

# Une généralisation de l'algorithme JADE

Eric MOREAU

MS-GESSY, ISITV, Université de Toulon et du Var  
av. G. Pompidou, BP56, 83162 La Valette du Var Cedex, France

moreau@isitiv.univ-tln.fr

**Résumé** – On considère le problème de la séparation de sources basée sur la diagonalisation conjointe d'un ensemble de matrices. L'algorithme JADE atteint cet objectif en utilisant des matrices de cumulants d'ordre quatre. Nous nous proposons dans cette communication de généraliser le contraste sur lequel repose JADE et son lien avec un critère de diagonalisation conjointe en l'étendant à des cumulants d'ordre quelconque supérieur ou égal à trois. L'intérêt principal est alors de pouvoir choisir l'ordre des cumulants ou de combiner des matrices de cumulants d'ordres différents.

**Abstract** – In the field of blind source separation, joint-diagonalization based approaches constitute an important framework leading to useful algorithms among which the popular JADE algorithm. This latter algorithm is derived using “only” fourth-order cumulants. In this paper, we are mainly interested in generalizing the JADE's underlying contrast and its link with a joint-diagonalization criterion involving cumulants of any order greater or equal to three. The main interest is then to be able to choose the cumulants order or to combine statistical information of different orders.

## 1 Introduction

Nous considérons le problème de la *séparation de sources* [4]. Les sorties d'un système réalisant un mélange de signaux aléatoires inconnus appelés sources, sont observées. Le mélange étant lui aussi supposé inconnu, on désire retrouver le plus fidèlement possible chacun des signaux d'entrée (les sources) grâce aux seules observations des sorties. Il s'agit donc de déterminer ou d'estimer un inverse du système mélangeant. Ce problème est souvent qualifié d'autodidacte, de non-supervisé ou d'aveugle dans la mesure où la séparation est réalisée avec pour “seule” information celle apportée par les observations. Pour cela il est nécessaire de faire l'hypothèse importante suivante : *les signaux sources sont statistiquement indépendants*. Notons que l'on pourra parfois avoir recours à des informations a priori sur les distributions des sources.

Deux des premières contributions importantes en séparation de sources sont les algorithmes JADE [1] et ICA [2]. Les deux sont des algorithmes blocs optimisant une certaine fonction de contraste (ceci de façon indirecte pour JADE). Cependant et contrairement à ICA, le contraste sous-jacent à JADE n'a été proposé que pour des cumulants d'ordre 4. Notre principal objectif dans cette contribution est de développer l'approche relative à JADE et de l'étendre à des cumulants d'ordre quelconque supérieur ou égal à trois.

## 2 Notations et hypothèses

Le modèle de mélange est donné par :

$$\mathbf{x}(n) = \mathbf{G}\mathbf{a}(n) \quad (1)$$

où  $\mathbf{x}(n)$  est le vecteur  $(N, 1)$  des observations,  $\mathbf{a}(n)$  le vecteur  $(N, 1)$  des sources,  $n$  le temps discret et  $\mathbf{G}$  une matrice carrée  $(N, N)$  *invertible* appelée matrice de mélange.

Le but étant de restituer les sources, il s'agit d'estimer une matrice de séparation  $\mathbf{H}$  agissant sur les observations  $\mathbf{x}$  :

$$\mathbf{y}(n) = \mathbf{H}\mathbf{x}(n) \quad (2)$$

de telle sorte que les composantes du vecteur  $(N, 1)$   $\mathbf{y}(n)$  correspondent aux différentes sources  $a_i(n)$ ,  $i \in \{1, \dots, N\}$ . On définit la matrice  $\mathbf{S}$  du système global par :

$$\mathbf{S} = \mathbf{H}\mathbf{G} \quad , \quad (3)$$

et ainsi grâce à (1) et (2), on a :

$$\mathbf{y}(n) = \mathbf{S}\mathbf{a}(n) \quad . \quad (4)$$

Les signaux sources étant inobservables, il subsiste deux indéterminations quant à leurs restitutions. En général, la puissance de même que l'ordre des sources ne peuvent pas être restitués. Elles seront donc considérées comme séparées lorsque :

$$\mathbf{S} = \mathbf{D}\mathbf{P} \quad (5)$$

où  $\mathbf{D}$  est une matrice diagonale inversible quelconque et  $\mathbf{P}$  une matrice de permutation elle aussi quelconque.

Il est nécessaire de faire les hypothèses suivantes :

**A1a.** Les sources  $a_i(n)$ ,  $i \in \{1, \dots, N\}$ , sont centrées, de puissance 1 et statistiquement indépendantes ;

**A1b.**  $\mathbf{a}(n)$  est un processus aléatoire stationnaire à l'ordre des cumulants considérés, *i.e.*  $\forall i \in \{1, \dots, N\}$ , Cum  $\underbrace{[a_i(n), \dots, a_i(n)]}_{R \times}$  est indépendant du temps discret

$n$ . Il est noté  $C_R[a_i]$ .

**A1c.** Au plus un des cumulants  $C_R[a_i]$ ,  $i \in \{1, \dots, N\}$  est nul.

**A2.**  $\mathbf{S}$  est une matrice unitaire, *i.e.*  $\mathbf{S}\mathbf{S}^T = \mathbf{I}$  où  $\mathbf{I}$  est la matrice identité.

Il est à noter que l'hypothèse A2 est faite sans perte de généralité. En effet elle peut toujours être adoptée moyennant un blanchiment préalable des observations.

L'ensemble des vecteurs aléatoires satisfaisant A1a à A1c est noté  $\mathcal{A}$ . L'ensemble des matrices satisfaisant A2 est noté  $\mathcal{U}$ . Le sous ensemble de  $\mathcal{U}$  des matrices satisfaisant la relation (5) est noté  $\mathcal{P}$  et le sous ensemble de  $\mathcal{P}$  des matrices diagonales est noté  $\mathcal{D}$ . Finalement l'ensemble des vecteurs aléatoires  $\mathbf{y}(n)$  déduit de (4) où  $\mathbf{a}(n) \in \mathcal{A}$  et  $\mathbf{S} \in \mathcal{U}$  est noté  $\mathcal{Y}_u$ .

### 3 Fonctions de contraste

Les fonctions de contraste ont été introduites dans [2] par P. Comon puis récemment généralisées dans [7]. Rappelons leur définition

**Définition 1.** *Un contraste sur  $\mathcal{Y}_u$  est une fonction multivariable  $\mathcal{I}(\cdot)$  de l'ensemble  $\mathcal{Y}_u$  sur  $\mathbb{R}$  satisfaisant les trois propriétés suivantes :*

- R1.  $\forall \mathbf{y} \in \mathcal{Y}_u, \forall \mathbf{D} \in \mathcal{D}, \mathcal{I}(\mathbf{D}\mathbf{y}) = \mathcal{I}(\mathbf{y})$ ;
- R2.  $\forall \mathbf{a} \in \mathcal{A}, \forall \mathbf{S} \in \mathcal{U}, \mathcal{I}(\mathbf{S}\mathbf{a}) \leq \mathcal{I}(\mathbf{a})$ ;
- R3.  $\forall \mathbf{a} \in \mathcal{A}, \forall \mathbf{S} \in \mathcal{U}, \mathcal{I}(\mathbf{S}\mathbf{a}) = \mathcal{I}(\mathbf{a}) \Rightarrow \mathbf{S} \in \mathcal{P}$ .

De tels contrastes sont des fonctions invariantes par changement d'échelle (R1) devant être maximisées (R2) afin de résoudre le problème de la séparation (R3). Un des premiers contrastes peut être trouvé dans [2]. Il s'agit de la fonction

$$\mathcal{I}_R^2(\mathbf{y}) = \sum_{i=1}^N |\mathcal{C}_R[y_i]|^2, \quad R \geq 3.$$

La maximisation de ce contraste a conduit à l'algorithme ICA. Un peu plus tard le contraste

$$\mathcal{I}_R^1(\mathbf{y}) = \sum_{i=1}^N |\mathcal{C}_R[y_i]|, \quad R \geq 3$$

a été proposé [5]. Ces deux derniers contrastes ont récemment été généralisés [7] : ainsi la fonction non symétrique

$$\mathcal{I}_R^f(\mathbf{y}) = \sum_{i=1}^N \gamma_i f(|\mathcal{C}_R[y_i]|), \quad R \geq 3$$

où  $f(\cdot)$  est une fonction convexe strictement croissante et où les  $\gamma_i, i \in \{1, \dots, N\}$  satisfont  $\gamma_1 \geq \dots \geq \gamma_N > 0$ , est un contraste.

On peut remarquer qu'aucun cumulants croisés n'intervient dans les contrastes précédents. Dans [1] un contraste comportant à la fois les auto-cumulants et des cumulants croisés est proposé. Il s'agit de la fonction

$$\mathcal{J}(\mathbf{y}) = \sum_{i_1, i_2, i_3=1}^N |\text{Cum}[y_{i_1}, y_{i_2}, y_{i_3}]|^2 \quad (6)$$

que l'on appellera le "contraste JADE".

Notre premier objectif est de généraliser le contraste précédent à des cumulants d'ordre quelconque supérieur ou égal à trois. Ceci est énoncé au niveau de la proposition suivante :

**Proposition 1.** *Soit  $R$  un entier supérieur ou égal à 3, la fonction*

$$\mathcal{J}_R(\mathbf{y}) = \sum_{i_1, \dots, i_{R-1}=1}^N |\text{Cum}[y_{i_1}, y_{i_2}, \dots, y_{i_{R-1}}]|^2 \quad (7)$$

est un contraste.

**Preuve.** Avec  $\mathbf{S} = (S_{i,j})$ , en utilisant (4) et grâce à la multi-linéarité des cumulants et à l'indépendance des sources, on a

$$\text{Cum}[y_{i_1}, \dots, y_{i_{R-1}}] = \sum_{\ell} S_{i_1, \ell}^2 S_{i_2, \ell} \cdots S_{i_{R-1}, \ell} \mathcal{C}_R[a_{\ell}]. \quad (8)$$

Ainsi

$$\begin{aligned} \mathcal{J}_R(\mathbf{y}) &= \sum_{i_1, \dots, i_{R-1}} \sum_{\ell_1 \ell_2} S_{i_1, \ell_1}^2 S_{i_2, \ell_2} \cdots S_{i_{R-1}, \ell_1} \\ &\quad S_{i_1, \ell_2}^2 S_{i_2, \ell_2} \cdots S_{i_{R-1}, \ell_2} \mathcal{C}_R[a_{\ell_1}] \mathcal{C}_R[a_{\ell_2}] \\ &= \sum_{\ell_1 \ell_2} \left( \sum_{i_1} S_{i_1, \ell_1}^2 S_{i_1, \ell_2}^2 \right) \left( \sum_{i_2} S_{i_2, \ell_1} S_{i_2, \ell_2} \right)^{R-2} \\ &\quad \mathcal{C}_R[a_{\ell_1}] \mathcal{C}_R[a_{\ell_2}]. \end{aligned} \quad (9)$$

De plus, comme  $\mathbf{S}$  est une matrice unitaire, on a  $\forall \ell_1, \ell_2, \sum_m S_{m, \ell_1} S_{m, \ell_2} = \delta_{\ell_1, \ell_2}$  où  $\delta_{i,j} = 1$  si  $i = j$  et 0 sinon. Donc à partir de (9) et parce que

$$\forall \ell_1, \quad \sum_{i_1} S_{i_1, \ell_1}^4 \leq \sum_{i_1} S_{i_1, \ell_1}^2 = 1,$$

on a

$$\begin{aligned} \mathcal{J}_R(\mathbf{y}) &= \sum_{\ell_1} \left( \sum_{i_1} S_{i_1, \ell_1}^4 \right) (\mathcal{C}_R[a_{\ell_1}])^2 \\ &\leq \sum_{\ell_1} (\mathcal{C}_R[a_{\ell_1}])^2 = \mathcal{J}_R(\mathbf{a}). \end{aligned} \quad (10)$$

On montre facilement maintenant que l'égalité dans (10) n'a lieu que si et seulement si  $\mathbf{S}$  satisfait (5). De plus  $\mathcal{J}_R(\mathbf{y})$  est invariante par toute multiplication de  $y_i, i = 1, \dots, N$ , par une constante de module 1. Donc  $\mathcal{J}_R(\mathbf{y})$  est une fonction de contraste.  $\square$

### 4 Diagonalisation conjointe

Un des principaux intérêts du contraste  $\mathcal{J}(\mathbf{y})$  dans (6) est son lien avec un critère de diagonalisation conjointe d'un ensemble de matrices. Un tel critère de diagonalisation conjointe est défini par

**Définition 2.** *Considérons un ensemble de  $M$  matrices carrées  $(N, N)$   $\mathbf{M}(m), m = 1, \dots, M$ , noté  $\mathcal{M}$ . Un opérateur de diagonalisation conjointe de cet ensemble est une matrice unitaire qui maximise la fonction*

$$\mathcal{D}(\mathbf{H}, \mathcal{M}) = \sum_{m=1}^M \left( \sum_{i=1}^N |M_{i,i}^H(m)|^2 \right) \quad (11)$$

où

$$M_{i,i}^H(m) = \sum_{n_1, n_2} H_{i, n_1} H_{i, n_2} M_{n_1, n_2}(m). \quad (12)$$

Il s'agit donc de déterminer une matrice unitaire maximisant la somme des carrés des éléments diagonaux de chacune des matrices de l'ensemble  $\mathcal{M}$  après changement de base. Il est facile de montrer que cela est équivalent à la minimisation de la somme des carrés des éléments hors diagonal, ce qui correspond alors à une "mesure naturelle" de l'écart par rapport à une matrice diagonale.

Le lien entre le contraste  $\mathcal{J}_R(\mathbf{y})$  dans (7) et un critère de diagonalisation conjointe est donné par la proposition suivante.

**Proposition 2.** Soit  $\mathcal{C}_R$  l'ensemble des  $M = N^{R-2}$  matrices  $\mathbf{C}(i_3, \dots, i_R) = (C_{i_1, i_2}(i_3, \dots, i_R))$  avec  $R \geq 3$  définies par

$$C_{i_1, i_2}(i_3, \dots, i_R) = \text{Cum}[x_{i_1}, x_{i_2}, x_{i_3}, \dots, x_{i_R}] \quad (13)$$

alors si  $\mathbf{H}$  est une matrice unitaire, on a

$$\mathcal{D}(\mathbf{H}, \mathcal{C}_R) = \mathcal{J}_R(\mathbf{H}\mathbf{x}) . \quad (14)$$

Cette proposition constitue une généralisation d'un résultat de [1] pour des cumulants d'ordre supérieur ou égal à trois.

Le nombre de matrices à diagonaliser croît avec l'ordre des cumulants considérés. Ainsi pour des cumulants d'ordre trois il y a  $N$  matrices tandis que pour des cumulants d'ordre quatre il y en a  $N^2$ . On peut toutefois montrer [8] que ce nombre de matrices peut toujours être ramené à  $N$  au moyen de la décomposition d'un certain opérateur linéaire.

Il est intéressant de remarquer que la diagonalisation conjointe d'un ensemble de matrices de cumulants d'ordres différents peut aussi être considérée. En rappelant que la somme de deux contrastes est un contraste, on a la proposition suivante.

**Proposition 3.** Soient  $\mathcal{C}_R$  et  $\mathcal{C}_S$ ,  $S \neq R$ , deux ensembles de matrices de cumulants d'ordre  $R$  et  $S$  respectivement et définis comme à la proposition 2, nous avons alors

$$\mathcal{D}(\mathbf{H}, \mathcal{C}_R \cup \mathcal{C}_S) = \mathcal{J}_R(\mathbf{H}\mathbf{x}) + \mathcal{J}_S(\mathbf{H}\mathbf{x}) . \quad (15)$$

## 5 Au sujet de l'algorithme

L'algorithme JADE est un algorithme de type Jacobi, *i.e.* la diagonalisation conjointe est réalisée par l'intermédiaire d'un produit de rotation plane (ou de Givens). Chacune des rotations planes travaille sur l'une des  $N' = N(N-1)/2$  paires du vecteur de sortie. On décompose ainsi un problème de dimension  $N$  en  $N'$  problèmes de dimension 2, lesquels sont censés être plus simples. C'est notamment le cas si l'on est capable d'obtenir la solution analytique exacte du problème de dimension 2. A cette fin on paramétrise la matrice de séparation ainsi :

$$\mathbf{H} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} . \quad (16)$$

Considérons maintenant la maximisation de  $\mathcal{J}_R(\mathbf{y})$ . Pour  $N = 2$ , on obtient la réécriture suivante du contraste

$$\mathcal{J}_R(\mathbf{y}) = \mathbf{u}_\theta^T \mathbf{A} \mathbf{u}_\theta \quad (17)$$

où  $\mathbf{u}_\theta^T = (\cos 2\theta \quad \sin 2\theta)$  et où  $\mathbf{A}$  est une matrice (2, 2) réelle et symétrique définie ainsi

$$A_{1,1} = t_1 ; \quad A_{1,2} = t_4 ; \quad A_{2,2} = \frac{1}{2}t_1 + 2t_2 + t_3 \quad (18)$$

avec, grâce à (13) et en définissant  $\mathbf{i} = (i_3, \dots, i_R)$ ,

$$\begin{aligned} t_1 &= \sum_{\mathbf{i}} (C_{1,1}(\mathbf{i}))^2 + (C_{2,2}(\mathbf{i}))^2 ; \\ t_2 &= \sum_{\mathbf{i}} (C_{1,2}(\mathbf{i}))^2 ; \\ t_3 &= \sum_{\mathbf{i}} C_{1,1}(\mathbf{i})C_{2,2}(\mathbf{i}) ; \\ t_4 &= \sum_{\mathbf{i}} C_{1,2}(\mathbf{i}) (C_{1,1}(\mathbf{i}) - C_{2,2}(\mathbf{i})) . \end{aligned}$$

On peut aussi écrire

$$\mathbf{u}_\theta^T \mathbf{A} \mathbf{u}_\theta = C + D \cos(4(\theta - \alpha)) \quad (19)$$

où  $C = (A_{1,1} + A_{2,2})/2$  est un terme constant puisque ne dépendant pas de  $\theta$  et

$$D = \sqrt{((A_{1,1} - A_{2,2})/2)^2 + A_{1,2}^2}$$

est positif. L'angle  $\alpha$  peut être déterminé par

$$\alpha = \frac{1}{4} \arctan \left( A_{1,2} , \frac{1}{2}(A_{1,1} - A_{2,2}) \right) \quad (20)$$

où  $\arctan(y, x)$  est, par définition, l'unique angle  $\beta$  dans  $(-\pi, \pi]$  tel que  $\cos \beta = \frac{x}{(x^2+y^2)^{1/2}}$  et  $\sin \beta = \frac{y}{(x^2+y^2)^{1/2}}$ . La valeur de  $\theta$  notée  $\theta_{\text{opt}}$  maximisant  $\mathbf{u}_\theta^T \mathbf{A} \mathbf{u}_\theta$  dans (19) est alors aisément calculable et l'on obtient

$$\theta_{\text{opt}} = \alpha . \quad (21)$$

## 6 Simulations

Afin d'illustrer les résultats précédents, nous présentons quelques simulations informatiques dans le cas de deux sources ( $N = 2$ ) et avec le mélange

$$\mathbf{G} = \begin{pmatrix} 1 & 0.9 \\ 0.8 & 1 \end{pmatrix} . \quad (22)$$

Nous allons considérer l'algorithme JADE et l'algorithme étendu noté eJADE diagonalisant à la fois des matrices de cumulants d'ordre trois et quatre. Les performances des algorithmes sont illustrées par un indice défini grâce à la matrice globale  $\mathbf{S}$  :

$$\begin{aligned} \text{ind}_\alpha(\mathbf{S}) &\triangleq \frac{1}{2} \left[ \sum_i \left( \sum_j \frac{|s_{ij}|^\alpha}{\max_\ell |s_{i\ell}|^\alpha} - 1 \right) \right. \\ &\quad \left. + \sum_j \left( \sum_i \frac{|s_{ij}|^\alpha}{\max_\ell |s_{\ell j}|^\alpha} - 1 \right) \right] \quad (23) \end{aligned}$$

où  $\alpha \geq 1$ . Cet indice positif est en effet nul si  $\mathbf{S}$  satisfait la relation (5). Nous considérons uniquement l'indice  $\text{ind}_2(\cdot)$ . Pour notre comparaison nous utilisons un signal i.i.d. noté PS prenant ses valeurs dans

$$\{-1, 0, \alpha\}$$

avec les probabilités respectives

$$\left\{ \frac{1}{1+\alpha}, \frac{\alpha-1}{\alpha}, \frac{1}{\alpha(1+\alpha)} \right\} .$$

Le paramètre réel  $\alpha$  est tel que  $\alpha \geq 1$ . On a alors pour un signal PS noté  $a$ ,  $\mathbb{E}[a] = 0$ ,  $\mathbb{E}[a^2] = 1$  et

$$C_3[a] = \alpha - 1, \quad C_4[a] = \alpha^2 - \alpha - 2 . \quad (24)$$

Voir Fig. 1 pour les courbes de  $C_3[a]$  et  $C_4[a]$  en fonction de  $\alpha$ . Pour chacune des figures on affiche à la fois la valeur moyenne et l'écart type de l'indice estimé sur 500 réalisations indépendantes.

On considère deux cas pour les sources  
(i) Les deux sources sont des signaux PS ;

(ii) La première source  $a_1$  est un signal PS tandis que la seconde  $a_2$  est un bruit blanc Gaussien.

Sur les deux figures 2 et 3, la moyenne et l'écart type de l'indice sont affichés en fonction de  $\alpha$ , pour respectivement les cas (i) et (ii). Le nombre de données pour évaluer les statistiques est  $N_d = 1000$  et il n'y a pas de bruit. Les figures montrent que les performances de eJADE fluctuent moins en fonction des statistiques des sources. Le gain de performance se situe principalement autour de  $\alpha = 2$ , *i.e.* lorsque le cumulante d'ordre 4 des sources est faible. Il faut aussi noté que lorsque le cumulante d'ordre 3 des sources est petit, *i.e.*  $\alpha$  proche de 1, les performances de eJADE ne sont pas dégradées par rapport à celles de JADE dans la mesure où les statistiques d'ordre trois n'apportent guère d'information.

## Références

- [1] J.-F. Cardoso and A. Souloumiac, "Blind Beamforming for non Gaussian Signals", *IEE Proceedings-F*, Vol. 40, pp 362-370, 1993.
- [2] P. Comon, "Independent Component Analysis, a New Concept?", *Signal Processing*, Vol. 36, pp 287-314, 1994.
- [3] L. De Lathauwer, B. De Moor and J. Vandewalle, "Blind source separation by simultaneous third-order tensor diagonalization", *European Signal Processing Conference (EUSIPCO'96)*, Trieste, Italy, pp 2089-2092, Sept. 1996.
- [4] C. Jutten and J. Herault, "Blind Separation of Sources, Part I: An Adaptative Algorithm Based on Neuromimetic Architecture", *Signal Processing*, Vol. 24, pp 1-10, 1991.
- [5] E. Moreau and O. Macchi, "High Order Contrasts for Self-Adaptive Source Separation", *International Journal of Adaptive Control and Signal Processing*, Vol. 10, No. 1, pp 19-46, Jan. 1996.
- [6] E. Moreau and J.-C. Pesquet, "Independence / Decorrelation Measures with Applications to Optimized Orthonormal Representations", in *Proc. ICASSP'97*, Munich, Germany, Vol. 5, pp 3425-3428, April 1997.
- [7] E. Moreau and N. Thirion-Moreau, "Non symmetrical contrasts for source separation", *IEEE Trans. Signal Processing*, accepted for publication, Feb. 1999.
- [8] E. Moreau, "A Generalization of Joint Diagonalization Criteria for Source Separation", submitted to *IEEE Trans. Signal Processing*, March 1999.

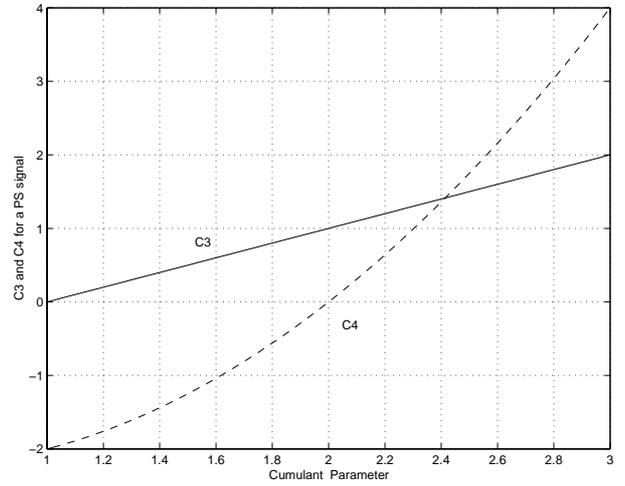


FIG. 1: Courbes des cumulants d'ordre trois et quatre pour un signal PS en fonction de  $\alpha$ .

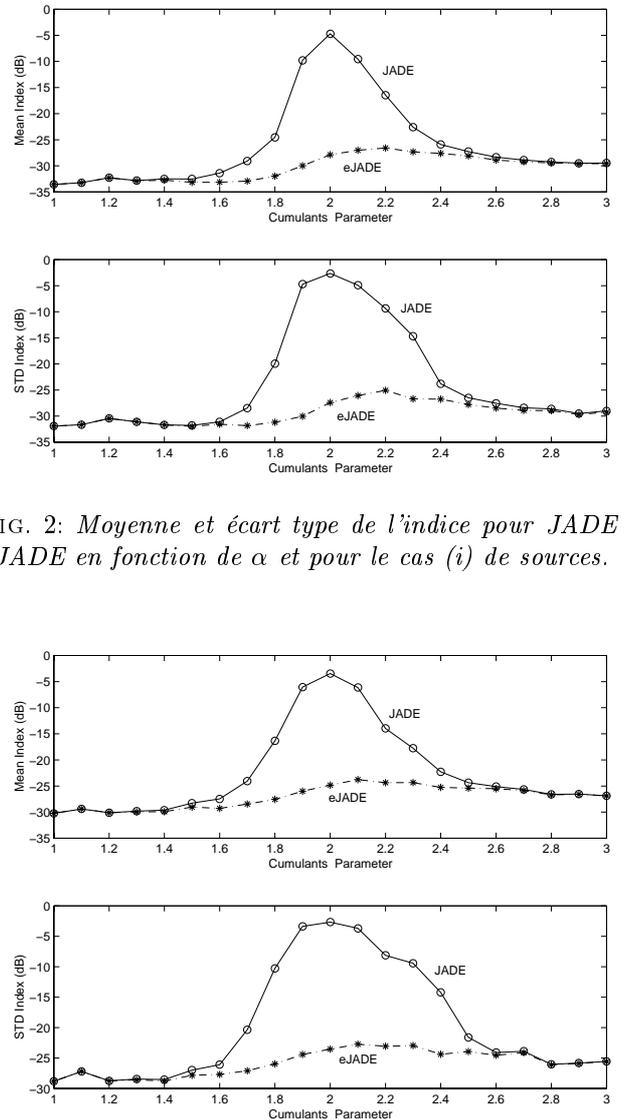


FIG. 2: Moyenne et écart type de l'indice pour JADE et eJADE en fonction de  $\alpha$  et pour le cas (i) de sources.

FIG. 3: Moyenne et écart type de l'indice pour JADE et eJADE en fonction de  $\alpha$  et pour le cas (ii) de sources.