

# Séparation de sources dans des Mélanges Post-Nonlinéaires

Anisse Taleb et Christian Jutten

TIRF-INPG. 46, Avenue Félix Viallet - 38000 Grenoble, France

## RÉSUMÉ

Cet article propose une nouvelle approche pour la séparation de mélanges post-nonlinéaires. Ces mélanges possèdent une propriété intéressante de séparabilité, et les solutions admettent presque les mêmes indéterminations que les mélanges linéaires. Cette nouvelle approche, basée sur la minimisation de l'information mutuelle, utilise un nouvel algorithme pour l'estimation des fonctions "score". Quelques résultats expérimentaux sont présentés à la fin de l'article.

## ABSTRACT

This paper proposes a new approach to the separation of postnonlinear mixtures. These mixtures have an interesting property of separability, and solutions have nearly the same indeterminacies than the linear case. The new approach, based on the minimum mutual information principle, implements a new algorithm for the estimation of output score functions. Some experimental results are provided at the end of the paper.

## 1 Introduction

Le problème de séparation de sources a été largement étudié au cours des dix dernières années, principalement dans le cas de mélanges linéaires instantanés, et plus récemment de mélanges linéaires convolutifs. En revanche, le cas de mélanges nonlinéaires n'a été que très peu abordé.

La raison de cette désaffection pour les mélanges nonlinéaires s'explique par une remarque simple. Soient deux variables aléatoires statistiquement indépendantes  $u$  et  $v$ , les variables aléatoires  $f(u)$  et  $g(v)$ , où  $f(\cdot)$  et  $g(\cdot)$  sont deux fonctions non-linéaires inversibles quelconques, sont aussi indépendantes. En conséquence, dans le cas général de mélanges non-linéaires, en se fondant sur la seule hypothèse d'indépendance des sources, on ne peut pas retrouver les sources mais seulement une fonction non-linéaire (inconnue) des sources : de telles distorsions sont en général inacceptables. Sans hypothèse supplémentaire sur les sources ou sur les mélanges, le problème a donc peu d'intérêt.

Quelques auteurs se sont penchés, malgré la difficulté, sur le problème de mélanges non-linéaires. Les contributions dans ce domaine sont rares. Citons le travail de Pajunen *et al.* [7], où les cartes de Kohonen ont été utilisées comme algorithme de séparation. Le travail de Deco [6], consistait en l'utilisation de structures conservant le volume, ceci en introduisant une paramétrisation particulière basée sur des concepts de géométrie symplectique. Récemment Yang *et al.* [13] ont proposé un algorithme pour des structures nonlinéaires faisant intervenir des nonlinéarités inter-canal, mais aucune étude n'a été faite sur les indéterminations liées à ce modèle.

Dans ce papier, nous nous intéressons à des mélanges nonlinéaires particuliers. Nous rappelons dans le paragraphe 2 les propriétés de ces mélanges. Dans la troisième partie, nous proposons un nouvel algorithme fondé sur la minimisation de l'information mutuelle pour résoudre ce problème. Des

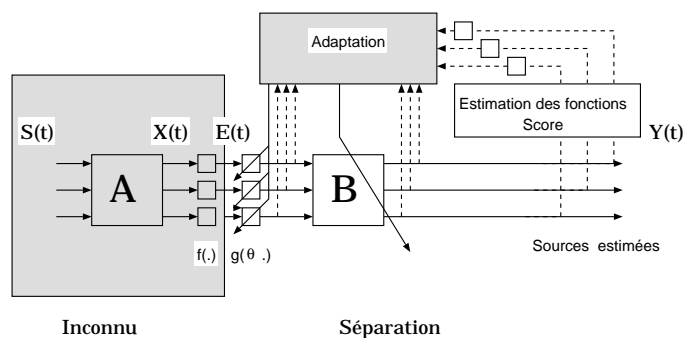


FIG. 1 — Schéma général de séparation.

résultats expérimentaux montrent l'intérêt de la méthode au paragraphe 4.

## 2 Mélanges post-nonlinéaires

On considère le mélange nonlinéaire particulier de  $n$  sources  $s_j(t)$  observées par  $n$  capteurs :

$$\begin{aligned} x_i(t) &= \sum_{j=1}^n a_{ij} s_j(t) \\ e_i(t) &= f_i(x_i(t)), \quad i = 1, \dots, n, \end{aligned} \quad (1)$$

où les  $f_i$  sont des fonctions nonlinéaires inversibles inconnues, les  $a_{ij}$  sont les coefficients réels d'une matrice de mélange instantanée  $\mathbf{A}$ , les  $s_j(t)$  sont les sources inconnues supposées statistiquement indépendantes. Dans la suite, ces mélanges  $e_i(t)$ , illustrés à la figure 1, seront appelés mélanges post-nonlinéaires (PNL). Le choix de ce modèle de mélange particulier repose sur un modèle physique assez réaliste, dans lequel le canal entraîne un mélange linéaire instantané, et les capteurs et leurs instrumentations (amplificateurs, etc.) sont responsables de la distorsion nonlinéaire.

Si l'on suppose un mélange PNL, la structure de séparation doit être constituée de deux étages (Figure 1) : le premier est

un ensemble de  $n$  blocs nonlinéaires, le second est une matrice de séparation  $\mathbf{B}$  pour mélanges linéaires instantanés. Les  $n$  blocs nonlinéaires fournissent des estimations paramétrique  $g_i(\theta_i, x)$ ,  $\theta_i$  étant le vecteur de paramètres. Chaque fonction  $g_i(\theta_i, x)$  doit compenser l'effet de la distorsion nonlinéaire due au modèle PNL, et ce en estimant l'inverse de  $f_i(x)$ .

Les mélanges PNL ont une propriété intéressante résumé par le lemme suivant [11] :

**Lemme 2.1** — *Soit un mélange PNL de  $n$  sources et  $n$  capteurs, si les composantes  $x_i$  sont non indépendantes, alors les sorties  $y_i$  de la structure séparante sont indépendantes deux à deux si et seulement si  $y_i = \alpha_i s_{\sigma(i)} + \beta_i$ , où  $\alpha_i$  et  $\beta_i$  sont des constantes réelles et  $\sigma(i)$  est une permutation sur  $\{1, \dots, n\}$ .*

Dans le cas où il existe au plus une source gaussienne, l'indépendance des  $x_i$  se traduit par une condition sur la matrice de mélange  $\mathbf{A}$ . En effet pour retrouver les sources sans distorsion, il faut que la matrice  $\mathbf{A}$  ait au moins 2 éléments non nuls par ligne ou par colonne. En pratique cette condition est souvent vérifiée. Ce lemme indique que, dans un mélange PNL, on peut donc retrouver les sources inconnues sans distorsion.

### 3 Estimation de la structure séparante

La structure séparante est ajustée de façon à ce que les signaux de sortie  $y_i(t)$ ,  $i = 1, \dots, n$ , deviennent statistiquement indépendants. L'indépendance statistique se traduit par  $\prod_i p_{Y_i}(y_i) = p_Y(\mathbf{y})$ . Comme proposé par plusieurs auteurs [5, 8, 14, 12], l'indépendance statistique vient en minimisant la divergence de Kullback-Liebler entre  $\prod_i p_{Y_i}(y_i)$  et  $p_Y(\mathbf{y})$ , ce qui est équivalent à minimiser l'information mutuelle entre les composantes  $y_i$  de  $\mathbf{y}$  :

$$I(\mathbf{y}) = \sum_{i=1}^n H(y_i) - H(\mathbf{y}) \quad (2)$$

La minimisation de (2) est délicate car l'entropie  $H(y_i)$  nécessite explicitement la connaissance des ddp  $p_{y_i}$ . Dans [11], nous avons proposé un premier critère dérivé d'une approche par maximum de vraisemblance, dans lequel nous approximations la log-vraisemblance des observations par un développement de Gram-Charlier des densités de probabilités (ddp) inconnues des sources. On obtient un critère constitué de deux termes : le premier est la fonction de contraste de Comon [5] pour les mélanges linéaires instantanés, le second est un terme supplémentaire lié à la nature nonlinéaire du mélange. En ajustant  $n$  réseaux de neurones (RN) de type perceptrons multi-couches (PMC) à une couche cachée pour estimer les fonctions  $g_i$  (par maximisation de ce critère) on obtient des résultats expérimentaux satisfaisants. Ceux-ci mettent cependant en évidence que l'approximation des ddp par le développement de Gram-Charlier nuit à la précision de la séparation dans le cas nonlinéaire. En fait, il s'avère très important d'avoir une bonne approximation de la ddp des sources pour pouvoir avoir une bonne estimation de l'inverse de la distorsions  $f_i(\cdot)$ .

Pour pallier à ce problème, on propose l'utilisation directe du critère (2). En tenant compte de la structure de séparation  $\mathbf{y} = \mathbf{B}[\mathbf{g}(\mathbf{e})]$ , l'information mutuelle du vecteur des sorties, s'écrit alors

$$I(\mathbf{y}) = \sum_{i=1}^n H(y_i) - H(\mathbf{e}) - \sum_{i=1}^n E[\log |g'_i(\theta_i, e_i)|] - \log |\det(\mathbf{B})| \quad (3)$$

#### Partie linéaire de la structure séparante

La dérivation de ce critère par rapport à la matrice de séparation  $\mathbf{B}$ , conduit directement à l'algorithme stochastique suivant [10] :

$$\mathbf{B}(t+1) = \mathbf{B}(t) + \mu_t \mathbf{K}(\mathbf{y}) \mathbf{B}^{-T}(t) \quad (4)$$

où  $\mu_t$  désigne un gain d'adaptation positif et  $\mathbf{K}(\mathbf{y})$  est une matrice  $n \times n$  qui ne dépend que des composantes de la sortie  $\mathbf{y}$  :

$$k_{ij}(\mathbf{y}) = \begin{cases} 0 & \text{si } i = j, \\ \psi_{Y_i}(y_i) y_j & \text{sinon.} \end{cases} \quad (5)$$

Dans (5) les fonctions  $\psi_{Y_i}(u) = \frac{d}{du} \log p_{Y_i}(u)$ , appelées fonctions "score", sont inconnues. Plusieurs auteurs, proposent une modification de ce type d'algorithmes, en introduisant la notion de *gradient relatif* [3], ou *naturel* [1]. Cette modification se traduit par une multiplication du gradient par le terme  $\mathbf{B}^T \mathbf{B}$ , le transformant en un algorithme dit *équivalent* :

$$\mathbf{B}(t+1) = (\mathbf{I} + \mu_t \mathbf{K}(\mathbf{y})) \mathbf{B}(t) \quad (6)$$

#### Partie nonlinéaire de la structure séparante

La dérivation du critère (3) par rapport à  $\theta_i$ , donne l'algorithme de gradient stochastique suivant :

$$\begin{aligned} \theta_i(t+1) &= \theta_i(t) + \lambda_t \Delta_i(t) \\ \Delta_i(t) &= \nabla_{\theta_i} \log |g'_i(\theta_i, e_i)| + \nabla_{\theta_i} g_i(\theta_i, e_i) \sum_{j=1}^n \psi_{Y_j}(y_j) b_{ji} \end{aligned} \quad (7)$$

Cet algorithme du gradient peut être amélioré par une procédure d'optimisation du second ordre. Cette procédure consiste à choisir une matrice symétrique définie positive  $\mathbf{G}_i$  et à adapter les paramètres de la manière suivante :

$$\theta_i(t+1) = \theta_i(t) + \lambda_t \mathbf{G}_i \Delta_i(t)$$

Pour les algorithmes du type Newton, la matrice  $\mathbf{G}_i$  correspond à l'inverse du hessien, qui est assez lourd à calculer. En se fondant sur le principe du gradient naturel d'Amari [1], on montre que le gradient naturel est obtenu avec  $\mathbf{G}_i$  égale à :

$$E \left[ \nabla_{\theta_i} \log |g'_i(\theta_i, e_i)| \nabla_{\theta_i}^T \log |g'_i(\theta_i, e_i)| \right]^{-1}$$

Les algorithmes (4) et (7) ne tiennent pas compte des indéterminations liées au modèle nonlinéaire. Plusieurs méthodes peuvent être utilisées pour pallier à ce problème. Les méthodes qui nous paraissent les plus efficaces consistent à introduire des termes de pénalisation pour que les sorties du bloc nonlinéaire soient centrées et normées, ainsi que les sorties du bloc linéaire. Par exemple, pour la partie linéaire, il suffit de mettre les termes diagonaux de  $\mathbf{K}(\mathbf{y})$  (nuls dans (5)) à :

$$k_{ii}(\mathbf{y}) = 1 - y_i^2.$$

Ceci a pour effet d'imposer aux signaux de sortie une variance égale à 1.

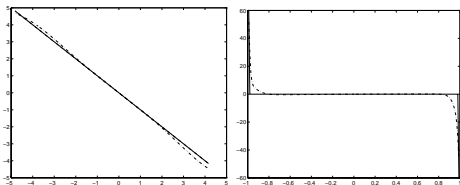


FIG. 2 — Estimation des fonctions score : à gauche gaussienne, à droite uniforme, (—) théorique, (- -) estimée

### Estimation de $\psi_{Y_i}$ :

Dans les deux algorithmes (5) et (7), les fonctions  $\psi_{Y_i}$ , provenant naturellement de la dérivation du logarithme des ddp marginales, interviennent explicitement. On retrouve ici un résultat montré par plusieurs auteurs. Récemment, Cardoso *et al.* [3] ont montré que, pour les algorithmes de la famille EASI, les fonction nonlinéaires réalisant le taux de rejection minimal sont de la forme  $\frac{p'_{s_i}}{p_{s_i}}$ . Bien avant, Pham *et al* [9] avaient déjà montré, dans le cas linéaire et par une approche du maximum de vraisemblance, que les fonctions nonlinéaires optimales pour la séparation de sources étaient de la même forme. Charkani *et al.* [4] ont aussi étendu ces résultats dans le cas de mélanges convolutifs. Dans ces deux dernières contributions, les auteurs utilisent une approche d'estimation paramétrique linéaire, en projetant  $\frac{p'_{s_i}}{p_{s_i}}$  sur un sous-espace de fonctions nonlinéaires.

Dans ce papier, nous proposons d'estimer directement les fonctions  $\psi_{Y_i}$ , fonctions nonlinéaires inconnues, par un modèle nonlinéaire paramétrique. Nous avons choisi d'utiliser des RN de type PMC. Chaque réseau fournit une estimation  $h_i(\mathbf{w}_i, u)$  de la fonction  $\psi_{Y_i}(u)$ . Le vecteur de paramètres  $\mathbf{w}_i$  est ajusté de façon à minimiser l'erreur quadratique moyenne :

$$\mathcal{E}_i = E[(h_i(\mathbf{w}_i, u) - \psi_{Y_i}(u))^2]. \quad (8)$$

En utilisant un algorithme de type gradient on a l'algorithme :

$$\mathbf{w}_i(t+1) = \mathbf{w}_i(t) - \mu_t \nabla_{\mathbf{w}_i} \mathcal{E}_i, \quad (9)$$

dans lequel le gradient de  $\mathcal{E}_i$  s'écrit :

$$\nabla_{\mathbf{w}_i} \mathcal{E}_i = 2E[h_i(\mathbf{w}_i, u) \nabla_{\mathbf{w}_i} h_i(\mathbf{w}_i, u) + \nabla_{\mathbf{w}_i} \frac{\partial h_i(\mathbf{w}_i, u)}{\partial u}]. \quad (10)$$

La dernière équation montre, de façon surprenante, que l'algorithme supervisé<sup>1</sup> écrit en (9) ne nécessite pas la connaissance *a priori* de la fonction désirée  $\psi_{Y_i}(u)$ . Cette méthode d'estimation est plus générale que celles proposées dans [9, 4], car elle s'affranchit du cadre restrictif de modèle paramétrique linéaire et du choix *a priori* de la base de projection.

Quelques résultats expérimentaux sur cette méthode d'estimation des fonctions "score" sont présentés sur la figure 3. On remarque que l'algorithme d'estimation fournit des résultats très satisfaisants, même en présence de densités à rupture franche, comme dans le cas uniforme.

Nous avons monté dans [10], que cette méthode peut être étendue à tout problème de minimisation ou maximisation de critère entropique. En effet, ces algorithmes demandent tous la

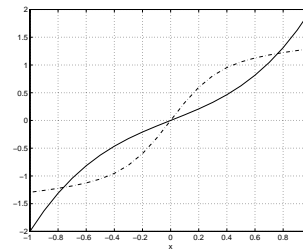


FIG. 3 — Distorsions, (—)  $f_1(x) = x + x^3$ , (- -)  $f_2(x) = 0.3x + \tanh(3x)$

connaissance des fonctions  $\psi_{Y_i}$  et plus généralement du gradient de  $\log p_Y$  dans le cas multivariable. Cette estimation est réalisée par apprentissage non supervisé d'un réseau de neurones de type MLP selon une règle d'adaptation analogue à (10) et (9). Dans [10], on applique la méthode à un problème de séparation de sources dans des mélanges linéaires : la méthode converge rapidement avec d'excellentes performances (de l'ordre de  $-40$  dB à  $-50$  dB de diaphonie résiduelle) et permet de s'affranchir d'hypothèses sur les kurtosis des sources. Un travail théorique préliminaire a été fait dans le cadre de mélanges complexes, quelques résultats sont exposés dans le papier de Bozinoski *et al.* [2].

## 4 Simulations

Dans cette partie nous présentons un exemple de simulation de l'approche proposée. Le mélange consiste en une sinusoïde et un bruit blanc uniforme (Fig. 4 gauche). La matrice de mélange est

$$\mathbf{A} = \begin{bmatrix} 8.598122e-01 & -5.681216e-01 \\ 9.991326e-01 & 9.621722e-02 \end{bmatrix},$$

les fonctions post-nonlinéaires (distorsions) sont du type saturante et dilatante (Fig. 3).

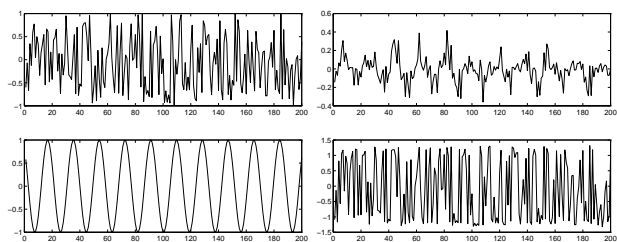


FIG. 4 — Sources et Mélanges Post-Nonlinéaires

L'observation  $\mathbf{e}(t)$  est obtenue en appliquant les distorsions sur le mélange linéaire (Fig. 4 droite), On obtient par ailleurs une forme très particulière de distribution (Fig. 5 gauche) qui indique la présence d'un mélange nonlinéaire.

L'algorithme a été appliqué sur ces mélanges, il converge en 20000 itérations. La distribution des sorties (Fig. 5 droite) du réseau nonlinéaire a la forme de parallélogramme, et indique que le mélange est devenu linéaire. On retrouve les sources originales (Fig. 6 gauche), et la distribution des signaux sources estimés (Fig. 6 droite). La diaphonie résiduelle

<sup>1</sup>L'erreur (8) nécessite la connaissance de la fonction désirée  $\psi_{Y_i}(u)$ .

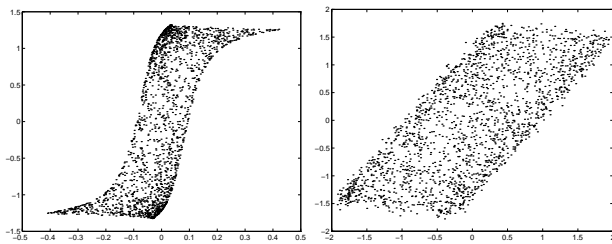


FIG. 5 — Distributions : mélanges post-nonlinéaires, sorties des fonctions nonlinéaires  $g_i$

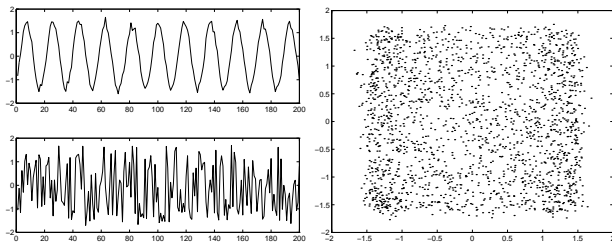


FIG. 6 — Sources estimées et distribution

est de l'ordre de  $-20$ dB. Des résultats expérimentaux ont été obtenus avec succès avec des mélanges fortement nonlinéaires de 5 sources.

## 5 Conclusion

Dans ce papier, nous avons proposé une méthode de séparation de sources pour mélanges nonlinéaires particuliers. Ces modèles appelés post-non linéaires, sont physiquement plausibles et présentent des propriétés de séparabilité sans distorsion très intéressantes. La méthode de séparation repose sur la minimisation d'un critère d'information mutuelle, qui nécessite l'estimation des dérivées logarithmique des ddp inconnues des sorties. L'estimation, effectuée au moyen d'un RN de type PMC en apprentissage non supervisé, conduit à d'excellents résultats.

Actuellement, nous étudions l'extension de cette approche à la séparation de sources dans des mélanges convolutifs linéaires et post-nonlinéaires.

## Références

- [1] S-I. Amari. Natural gradient works efficiently in learning. *Neural Computation*, 1997. To appear.
- [2] S. Bozinoski, A. Taleb, J-C. Guizzo, and C. Jutten. Séparation de sources, application à la séparation de signaux et de brouilleurs dans un satellite de télécommunications. In *Actes du XVIème colloque GRETSI*, Grenoble, Septembre 1997.
- [3] J-F. Cardoso and B. Laheld. Equivariant adaptive source separation. *IEEE Trans. on S.P.*, 44(12) :3017–3030, December 1996.
- [4] N. Charkani and Y. Deville. Optimization of the asymptotic performance of time-domain convolutive source

separation algorithms. In *ESANN'97*, pages 273–278, Bruges, Belgium, April 1997.

- [5] P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3) :287 – 314, April 1994.
- [6] G. Deco and W. Brauer. Nonlinear higher-order statistical decorrelation by volume-conserving neural architectures. *Neural Networks*, 8(4) :525–535, 1995.
- [7] P. Pajunen, A. Hyvarinen, and J. Karhunen. Nonlinear blind source separation by self-organizing maps. In *ICONIP 96*, volume 2, pages 1207–1210, Hong-Kong, September 1996.
- [8] D. T. Pham. Séparation aveugle de sources via une analyse en composantes indépendantes. In *Actes du XVème colloque GRETSI*, pages 289–292, Juan-Les-Pins, France, 18 - 21 september 1995.
- [9] D. T. Pham, P. Garat, and C. Jutten. Separation of a mixture of independent sources through a maximum likelihood approach. In J. Vandewalle, R. Boite, M. Moonen, and A. Oosterlinck, editors, *Signal Processing VI, Theories and Applications*, pages 771–774, Brussels, Belgium, August 1992. Elsevier.
- [10] A. Taleb and C. Jutten. Entropy optimization, application to blind source separation. In *ICANN'97*, Lausanne, October 1997. Submitted.
- [11] A. Taleb and C. Jutten. Nonlinear source separation : The post-nonlinear mixtures. In *ESANN'97*, pages 279–284, Bruges, April 1997.
- [12] L. Xu, C. C. Cheung, and S. Amari. Nonlinearity ans separation capability : Further justification for the ica algorithm with mixtures of densities. In *ESANN 97*, pages 291–296, Bruges, April 1997.
- [13] H. H. Yang, S. Amari, and A. Cichocki. Information back-propagation for blind separation of sources from non-linear mixture. In *ICNN'97*, Houston.
- [14] H.H. Yang and S.I. Amari. Adaptive on-line learning algorithms for blind separation– maximum entropy and minimum mutual information. *Neural Computation*, 1997. Accepted.