Séparation aveugle de sources par optimisation de l'information mutuelle sous contraintes, estimation et sélection de modèles

Mohamed Salem OULD MOHAMED[1], Amor KEZIOU[2], Hassan FENNIRI[1] & Georges DELAUNAY[1]

[1] CReSTIC, [2] UMR 6056 CNRS, Université de Reims Champagne-Ardenne
UFR Sciences Exactes Naturelles - Moulin de la Housse
BP 1039 - 51687 REIMS Cedex 2 - France
salem.ould-mohamed@univ-reims.fr

RÉSUMÉ. Dans cet article, nous proposons un nouvel algorithme de séparation aveugle de sources basé sur l'optimisation de l'information mutuelle sous contraintes. Le problème d'optimisation est résolue par passage au problème dual associé et la méthode de descente du gradient. L'estimateur du gradient stochastique proposé est basé sur l'estimation des densités par maximum de vraisemblance dans un modèle de lois exponentielles choisi par minimisation du critère AIC. Les résultats sont présentés dans le cas d'un mélange linéaire instantané de deux sources.

ABSTRACT. In this paper, we propose a new blind source separation algorithm based on the optimization of mutual information under constraints. The optimization problem is solved by using the dual problem and the descent gradient method. The estimator of stochastic gradient is based on the estimation of the densities by maximum likelihood method. The densities are chosen from exponential families using the AIC criterion. The results are presented in the case of instantaneous linear mixture.

MOTS-CLÉS : Séparation aveugle de sources; Information mutuelle; AIC; Sélection de modèles; Maximum de vraisemblance; Famille de lois exponentielles; Optimisation sous contraintes; Lagrangien; Problème inverse.

KEYWORDS : Blind source separation; Mutual information; AIC; Model selection; Maximum likelihood; Exponential family laws ; Optimization under constraints; Lagrangian; Inverse problem.

1. Introduction

Les premières méthodes de séparation aveugle de sources ont été abordées au milieu des années 80 [1, 2]. Ensuite, [3, 4] a défini rigoureusement le cadre mathématique de l'Analyse en Composantes Indépendantes, duquel dérivent plusieurs applications. La séparation aveugle de sources est une technique de traitement du signal, dont le principe peut être résumé de la façon suivante : "restituer un ensemble de signaux non-observables appelés sources à partir d'un ensemble de signaux mesurés appelés observations". La plupart des méthodes de séparation aveugle de sources supposent l'indépendance statistique de celles-ci ; elles consistent à rendre les composantes de la sortie indépendantes au sens statistique.

Dans le cas linéaire instantané, le mélange de deux sources est de la forme

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n},$$

où $\mathbf{s} = (s_1, s_2)^T$ sont les signaux sources et $\mathbf{A}$ est la matrice de mélange de dimension 2×2. $\mathbf{n}$ est le vecteur des signaux bruits, dont les éléments sont supposés centrés, mutuellement indépendants, et indépendants des signaux sources. La présence du bruit additif dans le modèle de mélange complique le problème de la séparation de sources. Pour simplifier le problème, le bruit est supposé négligeable (après prétraitement). Le but est d'estimer les sources $\mathbf{s}$ à partir des observations $\mathbf{x} = (x_1, x_2)^T$. L'estimateur, dans ce cas, est de la forme

$$\mathbf{y} = \mathbf{B}\mathbf{x},$$

où $\mathbf{B}$ est une matrice 2×2. Le problème consiste donc à chercher un estimateur $\widehat{\mathbf{B}}$, à partir des observations $\mathbf{x}$, qui conduit à une bonne estimation

$$\mathbf{y} = \widehat{\mathbf{B}}\mathbf{x},$$

des sources $\mathbf{s}$. Une solution à ce problème, à une indétermination du facteur d'échelle et à une permutation près, consiste à chercher une matrice $\widehat{\mathbf{B}}$ qui rend les composantes de $\mathbf{y}$ indépendantes au sens statistique, c.f. [5] et [4]. La minimisation de l'information mutuelle (MI) est considérée comme le critère idéal pour l'estimation de $\mathbf{B}$. Cela conduit à l'estimation des densités marginales p_{y_1} et p_{y_2} et des fonctions score marginales associées. Plusieurs méthodes d'estimation paramètriques et non parmètriques ont été proposées dans la littérature, c.f. [6] et [7]. Nous proposons dans ce travail, une autre approche d'estimation basée sur la sélection d'un modèle, via la minimisation du critère d'information de Akaïke (AIC), parmi une famille croissante de modèles de lois exponentielles de dimensions quelconques. Les paramètres du modèle sélectionné sont estimés par maximum de vraisemblance. Les estimateurs des fonctions score sont obtenus par dérivation des densités estimées. Enfin pour remédier au problème de l'indétermination du facteur d'échelle, et stabiliser l'algorithme de minimisation, dans le paragraphe suivant nous proposons de minimiser l'information mutuelle sous contraintes de normalisation des puissances des sources estimées.

2. Information mutuelle sous contraintes

L'approche que nous proposons est basée sur l'optimisation de l'information mutuelle

$$I(\mathbf{y}) = \int_{\mathbb{R}^2} \mathbf{p_y}(t_1,t_2) \ln \left(\frac{\mathbf{p_y}(t_1,t_2)}{p_{y_1}(t_1)p_{y_2}(t_2)} \right) dt_1 dt_2 \quad \text{sous les contraintes} \tag{1}$$

$$\text{var}(y_1) - 1 = 0 \quad \text{et} \quad \text{var}(y_2) - 1 = 0,$$

où

$$\text{var}(y_i) = \mathbb{E}\left[(y_i - \mathbb{E}(y_i))^2\right]$$

est la variance de y_i, i=1,2.

Pour résoudre le problème (1), nous allons utiliser le Lagrangien associé. Il s'écrit

$$L(\mathbf{B}, \lambda_1, \lambda_2) = I(\mathbf{y}) + \lambda_1 (\text{var}(y_1) - 1) + \lambda_2 (\text{var}(y_2) - 1) \,. \tag{2}$$

Selon le principe de dualité, la solution du problème sous contraintes (1) est donnée par celle du problème d'optimisation sans contraintes suivant

$$\min_{\mathbf{B}} \max_{\lambda 1, \lambda 2} L(\mathbf{B}, \lambda_1, \lambda_2) = \min_{\mathbf{B}} \max_{\lambda 1, \lambda 2} \{I(\mathbf{y}) + \lambda_1 (\text{var}(y_1) - 1) + \lambda_2 (\text{var}(y_2) - 1)\} \,. \tag{3}$$

Comme ce dernier est sans contrainte, la solution peut être calculée par utilisation des algorithmes de descente du gradient. À l'optimum, on a

$$\frac{\partial L}{\partial \mathbf{B}} = 0, \quad \frac{\partial L}{\partial \lambda_\mathbf{i}} = 0, \; i = 1, 2.$$

En exploitant des résultats de [8], voir aussi [9] et [10], il vient

$$\frac{\partial L}{\partial \mathbf{B}} = \frac{\partial I(\mathbf{Bx})}{\partial \mathbf{B}} + 2\mathbf{diag}(\lambda_1, \lambda_2)\mathbb{E}\left(\mathbf{yx}^T\right) \quad \text{et} \quad \frac{\partial L}{\partial \lambda_i} = \text{var}(y_i) - 1, \; i = 1, 2 \tag{4}$$

où

$$\frac{\partial I(\mathbf{Bx})}{\partial \mathbf{B}} = \mathbb{E}\left[\psi(\mathbf{y})\mathbf{x}^T\right] - \mathbf{B}^{-T}, \quad \psi(\mathbf{y}) = (\psi_1(y_1), \psi_2(y_2))^T$$

et $\psi_i(.) = -\frac{p'_{y_i}(.)}{p_{y_i}(.)}$, la fonction score marginale de $y_i, \; i = 1, 2$.

3. Sélection de modèles, pour l'estimation des densités, fonctions scores et gradient stochastique

Soit $\mathbf{y(1)} = (y_1(1), y_2(1))^T, \ldots, \mathbf{y(n)} = (y_1(n), y_2(n))^T$ un échantillon du vecteur aléatoire $\mathbf{y} = (y_1, y_2)^T$. Pour estimer les densités marginales $p_{y_1}(.)$ et $p_{y_2}(.)$, nous allons modéliser ces densités de probabilité par des familles de densités exponentielles de dimensions (nombre de paramètres) $d_1, d_2 \in \mathbb{N}^*$. Le modèle retenu est celui qui minimise le critère d'information de Akaïke (AIC) [11]. Dans ce qui suit, pour simplifier, on note $p_y(.)$ la densité marginale de y_1 ou y_2.

Une densité de loi exponentielle (de dimension d) et de support $[a, b]$ s'écrit

$$f_d(y, \theta, a, b) = \frac{\exp\{\theta_1 y + \cdots + \theta_d y^d\}}{\int_a^b \exp\{\theta_1 y + \cdots + \theta_d y^d\}\ dy},\ a, b \in \mathbb{R}, \theta \in \mathbb{R}^d \text{ and } d = 1, 2, \ldots \quad (5)$$

Notons $AIC(d)$ le critère AIC associé au modèle f_d et à l'échantillon $y(1), \ldots, y(n)$, i.e.,

$$AIC(d) = \min_{\theta \in \mathbb{R}^d} \left\{ -\frac{1}{n} \sum_{i=1}^{n} \log f_d(y(i), \theta, a, b) + \frac{d}{n} \right\}, \quad (6)$$

et $d^* = \arg\min_{d \in \mathbb{N}^*} AIC(d)$. Notons que le terme d/n est dû au biais de l'estimation de l'entropie de p_y, il dépend de la dimension d du modèle ; la quantité $AIC(d)$ est donc un estimateur corrigé du biais de l'entropie de $p_y(.)$ à l'intérieur du modèle f_d. Ainsi f_{d^*}, la famille exponentielle de dimension d^*, est le modèle retenu. Soit $\widehat{\theta}$ la solution de (6). Notons que $\widehat{\theta}$ est l'estimateur du maximum de vraisemblance du paramètre θ. Lorsque le support $[a, b]$ est inconnu, on l'estime par $[\widehat{a}, \widehat{b}] = [\min_{i=1,\ldots,n} y(i), \max_{i=1,\ldots,n} y(i)]$. Cette démarche conduit donc au choix du meilleur modèle et à une estimation optimale des paramètres de celui-ci. La densité $p_y(.)$ est estimée par $\widehat{p}_y(y) = f_{d^*}(y, \widehat{\theta}, \widehat{a}, \widehat{b})$, ce qui permet de déduire l'estimateur suivant des fonctions marginales score $\psi_y(.)$

$$\widehat{\psi}_y(y) = -\frac{\widehat{p}_y(y)'}{\widehat{p}_y(y)} = -\sum_{j=1}^{d^*} j\widehat{\theta}_j y^{j-1}. \quad (7)$$

3.1. Algorithme

Nous résumons la méthode de séparation proposée par l'algorithme suivant
Initialisation de $\mathbf{B},\ \lambda_1,\ \lambda_2$
Itérer :

1) $\mathbf{y} = \mathbf{B}\mathbf{x}$

2) calculer l'estimateur $\frac{\partial \widehat{L}}{\partial \mathbf{B}} = \mathbb{E}\left[\widehat{\psi}(\mathbf{y})\mathbf{x}^T\right] - \mathbf{B}^{-T} + 2\mathbf{diag}(\lambda_1, \lambda_2)\mathbb{E}\left(\mathbf{w}(\mathbf{y})\mathbf{x}^T\right)$

3) $\mathbf{B} \leftarrow \mathbf{B} - \mu_1 \frac{\partial \widehat{L}}{\partial \mathbf{B}}, \quad (\lambda_1, \lambda_2)^T \leftarrow (\lambda_1, \lambda_2)^T + \mu_2(\mathrm{var}(y_1) - 1, \mathrm{var}(y_2) - 1)^T$

fin

L'estimateur $\widehat{\psi}(.)$ dans (2) est fourni par (7).

4. Résultats de simulations

Pour illustrer la performance de l'algorithme proposé, nous présentons des résultats de simulations (voir la figure1), pour des signaux de loi uniforme centrée réduite, en utilisant la matrice de mélange suivante

$$A = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}.$$

La performance est évaluée par le rapport signal sur interférence suivant

$$SNR_i = 10 \log_{10} \frac{\mathbb{E}\{s_i^2\}}{\mathbb{E}\left\{(s_i - y_i)^2\right\}}.$$

La présente méthode est comparée avec la méthode d'estimation des fonctions score de [6]. Les simulations montrent une bonne performance de l'algorithme proposé.

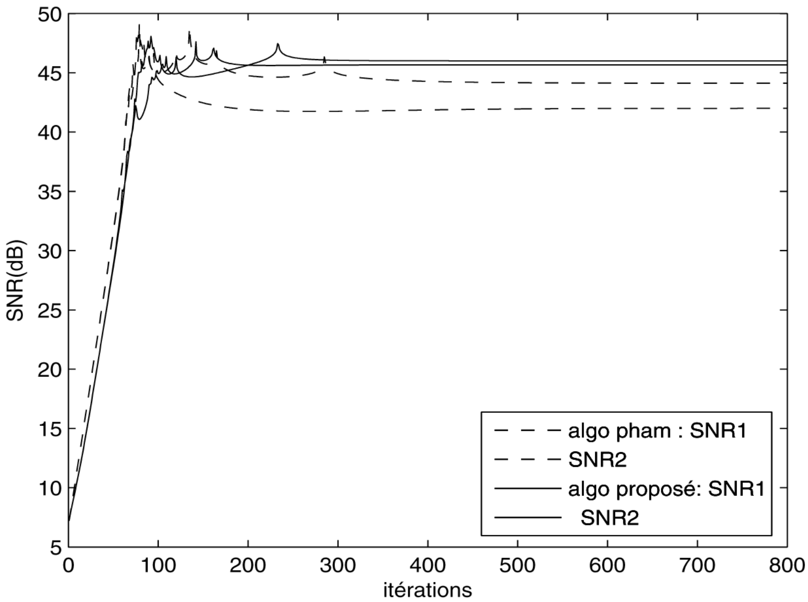

Figure 1. *Rapport signal sur interférence (SNR) de notre algorithme comparé avec celui de Pham en fonction du nombre d'itérations*

5. Conclusion

Dans ce papier, nous avons proposé une nouvelle méthode de séparation aveugle de sources dans le cadre d'un mélange linéaire instantané. La méthode proposée est basée sur l'optimisation de l'information mutuelle sous contraintes par un passage au problème dual et estimation du gradient stochastique par maximum de vraisemblance dans des familles de lois exponentielles. L'algorithme proposé s'applique pour une large classe de lois de probabilités pouvant être modélisées par une famille de lois exponentielles. De plus, la pénalisation permet de s'affranchir du problème de l'indétermination du facteur d'échelle et stabilise l'algorithme. La méthode peut se généraliser au cas de plusieurs sources.

6. Bibliographie

[1] J. HÉRAULT AND B. ANS, « Circuits neuronaux à synapses modifiables : Décodage de message composites par apprentissage non supervisé », *C. R. de l'Academie des Sciences, 299(III-13)*, p. 525-528, 1984.

[2] CHRISTIAN JUTTEN, J. HÉRAULT AND B. ANS, « Détection de grandeurs primitives dans un message composite par une architecture de calcul neurominétrique en apprentissage non supervisé », *Xème colloque GRETSI, Nice, France, 20-24 May*, p. 1017-1022, 1985.

[3] PIERRE, COMON, « Independent component analysis, a new concept ? », *Proc. Int. Workshop on Higher-Order Statistics, Chamrousse, France*, p. 111-120, 1991.

[4] PIERRE COMON, « Independent component analysis, a new concept ? », *Signal Processing, IEEE*, vol. 36, n^o 3, p. 287-314, 1994.

[5] JEAN-FRANÇOIS CARDOSO AND ANTOINE SOULOUMIAC, « Blind signal Beamforming for non Gaussian Signals », *Proceedings of the IEEE*, vol. 140, n^o 6, p. 362-370, 1993.

[6] DUNH-TUAN PHAM, « Fast algorithm for estimating mutual information, entropies and score functions. », *Proceedings of ICA 2003 Conference, Nara, Japan*, 2003.

[7] MASSOUD BABAIE-ZADEH AND CHRISTIAN JUTTEN, « A general approach for mutual information minimization and its application to blind source separation », *Signal Processing*, vol. 85, p. 975-995, 2005.

[8] MOHAMMED EL RHABI, GUILLAUME GELLE, HASSAN FENNIRI AND GEORGES DELAUNAY, « A penalized mutual information criterion for blind separation of convolutive mixtures », *Signal Processing*, vol. 84, p. 1979-1984, 2004.

[9] HASSAN FENNIRI, MOHAMMED EL RHABI, GUILLAUME GELLE AND GEORGES DELAUNAY, « Blind separation of rotating machine signals using penalized mutual information criterion and minimal distortion principle », *Mechanical systems and signal processing*, vol. 19, p. 1282-1992, 2005.

[10] MOHAMMED EL RHABI, HASSAN FENNIRI AND GEORGES DELAUNAY, « Blind sources separtion using penalized mutual information criterion and minimal distortion principle », *Second international symposium on communications, control and signal processing (ISCCSP 2006)*, Marrakech, Morocco, p. 13-15, March 2006.

[11] AKAIK E, H., « A new look at statistical model identification », *IEEE Transactions on Automatic Control*, vol. AU-19, p. 716-722, 1974.