HAL
open science

Une extension avec variables latentes du modèle d'analyse en composantes indépendantes

Selwa Rafi, Marc Castella, Wojciech Pieczynski

To cite this version:

Selwa Rafi, Marc Castella, Wojciech Pieczynski. Une extension avec variables latentes du modèle d'analyse en composantes indépendantes. GRETSI 2011 : XXIIIe Colloque, Sep 2011, Bordeaux, France. pp.1 - 4. hal-01465735

HAL Id: hal-01465735

https://hal.science/hal-01465735v1

Submitted on 13 Feb 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Une extension avec variables latentes du modèle d'analyse en composantes independantes

Selwa Rafi, Marc Castella, Wojciech Pieczynski

Institut Télécom; Télécom SudParis; Département CITI; UMR-CNRS 5157
9 rue Charles Fourier, 91011 Evry Cedex, France
selwa.rafi@it-sudparis.eu, marc.castella@it-sudparis.eu
wojciech.pieczynski@it-sudparis.eu

Résumé – Cet article traite du problème de séparation aveugle de sources (SAS) dans le cas d'un mélange linéaire instantané. Nous étendons le modèle d'analyse en composantes indépendantes (ACI) en un nouveau modèle où les sources peuvent être dépendantes. Conditionnellement à un processus latent, l'indépendance ou non des composantes des sources est contrôlée. Nous proposons une méthode de séparation qui combine: (i) une séparation par ACI classique réalisée sur l'ensemble des échantillons dont les composantes sont conditionnellement indépendantes, et (ii) une méthode d'estimation du processus latent qui repose sur une procédure d'estimation conditionnelle itérative (ECI). Des simulations confirment la plus grande généralité de notre modèle et montrent que notre méthode permet la prise en compte d'une dépendance partielle des sources.

Abstract – This paper deals with the problem of blind source separation of instantaneous linear mixtures. The Independent Component Analysis (ICA) model is extended to a new model where the sources can be dependent. Depending on the value of a latent process, the unknown sources are assumed either mutually independent or dependent. We propose a separation method which combines: (i) a classical ICA separation performed using the set of samples whose components are conditionally independent, and (ii) a method for estimation of the latent process using the Iterative Conditional Estimation (ICE) method. Simulations confirm the great generality of our model and show that our method takes into account a partial dependence between the sources.

1 Introduction

La séparation aveugle de sources (SAS) est un domaine de recherche actif en raison d'un large spectre d'applications potentielles (audio, télécommunications, biologie,...).

Dans le cas des systèmes linéaires instantanés multi-entrées/multi-sorties, la SAS correspond à la technique d'analyse en composantes indépendantes (ACI), qui est désormais bien connue [5]. L'hypothèse de base de l'ACI consiste en l'indépendance mutuelle des sources. Toutefois, cette hypothèse n'est pas toujours nécessaire pour la SAS et peut être remplacée par d'autres hypothèses. À titre d'exemple, dans le cas des mélanges statiques, les sources peuvent être séparées si elles sont décorrélées seulement, à condition que leur non-stationnarité ou leur couleur puisse être exploitée. D'autres propriétés telles que l'appartenance à un alphabet fini peuvent aussi être utilisées [4]. Nous considérons dans ce travail des sources dépendantes mais ne ferons aucune hypothèse, ni sur la non-stationarité, ni sur la couleur. À notre connaissance, peu de références ont abordé la SAS dans un tel contexte [1,3], bien que l'intérêt des sources dépendantes ait été mis en évidence dans certaines applications.

Notre apport est de proposer un modèle qui combine celui d'ACI avec un modèle probabiliste sur les sources. De manière stochastique, elles sont ainsi dépendantes ou indépendantes à différents instants. Notre méthode exploite les sous-échantillons où les sources sont générées de façon «indépendante». Un processus caché définit les instants où les sources sont indépendantes et nous l'estimons grâce à un algorithme d'estimation conditionnelle itérative (ECI). Enfin, le processus caché est ou bien une suite de variables aléatoires indépendantes et identiquement distribuées, ou bien une chaîne de Markov.

Dans l'ensemble de l'article $\mathcal{N}(\mathbf{0}, \Sigma)$ désignera la loi gaussienne de moyenne nulle et de matrice de covariance Σ, $\mathcal{L}(\lambda)$ est la distribution scalaire de Laplace définie par le paramètre λ. Nous utiliserons le signe $\sim$ pour indiquer la distribution suivie par une variable aléatoire et $\mathbb{P}(\mathbf{r} \,|\, \mathbf{X}; \theta)$ représente la loi de $\mathbf{r}$ conditionnellement à $\mathbf{X}$ sous la valeur θ du paramètre.

2 Extension du modèle ACI

2.1 Mélange linéaire

Nous considérons un vecteur observé de taille N, $\mathbf{x}(t) \triangleq (x_1(t), \ldots, x_N(t))^{\mathrm{T}}$, qui résulte d'un mélange linéaire de

N signaux inconnus et non observés appelés *sources*. En d'autres termes, il existe une matrice $\mathbf{A} \in \mathbb{R}^{N\times N}$ et un processus à valeurs vectorielles, $\mathbf{s}(t) \triangleq (s_1(t), \ldots, s_N(t))^{\mathrm{T}}$, tel que pour chacun des T échantillons $t \in \{1, \ldots, T\}$ nous avons :

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t). \tag{1}$$

Soit $\mathbf{X} \triangleq (\mathbf{x}(1), \ldots, \mathbf{x}(T))$ la matrice $N \times T$ contenant les *observations* et $\mathbf{S} \triangleq (\mathbf{s}(1), \ldots, \mathbf{s}(T))$ la matrice $N \times T$ des sources. La matrice $\mathbf{A}$ est inconnue, et l'objectif de la SAS consiste à restaurer $\mathbf{S}$ à partir de $\mathbf{X}$. Nous supposons $\mathbf{A}$ inversible et le problème revient donc à estimer $\mathbf{A}$ ou son inverse $\mathbf{B}$, dite *matrice de séparation*. Une solution connue et appelée ACI [5], nécessite généralement deux hypothèses : les composantes des sources sont non gaussiennes —sauf éventuellement une— et sont indépendantes. Avec ces hypothèses, on peut estimer une matrice $\mathbf{B} \in \mathbb{R}^{N\times N}$ telle que $\mathbf{y}(t) = \mathbf{B}\mathbf{x}(t)$ restaure les sources à des ambiguïtés de permutation et de facteur d'échelle près.

2.2 Variables latentes

Dans ce travail, nous proposons un modèle et une méthode qui étendent ACI afin d'affaiblir l'hypothèse d'indépendance. L'idée consiste à introduire un processus caché $r(t)$ qui selon sa valeur à l'instant t, contrôle l'indépendance ou la dépendance des composantes de $\mathbf{s}(t)$. Soit $\mathbf{r} \triangleq (r(1), \ldots, r(T))$. On suppose :

A1. Conditionnellement à $\mathbf{r}$, les variables aléatoires vectorielles $\mathbf{s}(1), \ldots, \mathbf{s}(T)$ à différents instants sont indépendantes.

A2. Le processus caché $r(t)$ prend ses valeurs dans $\{0, 1\}$ et, conditionnellement à $r(t)$:
   (i) si $r(t) = 0$, les composantes de $\mathbf{s}(t)$ sont mutuellement indépendantes et non gaussiennes, sauf éventuellement une au plus ;
   (ii) si $r(t) = 1$, les composantes de $\mathbf{s}(t)$ sont dépendantes.

Dans un contexte SAS, si le processus $\mathbf{r}$ était connu, on pourrait appliquer les techniques d'ACI en ignorant les échantillons à des instants où les sources sont dépendantes. Plus précisément, soit $\mathcal{I}_0 \triangleq \{t \in \{1, \ldots, T\} \mid r(t) = 0\}$ l'ensemble des instants où les composantes de $\mathbf{s}(t)$ sont indépendantes. Alors, le sous-ensemble $\mathbf{X}_0 \triangleq (\mathbf{x}(t))_{t\in\mathcal{I}_0}$ de l'ensemble des observations $\mathbf{X}$ satisfait les hypothèses habituellement exigées par les techniques d'ACI. L'idée de notre travail consiste à effectuer alternativement et de manière itérative une estimation de $\mathbf{B}$ (correspondant à $\mathbf{A}^{-1}$) et des données cachées $\mathbf{r}$.

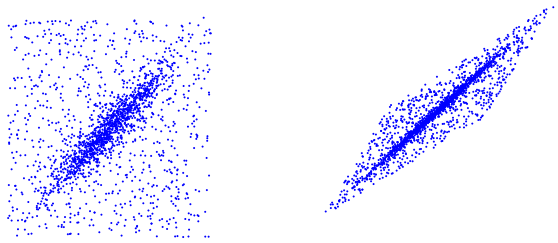
2.3 Modèle des sources

Notre méthode est valide quel que soit le nombre N de sources. Cependant, nous considérons le cas simple de $N = 2$ sources, qui correspond également aux données considérées dans nos simulations. Dans l'hypothèse A2(ii), les composantes dépendantes de $\mathbf{S}$ sont générées avec deux caractéristiques différentes selon les cas étudiés :

– *Cas 1* :Soit $\mathbf{u}$ le vecteur aléatoire défini par :

$$\mathbf{u} \triangleq \frac{1}{\sqrt{2}} \begin{pmatrix} s_1(t) + s_2(t) \\ s_1(t) - s_2(t) \end{pmatrix}. \tag{2}$$

Le vecteur $\mathbf{u}$ est alors distribué selon une loi $\mathcal{N}(0, \sigma) \times \mathcal{L}(\lambda)$. Les composantes de $\mathbf{u}$ sont indépendantes, $u_1 \sim \mathcal{N}(0, \sigma)$ et $u_2 \sim \mathcal{L}(\lambda)$. Un exemple illustratif est donné figure 1 ;



(a) (b)

FIG. 1 – (a) Réalisation des sources $\mathbf{s} = (s_1, s_2)$ correspondant au *Cas 1* avec $\lambda = 9$ et $\sigma = 0.7$. (b) Observations $\mathbf{x} = \mathbf{A}\mathbf{s}$ correspondantes.

– *Cas 2* : Pour tout $t \in \{1, \ldots, T\}$: $s_1(t) = s_2(t)$.

Intuitivement, notre méthode consiste à éliminer l'ensemble du nuage des points dépendants dans les distributions de la figure 1 et utiliser les données restantes pour une séparation basée sur l'ACI classique.

3 Estimation des paramètres et segmentation du processus latent

Notons θ les paramètres à estimer, c'est-à-dire la matrice $\mathbf{B}$ et les paramètres qui définissent la distribution de $\mathbf{r}$. L'ensemble $(\mathbf{r}, \mathbf{X})$ est appelé données *complètes*, tandis que $\mathbf{X}$ constitue les données *incomplètes*. Le modèle de la partie 2 correspond au cas où nous ne disposons que des données *incomplètes* pour estimer θ. Il est à noter que l'adjectif *aveugle* est utilisé pour indiquer que $\mathbf{S}$ n'est pas disponible, tandis que *incomplète* indique que $\mathbf{r}$ est inconnu.

Estimation Conditionnelle Itérative (ECI)

ECI est une méthode d'estimation utilisée dans le contexte de données incomplètes [7,8] et qui consiste à définir de façon itérative une séquence $(\hat{\theta}^{[q]})_{q=1,2,\ldots}$ de valeurs des paramètres en partant d'une valeur initiale. A chaque itération $[q]$, les paramètres $\hat{\theta}^{[q]}$ sont obtenus à partir des

observations et des valeurs à l'itération précédente $[q-1]$. L'estimation par la méthode ECI nécessite :

- l'existence d'un estimateur à partir des données complètes $\hat{\theta}(\mathbf{r}, \mathbf{X})$;
- la possibilité de calculer $\mathbb{E}\{\hat{\theta}(\mathbf{r}, \mathbf{X}) \mid \mathbf{X}; \hat{\theta}^{[q-1]}\}$ ou de simuler des réalisations de variables aléatoires selon la loi $\mathbb{P}(\mathbf{r}|\mathbf{X}; \hat{\theta}^{[q-1]})$.

La méthode ECI se déroule de la manière suivante :

- Pour les estimateurs $\hat{\theta}_j$ dont on peut calculer l'espérance conditionnelle, l'estimation est donnée par :

$$\hat{\theta}_j^{[q]} = \mathbb{E}\{\hat{\theta}_j(\mathbf{r}, \mathbf{X}) \mid \mathbf{X}; \hat{\theta}_j^{[q-1]}\}; \tag{3}$$

- Pour les estimateurs $\hat{\theta}_j$ dont on ne peut pas calculer l'espérance conditionnelle, un nombre fixé $K \in \mathbb{N}^*$ de réalisations de $\mathbf{r}$ sont tirées selon la loi *a posteriori* $\mathbb{P}(\mathbf{r}|\mathbf{X}; \hat{\theta}^{[q-1]}$. En notant $\mathbf{r}^{(k)}$ ces réalisations, l'expression de l'estimation est donnée par :

$$\hat{\theta}^{[p]} = \frac{1}{K} \sum_{k=1}^{K} \hat{\theta}(\mathbf{r}^{(k)}, \mathbf{X}) \tag{4}$$

4 Modélisation de $(\mathbf{r}, \mathbf{S})$

Notre modèle est un modèle original qui combine un modèle statistique de $(\mathbf{r}, \mathbf{S})$ avec un modèle ACI sur $(\mathbf{S}, \mathbf{X})$. Cette combinaison nous place dans un contexte plus général que le contexte de la séparation ou de la segmentation :

- dans le cas de la séparation, l'hypothèse d'indépendance des sources $\mathbf{s}(t)$ du modèle ACI est plus restrictive que l'hypothèse A2.
- dans le cas de la segmentation, le modèle probabiliste de $(\mathbf{r}, \mathbf{X})$ est posé sans aucune relation particulière, c'est-à-dire, sans l'équation (1).

4.1 Modèle pour le calcul de la loi $\mathbb{P}(\mathbf{r} \mid \mathbf{X}; \theta)$

Le modèle supposé sur $(\mathbf{r}, \mathbf{S})$ pour le calcul de la loi $\mathbb{P}(\mathbf{r} \mid \mathbf{X}; \theta)$ est distinct de celui indiqué en 2.3 et remplacé par une distribution moins informative mais plus robuste que celle vérifiée par les données.

Dans l'algorithme d'estimation, on suppose que la loi de probabilité $\mathbb{P}(\mathbf{r}, \mathbf{S})$ s'écrit de la manière suivante :

$$\mathbb{P}(\mathbf{r}, \mathbf{S}) = \mathbb{P}(\mathbf{r}) \prod_{t=1}^{T} \mathbb{P}(\mathbf{s}(t)|r(t))$$

Où $\mathbb{P}(\mathbf{s}(t)|r(t) = 1)$ est telle que les deux composantes $u_i, i = 1, 2$ du vecteur $\mathbf{u}$ dans (2) sont distribuées selon $u_i \sim \frac{1}{2}\mathcal{L}(\lambda) + \frac{1}{2}\mathcal{N}(0, \sigma)$. En d'autres termes, lorsque r(t) = 1, chaque composante de $\mathbf{u}$ est modélisée par un mélange de densités de probabilité gaussiennes et de Laplace. La distribution choisie pour $\mathbf{u}$ permet de remédier aux ambiguïtés de permutations du modèle d'ACI. Finalement, lorsque r(t) =0, $\mathbb{P}(\mathbf{s}(t)|r(t) = 0)$ est modélisée par une distribution gaussienne.

4.2 Modèle du processus caché r

Dans la version la plus simplifiée de notre méthode, $\mathbf{r}$ est modélisé par un modèle de Bernoulli i.i.d. Nous avons donc $\mathbb{P}(\mathbf{r}) = \prod_{t=1}^{T} \mathbb{P}(r(t))$ avec $\mathbb{P}(r(t) = 0) = p$ et $\mathbb{P}(r(t) = 1) = 1 - p$. Notre méthode peut être étendue au cas où $\mathbf{r}$ est modélisée par une chaîne de Markov stationnaire. Dans ce cas, la distribution de probabilité de $\mathbf{r}$ est donnée par : $\mathbb{P}(\mathbf{r}) = \mathbb{P}(r(1)) \prod_{t=2}^{t=T} \mathbb{P}(r(t) \mid r(t-1))$ où $\mathbb{P}(r(t) \mid r(t-1))$ est donnée par une matrice de transition indépendante de t. Un intérêt de la modélisation par une chaîne de Markov réside dans sa plus grande généralité, tout en ayant la possibilité de calculer $\mathbb{P}(\mathbf{r} \mid \mathbf{X}; \theta)$ grâce à l'algorithme forward-backward [6], permettant ainsi l'application de la méthode ECI [7].

5 Expérimentations

Notre méthode est une combinaison des deux algorithmes ACI et ECI. L'estimateur en données complètes $\hat{\theta}(\mathbf{r}, \mathbf{X})$ repose sur l'une des méthodes ACI classiques appliquée à un sous ensemble des échantillons et que l'on note `ACI` dans la description de notre algorithme donnée ci-dessous.

Initialisation des paramètres $\hat{\theta}^{[0]} = (\hat{\mathbf{B}}^{[0]}, \hat{p}^{[0]})$.
Pour $q = 1, 2, \ldots, q_{\max}$, répéter :
- calculer $\mathbb{P}(\mathbf{r} \mid \mathbf{X}; \hat{\theta}^{[q]})$ et tirer $\hat{\mathbf{r}}^{[q]}$ selon cette distribution,
- définir : $\hat{\mathcal{I}}_0^{[q]} = \{t \mid \hat{r}^{[q]}(t) = 0\}$ et $\hat{\mathbf{X}}_0^{[q]} = (x(t))_{t \in \hat{\mathcal{I}}_0^{[q]}}$
- $\hat{\mathbf{B}}^{[q+1]} = \texttt{ACI}(\hat{\mathbf{X}}_0^{[q]})$
- mettre à jour les paramètres du processus $\mathbf{r}$.

Dans le cas le plus simple où $\mathbf{r}$ est i.i.d la dernière étape de l'algorithme décrit ci-dessus consiste à mettre à jour la valeur estimée de p donnée par $\hat{p}^{[q+1]} = \frac{1}{T} \sum_{t=1}^{T} \mathbb{P}(r(t) = 0 \mid \mathbf{x}(t), \hat{\theta}^{[q]})$. On précise donc que nous estimons p par l'equation (3) tandis que $\mathbf{B}$ est estimée par l'equation (4) avec $K = 1$.

5.1 Les paramètres des données

Dans nos simulations, le processus $\mathbf{r}$ est soit une suite i.i.d avec $\mathbb{P}(r(t) = 0) = p$ et $\mathbb{P}(r(t) = 1) = 1 - p$, soit une chaîne de Markov. Les sources sont générées suivant le modèle décrit dans le paragraphe 2.3 sur les données dépendantes et mélangées selon l'équation (1). La matrice de mélange est tirée aléatoirement. Nous comparons la qualité de séparation obtenue par notre méthode (ACI+ECI) avec celle obtenue par une méthode classique d'ACI qui ignore la dépendance des sources. Pour $N = 2$, nous présentons les valeurs moyennes sur 1000 réalisations de Monte-Carlo de l'erreur quadratique moyenne (EQM) en fonction du paramètre p correspondant à la probabilité d'indépendance des sources. Nous avons choisi l'algorithme Jade [2] comme méthode de séparation classique,

et le nombre d'itérations de l'algorithme ECI a été fixé empiriquement à $q_{max} = 20$.

5.2 Variable latente i.i.d.

Dans un premier temps, nous avons testé notre méthode dans le cas où le processus $\mathbf{r}$ est i.i.d. Sur la figure 2 sont tracées les valeurs de l'EQM en fonction des valeurs de la probabilité p d'indépendance des sources, sur des données dont les composantes dépendantes sont générées selon le *Cas 1* du paragraphe 2.3 avec $\lambda = \sqrt{2}$ et $\sigma = 1$. Nous comparons les différentes méthodes : Jade appliquée sur $\mathbf{X}$, notre méthode (ECI+ACI) et Jade appliquée sur $\mathbf{X}_0$ noté (Jade+r). Le dernier cas est un cas idéal que nous ne pouvons appliquer que si nous disposons des données complètes. La figure 2 montre clairement l'efficacité de notre méthode de séparation pour des valeurs de p supérieures à 0.3. L'amélioration de la qualité de séparation est nettement observée pour des valeurs de p comprises entre 0.4 et 0.9.

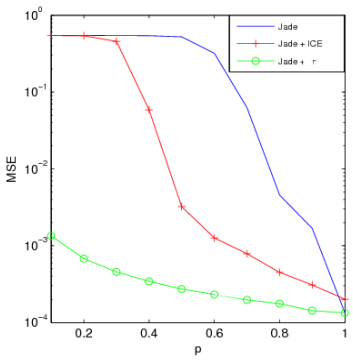


FIG. 2 – Pour $T = 5000$ échantillons, comparaison en fonction de p des méthodes de séparation : Jade seul / Jade+ estimation de $\mathbf{r}$ par ECI / Jade + $\mathbf{r}$ connu (données complètes). En abscisse, $p = \mathbb{P}(r(t) = 0)$ (sources indépendantes pour $p = 1$).

Dans un deuxième temps, nous avons étudié le cas où la distribution *a posteriori* considérée pour $(\mathbf{r}, \mathbf{S})$ est différente de celle vérifiée effectivement par les données. L'étude a été effectuée pour différentes tailles d'échantillon et pour une valeur de p égale à 0.5. Les données considérées sont telles que les composantes dépendantes sont générées selon le *Cas 2*. Les distributions supposées pour $(\mathbf{r}, \mathbf{S})$ sont les mêmes que pour l'expérience précédente, avec cette fois $\lambda = 5$. Les résultats sont présentés dans le tableau 1 et montrent l'efficacité et la robustesse de notre méthode vis à vis d'une erreur de modélisation de $(\mathbf{r}, \mathbf{S})$.

5.3 Variable latente de Markov

Nous avons testé notre méthode dans le cas où le processus $\mathbf{r}$ est une chaîne de Markov stationnaire avec la matrice de transition $\left(\begin{smallmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{smallmatrix}\right)$. Nous avons comparé les valeurs de l'EQM obtenues avec notre méthode "Jade+ECI" basée sur le modèle de chaîne de Markov cachée pour le processus $\mathbf{r}$ et "Jade+ECI" basée sur un modèle i.i.d pour $\mathbf{r}$. Les résultats du tableau 2 montrent que la qualité de séparation est améliorée lorsque l'on prend en compte la dépendance markovienne du processus $\mathbf{r}$.

T	Jade	Jade + ECI	Jade + $\mathbf{r}$
1000	$6.0\ 10^{-1}$	$5.4\ 10^{-2}$	$1.3\ 10^{-3}$
2000	$6.2\ 10^{-1}$	$3.0\ 10^{-2}$	$0.6\ 10^{-3}$
5000	$6.4\ 10^{-1}$	$2.7\ 10^{-2}$	$0.2\ 10^{-3}$
10000	$6.3\ 10^{-1}$	$2.7\ 10^{-2}$	$0.1\ 10^{-3}$

TAB. 1 – Valeurs moyennes de l'EQM pour différentes tailles de l'échantillon et pour $p = 0.5$.

T	modèle de Markov	modèle i.i.d.
1000	$1.2\ 10^{-1}$	$1.4\ 10^{-1}$
5000	$1.1\ 10^{-2}$	$2.6\ 10^{-2}$

TAB. 2 – Cas où $\mathbf{r}$ est généré selon une chaîne de Markov : EQM pour un modèle i.i.d. et un modèle de Markov de $\mathbf{r}$ dans l'algorithme.

Références

[1] J.-F. Cardoso. Multidimensional independent component analysis. In Proc. ICASSP '98. Seattle, 1998.

[2] J.-F. Cardoso and A. Souloumiac. Blind beamforming for non gaussian signals. IEE-Proceeding-F, 140(6) :362–370, December 1993.

[3] M. Castella and P. Comon. Blind separation of instantaneous mixtures of dependent sources. In Proc. of ICA'07, volume 4666 of LNCS, pages 9–16, London, UK, September 2007.

[4] P. Comon. Contrasts, independent component analysis, and blind deconvolution. Int. Journal Adapt. Control Sig. Proc., 18(3) :225–243, Apr. 2004.

[5] P. Comon and C. Jutten, editors. Handbook of Blind Source Separation, Independent Component Analysis and Applications. Academic Press, 2010.

[6] P. A. Devijver. Baum's forward-backward algorithm revisited. Pattern Recognition Letters, 3 :369–373, 1985.

[7] P. Lanchantin, J. Lapuyade-Lahorgue, and W. Pieczynski. Unsupervised segmentation of randomly switching data hidden with non-gaussian correlated noise. Signal Processing, 91 :163–175, February 2011.

[8] A. Peng and W. Pieczynski. Adaptive mixture estimation and unsupervised local bayesian image segmentation. CVGIP : Graphical Models and Image Processing, 57(5) :389–399, 1995.