

# Séparation de sources modélisées par des ondelettes

Mahieddine M. ICHIR, Ali MOHAMMAD-DJAFARI

Laboratoire des Signaux et Systèmes  
3, rue Joliot Curie, Supélec, Plateau de Moulon, 91192 Gif sur Yvette, France  
ichir@lss.supelec.fr, djafari@lss.supelec.fr

**Résumé** – Dans ce travail nous considérons le problème de séparation de sources aveugle dans le cadre de l'estimation bayésienne. Nous transportons le problème dans le domaine d'ondelettes où la distribution des coefficients d'ondelettes des signaux est modélisée par une loi du type exponentiel double. Dans un premier temps, les coefficients sont supposés indépendants et identiquement distribués. Dans un second temps nous leur attribuons un modèle hiérarchique markovien du premier ordre pour prendre en compte une corrélation inter-échelles. Un algorithme MCMC du type Hastings-Metropolis/Gibbs est développé pour l'estimation des grandeurs d'intérêts (sources, matrice de mélange, variance du bruit et hyperparamètres des lois *a priori*). Des résultats de simulations et des comparaisons sont présentés.

**Abstract** – In this work, we consider the problem of blind source separation in a Bayesian framework. We transport the problem to the wavelet domain where the distribution of the wavelet coefficients of the sources is modeled by a double exponential probability law. In a first step, the wavelet coefficients are supposed independent and identically distributed. In a second step, we assign to the coefficients a hierarchical first order markovian model to account for an inter-scale correlation. A MCMC Hastings-Metropolis/Gibbs algorithm is developed for the estimation of the interest quantities (sources, mixing matrix, noise variance and hyperparameters of the prior laws). Simulation results and comparisons are presented.

## 1 Introduction

La séparation de sources est un domaine de recherche important en traitement du signal et en analyse de données. L'analyse en composantes indépendantes (ACI) [3] est l'une des méthodes les plus développée. Cependant, dans certaines applications, l'ACI s'avère inadaptée et inefficace, particulièrement dans un mélange bruité et/ou dans la cas où le modèle de mélange ne correspond pas totalement à la réalité du processus de mesure.

L'estimation bayésienne a déjà fait ses preuves pour la résolution du problème de séparation de sources [4, 9, 8]. Elle permet d'une manière naturelle de prendre en compte toutes les incertitudes et toute la connaissance *a priori* sur le modèle d'observation des signaux: indépendants ou corrélés temporellement et/ou spatialement, qui est une étape importante dans cette approche.

Le problème de séparation de sources a été traité soit directement dans le domaine des signaux, soit dans un domaine transformé: Fourier [7], Ondelettes [10, 2]. L'idée de traiter le problème dans un domaine transformé est basée sur le fait que souvent une transformée linéaire et inversible restructure les données, donnant aux coefficients une structure plus simple à modéliser.

La transformée en ondelettes se trouve être une représentation particulièrement parcimonieuse des signaux/images (non stationnaires). Cette propriété lui vaut son succès dans diverses domaines du traitement du signal: codage, compression et débruitage, mais son utilisation pour la séparation des sources est très récente [10, 2].

## 2 Approche bayésienne

Le modèle de séparation de sources considéré est un modèle de mélange linéaire, instantané et bruité de la forme:

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

pour  $t = 1, \dots, T$ , où  $\mathbf{x}(t)$  représente le vecteur des données observées,  $\mathbf{A}$  et  $\mathbf{s}(t)$  représentent la matrice de mélange et le vecteur des sources et  $\epsilon(t)$  représente le bruit de mesure. Le même modèle (eq. 1) est valide dans le domaine transformé des ondelettes.

Dans ce qui suit, nous désignerons par  $\mathbf{x}_j(k)$  le vecteur des coefficients d'ondelettes ( $k$ ) à l'échelle ( $j$ ) du vecteur  $\mathbf{x}(t)$ , et (eq. 1) s'écrit alors:

$$\mathbf{x}_j(k) = \mathbf{A}\mathbf{s}_j(k) + \epsilon_j(k) \quad (2)$$

pour  $j = 1, \dots, J$  et  $k = 1, \dots, T/2^j$ .

Dans le cadre de l'approche bayésienne, la distribution *a posteriori* s'écrit:

$$p(\mathbf{s}, \mathbf{A}, \theta | \mathbf{x}) \propto l(\mathbf{s}, \mathbf{A}, \theta | \mathbf{x}) \pi(\mathbf{s}, \mathbf{A} | \theta) \pi(\theta) \quad (3)$$

où  $(\mathbf{s}, \mathbf{x}) = \{\mathbf{s}_j(k), \mathbf{x}_j(k)\}$ ,  $l(\mathbf{s}, \mathbf{A}, \theta | \mathbf{x})$  est la vraisemblance du modèle (eq. 2) donnée par la distribution du bruit et  $\pi(\mathbf{s}, \mathbf{A} | \theta)$  est la loi qui reflète notre connaissance *a priori* sur les coefficients des sources et de la matrice de mélange.  $\pi(\theta)$  est la loi qui traduit un certain comportement des hyperparamètres introduit pour la description statistique du modèle (eq. 2) (positivité de la variance du bruit par exemple).

### 3 Modélisation statistique des coefficients d'ondelettes

La transformée en ondelettes permet d'avoir une représentation parcimonieuse des signaux [6].



FIG. 1: a) image de Lena, b) sa transformée en ondelettes

La transformée en ondelettes d'un signal/image comme on le voit dans la figure 1, résulte en un grand nombre de petits coefficients et en un petit nombre de grands coefficients. Cette propriété peut être modélisée par le choix d'une loi de probabilité piquée et à longues queues tel que le mélange de gaussiennes [1] ou la loi exponentielle généralisée [5]. Dans cette communication nous mettons l'accent sur les distributions exponentielles généralisées :

$$\pi(x|\gamma, \alpha) = K \exp\left(-\frac{1}{2\gamma}|x|^\alpha\right) \quad (4)$$

où  $K$  est une constante de normalisation et  $(\alpha, \gamma)$  sont tel que  $(1 \leq \alpha \leq 2, \text{ et } \gamma > 0)$ .

Dans la figure 2, nous avons représenté les coefficients d'ondelettes d'un signal unidimensionnel (en valeurs absolues) et nous remarquons que les coefficients ont tendance à se propager à travers les échelles. C'est une propriété que nous prenons en compte pour enrichir le modèle (eq. 4). Nous modéliserons ainsi les coefficients d'ondelettes par une loi *a priori* de la forme:

$$\pi(s_j(k)|\mathcal{S}_{j-1}) \propto \exp\left(-\frac{1}{2}d_j^*|s_j(k) - \mathcal{A}_j\mathcal{P}_j(k)|^{\alpha_j}\right) \quad (5)$$

où l'ensemble  $(\mathcal{S}_{j-1})$  définit les parents de  $s_j(k)$  dans une modélisation hiérarchique (figure 3).

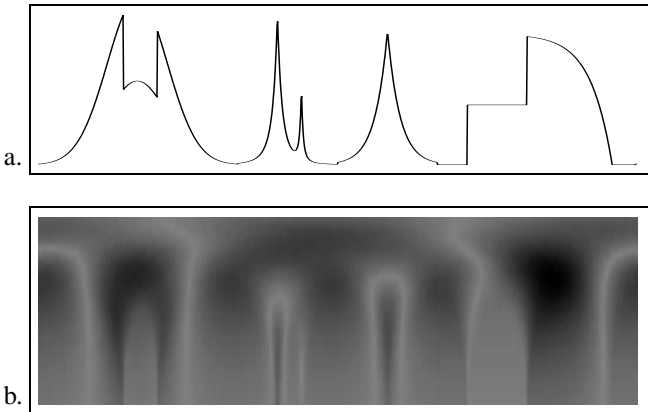


FIG. 2: a. signal unidimensionnel, b. transformée en ondelettes continue du signal (valeurs absolues)

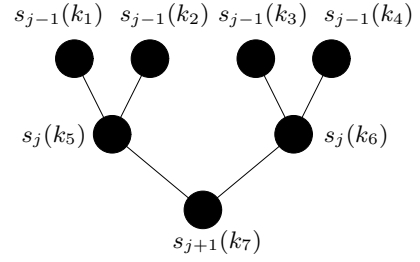


FIG. 3: modèle graphique décrivant la corrélation inter-échelle de la représentation multi-échelle (1-D)

Un coefficient  $s_j(k)$  à l'échelle  $(j)$  dépend d'un ensemble de coefficients  $\{s_{j-1}\}_{k \in \mathcal{P}}$ , à l'échelle  $(j-1)$  désignés par  $\mathcal{P}_j(k)$  (les parents à l'échelle plus fine).  $\mathcal{A}_j$  définit une transition markovienne que nous décrivons plus formellement comme étant un opérateur d'échelle d'une résolution à une autre.

Cependant les coefficients à une résolution donnée seront supposés i.i.d. On peut trouver une description plus détaillée et plus complète d'un tel modèle dans [1]. Nous remarquons que le modèle décrit dans [1] est un modèle hiérarchique (connu sous le nom de "Quad Tree"), qui commence par lier le coefficient de la résolution la plus grossière (l'ancêtre) aux coefficients des résolutions les plus fines (les descendants).

### 4 Algorithme MCMC

Comme estimés des sources  $s_j(k)$ , de la matrice de mélange  $\mathbf{A}$  et des hyperparamètres  $\theta = [\mathbf{R}_\epsilon, \{\gamma_j\}]$ , nous prenons l'espérance de la loi *a posteriori* (eq. 3), où:

1. les coefficients sont modélisés par des lois exponentielles *a priori* (eq. 5), indépendantes de la matrice de mélange,
2. les éléments de la matrice de mélange sont modélisés par une loi *a priori* gaussienne de moyenne  $\mu_A$  et de matrice de covariance  $\mathbf{R}_A$ :

$$\pi(\mathbf{A}|\mu_A, \mathbf{R}_A) = \mathcal{N}(\mu_A, \mathbf{R}_A), \quad (6)$$

3. les paramètres  $(\sigma_i^2, \{\gamma_j\})$  sont modélisés par des lois *a priori* inverse gamma (traduisant le caractère positif de ces paramètres):

$$\pi(x|\nu, \beta) = \mathcal{IG}(x|\nu, \beta) \propto \frac{e^{-\beta x}}{x^{\nu+1}} \quad (7)$$

Sous l'hypothèse d'un bruit blanc gaussien de matrice de covariance  $\mathbf{R}_\epsilon = \text{diag}(\{\sigma_i^2\})$ , la vraisemblance est donnée par:

$$l(s_j(k), \mathbf{A}, \theta|x_j(k)) = \mathcal{N}(\mathbf{A}s_j(k), \mathbf{R}_\epsilon) \quad (8)$$

La loi *a posteriori* (eq. 3) s'écrit alors:

$$p(s_j(k), \mathbf{A}, \theta|x) \propto \mathcal{N}(\mathbf{A}s_j(k), \mathbf{R}_\epsilon) \mathcal{N}(\mu_A, \mathbf{R}_A) \mathcal{Exp}(s_j(k)|\gamma_j) \mathcal{IG}(\theta|\nu, \beta) \quad (9)$$

pour  $j = 1, \dots, J$  et  $k = 1, \dots, T/2^j$

Nous faisons recours à un algorithme MCMC (Monté Carlo par Chaîne de Markov) pour générer des échantillons suivant (asymptotiquement) la loi *a posteriori* (eq. 9). Mais vu la forme de cette loi (pas directement échantillonnable), nous utilisons un algorithme MCMC hybride, l'algorithme de Hastings - Metropolis / Gibbs.

## L'échantillonnage par Gibbs

L'échantillonnage des grandeurs d'intérêts  $[s_j(k), \mathbf{A}$  et  $\theta]$  est fait d'une manière alternée suivant leurs lois conditionnelles comme décrit ci-dessous.

### Algorithme de Gibbs

à l'itération (t),

1.  $s_j^t(k) | \{\mathbf{A}^{t-1}, \theta^{t-1}\} \sim \mathcal{N}(\mathbf{A}s_j(k), \mathbf{R}_\epsilon) \mathcal{E}xp(s_j(k) | \gamma_j)$   
voir l'étape de Hastings-Metropolis

2.  $\mathbf{A}^t | \{\mathcal{S}^t, \theta^{t-1}\} \sim \mathcal{N}(\mu, \mathbf{R})$

où

$$\mathbf{R} = \left( \sum_{j,k} \mathbf{R}_\epsilon^{-1} \otimes C_{ss} + \mathbf{R}_A^{-1} \right)^{-1},$$

$$\mu = \mathbf{R} \left( (\mathbf{R}_\epsilon^{-1} \otimes \mathbb{I}_n) \sum_{j,k} C_{xs} + \mu_A \right)$$

$$\mathcal{S} = \bigcup_{j,k} s_j(k), C_{ss} = s_j(k) s_j^*(k), C_{xs} = \mathbf{x}_j(k) \otimes s_j(k)$$

3.  $\{\sigma_i^2\}^t | \{\mathcal{S}^t, \mathbf{A}^t\} \sim \mathcal{IG}(\nu', \beta')$

où

$$\nu' = T/2 + \nu,$$

$$\beta' = \left( \frac{1}{2} \sum_k (\mathbf{x}_i(k) - [\mathbf{A}\mathbf{s}(k)]_i)^2 + \beta \right)$$

4.  $\{\gamma_j\}_i^t | \{\mathcal{S}^t, \mathbf{A}^t\} \sim \mathcal{IG}(\nu', \beta')$

où

$$\nu' = \frac{T/2^j}{\alpha_j} + \nu,$$

$$\beta' = \left( \frac{1}{2} \sum_k |s_j(k) - \mathcal{A}_j \mathcal{P}_j(k)|_i^{\alpha_j} + \beta \right)$$

### L'Étape de Hastings-Metropolis

La première étape de Gibbs pour l'échantillonnage des coefficients des sources consiste à générer des échantillons suivant une loi *a posteriori* conditionnelle qui est le produit d'une gaussienne (vraisemblance) et de l'*a priori* exponentielle des sources  $\pi(s_j(k))$ . De ce fait, cette loi n'est pas directement échantillonnable. Nous introduisons donc une étape de Hastings-Metropolis pour approcher cette loi. Nous approximons l'*a priori* exponentiel des coefficients (eq. 5) par une loi de la même famille, la gaussienne:

$$\pi(s_j(k)) \sim \tilde{\pi}(s_j(k)) = \mathcal{N}(\mathcal{A}_j \mathcal{P}_j, \mathbf{R}) \quad (10)$$

où  $\mathbf{R} = \text{diag}(\gamma_1, \dots, \gamma_n)$ .

La loi *a posteriori* approchée est donc une gaussienne:

$$\tilde{p}(s_j(k)) \propto \mathcal{N}(\tilde{\mu}, \tilde{\mathbf{R}}) \quad (11)$$

où

$$\tilde{\mathbf{R}} = \left( \mathbf{A}^* \mathbf{R}_\epsilon^{-1} \mathbf{A} + \mathbf{R}_{\gamma_j}^{-1} + \mathcal{A}_{j+1}^* \mathbf{R}_{\gamma_{j+1}}^{-1} \mathcal{A}_{j+1} \right)^{-1},$$

$$\begin{aligned} \tilde{\mu} &= \tilde{\mathbf{R}} \left( \mathbf{A}^* \mathbf{R}_\epsilon^{-1} \mathbf{x}_j(k) + \mathbf{R}_{\gamma_j}^{-1} \mathcal{A}_j \mathcal{P}_j(k) \right. \\ &\quad \left. + \mathcal{A}_{j+1}^* \mathbf{R}_{\gamma_{j+1}}^{-1} (s_{j+1}(k) - \mathcal{A}_{j+1} \mathcal{N}_j(k)) \right) \end{aligned}$$

où  $\mathcal{N}_j(k)$  définit un système de voisinage de  $s_j(k)$ .

Les expressions de  $\tilde{\mu}$  et de  $\tilde{\mathbf{R}}$  se simplifient dans le cas où les coefficients sont supposés indépendants, et s'écrivent:

$$\tilde{\mathbf{R}} = \left( \mathbf{A}^* \mathbf{R}_\epsilon^{-1} \mathbf{A} + \mathbf{R}_{\gamma_j}^{-1} \right)^{-1},$$

$$\tilde{\mu} = \tilde{\mathbf{R}} \mathbf{A}^* \mathbf{R}_\epsilon^{-1} \mathbf{x}_j(k)$$

L'étape de Hastings-Metropolis est décrite ci-dessous.

### Étape de Hastings-Metropolis

à l'itération (t),

1.1.  $\mathbf{y} | \{\mathbf{A}^{t-1}, \theta^{t-1}\} = \mathbf{U}\mathbf{z} + \tilde{\mu}$

où

$$\mathbf{z} \sim \exp\left(-\frac{1}{2} \mathbf{d}^* |\mathbf{z}|\right),$$

$$\tilde{\mathbf{R}} = \mathbf{U} \mathbf{A} \mathbf{U}^*, \mathbf{A} = \text{diag}(\gamma_1, \dots, \gamma_n),$$

$$\mathbf{d}^* = [\gamma_1^{-1}, \dots, \gamma_n^{-1}]$$

1.2.

$$s_j^t(k) = \begin{cases} \mathbf{y} & \text{avec prob. } \rho, \\ s_j^{t-1}(k) & \text{avec prob. } 1 - \rho \end{cases}$$

où

$$\rho = \left\{ 1 \wedge \left( \frac{p(\mathbf{y})}{g(\mathbf{y})} \bigg/ \frac{p(s_j^{t-1}(k))}{g(s_j^{t-1}(k))} \right) \right\},$$

$$p(s_j(k) | \mathbf{A}^{t-1}, \theta^{t-1}) \propto \mathcal{N}(\mathbf{A}s_j(k), \mathbf{R}_\epsilon) \mathcal{E}xp(s_j(k) | \gamma_j)$$

$$g(s_j(k)) \propto \exp\left(-\frac{1}{2} \mathbf{d}^* |\mathbf{U}^*(s_j(k) - \tilde{\mu})|\right)$$

## 5 Simulations

Nous avons testé l'algorithme décrit ci-dessus sur des données simulées. Les résultats sont présentés dans la figure 4. Deux images (figure 4.a) sont mélangées par une matrice

$$\mathbf{A} = \begin{bmatrix} 0.875 & 0.508 \\ 0.484 & 0.861 \end{bmatrix}$$

et un bruit de 20dB est ajouté sur chacune des images obtenues (figure 4.b). Les estimés des sources obtenues avec une modélisation i.i.d sont présentées dans la figure 4.c, et les estimés obtenues avec prise en compte d'une corrélation inter-échelles sont présentées dans la figure 4.d.

Pour quantifier les estimés, nous avons choisi une mesure de puissance de l'erreur de l'estimation normée donnée par:

$$P_\beta(\tilde{\mathbf{S}}) = \frac{\|\mathbf{S} - \tilde{\mathbf{S}}\|_\beta}{\|\mathbf{S}\|_\beta}, \quad 1 \leq \beta \leq 2 \quad (12)$$

Nous avons présentés des résultats numériques dans le tableau 1. Nous remarquons qu'en ce qui concerne les sources, c'est difficile de dire si une modélisation est meilleure qu'une autre. Cependant, pour l'estimation de la matrice de mélange, la modélisation qui prend en compte une corrélation inter-échelles offre une estimation légèrement meilleure.

## 6 Remarques et conclusion

Dans ce travail nous avons proposé une approche bayésienne pour la séparation de sources en modélisant les distributions des coefficients d'ondelettes des sources (signaux/images) par des distributions *a priori* du type exponentiel généralisé. Nous



FIG. 4: a. images sources, b. images mélangées, c. estimés dans le cas i.i.d, d. estimés dans le cas corrélé

avons considéré deux cas de figures, une loi séparable quand les coefficients sont supposés indépendants, et une loi non séparable qui prend en compte une corrélation inter-échelles de ces coefficients. Nous avons proposé un algorithme MCMC hybride du type Hastings-Metropolis/Gibbs approprié et avons montrés les performances de cette méthode.

Bien que les résultats de séparation sont meilleurs que ceux obtenus avec certaines méthodes (ICA), la modélisation par des exponentielles doubles ne semble pas fournir la diversité nécessaire pour la séparation de sources. Nous nous tournerons, dans nos futurs travaux, vers la modélisation de la distribution piquée des coefficients d'ondelettes par un mélange de gaussiennes. Les modèles de mélanges offrent une certaine facilité du point de vue algorithmique, et des algorithmes tels que le MCMC ou l'EM peuvent être implémentés d'une manière très rapide (rapidité de la convergence). La corrélation inter-échelles peut être prise en compte, avec ce genre de modèles,

TAB. 1: résultats numériques de simulation

	$P_1(\tilde{S})$		$P_2(\tilde{S})$		$P_2(\tilde{A})$
	Src. 1	Src. 2	Src. 1	Src. 2	
Coef. Indp.	0.129	0.146	0.148	0.167	0.037
Coef. Cor.	0.129	0.124	0.151	0.144	0.030

d'une manière très ingénieuse et efficace par des modèles markoviens à variable cachées (HMM).

## Références

- [1] M. S. Crouse, R. D. Nowak et R. G. Baraniuk. *Wavelet-Based Statistical Signal Processing Using Hidden Markov Models*. IEEE Trans. Sig. Proc., Avril 1998, pages 886-902. Vol 46, NO 4.
- [2] A. Mohammad-Djafari and M. M. Ichir. *Wavelet Domain Image Separation*. Bayesian Inference and Maximum Entropy Methods. MaxEnt Workshops, Aout 2002.
- [3] A. Hyvärinen, J. Karhunen et E. Oja. *Independent Component Analysis*. John Wiley, 2001, New York.
- [4] K. Knuth. *A Bayesian approach to source separation*. Proceedings of Independent Component Analysis Workshop, 1999, pages 283-288.
- [5] S. G. Mallat. *A Theory of Multiresolution Signal Decomposition: The Wavelet Representation*. IEEE Trans. on Patt. Anal. and Mach. Intell., Juillet 1989, pages ,Vol 11.
- [6] -. *a Wavelet Tour of Signal Processing*. Academic Press, 1999.
- [7] H. Snoussi, G. Patanchon, J.F. Macías-Pérez, A. Mohammad-Djafari et J. Delabrouille. *Bayesian blind component separation for cosmic microwave background observations*. Bayesian Inference and Maximum Entropy Methods, MaxEnt Workshops, Aout 2002, pages 125-140.
- [8] H. Snoussi et A. Mohammad-Djafari. *Separation of mixed HMM sources*. Bayesian Inference and Maximum Entropy Methods. MaxEnt Workshops, Aout 2001, pages 77-88.
- [9] -. *Bayesian unsupervised learning for source separation with mixture of Gaussians prior*. To appear in Int. Journal of VLSI Signal Processing Systems, 2003.
- [10] M. Zibulevsky et B.A. Pearlmutter. *Blind Source Separation by Sparse Decomposition in a Signal Dictionary*. MIT Letters on Neural Computation, 2001, pages 863-882, Vol 13.