Minimisation de critères de moindres carrés pénalisés par la norme ℓ^1 dans le cas complexe

Sébastien Bourguignon[1], Hervé Carfantan[1], Jérôme Idier[2]

[1]Laboratoire d'Astrophysique de Toulouse et Tarbes - 14 avenue Edouard Belin, 31400 Toulouse, France

[2]IRCCyN - 1 rue de la Noë, BP 92 101 - 44321 Nantes Cedex 03, France

bourgui@ast.obs-mip.fr, Herve.Carfantan@ast.obs-mip.fr, Jerome.Idier@irccyn.ec-nantes.fr

Résumé – Une approche classique des représentations parcimonieuses consiste à minimiser un critère quadratique pénalisé par la norme ℓ^1 du vecteur inconnu. Si ce vecteur est réel, l'optimisation peut être abordée efficacement par programmation quadratique. Le problème est plus délicat quand le vecteur est complexe. C'est le cas étudié ici, car il intéresse des applications importantes comme l'analyse spectrale. Nous examinons le comportement, pour des problèmes de grande taille, de deux algorithmes récemment proposés dans ce cadre, de type *Iterative Coordinate Descent* et *Iterative Reweighted Least-Squares*. Nous proposons ensuite une procédure d'optimisation mixte tirant parti des avantages complémentaires de ceux-ci. Les performances obtenues surpassent nettement celles d'une approche de type *Second-Order Cone Programming*, également proposée pour ce type de problème.

Abstract – The minimisation of a quadratic criterion penalised by the ℓ^1 norm of the unknown vector is a classical approach to sparse representations. For real-valued variables, optimisation can be tackled efficiently by quadratic programming. The case of a complex-valued vector is more tricky. It is studied here, with a view to important applications such as spectral analysis. We examine two optimisation strategies, namely *Iterative Coordinate Descent* (ICD) and *Iterative Reweighted Least-Squares* (IRLS). Then, we propose an hybrid algorithm that benefits from the specificities of ICD and IRLS. The resulting procedure yields much better performances than *Second-Order Cone Programming* based algorithms, that were recently proposed for such a problem.

1 Introduction

Le domaine des représentations parcimonieuses est devenu une thématique particulièrement active au cours des dix dernières années. Le but est de représenter un signal par une combinaison linéaire du plus petit nombre d'atomes, pris dans un dictionnaire habituellement composé de différents signaux aux caractéristiques complémentaires. Ce dictionnaire étant généralement redondant, le modèle linéaire reliant les données aux paramètres inconnus est sous-déterminé. En présence de bruit, l'approche classique de ce problème consiste à minimiser un critère quadratique d'attache aux données, pénalisé par un terme de norme ℓ^1 favorisant les solutions parcimonieuses [1].

Lorsque les fonctions du dictionnaire sont réelles, le problème d'optimisation se ramène à un programme quadratique, pouvant être résolu par des algorithmes particulièrement efficaces [2, 3]. Cependant, l'utilisation de fonctions à valeurs complexes est parfois plus réaliste en termes de modélisation. Ainsi, pour l'estimation de raies spectrales à partir de séries temporelles, le caractère parcimonieux porte sur le module des amplitudes complexes du spectre à reconstruire [4]. Il en va de même pour certains problèmes de déconvolution impulsionnelle, où les sources sont naturellement décrites par des variables complexes [5].

Avec des variables complexes, l'optimisation n'est plus un programme quadratique. Le problème peut cependant être formulé comme un programme conique du second ordre (SOCP, *Second Order Cone Programming*, [6]), permettant d'aborder l'optimisation par des méthodes de point intérieur [7]. Les algorithmes de type *Iterative Coordinate Descent* (ICD, [8]) et *Iterative Reweighted Least-Squares* (IRLS, [9]), récemment proposés en application à ce type de problème, sont également valides dans le cas complexe.

Notre étude porte sur le comportement des algorithmes ICD et IRLS dans le cas de problèmes de grande taille. Nous proposons ensuite une procédure d'optimisation mixte tirant parti des avantages complémentaires des algorithmes ICD et IRLS, permettant de réduire considérablement le coût de calcul. Nous montrons enfin que les performances obtenues sont supérieures à celles d'une approche de type SOCP, notamment lorsque la taille du problème est grande.

2 Représentations parcimonieuses

2.1 Pénalisation ℓ^1

Nous recherchons une représentation parcimonieuse d'un jeu de données $\boldsymbol{y} \in \mathbb{C}^N$ parmi les éléments d'un dictionnaire de M signaux à valeurs complexes. Dans le cas de données bruitées, ce problème est classiquement abordé [1] sous la forme de la minimisation du critère :

$$J(\boldsymbol{x}) = \frac{1}{2} \left\| \boldsymbol{y} - \mathbf{A}\boldsymbol{x} \right\|_2^2 + \lambda \left\| \boldsymbol{x} \right\|_1 , \qquad (1)$$

où la matrice $\mathbf{A}$ de taille $N \times M$ est le dictionnaire discrétisé, $\boldsymbol{x} \in \mathbb{C}^M$ contient les coefficients de la représentation et $\lambda > 0$ contrôle le degré de parcimonie de la solution $\widehat{\boldsymbol{x}} = \arg\min J(\boldsymbol{x})$, c'est-à-dire son nombre de composantes non nulles. Le critère J étant convexe, au sens large seulement, l'existence de minima locaux est exclue, mais

pas celle d'un ensemble (convexe) de minimiseurs globaux non réduit à un singleton. Des travaux récents [10, 11, 12] établissent des conditions, portant sur le dictionnaire $\mathbf{A}$, garantissant l'unicité du minimiseur si celui-ci est suffisamment parcimonieux.

2.2 Intérêt de dictionnaires complexes

La représentation par un dictionnaire de signaux complexes complique substantiellement le problème de minimisation du critère (1). Le cas de signaux réels, pour lequel l'optimisation peut s'écrire comme un programme quadratique, a été largement abordé. Différentes approches algorithmiques très efficaces ont été proposées [2, 3], qui ne peuvent cependant pas s'étendre au cas complexe.

Dans certains cas, l'utilisation d'un dictionnaire à valeurs complexes reste néanmoins plus judicieuse en termes de modélisation *a priori*. Prenons pour exemple le problème classique de l'analyse spectrale d'une série temporelle, où l'on cherche à estimer un spectre de raies. La formulation (1) est particulièrement adaptée, le dictionnaire $\mathbf{A}$ étant alors formé d'exponentielles complexes, de fréquences discrétisées sur une grille arbitrairement fine : $\mathbf{A}$ est une matrice de type « Fourier inverse ». Il peut paraître superflu d'utiliser des fonctions à valeurs complexes, le dictionnaire pouvant être décomposé de manière équivalente en la concaténation de fonctions sinus et cosinus. Le vecteur de coefficients associés s'identifie alors à $[\Re(\boldsymbol{x}); \Im(\boldsymbol{x})]$, de même taille, mais à valeurs réelles. Cependant, la recherche d'une représentation parcimonieuse d'un jeu de données dans un tel dictionnaire s'avère moins efficace. En effet, la pénalisation en norme ℓ^1 s'écrit alors $R_{\mathbb{R}}(\boldsymbol{x}) = \|\Re(\boldsymbol{x})\|_1 + \|\Im(\boldsymbol{x})\|_1$, ce qui n'est pas équivalent à $\|\boldsymbol{x}\|_1$. La Fig. 1 montre que les résultats obtenus dans les deux cas peuvent être sensiblement différents. En particulier, la pénalisation par $R_{\mathbb{R}}$ donne des solutions parcimonieuses pour $\Re(\boldsymbol{x})$ et $\Im(\boldsymbol{x})$, mais dont les composantes non nulles ne coïncident pas. Ce n'est pas le cas pour la pénalisation par $\|\boldsymbol{x}\|_1$, où $\Re(\boldsymbol{x})$ et $\Im(\boldsymbol{x})$ sont non nuls aux mêmes endroits, ce qui est plus satisfaisant en termes de localisation fréquentielle. Notons encore que, dans le cas de la pénalisation par $R_{\mathbb{R}}$, une valeur différente du paramètre de régularisation λ n'améliore pas l'estimation.

3 Algorithmes d'optimisation

Si les coefficients de la représentation sont à valeurs complexes, l'optimisation ne peut plus être formulée comme un programme quadratique. Parmi les algorithmes récemment introduits dans ce domaine, cependant, les méthodes de type ICD [8] et IRLS [9] sont applicables aux variables complexes. Nous étudions en détail le comportement de ces algorithmes, qui présentent des propriétés complémentaires pour l'estimation de solutions parcimonieuses.

3.1 L'algorithme ICD

L'algorithme ICD consiste à effectuer des minimisations successives par rapport à chaque composante. Il peut être vu comme un cas particulier de l'optimisation par blocs

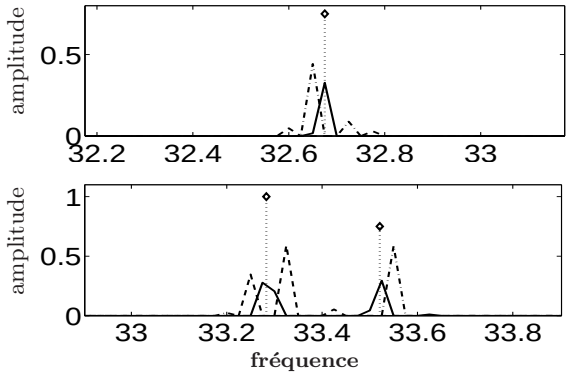


FIG. 1 – Estimation de raies spectrales par minimisation du critère (1). Trait plein : estimateur $|\widehat{\boldsymbol{x}}|$ obtenu avec un dictionnaire d'exponentielles complexes. Ligne discontinue et pointillé : $|\Re(\widehat{\boldsymbol{x}}_{\mathbb{R}})|$ et $|\Im(\widehat{\boldsymbol{x}}_{\mathbb{R}})|$ pour l'estimateur $\widehat{\boldsymbol{x}}_{\mathbb{R}}$ obtenu avec un dictionnaire de sinus et cosinus. Les losanges correspondent aux fréquences théoriques.

proposée dans [8], où les blocs sont pris de dimension unité, ainsi que comme un cas particulier de l'algorithme proposé dans [13], où l'information de parcimonie opère directement sur les inconnues $\boldsymbol{x}$. Pour le critère (1), la solution de chaque minimisation mono-composante est explicite et fait naturellement apparaître des composantes nulles au fil des itérations. En effet, en notant $\boldsymbol{a}_k$ les colonnes de $\mathbf{A}$, supposées de norme unité, on a [11] :

$$\arg\min_{x_k} J(\boldsymbol{x}) = \phi_\lambda(\boldsymbol{a}_k^\dagger \boldsymbol{e}_k), \tag{2}$$

$\boldsymbol{e}_k = \boldsymbol{y} - \sum_{\ell \neq k} \boldsymbol{a}_\ell x_\ell$, et ϕ_λ la fonction de *seuillage doux* :

$$\forall u = \rho e^{j\phi},\ \ \phi_\lambda(u) = (\rho - \lambda)\, e^{j\phi} \text{ si } \rho > \lambda,\ 0 \text{ sinon.}$$

Le coût de chaque itération est donc faible, même pour des problèmes de grande taille. De plus, la convergence de l'ICD vers le minimum du critère (1) est garantie [8, 13].

3.2 L'algorithme IRLS

L'algorithme IRLS se décompose en deux étapes constituées, à chaque itération t, par :

i) mise à jour de la matrice $\mathbf{Q}^{(t)} = \mathrm{diag}\{|\boldsymbol{x}^{(t)}|\}$,

ii) calcul de la nouvelle itérée :

$$\boldsymbol{x}^{(t+1)} = \mathbf{Q}^{(t)}(\lambda \mathbf{I} + \mathbf{Q}^{(t)}\mathbf{A}^\dagger \mathbf{A}\mathbf{Q}^{(t)})^{-1}\mathbf{Q}^{(t)}\mathbf{A}^\dagger \boldsymbol{y}.$$

Une telle procédure est particulièrement intéressante pour des solutions parcimonieuses, puisque le calcul de ii) se ramène à l'inversion d'un système linéaire, dont la taille correspond au nombre de composantes non nulles de $\boldsymbol{x}^{(t)}$, noté par la suite $M^{(t)}$.

La mise à jour simultanée de l'ensemble de ces composantes par IRLS se révèle de fait bien plus efficace pour estimer les amplitudes que les mises à jour monodimensionnelles successives de l'ICD. Cependant, le coût de chaque itération reste élevé tant que le nombre de composantes non nulles n'est pas suffisamment réduit. À ce propos, notons que le calcul de ii) sous la forme équivalente $\boldsymbol{x}^{(t+1)} = \mathbf{Q}^{(t)}\mathbf{A}^\dagger(\lambda \mathbf{I} + \mathbf{A}\mathbf{Q}^{(t)}\mathbf{A}^\dagger)^{-1}\boldsymbol{y}$ permet de se ramener à l'inversion d'un système de la taille N des données, réduisant ainsi le coût des itérations tant que $M^{(t)} > N$.

Par ailleurs, la convergence de l'IRLS n'est pas assurée : il est facile de voir d'après l'étape ii) que l'algorithme reste

bloqué en un point non optimal si, à une itération t, une composante de $\boldsymbol{x}^{(t)}$ est nulle alors que le vrai minimiseur est non nul en cet endroit. Néanmoins, l'IRLS assure la décroissance du critère au sens large à chaque itération [9].

3.3 Un algorithme hybride

Le comportement des algorithmes ICD et IRLS est représenté en Fig. 2, sur un problème d'analyse spectrale à haute résolution, avec $N = 500$ données et $M = 4000$ inconnues. L'algorithme ICD est efficace pour faire décroître le nombre de composantes non nulles de la solution, mais est plutôt lent pour estimer correctement les amplitudes associées. À l'inverse, la décroissance du critère pour l'algorithme IRLS est très lente tant que le nombre de composantes non nulles est élevé, puis devient plus rapide que celle obtenue avec ICD sur la fin, lorsque les itérées ont peu de composantes non nulles.

Nous proposons un algorithme mixte pour la minimisation du critère (1), spécialement efficace pour l'estimation de solutions parcimonieuses. Le schéma général est celui de l'ICD, entrecoupé d'un nombre d'itérations d'IRLS tel que le coût de calcul associé à celles-ci soit du même ordre qu'un balayage de l'ensemble des inconnues par ICD. Puisque l'ICD est un algorithme convergent, et comme l'IRLS assure la décroissance du critère au sens large, un argument de type *spacer steps* [14] permet de conclure que la version hybride converge vers le minimum de (1).

La Fig. 2 montre que l'algorithme hybride se comporte d'abord comme l'ICD, quand celui-ci est le plus efficace, puis comme l'IRLS, quand le nombre de composantes non nulles a diminué. Sur cet exemple, il permet d'obtenir un gain d'un facteur dix sur le nombre total de multiplications par rapport à l'IRLS, celui-ci nécessitant à son tour environ deux fois moins de multiplications que l'ICD.

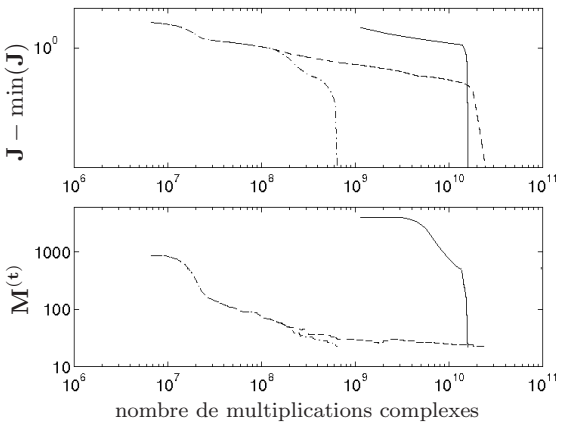


FIG. 2 – Comportement des algorithmes ICD (pointillé), IRLS (trait plein) et mixte (ligne discontinue). Haut : Évolution du critère en fonction du temps CPU. Bas : Évolution correspondante du nombre de composantes non nulles.

3.4 Approche de type SOC

Une approche algorithmique différente, s'appuyant sur la formulation de l'optimisation comme un SOCP (*Second-Order Cone Program*), a été récemment proposée, qui est par exemple expliquée pédagogiquement dans [6]. Celle-ci représente, à notre connaissance, la seule alternative à ce jour aux approches ICD et IRLS pour la minimisation du critère (1) lorsque les inconnues $\boldsymbol{x}$ sont complexes. Nous détaillons ici les pas permettant d'écrire cette optimisation sous la forme d'un SOCP.

La forme générale d'un SOCP s'écrit :

$$\min_{\boldsymbol{x}} \ \boldsymbol{c}^T\boldsymbol{x} \text{ avec } \mathbf{A}\boldsymbol{x} = \boldsymbol{b} \text{ et } \boldsymbol{x} \in \mathcal{K}$$

où $\mathcal{K}$ est un cône quadratique, c'est-à-dire tel que :

$$\{(x_1, \boldsymbol{x}_2) \in \mathbb{R} \times \mathbb{R}^{N-1} \text{ tel que } ||\boldsymbol{x}_2||_2 \leqslant x_1\}.$$

Les minimisations des deux termes de (1) s'écrivent comme des SOCP. En effet, d'après [6] :

$$||\boldsymbol{x}||_1 = \min_{\boldsymbol{t}} \ 1^T\boldsymbol{t} \text{ avec } \forall k, ||(\Re\{x_k\}, \Im\{x_k\})||_2 \leqslant t_k \quad (3)$$

et $||\boldsymbol{y} - \mathbf{A}\boldsymbol{x}||_2^2 = \min_{s,\boldsymbol{z}} \ s$ avec $\boldsymbol{z} = \boldsymbol{y} - \mathbf{A}\boldsymbol{x}$, $\boldsymbol{z}^T\boldsymbol{z} \leqslant s$ et $s \geqslant 0$. Or $s = (s+1)^2/4 - (s-1)^2/4$, d'où

$$\boldsymbol{z}^T\boldsymbol{z} \leqslant s \iff ||\,((s-1)/2, \boldsymbol{z})\,||_2 \leqslant (s+1)/2.$$

Avec $u = (s+1)/2$ et $v = (s-1)/2$, on obtient :

$$||\boldsymbol{y}-\mathbf{A}\boldsymbol{x}||_2^2 = \min_{s,\boldsymbol{z},u,v} \ s \text{ avec } \begin{cases} s \geqslant 0, \ \boldsymbol{z} = \boldsymbol{y} - \mathbf{A}\boldsymbol{x}, \\ u = (s+1)/2, \\ v = (s-1)/2 \\ ||(v, \boldsymbol{z})||_2 \leqslant u \end{cases} \quad (4)$$

La minimisation du critère (1) s'effectue alors en minimisant $s/2+\lambda 1^T\boldsymbol{t}$ sous l'ensemble des contraintes imposées par (3) et (4).

4 Résultats

Nous comparons ici le coût de l'optimisation par l'algorithme proposé en § 3.3 avec celui d'une approche de type SOCP décrite en § 3.4. Les algorithmes sont comparés sur un problème particulièrement difficile d'analyse spectrale [12] : $\mathbf{A} = \{\exp(j2\pi t_n f_k)\}_{n,k}$, où f_k sont les fréquences (positives et négatives) auxquelles le spectre $\boldsymbol{x}$ est discrétisé. Les $N = 514$ instants d'échantillonnage t_n sont irréguliers et correspondent à 5 nuits d'observations d'objets astrophysiques (les observations ne pouvant être effectuées que quelques heures par nuit).

Notons que pour une telle application, si les données sont à valeurs réelles, le spectre reconstruit présente une symétrie hermitienne. En prenant en compte des amplitudes de cosinus et de sinus plutôt que d'exponentielles complexe et en régularisant conjointement ces amplitudes, on peut alors de façon équivalente minimiser le critère :

$$J_R(\boldsymbol{x}) = \frac{1}{2}\,\|\boldsymbol{y} - \mathbf{B}\boldsymbol{x}\|^2 + \lambda\left(|x_1| + \sum_{k=1}^{P}\sqrt{x_{2k}^2 + x_{2k+1}^2}\right) \quad (5)$$

où $\mathbf{B}_{n,1} = 1$, $\mathbf{B}_{n,2k} = \cos 2\pi f_k t_n$ et $\mathbf{B}_{n,2k+1} = \sin 2\pi f_k t_n$, pour des fréquences positives uniquement, ce qui réduit de moitié la taille des inconnues par rapport à (1). La minimisation de (5) s'écrit encore comme un SOCP. En revanche, une adaptation efficace des algorithmes ICD et IRLS, et donc de l'approche hybride du § 3.3, n'est pas immédiate et des travaux sont actuellement en cours dans cette voie.

Nous étudions donc le coût de l'optimisation :

– du critère (1) par les algorithmes ICD, IRLS et hybride, où $\mathbf{A} = \{\exp(j2\pi t_n f_k)\}_{n,k}$,
– des critères (1) (algorithme $\text{SOCP}_{\mathbb{C}}$) et (5) (algorithme $\text{SOCP}_{\mathbb{R}}$) par l'approche SOCP.

L'ICD est programmé en langage C, l'IRLS en Matlab et l'algorithme hybride fait appel aux deux algorithmes précédents. Les algorithmes de type SOCP utilisent la bibliothèque SeDuMi et le code fourni par D. Malioutov (`http://ssg.mit.edu/group/dmm/dmm.shtml`), également programmé en Matlab et en C. Dans [12], il est montré que le minimiseur $\widehat{\boldsymbol{x}}$ vérifie :

$$\begin{cases} \forall k \text{ tel que } |\widehat{x}_k| = 0 & : \quad |r_k| \leqslant \lambda, \\ \forall k \text{ tel que } |\widehat{x}_k| \neq 0 & : \quad r_k + \lambda \widehat{x}_k / |\widehat{x}_k| = 0, \end{cases}$$

avec $\boldsymbol{r} = \mathbf{A}^{\dagger}(\mathbf{A}\widehat{\boldsymbol{x}} - \boldsymbol{y})$. Nous avons utilisé cette propriété comme test d'arrêt des algorithmes ICD, IRLS et hybride, avec une tolérance de 10^{-6}. À l'arrêt des algorithmes SOCP, ce test était vérifié à environ 10^{-5} près. Le Tableau 1 donne les coûts de calcul obtenus, pour différents nombres d'inconnues (1000, 2000 et 3000 fréquences positives) et deux degrés de parcimonie : les données sont constituées de 5 ou de 40 sinusoïdes, et la solution a respectivement 20 ou 200 éléments non nuls environ[1]. Les coûts correspondent aux valeurs retournées par la fonction `cputime` de Matlab, exprimées en secondes.

# fréq. $\geqslant 0$	1000		2000		3000	
# sinus	5	40	5	40	5	40
Hybride	4	66	15	162	40	965
IRLS	32	1313	164	8834	377	>6h
ICD	14	65	109	163	563	987
$\text{SOCP}_{\mathbb{C}}$	585	616	1185	1282	1805	2067
$\text{SOCP}_{\mathbb{R}}$	78	82	151	172	242	263

Tab. 1 – Temps de calcul des différents algorithmes.

Ces résultats permettent de conclure sur trois points.

Le gain apporté par l'algorithme hybride par rapport aux algorithmes ICD et IRLS est d'autant plus efficace que la solution est parcimonieuse. Lorsque celle-ci l'est moins, le coût de l'optimisation par la version hybride est sensiblement voisin de celui de l'ICD.

Sur cet exemple, l'algorithme IRLS seul est souvent le plus coûteux, et est d'autant plus pénalisé que le nombre de composantes non nulles est élevé.

L'algorithme $\text{SOCP}_{\mathbb{C}}$ est beaucoup plus coûteux que les algorithmes précédents pour minimiser le critère (1). Le passage à la version rélle, qui ne peut être généralisé pour tout problème à variable complexe, améliore cependant nettement l'approche par SOCP. Notons que ces algorithmes sont quasiment insensibles à la parcimonie de la solution. Ainsi, pour les solutions moins parcimonieuses, $\text{SOCP}_{\mathbb{R}}$ est de coût moindre ou comparable à celui de l'algorithme hybride que nous proposons.

Des simulations à plus grand nombre de données (N doublé) ont également été effectuées. Dans le cas de solutions très parcimonieuses, le temps de calcul des algorithmes ICD, IRLS et hybride n'est guère affecté par le nombre de données. Les algorithmes SOCP, par contre, se montrent très sensibles au nombre de données (lorsque N est doublé, le coût est multiplié par plus de 3 dans le cas réel et plus de 4 dans le cas complexe), ce qui se comprend aisément puisque cela ajoute des inconnues supplémentaires en proportion.

L'algorithme hybride proposé, étant spécifiquement conçu pour exploiter la parcimonie, améliore donc sensiblement les performances d'algorithmes pré-existants, qu'ils soient de type ICD, IRLS ou encore SOCP. Ces performances sont d'ailleurs d'autant plus efficaces que la solution est parcimonieuse, et ce quelle que soit la taille du problème considéré (nombre d'inconnues et de données).

[1] Chaque sinusoïde correspond à deux raies (fréquences positive et négative), elles-mêmes souvent doublées pour cause de discrétisation inexacte (cf Fig. 1). Les composantes non nulles supplémentaires sont dues à des artéfacts de trés faible amplitude.

Références

[1] J.-J. Fuchs, Une approche à l'estimation et à l'identification simultanées, in *Actes du 16e colloque GRETSI*, Grenoble, sep. 1997, pp. 1273–1276.

[2] S. Chen, D. Donoho et M. Saunders, Atomic decomposition by basis pursuit, *SIAM J. Sci. Comput.*, vol. 20, No 1, pp. 33–61, 1999.

[3] B. Efron, T. Hastie, I. Johnstone et R. Tibshirani, Least Angle Regression, *Ann. Stat.*, vol. 32, No 2, pp. 407–499, 2004.

[4] M. Sacchi, T. Ulrych et C. Walker, Interpolation and extrapolation using a high-resolution discrete Fourier transform, *IEEE Trans. Signal Processing*, vol. 46, No 1, pp. 31–38, jan. 1998.

[5] O. Rabaste, *Estimation de canaux multitrajets. Application à la tomographie acoustique océanique active discrète*, Thèse de doctorat, ENST Bretagne, Brest, nov. 2006.

[6] D. M. Malioutov, *A Sparse Signal Reconstruction Perspective for Source Localization with Sensor Arrays*, Masters thesis, MIT, juil. 2003.

[7] J. Sturm, Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones, *Optim. Meth. Soft.*, vol. 11–12, pp. 625–653, 1999.

[8] S. Sardy, A. Bruce et P. Tseng, Block coordinate relaxation methods for nonparametric wavelet denoising, *Journal of Computational and Graphical Statistics*, vol. 9, pp. 361–379, 2000.

[9] M. Figueiredo, Adaptive sparseness for supervised learning, *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1150–1159, sep. 2003.

[10] J.-J. Fuchs, Recovery of exact sparse representations in the presence of bounded noise, *IEEE Trans. Inf. Theory*, vol. 51, No 10, pp. 3601–3608, oct. 2005.

[11] J. Tropp, Just relax : convex programming methods for identifying sparse signals in noise, *IEEE Trans. Inf. Theory*, vol. 52, No 3, pp. 1030–1051, mars 2006.

[12] S. Bourguignon, H. Carfantan et J. Idier, A sparsity-based method for the estimation of spectral lines from irregularly sampled data, Rapport technique, OMP/IRCCyN, fév. 2007.

[13] I. Daubechies, M. Defrise et C. D. Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, *Comm. Pure Appl. Math.*, vol. 57, pp. 1413–1457, 2004.

[14] D. G. Luenberger, *Linear and Nonlinear Programming*, Addison-Wesley, 1989.