Séparation aveugle de sources dans les mélanges post non linéaires.

Sophie ACHARD[1] Dinh Tuan PHAM[1] Christian JUTTEN[2]

[1]Laboratoire de Modélisation et Calcul, IMAG, C.N.R.S.
Univ. de Grenoble, B.P. 53X, 38041 Grenoble Cedex, France

[2]Laboratoire des Images et des Signaux, INPG
46 avenue Félix Viallet, 38031 Grenoble Cedex, France

Sophie.Achard@imag.fr, Dinh-Tuan.Pham@imag.fr, Christian.Jutten@inpg.fr

Résumé – Partant des travaux de Taleb et Jutten sur la séparation aveugle de mélanges post non linéaires de sources, nous développons de nouvelles méthodes utilisant les dérivées des transformations non linéaires et une paramétrisation de celles-ci par des fonctions constantes par morceaux. On exprime tout d'abord le gradient du critère de l'information mutuelle en fonction de ces dérivées. Puis par la méthode de descente opposée au gradient, on met en place un algorithme de séparation. Enfin, nous illustrons la méthode par des simulations.

Abstract – Following the work of Taleb and Jutten on blind sources separation in post non linear mixtures, we develop new methods using the derivatives of non linear transformations and a parametrization of them by piecewise constant fonctions. We begin by expressing the relative gradient of mutual information in terms of the derivatives of non linear transformations. Then using the gradient descent method, we develop an algorithm of separation. Finally, we illustrate the method by simulations.

1 Introduction

Durant ces dernières années, la recherche sur la séparation aveugle de sources s'est activement développée. Bien que le cas des mélanges linéaires instantanés ou convolutifs a été beaucoup étudié [6], [8], [2], [1], le cas des mélanges non linéaires est encore peu abordé. Pour ce cas, une grande difficulté est la non identifiabilité : il existe en effet des mélanges non linéaires de sources qui conservent leurs indépendances (Darmois [4]). Il est donc impossible de réaliser la séparation aveugle de sources avec la seule hypothèse d'indépendance et des contraintes de type structurel sont nécessaires. Pour cette raison, nous considérons ici le cas des mélanges post non linéaires. Ce genre de mélange est assez réaliste. De plus, comme l'ont montré Taleb et Jutten [5], il permet la séparation aveugle des sources.

Le modèle de mélange post non linéaire peut être décrit comme suit: On dispose des enregistrements de capteurs $X_1, X_2, \dots, X_K$ provenant des transformations non linéaires sur des mélanges linéaires de K sources *indépendantes*, $S_1, S_2, \dots, S_K$, c'est-à-dire :

$$X_i = f_i(\sum_{k=1}^{K} \mathbf{A}_{ik} S_k), \qquad i = 1, \dots, K$$

oú $\mathbf{A}_{ik}$ désigne le terme général de la matrice de mélange $\mathbf{A}$ *inversible*, et $f_1, f_2, \dots, f_K$ sont K applications *monotones*. Le but est de déterminer K applications $g_1, g_2, \dots, g_K$ et une matrice $\mathbf{B}$ telles que les variables $Y_1, Y_2, \dots, Y_K$ définies par,

$$Y_i = \sum_{k=1}^{K} \mathbf{B}_{ik} Z_k, \text{ où } Z_k = g_k(X_k), \text{ pour tout } i = 1, \dots, K.$$

représentent une estimation des sources $S_1, S_2, \dots, S_K$.

Dans un premier temps, nous présenterons deux nouvelles méthodes de résolution du problème qui conduiront à deux algorithmes différents. Puis la partie 3 sera consacrée à l'implémentation numérique des algorithmes. Enfin, quelques résultats d'expérimentation seront présentés dans la partie 4.

2 Méthodes de résolution

Le principe de séparation de sources consiste à rendre le plus indépendantes possibles les sources reconstituées, $Y_1, Y_2, \dots, Y_K$. En se servant de l'information mutuelle comme fonction de contraste [2], Taleb et Jutten [5] ont obtenu le critère à minimiser suivant:

$$C(B, g_1, \dots, g_K) = \sum_{i=1}^{K} (H(Y_i) - H(Z_i)) - \log|\det \mathbf{B}|$$

où H désigne l'entropie de Shannon (pour toute variable aléatoire T de densité p_T, $H(T) = -E[\log\ p_T(T)]$) et $\mathbf{B}$ la matrice de séparation.

A partir de la définition de ce critère, ils ont proposé un algorithme de séparation utilisant la méthode de descente selon le gradient relatif du critère (Cardoso [1]), exprimé par rapport aux fonctions non linéaires $g_1, \dots, g_K$.

2.1 Approche non paramétrique

Nous pouvons remarquer que le critère défini par Taleb et Jutten ne dépend en fait que des dérivées $g'_1, \dots, g'_K$ de $g_1, \dots, g_K$. En effet, il est invariant par ajout d'une constante aux transformations non linéaires $g_1, \dots, g_K$. Ceci nous a donc motivé à exprimer le gradient relatif du critère par rapport aux arguments $\mathbf{B}$ et $g'_1, \dots, g'_K$ et à développer des méthodes basées directement sur cette paramétrisation.

2.1.1 Expression du gradient relatif du critère

Nous avons calculé le développement de Taylor au 1^{er} ordre de $C(\mathbf{B}+\varepsilon\mathbf{B}, g_1+\delta_1\circ g_1, g_2+\delta_2\circ g_2, \dots, g_K+\delta_K\circ g_K)$ par rapport à $\varepsilon, \delta_1', \delta_2', \dots, \delta_K'$ autour de 0. Nous obtenons alors comme expression du gradient relatif de C, la fonctionnelle linéaire suivante,

$$\varepsilon, \delta_1', \delta_2', \dots, \delta_K' \mapsto \sum_{1\leq i\neq k\leq K} \varepsilon_{ik} E[\psi_{Y_i}(Y_i)Y_k] +$$

$$\sum_{k=1}^{K} \int \left\{ E\left[\mathbb{I}_+(Z_k - z) \sum_{i=1}^{K} \psi_{Y_i}(Y_i)\mathbf{B}_{ik} \right] - p_{Z_k}(z) \right\} \delta_k'(z) dz$$

où ψ_{Y_i} désigne la fonction score de Y_i (i.e. $\psi_{Y_i} = -(\log p_{Y_i})'$, p_{Y_i} étant la densité de Y_i) et $\mathbb{I}_+$ désigne l'indicatrice de $[0,\infty)$($\mathbb{I}_+(x) = 1$ si $x \geq 0$, 0 sinon).

2.1.2 Equations d'estimation

Utilisant le fait que le gradient s'annule lorsque le critère a atteint son minimum, nous en déduisons les équations d'estimation suivantes

$$\begin{aligned} E[Y_j\psi_{Y_i}(Y_i)] &= 0,\ 1 \leq i \neq j \leq K, \\ E\left[\mathbb{I}_+(Z_k - z) \sum_{i=1}^{K} \psi_{Y_i}(Y_i)\mathbf{B}_{ik} \right] &= p_{Z_k}(z), 1 \leq k \leq K. \end{aligned}$$

Notons que ces équations d'estimation sont bien satisfaites si les variables $Y_1, Y_2, \dots, Y_K$ sont effectivement indépendantes.

2.1.3 Méthode de descente opposée au gradient

La minimisation peut-être réalisée grâce à la méthode de descente du gradient. On ajuste les variables $\mathbf{B}$ et $g_1', \dots, g_K'$ par le schéma itératif suivant, λ et μ représentent les constantes de réglage:

$$\left\{ \begin{array}{lll} \mathbf{B} & \mapsto & \mathbf{B} - \lambda\mathbf{D}\mathbf{B} \\ g_k' & \mapsto & g_k'(1+\mu h_k' \circ g_k) \end{array} \right. ,\ 1 \leq k \leq K.$$

où $\mathbf{D}$ et h_k' sont déduites des expressions du gradient. On a

$$\mathbf{D}_{ij} = \begin{cases} E[Y_j\psi_{Y_i}(Y_i)] & 1 \leq i \neq j \leq K, \\ 0 & \text{sinon.} \end{cases}$$

Quant au h_k', pour $1 \leq k \leq K$, on les obtient de deux manières différentes en fonction de la métrique adoptée :

- avec la métrique normale (associée au produit scalaire de l'espace de Hilbert des fonctions de carré intégrable par rapport à la mesure de Lebesgue), on a :

$$h_k^{n\prime}(z) = p_{Z_k}(z) - E\left[\mathbb{I}_+(Z_k - z) \sum_{i=1}^{K} \psi_{Y_i}(Y_i)\mathbf{B}_{ik} \right]$$

- avec la métrique probabiliste (associée au produit scalaire de l'espace de Hilbert des fonctions de carré intégrable par rapport à la mesure de probabilité), on a :

$$h_k^{p\prime}(z) = 1 - \frac{1}{p_{Z_k}(z)} E\left[\mathbb{I}_+(Z_k - z) \sum_{i=1}^{K} \psi_{Y_i}(Y_i)\mathbf{B}_{ik} \right]$$

2.2 Approche semi-paramétrique

Dans la méthode développée ci-dessus, seule la dérivée des transformations non linéaires intervient. Il est alors intéressant d'approcher $g_1, g_2, \dots, g_K$ par des fonctions continues linéaires par morceaux dont les intervalles de linéarité sont notés $[\xi_{k,m}, \xi_{k,m+1}],\ m = 1 \dots M-1$. On peut supposer sans perte de généralité que les fonctions $g_1, g_2, \dots, g_K$ sont croissantes. Nous verrons dans le § 3.2 comment cette condition peut-être aisément vérifiée. Remarquons à présent que la composée de deux fonctions linéaires par morceaux est encore linéaire par morceaux. Nous prendrons donc les fonctions δ_k du § 2.1.1 continues linéaires par morceaux dont les intervalles de linéarité sont $[\zeta_{k,m}, \zeta_{k,m+1}],\ 1 \leq m \leq M-1$, où $\zeta_{k,i} = g_k(\xi_{k,i}),\ i = 1 \dots M,\ k = 1 \dots K$. Leurs dérivées s'écrivent

$$\delta_k'(z) = \sum_{m=1}^{M-1} d_{k,m} \mathbb{I}_{[\zeta_{k,m};\zeta_{k,m+1}[}(z).$$

2.2.1 Gradient relatif

En remplaçant les dérivées δ_k' de la formule ci dessus, dans l'expression du gradient relatif obtenu précédemment, on a,

- Gradient relatif de C par rapport à $\mathbf{B}$:
$$\varepsilon \mapsto \sum_{1\leq i\neq k\leq K} \varepsilon_{ik} E[\psi_{Y_i}(Y_i)Y_k].$$
- Gradient relatif de C par rapport à $d_{k,1}, \dots, d_{k,M-1}$:
$d_{k,1}, \dots, d_{k,M-1} \mapsto$
$$\sum_{m=1}^{M-1} \int_{\zeta_{k,m}}^{\zeta_{k,m+1}} E\left[\mathbb{I}_+(Z_k - z) \sum_{i=1}^{K} \psi_{Y_i}(Y_i)\mathbf{B}_{ik} \right] dz$$
$$-P(\zeta_{k,m} \leq Z_k \leq \zeta_{k,m+1}) d_{k,m},\quad 1 \leq k \leq K.$$

2.2.2 Méthode de descente opposée au gradient

La minimisation du critère est réalisée alors de la même façon que dans l'approche non paramétrique en ajustant cette fois les paramètres $d_{k,m}$ selon le même schéma itératif qu'au § 2.1.3 mais avec les fonctions h_k' du § 2.1.3 définies par

$$h_k'(z) = \sum_{m=1}^{M-1} d_{k,m} \mathbb{I}_{[\zeta_{k,m};\zeta_{k,m+1}[}(z).$$

Comme auparavant, les $d_{k,m},\ 1 \leq k \leq K,\ 1 \leq m \leq M-1$, sont obtenus de deux manières différentes selon la métrique adoptée :

- avec la métrique de Lebesgue (associée au produit scalaire $< d_{k,.}, d'_{k,.} >= \sum d_{k,m} d'_{k,m}(\zeta_{k,m+1} - \zeta_{k,m})$), on a :

$$\begin{aligned} d_{k,m} &= \frac{1}{\zeta_k^{(m+1)} - \zeta_k^{(m)}} \Bigg\{ P(\zeta_{k,m} \leq Z_k \leq \zeta_{k,m+1}) \\ &\quad - \int_{\zeta_{k,m}}^{\zeta_{k,m+1}} E\left[\mathbb{I}_+(Z_k - z) \sum_{i=1}^{K} \psi_{Y_i}(Y_i)\mathbf{B}_{ik} \right] dz \Bigg\}. \end{aligned}$$

- avec la métrique probabiliste (associée au produit scalaire $< d_{k,.}, d'_{k,.} >= \sum d_{k,m} d'_{k,m} P(\zeta_{k,m} \leq Z_k \leq \zeta_{k,m+1})$), on a :

$$\begin{aligned} d_{k,m} &= 1 - \frac{1}{P(\zeta_{k,m} \leq Z_k \leq \zeta_{k,m+1})} \times \\ &\quad \int_{\zeta_{k,m}}^{\zeta_{k,m+1}} E\left[\mathbb{I}_+(Z_k - z) \sum_{i=1}^{K} \psi_{Y_i}(Y_i)\mathbf{B}_{ik} \right] dz. \end{aligned}$$

2.3 Estimations de la fonction score

Notons que dans les algorithmes ci-dessus, seule la fonction score des sources reconstituées $Y_1, Y_2, \ldots, Y_K$ intervient. Mais, les équations d'estimation ne sont satisfaites que dans le cas où les fonctions scores sont exactes. C'est pourquoi, l'estimation de celles-ci doit être a priori réalisée de manière très précise. Plusieurs méthodes d'estimation des fonctions scores peuvent être utilisées, la méthode des noyaux (Taleb et Jutten [5]), la méthode des splines (Cox [3], Pin [7]). Nous présentons ici une nouvelle méthode d'estimation à partir de l'estimation de l'entropie. En effet, la variation de l'entropie d'une variable aléatoire T par rapport à une *"petite"* variable aléatoire Δ s'écrit,

$$\begin{aligned} H(T+\Delta) - H(T) &= E[\Delta \psi_T(T)] \\ &\quad + \text{termes d'ordre supérieur en} \Delta \end{aligned}$$

où ψ_T est la fonction score de T. Pour toute variable aléatoire T, on note $t_1, \ldots, t_N$ un échantillon de T, $\mathcal{H}(t_1, t_2, \ldots, t_N)$ un estimateur de l'entropie de T. En se basant sur l'égalité précédente, nous proposons d'estimer la fonction score de T en t_n par

$$\hat{\psi}_T(t_n) = N \frac{\partial \mathcal{H}(t_1, \ldots, t_N)}{\partial t_n}, \;\; 1 \leq n \leq N.$$

Pour $\mathcal{H}$, nous prenons une discrétisation de l'intégrale, $-\int \hat{p}_T(T) \log \hat{p}_T(T)$, où $\hat{p}_T$ est un estimateur à noyau de la densité p_T de T. Par ailleurs, en remarquant que p_T vérifie $p_T(t) = -E[\psi_T(T) \mathbb{1}_+(T-t)]$ nous proposons d'estimer la fonction de répartition de T par,

$$\hat{P}_T(t^{(M)} \leq T \leq t^{(M+1)}) = \frac{t^{(M)} - t^{(M+1)}}{N} \sum_{n=1}^{M} \hat{\psi}_T(t^{(n)}).$$

où $t^{(1)} \leq \ldots \leq t^{(N)}$ représentent les statistiques d'ordre de T.

3 Implémentation numérique

3.1 Algorithmes

Soient $x_{k,n},\ n = 1, \ldots, N$, les observations enregistrées par les capteurs et $z_{k,n} = g_k(x_{k,n})$, l'algorithme développé par Taleb et Jutten [9], modifie à chaque pas la matrice $\mathbf{B}$ et les $z_{k,n}$. En se basant sur l'expression précédente du gradient relatif, la nouvelle méthode que nous proposons procède par ajustement de la matrice $\mathbf{B}$ et des différences $z_k^{(2)} - z_k^{(1)}, \ldots, z_k^{(N)} - z_k^{(N-1)}$, où $z_k^{(m)} = g_k(x_k^{(m)})$ et $x_k^{(1)} \leq \ldots \leq x_k^{(N)}$ sont les statistiques d'ordre de X_k. Dans le cas de la métrique probabiliste, l'algorithme proposé est décrit dans l'encadré ci-contre, dans lequel $\hat{P}_{k,m}$ désigne une estimation de $P_{k,m} = P(x_k^{(m)} \leq X_k < x_k^{(m+1)}) = P(z_k^{(m)} \leq Z_k < z_k^{(m+1)})$, $\{\pi_k(n)\}$ désigne la permutation définie par $z_k^{(m)} = z_{k,\pi_k(m)}$, et μ, λ sont deux *"petites"* constantes de réglage.

Initialisations :

$\mathbf{B} = I$

$Y = Z = X$

Boucle :

$\hat{\psi}_{Y_i}$ estimation de $\psi_{Y_i}, i = 1, \ldots, K$.

$z_k^{(m+1)} - z_k^{(m)} \mapsto [z_k^{(m+1)} - z_k^{(m)}] \times$

$\left[1 + \mu - \mu \frac{z_k^{(m+1)} - z_k^{(m)}}{\hat{P}_{k,m} N} \sum_{n=m+1}^{N} \sum_{i=1}^{K} \hat{\psi}_{Y_i}(y_{i,\pi_k(n)}) \mathbf{B}_{ik}\right]$

$k = 1, \ldots, K,\ n = 1, \ldots, N$

Normalisation de Z.

$\mathbf{B} \mapsto \mathbf{B} - \lambda \hat{\mathbf{D}} \mathbf{B}$.

où $\hat{\mathbf{D}}_{ij} = \begin{cases} \frac{1}{N} \sum_{n=1}^{N} \hat{\psi}_{Y_i}(y_{i,n}) y_{j,n} & 1 \leq i \neq j \leq K, \\ 0 & \text{sinon.} \end{cases}$

$Y = \mathbf{B} Z$.

Normalisation de Y.

Jusqu'à convergence.

Le cas de la métrique normale étant tout à fait semblable, nous ne le détaillerons pas.

3.2 Normalisations

Comme il n'est possible de retrouver les sources qu'à un facteur d'échelle près et à une permutation près, nous imposons la contrainte suivante, $\int \log g_k'(x) p_{X_k}(x) dx = 0$, $k = 1, \ldots, K$, qui fixe les indéterminations sur les facteurs d'échelle de $g_1, g_2, \ldots, g_K$. De plus, sans perte de généralité, nous pouvons imposer que les fonctions $g_1, g_2, \ldots, g_K$ soient strictement croissantes, ce qui se traduit simplement par les conditions $z_k^{(m+1)} - z_k^{(m)} > 0$, $m = 1, \ldots, N-1$. Ces changements ne modifient pas le critère à condition de multiplier à droite la matrice $\mathbf{B}$ par une matrice diagonale appropriée.
Au lieu de procéder à la normalisation des sources, nous pouvons ajuster la matrice $\mathbf{B}$ par une méthode invariante par changement d'échelle. La matrice $\mathbf{D}$ de l'algorithme est alors modifiée comme suit :

$$\begin{bmatrix} \hat{D}_{ij} \\ \hat{D}_{ji} \end{bmatrix} = \begin{bmatrix} \hat{E}[\hat{\psi}_{Y_i}(Y_i) Y_j] / \widehat{\text{var}}(Y_j) \hat{E}[\hat{\psi}^2_{Y_i}(Y_i)] \\ \hat{E}[\hat{\psi}_{Y_j}(Y_j) Y_i] / \widehat{\text{var}}(Y_i) \hat{E}[\hat{\psi}^2_{Y_j}(Y_j)] \end{bmatrix}.$$

Les dénominateurs représentent les termes diagonaux du Hessien du critère relativement à la partie linéaire.

4 Résultats expérimentaux

Nous présentons sur la figure 1, les résultats d'une simulation de la méthode semi-paramétrique avec

- Une grille de 10 points parmi les observations.
- La métrique probabiliste avec 32 itérations.
- L'estimation de la fonction score par la méthode présentée ci-dessus.
- $\mu = 0.2$, $\lambda = 0.5$.

– $\mathbf{A} = \begin{pmatrix} 1 & 0.6 \\ 0.7 & 1 \end{pmatrix}$ et $f_1(x) = f_2(x) = \tanh(4x) + 0.1x$

Seul est présenté ici le résultat d'une simulation faite sous des conditions particulières. D'autres simulations effectuées nous ont permis de remarquer que les performances de l'algorithme sont très influencées par les composantes du mélange et la méthode d'estimation de la fonction score. Les deux métriques ont également des effets très différents sur la convergence de l'algorithme. Avec la métrique normale, l'algorithme est plus stable à la convergence. Nous pouvons aussi souligner la convergence rapide de ces algorithmes en comparaison à celui de Taleb et Jutten [5], par contre, ils présentent une grande instabilité quand le nombre d'itération augmente. Enfin, les coefficients μ et λ ont été fixés empiriquement, on peut envisager de les adapter à chaque itération afin de forcer la convergence. Nous avons l'intention d'étudier précisément dans quelles mesures les paramètres de l'algorithme affectent ses performances.

5 Conclusion

En étudiant les invariances du critère utilisé par Taleb et Jutten comme fonction de contraste, nous avons trouvé intéressant d'exprimer le gradient relatif du critère par rapport aux dérivées des transformations non linéaires. Puis, à partir de ceci nous avons pu développer une approche semi-paramétrique, consistant en l'approximation des transformations non linéaires par des fonctions continues linéaires par morceaux. Au cours de ce travail nous avons aussi envisager une autre procédure d'estimation des fonctions scores. Ainsi d'autres algorithmes permettant de séparer les sources ont pu être mis en place.

Références

[1] J.-F. CARDOSO. Blind signal separation : Statistical principles. *Proceedings IEEE*, 10(86):2009–2025, Oct. 1998.

[2] P. COMON. Independent component analysis, a new concept ? *Signal Processing*, 3(36):287–314, Avr. 1994.

[3] D.D. COX. A penalty method for nonparametric estimation of the logarithmic derivative of density function. *Ann. Inst. Statist. Math.*, 37(Part A):271–288, 1985.

[4] G. DARMOIS. Analyse générale des liaisons stochastiques. *Rev. Inst. Internat. Stat.*, 21:2–8, 1953.

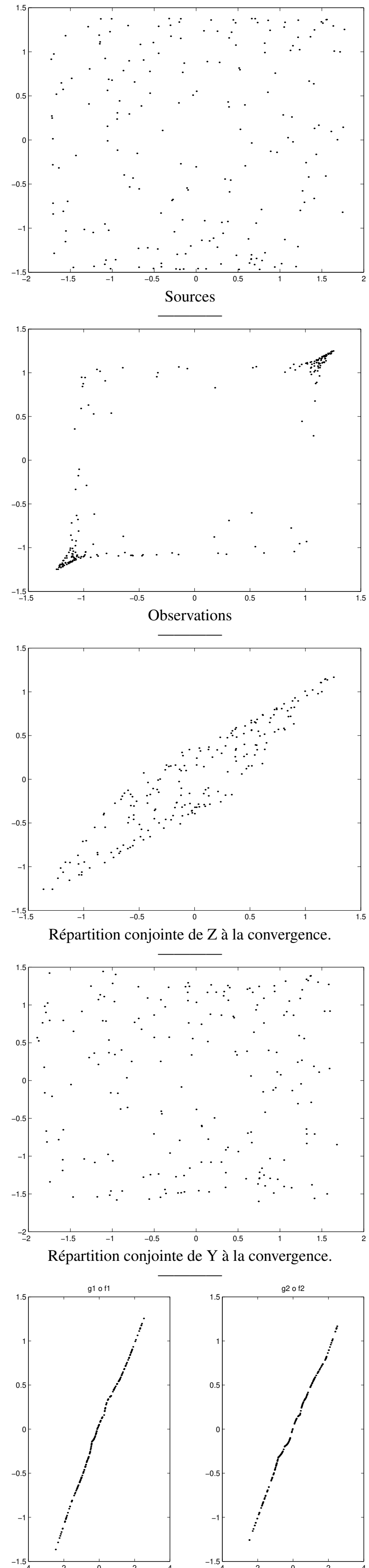
[5] A. TALEB et C. JUTTEN. Sources separation in post-nonlinear mixtures. *IEEE Transactions on Signal Processing*, 10(47):2807–2820, Oct. 1999.

[6] A. BELL et T. SEJNOWSKI. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6), 1995.

[7] PIN T. NG. Smoothing spline score estimation. *S.I.A.M. J. Sci. Compt.*, 15(5):1003–1025, Sept 1994.

[8] D.T. PHAM. Blind separation of instantaneous mixture of sources via an independent component analysis. *IEEE Transactions on Signal Processing*, 44(11):2768–2779, Nov. 1996.

[9] A. TALEB. *Séparation de Sources dans les Mélanges Non Linéaires*. PhD thesis, I.N.P.G. - Laboratoire L.I.S., 1999.



Sources

Observations

Répartition conjointe de Z à la convergence.

Répartition conjointe de Y à la convergence.

Composition des fonctions non linéaires à la convergence.

FIG. 1: *Simulation*