  est alors calculée par :

$$\hat{W} = \Delta' L'^H \quad (10)$$

où  $\Delta' = \text{diag}((\lambda_1 - \hat{\sigma})^{-\frac{1}{2}}, \dots, (\lambda_m - \hat{\sigma})^{-\frac{1}{2}})$  et  $L' = [l_1, \dots, l_m]$ .

L'ensemble des données est alors blanchi par  $\hat{W}$ , les données  $\hat{W}x(t)$  ainsi obtenus permettent de former les estimées  $\hat{R}(\tau)$  des matrices de corrélation blanchies de dimension  $m \times m$  pour  $\tau \in \{\tau_j | j = 1, \dots, K\}$ , un ensemble fixé de retards. Une matrice unitaire  $\hat{U}$  est finalement obtenue par diagonalisation conjointe approchée de l'ensemble des matrices  $\{\hat{R}(\tau_j) | j = 1, \dots, K\}$ . Les signaux sources sont alors estimés par  $\hat{s}(t) = \hat{U}^H \hat{W}x(t)$ .

## 4 Performances

Après blanchiment les signaux capteurs sont de la forme :

$$\underline{x}(t) = \hat{W}x(t) = \sum_{p=1}^m s_p(t) \hat{W}a_p + \hat{W}b(t). \quad (11)$$

Si on note par  $\hat{a}_p$  les estimées des signatures blanches obtenues par l'algorithme précédent, l'estimée du  $p$ -ème signal source est alors donnée par :

$$\hat{s}_p(t) = \hat{a}_p^* \underline{x}(t) = \sum_{p=1}^m s_p(t) \hat{a}_p^* \hat{W}a_p + \hat{a}_p^* \hat{W}b(t). \quad (12)$$

Les sources étant de puissance unité, on définit alors le taux de réjection de la source  $q$  par rapport à la source  $p$  par :

$$i_{pq} = E(|\hat{a}_p^* \hat{W}a_q|^2) = E(|\hat{a}_p^* \hat{W}W^{\#} \underline{a}_q|^2) \quad (13)$$

où les  $\underline{a}_p$  désignent les vecteurs  $a_p$  blanchis :  $\underline{a}_p \stackrel{\text{def}}{=} W a_p$ .

Ici,  $i_{pq}$  représente la puissance moyenne de l'interférence de la source  $q$  dans l'estimée  $\hat{s}_p$  de  $s_p$ .

Un critère global de mesure des performances est donné par la somme des interférences :

$$I_{perf} = \sum_{q \neq p} i_{pq} \quad (14)$$

Pour les simulations, on se place dans le cas d'un réseau linéaire à trois capteurs équidistant d'une demi longueur d'onde, et de deux sources gaussiennes indépendantes localisées à  $\phi_1 = 0$  et  $\phi_2 = 0.4$ , tel que la  $q$ -ème composante

de la  $p$ -ème colonne de  $A$  est donnée par  $\exp(j\pi q \phi_p)$ . Le bruit additif gaussien est de variance  $\sigma I$  tandis que les signaux sources sont de variance unité et que chacun d'eux est généré par filtrage d'un processus gaussien, blanc, complexe et circulaire selon un modèle AR1. On choisit le coefficient du modèle auto régressif de la  $p$ -ème source sous la forme  $a_p = 0.85e^{j\theta_p}$ . Les statistiques sont évaluées sur des blocs de données de 500 échantillons et les taux de rejection empiriques sont calculés sur 100 réalisations. Les courbes présentées dans cette section sont indicées par le nombre de matrices de corrélation diagonalisées conjointement. Sur la figure 1, on représente le taux de rejection  $I_{perf}$  en dB en fonction de  $\sigma$  en dB. Ce tracé correspond aux choix  $\theta_1 = 0.5$  et  $\theta_2 = 0.55$ , représentant le cas d'une faible différence spectrale. A faible niveau du RSB, i.e. pour  $\sigma = 5dB$ , le taux de rejection est proche de  $-6dB$  signifiant que la méthode est essentiellement inutile dans ce cas de figure. Ainsi pour un niveau de bruit de  $0dB$ , la diagonalisation conjointe de 3 ou 4 matrices abaisse le taux de rejection à  $-13dB$  ce qui montre clairement le gain apporté dans de telles situations par l'utilisation de plusieurs matrices de corrélation. Pour de forts niveaux du RSB, il apparait que la diagonalisation d'une seule matrice est suffisante pour atteindre un taux de rejection de  $-18dB$ . Sur la figure 2 on maintient le niveau de bruit constant à  $\sigma = 0dB$  et  $\theta_1 = 0.5$  et on fait varier  $\theta_2 = \theta_1 + \delta\theta$ . Le tracé représente le taux de rejection  $I_{perf}$  en dB en fonction du décalage spectral  $\delta\theta$ . On observe ainsi que les performances sont significativement accrues pour un faible écart spectral en utilisant 3 ou 4 matrices de corrélation dans le critère de diagonalisation conjointe.

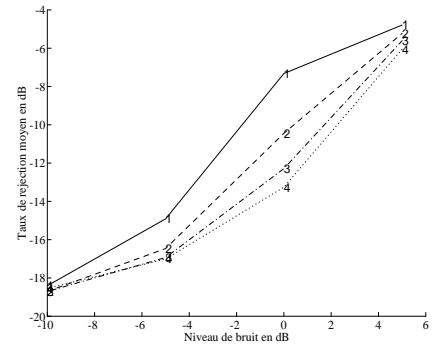


Fig. 1 Performance en fonction de  $\sigma$ .

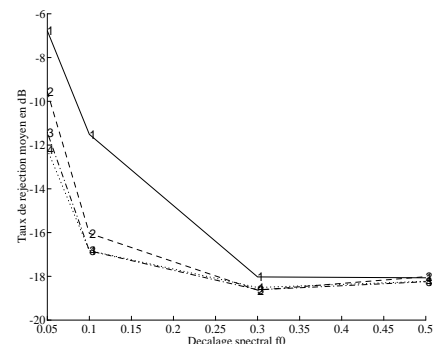


Fig. 2 Performance en fonction du décalage spectral  $\delta\theta$ .

## Conclusion

Nous avons proposé dans cet article une technique de séparation aveugle de sources corrélées temporellement. Cette technique est fondée sur la diagonalisation conjointe d'un ensemble de matrices de corrélation. L'avantage de la méthode présentée est d'accroître l'efficacité statistique. Ceci a été illustré par des simulations numériques qui montrent la nécessité de l'exploitation de plusieurs matrices de corrélation dans des contextes difficiles (faible RSB, sources avec un faible décalage spectral). Une analyse asymptotique des performances de la technique proposée est en cours et fera l'objet d'une publication ultérieure.

## Annexe : Algorithme de diagonalisation conjointe approchée

Nous généralisons [9] la technique de Jacobi de diagonalisation d'une seule matrice hermitienne au problème de la diagonalisation conjointe approchée d'un ensemble de matrices carrées arbitraires. La méthode proposée consiste alors à minimiser le critère (8) par une succession de rotations de Givens, ce qui nous conduit à la résolution du même problème pour  $K$  matrices  $2 \times 2$  :

$$H_k = \begin{bmatrix} a_k & b_k \\ c_k & d_k \end{bmatrix} \quad (15)$$

pour  $k = 1, \dots, K$ .

On cherche donc une matrice unitaire  $V$  telle que  $H'_k = V^H H_k V$  ( $k = 1, \dots, K$ ) maximisant le critère (8).

La transformation unitaire  $V$  est paramétrisée par :

$$V = \begin{bmatrix} \cos\theta & e^{i\phi}\sin\theta \\ -e^{-i\phi}\sin\theta & \cos\theta \end{bmatrix}, \quad (16)$$

En notant  $a'_k, b'_k, c'_k$  et  $d'_k$  les coefficients de  $H'_k$ , l'optimisation du critère (8) revient à chercher  $\theta$  et  $\phi$  qui maximiseront la quantité  $\sum_k |a'_k|^2 + |d'_k|^2$ . Cette maximisation est encore équivalente à celle de  $|a'_k - d'_k|^2$  puisque  $2(|a'_k|^2 + |d'_k|^2) = |a'_k - d'_k|^2 + |a'_k + d'_k|^2$  et que la trace  $a'_k + d'_k$  est invariante dans une transformation unitaire. Donc on est ramené à la maximisation de la seule quantité :

$$Q = \sum_k |a'_k - d'_k|^2 \quad (17)$$

Or on trouve :

$$a'_k - d'_k = (a_k - d_k)\cos 2\theta - (b_k + c_k)\sin 2\theta \cos\phi - (j(c_k - b_k))\sin 2\theta \sin\phi. \quad (18)$$

pour  $k=1, \dots, K$ . En définissant alors les vecteurs

$$u^T \stackrel{\text{def}}{=} [a'_1 - d'_1, \dots, a'_K - d'_K] \quad (19)$$

$$v^T \stackrel{\text{def}}{=} [\cos 2\theta, -\sin 2\theta \cos\phi, -\sin 2\theta \sin\phi] \quad (20)$$

$$g_k^T \stackrel{\text{def}}{=} [a_k - d_k, b_k + c_k, j(c_k - b_k)] \quad (21)$$

les  $K$  équations (18) s'écrivent sous la forme  $u = Gv$  où  $G^T = [g_1, \dots, g_K]$ . La quantité (17) à maximiser est égale

à la norme de  $u$  :

$$Q = u^H u = v^T (G^H G) v = v^T \text{Re}(G^H G) v \quad (22)$$

où on a utilisé le fait que la matrice  $G^H G$  étant hermitienne par construction, sa partie imaginaire est antisymétrique et par suite la contribution de cette dernière est nulle dans la forme quadratique (22).

La dernière étape consiste à remarquer que la paramétrisation particulière (20) de  $v$  est équivalente à la condition  $v^T v = 1$ . La maximisation de la forme quadratique  $Q$  sous la contrainte d'unitarité est alors obtenue en prenant  $v$  comme le vecteur propre associé à la plus grande valeur propre de  $\text{Re}(G^H G)$ .

## References

- [1] M. Gaeta and J.-L. Lacoume, "Source separation without a priori knowledge: the maximum likelihood solution," in *Proc. EUSIPCO*, pp. 621–624, 1990.
- [2] P. Comon, "Independent component analysis," in *Proc. Int. Workshop on Higher-Order Stat., Chamrousse, France*, pp. 111–120, 1991.
- [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*, 1993.
- [4] C. Jutten and J. Héroult, "Independent component analysis versus PCA," in *Proc. EUSIPCO*, pp. 643–646, 1988.
- [5] D. T. Pham, P. Garrat, and C. Jutten, "Separation of a mixture of independent sources through a maximum likelihood approach," in *Proc. EUSIPCO*, pp. 771–774, 1992.
- [6] L. Féty, "Méthodes de traitement d'antenne adaptées aux radio-communications," *Thèse de docteur-ingénieur de l'ENST*, June 1988.
- [7] R. L. L. Tong and Y. H. V. C. Soon, "Indeterminacy and identifiability of blind identification," *IEEE Tr. on CS*, vol. 38, pp. 499–509, May 1991.
- [8] Y. I. L. Tong and R. Liu, "A finite-step global convergence algorithm for the cumulant-based parameter estimation of multichannel moving average processes," in *Proc. ICASSP*, pp. 3445–3448, 1991.
- [9] A. Souloumiac and J.-F. Cardoso, "Comparaison de méthodes de séparation de sources," in *Proc. GRETSI, Juan les Pins, France*, 1991.
- [10] G. H. Golub and C. F. V. Loan, *Matrix computations*. The Johns Hopkins University Press, 1989.