



# SEPARATION AUTO-ADAPTATIVE DE SOURCES SANS BLANCHIMENT PREALABLE

Eric MOREAU et Odile MACCHI

Laboratoire des Signaux et Systèmes, CNRS-ESE, 91192 Gif sur Yvette,  
et Groupement de Recherche TdSI

## RÉSUMÉ

Les critères actuels de séparation de sources, basés sur une fonction de contraste, nécessitent une étape de blanchiment du mélange. Nous présentons une nouvelle fonction de contraste affranchie de cette étape. Un algorithme adaptatif du gradient en découle. A performances égales en régime permanent, il se compare favorablement à l'algorithme de Jutten-Hérault au niveau vitesse de convergence (tout en étant de complexité supérieure).

## ABSTRACT

In order to perform source separation, the usual approach is to use a contrast function, after a preprocessing has whitened the source mixture. Here a novel contrast function is defined that is free of the prewhitening step. An adaptive gradient algorithm can be associated to this contrast. For identical steady-state performances its convergence speed outperforms the Jutten-Herault algorithm (with a higher computational complexity).

## 1 Introduction

Le problème appelé *séparation de sources* est récent et permet d'entrevoir de nouveaux développements dans les domaines des communications et du traitement d'antenne. Il a pour but, de restituer un vecteur source  $\mathbf{a}$  dont les composantes  $a_i$  sont supposées statistiquement indépendantes, centrées et de cumulants d'ordre 4 non nul, à partir de la seule observation de leur mélange linéaire instantané  $\mathbf{x} = \mathbf{G}\mathbf{a}$  où  $\mathbf{a}$  et  $\mathbf{x}$  sont des vecteurs de dimension  $n$  et  $\mathbf{G}$  une matrice carrée inversible à coefficients constants. Ce problème est souvent qualifié d'*autodidacte* dans la mesure où seule l'information apportée par l'observation de  $\mathbf{x}$  est utilisée pour le traitement.

Il s'agit donc de la recherche d'un système linéaire :

$$\mathbf{y} = \mathbf{H}\mathbf{x} \quad (1)$$

où la matrice  $\mathbf{H}$  est telle que  $\mathbf{y}$  ait des composantes indépendantes. Une telle matrice  $\mathbf{H}$  sera dite séparante. Le problème ainsi posé n'admet pas une solution unique. En effet  $\mathbf{H}$  est déterminée au produit près d'une matrice  $\mathbf{AP}$ , où  $\mathbf{A}$  est une matrice diagonale inversible et  $\mathbf{P}$  une matrice de permutation. Ainsi on doit estimer une matrice  $\mathbf{H}$  telle que

$$\mathbf{S} \triangleq \mathbf{H}\mathbf{G} = \mathbf{A}\mathbf{P}; \quad (2)$$

$\mathbf{A}$  et  $\mathbf{P}$  correspondent respectivement à une indétermination sur la puissance et sur l'ordre des sources restituées. L'indétermination sur  $\mathbf{A}$  entraîne qu'il y a  $n$  indéterminées parmi les  $n^2$  inconnues de  $\mathbf{H}$ . Ainsi, pour avoir un problème bien posé, il est nécessaire d'imposer  $n$  conditions supplémentaires.

Une première approche introduit ces conditions dans la matrice  $\mathbf{H}$ . Dans la contribution originale de Jutten et

Hérault [1], ceci est fait en écrivant  $\mathbf{H}$  sous une forme inverse  $\mathbf{H} = \mathbf{F}^{-1}$ , avec la contrainte que  $f_{ii} = 1$ . On peut aussi écrire directement

$$h_{ii} = 1, \quad (3)$$

comme dans [2]. Une deuxième approche procède en deux étapes grâce à un blanchiment préalable [3]-[7], ce qui normalise les puissances des sorties  $y_i$  et par là fixe les  $n$  conditions.

Notre travail s'inscrit dans la première approche : nous écrivons directement la contrainte (3), ce qui évite la cascade prétraitement/séparation. Dans cette approche, il existe un algorithme adaptatif qui pilote les éléments  $h_{ij}$  de  $\mathbf{H}$ , avec pour incrément la loi très simple

$$\Delta h_{ij} = -\mu y_i^3 y_j, \quad i \neq j, \quad \mu > 0, \quad (4)$$

où  $\mu$  est le pas d'adaptation.

Dans le cas de deux sources, il a été prouvé [2] que les matrices séparantes  $\mathbf{H}$  sont les points d'équilibre stable de l'algorithme (4) sous réserve que les sources soient globalement sous-gaussiennes. Mais cela n'est pas prouvé pour  $n \geq 3$ . Ainsi, il se pourrait que (4) converge vers une matrice  $\mathbf{H}$  non séparante. C'est pourquoi notre objectif ici est de concevoir un autre algorithme adaptatif, dont le lien théorique avec les matrices séparantes soit clair.

## 2 Contraste et blancheur

Une fonction de contraste [3] est une application  $J(\cdot)$  dans  $\mathbf{R}$  d'un espace de vecteurs aléatoires  $\mathbf{y}$  de  $\mathbf{R}^n$ , dépendant seulement de la loi de probabilité de  $\mathbf{y}$  et telle que :



**P-1.**  $J(\mathbf{y})$  est symétrique par rapport aux composantes  $y_i$  de  $\mathbf{y}$  :

$$\forall \mathbf{P} \text{ matrice de permutation, } J(\mathbf{P}\mathbf{y}) = J(\mathbf{y}) ; \quad (5)$$

**P-2.**  $J(\mathbf{y})$  est invariante par changement d'échelle :

$$\forall \mathbf{A} \text{ matrice diagonale inversible, } J(\mathbf{A}\mathbf{y}) = J(\mathbf{y}) ; \quad (6)$$

**P-3.** Le mélange linéaire de composantes indépendantes ne peut que dégrader le contraste :

$$\forall \mathbf{a} \text{ vecteur indépendant, } \forall \mathbf{S}, J(\mathbf{S}\mathbf{a}) \leq J(\mathbf{a}) ; \quad (7)$$

**P-4.** Seules les permutations et changements d'échelle conservent le contraste de sources indépendantes :

$$\forall \mathbf{a} \text{ vecteur indépendant, } J(\mathbf{S}\mathbf{a}) = J(\mathbf{a}) \Leftrightarrow \mathbf{S} = \mathbf{A}\mathbf{P} . \quad (8)$$

Ainsi pour un mélange  $\mathbf{x}$  de sources indépendantes  $a_i$ , le contraste  $J(\mathbf{y})$  du vecteur  $\mathbf{y} = \mathbf{H}\mathbf{x}$  est maximum ssi  $\mathbf{H}$  est une matrice séparante au sens de l'équation (2). Ainsi, pour séparer le mélange  $\mathbf{x}$ , il faut et il suffit de chercher une matrice pour laquelle le contraste  $J(\mathbf{y})$  est maximum. Le premier exemple de contraste fut donné par [3] à partir de statistiques d'ordre 4 de  $\mathbf{y}$ , en travaillant sur des vecteurs "blancs" pour lesquels  $\forall i, j \quad E[y_i y_j] = 1$  si  $i = j$  et 0 sinon. Le cumulants d'ordre 4 de variables aléatoires centrées étant défini par

$$\text{cum}(y_1 y_2 y_3 y_4) \triangleq E[y_1 y_2 y_3 y_4] - E[y_1 y_2]E[y_3 y_4] - E[y_1 y_3]E[y_2 y_4] - E[y_1 y_4]E[y_2 y_3] \quad (9)$$

le théorème suivant a été prouvé dans [7] :

**Théorème 1** Si tous les cumulants d'ordre 4 des sources ont le même signe :

$$\forall k, \text{sgn}[\text{cum}(a_k^4)] = \epsilon, \quad \epsilon \triangleq \pm 1, \quad (10)$$

et si  $\beta, \gamma \geq 0$ , alors la fonction

$$J_{\beta\gamma}(\mathbf{y}) = \sum_{i=1}^n |\text{cum}(y_i^4)| - \beta \sum_{i < j=1}^n |\text{cum}(y_i^2 y_j^2)| - \gamma \sum_{i \neq j=1}^n |\text{cum}(y_i y_j^3)| \quad (11)$$

est un contraste pour les vecteurs  $\mathbf{y}$  blancs.

Nous allons maintenant généraliser ce résultat.

### 3 Contraste et normalisation

Nous dirons qu'un vecteur  $\mathbf{y}$  est "normalisé" ssi toutes ses composantes sont de puissance égale à 1 :

$$E[y_i^2] = 1, \quad \forall i. \quad (12)$$

Naturellement un vecteur blanc est normalisé. Il est facile de voir, grâce à (2), que blanchir  $\mathbf{a}$  est équivalent à changer la matrice  $\mathbf{A}$ . Ainsi sans perte de généralité, on peut supposer  $\mathbf{a}$  blanc. Il s'ensuit qu'un vecteur  $\mathbf{y} = \mathbf{S}\mathbf{a}$  est normalisé ssi

$$\forall i, \sum_{k=1}^n s_{ik}^2 = 1. \quad (13)$$

Les théorèmes suivants généralisent le théorème 1.

**Théorème 2** Sous la condition (10), la fonction  $J_{\beta 0}(\mathbf{y})$  est un contraste pour les vecteurs  $\mathbf{y}$  normalisés, pourvu que  $\beta \geq 1$ .

**Démonstration :** Les propriétés P-1 et P-2 sont évidentes. Examinons P-3. D'après (9)-(11), on a

$$J_{\beta 0}(\mathbf{y}) = \sum_{k=1}^n f_n(s_{1k}^2, \dots, s_{nk}^2) |\text{cum}(a_k^4)|, \quad (14)$$

où

$$f_n(z_1, \dots, z_n) \triangleq \sum_{i=1}^n z_i^2 - \beta \sum_{i < j=1}^n z_i z_j. \quad (15)$$

La propriété P-3 est donc prouvée si l'on montre que

$$(z_1, \dots, z_n) \in [0, 1]^n \Rightarrow f_n(z_1, \dots, z_n) \leq 1, \quad (16)$$

la normalisation (13) assurant que  $(s_{1k}^2, \dots, s_{nk}^2) \in [0, 1]^n$ . Le cas  $n = 1$  est évident puisque  $f_1(z_1) = z_1^2$ . Supposons la propriété (16) vraie à l'ordre  $n - 1$ . On a

$$f_n(z_1, \dots, z_{n-1}, 0) = f_{n-1}(z_1, \dots, z_{n-1}) \leq 1, \quad (17)$$

$$f_n(z_1, \dots, z_{n-1}, 1) = 1 + \sum_{i=1}^{n-1} z_i^2 - \beta \sum_{i=1}^{n-1} z_i - \beta \sum_{i < j=1}^{n-1} z_i z_j. \quad (18)$$

Or  $z_i^2 \leq z_i$  et  $\beta \geq 1$  de sorte que

$$f_n(z_1, \dots, z_{n-1}, 1) \leq 1; \quad (19)$$

$f_n(z_1, \dots, z_n)$  est un trinôme en  $z_n$  avec un minimum; donc (17) et (19) terminent la preuve de (16).

D'après ce qui précède, l'égalité  $f_n(z_1, \dots, z_n) = 1$  ne peut donc être vraie que si  $\forall i, z_i = 0$  ou 1. On vérifie que

$$f_n(0, \dots, 0, 0) = 0. \quad (20)$$

Examinons enfin P-4. L'égalité  $J_{\beta 0}(\mathbf{y}) = J_{\beta 0}(\mathbf{a})$  exige que toutes les inégalités (16) deviennent des égalités. Donc  $\forall (i, k), s_{ik}^2 = 0$  ou 1. Il découle des contraintes de normalisation (13) sur les lignes de  $\mathbf{S}$  que chacune contient un et un seul élément  $s_{ik(i)}$  non nul, qui vaut  $\pm 1$ .

Pour prouver P-4, il reste à montrer que la matrice  $[s_{ik}^2]$  est une matrice de permutation. D'après (13), si  $\mathbf{S}$  avait deux éléments non nuls dans une même colonne, elle aurait au moins une colonne de zéros, soit la colonne  $\ell$ . Alors  $f_n(s_{1\ell}^2, \dots, s_{n\ell}^2) = 0$ . Il découle de (14) et de  $\text{cum}(a_\ell^4) \neq 0$  que  $J_{\beta 0}(\mathbf{y}) < J_{\beta 0}(\mathbf{a})$ . Cela termine la preuve de P-4. Ainsi  $J_{\beta 0}(\mathbf{y})$  est bien un contraste et le maximum est atteint ssi

$$\mathbf{S} = \text{diag}(\epsilon_1, \dots, \epsilon_n) \mathbf{P}, \quad (21)$$

où  $\mathbf{P}$  est une matrice de permutation et  $\epsilon_i = \pm 1$ .  $\diamond$

**Théorème 3** Sous la condition (10) et si  $\beta \geq 1, \gamma \geq 0$ , la fonction  $J_{\beta\gamma}(\mathbf{y})$  est un contraste pour les  $\mathbf{y}$  normalisés.

**Démonstration :** Les propriétés P-1 et P-2 sont évidentes. P-3 provient des relations :

$$J_{\beta\gamma}(\mathbf{y}) \leq J_{\beta 0}(\mathbf{y}) \quad (22)$$

$$J_{\beta 0}(\mathbf{y}) \leq J_{\beta 0}(\mathbf{a}) \quad (23)$$

$$J_{\beta 0}(\mathbf{a}) = J_{\beta\gamma}(\mathbf{a}) \quad (24)$$

L'inégalité (23) est vraie car  $J_{\beta 0}(\mathbf{y})$  est un contraste; (24) est vraie car les sources sont indépendantes, donc leurs cumulants croisés sont nuls. Comme  $J_{\beta 0}(\mathbf{y})$  est un contraste, l'inégalité (23) devient une égalité ssi  $\mathbf{S}$  satisfait (21). Dans ce cas (22) devient elle-même une égalité. De la sorte  $J_{\beta\gamma}(\mathbf{y})$  est bien un contraste pour  $\mathbf{y}$  normalisé.  $\diamond$



Nous allons maintenant éliminer l'hypothèse de normalisation et ainsi disposer d'un contraste pour des vecteurs  $\mathbf{y}$  quelconques.

**Théorème 4** Sous la condition (10), la fonction

$$I'_{\beta\gamma}(\mathbf{y}) = \sum_{i=1}^n \frac{|\text{cum}(y_i^4)|}{E^2[y_i^2]} - \beta \sum_{i<j=1}^n \frac{|\text{cum}(y_i^2 y_j^2)|}{E[y_i^2]E[y_j^2]} - \gamma \sum_{i \neq j=1}^n \frac{|\text{cum}(y_i y_j^3)|}{E^{1/2}[y_i^2]E^{3/2}[y_j^2]} \quad (25)$$

est un contraste pour tout vecteur  $\mathbf{y}$ , si  $\beta \geq 1$ ,  $\gamma \geq 0$ .

**Démonstration :** Cela est immédiat en prenant comme vecteur  $\tilde{\mathbf{y}}$  normalisé de la fonction  $J_{\beta\gamma}(\cdot)$ , le vecteur  $\tilde{\mathbf{y}} = \mathbf{W}\mathbf{y}$  où  $\mathbf{W}$  est une matrice diagonale dont les éléments sont les racines carrées des puissances respectives de chacune des composantes du vecteur  $\mathbf{y}$  maintenant quelconque.  $\diamond$

**Remarque :** on peut réécrire  $I'_{\beta\gamma}(\mathbf{y})$  en explicitant les cumulants. On a

$$|\text{cum}(y_i^4)| = \epsilon \text{cum}(y_i^4), \quad |\text{cum}(y_i^2 y_j^2)| = \epsilon \text{cum}(y_i^2 y_j^2). \quad (26)$$

Au contraire le signe de  $\text{cum}(y_i y_j^3)$  varie avec le mélange inconnu. Nous prendrons ainsi  $\gamma = 0$  dans la suite

$$I'_{\beta 0}(\mathbf{y}) = \epsilon \left\{ \sum_{i=1}^n \frac{\text{cum}(y_i^4)}{E^2[y_i^2]} - \beta \sum_{i<j=1}^n \frac{\text{cum}(y_i^2 y_j^2)}{E[y_i^2]E[y_j^2]} \right\}. \quad (27)$$

Comme

$$\text{cum}(y_i^4) = E[y_i^4] - 3E^2[y_i^2], \quad (28)$$

$$\text{cum}(y_i^2 y_j^2) = E[y_i^2 y_j^2] - 2E^2[y_i y_j] - E[y_i^2]E[y_j^2], \quad (29)$$

en négligeant les constantes, on obtient le contraste

$$I_{\beta}(\mathbf{y}) = \epsilon \left\{ \sum_{i=1}^n \frac{E[y_i^4]}{E^2[y_i^2]} - \beta \sum_{i<j=1}^n \frac{E[y_i^2 y_j^2] - 2E^2[y_i y_j]}{E[y_i^2]E[y_j^2]} \right\}. \quad (30)$$

## 4 Algorithme adaptatif

Pour séparer les sources, il faut estimer une matrice  $\mathbf{H}$  maximisant l'expression (30). L'indétermination sur les puissances des sources est levée par la contrainte (3). Chaque élément  $h$  de  $\mathbf{H}$  peut être optimisé par l'algorithme du gradient déterministe en temps discret

$$h(k) = h(k-1) + \frac{\mu}{2} \cdot \frac{\partial I_{\beta}}{\partial h} \Big|_{H=H(k-1)}, \quad \mu > 0, \quad (31)$$

évalué avec la valeur  $\mathbf{H}(k-1)$  pour  $\mathbf{H}$ . Or

$$\frac{\mu}{2} \cdot \frac{\partial}{\partial h} \left[ \frac{E[y_i^4]}{E^2[y_i^2]} \right] = 2\mu \frac{E[\frac{\partial y_i}{\partial h} y_i^3]E[y_i^2] - E[\frac{\partial y_i}{\partial h} y_i]E[y_i^4]}{E^3[y_i^2]}, \quad (32)$$

avec des équations analogues pour les dérivées partielles des autres termes de (30). Une version stochastique de (31) est

$$h(k) = h(k-1) + \mu \epsilon \left\{ 2 \sum_{i=1}^n \frac{\partial y_i}{\partial h} \frac{y_i^3 E[y_i^2] - E[y_i^4]}{E^3[y_i^2]} - \beta \sum_{i<j=1}^n \dots \right\}, \quad (33)$$

où l'on n'a pas écrit tous les termes pour abrégé. Or

$$\frac{\partial y_i}{\partial h_{\ell m}} = \delta_{i\ell} x_m, \quad m \neq i. \quad (34)$$

Les espérances intervenant dans (33), sont estimées par un simple moyennage à horizon glissant, selon

$$\hat{E}[y_i^{\alpha_1} y_j^{\alpha_2}](k) = (1-\delta) \hat{E}[y_i^{\alpha_1} y_j^{\alpha_2}](k-1) + \delta y_i^{\alpha_1} y_j^{\alpha_2}(k-1), \quad (35)$$

avec  $(\alpha_1, \alpha_2) = (2, 0), (1, 1), (4, 0), (2, 2)$ ,  $\delta > 0$ .

## 5 Simulations

Dans les simulations les  $a_i$  sont des signaux i.i.d. centrés prenant les deux valeurs  $\pm 1$  avec probabilité 1/2. Leur cumulants d'ordre 4 est négatif ( $\epsilon = -1$ ).

La figure 1 montre l'évolution au cours du temps ( $k$ ) de la sortie  $y_1$ , dans le cas de trois sources et du mélange

$$\mathbf{G} = \begin{pmatrix} 1 & 0,3 & 0,3 \\ 0,3 & 1 & 0,3 \\ 0,3 & 0,3 & 1 \end{pmatrix} \quad (36)$$

avec  $\mathbf{H}(0) = \mathbf{I}$ ,  $\beta = 1$ ,  $\mu = 0,0002$  et  $\delta = 0,001$ . On voit que l'algorithme reconstitue, sur la sortie  $y_1$  les niveaux pris par  $a_1$  (à un facteur d'échelle près). Dans ce cas  $\mathbf{P} = \mathbf{I}$  et la matrice  $\mathbf{S}$  doit converger vers une matrice diagonale. On peut donc [5] considérer comme mesure de qualité la quantité :

$$\eta(k) = \sum_{i \neq j} s_{ij}^2(k). \quad (37)$$

La figure 2 montre la convergence de  $\eta(k)$  vers 0. Ainsi la séparation est bien réalisée.

Effectuons maintenant une comparaison du nouvel algorithme (33) avec l'algorithme (4) de Jutten, dans le cas de deux sources, pour

$$\mathbf{G} = \begin{pmatrix} 1 & 0,6 \\ 0,3 & 1 \end{pmatrix} \quad (38)$$

La figure 3 représente l'évolution de la moyenne de la mesure de qualité  $\eta(k)$  pour  $k \in [2000, 3000]$ , en fonction du pas d'adaptation  $\mu$ , pour  $\delta = 0,001$  et  $\beta = 1$ . La partie décroissante de chaque courbe correspond à des valeurs de  $\mu$  trop petites pour assurer que le transitoire est achevé avant l'itération  $k = 2000$ . La partie utile est donc la partie croissante. Le minimum de chacune des courbes correspond ainsi approximativement à une période transitoire de 2000 itérations. Cette figure montre que

- pour une même durée du transitoire, l'algorithme (33) a de meilleures performances que (4). Ceci est confirmé visuellement sur la figure 4 qui montre qu'au-delà de  $k = 2000$ , l'algorithme (33) a une plus faible erreur que (4) (voir la grandeur  $e_1(k) = y_1(k) - a_1(k)$ );
- pour un niveau de performances identique en régime permanent, l'algorithme (33) converge plus vite que (4).

## 6 Conclusion

Les critères de séparation de sources existants, basés sur une fonction de contraste, nécessitent une étape préalable de blanchiment. Nous avons présenté ici un nouveau contraste qui est affranchi de cette étape. Nous en déduisons un algorithme adaptatif du gradient stochastique qui met conjointement en œuvre l'estimation de moments d'ordre 2 et 4. Les simulations confirment la théorie et montrent que, dans le cas de deux sources, l'algorithme présenté a de meilleures performances que l'algorithme de Jutten, avec cependant une complexité plus importante. Il est à



noter qu'il n'est pas prouvé l'existence ou la non existence de maxima locaux du contraste pouvant faire converger l'algorithme adaptatif vers une solution non séparante. Cependant aucune des simulations effectuées n'a jamais rencontré ce problème, dont l'étude théorique semble très difficile.

## References

- [1] J. Herault, C. Jutten et B. Ans, "Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé", *10ième colloque GRETSI*, Nice, France, 20-24 Mai 1985, pp 1017-1022.
- [2] E. Moreau and O. Macchi, "Two Novel Architectures for the Self-Adaptive Separation of Signals", *ICC'93, IEEE International Conference on Communication*, Geneva, Switzerland, 23-26 May, 1993, II pp 1154-1159.
- [3] P. Comon, "Blind Identification in Presence of Noise", *EUSIPCO'92, European Signal Processing Conference*, Brussels, Belgium, August 24-27, 1992, pp 835-838.
- [4] P. Comon, "Séparation de mélanges de signaux", *12ième colloque GRETSI*, Juan les Pins, France, 12-16 Juin 1989, pp 137-140.
- [5] J.F. Cardoso and B. Laheld, "Adaptive Blind Source Separation for Channel Spatial Equalization", *Proc of the 2nd COST 299 workshop on adaptive algorithms in communications*, Bordeaux, 1992.
- [6] N. Delfosse et P. Loubaton, "Séparation adaptative de sources indépendantes par une approche de déflation", *14ième colloque GRETSI*, Juan les Pins, France, 13-16 Septembre 1993.
- [7] E. Moreau and O. Macchi, "New Self-adaptive Algorithms for Source Separation Based on Contrast Functions", *HOS'93, IEEE Signal Processing Workshop on Higher-Order Statistics*, South Lake Tahoe, California, USA, June 7-9, 1993, pp 215-219.

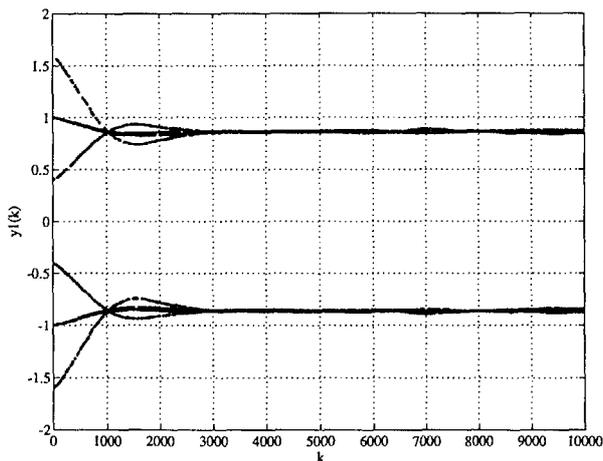


Figure 1: Reconstitution sur  $y_1$  des niveaux pris par  $a_1$ ,  $n = 3$ ,  $\beta = 1$

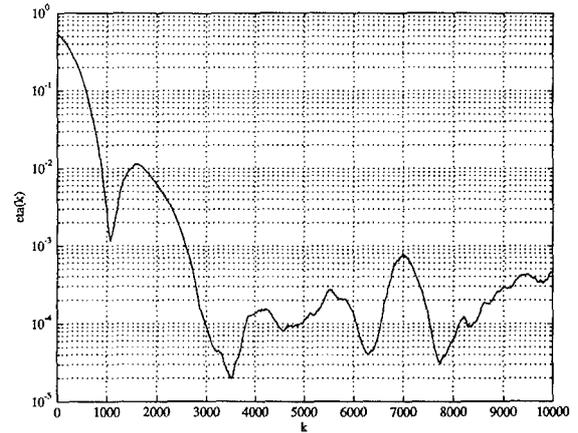


Figure 2: Mesure de qualité de la séparation de sources,  $n = 3$ ,  $\beta = 1$

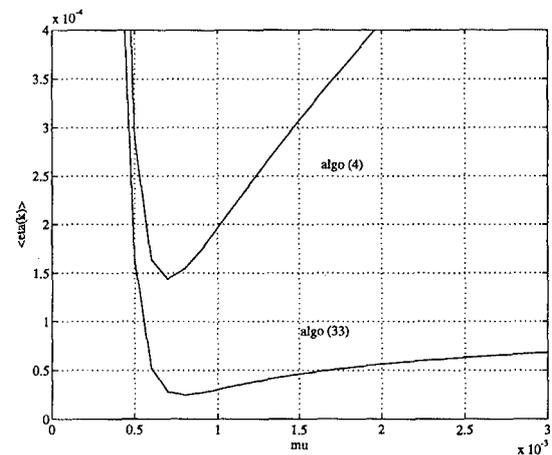


Figure 3: Comparaison des performances,  $n = 2$ ,  $\beta = 1$

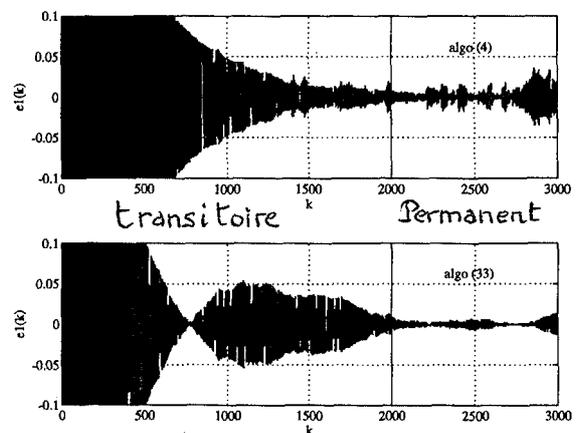


Figure 4: Comparaison des régimes permanents à transitoire d'égale durée,  $n = 2$ ,  $\beta = 1$