

# Séparation de sources par une approche bayésienne hiérarchique

Hichem SNOUSSI<sup>1</sup>, Ali MOHAMMAD-DJAFARI<sup>1</sup>

<sup>1</sup>Laboratoire des Signaux et Systèmes (CNRS – SUPÉLEC – UPS)  
SUPÉLEC, Plateau de Moulon, 91192 Gif-sur-Yvette Cedex, France  
snoussi@lss.supelec.fr, djafari@lss.supelec.fr

**Résumé** – Dans cette contribution, nous présentons un nouvel algorithme de séparation de sources dans le cas d'un mélange linéaire instantané bruité dans un cadre bayésien. La distribution *a priori* des sources est modélisée par un mélange de gaussiennes et celle de la matrice de mélange par une gaussienne. Le mélange de gaussiennes est modélisé hiérarchiquement grâce à l'introduction de variables cachées représentant les étiquettes du mélange. Nous considérons la distribution *a posteriori* des sources, de la matrice de mélange, des étiquettes du mélange gaussien et des hyperparamètres du mélange avec des *a priori* appropriés afin d'éliminer la dégénérescence des variances et nous proposons un algorithme itératif pour estimer conjointement les sources, la matrice de mélange et les hyperparamètres.

**Abstract** – In this contribution, we present a new algorithm for source separation in the case of noisy instantaneous linear mixture, within the Bayesian statistical framework. The source distribution prior is modeled by a mixture of Gaussians and the mixing matrix elements distributions by Gaussians. We model the mixture of Gaussians hierarchically by mean of hidden variables representing the labels of the mixture. Then, we consider the joint *a posteriori* distribution of sources, mixing matrix elements, labels of the mixture and other parameters of the mixture with appropriate prior probability laws to eliminate degeneracy of the likelihood function of variance parameters and we propose an iterative algorithm to estimate jointly sources, mixing matrix and hyperparameters.

## 1 Introduction

On considère un mélange linéaire instantané de  $n$  sources. Les observations sont entachées d'un bruit blanc additif gaussien :

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \epsilon(t), \quad t = 1, \dots, T \quad (1)$$

avec  $\mathbf{x}(t)$  le vecteur ( $m \times 1$ ) des observations,  $\mathbf{s}(t)$  le vecteur ( $n \times 1$ ) des sources dont les composantes doivent être séparées,  $\mathbf{A}$  est la matrice de mélange ( $m \times n$ ) et  $\epsilon(t)$  représente les erreurs de modélisation le bruit affectant les mesures. Nous supposons que le bruit est statistiquement indépendant des sources, centré, blanc et gaussien de variance connue  $\sigma_\epsilon^2 \mathbf{I}$ . On note  $\mathbf{s}_{1..T}$  la matrice  $n \times T$  des sources et  $\mathbf{x}_{1..T}$  la matrice  $m \times T$  des observations. Le problème de séparation de sources est composé de deux sous-problèmes : Reconstruction des sources et identification de la matrice de mélange. Trois directions peuvent être explorées :

1. *Apprentissage supervisé* : Identification de la matrice  $\mathbf{A}$  connaissant une séquence d'apprentissage et l'utiliser pour reconstruire les sources.
2. *Identification aveugle* : Estimer  $\mathbf{A}$  dans un premier temps à partir des observations  $\mathbf{x}_{1..T}$  et reconstruire les sources dans un deuxième temps.
3. *Estimation conjointe* : Estimer conjointement  $\mathbf{A}$  et  $\mathbf{s}_{1..T}$ .

Dans ce travail, nous allons explorer la troisième direction.

Le papier est organisé comme suit : Nous commençons dans la section 2 par introduire l'approche bayésienne pour

la séparation de sources. Nous présentons l'algorithme JMAP connaissant les hyperparamètres. Nous introduisons, dans la section 3, la modélisation hiérarchique des sources à l'aide des variables cachées représentant les étiquettes du mélange gaussien. Dans la section 4, nous incluons l'estimation des hyperparamètres dans l'algorithme JMAP en profitant de la structure hiérarchique de l'*a priori*. Nous présentons quelques résultats de simulations dans la section 5 et nous concluons dans la section 6.

## 2 Approche bayésienne

Connaissant les données  $\mathbf{x}_{1..T}$ , la distribution *a posteriori* conjointe des sources  $\mathbf{s}_{1..T}$  et de la matrice de mélange  $\mathbf{A}$  est :

$$p(\mathbf{A}, \mathbf{s}_{1..T} | \mathbf{x}_{1..T}) \propto p(\mathbf{x}_{1..T} | \mathbf{A}, \mathbf{s}_{1..T}) p(\mathbf{A}) p(\mathbf{s}_{1..T}) \quad (2)$$

où  $p(\mathbf{x}_{1..T} | \mathbf{A}, \mathbf{s}_{1..T})$ ,  $p(\mathbf{s}_{1..T})$  et  $p(\mathbf{A})$  modélisent les connaissances *a priori* sur le bruit, sur les sources et sur la matrice de mélange. Nous avons, désormais, trois directions :

- (i) Intégrer (2) par rapport à  $\mathbf{s}_{1..T}$  pour obtenir la marginale en  $\mathbf{A}$  à partir de laquelle on déduit une estimation de la matrice de mélange [1].
- (ii) Intégrer (2) par rapport à  $\mathbf{A}$  pour obtenir la marginale en  $\mathbf{s}_{1..T}$  à partir de laquelle on déduit une estimée des sources.
- (iii) Estimer conjointement  $\mathbf{A}$  et  $\mathbf{s}_{1..T}$ .

## 2.1 Choix des distributions *a priori*

Les sources sont modélisées par un mélange de gaussiennes [2] :

$$p(s_j) = \sum_{i=1}^{q_j} \alpha_{ji} \mathcal{N}(m_{ji}, \sigma_{ji}^2), \quad j = 1..n \quad (3)$$

L'hyperparamètre  $q_j$  est supposé connu.

Afin de tenir compte d'une éventuelle information sur la matrice de mélange, nous avons choisi un *a priori* gaussien pour chaque coefficient de la matrice  $\mathbf{A}$  [1] :

$$p(\mathbf{A}_{ij}) = \mathcal{N}(\mathbf{M}_{ji}, \sigma_{a,ij}^2) \quad (4)$$

Ce qui peut être interprété comme connaissant chaque élément ( $\mathbf{M}_{ji}$ ) avec une incertitude ( $\sigma_{a,ij}^2$ ). On note ici l'avantage d'estimer directement la matrice  $\mathbf{A}$  et non la matrice séparatrice  $\mathbf{B}$  (inverse de  $\mathbf{A}$ ). Cette approche possède au moins deux avantages : (i)  $\mathbf{A}$  peut être rectangulaire ( $n \neq m$ ), (ii) en pratique, on possède des informations *a priori* sur la matrice de mélange et non sur son inverse qui peut ne pas exister.

## 2.2 Algorithme JMAP

Nous proposons un algorithme itératif pour estimer conjointement  $\mathbf{s}_{1..T}$  et  $\mathbf{A}$  en maximisant le logarithme de la distribution *a posteriori* par relaxation :

$$\begin{cases} \widehat{\mathbf{s}}_{1..T}^{(k)} &= \arg \max_{\mathbf{s}_{1..T}} \ln p(\widehat{\mathbf{A}}^{(k-1)}, \mathbf{s}_{1..T} | \mathbf{x}_{1..T}) \\ \widehat{\mathbf{A}}^{(k)} &= \arg \max_{\mathbf{A}} \ln p(\mathbf{A}, \widehat{\mathbf{s}}_{1..T}^{(k)} | \mathbf{x}_{1..T}) \end{cases}$$

Dans la suite, on suppose que les sources sont spatialement indépendantes. Nous faisons aussi l'hypothèse que les sources sont temporellement blanches bien que cette hypothèse ne soit pas nécessaire dans notre approche. L'optimisation est donc faite indépendamment à chaque instant  $t$  :

$$\widehat{\mathbf{s}}(t)^{(k+1)} = \arg \max_{\mathbf{s}(t)} \{ \ln p(\mathbf{x}(t) | \widehat{\mathbf{A}}^{(k)}, \mathbf{s}) + \sum_{j=1}^n \ln p_j(\mathbf{s}_j(t)) \}$$

La distribution *a posteriori* de  $\mathbf{s}$  est un mélange de  $\prod_{j=1}^n q_j$  gaussiennes, menant ainsi à un coût de calcul élevé. Ce qui nous a conduit à introduire une étape de relaxation dans l'estimation des composantes du vecteur  $\mathbf{s}$  :

$$\widehat{\mathbf{s}}_j(t)^{(k+1)} = \arg \max_{\mathbf{s}_j(t)} \left\{ \ln p(\mathbf{s}_j(t) | \mathbf{x}(t), \widehat{\mathbf{A}}^{(k)}, \widehat{\mathbf{s}}_{l < j}(t)^{(k+1)}, \widehat{\mathbf{s}}_{l > j}(t)^{(k)}) \right\}$$

La distribution *a posteriori* de  $\mathbf{s}_j$  est un mélange de  $q_j$  gaussiennes :  $\sum_{z=1}^{q_j} \alpha'_{jz} \mathcal{N}(m'_{jz}, \sigma'_{jz}{}^2)$ . Si les moyennes  $m'_{jz}$  ne sont trop proches les unes des autres, on se trouve dans le cas d'une distribution multi-modale et pour estimer  $\mathbf{s}_j$ , on choisit  $m'_{jz}$  pour lequel le rapport  $\frac{\alpha'_{jz}}{\sigma'_{jz}}$  est le plus grand.

Après la remise à jour des sources  $\mathbf{s}_{1..T}$ , l'estimée de  $\mathbf{A}$  à cette même itération est obtenue en maximisant :

$$J(\mathbf{A}) = \sum_{t=1}^T \ln p(\mathbf{x}(t) | \mathbf{A}, \widehat{\mathbf{s}}^{k+1}(t)) + \ln p(\mathbf{A}(t)) + cte$$

qui est une fonction quadratique en  $\mathbf{A}$ . En annulant la dérivée et en définissant  $\mathbf{\Lambda}_{i,j} = \frac{\sigma_c^2}{\sigma_{a,i,j}^2}$ , la remise à jour de  $\mathbf{A}$  est :

$$\widehat{\mathbf{A}}^{k+1} = \text{Mat}([\mu + \mathbf{S}^*]^{-1} [\mu \text{Vect}(\mathbf{M}) + \text{Vect}(\mathbf{x}_{1..T}(\widehat{\mathbf{s}}_{1..T}^{k+1})^*)])$$

où  $\text{Vect}$  est l'opérateur transformant une matrice en un vecteur et  $\text{Mat}$  l'opérateur inverse.  $\mu$  est la matrice diagonale ( $nm \times nm$ ) dont le vecteur diagonale est  $\text{Vect}((\mathbf{\Lambda}_{i,j})_{i=1..m, j=1..n})$  et  $\mathbf{S}^*$  la matrice ( $nm \times nm$ ) formée de blocs diagonales  $\widehat{\mathbf{s}}_{1..T} \widehat{\mathbf{s}}_{1..T}^*$  estimés à l'itération ( $k+1$ ).

Nous allons maintenant examiner de plus près l'expression de la distribution *a posteriori* des sources. C'est une distribution multi-modale si les variances des gaussiennes sont petites devant les distances entre les moyennes. Le maximum de la distribution, dans ce cas, ne correspond pas en général à la moyenne de la gaussienne la plus probable. Nous allons, dans un premier temps, sélectionner la gaussienne la plus probable à chaque instant  $t$  (classification) puis estimer la source  $\mathbf{s}_t$  comme la moyenne *a posteriori* de cette gaussienne. Cela nous conduit à l'introduction des variables cachées et à la modélisation hiérarchique.

## 3 Modélisation hiérarchique

La distribution *a priori* de la composante  $s_j$  est :

$$p(s_j) = \sum_{i=1}^{q_j} \alpha_{ji} \mathcal{N}(m_{ji}, \sigma_{ji}^2)$$

On considère la variable cachée  $z_j$  prenant ses valeurs dans l'ensemble discret  $\mathcal{Z}_j = (1, \dots, q_j)$  avec des probabilités  $\alpha_{ji} = p(z_j = i)$ . Conditionnellement à  $z_j = i$ ,  $\mathbf{s}_j$  suit la loi normale  $\mathcal{N}(m_{ji}, \sigma_{ji}^2)$ . On peut étendre cette notion pour le cas vectoriel en considérant le vecteur des étiquettes  $\mathbf{z} = [z_1, \dots, z_n]$  à valeurs dans l'ensemble  $\mathcal{Z} = \prod_{j=1}^n \mathcal{Z}_j$ . La distribution de  $\mathbf{s}$  connaissant  $\mathbf{z}$  est la loi normale  $p(\mathbf{s} | \mathbf{z}) = \mathcal{N}(\mathbf{m}_z, \mathbf{\Gamma}_z)$  avec :

$$\mathbf{m}_z = [m_{1z_1}, m_{2z_2}, \dots, m_{nz_n}] \quad (5)$$

$$\mathbf{\Gamma}_z = \mathbf{diag}(\sigma_{1z_1}^2, \sigma_{2z_2}^2, \dots, \sigma_{nz_n}^2) \quad (6)$$

La loi *a priori* marginale de  $\mathbf{s}$  est un mélange de  $\prod_{j=1}^n q_j$  gaussiennes :

$$p(\mathbf{s}) = \sum_{\mathbf{z} \in \mathcal{Z}} p(\mathbf{z}) p(\mathbf{s} | \mathbf{z}) \quad (7)$$

On peut ré-interpréter ce mélange comme un ensemble discret de couples  $(\mathcal{N}_z, p(\mathbf{z}))$ . Les sources qui appartiennent à cette classe sont générées de la façon suivante : générer les variables cachées  $\mathbf{z} \in \mathcal{Z}$  selon  $p(\mathbf{z})$  et ensuite générer  $\mathbf{s}$  selon la normale  $\mathcal{N}_z$ . Ce modèle peut être étendu en considérant un ensemble dénombrable ou continue de valeurs de  $\mathbf{z}$ .

### 3.1 Distribution *a posteriori* des sources

Dans ce paragraphe, on suppose que la matrice de mélange est connue. La loi *a posteriori* de  $\mathbf{s}$ ,  $\mathbf{z}$  et  $\mathbf{x}$  peut être

factorisée sous deux formes :  $p(\mathbf{s}, \mathbf{z}, \mathbf{x}) = p(\mathbf{x}|\mathbf{s})p(\mathbf{s}|\mathbf{z})p(\mathbf{z})$  ou  $p(\mathbf{s}, \mathbf{z}, \mathbf{x}) = p(\mathbf{s}|\mathbf{x}, \mathbf{z})p(\mathbf{z}|\mathbf{x})p(\mathbf{x})$ . Donc, la loi marginale *a posteriori* possède deux formes :

$$p(\mathbf{s}|\mathbf{x}) = \sum_{\mathbf{z} \in \mathcal{Z}} \frac{p(\mathbf{z})p(\mathbf{x}|\mathbf{s})p(\mathbf{s}|\mathbf{z})}{p(\mathbf{x})} \quad (8)$$

ou

$$p(\mathbf{s}|\mathbf{x}) = \sum_{\mathbf{z} \in \mathcal{Z}} p(\mathbf{z}|\mathbf{x})p(\mathbf{s}|\mathbf{x}, \mathbf{z}) \quad (9)$$

On note, dans la seconde forme, que l'*a posteriori* reste dans la même famille que l'*a priori* (même expressions conditionnellement à  $\mathbf{x}$ ). La stratégie d'estimation est basée sur la remarque suivante : Les sources étant modélisées hiérarchiquement, on va les estimer hiérarchiquement ; on commence par estimer les variables cachées à partir de  $p(\mathbf{z}|\mathbf{x})$  puis on estime les sources à partir de  $p(\mathbf{s}|\mathbf{x}, \mathbf{z})$  qui est gaussienne.

La probabilité  $p(\mathbf{z}|\mathbf{x})$  est obtenue en intégrant par rapport à  $\mathbf{s}$  la loi *a posteriori* conjointe de  $\mathbf{z}$  et  $\mathbf{s}$  :

$$p(\mathbf{z}|\mathbf{x}) = \int p(\mathbf{z}, \mathbf{s}|\mathbf{x})d\mathbf{s} \propto p(\mathbf{z}) \int p(\mathbf{x}|\mathbf{s})p(\mathbf{s}|\mathbf{z})d\mathbf{s} \quad (10)$$

L'approche bayésienne nous permet d'exprimer notre information *a priori* via des modèles paramétriques. Cependant, en général, on ne connaît pas les paramètres de la loi *a priori*. Dans la prochaine section, nous allons voir comment on peut aussi estimer les hyperparamètres inconnus.

## 4 Estimation d'hyperparamètres

Les hyperparamètres considérés ici sont les poids  $\alpha_{jz}$ , les moyennes  $m_{jz}$  et les variances  $\frac{1}{\psi_{jz}}$  du mélange de gaussiennes :

$\mathbf{s}_j \sim \sum_{z=1}^{q_j} \alpha_{jz} \mathcal{N}\left(m_{jz}, \frac{1}{\psi_{jz}}\right)$ ,  $j = 1, \dots, n$ . Nous développons dans la suite une méthode d'estimation de ces hyperparamètres. L'idée principale est la suivante : Conditionnellement aux variables cachées  $(z_j)_{1..T} = [z_j(1), \dots, z_j(T)]$ , les hyperparamètres  $m_{jz}$  et  $\psi_{jz}$  pour  $z \in \mathcal{Z}_j = (1, \dots, q_j)$  sont les moyennes et variances d'une distribution gaussienne. A partir de  $(z_j)_{1..T} = [z_j(1), \dots, z_j(T)]$ , on effectue une partition de l'ensemble  $\mathcal{T} = [1, \dots, T]$  en sous-ensembles  $\mathcal{T}_z$  :

$$\mathcal{T}_z = \{t \mid z_j(t) = z\}, \quad z \in \mathcal{Z}_j \quad (11)$$

Ceci constitue l'étape de classification.

Supposons que la matrice de mélange  $\mathbf{A}$  et les composantes  $\mathbf{s}_{l \neq j}$  sont fixées et que la classification est effectuée. L'estimation de  $m_{jz}$  et  $\psi_{jz}$  se ramène au cas simple et connue de l'estimation des paramètres d'une gaussienne [3].

Soit  $\theta_{jz} = (m_{jz}, \psi_{jz})$ , la loi *a posteriori* jointe de  $\mathbf{s}_j$  et  $\theta_{jz}$  sachant  $z_j$  à chaque instant  $t$  est :

$$p(\mathbf{s}_j, \theta_{jz} \mid \mathbf{x}, z_j) \propto p(\mathbf{x} \mid \mathbf{s})p(\mathbf{s}_j \mid \theta_{jz}, z_j)p(\theta_{jz} \mid z_j) \quad (12)$$

$p(\mathbf{s}_j \mid \theta_{jz}, z_j)$  est une gaussienne de moyenne  $m_{jz}$  et de variance  $\frac{1}{\psi_{jz}}$ .

$p(\theta_{jz} \mid z_j) = p(\theta_{jz}) = p(m_{jz})p(\psi_{jz})$  est la loi *a priori* des

hyperparamètres. La loi *a posteriori* marginale de  $\theta_{jz}$  est obtenue à partir de l'expression précédente en intégrant par rapport à  $\mathbf{s}_j$  :

$$p(\theta_{jz} \mid \mathbf{x}, z_j) \propto p(\theta_{jz}) \int_{\mathbf{s}_j} p(\mathbf{x} \mid \mathbf{s})p(\mathbf{s}_j \mid \theta_{jz}, z_j) d\mathbf{s}_j. \quad (13)$$

En posant  $\phi_j = 1/\sigma_j^2$  et  $\psi_{jz} = 1/\sigma_{jz}^2$  :

$$p(\theta_{jz} \mid \mathbf{x}, z_j) \propto p(\theta_{jz}) \sqrt{\frac{\phi_j \psi_{jz}}{\phi_j + \psi_{jz}}} \times \exp\left[-\frac{1}{2} \frac{\phi_j \psi_{jz}}{\phi_j + \psi_{jz}} (m_{jz} - m_j)^2\right]$$

Notons que la vraisemblance est normale pour la moyenne  $m_{jz}$  et gamma pour  $\lambda_{jz} = (\phi_j \psi_{jz}) / (\phi_j + \psi_{jz})$ .

En prenant un *a priori* uniforme pour les poids et les moyennes, l'estimée de  $\alpha_{jz}$  et  $m_{jz}$  sont :

$$\hat{\alpha}_{jz}^{MAP} = \frac{\text{Card}(\mathcal{T}_z)}{T} \quad (14)$$

$$\hat{m}_{jz}^{MAP} = \frac{\sum_{t \in \mathcal{T}_z} m_j(t)}{T_z} \quad (15)$$

Pour les variances, nous choisissons un *a priori* inverse gamma  $\mathcal{G}(\alpha, \beta)$ . Ce choix est motivé par deux raisons : (i) c'est un *a priori* propre qui permet d'éliminer la dégénérescence des variances vers zéro (il a été montré dans [4] (cas sans bruit et sans mélange) et dans [5] que la vraisemblance n'est pas bornée ce qui cause la dégénérescence des variances vers zéro), (ii) c'est un *a priori* conjugué ce qui conduit à des expressions de ré-estimation simples. L'estimation de  $\psi_{jz}$  au sens du MAP est donc :

$$\hat{\psi}_{jz}^{MAP} = \frac{\alpha^{post} - 1}{\beta^{post}} \quad (16)$$

avec  $\alpha^{post} = \alpha + \frac{T_z}{2}$  et  $\beta^{post} = \beta + \frac{\sum_{t \in \mathcal{T}_z} (m_j(t) - \hat{m}_{jz}^{MAP})^2}{4}$ .

### 4.1 Algorithme JMAP

En incluant l'estimation des hyperparamètres, l'algorithme JMAP hiérarchique est composé de quatre étapes :

1. Estimation des variables cachées  $(\hat{z}_j)_{1..T}^{MAP}$  et calcul des partitions :  $\hat{\mathcal{T}}_z = \{t \mid (\hat{z}_j)^{MAP}(t) = z\}$ .
2. Remise à jour des hyperparamètres  $\hat{\alpha}_{jz}^{MAP}$ ,  $\hat{\psi}_{jz}^{MAP}$  et  $\hat{m}_{jz}^{MAP}$  selon les équations (14), (15) et (16).
3. Estimation des sources  $(\hat{\mathbf{s}})_{1..T}^{MAP}$  selon (2.2).
4. Estimation de la matrice de mélange  $(\hat{\mathbf{A}})^{MAP}$  (2.2).

## 5 Résultats de simulations

Afin de tester les performances de l'algorithme JMAP hiérarchique, nous considérons 1000 observations générées comme dans [6]. L'*a priori* des sources est un mélange de deux gaussiennes centrées en  $-1$  et  $1$  avec la même variance  $0.01$  et même poids  $0.5$ . La matrice de mélange inconnue est  $\mathbf{A} = \begin{pmatrix} 1 & -0.6 \\ 0.4 & 1 \end{pmatrix}$ . Le rapport signal sur

bruit SNR est 8 dB. L'indice de performance [7] est utilisé pour caractériser le résultat d'identification :

$$ind(S = \hat{\mathbf{A}}^{-1} \mathbf{A}) = \frac{1}{2} \sum_i \left( \sum_j \frac{|S_{ij}|^2}{\max_l |S_{il}|^2} - 1 \right) + \frac{1}{2} \sum_j \left( \sum_i \frac{|S_{ij}|^2}{\max_l |S_{lj}|^2} - 1 \right)$$

La figure 1 illustre les résultats de séparation. Les sources sont reconstruites avec une erreur quadratique moyenne EQM= 0.01. La figure 2a représente l'évolution de l'indice de performance au cours des itérations. On note la convergence de JMAP à partir de l'itération 30 vers une valeur satisfaisante de  $-40$  dB. Pour un SNR= 15 dB, les algorithmes PWS, NS [6] et EASI [8] convergent vers une valeur de  $-35$  dB pour 6000 observations. Les figures 2c et 2d illustrent les résultats de l'identification des hyperparamètres. On note la convergence de l'algorithme vers les valeurs originales ( $-1$  pour  $m_{11}$  et 0.01 pour  $\sigma_{11}^2 = 1/\psi_{11}$ ). Afin de valider l'idée de la classification des données avant l'estimation des hyperparamètres, nous pouvons visualiser l'évolution des erreurs de classification (nombre de données mal classifiées). La figure 2b montre la convergence de ces erreurs vers zéro à partir de l'itération 5. Ainsi, après cette itération, l'identification des hyperparamètres est effectuée sur les données correctement classées. L'estimation de  $\alpha_{jz}$ ,  $m_{jz}$  et  $\psi_{jz}$  prend en compte uniquement les données appartenant à la bonne classe et n'est pas ainsi corrompue par des données mal classées. La comparaison de l' algorithme JMAP avec l'algorithme EM a été réalisée dans [1] où nous avons noté que l'algorithme JMAP est moins sensible au choix de l'initialisation que l'algorithme EM.

## 6 Conclusion

Nous avons proposé une solution pour le problème de séparation de sources dans un cadre bayésien. L'intérêt de cette approche réside dans les aspects suivants :

- Prise en compte d'incertitude de modèle et du bruit.
- Introduction des distributions *a priori* pour la matrice de mélange et pour les hyperparamètres. Ceci est motivé par deux raisons différentes : l'*a priori* sur la matrice de mélange nous permet d'exploiter des informations *a priori* sur problème physique direct et l'*a priori* sur les variances permet d'éliminer la dégénérescence et de borner la vraisemblance.

Nous exploitons la structure à données manquantes de l'*a priori* des sources pour développer une stratégie d'estimation hiérarchique.

## Références

- [1] H. Snoussi et A. Mohammad-Djafari, « Bayesian source separation with mixture of gaussians prior for sources and gaussian prior for mixture coefficients », in *Bayesian Inference and Maximum Entropy Methods*, A. Mohammad-Djafari, Ed., Gif-sur-Yvette, juillet 2000, Proc. of MaxEnt, pp. 388–406, Amer. Inst. Physics.
- [2] E. Moulines, J. Cardoso et E. Gassiat, « Maximum likelihood for blind separation and deconvolution of noisy si-

gnals using mixture models », in *ICASSP-97*, Munich, Allemagne, avril 1997.

- [3] A. Mohammad-Djafari, « Model selection for inverse problems : Best choice of basis function and model order selection », in *Bayesian Inference and Maximum Entropy Methods*, J. R. G. Erikson et C. Smith, Eds. juillet 1999, p. to appear on may 2001, à paraître dans Amer. Inst. Physics.
- [4] A. Ridolfi et J. Idier, « Penalized maximum likelihood estimation for univariate normal mixture distributions », in *Actes 17<sup>e</sup> coll. GRETSI*, Vannes, septembre 1999, pp. 259–262.
- [5] H. Snoussi et A. Mohammad-Djafari, « Dégénérescences des estimateurs MV en séparation de sources », Technical report ri-s0010, GPI-L2S.
- [6] O. Macchi et E. Moreau, « Adaptive unsupervised separation of discrete sources », in *Signal Processing*, 1999, vol. 73, pp. 49–66.
- [7] E. Moreau et O. Macchi, « High-order contrasts for self-adaptative source separation », in *Adaptative Control Signal Process. 10*, 1996, pp. 19–46.
- [8] J. Cardoso et B. Label, « Equivariant adaptative source separation », *Signal Processing*, vol. 44, pp. 3017–3030, 1996.

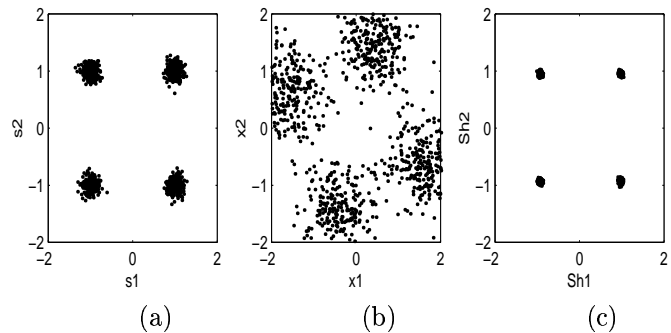


Figure 1- Résultats de séparation avec SNR= 8 dB  
(a) Distribution des sources,  
(b) signaux mélangés et (c) sources séparées

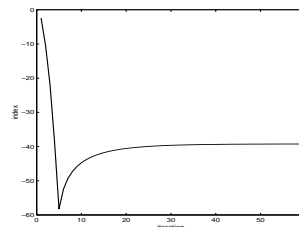


Figure 2a- Evolution  
de l'indice de performance

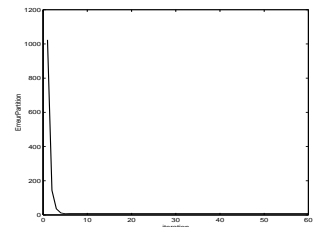


Figure 2b-Evolution  
des erreurs de classification

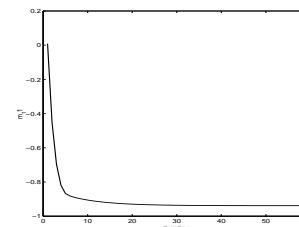


Figure 2c- Identification de  
 $m_{11}$

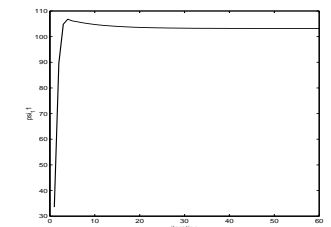


Figure 2d- Identification de  
 $\psi_{11}$