

De la Séparation de Sources à l'Analyse en Composantes Indépendantes

Christian JUTTEN, PIERRE COMON

version 1.2, janvier 2005

Approche bayésienne en séparation de sources

Avant de détailler l'utilisation de cette approche dans le cadre de la séparation de sources, il est nécessaire de préciser un point important : Lorsque nous considérons un signal $x(t)$ (une quantité x qui varie avec un index t), ou lorsque nous considérons un paramètre θ qui représente une quantité quelconque, de dire qu'il s'agit d'une variable aléatoire ou non, n'a pas de sens réel. Lorsque nous attribuons une loi de probabilité à ce signal (ou à ce paramètre), c'est notre connaissance sur ce signal (ou sur ce paramètre) qui est modélisée. Cette attribution ne change pas la réalité des choses. Cette loi de probabilité reflète juste notre état de connaissance sur ce signal ou sur ce paramètre. Un autre point important est que la règle de Bayes est juste l'outil de base pour mettre à jour notre état de connaissance *a priori* (avant avoir observé les données) en état de connaissance *a posteriori* (après avoir observé les données) en utilisant la vraisemblance qui mesure l'apport des observations. être bayésien est donc une démarche pour résoudre un problème. Ce chapitre est justement une démonstration de l'application de cette démarche pour proposer des solutions aux problèmes de la séparation de sources.

0.1. Introduction et notations

Avec ces remarques préliminaires, et le fait que nous nous intéressons à la fois à la séparation des signaux et des images, dans la suite de ce chapitre, les notations suivantes seront utilisées :

- $x_i(t)$ ou $x_i(\mathbf{r})$ l'observation faite par le capteur numéro i à l'instant t ou à la position \mathbf{r} , $i = 1, \dots, m$
- $s_j(t)$ ou $s_j(\mathbf{r})$ la source numéro j à l'instant t ou à la position \mathbf{r} , $j = 1, \dots, n$
- $\mathbf{x}(t)$ ou $\mathbf{x}(\mathbf{r})$ le vecteur des observations à l'instant t ou à la position \mathbf{r}
- $\mathbf{s}(t)$ ou $\mathbf{s}(\mathbf{r})$ le vecteur des sources à l'instant t ou à la position \mathbf{r}

4 Séparation de Sources

$\mathcal{T} = \{1, \dots, T\}$ représente un intervalle du temps

$\mathcal{R} = \{\mathbf{r}_i, = (x_i, y_i)\}$ représente l'ensemble des positions des pixels d'une image

$\mathbf{X} = \{\mathbf{x}(t), t \in \mathcal{T}\}$ ou $\{\mathbf{x}(\mathbf{r}), \mathbf{r} \in \mathcal{R}\}$ un tableau représentant toutes les observations

$\mathbf{S} = \{\mathbf{s}(t), t \in \mathcal{T}\}$ ou $\{\mathbf{s}(\mathbf{r}), \mathbf{r} \in \mathcal{R}\}$ un tableau représentant toutes les sources

Nous avons alors, dans le cas des signaux 1D :

$$x_i(t) = \sum_{j=1}^n A_{i,j} s_j(t) + \epsilon_i(t), t \in \mathcal{T} \longrightarrow \mathbf{x}(t) = \mathbf{A} \mathbf{s}(t) + \boldsymbol{\epsilon}(t) \quad [1]$$

et dans le cas des signaux n-D, et en particulier dans le cas des images :

$$x_i(\mathbf{r}) = \sum_{j=1}^n A_{i,j} s_j(\mathbf{r}) + \epsilon_i(\mathbf{r}), \mathbf{r} \in \mathcal{R} \longrightarrow \mathbf{x}(\mathbf{r}) = \mathbf{A} \mathbf{s}(\mathbf{r}) + \boldsymbol{\epsilon}(\mathbf{r}) \quad [2]$$

Dans les deux cas, nous avons

$$\mathbf{X} = \mathbf{A} \mathbf{S} + \mathbf{E} \quad [3]$$

où $\mathbf{E} = \{\boldsymbol{\epsilon}(t), t \in \mathcal{T}\}$ ou $\{\boldsymbol{\epsilon}(\mathbf{r}), \mathbf{r} \in \mathcal{R}\}$ représente les erreurs de la modélisation et/ou de mesures, communément appelé le bruit.

Par la suite, par défaut, nous considérons le cas 1D, et seulement lorsqu'il est nécessaire, nous précisons les différences et les spécificité du cas n-D.

0.2. Schéma général de l'approche bayésienne

Le schéma général de l'inférence bayésienne dans un problème de séparation de sources peut se résumer par les étapes suivantes :

– Bien décrire le modèle qui lie les observations aux sources et en déduire la loi $p(\mathbf{X}|\mathbf{A}, \mathbf{S})$ qui est appelée la vraisemblance des inconnues. Dans un modèle du type $\mathbf{X} = \mathbf{A} \mathbf{S} + \mathbf{E}$, où on suppose le bruit additif, on a $p(\mathbf{X}|\mathbf{A}, \mathbf{S}) = p_{\mathbf{E}}(\mathbf{X} - \mathbf{A} \mathbf{S})$ où $p_{\mathbf{E}}$ est la loi du bruit.

– Attribuer des loi *a priori* à tous les inconnus du problème, c'est-à-dire aux sources $p(\mathbf{S})$ et à la matrice de séparation $p(\mathbf{A})$ pour décrire notre connaissance *a priori* ou pour contraindre l'espace des solutions pour rendre le problème bien posé.

– En utilisant la règle de Bayes, déduire la loi *a posteriori* $p(\mathbf{A}, \mathbf{S}|\mathbf{X})$

$$p(\mathbf{A}, \mathbf{S}|\mathbf{X}) = \frac{p(\mathbf{X}|\mathbf{A}, \mathbf{S}) p(\mathbf{A}) p(\mathbf{S})}{p(\mathbf{X})} \propto p(\mathbf{X}|\mathbf{A}, \mathbf{S}) p(\mathbf{A}) p(\mathbf{S}) \quad [4]$$

où

$$p(\mathbf{X}) = \int p(\mathbf{X}|\mathbf{A}, \mathbf{S}) p(\mathbf{A}) p(\mathbf{S}) d\mathbf{S} \quad [5]$$

Cette loi résume l'état de notre connaissance sur les inconnues du problème après avoir pris en compte les *a priori* et les observations.

– Utiliser cette loi *a posteriori* pour définir une solution ou un ensemble de solutions (des estimateurs pour \mathbf{A} et pour \mathbf{S}) pour le problème.

Notons que la règle de Bayes nous fournit l'expression de la loi *a posteriori* d'une grandeur inconnue \mathbf{Z} après avoir observé les données \mathbf{X} . Si \mathbf{Z} était juste un scalaire ou un vecteur de dimension deux, on pourrait représenter la courbe ou la surface de sa loi $p(\mathbf{Z}|\mathbf{X})$. Mais, en pratique, \mathbf{Z} est un vecteur de très grande dimension, et l'utilisation directe et la représentation de $p(\mathbf{Z})$ n'est pas facile. On est alors amené à résumer $p(\mathbf{Z}|\mathbf{X})$ par quelques grandeurs comme par exemple, son mode, sa moyenne, sa médiane, les modes de ses lois marginales, etc. [JAY 68, ROB 92, ROB 97]. C'est ainsi qu'on se trouve face aux estimateurs dites le maximum *a posteriori* (MAP), la moyenne *a posteriori* (MP), etc.

Dans un problème de séparation de sources, nous avons deux jeux d'inconnus : \mathbf{A} et \mathbf{S} . C'est pourquoi, à partir d'ici, trois directions s'ouvrent à nous :

1. Estimation jointe :

On cherche à calculer $(\hat{\mathbf{A}}, \hat{\mathbf{S}})$ en utilisant la loi jointe $p(\mathbf{A}, \mathbf{S}|\mathbf{X})$, par exemple

$$(\hat{\mathbf{A}}, \hat{\mathbf{S}}) = \arg \max_{(\mathbf{A}, \mathbf{S})} \{J(\mathbf{A}, \mathbf{S}) = \ln p(\mathbf{A}, \mathbf{S}|\mathbf{X})\} \quad [6]$$

2. Estimation de \mathbf{A} :

On s'intéresse à l'identification de la matrice de mélange en utilisant la loi $p(\mathbf{A}|\mathbf{X})$ qui peut être obtenue en intégrant $p(\mathbf{A}, \mathbf{S}|\mathbf{X})$ par rapport à \mathbf{S} , par exemple :

$$\hat{\mathbf{A}} = \arg \max_{\mathbf{A}} \{J(\mathbf{A}) = \ln p(\mathbf{A}|\mathbf{X})\} \quad [7]$$

3. Estimation de \mathbf{S} :

On s'intéresse à la séparation des sources proprement dite en utilisant la loi $p(\mathbf{S}|\mathbf{X})$

qui peut être obtenue en intégrant $p(\mathbf{A}, \mathbf{S}|\mathbf{X})$ par rapport à \mathbf{A} , par exemple :

$$\hat{\mathbf{S}} = \arg \max_{\mathbf{S}} \{J(\mathbf{S}) = \ln p(\mathbf{S}|\mathbf{X})\} \quad [8]$$

Dans chacun de ces cas, le MAP peut aussi être remplacé par la moyenne *a posteriori* ou par d'autres estimateurs.

0.3. Sources indépendantes et blanches et liens avec ACP, ACI et autres méthodes classiques

Pour illustrer le schéma général de l'approche bayésienne et montrer son lien avec les méthodes classiques de la séparation de sources, rien de mieux que de l'appliquer dans un cas où on puisse faire des calculs d'une manière simple. Prenons le modèle $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \boldsymbol{\epsilon}(t)$ et faisons tout d'abord l'hypothèse que les bruits $\boldsymbol{\epsilon}$ sont indépendants, blancs, centrés, gaussiens de variances connues $\sigma_{\epsilon_i} = \sigma_{\epsilon}$. On a alors

$$p(\mathbf{X}|\mathbf{A}, \mathbf{S}) = \prod_t \mathcal{N}(\mathbf{x}(t) - \mathbf{A}\mathbf{s}(t), \boldsymbol{\Sigma}_{\epsilon}) \quad [9]$$

avec $\boldsymbol{\Sigma}_{\epsilon} = \sigma_{\epsilon}\mathbf{I}$. Il s'agit de l'application de la première étape de l'approche bayésienne. L'étape suivante est l'attribution des loi *a priori* pour les sources et pour la matrice de mélange.

Rappelons que les sources $\mathbf{S} = \mathbf{s}_{1..T}$ sont caractérisées par leur lois de probabilité $p(\mathbf{s}_{1..T})$. Une hypothèse principale dans la plupart des méthodes classiques de séparation de sources est l'indépendance des sources. Cette indépendance, strictement parlant, signifie $p(\mathbf{S}) = \prod_j p(s_{j1..T})$. Une autre hypothèse est la blancheur (l'indépendance dans le temps), ce qui signifie $p(\mathbf{s}_{1..T}) = \prod_t p(\mathbf{s}(t))$. La combinaison de ces deux hypothèses donne :

$$p(\mathbf{S}) = p(\mathbf{s}_{1..T}) = \prod_t p(\mathbf{s}(t)) = \prod_t \prod_j p_j(s_j(t)) \quad [10]$$

où p_j est la loi *a priori* de la source $s_j(t)$.

Nous verrons par la suite que dans l'approche bayésienne on n'a pas besoin de ces hypothèses. Mais pour l'instant nous allons faire ces hypothèses pour voir le lien entre l'ACP et l'ACI avec l'approche bayésienne [MOH 99].

Concernant la matrice de mélange, nous faisons pour l'instant l'hypothèse que nous ne connaissons *a priori* pas grand chose sur la matrice de mélange \mathbf{A} , ce qui peut être traduit par l'attribution d'une loi *a priori* uniforme $p(\mathbf{A}) = \text{cte}$ ou une gaussienne $p(a_{ij}) = \mathcal{N}(0, \sigma_a^2)$ avec une variance σ_a^2 très grande [KNU 99, KNU 01, ROB 99, ROB 01, MOH 99].

Bien entendu, dans la suite de ce chapitre, nous reviendrons sur chacune de ces hypothèses, mais pour l'instant, l'objectif est de montrer le principe de la méthode. Nous avons alors tous les ingrédients pour trouver l'expression de la loi *a posteriori*

$$\begin{aligned} p(\mathbf{A}, \mathbf{S} | \mathbf{X}) &\propto p(\mathbf{X} | \mathbf{A}, \mathbf{S}) p(\mathbf{S}) p(\mathbf{A}) \\ &\propto \prod_t [\mathcal{N}(\mathbf{x}(t) - \mathbf{A}\mathbf{s}(t), \Sigma_\epsilon) p(\mathbf{s}(t))] p(\mathbf{A}) \end{aligned} \quad [11]$$

Analysons alors les différentes solutions :

1. \mathbf{A} connue, estimation de \mathbf{S} :

$$p(\mathbf{S} | \mathbf{A}, \mathbf{X}) \propto p(\mathbf{X} | \mathbf{A}, \mathbf{S}) p(\mathbf{S}) \propto \prod_t [\mathcal{N}(\mathbf{x}(t) - \mathbf{A}\mathbf{s}(t), \Sigma_\epsilon) p(\mathbf{s}(t))] \quad [12]$$

Il est alors facile de montrer que

$$-\ln p(\mathbf{S} | \mathbf{A}, \mathbf{X}) = \sum_t \frac{1}{\sigma_\epsilon^2} \|\mathbf{x}(t) - \mathbf{A}\mathbf{s}(t)\|^2 - \ln p(\mathbf{s}(t)) + c$$

où c est une constante indépendante de \mathbf{S} . Pour pouvoir aller plus loin dans les calculs, il faut préciser l'expression de $p(\mathbf{s}(t))$. Bien que naïf, supposons $p(\mathbf{s}(t)) = \mathcal{N}(\mathbf{0}, \sigma_s^2 \mathbf{I})$. Alors $p(\mathbf{S} | \mathbf{A}, \mathbf{X})$ est gaussienne

$$p(\mathbf{S} | \mathbf{A}, \mathbf{X}) = \prod_t p(\mathbf{s}(t) | \mathbf{A}, \mathbf{x}(t)) = \prod_t \mathcal{N}(\hat{\mathbf{s}}(t), \hat{\Sigma}) \quad [13]$$

et l'estimé au sens du MAP ou de la moyenne *a posteriori* de \mathbf{S} s'obtient facilement en cherchant $\hat{\mathbf{S}}$ qui maximise $-\ln p(\mathbf{S} | \mathbf{A}, \mathbf{X})$ par rapport à \mathbf{S} et qui s'écrit :

$$\hat{\mathbf{s}}(t) = (\mathbf{A}'\mathbf{A} + \lambda_s \mathbf{I})^{-1} \mathbf{A}'\mathbf{x}(t) \quad \text{ou encore} \quad \hat{\mathbf{S}} = (\mathbf{A}'\mathbf{A} + \lambda_s \mathbf{I})^{-1} \mathbf{A}'\mathbf{X} \quad [14]$$

où $\lambda_s = \sigma_\epsilon^2 / \sigma_s^2$. Notons aussi que $\widehat{\Sigma} = \sigma_\epsilon (\mathbf{A}'\mathbf{A} + \lambda_s \mathbf{I})^{-1}$ et que lorsque σ_s^2 est très grand par rapport à σ_ϵ^2 , alors $\lambda_s \simeq 0$ et $\widehat{\mathbf{s}}(t) = \mathbf{A}^+ \mathbf{x}(t)$ où \mathbf{A}^+ est l'inverse généralisée de \mathbf{A} lorsqu'elle existe.

2. Sources \mathbf{S} connues, estimation de \mathbf{A} :

$$p(\mathbf{A}|\mathbf{S}, \mathbf{X}) \propto p(\mathbf{X}|\mathbf{A}, \mathbf{S}) p(\mathbf{A}) \propto \prod_t \mathcal{N}(\mathbf{x}(t) - \mathbf{A}\mathbf{s}(t), \Sigma_\epsilon) p(\mathbf{A}) \quad [15]$$

Il est alors facile de montrer que $p(\mathbf{A}|\mathbf{S}, \mathbf{X})$ est gaussienne et que

$$-\ln p(\mathbf{A}|\mathbf{S}, \mathbf{X}) = \sum_t \frac{1}{\sigma_\epsilon^2} \|\mathbf{x}(t) - \mathbf{A}\mathbf{s}(t)\|^2 - \ln p(\mathbf{A}) + c$$

où c est une constante indépendante de \mathbf{A} . Alors, l'estimé au sens du MAP ou de la moyenne *a posteriori* de \mathbf{A} s'obtient facilement en cherchant $\widehat{\mathbf{A}}$ qui maximise $-\ln p(\mathbf{A}|\mathbf{S}, \mathbf{X})$ par rapport à \mathbf{A} et qui s'écrit :

$$\widehat{\mathbf{A}} = \left(\sum_t \mathbf{s}(t)\mathbf{s}'(t) + \lambda_a \mathbf{I} \right)^{-1} \sum_t \mathbf{x}(t)\mathbf{s}'(t) = (\mathbf{S}\mathbf{S}' + \lambda_a \mathbf{I})^{-1} \mathbf{X}\mathbf{S}' \quad [16]$$

où $\lambda_a = \sigma_\epsilon^2 / \sigma_a^2$. Notons que lorsque σ_a^2 est très grand par rapport à σ_ϵ^2 , alors $\lambda_a \simeq 0$.

3. Estimation jointe de \mathbf{A} et de \mathbf{S} :

Ecrivant l'expression de $p(\mathbf{A}, \mathbf{S}|\mathbf{X})$ on a :

$$-\ln p(\mathbf{A}, \mathbf{S}|\mathbf{X}) = \sum_t \left[\frac{1}{\sigma_\epsilon^2} \|\mathbf{x}(t) - \mathbf{A}\mathbf{s}(t)\|^2 + \ln p(\mathbf{s}(t)) \right] + \ln p(\mathbf{A}) + c \quad [17]$$

où c est une constante indépendante de \mathbf{A} et de \mathbf{S} . Si on s'intéresse à l'estimation jointe au sens du MAP, on aura à maximiser $J(\mathbf{A}, \mathbf{S}) = -\ln p(\mathbf{A}, \mathbf{S}|\mathbf{X})$ par rapport à ses arguments. Un algorithme simple et souvent utilisé est maximiser ce critère alternativement par rapport à \mathbf{S} et ensuite par rapport à \mathbf{A} , ce qui revient à :

$$\begin{cases} \widehat{\mathbf{S}} &= \arg \max_{\mathbf{S}} \left\{ p(\mathbf{S}|\widehat{\mathbf{A}}, \mathbf{X}) \right\} = \arg \min_{\mathbf{S}} \left\{ \|\mathbf{X} - \widehat{\mathbf{A}}\mathbf{S}\|^2 - \ln p(\mathbf{S}) \right\} \\ \widehat{\mathbf{A}} &= \arg \max_{\mathbf{A}} \left\{ p(\mathbf{A}|\widehat{\mathbf{S}}, \mathbf{X}) \right\} = \arg \min_{\mathbf{A}} \left\{ \|\mathbf{X} - \mathbf{A}\widehat{\mathbf{S}}\|^2 - \ln p(\mathbf{A}) \right\} \end{cases} \quad [18]$$

Cependant, il n'y a, en général, aucune garantie qu'un tel algorithme converge vers la solution MAP jointe. D'autre part, la difficulté de l'optimisation dans chaque étape dépend bien sûr des choix pour $p(\mathbf{s}(t))$ et pour $p(\mathbf{A})$. Dans le cas des hypothèses gaussiennes des deux exemples précédents, on a :

$$-\ln p(\mathbf{A}, \mathbf{S} | \mathbf{X}) = \sum_t \left[\frac{1}{\sigma_c^2} \|\mathbf{x}(t) - \mathbf{A}, \mathbf{s}(t)\|^2 - \frac{1}{\sigma_s^2} \|\mathbf{s}(t)\|^2 \right] + \frac{1}{\sigma_a^2} \|\mathbf{A}\|^2 + c$$

On peut alors constater que ce critère est quadratique en \mathbf{S} à \mathbf{A} fixée, et quadratique en \mathbf{A} à \mathbf{S} fixée, ce qui permet d'obtenir une solution analytique pour chaque étape :

$$\begin{cases} \widehat{\mathbf{S}} = (\widehat{\mathbf{A}}' \widehat{\mathbf{A}} + \lambda_1 \mathbf{I})^{-1} \widehat{\mathbf{A}}' \mathbf{X}, \\ \widehat{\mathbf{A}} = (\widehat{\mathbf{S}} \widehat{\mathbf{S}}' + \lambda_2 \mathbf{I})^{-1} \mathbf{X} \widehat{\mathbf{S}}' \end{cases} \quad [19]$$

ou encore

$$\begin{cases} \widehat{\mathbf{s}}(t) = (\widehat{\mathbf{A}}' \widehat{\mathbf{A}} + \lambda_s \mathbf{I})^{-1} \widehat{\mathbf{A}}' \mathbf{x}(t), \\ \widehat{\mathbf{A}} = (\sum_t \widehat{\mathbf{s}}(t) \widehat{\mathbf{s}}'(t) + \lambda_a \mathbf{I})^{-1} \sum_t \mathbf{x}(t) \widehat{\mathbf{s}}'(t) \end{cases} \quad [20]$$

Cependant, bien que ce type d'algorithmes intuitifs sont souvent utilisés, non seulement il n'y a aucune garantie de convergence, mais aussi qu'il est difficile de caractériser la solution obtenue, car bien que ce critère est quadratique en \mathbf{S} à \mathbf{A} fixée, et quadratique en \mathbf{A} à \mathbf{S} fixée, il n'est pas convexe en couple (\mathbf{A}, \mathbf{S}) car il est bilinéaire.

Cet algorithme avec $\lambda_a = \lambda_s = 0$ correspond à l'optimisation alternée du critère des moindres carrés $J(\mathbf{A}, \mathbf{S}) = \|\mathbf{X} - \mathbf{A}\mathbf{S}\|^2$ par rapport à \mathbf{S} et à \mathbf{A} utilisé couramment en spectrométrie [BRO 97, CAM 81, ATT 98, ATT 99, MOU 03b, MOU 03a, MOU 04b, MOU 04a]. En effet, la mise en oeuvre de tels algorithmes est très simple, mais si on ne contraint pas la solution, ce genre d'algorithme ne convergera pas. C'est pourquoi, ces algorithmes sont utilisés en imposant des contraintes du genre positivité :

$$\begin{cases} \widehat{\mathbf{s}}(t) = (\widehat{\mathbf{A}}' \widehat{\mathbf{A}})^{-1} \widehat{\mathbf{A}}' \mathbf{x}(t), & \text{puis appliquer des contraintes sur } \widehat{\mathbf{s}}(t) \\ \widehat{\mathbf{A}} = (\sum_t \widehat{\mathbf{s}}(t) \widehat{\mathbf{s}}'(t))^{-1} \sum_t \mathbf{x}(t) \widehat{\mathbf{s}}'(t), & \text{puis appliquer des contraintes sur } \widehat{\mathbf{A}} \end{cases}$$

Cependant, étudier la convergence de tels algorithmes et les propriétés de la solution obtenue n'est pas facile. La prise en compte de $p(\mathbf{s}(t))$ et de $p(\mathbf{A})$ n'est réellement pas plus compliquée tant qu'on ne cherche pas la rigueur mathématique. Par exemple, si on choisit des lois gaussiennes tronquées pour $p(\mathbf{s}(t))$ et pour $p(\mathbf{A})$ pour imposer la positivité, la modification à apporter à l'algorithme précédent reste assez simple :

$$\begin{cases} \widehat{\mathbf{s}}(t) = (\widehat{\mathbf{A}}' \widehat{\mathbf{A}} + \lambda_s \mathbf{I})^{-1} \widehat{\mathbf{A}}' \mathbf{x}(t), & \text{puis appliquer des contraintes sur } \widehat{\mathbf{s}}(t) \\ \widehat{\mathbf{A}} = (\sum_t \widehat{\mathbf{s}}(t) \widehat{\mathbf{s}}'(t) + \lambda_a \mathbf{I})^{-1} \sum_t \mathbf{x}(t) \widehat{\mathbf{s}}'(t), & \text{puis appliquer des contraintes sur } \widehat{\mathbf{A}} \end{cases}$$

où λ_s et λ_a sont, respectivement, inversement proportionnelles aux variances σ_s^2 et σ_a^2 des loi gaussiennes tronquées des éléments de la matrice de mélange

$$p(a_{ij}) = 0 \text{ si } a_{ij} < 0 \text{ et } p(a_{ij}) = 2\mathcal{N}(0, \sigma_a^2) \text{ si } a_{ij} \geq 0$$

et des sources

$$p(s_j(t)) = 0 \text{ si } s_j(t) < 0 \text{ et } p(s_j(t)) = 2\mathcal{N}(0, \sigma_s^2) \text{ si } s_j(t) \geq 0.$$

3. Estimation de \mathbf{A} en marginalisant \mathbf{S} :

L'idée est, d'intégrer $p(\mathbf{A}, \mathbf{S} | \mathbf{X})$ par rapport à \mathbf{S} pour obtenir la loi marginale $p(\mathbf{A} | \mathbf{X})$ et l'utiliser ensuite pour estimer \mathbf{A} . Ceci peut se faire d'une autre manière en remarquant que $p(\mathbf{A} | \mathbf{X}) \propto p(\mathbf{X} | \mathbf{A}) p(\mathbf{A})$. On a alors besoin d'exprimer $p(\mathbf{X} | \mathbf{A})$. Pour cela, rappelons que, si on suppose les signaux $\mathbf{s}(t)$ et $\boldsymbol{\epsilon}(t)$ indépendants et blancs, alors $\mathbf{x}(t)$ est aussi blanc, c'est-à-dire $p(\mathbf{X} | \mathbf{A}) = \prod_t p(\mathbf{x}(t) | \mathbf{A})$ où

$$p(\mathbf{x} | \mathbf{A}) = \int p(\mathbf{x}, \mathbf{s} | \mathbf{A}) d\mathbf{s} = \int p(\mathbf{x} | \mathbf{A}, \mathbf{s}) p(\mathbf{s}) d\mathbf{s}$$

Notons aussi que lorsque $p(\mathbf{A})$ est choisi uniforme, on a $p(\mathbf{A} | \mathbf{X}) \propto p(\mathbf{X} | \mathbf{A})$, c'est à dire la vraisemblance de \mathbf{A} dans les données \mathbf{X} .

On peut alors montrer que la plupart des méthodes classiques de l'ACP et de l'ACI reviennent à estimer \mathbf{A} en approximant l'expression de la vraisemblance $p(\mathbf{X} | \mathbf{A})$ et en le maximisant par rapport à \mathbf{A} . En effet trouver une expression simple pour cette vraisemblance n'est possible que dans deux cas : i) le cas gaussien et ii) le cas d'un modèle sans bruit.

Cas gaussien et lien avec ACP :

En effet, si nous supposons les sources $\mathbf{s}(t)$ et les bruits $\boldsymbol{\epsilon}(t)$ gaussiennes, on a :

$$\begin{aligned} p(\mathbf{x}(t) | \mathbf{A}, \mathbf{s}(t), \boldsymbol{\Sigma}_\epsilon) &= \mathcal{N}(\mathbf{A}\mathbf{s}(t), \boldsymbol{\Sigma}_\epsilon) \\ p(\mathbf{s}(t) | \boldsymbol{\Sigma}_s) &= \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_s) \end{aligned} \longrightarrow p(\mathbf{x}(t) | \mathbf{A}, \boldsymbol{\Sigma}_\epsilon, \boldsymbol{\Sigma}_s) = \mathcal{N}(\mathbf{0}, \mathbf{A}\boldsymbol{\Sigma}_s\mathbf{A}' + \boldsymbol{\Sigma}_\epsilon)$$

Chercher alors la matrice de mélange \mathbf{A} qui maximise la vraisemblance

$$p(\mathbf{X} | \mathbf{A}, \boldsymbol{\Sigma}_\epsilon, \boldsymbol{\Sigma}_s) = \prod_t p(\mathbf{x}(t) | \mathbf{A}, \boldsymbol{\Sigma}_\epsilon, \boldsymbol{\Sigma}_s)$$

devient équivalent aux algorithmes de l'ACP qui consiste en estimation de la matrice de covariance des observations $\mathbf{x}_{1..T}$ par $\Sigma_x = \frac{1}{T} \sum_t \mathbf{x}(t)\mathbf{x}'(t)$ et son identification avec $\mathbf{A}\Sigma_s\mathbf{A}'$ dans le cas d'un modèle sans bruit ($\Sigma_\epsilon = 0$) ou avec $\mathbf{A}\Sigma_s\mathbf{A}' + \Sigma_\epsilon$ dans le cas plus général.

Cas non-gaussien et lien avec ACI :

Dans le cas non-gaussien, il n'est pas possible d'obtenir une expression analytique pour l'estimée au sens du MV, car premièrement, il est rare qu'on puisse obtenir une expression analytique pour la vraisemblance, et même lorsque cela est possible, cette expression n'est pas quadratique en \mathbf{A} . Cependant, il y a un cas où on peut au moins obtenir une expression pour la vraisemblance est lorsque le modèle d'observation est supposée sans bruit ($\epsilon = \mathbf{0}$) et inversible. C'est dans ce cas que la plupart des méthodes de l'ACI ont été développées [HYV 00, JUT 91, COM 91, COM 94, PHA 96, CAR 99, JUT 00, BEL 94, BEL 95, CAR 96, CAR 97, CAR 98]. En effet, dans ce cas $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$ et si on suppose \mathbf{A} inversible : $\mathbf{A} = \mathbf{B}^{-1}$, il est alors facile de montrer que

$$\ln p(\mathbf{X}|\mathbf{B}) = \frac{T}{2} \ln |\mathbf{B}| + \sum_t \sum_j \ln p_j([\mathbf{B}\mathbf{x}(t)]_j) + c \quad [21]$$

où c est une constante ne dépendant pas de \mathbf{B} et où p_j est la loi *a priori* de la source s_j . Si maintenant on note par $\mathbf{y}(t) = \mathbf{B}\mathbf{x}(t)$, le schéma d'un algorithme itératif du type gradient pour la recherche de la matrice de séparation \mathbf{B} qui maximise $\ln p(\mathbf{x}_{1..T}|\mathbf{B})$ devient

$$\mathbf{B}^{(k+1)} = \mathbf{B}^{(k)} - \gamma \mathbf{H}(\mathbf{y}) \quad \text{avec} \quad \mathbf{H}(\mathbf{y}) = \phi(\mathbf{y}) \mathbf{y}^t - \mathbf{I}, \quad [22]$$

où $\phi(\mathbf{y}) = [\phi_1(y_1), \dots, \phi_n(y_n)]^t$ avec $\phi_j(z) = -\frac{p'_j(z)}{p_j(z)}$.

Il est à noter que l'expression des opérateur $\phi_j(y_j)$ dépend de la l'expression de la loi *a priori* des sources. Le tableau suivant donne un certain nombre de ces expressions.

Gauss	$p(z) \propto \exp[-\alpha z^2]$	$\phi(z) = 2\alpha z$
Laplace	$p(z) \propto \exp[-\alpha z]$	$\phi(z) = \alpha \text{sign}(z)$
Cauchy	$p(z) \propto \frac{1}{1+(z/\alpha)^2}$	$\phi(z) = \frac{2z/\alpha^2}{1+(z/\alpha)^2}$
Gamma	$p(z) \propto z^\alpha \exp[-\beta z]$	$\phi(z) = -\alpha/z + \beta$
loi sous-gaussiennes	$p(z) \propto \exp\left[-\frac{1}{2}z^2\right] \text{sech}^2(z)$	$\phi(z) = z + \tanh(z)$
mélange de gaussiennes	$p(z) \propto \exp\left[-\frac{1}{2}(z-\alpha)^2\right] + \exp\left[-\frac{1}{2}(z+\alpha)^2\right]$	$\phi(z) = \alpha z - \alpha \tanh(\alpha z)$

Ce schéma avec d'autres variants (autres algorithmes que du gradient et éventuellement avec d'autres contraintes supplémentaires) forme la base d'un grand nombre d'algorithmes classiques de séparation de sources.

Cas non-gaussien et lien avec l'algorithme EM :

Nous venons de voir que dans le cas général, obtenir une expression analytique pour la vraisemblance n'est pas possible. C'est alors ici que l'algorithme EM intervient comme un algorithme itératif qui converge localement vers la solution au sens du MV [DEL 99]. Sans entrer dans le détail de la présentation de cet algorithme, nous donnons ici une interprétation qui considère les sources comme des variables cachées et donc les observations \mathbf{X} comme des données incomplètes et l'ensemble (\mathbf{X}, \mathbf{S}) comme des données complètes.

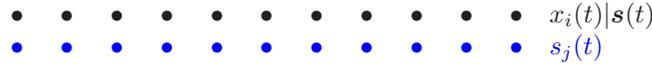


Figure 1. Modèle de séparation de sources où les sources sont considérées comme des variables cachées.

L'algorithme EM devient alors

$$\begin{cases} \text{E: } Q(\mathbf{A}, \hat{\mathbf{A}}) = \int J(\mathbf{x}, \mathbf{s}) p(\mathbf{x}, \mathbf{s} | \hat{\mathbf{A}}) d\mathbf{s} \\ \text{M: } \hat{\mathbf{A}} = \arg \max_{\mathbf{A}} \{Q(\mathbf{A}, \hat{\mathbf{A}})\} \end{cases} \quad [23]$$

où $J(\mathbf{x}, \mathbf{s}) = \ln p(\mathbf{x}, \mathbf{s} | \mathbf{A}) = \ln p(\mathbf{x} | \mathbf{A}, \mathbf{s}) + \ln p(\mathbf{s})$ est la log-vraisemblance des données complète. Mais, malheureusement, là encore l'étape de calcul de l'espérance n'est possible analytiquement que dans le cas gaussien.

Dans le cas gaussien $J(\mathbf{x}, \mathbf{s}) = \|\mathbf{x} - \mathbf{A}\mathbf{s}\|^2 + \lambda_s \|\mathbf{s}\|^2$. Alors, en le remplaçant et en calculant l'expression de $Q(\mathbf{A}, \hat{\mathbf{A}})$ en ne gardant que les termes qui dépendent de \mathbf{A} , on montre que cet algorithme devient équivalent à l'algorithme de *restauration-maximisation* suivant :

$$\begin{cases} \text{E: } \hat{\mathbf{S}} = \mathbb{E} \left\{ \mathbf{S} | \hat{\mathbf{A}}, \mathbf{X} \right\} \\ \text{M: } \hat{\mathbf{A}} = \arg \max_{\mathbf{A}} \left\{ p(\mathbf{A} | \hat{\mathbf{S}}, \mathbf{X}) \right\} \end{cases} \quad [24]$$

ce qui peut être comparé avec l'algorithme d'optimisation conjointe de $p(\mathbf{A}, \mathbf{S} | \mathbf{X})$ où la première maximisation par rapport à \mathbf{S} est remplacée par le calcul de la moyenne *a posteriori*. Les deux algorithmes sont équivalents dans le cas gaussien.

4. Intégration par rapport à \mathbf{A} :

De la même manière que nous avons intégré $p(\mathbf{A}, \mathbf{S} | \mathbf{X})$ par rapport à \mathbf{S} pour obtenir la loi marginale $p(\mathbf{A} | \mathbf{X})$, on peut envisager l'intégration par rapport à \mathbf{A} pour obtenir $p(\mathbf{S} | \mathbf{X})$ et estimer directement \mathbf{S} . Cette voie n'est pas très utilisée car l'intégration par rapport à \mathbf{A} n'est pas aisée, mais on peut montrer que dans le cas gaussien, on retrouve un algorithme du type *restauration-maximisation* du paragraphe précédent.

Dans la suite de ce chapitre, nous nous focalisons particulièrement sur l'estimation jointe qui nous semble la plus complète, et étudierons surtout les effets des différentes modélisation des sources.

0.4. Choix des loi *a priori* et expressions des loi *a posteriori*

Une étape essentielle dans n'importe quelle méthode de séparation de sources est la modélisation *a priori* des sources. Dans la suite de ce chapitre, nous verrons comment les modélisation de plus en plus élaborée des sources et de la matrice de mélange conduit à des méthodes et algorithmes de séparation de sources de plus en plus précises.

Nous venons aussi de voir que dans l'approche bayésienne, pour être capable d'exprimer la loi *a posteriori*, il faut déterminer la vraisemblance $p(\mathbf{X} | \mathbf{A}, \mathbf{S})$ et les loi *a priori* $p(\mathbf{A})$ et $p(\mathbf{S})$. Mais, évidemment ces différentes loi dépendent aussi des hyper paramètres. Par exemple, si on suppose le bruit centré et Gaussien, $p(\mathbf{X} | \mathbf{A}, \mathbf{S})$ dépend de la matrice de covariance $\boldsymbol{\theta}_\epsilon = \boldsymbol{\Sigma}_\epsilon$ du bruit. De même, $p(\mathbf{S})$ dépendra d'un ensemble de paramètres $\boldsymbol{\theta}_s$ et $p(\mathbf{A})$ de $\boldsymbol{\theta}_A$. Nous noterons par $\boldsymbol{\theta} = (\boldsymbol{\theta}_\epsilon, \boldsymbol{\theta}_s, \boldsymbol{\theta}_A)$ l'ensemble de ces hyper paramètres, qu'il faut aussi estimer. Naturellement, dans une

approche bayésienne, on attribue aussi une loi *a priori* $p(\boldsymbol{\theta})$ à ces paramètres. On a alors

$$p(\mathbf{A}, \mathbf{S}, \boldsymbol{\theta} | \mathbf{X}) \propto p(\mathbf{X} | \mathbf{A}, \mathbf{S}, \boldsymbol{\Sigma}_\epsilon) p(\mathbf{A} | \boldsymbol{\theta}_A) p(\mathbf{S} | \boldsymbol{\theta}_s) \quad [25]$$

Par la suite, nous faisons les choix suivants :

— Les bruit ϵ_i sont supposés mutuellement indépendants, centrés et gaussiennes, ce qui donne :

$$\begin{aligned} p(\mathbf{X} | \mathbf{A}, \mathbf{S}, \boldsymbol{\Sigma}_\epsilon) &= \prod_t \mathcal{N}(\mathbf{A}\mathbf{s}(t), \boldsymbol{\Sigma}_\epsilon) \\ &\propto \exp \left[\frac{-1}{2} \sum_t ((\mathbf{x}(t) - \mathbf{A}\mathbf{s}(t))' \boldsymbol{\Sigma}_\epsilon^{-1} (\mathbf{x}(t) - \mathbf{A}\mathbf{s}(t))) \right] \\ &\propto \exp \left[\frac{-1}{2} \left(\sum_t \text{Tr} \{ (\mathbf{x}(t) - \mathbf{A}\mathbf{s}(t)) \boldsymbol{\Sigma}_\epsilon^{-1} (\mathbf{x}(t) - \mathbf{A}\mathbf{s}(t))' \} \right) \right] \end{aligned}$$

— Pour les éléments de la matrice de mélange, suivant l'application, on choisit :

– Gaussiennes :

$$p(\mathbf{A}_{ij}) = \mathcal{N}(M_{ji}, \sigma_{a,ij}^2) \longrightarrow p(\mathbf{A} | \boldsymbol{\theta}_A) = \mathcal{N}(\mathbf{M}_A, \boldsymbol{\Sigma}_A)$$

– Gaussiennes généralisées :

$$p(A_{ij}) \propto \exp[-\gamma |A_{ij}|]$$

pour prendre en compte la sparcité des éléments de la matrice ;

– Gamma :

$$p(A_{ij}) = \mathcal{G}(\alpha_j, \beta_j) \propto A_{ij}^{\alpha_j - 1} \exp[-\beta_j A_{ij}]$$

pour assurer leurs positivités (par exemple en spectrométrie) ; ou

– uniforme pour le cas général.

— Pour les hyper paramètres, nous choisissons des loi conjuguées, c'est-à-dire des loi inverse gamma pour les variance, des loi inverse Wishart pour les matrices de covariance et des loi normales pour les moyennes [SNO 02].

Par la suite, nous discuterons surtout les différents choix pour la modélisation des sources. Mais auparavant, nous allons mentionner les différentes possibilités de choix des estimateurs et des algorithmes possibles pour leurs calcul. Le point essentiel est qu'à partir de l'expression de la loi *a posteriori* $p(\mathbf{S}, \mathbf{A}, \boldsymbol{\theta} | \mathbf{X})$, on peut envisager différents schémas :

$$\begin{aligned}
 \text{Alg. 1 : } \widehat{S} &\sim p(S|\widehat{A}, \widehat{\theta}, X) \longrightarrow \widehat{A} \sim p(A|\widehat{S}, \widehat{\theta}, X) \longrightarrow \widehat{\theta} \sim p(\theta|\widehat{A}, \widehat{S}, X) \\
 \text{Alg. 2 : } \widehat{S} &\sim p(S|\widehat{A}, \widehat{\theta}, X) \longrightarrow \widehat{A} \sim p(A|\widehat{\theta}, X) \longrightarrow \widehat{\theta} \sim p(\theta|\widehat{A}, X) \\
 \text{Alg. 3 : } \widehat{S} &\sim p(S|\widehat{A}, \widehat{\theta}, X) \longrightarrow \widehat{A} \sim p(A|\widehat{\theta}, X) \longrightarrow \widehat{\theta} \sim p(\theta|X)
 \end{aligned}$$

Dans chacun de ces schémas, \sim signifie soit *égale à l'argument qui maximise* ou *échantillonner suivant*. Le premier correspond à un estimateur du type MAP et nécessite donc un algorithme d'optimisation. Le deuxième cas correspond à un estimateur du type la moyenne *a posteriori* et l'utilisation des techniques d'intégration par MCMC. Évidemment, la mise en œuvre de **Alg. 1** est la plus simple car nous disposons déjà l'expression de toutes les loi conditionnelles nécessaires :

$$\begin{aligned}
 p(S|A, \theta, X) &\propto p(X|S, A, \theta_\epsilon) p(S|\theta_s), \\
 p(A|S, \theta, X) &\propto p(X|S, A, \theta_\epsilon) p(A|\theta_A), \\
 p(\theta|A, S, X) &\propto p(X|S, A, \theta_\epsilon) p(\theta)
 \end{aligned} \tag{26}$$

Alg. 2 nécessite l'intégration par rapport à S pour obtenir $p(A|\theta, X)$ et $p(\theta|A, X)$:

$$\begin{aligned}
 p(A|\theta, X) &\propto p(X|A, \theta_\epsilon) p(A|\theta_A), \\
 p(\theta|A, X) &\propto p(X|A, \theta_\epsilon) p(\theta) \\
 \text{avec } p(X|A, \theta_\epsilon) &= \int p(X|S, A, \theta_\epsilon) p(S|\theta_s) dS
 \end{aligned} \tag{27}$$

qui n'est pas toujours possible.

Alg. 3 nécessite une étape de marginalisation par rapport à S pour obtenir $p(A|\theta, X)$ et une étape de marginalisation supplémentaire par rapport à A pour obtenir $p(\theta|X)$:

$$\begin{aligned}
 p(\theta|X) &\propto p(X|\theta_\epsilon) p(\theta), \\
 \text{avec } p(X|\theta_\epsilon) &= \int p(X|A, \theta_\epsilon) p(A) dA
 \end{aligned} \tag{28}$$

qui ne sont pas toujours possible.

0.5. Modélisation des sources

Nous avons déjà étudié le cas des sources gaussiennes et le lien avec l'ACP. Il est évident que sans la prise en compte de la dynamique temporelle des sources (différences entre les fonctions de corrélations ou entre les densités spectrales), on ne peut séparer les sources gaussiennes. Par contre, dès que les sources sont non-gaussiennes, la séparation est possible si les sources ont des loi de probabilités différentes. Parmi les différents choix de loi de probabilités pour les sources on peut mentionner les loi gaussiennes généralisées (GG), [SNO 03, SNO 04, ICH 03b], des loi Gammas [MOU 03a] et des loi de mélange de gaussiennes (MoG) [SNO 04, ROB 03, ICH 03b, ICH 03a, ICH 03c].

Par la suite, nous distinguons donc deux cas : le cas des *signaux stationnaires* et le cas des *signaux non stationnaires*.

Dans le premier cas, nous considérons la modélisation des loi des sources par des gaussiennes généralisées, Gamma ou par des mélange de gaussiennes que nous les nommerons IGG, IG et IGM. (le préfix I s'entend iid). Nous nous focaliserons surtout sur IGM, car naturellement elles introduisent la notion de variable cachée de classification, et fournissons un certain nombre d'algorithmes que nous avons développés avec cette modélisation.

Dans le deuxième cas, nous considérons la modélisation des loi des sources non stationnaires à l'aide des variables cachées qui nous permettrons de modéliser d'une manière bien contrôlée la non stationnarité des sources. En particulier, nous étudierons les cas de la modélisation par chaîne ou par champs de Markov cachés.

0.6. Modélisation des sources blanches non gaussiennes mais stationnaires

0.6.1. Gaussienne généralisée

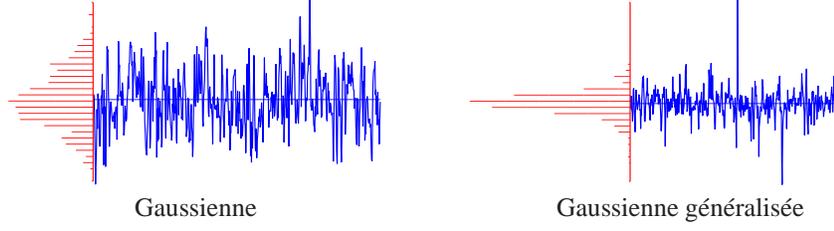
Dans cette modélisation, la loi de chaque source $s_j(t)$ est :

$$p(s_j | \gamma_j, \beta_j) \propto \exp[-\beta_j |s_j|^{\gamma_j}], \quad 1 \leq \beta_j \leq 2 \quad [29]$$

Ainsi, les loi des sources sont paramétrisées par $\beta = \{\beta_1, \dots, \beta_n\}$ et $\gamma = \{\gamma_1, \dots, \gamma_n\}$. La figure suivante montre un exemple de réalisation d'un signal gaussienne en comparaison avec une réalisation d'un signal GG.

Avec cette modélisation, il est facile de montrer que

$$-\ln p(\mathbf{S}, \mathbf{A} | \mathbf{X}, \boldsymbol{\gamma}, \boldsymbol{\beta}) \propto - \sum_t [(\mathbf{x}(t) - \mathbf{A}\mathbf{s}(t))' \boldsymbol{\Sigma}_\epsilon^{-1} (\mathbf{x}(t) - \mathbf{A}\mathbf{s}(t)) + \sum_j \beta_j |s_j(t)|^{\gamma_j}]$$



$$\begin{aligned}
 p(s_j(t)) &= \mathcal{N}(0, \sigma_j^2) \propto \exp\left[-\frac{1}{\sigma_j^2}|s_j(t)|^2\right] & p(s_j(t)) &\propto \exp[-\beta_j|s_j(t)|^{\gamma_j}] \\
 p(s_j(t), t = 1..T) &\propto \exp\left[-\frac{1}{\sigma_j^2}\sum_t |s_j(t)|^2\right] & p(s_j(t), t = 1..T) &\propto \exp[-\beta_j\sum_t |s_j(t)|^{\gamma_j}]
 \end{aligned}$$

Figure 2. Exemples de réalisations d'un signal gaussien et d'un signal GG.

et si on est capable de fixer *a priori* les hyper paramètres, on peut alors utiliser l'algorithme suivant :

$$\begin{cases}
 \hat{s}_j^{(k+1)}(t) = \arg \min_{s_j(t)} \left\{ J_1(s_j(t)) = -\ln p\left(s_j(t)|\mathbf{x}(t), \hat{\mathbf{A}}^{(k)}, \hat{\mathbf{s}}_{k \neq j}^{(k)}(t)\right) \right\} \\
 \hat{\mathbf{A}}^{(k+1)} = \arg \min_{\mathbf{A}} \left\{ J_2(\mathbf{A}) = -\ln p(\mathbf{A}|\hat{\mathbf{S}}^{(k)}, \mathbf{X}) \right\}
 \end{cases}$$

où

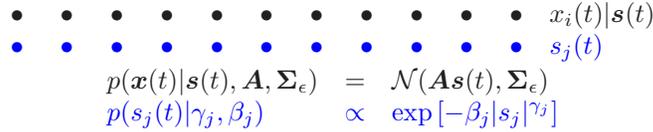
$$\begin{cases}
 J_1(s_j(t)) = \sum_i |x_i(t) - A_{ij}s_j(t) - \sum_{j' \neq j} A_{ij'} \hat{s}_{j'}^{(k)}(t)|^2 + \beta_j \sum_j |s_j(t)|^{\gamma_j} \\
 J_2(\mathbf{A}) = \sum_t \|\mathbf{x}(t) - \mathbf{A}\hat{\mathbf{s}}t\|^2 + \ln p(\mathbf{A})
 \end{cases}$$

On peut utiliser différents algorithmes d'optimisation (point fixe ou de descente de gradient) pour l'optimisation de $J_1(s_j(t))$ tant que $2 \geq \gamma_j > 1$, et si on choisit une loi uniforme ou Gaussienne pour $p(\mathbf{A})$, le critère $J_2(\mathbf{A})$ a une solution explicite. Avec d'autres choix pour $p(\mathbf{A})$, par exemple la loi gamma, on peut aussi utiliser un algorithme de descente pour l'optimisation de $J_2(\mathbf{A})$.

Mais, on peut aussi attribuer des lois *a priori* conjuguées à l'ensemble des paramètres et hyper-paramètres $\theta = (\mathbf{A}, \Sigma_\epsilon, \gamma, \beta)$ du problème pour obtenir une expression pour la loi *a posteriori*

$$p(\mathbf{A}, \Sigma_\epsilon, \gamma, \beta | \mathbf{x}_{1..T}) \propto p(\mathbf{x}_{1..T} | \mathbf{A}, \Sigma_\epsilon, \gamma, \beta) p(\mathbf{A}) p(\Sigma_\epsilon) p(\gamma) p(\beta). \quad [30]$$

On peut alors estimer d'abord $(\mathbf{A}, \Sigma_\epsilon, \gamma, \beta)$ et les utiliser ensuite pour estimer les sources. Mais, l'obtention de l'expression exacte de $p(\mathbf{A}, \Sigma_\epsilon, \gamma, \beta | \mathbf{x}_{1..T})$ est difficile et on peut alors recourir de nouveau à l'algorithme EM où les sources sont considérées comme des variables cachées (données manquantes) comme c'est illustré sur la figure suivante.



$$\begin{aligned}
 & \bullet \quad x_i(t)|s(t) \\
 & \bullet \quad s_j(t) \\
 & p(\mathbf{x}(t)|\mathbf{s}(t), \mathbf{A}, \Sigma_\epsilon) = \mathcal{N}(\mathbf{A}\mathbf{s}(t), \Sigma_\epsilon) \\
 & p(s_j(t)|\gamma_j, \beta_j) \propto \exp[-\beta_j|s_j|^{\gamma_j}]
 \end{aligned}$$

Figure 3. Un modèle de séparation de sources où les sources sont considérées comme des variables cachées.

La principale difficulté réside alors dans l'étape de l'espérance qui nécessite une intégration. Trouver une expression exacte et analytique pour ces intégrales est rarement possible. On utilise alors de nouveau soit une approximation, soit une estimation via MCMC (Méthodes du type approximation de Rao-Blackwell ou du type SEM [SNO 03]).

Une autre possibilité est recourir à des techniques variationnelles qui cherchent à l'approximer cette loi par une loi séparable :

$$p(\mathbf{A}, \Sigma_\epsilon, \gamma, \beta | \mathbf{x}_{1..T}) \simeq \prod_t p(\mathbf{A}, \Sigma_\epsilon | \mathbf{x}(t), \mathbf{s}(t)) p(\gamma | \mathbf{s}(t)) p(\beta | \mathbf{s}(t))$$

ce qui permettrait alors de développer des techniques d'échantillonnage plus simples. Les lecteurs intéressés par l'utilisation de cette modélisation et les différentes utilisations et implémentations se référer aux [ICH 03c].

0.6.2. Mélange de gaussiennes

Dans cette modélisation, la loi de chaque source $s_j(t)$ est :

$$p(s_j | \alpha_{j_k}, m_{j_k}, \sigma_{j_k}^2) = \sum_{k=1}^{K_j} \alpha_{j_k} \mathcal{N}(m_{j_k}, \sigma_{j_k}^2) \quad \text{avec} \quad \sum_{k=1}^{K_j} \alpha_{j_k} = 1 \quad [31]$$

Ainsi, la loi de chaque source s_j est paramétrée par

$$(K_j, \mathbf{m}_j = \{m_{jk}, k = 1, \dots, K_j\}, \boldsymbol{\sigma}_j = \{\sigma_{jk}^2, k = 1, \dots, K_j\})$$

et les loi des sources sont donc paramétrisées par $\mathbf{K} = \{K_1, \dots, K_n\}$, $\boldsymbol{\alpha} = \{\alpha_{jk}\}$, $\mathbf{m} = \{\mathbf{m}_j, j = 1, \dots, n\}$ et $\boldsymbol{\sigma} = \{\boldsymbol{\sigma}_j, j = 1, \dots, n\}$.

Le principal intérêt de cette modélisation se trouve dans l'interprétation de $\alpha_{jk} = P(z_j(t) = k)$ ce qui nous permet d'introduire la variable cachée $z_j(t)$ qui peut prendre des valeurs discrètes $k = 1, \dots, K_j$.

Ainsi, on peut écrire

$$p(s_j(t)|z_j(t) = k, m_{jk}, \sigma_{jk}^2) = \mathcal{N}(m_{jk}, \sigma_{jk}^2) \text{ et } P(z_j(t) = k) = \alpha_{jk}. \quad [32]$$

et le problème se ramène à l'estimation de $\mathbf{s}(t) = \{s_1(t), \dots, s_n(t)\}$, de $\mathbf{z}(t) = \{z_1(t), \dots, z_n(t)\}$ et les hyperparamètres $(\mathbf{A}, \boldsymbol{\Sigma}_\epsilon)$ et $\boldsymbol{\theta}_s = (\mathbf{m}, \boldsymbol{\sigma})$.

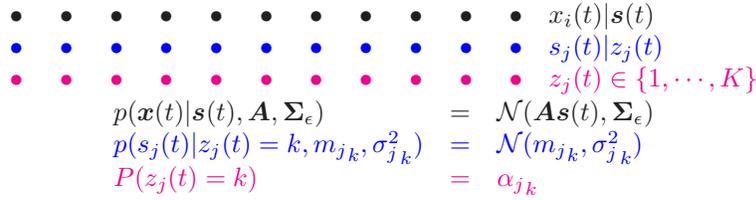


Figure 4. Modèle de mélange et des variables cachées

Notons aussi que les sources étant modélisées par des mélanges de gaussiennes et les observations étant mélanges instantanées des sources, on a

$$p(x_i(t)|\mathbf{A}, \mathbf{z}(t), \mathbf{m}, \boldsymbol{\sigma}) = \mathcal{N}\left(\sum_j A_{ij}m_{jz_j(t)}, \sum_j A_{ij}\sigma_{jz_j(t)}A_{ji} + \sigma_{\epsilon_i}^2\right) \quad [33]$$

ce qui signifie que la loi de chaque observation $x_i(t)$ est aussi une loi de mélange avec $K_i = \prod_j K_j$ composantes.

Nous avons utilisé cette modélisation et avons développé un certain nombre d'algorithmes de séparation. Ici, nous mentionnons un algorithme qui est basé sur l'estimation jointe des sources \mathbf{S} , des variables cachées \mathbf{Z} , de la matrice de mélange \mathbf{A} , et

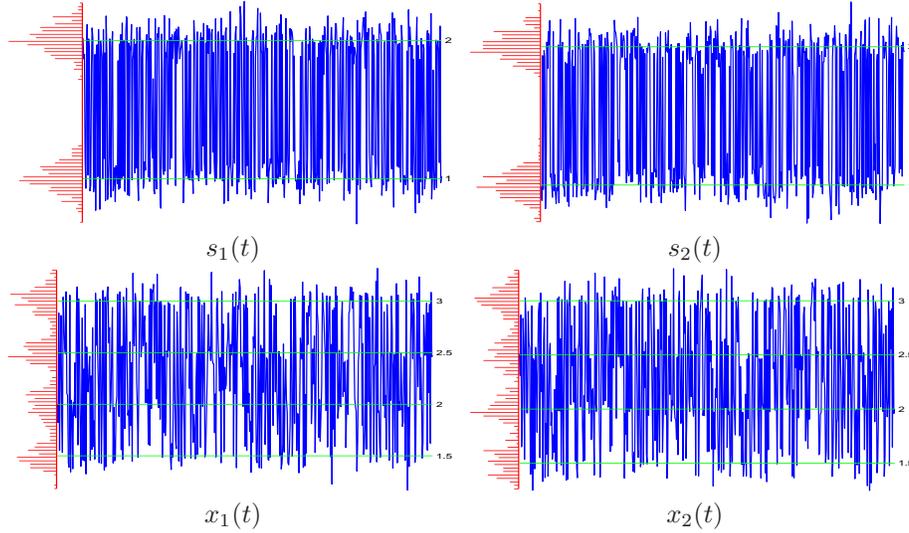


Figure 5. Signaux sources s_1 et s_2 sont des mélange de 2 gaussiennes. Signaux observés x_1 et x_2 sont aussi des mélanges de gaussiennes, mais chacun a, au plus 4 composants.

des hyperparamètres m_{j_k} et $\sigma_{j_k}^2$ en maximisant la loi *a posteriori* jointe $p(\mathbf{S}, \mathbf{Z}, \mathbf{A}, \boldsymbol{\theta} | \mathbf{X})$, respectivement par rapport à ses arguments :

$$\begin{cases} \hat{\mathbf{s}}(t) = \arg \min_{\mathbf{s}(t)} \left\{ J_1(\mathbf{s}(t)) = -\ln p(\mathbf{s}(t) | \mathbf{x}(t), \hat{\mathbf{z}}(t), \hat{\mathbf{A}}, \hat{\mathbf{s}}(t)) \right\} \\ \hat{\mathbf{A}} = \arg \min_{\mathbf{A}} \left\{ J_2(\mathbf{A}) = -\ln p(\mathbf{A} | \hat{\mathbf{S}}, \mathbf{X}) \right\} \\ \hat{\mathbf{z}}(t) = \arg \min_{\mathbf{z}(t)} \left\{ J_3(\mathbf{z}(t)) = -\ln p(\mathbf{z}(t) | \hat{\mathbf{s}}(t), \hat{\boldsymbol{\theta}}) \right\} \\ \hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \left\{ J_4(\boldsymbol{\theta}) = -\ln p(\boldsymbol{\theta} | \hat{\mathbf{s}}(t), \hat{\mathbf{z}}(t)) \right\} \end{cases} \quad [34]$$

Pour d'autres implémentations basées sur l'échantillonnage de la loi *a posteriori* et les loi marginales associées $p(\mathbf{S} | \mathbf{A}, \mathbf{Z}, \boldsymbol{\theta}, \mathbf{X})$, $p(\mathbf{Z} | \mathbf{S}, \boldsymbol{\theta})$, $p(\mathbf{A} | \mathbf{S}, \boldsymbol{\theta}, \mathbf{X})$ et $p(\boldsymbol{\theta} | \mathbf{A}, \mathbf{S}, \mathbf{X})$:

$$\begin{cases} \hat{\mathbf{s}}(t) \sim p(\mathbf{s}(t) | \hat{\mathbf{z}}(t), \hat{\mathbf{A}}, \hat{\boldsymbol{\theta}}, \mathbf{x}(t)) \\ \hat{\mathbf{A}} \sim p(\mathbf{A} | \hat{\mathbf{S}}, \mathbf{X}) \\ \hat{\mathbf{z}}(t) \sim p(\mathbf{z}(t) | \hat{\mathbf{s}}(t), \hat{\boldsymbol{\theta}}) \\ \hat{\boldsymbol{\theta}} \sim p(\boldsymbol{\theta} | \hat{\mathbf{s}}(t), \hat{\mathbf{z}}(t)) \end{cases} \quad [35]$$

ou encore ceux qui sont basés sur l'estimation de la matrice de mélange et des hyperparamètres en intégrant les sources en utilisant un algorithme du type EM se référer à [SNO 00, SNO 01, SNO 03, SNO 04].

0.7. Prise en compte de la corrélation temporelle

Supposer que les signaux sources $s(t)$ sont temporellement blanche est une hypothèse irréaliste. Dans la plupart des méthodes de séparation de sources, avant appliquer l'algorithme de séparation, on "blanchi" des signaux. Mais, il est possible d'inclure cette étape d'une manière optimale dans la méthode de séparation en modélisant les sources. Un outil simple pour cela est la modélisation markovienne qui peut être utilisée à la fois pour des signaux 1D ou nD. Le cas des modèles de Gauss-Markov est intéressant car, premièrement simple à implémenter, et deuxièmement liés aux notions de la fonction d'auto-corrélation et la densité spectrale de puissance des signaux. Ces modèles ont été utilisés avec succès pour la séparation de sources des signaux audio ou en séparation des images en CMB.

Un modèle de Gauss-Markov d'ordre 1 pour le cas 1D

$$\begin{aligned} p(s_j(t)|s_j(t-1)) &= \mathcal{N}(\alpha_j s_j(t-1), \sigma_j^2) \\ &\propto \exp \left[\frac{-1}{2\sigma_j^2} (s_j(t) - \alpha_j s_j(t-1))^2 \right] \end{aligned} \quad [36]$$

est équivalent d'un processus AR d'ordre 1

$$s_j(t) = \alpha_j s_j(t-1) + \eta_j(t), \quad p(\eta_j(t)) = \mathcal{N}(0, \sigma_j^2). \quad [37]$$

Ces modèles se caractérisent avec deux paramètres α et σ^2 . Mais souvent, on utilise un modèle très proche qui est caractérisé seulement avec un seul paramètre

$$\begin{aligned} p(s_j(t)|s_j(t-1)) &= \mathcal{N}(s_j(t-1), 1/(2\beta_j)) \\ &\propto \exp \left[-\beta_j (s_j(t) - s_j(t-1))^2 \right] \end{aligned} \quad [38]$$

ce qui permet d'écrire

$$p(s_j(t), t \in \mathcal{T}) \propto \exp \left[-\beta_j \sum_j (s_j(t) - s_j(t-1))^2 \right] \quad [39]$$

L'extension au cas des images s'écrit :

$$p(s_j(\mathbf{r})|s_j(\mathbf{r}'), \mathbf{r}' \in \mathcal{V}(\mathbf{r})) \propto \exp \left[-\beta_j \left(s_j(\mathbf{r}) - \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} s_j(\mathbf{r}') \right)^2 \right] \quad [40]$$

$$p(s_j(\mathbf{r}), \mathbf{r} \in \mathcal{R}) \propto \exp \left[-\beta_j \sum_{\mathbf{r} \in \mathcal{R}} \left(s_j(\mathbf{r}) - \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} s_j(\mathbf{r}') \right)^2 \right] \quad [41]$$

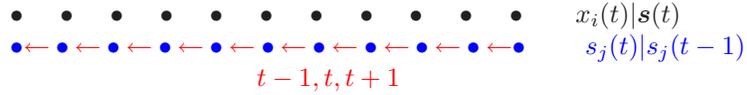


Figure 6. Un modèle de Markov pour des sources

0.8. Modélisation des sources non stationnaires

Supposer que les signaux $s(t)$ et par conséquent les observations $\mathbf{x}(t)$ sont stationnaires sur l'intervalle de l'observation $\mathcal{T} = \{1..T\}$ est trop idéal. Ceci est encore moins vrai en séparation d'images lorsque $\mathbf{s}(\mathbf{r})$ et $\mathbf{x}(\mathbf{r})$ représentent, respectivement, des images sources et des images observées. En effet, dans une image, il y a souvent des zones homogènes séparées par des contours. Un outil simple de la prise en compte de la non stationnarité est la modélisation markovienne avec des variables cachées de contours ou de régions. Une présentation très succinctes de ces différentes modélisations se suit.

0.8.1. chaîne de Markov avec variable cachée représentant des ruptures

Dans cette modélisation on introduit une variable binaire $q_j(t)$ qui représente les instants de ruptures. Deux modélisations différentes ont été proposées :

$$p(s_j(t)|s_j(t-1), q_j(t)) = \begin{cases} \mathcal{N}(\alpha_j s_j(t-1), \sigma_j^2) & \text{si } q_j(t) = 0, \\ \mathcal{N}(0, \sigma_j^2) & \text{si } q_j(t) = 1, \end{cases} \quad [42]$$

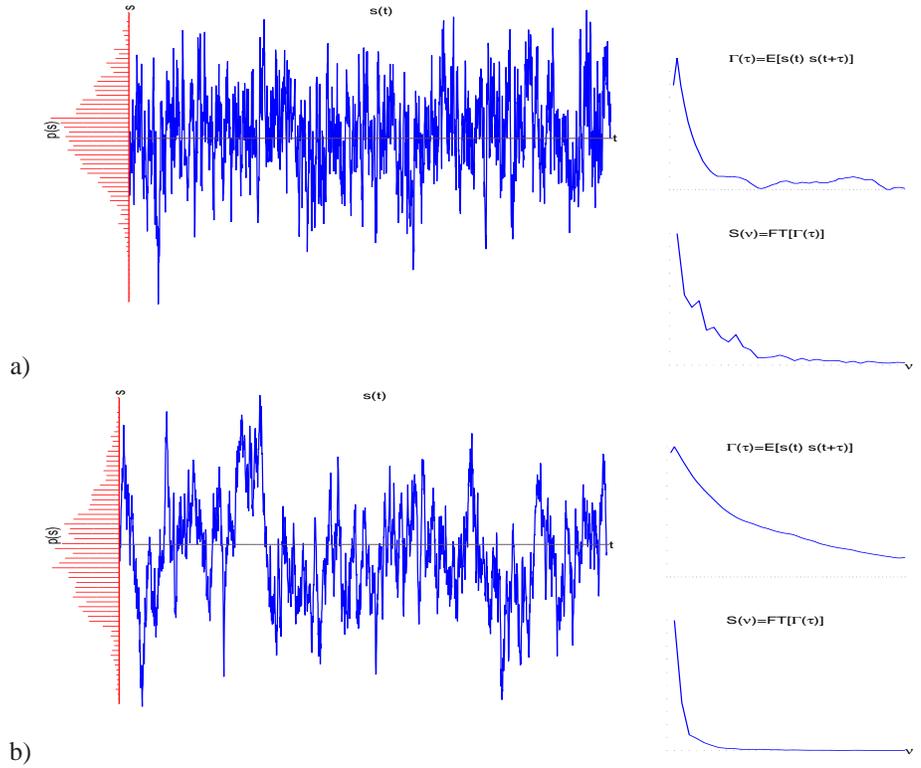


Figure 7. Deux signaux AR1 avec des coefficients, respectivement, .7 et .95. A droite de chaque signal sont représentées leurs fonctions d'autocorrélation $\Gamma(\tau)$ et leurs densité spectrale de puissances $S(\nu) = TF[\Gamma(\tau)]$.

avec deux paramètres (α_j, σ_j^2) , et

$$\begin{aligned}
 & p(s_j(t) | s_j(t-1), q_j(t)) \\
 &= \begin{cases} \mathcal{N}(s_j(t-1), 1/(2\beta_j)) & \text{si } q_j(t) = 0, \\ \mathcal{N}(0, 1/(2\beta_j)) & \text{si } q_j(t) = 1, \end{cases} \\
 &\propto \begin{cases} \exp[-\beta_j(s_j(t) - s_j(t-1))^2] & \text{si } q_j(t) = 0, \\ \exp[-\beta_j(s_j(t))^2] & \text{si } q_j(t) = 1, \end{cases} \quad [43]
 \end{aligned}$$

avec un seul paramètre $\beta_j = 2/\sigma_j^2$.

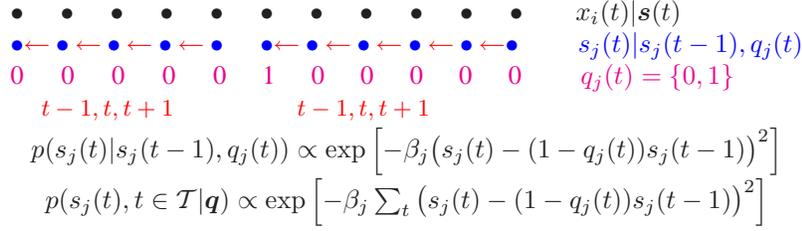
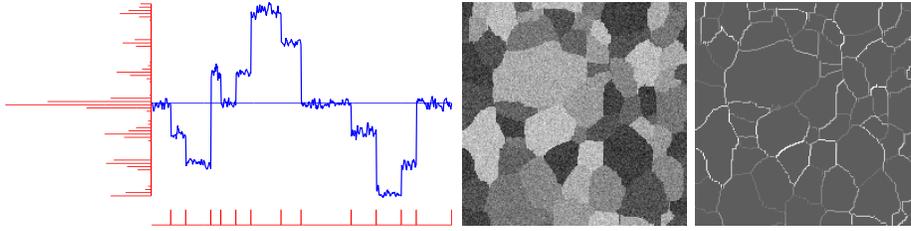


Figure 8. Chaîne de Markov avec des variables cachées représentant des ruptures (contours)



$$p(s_j(t)|s_j(t-1), q_j(t)) = \mathcal{N}((1 - q_j(t))s_j(t-1), \sigma_j^2)$$

$$\propto \exp \left[\frac{-1}{2\sigma_j^2} (s_j(t) - (1 - q_j(t))s_j(t-1))^2 \right]$$

$$p(s_j(\mathbf{r})|s_j(\mathbf{r}'), \mathbf{r}' \in \mathcal{V}(\mathbf{r})q_j(\mathbf{r})) = \mathcal{N}((1 - q_j(\mathbf{r}))\alpha_j \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} s_j(\mathbf{r}'), \sigma_j^2)$$

$$\propto \exp \left[\frac{-1}{2\sigma_j^2} (s_j(\mathbf{r}) - (1 - q_j(\mathbf{r})) \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} s_j(\mathbf{r}'))^2 \right]$$

Figure 9. Modèles markoviens avec variables contours : à gauche le cas 1D et à droite le cas 2D.

Deux cas de figures sont alors possibles :

i) on suppose les $q_j(t)$ pour des différents t indépendantes :

$$p(q_j(t) = 1) = \lambda_j, \quad p(q_j(t) = 0) = 1 - \lambda_j, \quad \forall t \longrightarrow \quad [44]$$

$$p(q_j(t), t \in \mathcal{T}) = \lambda_j^{\sum_t q_j(t)} (1 - \lambda_j)^{\sum_t (1 - q_j(t))}. \quad [45]$$

Cette modélisation est appelée Bernoulli-Gaussien.

ii) on suppose les $q_j(t)$ pour des différents t sont markoviens, ce qui nous ramène au modèle de chaîne de Markov cachée :

$$p(q_j(t) = k | q_j(t-1) = l) = p_{jkl}, \quad k, l = \{0, 1\}, \quad p_{jkl} = \begin{bmatrix} \alpha_j & 1 - \beta_j \\ 1 - \alpha_j & \beta_j \end{bmatrix}$$

pour le cas 1D et de champs de Markov caché :

$$p(q_j(\mathbf{r}), \mathbf{r} \in \mathcal{R}) \propto \exp \left[-\alpha_j \sum_{\mathbf{r}} \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} \delta(q_j(\mathbf{r}) - q_j(\mathbf{r}')) \right] \quad [46]$$

dans le cas des images.

0.8.2. Modélisation hiérarchique avec variables cachées étiquettes de classification

Une classe importante des signaux et des images est la classe des signaux homogènes par morceaux ou des images homogènes par régions. Par homogénéité des pixels d'une image se trouvant dans une région, nous entendons, la partage d'un nombre fini de propriétés statistiques, par exemple la moyenne. De plus, souvent, on peut classer ces pixels en nombre fini K de classes. Alors, pour modéliser ces signaux, on introduit une variable discrète $z_j(t)$ qui représente l'étiquette (label) de la classe et qui prends des valeurs discrètes $\{1, 2, \dots, K\}$. Le modèle le plus simple pour caractériser l'homogénéité des pixels appartenant à une classe k est :

$$p(s_j(t) | z_j(t) = k) = \mathcal{N}(m_{j_k}, \sigma_{j_k}^2) \quad [47]$$

ce qui, d'une manière équivalente revient à écrire :

$$p(s_j(t)) = \sum_{k=1}^K p(z_j(t) = k) \mathcal{N}(m_{j_k}, \sigma_{j_k}^2), \quad [48]$$

qui est une mélange de gaussiennes.

Ainsi, tous les pixels d'une image qui ont l'étiquette k partagent la même moyennes et la même variance, et on souhaiterait qu'ils soit regroupés dans une ou plusieurs régions connexes. Ce dernier souhait peut être exhaussé en imposant un lien entre les valeurs de $z_j(t)$ pour deux instants t et t' proche ou pour deux positions voisins \mathbf{r} et \mathbf{r}' . Il s'agit de modéliser $z_j(t)$ par une chaîne de Markov :

$$p(z_j(t) = k) = p_{jk}, \quad p(z_j(t) = k | z_j(t-1) = l) = p_{jkl} \quad [49]$$

et $z_j(\mathbf{r})$ par un champs de Markov :

$$p(z_j(\mathbf{r}), \mathbf{r} \in \mathcal{R}) \propto \exp \left[-\alpha_j \sum_{\mathbf{r}} \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} \delta(z_j(\mathbf{r}) - z_j(\mathbf{r}')) \right] \quad [50]$$

Notant aussi que z_j prennent des valeurs discrètes, les champs considérés sont du type Ising (K=2) ou de Potts pour le cas plus général.

Notons aussi que, dans ce cas, $z_j(\mathbf{r})$ représente une segmentation de l'image, et il y a bien sur un lien entre $z_j(t)$ et $q_j(t)$, dans le sens que, un changement de valeurs dans $z_j(t)$ mets à un la valeur de $q_j(t)$: $q_j(t) = \delta(z_j(t) - z_j(t-1))$, où $\delta(t) = 1$ si $t = 0$ et $\delta(t) = 0$ si $t \neq 0$. Les deux figures qui suivent illustrent ces propos.

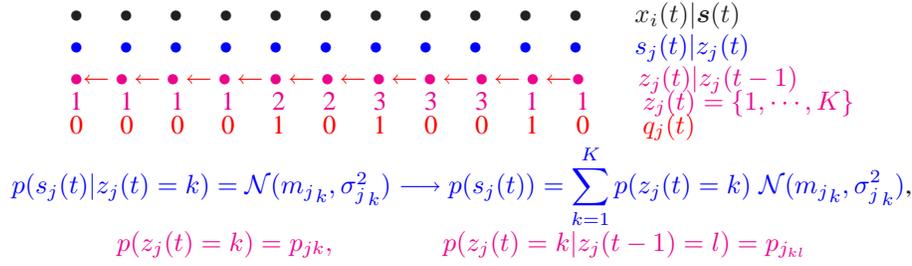
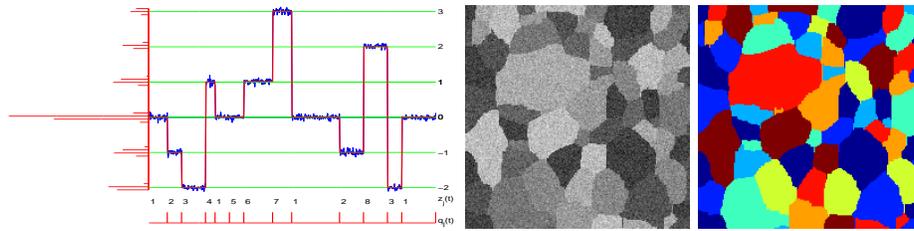


Figure 10. chaîne de Markov avec des variables cachées représentant des labels (régions).



$$p(s_j(\mathbf{r})|z_j(\mathbf{r}) = k) = \mathcal{N}(m_{jk}, \sigma_{jk}^2)$$

$$p(z_j(\mathbf{r}), \mathbf{r} \in \mathcal{R}) \propto \exp \left[-\alpha_j \sum_{\mathbf{r} \in \mathcal{R}} \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} \delta(z_j(\mathbf{r}) - z_j(\mathbf{r}')) \right]$$

Figure 11. Champs de Markov avec des variables cachées représentant des labels (régions).

0.8.3. Modélisation doublement markovienne

Dans la modélisation précédente, les pixels ayant la même étiquette ont la même moyenne et la même variance, mais il n'y a pas de corrélation locale entre ces pixels : les pixels se trouvant dans une région donnée sont considérés indépendants. Une modélisation qui remédie à cette lacune est présentée dans le schéma suivant :

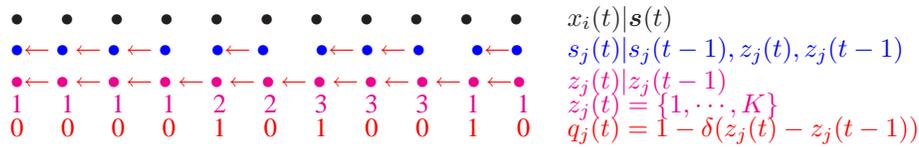
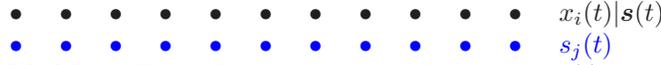
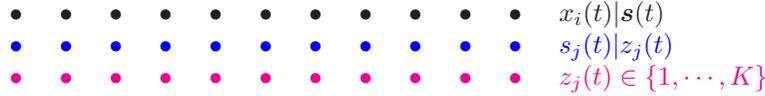


Figure 12. chaîne ou champs de Gauss-Markov avec variable cachée étiquettes de classification

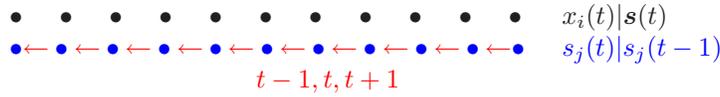
La figure suivante résume ces différentes modélisations :



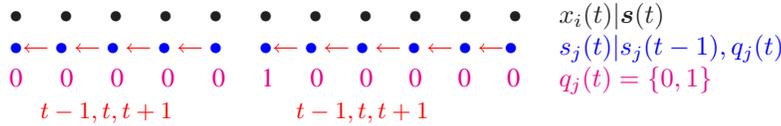
Modèle 1 : Un modèle iid simple pour les sources $s_j(t)$



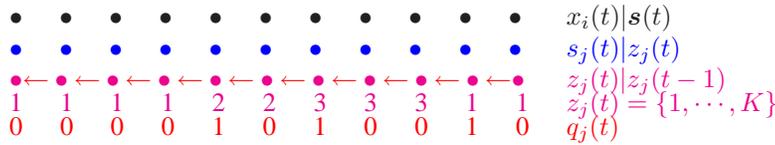
Modèle 2 : Un modèle iid pour les sources $s_j(t)|z_j(t)$ avec des variables cachées $z_j(t)$ iid, ce qui devient équivalent du **Modèle 1** : avec une lois de mélange pour les sources $s_j(t)$.



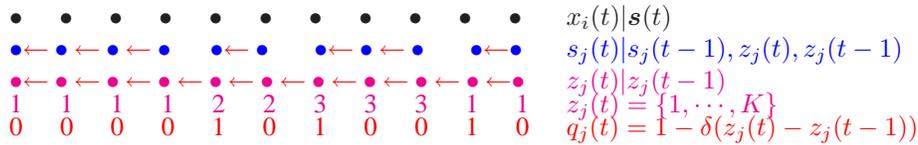
Modèle 3 : Un modèle markovian simple pour les sources $s_j(t)$ qui prend en compte la dynamique temporelle des sources, et par conséquent, des observations.



Modèle 4 : Un modèle markovien composite avec des variable cachée contours $q_j(t)$. Ce modèle permet de prendre en compte des ruptures dans des signaux.



Modèle 5 : Un modèle composite où $s_j(t)|z_j(t)$ sont iid, mais des variable cachée étiquettes $z_j(t)$ sont markoviennes (Potts). Ce modèle traduit la notion de l'homogénéité par région. Notons que les contours $q_j(t)$ sont obtenus d'une manière déterministe à l'aide de la carte des segmentation $z_j(t)$. Ces contours sont par construction fermés.



Modèle 6 : Un modèle doublement markovien avec des variable cachée étiquettes (Markov-Potts).

Figure 13. Différentes modélisation pour les sources :

Modèle 1 correspond à la modélisation classique des signaux dans la plupart des méthodes de l'ACP et de l'ACI où les sources sont considérées mutuellement indépendantes et temporellement blanches. Différents choix pour $p(s_j)$ aboutissent à des algorithmes différents. Comme nous avons vu, le choix de gaussienne généralisée (GG) est assez classique pour traduire l'éparsité des signaux sources. Cette modélisation est aussi utilisée surtout lorsque le problème de la séparation de sources est ramené dans un espace transformé (Fourier, ondelettes ou splines).

Modèle 2 aussi peut correspondre à la modélisation classique des signaux, mais ici la loi de chaque source est modélisée par un mélange de lois, par exemple la mélange de gaussiennes (MoG). En effet, la loi marginale $p(s_j(t))$ est un mélange de lois. Notons aussi que les sources $s_j(t)$ ne sont pas blanches car $s_j(t)$ dépend de $s_j(t')$, bien que conditionnellement aux z_j les s_j soient indépendantes.

Modèle 3 commence à prendre en compte la dynamique temporelle des sources, mais les sources sont toujours supposées stationnaires (au moins à l'ordre deux). Beaucoup de méthodes profitent de cette propriété de stationnarité à l'ordre deux pour proposer des méthodes de séparation dans le domaine spectral.

Modèle 4 commence à prendre en compte la présence de discontinuité dans les sources et convient pour modéliser des sources régulières par morceaux ou par régions dans le cas des images. Dans ce cas, les variables cachées $q_j(t)$ représentent des contours dans des images sources. On peut établir un lien entre ce modèle et le **Modèle 3**. En effet, dans le **Modèle 3**, lorsqu'on choisit des lois non-gaussiennes pour $p(s_j(t)|s_j(t-1))$, par exemples dans les modèles markoviens avec des potentiels non-quadratiques ou non-convexes, il apparaît, d'une manière implicite, des variables cachées liées aux contours.

Modèle 5 aussi prend en compte la présence de discontinuité dans les sources et convient pour modéliser des sources homogènes par morceaux ou par régions dans le cas des images. Dans ce cas, les variables cachées représentent des étiquettes des classifications attribuées à chaque pixel de l'image. L'ensemble des pixels d'une image ayant la même valeur d'étiquette se regroupent dans des segments (régions) connexes grâce à la modélisation markovienne de ces étiquettes. Ainsi, dans cette modélisation on estime à la fois les sources $s_j(\mathbf{r})$ et leurs segmentations $z_j(\mathbf{r})$. Tous les pixels $s_j(\mathbf{r})$ qui ont la même valeur d'étiquette partagent des paramètres communs (moyennes et variance par exemple). Ces pixels, par construction se trouvent dans un ou plusieurs régions. Cette modélisation est très riche en particulier pour les images. Nous avons utilisé différentes variations de cette modélisation pour la classification, segmentation et réduction de données dans des images hyper spectrales.

Modèle 6 est une version plus sophistiquée du **Modèle 5**. En effet, dans ce dernier, les pixels de l'image ayant la même valeur d'étiquette partagent seulement les mêmes paramètres (moyennes et variances), mais ils sont considérés indépendants. **Modèle 6**

apporte une modélisation markovienne aussi pour ces pixels. Dans cette modélisation, une fois $z_j(\mathbf{r})$ estimée, on a aussi accès aux contours $q(\mathbf{r})$ qui sont déduite d'une manière déterministes et simple à partir de $z_j(\mathbf{r})$. Cependant, la mise en œuvre de telle modélisation devienne assez complexe pour la séparation de sources.

Nous avons utilisé ces différentes modélisations dans divers domaines d'applications de la séparation de sources. Par exemple **Modèle 1** avec des lois gaussiennes généralisées (GG) est utilisé dans [ICH 03b] et un exemple de résultat est donné dans la figure qui suit.

Modèle 2 avec des lois gaussiennes généralisées (GG) est aussi utilisé dans [ICH 03b] pour modéliser les coefficients d'ondelettes des sources. En effet, les modèles de séparation de sources restent inchangés lorsqu'on effectue une transformation orthogonal sur des sources et sur des observations. On peut alors utiliser les mêmes techniques d'estimation mais dans le domaine transformées (domaines de Fourier ou ondelettes). Un avantage important de passer dans un domaine transformée, par exemple le domaine d'ondelettes, est que souvent la modélisation adaptée peut être plus simple. Par exemple, pour modéliser la lois des pixels d'une image source par un mélange de gaussiennes nécessite un nombre important de classes K , alors que, dans le domaine des ondelettes, une modélisation avec seulement deux classes peut suffire. La figure suivante montre un exemple de résultat que l'on peut obtenir en utilisant une telle modélisation.

0.9. Exemples de résultats de séparation

Dans cette section, une sélection de résultats de la séparation de sources dans divers domaines sont présentés pour montrer l'utilité de l'approche bayésienne. Ces résultats sont obtenus en utilisant les différentes modélisations décrites précédemment soit directement sur des signaux et des images directement soit sur des coefficients dans un espace transformée.

0.9.1. Séparation de sources en spectrométrie

Ces résultats concernent la spectroscopie proche infra-rouge. La figure suivante montre un exemple de simulation de 10 signaux que l'on peut observer en mélangeant 3 substances chimiques (cyclopentane (A), cyclohexane (B) and cyclohexane (C)) en supposant qu'il n'y a pas d'interaction entre ces substances. Le modèle linéaire de mélange est alors valide. Ici, il fallait prendre en compte le fait que les spectres ainsi que les éléments de la matrices de mélanges à trouver sont positifs. Le modèle utilisé est du type **Modèle 1** avec des lois *a priori* gamma pour les sources et pour les éléments de la matrices de mélange. L'algorithme implémenté ici cherche à estimer conjointement les sources, la matrice de mélange ainsi que les hyper-paramètres en maximisant, la lois *a posteriori* jointe successivement. Ces résultats sont fournis par S. Moussaoui et al. [MOU 04c].

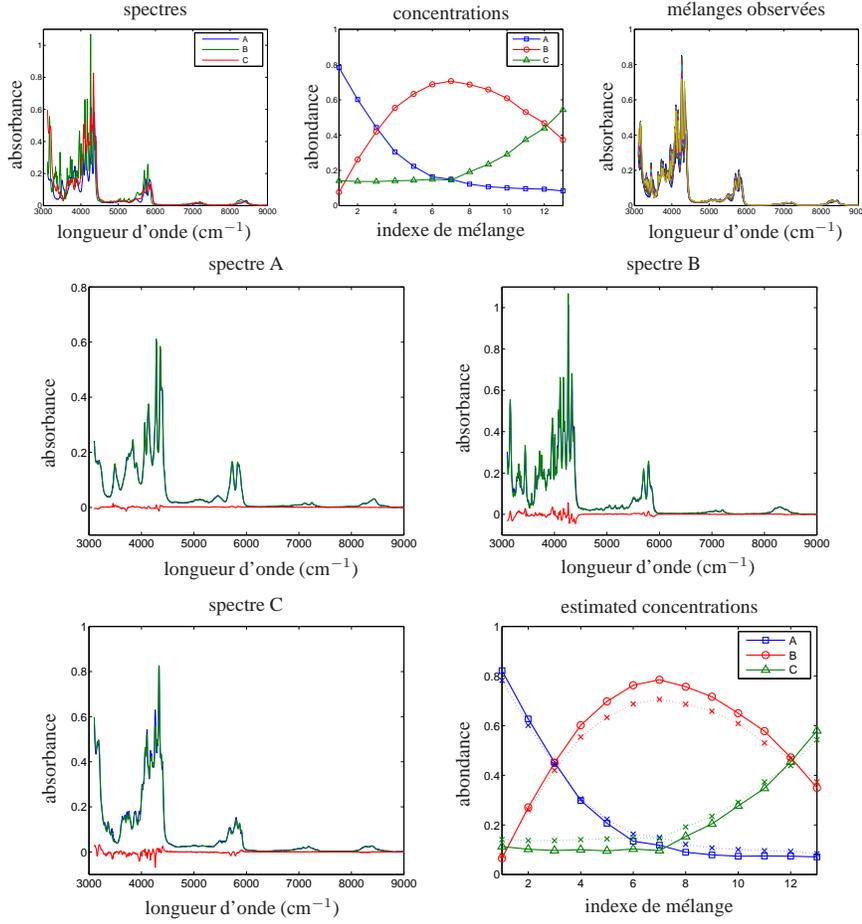


Figure 14. Séparation de sources en spectrométrie : a) les trois sources, les éléments de la matrice de mélange, et les dix observations, b) les trois sources et la matrices de mélange estimées.

Critères de performance :

$$PI = \frac{1}{2} \sum_{i=1}^n \left\{ \left(\sum_{k=1}^n \frac{|G_{(i,k)}|^2}{\max_{\ell} |G_{(i,\ell)}|^2} - 1 \right) + \left(\sum_{k=1}^n \frac{|G_{(k,i)}|^2}{\max_{\ell} |G_{(\ell,i)}|^2} - 1 \right) \right\}$$

$$CT_{S_j} = \sum_{k=1}^N (S_j(\nu_k) - \hat{S}_j(\nu_k))^2 \quad \text{avec } G = A^{\#} A.$$

	NNICA	NMF	ALS	BPSS
CT_{S_1} (dB)	-4.82	-14.20	-15.18	-33.23
CT_{S_2} (dB)	-5.64	-17.50	-23.43	-24.98
CT_{S_3} (dB)	-4.77	-17.88	-14.01	-26.05
PI(dB)	-1.02	-11.60	-8.10	-19.22

0.9.2. Séparation de sources en astrophysique

Un des problèmes de traitement d'image en astrophysique est la séparation des différents rayonnements et en particulier l'extraction de l'activité du fond cosmologique (Cosmological Microwave Background). La modélisation des observations, une fois passée dans le domaine de Fourier, s'écrit mathématiquement comme un problème de séparation de sources. L'exemple de résultat de séparation présentée ici est une simulation dont l'objectif est la séparation de trois sources à partir de six images. La modélisation utilisée ici est **Modèle 3** mais dans le domaine de Fourier. La figure suivante montre ce résultat.

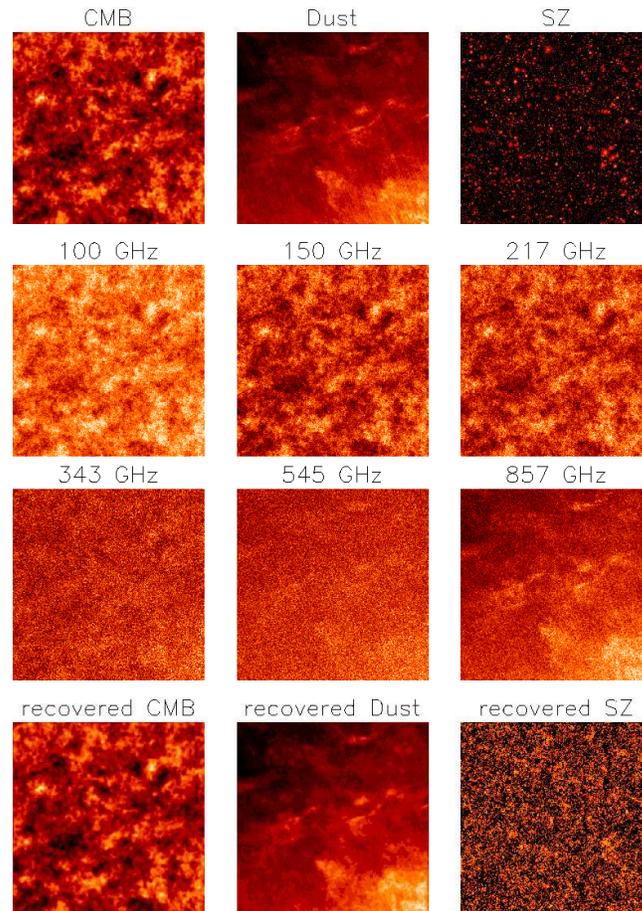


Figure 15. Séparation de sources en astrophysique : a) les 3 sources S , b) les 6 observations X , c) les trois sources estimées \hat{S} .

0.9.3. Séparation de sources en imagerie satellitaire

Il s'agit de deux exemples de résultats de simulation pour la séparation de nuages dans une imagerie satellitaire. Dans le premier exemple, la modélisation utilisée est du type **Modèle 4** où les pixels de chaque image source est supposée appartenir à deux classes [SNO 03].

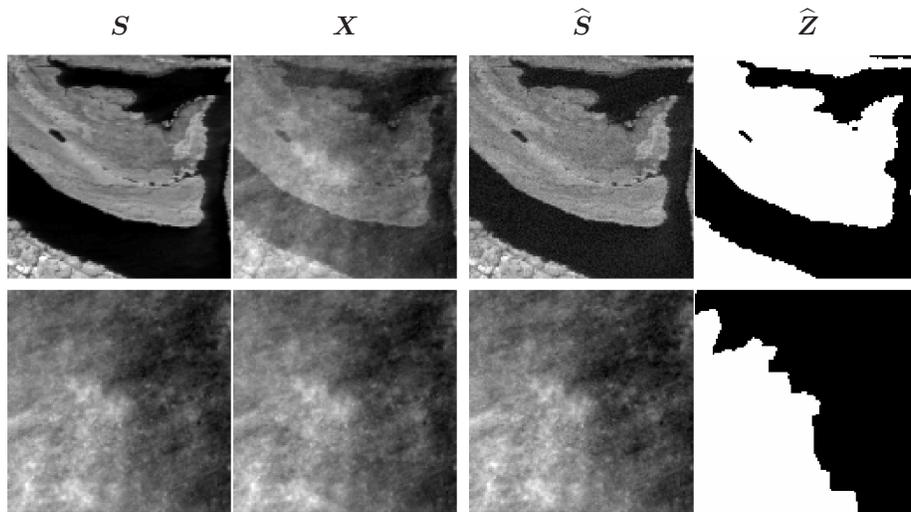


Figure 16. Séparation des nuages en imagerie satellitaire : a) les sources S , b) les observations X , c) résultats de séparation \hat{S} et d) résultats de segmentation pour chaque source \hat{Z} .

Le deuxième exemple traite ce même genre de problème, mais en effectuant la séparation dans le domaine d'ondelettes. En effet, si on applique une transformation unitaire comme par exemple une transformée en ondelettes orthogonale à l'équation de la séparation de sources, on obtient la même relation dans le domaine transformée. On peut alors effectuer la séparation de sources sur ces coefficients et ensuite reconstruire les signaux. Un avantage essentiel, souvent valide, est que la modélisation *a priori* appropriée sur ces coefficients peut être plus simple. Par exemple, un modèle du type **Modèle 1** ou **Modèle 2** peut très bien convenir. La figure qui suit montre un exemple de tels résultats avec une lois *a priori* gaussienne généralisée (GG) ou une mélange de gaussienne (MoG) sur les coefficients d'ondelettes des sources.

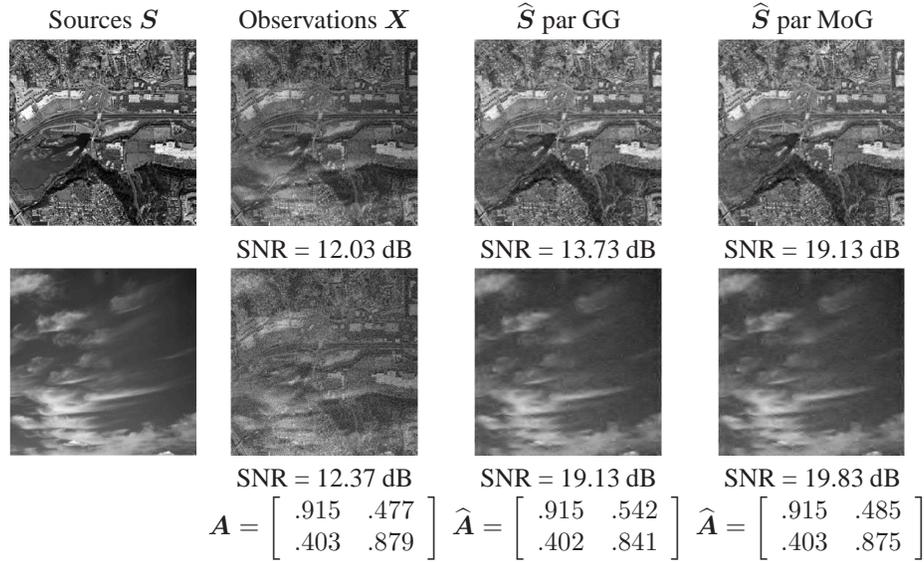


Figure 17. Séparation de sources utilisant la modélisation 1 et 2 : a) les sources \mathbf{S} , b) les observations $\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{E}$, c) résultats de séparation par la modélisation GG et d) résultats de séparation par la modélisation MoG.

0.9.4. Réduction de données et classification des images hyper-spectrales

Les données hyper-spectrales forment une cube, souvent présentées comme un ensemble d'images $x_{\omega}(\mathbf{r})$. Mais, une particularité de ces images est que la suite des pixels en une position \mathbf{r} donnée le long des bandes spectrale ω forment un spectre. C'est pourquoi, on peut aussi considérer les données hyper-spectrales comme un ensemble de spectres $x_{\mathbf{r}}(\omega)$ [SAS 87, PAR 00, BAL 05b, BAL 05a]. Évidemment, ces images ou ces spectres ne sont rarement indépendants des une et des autres. C'est pourquoi, un des problème de traitement de données dans ce domaine d'application est la réduction de données.

Si on considère ces données comme un ensemble de spectres, alors le problème de la réduction de données devient :

$$x_{\mathbf{r}}(\omega) = \sum_{k=1}^K A_{\mathbf{r},k} s_k(\omega) + \epsilon_{\mathbf{r}}(\omega) \quad [51]$$

où les $s_k(\omega)$ sont les K spectres sources que l'on souhaite mutuellement les plus indépendantes que possibles. Notons ici que les colonnes de la matrices de mélange

\mathbf{A} sont en fait des images $A_k(\mathbf{r})$. Dans un cas idéal, chacune de ses images doivent avoir tous les pixels à zéro sauf les pixels appartenant à des régions représentant un type (une classe k) de matière, et donc, associée à un spectre $s_k(\omega)$.

Si, par contre, on considère ces données comme un ensemble d'images $x_\omega(\mathbf{r})$, alors le problème de la réduction de données devient :

$$x_\omega(\mathbf{r}) = \sum_{j=1}^N A_{\omega,j} s_j(\mathbf{r}) + \epsilon_\omega(\mathbf{r})$$

où les $s_\omega(\mathbf{r})$ sont les N images sources que l'on souhaite les plus indépendantes mutuellement que possibles. Notons ici que les colonnes de la matrices de mélange \mathbf{A} sont en fait des spectres $A_j(\omega)$. Dans un cas idéal dans ce cas, chacune des images sources $s_j(\mathbf{r})$ doivent avoir tous les pixels à zéro sauf les pixels appartenant à des régions représentant un type (une classe j) de matière associée à un spectre $A_j(\omega)$. Cette explication devient plus claire si dans ces deux modèles on choisit $K = N$, ce qui signifie que les colonnes $A_{\mathbf{r},k}$ de la matrice de mélange dans le premier modèle et les images sources $s_j(\mathbf{r})$ de la deuxième modèle représentent la même chose. De même, les colonnes $A_{\omega,j}$ de la matrice de mélange dans le deuxième modèle et les images sources $s_k(\omega)$ du premier modèle représentent aussi la même chose.

A priori, on peut s'intéresser à ces deux problèmes d'une manière indépendantes. Mais, on peut aussi prendre en compte cette particularité. En effet, dans ce dernier cas, les colonnes de la matrices de mélange qu'il faut estimer sont des spectres. Elles ont donc des particularités, par exemple, elles doivent, au moins, être positives. Et dans le premier cas, les colonnes de la matrices de mélange qu'il faut estimer sont des images qui ont elles aussi des spécificités d'être homogènes par régions.

Dans les résultats que nous allons présentés ici, nous considérons ce dernier modèle. Évidemment, on impose *a priori* que les sources $s_j(\mathbf{r})$ soient mutuellement indépendantes. Mais une particularité des images hyper-spectrales est qu'elles ont toutes composée d'un ensemble fini de régions et un ensemble fini de K composants caractéristiques. Alors, si on impose aussi aux sources $s_j(\mathbf{r})$ d'avoir ces mêmes caractéristiques, on voit tout de suite qu'un des modèles *a priori* **Modèle 5** ou **Modèle 6** peut convenir.

En effet, nous avons développé une méthode basée sur le **Modèle 5** avec une particularité supplémentaire qui est de supposer que les images observées et les sources ont une segmentation commune $z(\mathbf{r})$ mais inconnue. Par ailleurs, un des objectifs des utilisateurs de telles images est de trouver cette segmentation d'une manière supervisée ou non. La figure suivante montre deux exemples de résultats que l'on obtient par ce type de méthode.

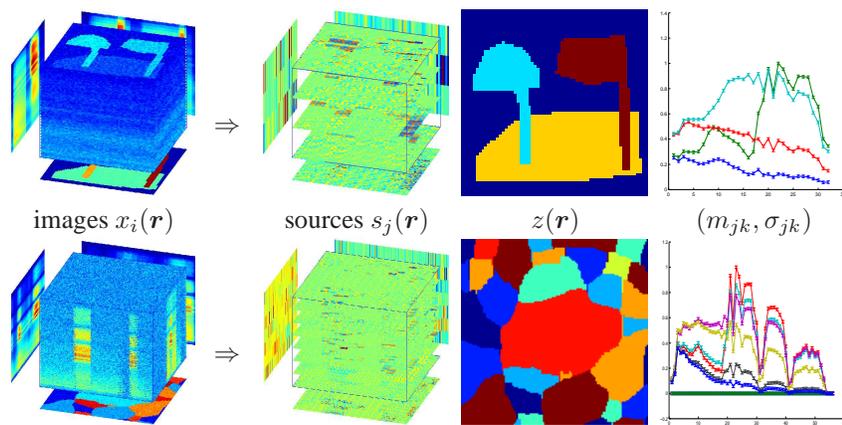


Figure 18. Réduction de données et classification des images hyper-spectrales : a) les 32 images $x_i(\mathbf{r})$ d'un cube hyperspectrale, b) les 8 sources $s_j(\mathbf{r})$ estimées, c) la segmentation commune $z(\mathbf{r})$ estimée, d) les moyennes m_{jk} et les variances σ_{jk}^2 estimées des spectres associé à chaque classe. Les moyennes m_{jk} représentent les colonnes de la matrice \mathbf{A} et les σ_{jk} leurs écarts-types.

Bibliographie

- [ATT 98] ATTIAS H., « Independent Factor Analysis », *Neural Computation*, vol. 11, p. 803–851, 1998.
- [ATT 99] ATTIAS H., « Independent factor analysis », *Neural Computation*, vol. 11, n°4, p. 803-851, 1999.
- [BAL 05a] BALI N., MOHAMMAD-DJAFARI A., « Approxiamtion en champ moyen pour la séparation de sources appliqué aux images hyperspectrales », *Actes 20^e coll. GRETSI*, Louvains Belgium, Sep. 2005.
- [BAL 05b] BALI N., MOHAMMAD-DJAFARI A., « Mean Field Approximation for BSS of images with compound hierarchical Gauss-Markov-Potts model », *MaxEnt05, San José CA, US*, American Institute of Physics (AIP), Aug. 2005.
- [BEL 94] BELOUHRANI A., CARDOSO J.-F., « Maximum likelihood source separation for discrete sources », *EUSIPCO'94*, 1994.
- [BEL 95] BELOUHRANI A., Séparation autodidacte de sources : Algorithmes, Performances et Application à des signaux expérimentaux, PhD thesis, Ecole Nationale Supérieure des Télécommunications, 1995.
- [BRO 97] BRO R., DE JONG S., « A Fast Non-negativity Constrained Least Squares Algorithm », *Journal of Chemometrics*, vol. 11, p. 393–401, 1997.
- [CAM 81] CAMPBELL S.-I., POOLE G.-D., « Computing non-negative rank factorizations », *Linear Algebra and its Applications*, vol. 35, p. 175–182, 1981.
- [CAR 96] CARDOSO J., LABELD B., « Equivariant adaptative source separation », *Signal Processing*, vol. 44, p. 3017-3030, 1996.
- [CAR 97] CARDOSO J. F., « Infomax and maximum likelihood for source separation », *IEEE Letters on Signal Processing*, vol. 4, p. 112-114, avril 1997.
- [CAR 98] CARDOSO J., « Blind signal separation : statistical principles », *Proc. IEEE*, vol. 9, p. 2009-2025, 1998.
- [CAR 99] CARDOSO J.-F., « High-Order Contrasts for Independent Component Analysis », *Neural Computation*, vol. 11, n°1, p. 157-192, 1999.

- [COM 91] COMON P., JUTTEN C., HÉRAULT J., « Blind Separation of Sources, Part II : Problems Statement », *Signal Processing*, vol. 24, p. 11-20, 1991.
- [COM 94] COMON P., « Independent Component Analysis – a new concept ? », *Signal Processing*, vol. 36, p. 287–314, 1994.
- [DEL 99] DELYON B., LAVIELLE M., MOULINES E., « Convergence of a stochastic approximation version of the EM algorithm », *The Annals of Statistics*, vol. 27, p. 94-128, 1999.
- [HYV 00] HYVARINEN A., OJA E., « Independent Component Analysis : Algorithms and Applications », *Neural Networks*, vol. 13, p. 411–430, 2000.
- [ICH 03a] ICHIR M. M., MOHAMMAD-DJAFARI A., « Bayesian Wavelet Based Statistical Signal and Image Separation », *Bayesian Inference and Maximum Entropy Methods*, Jackson Hole (WY), USA, MaxEnt Workshops, Amer. Inst. Physics, p. 417-428, Aug. 2003, Maxent03.
- [ICH 03b] ICHIR M. M., MOHAMMAD-DJAFARI A., « Séparation de sources modélisées par des ondelettes », *Actes 19^e coll. GRETSI*, Paris, France, Sep. 2003.
- [ICH 03c] ICHIR M. M., MOHAMMAD-DJAFARI A., « Wavelet Domain Blind Image Separation », UNSER M. A., ALDROUBI A., LAINE A. F., Eds., *Wavelets : Applications in Signal and Image Processing X*, vol. 5207 de *International Conference on Electronic Imaging*, SPIE, Bellingham, WA, p. 361-370, Aug. 2003.
- [JAY 68] JAYNES E. T., « Prior Probabilities », *IEEE Trans. Systems Science and Cybern.*, vol. SSC-4, n°3, p. 227–241, 1968.
- [JUT 91] JUTTEN C., HÉRAULT J., « Blind Separation of Sources .1. an Adaptive Algorithm Based on Neuromimetic Architecture », *Signal Processing*, vol. 24, n°1, p. 1-10, 1991.
- [JUT 00] JUTTEN C., « Source separation : from dusk till dawn », *Proc. of 2nd Int. Workshop on Independent Component Analysis and Blind Source Separation (ICA'2000)*, Helsinki, Finland, p. 15–26, 2000.
- [KNU 99] KNUTH K., « A Bayesian approach to source separation », CARDOSO J.-F., JUTTEN C., LOUBATON P., Eds., *Proceedings of International Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, Aussios, France, p. 283–288, 1999.
- [KNU 01] KNUTH K., « Source separation as an exercise in logical induction. », MOHAMMAD-DJAFARI A., Ed., *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, Paris, France, American Institute of Physics, p. 340–349, 2001.
- [MOH 99] MOHAMMAD-DJAFARI A., « A Bayesian approach to source separation », *proceedings of International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt'99)*, vol. 567, American Institute of Physics (AIP) proceedings, p. 221–244, 1999.
- [MOU 03a] MOUSSAOUI S., CARTERET C., MOHAMMAD-DJAFARI A., CASPARY O., BRIE D., HUMBERT B., « Approche bayésienne pour l'analyse de mélanges en spectroscopies. », *Chimiométrie'2003*, Paris, France, Décembre 2003.
- [MOU 03b] MOUSSAOUI S., DJAFARI A. M., BRIE D., CASPARY O., CARTERET C., HUMBERT B., « A Bayesian Method for Positive Source Separation : Application to Mixture Analysis in Spectroscopy », *IAR'2003*, Duisburg, Germany, novembre 2003.

- [MOU 04a] MOUSSAOUI S., BRIE D., CARTERET C., MOHAMMAD-DJAFARI A., « Application of Bayesian Non-negative Source Separation to Mixture Analysis in Spectroscopy », FISCHER R., PREUSS R., VON TOUSSAINT U., Eds., *International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt'2004)*, vol. 735, Munich, Allemagne, American Institute of Physics (AIP), p. 237-244, 25-30 Juillet 2004.
- [MOU 04b] MOUSSAOUI S., MOHAMMAD-DJAFARI A., BRIE D., CASPARY O., « A Bayesian Method for Positive Source Separation », *proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'2004)*, Montreal, Canada, p. 485–488, May 2004.
- [MOU 04c] MOUSSAOUI S., BRIE D., CASPARY O., MOHAMMAD-DJAFARI A., « A Bayesian method for positive source separation », *ICASSP04*, Oct. 2004.
- [PAR 00] PARRA L., SPENCE C., ZIEHE A., MUELLER K.-R., SAJDA P., « Unmixing hyper-spectral data », *Advances in Neural Information Processing Systems 13, (NIPS'2000)*, MIT Press, p. 848–854, 2000.
- [PHA 96] PHAM D.-T., « Blind separation of instantaneous mixture sources via independent component analysis », *IEEE Trans. Signal Processing*, vol. 44, 1996.
- [ROB 92] ROBERT C., *L'analyse statistique bayésienne*, Economica, Paris, France, 1992.
- [ROB 97] ROBERT C. P., *The Bayesian Choice. A Decision-Theoretic Motivation*, Springer Texts in Statistics, Springer Verlag, New York, NY, 1997.
- [ROB 99] ROBERT C., *Monte Carlo Statistical Methods*, SPRINGER-VERLAG, BERLIN, 1999.
- [ROB 01] ROBERT C., *The Bayesian Choice*, SPRINGER-VERLAG, 2ND ÉDITION, 2001.
- [ROB 03] ROBERTS S., CHOUDREY R., « DATA DECOMPOSITION USING INDEPENDENT COMPONENT ANALYSIS WITH PRIOR CONSTRAINTS », *Pattern Recognition*, VOL. 36, p. 1813–1825, 2003.
- [SAS 87] SASAKI K., KAWATA S., MINAMI S., « COMPONENT ANALYSIS OF SPATIAL AND SPECTRAL PATTERNS IN MULTISPECTRAL IMAGES. I. BASICS », *Journal of the Optical Society of America. A*, VOL. 4, N°11, p. 2101–2106, 1987.
- [SNO 00] SNOUSSI H., MOHAMMAD-DJAFARI A., « BAYESIAN SOURCE SEPARATION WITH MIXTURE OF GAUSSIANS PRIOR FOR SOURCES AND GAUSSIAN PRIOR FOR MIXTURE COEFFICIENTS », MOHAMMAD-DJAFARI A., ED., *Bayesian Inference and Maximum Entropy Methods*, GIF-SUR-YVETTE, FRANCE, PROC. OF MAXENT, AMER. INST. PHYSICS, p. 388–406, JULY 2000.
- [SNO 01] SNOUSSI H., MOHAMMAD-DJAFARI A., « UNSUPERVISED LEARNING FOR SOURCE SEPARATION WITH MIXTURE OF GAUSSIANS PRIOR FOR SOURCES AND GAUSSIAN PRIOR FOR MIXTURE COEFFICIENTS. », J.MILLER D., ED., *Neural Networks for Signal Processing XI*, IEEE WORKSHOP, p. 293-302, SEP. 2001.
- [SNO 02] SNOUSSI H., MOHAMMAD-DJAFARI A., « INFORMATION GEOMETRY AND PRIOR SELECTION. », WILLIAMS C., ED., *Bayesian Inference and Maximum Entropy Methods*, MAXENT WORKSHOPS, AMER. INST. PHYSICS, p. 307–327, AUG. 2002.

[SNO 03] SNOUSSI H., APPROCHE BAYÉSIENNE EN SÉPARATION DE SOURCES. APPLICATIONS EN IMAGERIE., PHD THESIS, UNIVERSITÉ DE PARIS-SUD, ORSAY, FRANCE, SEP. 2003.

[SNO 04] SNOUSSI H., MOHAMMAD-DJAFARI A., « FAST JOINT SEPARATION AND SEGMENTATION OF MIXED IMAGES », *Journal of Electronic Imaging*, VOL. 13, N°2, P. 349-361, APR. 2004.