

# Décomposition de spectres en motifs paramétriques par approximation parcimonieuse

Vincent MAZET<sup>1</sup>, Charles SOUSSEN<sup>2</sup>, El-Hadi DJERMOUNE<sup>2</sup>

<sup>1</sup>ICube, Université de Strasbourg, CNRS (UMR 7357) ; 300 boulevard Sébastien Brant, BP 10413, 67412 Illkirch

<sup>2</sup>CRAN, Université de Lorraine, CNRS (UMR 7039) ; Campus sciences, BP 70239, 54506 Vandœuvre  
vincent.mazet@unistra.fr, {charles.soussen, el-hadi.djermoune}@univ-lorraine.fr

**Résumé** – On considère le problème de décomposition d’une séquence de spectres telle qu’on peut en rencontrer en spectroscopie de photo-électrons [16] : les données sont constituées de plusieurs spectres, acquis à des temps différents. Les raies des spectres évoluent lentement et sont modélisées par des gaussiennes d’amplitudes positives. L’objectif de ce travail est d’estimer les paramètres des motifs et leurs trajectoires d’évolution afin d’aider à l’interprétation des données. Nous proposons deux méthodes originales utilisant des approches parcimonieuses. Dans la première, les spectres sont décomposés de façon séquentielle avec une contrainte de positivité sur les amplitudes en utilisant une implémentation non-négative de l’algorithme  $IR\ell_1$  [19]. Dans la seconde approche, la décomposition est conjointe avec une contrainte de douceur sur l’évolution des positions des motifs. Dans les deux cas, l’algorithme hongrois [14] est appliqué pour relier les atomes sélectionnés afin de retrouver les trajectoires des raies. Les méthodes développées sont comparées à une approche bayésienne utilisant l’algorithme RJMCMC [16] : les résultats sont comparables en qualité mais le temps de calcul est sensiblement réduit.

**Abstract** – We consider the decomposition of a series of spectra such as those acquired in photoelectron spectroscopy [16]: the data consist of several spectra acquired at different times. The peaks in the spectra are assumed to be Gaussian with positive amplitudes and their characteristics vary slowly between two consecutive spectra. This work aims at estimating the peak parameters and their evolution to help the data interpretation. We propose two original methods based on sparse approximation. In the first method, the spectra are decomposed sequentially with a positivity constraint on the amplitudes using an implementation of the non-negative  $IR\ell_1$  algorithm [19]. In the second approach, we opt for a joint approach with a smooth constraint of the peak positions. In both approaches, the Hungarian algorithm [14] is used to obtain the trajectories of the selected peaks. The proposed methods are compared to a Bayesian approach with RJMCMC [16]: the results are comparable but the calculation time is dramatically reduced.

## 1 Introduction

On s’intéresse à la décomposition d’un signal spectroscopique en une somme de raies, comme par exemple dans [16]. Les raies fournissent des informations sur la composition de l’échantillon analysé et sont modélisées par des motifs paramétriques. La décomposition paramétrique d’un spectre unique a déjà fait l’objet de plusieurs études [10, 12, 13, 17]. Nous considérons ici le problème de la décomposition d’une séquence temporelle de spectres dans le but de suivre l’évolution des motifs : un exemple simulé d’une telle séquence est représenté figure 1.

Nous avons montré dans [15] qu’il est préférable de décomposer les spectres de manière conjointe plutôt que séquentielle, c’est-à-dire de traiter les données dans leur ensemble plutôt que chaque spectre indépendamment de ses voisins. La raison est que seule une approche conjointe permet de garantir une évolution douce des motifs à travers la séquence en guidant les décompositions des spectres grâce aux décompositions des spectres voisins. Une approche conjointe peut également fournir un suivi des raies dans la séquence.

Dans [16], nous utilisons un algorithme de Monte Carlo

par chaînes de Markov (MCMC) : cette approche est gourmande en temps de calcul (cf. section 5) et devient inexploitable lorsque les données sont trop volumineuses. Or, plusieurs auteurs [5, 17] ont montré que les méthodes d’approximation parcimonieuse par des algorithmes déterministes sont une alternative efficace pour décomposer un spectre de raies. Une démarche conjointe conduit à minimiser un critère non convexe et en partie discret (du fait de la mise en correspondance des raies d’un spectre avec celles des autres spectres). Aussi, dans un premier temps, nous présentons une méthode séquentielle (chaque spectre est décomposé de façon indépendante) tout en tenant compte de la positivité du vecteur des amplitudes : l’optimisation est continue (section 2). Dans un second temps, nous présentons une démarche conjointe où, cette fois, l’optimisation est discrète et la positivité des amplitudes n’est pas modélisée (section 3). Dans les deux cas, les raies estimées sont classifiées afin de reconstituer leurs trajectoires (section 4). La section 5 présente les résultats obtenus avec ces deux approches sur une séquence simulée, puis la perspective d’une approche à la fois conjointe et prenant en compte la positivité des raies est discutée.

On note  $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_S] \in \mathbb{R}^{N \times S}$  la séquence des  $S$  spectres de  $N$  échantillons chacun. Chaque spectre  $\mathbf{y}_s$  est mo-

délicé comme une somme bruitée d'un faible nombre de raies. On peut donc écrire  $\mathbf{y}_s = \mathbf{A}\mathbf{x}_s + \mathbf{b}_s$  où  $\mathbf{x}_s$  est un vecteur parcimonieux,  $\mathbf{b}_s$  regroupe les erreurs de modélisation et le bruit d'observation et  $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_M] \in \mathbb{R}^{N \times M}$  est le dictionnaire dont chaque atome  $\mathbf{a}_m$  représente une raie potentiellement présente dans les spectres. Nous considérons ici des raies gaussiennes :

$$\forall m, \quad \mathbf{a}_m \propto \left\{ \exp \left( -\frac{(n - c_m)^2}{2w_m^2} \right) \right\}_{n \in \{1, \dots, N\}} \quad (1)$$

dont les centres et largeurs sont définis par  $c_m$  et  $w_m$  (par ailleurs, les atomes sont normalisés :  $\|\mathbf{a}_m\|_2 = 1$ ). Afin d'obtenir des décompositions satisfaisantes, il faut considérer beaucoup de valeurs pour  $c_m$  et  $w_m$  : le dictionnaire  $\mathbf{A}$  est donc surdimensionné et contient un grand nombre d'atomes ( $M \gg N$ ). Ces atomes sont par ailleurs très corrélés.

## 2 Décomposition séquentielle non-négative

### 2.1 Optimisation de critère pénalisé

Le problème de l'approximation parcimonieuse est formulé comme l'optimisation, sous contrainte de positivité, d'un critère pénalisé comportant un terme quadratique qui représente l'erreur d'approximation et un terme de régularisation favorisant la parcimonie de  $\mathbf{x}_s$  :

$$\min_{\mathbf{x}_s \geq 0} \left\{ F(\mathbf{x}_s) = \|\mathbf{y}_s - \mathbf{A}\mathbf{x}_s\|_2^2 + \lambda \sum_{m=1}^M \varphi(\mathbf{x}_s(m)) \right\} \quad (2)$$

Pour obtenir une solution parcimonieuse, la fonction  $\varphi$  doit être non différentiable en 0. Le choix le plus naturel est celui de la pseudo-norme  $\ell_0$  et nécessite des algorithmes d'optimisation combinatoire (algorithmes gloutons par exemple). À l'inverse, la norme  $\ell_1$   $\varphi(t) = |t|$  rend le critère  $F(\mathbf{x}_s)$  convexe (il est nommé  $\ell_2$ - $\ell_1$  dans la suite). Son optimisation conduit à un problème de programmation quadratique et permet de prendre en compte plus facilement la positivité des amplitudes. Par ailleurs, les méthodes  $\ell_2$ - $\ell_1$  sont conçues pour gérer des dictionnaires surdimensionnés. Mais comme les colonnes du dictionnaire sont très corrélées, la régularisation  $\ell_1$  a tendance à produire plusieurs raies proches ou des pics de faibles amplitudes [11], alors qu'une fonction  $\varphi(t)$  non convexe conduit à des approximations plus parcimonieuses.

Aussi, nous choisissons la fonction non convexe  $\varphi(t) = \log(|t| + \varepsilon)$  où  $\varepsilon > 0$  est faible. Ce choix a été guidé par l'utilisation de l'algorithme *Iterative Reweighted*  $\ell_1$  ( $\text{IR}\ell_1$ ) [6, 19, 21], conçu pour minimiser le critère  $F(\mathbf{x}_s)$ . Comme  $\varphi$  est une fonction paire et concave sur  $\mathbb{R}^+$ , le critère  $F(\mathbf{x}_s)$  est généralement non convexe et multimodal. Bien que sous-optimal,  $\text{IR}\ell_1$  conduit à des solutions de bonne qualité, même lorsque les atomes du dictionnaire sont corrélés les uns aux autres [18]. De plus, la structure de l'algorithme  $\text{IR}\ell_1$  permet naturellement de gérer la contrainte de positivité.

## 2.2 Algorithme $\text{IR}\ell_1$ positif

L'algorithme  $\text{IR}\ell_1$  résout une suite de problèmes d'optimisation  $\ell_2$ - $\ell_1$ . Les critères  $\ell_2$ - $\ell_1$  s'écrivent  $\|\mathbf{y}_s - \mathbf{A}\mathbf{x}_s\|_2^2 + \lambda \|\mathbf{W}\mathbf{x}_s\|_1$  où  $\mathbf{W}$  est une matrice diagonale dont les éléments dépendent de la solution  $\mathbf{x}_s$  trouvée à l'itération précédente [6].  $\text{IR}\ell_1$  s'interprète comme un algorithme de majorisation-minimisation pour optimiser le critère  $F(\mathbf{x}_s)$  : celui-ci est majoré par un critère auxiliaire obtenu en majorant la fonction concave  $\varphi(|t|)$  autour de la solution courante  $\mathbf{x}_s$  par sa tangente [6]. De façon à limiter le nombre de paramètres à régler empiriquement, nous utilisons l'algorithme Adaptive Lasso [21] qui correspond au cas particulier où  $\varepsilon \mapsto 0$  : il n'y a donc pas de réglage à effectuer pour  $\varepsilon$ .

Comme l'a souligné Wipf [19, section IV],  $\text{IR}\ell_1$  peut s'adapter naturellement au problème d'optimisation sous contrainte de positivité [19]. Plusieurs algorithmes de minimisation  $\ell_2$ - $\ell_1$  ont été utilisés dans le contexte des approches parcimonieuses : citons par exemple les algorithmes ADMM (*alternating direction method of multipliers*) [7, 8], de gradient projeté [9] ou seuillage itératif [4, 20]. Nous avons choisi une implémentation basée sur l'algorithme d'homotopie [1].

## 3 Décomposition conjointe non contrainte

Seule une démarche conjointe permet de prendre en compte l'évolution douce des paramètres des raies. Cette information peut être codée sous la forme d'une régularisation qui favorisera la similarité entre deux vecteurs d'amplitude voisins  $\mathbf{x}_s$  et  $\mathbf{x}_{s+1}$  pour tout  $s \in \{1, \dots, S-1\}$ . La similarité entre ces deux vecteurs peut être mesurée à l'aide de la distance de Hausdorff [3]. On note  $\Gamma_s$  (resp.  $\Gamma_{s+1}$ ) le support du vecteur  $\mathbf{x}_s$  (resp.  $\mathbf{x}_{s+1}$ ). Les supports  $\Gamma_s$  et  $\Gamma_{s+1}$  représentent des nuages de points sur une grille bidimensionnelle (positions et largeurs des raies), où chaque point correspond à un motif gaussien. En notant  $\gamma_s$  l'un des points du support  $\Gamma_s$  et  $\gamma_{s+1}$  l'un des points du support  $\Gamma_{s+1}$ , le critère à minimiser devient :

$$\min_{\mathbf{x}_1, \dots, \mathbf{x}_S} \left\{ G(\mathbf{x}) = \sum_{s=1}^S \|\mathbf{y}_s - \mathbf{A}\mathbf{x}_s\|_2^2 + \lambda \sum_{s=1}^S \|\mathbf{x}_s\|_0 + \mu \sum_{s=1}^{S-1} \max(d(\Gamma_s, \Gamma_{s+1}), d(\Gamma_{s+1}, \Gamma_s)) \right\} \quad (3)$$

où  $d(\Gamma_s, \Gamma_{s+1})$  est la plus grande des plus petites distances entre les atomes de  $\Gamma_s$  vers ceux de  $\Gamma_{s+1}$  :

$$d(\Gamma_s, \Gamma_{s+1}) = \max_{\gamma_s \in \Gamma_s} \left( \min_{\gamma_{s+1} \in \Gamma_{s+1}} \|\gamma_s - \gamma_{s+1}\|_2 \right). \quad (4)$$

Dans le cas particulier où l'un des deux supports est vide, alors la distance  $d(\Gamma_s, \Gamma_{s+1})$  est fixée à une valeur non nulle.

La minimisation de ce critère est un problème combinatoire de par l'aspect discret des deux derniers termes. Pour cette raison, nous utilisons l'algorithme de minimisation  $\ell_2$ - $\ell_0$  SBR

(single best replacement) [18]. Cet algorithme a fait preuve de bons résultats dans des applications en déconvolution impulsionnelle où, comme dans notre problème, le dictionnaire est grand et redondant. SBR est un algorithme itératif où, à chaque itération, on teste l'ajout ou le retrait d'un atome parmi tous les atomes du dictionnaire : l'atome permettant la plus grande descente du critère est ajouté ou retiré. Il est ainsi trivial d'adapter l'algorithme à l'insertion du terme de régularisation interspectres.

## 4 Estimation des trajectoires des raies

Une fois tous les spectres  $\mathbf{y}_s$  décomposés, un post-traitement affecte les raies obtenues à une trajectoire. Chaque atome du premier spectre ( $s = 1$ ) est affecté à une trajectoire différente, de façon arbitraire. On propage ensuite ces affectations aux spectres voisins de la façon suivante. À l'étape  $s$ , on considère les vecteurs  $\mathbf{x}_s$  et  $\mathbf{x}_{s+1}$  et on construit la matrice  $\mathbf{C} \in \mathbb{R}^{M \times M}$  dont chaque élément  $C_{i,j}$  représente le coût d'associer l'atome  $j$  du vecteur  $\mathbf{x}_{s+1}$  à l'atome  $i$  du vecteur  $\mathbf{x}_s$  pour qu'ils soient dans la même trajectoire. Ce coût dépend de la distance euclidienne  $d_{i,j}$  entre les paramètres des atomes :

$$C_{i,j} = \begin{cases} d_{i,j} & \text{si } d_{i,j} < D, \mathbf{x}_s(i) \neq 0, \mathbf{x}_{s+1}(j) \neq 0, \\ +\infty & \text{sinon} \end{cases} \quad (5)$$

avec

$$d_{i,j} = \frac{(c_i - c_j)^2}{\sigma_c^2} + \frac{(\mathbf{x}_s(i) - \mathbf{x}_{s+1}(j))^2}{\sigma_x^2} + \frac{(w_i - w_j)^2}{\sigma_w^2} \quad (6)$$

où  $c_i$  et  $c_j$  sont respectivement les centres des atomes  $i$  et  $j$  (de même pour les largeurs  $w_i$  et  $w_j$ ), et  $\sigma_c$ ,  $\sigma_x$  et  $\sigma_w$  sont des hyperparamètres à régler (ils sont proportionnels à l'écart qu'on autorise sur les centres, amplitudes et largeurs pour associer deux raies dans la même trajectoire). Si l'un au moins des deux atomes a une amplitude nulle ou si les paramètres des deux atomes sont trop éloignés (d'une valeur  $D$ ), le coût correspondant est infini et l'affectation est donc impossible. Nous utilisons l'algorithme hongrois [14] qui est rapide et efficace pour trouver les affectations qui minimisent le coût global, défini comme la somme des coûts des affectations  $\sum_{i,j} C_{i,j} \delta_{i,j}$  où  $\delta_{i,j} = 1$  si l'atome  $j$  du vecteur  $\mathbf{x}_{s+1}$  est associé à l'atome  $i$  du vecteur  $\mathbf{x}_s$ , 0 sinon.

## 5 Résultats et discussion

Pour quantifier les performances des méthodes proposées, nous avons synthétisé une séquence de  $S = 10$  spectres de  $N = 300$  échantillons chacun avec un rapport signal-à-bruit d'environ 3,6 dB (figure 1). Le dictionnaire contient  $M = 300$  atomes de même largeur et de centres différents. À titre de comparaison, nous avons également testé l'approche bayésienne avec l'algorithme RJMCMC proposé dans [16] où la décomposition est effectuée de manière conjointe. Les méthodes

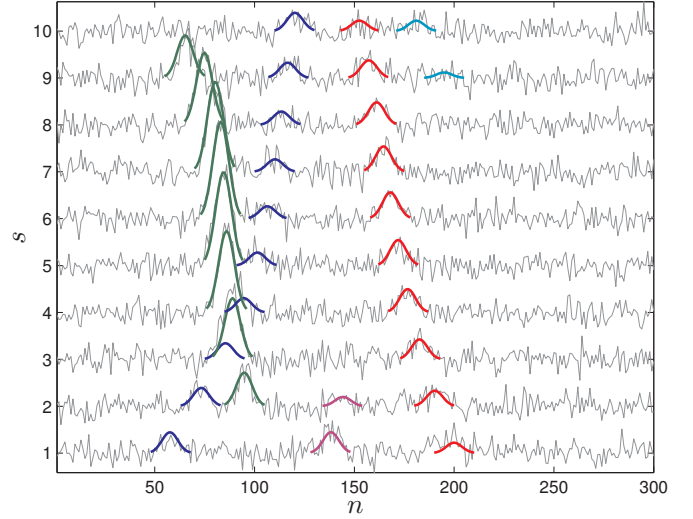


FIGURE 1 – Exemple de séquence simulée de  $S = 10$  spectres de  $N = 300$  échantillons et 5 trajectoires (représentées en couleur).

présentées ici sont significativement plus rapides que l'algorithme RJMCMC (rapport d'environ 10). Par ailleurs, lorsque le bruit est assez important, comme c'est le cas dans la simulation présentée ici, les méthodes d'approximation parcimonieuse rivalisent avec l'estimation obtenue par RJMCMC : les erreurs quadratiques moyennes des différentes approches sont très proches (de l'ordre de 0,5) ; voir aussi les paramètres des raies estimées figure 2. Toutefois, certaines difficultés, comme le croisement entre deux trajectoires, ne sont pas résolues. Enfin, l'algorithme SBR, même s'il ne prend pas en compte la positivité des raies (une raie est estimée avec une amplitude négative) fournit de très bons résultats par rapport aux autres méthodes : c'est en effet la seule méthode qui identifie l'une des trajectoires (en bleu figure 2). En conclusion, ces premiers résultats confirment que les méthodes d'approximation parcimonieuse sont des alternatives intéressantes face aux algorithmes MCMC pour décomposer des séquences de spectres.

Pour la suite de ce travail, nous nous concentrons sur le développement d'une approche conjointe qui tienne compte de la positivité des raies. L'ajout à la fois d'une contrainte de douceur sur les paramètres des raies et de positivité de la décomposition n'est pas simple. Nous avons étudié une première piste dans laquelle la douceur n'est pas exprimée sur les raies elles-mêmes mais sur le spectre reconstruit : ainsi, le critère permet l'utilisation de l'implémentation non-négative de  $\text{IRL}_1$ . Les résultats sur les données simulées sont présentés figure 2 : ils montrent que ce type de régularisation n'introduit pas suffisamment de douceur sur l'estimation et n'apporte pas plus d'amélioration par rapport aux méthodes présentées.

Une autre piste d'étude concerne l'utilisation de normes mixtes (voir par exemple [2]). En effet, la douceur sur l'évolution des paramètres peut être modélisée en imposant une structure sur le vecteur des amplitudes d'un spectre par rapport à celui du spectre voisin.

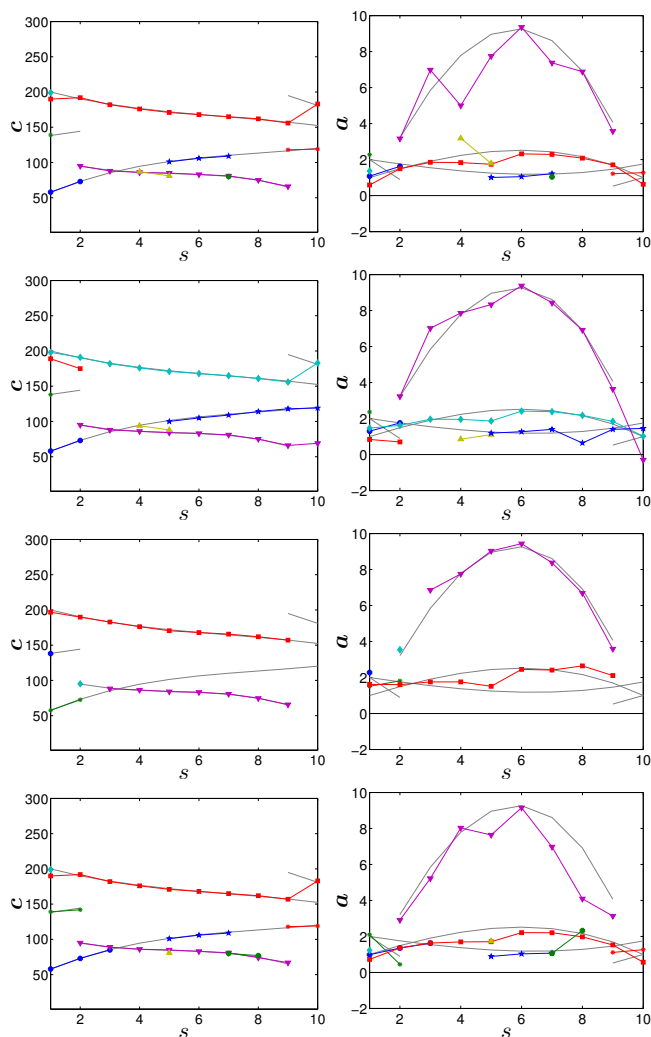


FIGURE 2 – Positions (à gauche) et amplitudes (à droite) des raies estimées pour les données de la figure 1. De haut en bas : première approche (séquentielle avec  $IR\ell_1$ ), conjointe (avec SBR), RJMCMC [16], conjointe (avec  $IR\ell_1$ ). Les lignes grises et colorées correspondent respectivement aux paramètres réels et estimés.

## Remerciements

Les auteurs tiennent à remercier Jérôme Idier (IRCCyN) pour les discussions fructueuses, ainsi que les deux experts anonymes pour leurs remarques qui ont permis d'améliorer les études présentées dans cet article.

## Références

[1] M.S. ASIF et J. ROMBERG : Fast and accurate algorithms for re-weighted  $\ell_1$ -norm minimization. Rapport technique, 2012.  
 [2] F. BACH, R. JENATTON, J. MAIRAL et G. OBOZINSKI : Structured sparsity through convex optimization. *Statistical Science*, 27(4):450–468, 2012.  
 [3] M.F. BARNESLEY : *Fractals Everywhere*. Academic Press, 2<sup>e</sup> édition, 1993.

[4] A. BECK et M. TEBoulLE : A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1), 2009.  
 [5] S. BOURGUIGNON, H. CARFANTAN et J. IDIER : A sparsity-based method for the estimation of spectral lines from irregularly sampled data. *IEEE J. Sel. Top. Signal Process.*, 2007.  
 [6] E.J. CANDÈS, M.B. WAKIN et S.P. BOYD : Enhancing sparsity by reweighted  $\ell_1$  minimization. *J. Fourier Anal. Appl.*, 14:877–905, 2008.  
 [7] M.J. FADILI et J.L. STARCK : Monotone operator splitting for optimization problems in sparse recovery. *In ICIP*, 2009.  
 [8] M.A.T. FIGUEIREDO et J.M. BIUCAS-DIAS : Restoration of Poissonian images using alternating direction optimization. *IEEE Transactions on Image Processing*, 19(12), 2010.  
 [9] M.A.T. FIGUEIREDO, R.D. NOWAK et S.J. WRIGHT : Gradient projection for sparse reconstruction : Application to compressed sensing and other inverse problems. *IEEE Journal of Selected Topics in Signal Processing*, 1(4), 2007.  
 [10] R. FISCHER et V. DOSE : *Monographs of Official Statistics : Bayesian Methods With Applications to Science, Policy, and Official Statistics*, chapitre Analysis of Mixtures in Physical Spectra, pages 145–154. 2001.  
 [11] J.J. FUCHS : Convergence of a sparse representations algorithm applicable to real or complex data. *IEEE Journal of Selected Topics in Signal Processing*, 1(4), 2007.  
 [12] S. GULAM RAZUL, W.J. FITZGERALD et C. ANDRIEU : Bayesian model selection and parameter estimation of nuclear emission spectra using RJMCMC. *Nucl. Instrum. Meth. A*, 497(2-3):492–510, 2003.  
 [13] Y. GUO : *Représentations parcimonieuses pour la séparation des signaux : quantification en IRM spectroscopique pour l'étude du métabolisme*. Thèse de doctorat, Université de Reims Champagne-Ardenne, 2011.  
 [14] H.W. KUHN : The Hungarian method for the assignment problem. *Nav. Res. Log.*, 2:83–97, 1955.  
 [15] V. MAZET : Joint Bayesian decomposition of a spectroscopic signal sequence. *IEEE Signal Proc. Lett.*, 18(3):191–184, 2011.  
 [16] V. MAZET, S. FAISAN, A. MASSON, M.-A. GAVEAU, L. POISSON et J.-M. MESTDAGH : Joint Bayesian decomposition of a spectroscopic signal sequence with RJMCMC. *In IEEE SSP Workshop*, Ann Arbor, Michigan, 2012.  
 [17] S. SAHNOUN, E.-H. DJERMOUNE, C. SOUSSEN et D. BRIE : Sparse multidimensional modal analysis using a multigrid dictionary refinement. *Eurasip Journal on Advances in Signal Processing*, 2012(60):1–10, 2012.  
 [18] C. SOUSSEN, J. IDIER, D. BRIE et J. DUAN : From Bernoulli-Gaussian deconvolution to sparse signal restoration. *IEEE Trans. Signal Proces.*, 59(10):4572–4584, 2011.  
 [19] D. WIPF et S. NAGARAJAN : Iterative reweighted  $\ell_1$  and  $\ell_2$  methods for finding sparse solutions. *IEEE J. Sel. Top. Signal Process.*, 2010.  
 [20] M. ZIBULEVSKY et M. ELAD :  $\ell_1$ - $\ell_2$  optimization in signal and image processing. *IEEE Signal Processing Magazine*, 27(3):76–88, mai 2010.  
 [21] H. ZOU : The adaptive Lasso and its oracle properties. *J. Am. Stat. Assoc.*, 101(476):1418–1429, 2006.