

Acquisition Comprimée intégrant une Connaissance Partielle du Modèle d'Observation: Analyse Asymptotique

Guillaume BOULEUX¹, Rémy BOYER^{*2}

¹Université de Lyon, Université de Saint Etienne, DISP EA-4570, INSA de Lyon
20 Av. de Paris, 42334 Roanne Cedex, France, guillaume.bouleux@insa-lyon.fr

²Laboratoire des Signaux et Systèmes (L2S), Université Paris Sud-CNRS-CentraleSupélec
Gif-Sur-Yvette, France, remy.boyer@l2s.centralesupelec.fr

Résumé – De nombreux contextes applicatifs d'importance présupposent une connaissance partielle des paramètres du modèle d'observation issus de mesures réalisées au préalable. Le modèle d'observation est alors composé d'un signal d'intérêt, d'un signal interférence partiellement connu et d'un bruit additif. Ce travail propose dans le cadre de l'acquisition comprimée, une étude asymptotique des performances d'estimation du vecteur parcimonieux des amplitudes pour les estimateurs parcimonieux exploitant par déflation la connaissance *a priori* du sous-espace "interférence". Sous ces hypothèses, nous proposons l'expression de la Borne de Cramér-Rao Moyennée (BCRM) dans un contexte asymptotique c'est-à-dire lorsque le nombre de mesures et le valeurs non-nulles d'amplitude croient avec un ratio asymptotiquement constant. L'expression qui en résulte est particulièrement compacte en terme d'implémentation et d'interprétation et est valable pour tout type de prior sur le vecteur d'amplitudes. Enfin, quelques simulations numériques illustrent l'intérêt de la borne Bayésienne proposée au regard des estimateurs parcimonieux de l'état de l'art.

Abstract – Numbers of applicative contexts presuppose a partial knowledge of the observation model induced by post-processing acquisitions. The observation model is then composed of a signal of interest, a partially known interfering signal and an additive noise. Based on the compressed sensing theory, the presented work proposes an asymptotic performances analysis of the sparse amplitudes vector for sparse popular estimators straightforwardly extended in this article to exploitation of the a priori knowledge of the interfering subspace. With these assumptions, we propose closed-form expressions of the Expected Cramer-Rao Bound (ECRB) in an asymptotic regime, *i.e.* when the number of measurements and non-zero amplitude values increase with constant asymptotic ratio. The derived closed-form expressions are easily computable and easy to interpret. In addition, the proposed Bayesian lower is valid for any amplitudes vector priors. Finally, several numerical simulations illustrate the practical interest of the proposed Bayesian bound regarding the state-of-art sparse estimators.

1 Introduction

L'acquisition comprimée ou « compressed sensing » en anglais est un domaine en plein essor [2, 8]. Cette thématique de recherche pose le postulat qu'un signal parcimonieux, c'est-à-dire, ayant une représentation composée par un faible nombre de valeurs non-nulles dans une base de décomposition donnée (Fourier, Ondelette, Base canonique réelle, etc) est échantillonnable à un rythme inférieur à celui de Nyquist. Ce paradigme a donc été exploité avec succès dans de nombreuses applications telles que la tomographie, les applications d'imagerie RADAR, le traitement d'antennes, etc. Les performances des algorithmes d'estimation pour l'acquisition comprimée (par exemple [9, 14]) doivent vérifier les conditions RIP (Restricted Isometry Property) [8]. Ces conditions nous in-

diquent que les propriétés structurelles du dictionnaire de représentation sont essentielles. En particulier, les dictionnaires générés de manière aléatoire suivant une distribution sub-Gaussienne vérifient les conditions RIP avec une probabilité élevée [2]. Ce cadre sera le contexte de travail de cet article. Plus précisément, nous supposerons que la matrice de dictionnaire sur-complet est (*i*) de type non-stochastique, c'est-à-dire, vue comme une réalisation unique générée selon une distribution donnée et (*ii*) pour un signal parcimonieux, les dimensions de la matrice dictionnaire croissent selon un rapport asymptotiquement constant. Ce contexte est celui postulé par la théorie des matrices aléatoires de grandes dimensions, qui s'est avérée un cadre foisonnant tant du point de théorique qu'applicatif [1, 10, 13]. Dans ce contexte, nous proposons une étude de la Borne de Cramér-Rao Moyennée (BCRM) [3, 16] caractérisant les performances optimales d'estimation des amplitudes non-nulles supposées aléatoires dans le cas où

*Ce travail est soutenu par les projets NEWCOM# (Grant agreement no. 318306), ANR MAGELLAN et MI-CNRS Titan

il existe une connaissance *a priori* et partielle du signal d'interférence. Ce cas de figure est d'intérêt pratique (cf. [5–7, 11, 12] et les références dans ces travaux). L'intérêt de la BCR dans le cadre de l'acquisition comprimée a été démontré dans les références [4, 15].

2 Acquisition comprimée exploitant une connaissance *a priori*

2.1 Description du modèle d'acquisition comprimée

Soit \mathbf{y} un vecteur composé de N mesures corrompues par un bruit \mathbf{n} supposé centré circulaire Gaussien et blanc de variance σ^2 . Le modèle d'acquisition comprimée [8] est alors défini selon

$$\mathbf{y} = \Psi \mathbf{s} + \mathbf{n} = \Psi \Phi \mathbf{x} + \mathbf{n} \quad (1)$$

où Ψ est la matrice de mesure connue de dimensions $N \times K$ où $N < K$, le vecteur $\mathbf{x} = \Phi^T \mathbf{s}$ de dimensions $K \times 1$ admet une représentation L -parcimonieuse dans la base Φ (e.g. Fourier, Ondelettes, base canonique, etc) avec $L < N$ et $\mathbf{H} \stackrel{\text{def.}}{=} \Psi \Phi$ est typiquement nommé matrice dictionnaire sur-complet.

Une question centrale à la théorie de l'acquisition comprimée est combien de mesures, N , doivent être acquises pour permettre l'estimation du vecteur \mathbf{x} . Pour répondre à cette question, les propriétés de la matrice dictionnaire \mathbf{H} sont fondamentales afin de garantir un niveau de performance satisfaisant. Dans [2, 8], les auteurs introduisent la notion de conditions dites "Restricted Isometry Property (RIP)" et prouvent le rôle primordial que ces conditions RIP jouent. Plus précisément, on dit que la matrice \mathbf{H} vérifie les conditions RIP d'ordre L si il existe une constante $\epsilon_L \in (0, 1)$ vérifiant :

$$\|\mathbf{x}\|^2 (1 - \epsilon_L) \leq \|\mathbf{H}\mathbf{x}\|^2 \leq \|\mathbf{x}\|^2 (1 + \epsilon_L) \quad (2)$$

pour tout vecteur \mathbf{x} de parcimonie d'ordre L . Les conditions RIP caractérisant les matrices \mathbf{H} bénéficiant de la propriété de quasi-isométrie pour un vecteur \mathbf{x} de parcimonie d'ordre L . Une stratégie standard pour vérifier les conditions RIP avec une probabilité haute est de considérer les entrées de la matrice de mesure, Φ , suivant une distribution sub-Gaussienne i.i.d., de moyenne nulle et de variance $1/N$.

2.2 Acquisition comprimée et connaissance *a priori*

Dans les applications de contrôle non destructif de défauts [5, 6], de traitement d'antennes [7], communications numériques [11] ou encore en traitement RADAR [12], par exemple, il est souvent possible d'avoir une connaissance *a priori* d'un certain nombre de paramètres caractérisant le modèle. Ceux-ci sont issus par exemple de mesures réalisées au préalable afin de caractériser le médium de transmission. Par exemple, en RADAR, il est usuel de disposer d'un ensemble de données où on a la connaissance *a priori* que les cibles sont absentes. Il est alors possible d'estimer un ensemble de paramètres de retards, notés $\mathcal{T} = \{\tau_1, \dots, \tau_{L_B}\}$, décrivant les temps d'arrivée des échos RADAR de type clutter. Le modèle en présence de L_A cibles d'intérêt est alors décomposable en deux contributions selon $s(k) = \sum_{\ell=1}^{L_A} \alpha_\ell g(kT_S - \tau_\ell) + i(k)$ où $g(t)$ est une forme d'onde connue, $i(k) = \sum_{\ell=1}^{L_B} \beta_\ell g(kT_S - \tau_\ell)$ est un signal d'interférence partiellement connu, c'est-à-dire que l'ensemble \mathcal{T} est connu *a priori* mais pas les amplitudes β_ℓ et α_ℓ qui sont les paramètres d'intérêts.

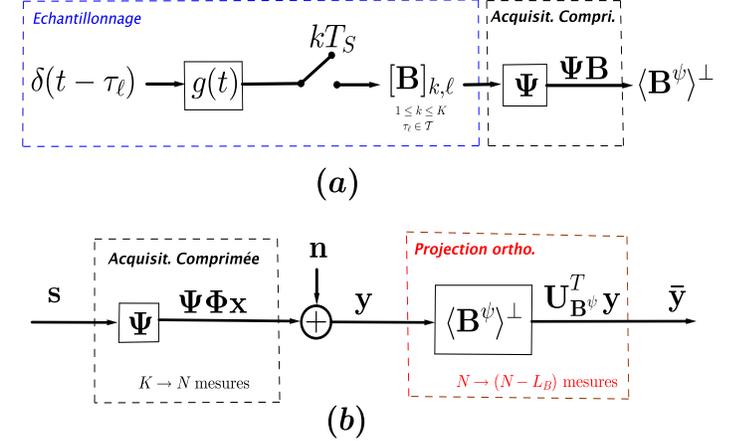


FIGURE 1 – (a) Schéma de construction du sous-espace $\langle \mathbf{B}^\psi \rangle^\perp$, (b) Chaîne globale de traitements de l'information

térisant le modèle. Ceux-ci sont issus par exemple de mesures réalisées au préalable afin de caractériser le médium de transmission. Par exemple, en RADAR, il est usuel de disposer d'un ensemble de données où on a la connaissance *a priori* que les cibles sont absentes. Il est alors possible d'estimer un ensemble de paramètres de retards, notés $\mathcal{T} = \{\tau_1, \dots, \tau_{L_B}\}$, décrivant les temps d'arrivée des échos RADAR de type clutter. Le modèle en présence de L_A cibles d'intérêt est alors décomposable en deux contributions selon $s(k) = \sum_{\ell=1}^{L_A} \alpha_\ell g(kT_S - \tau_\ell) + i(k)$ où $g(t)$ est une forme d'onde connue, $i(k) = \sum_{\ell=1}^{L_B} \beta_\ell g(kT_S - \tau_\ell)$ est un signal d'interférence partiellement connu, c'est-à-dire que l'ensemble \mathcal{T} est connu *a priori* mais pas les amplitudes β_ℓ et α_ℓ qui sont les paramètres d'intérêts.

Selon le modèle ci-dessus, il est possible de spécifier la matrice de base selon $[\Phi]_{k,k'} = g(kT_S - \tau_{k'})$ où $1 \leq k, k' \leq K$. Sans perte de généralité, la matrice de base suit la réécriture suivante $\Phi = [\mathbf{A} \ \mathbf{B} \ \dots]$ et $\mathbf{x} = [\boldsymbol{\alpha}^T \ \boldsymbol{\beta}^T \ \dots]^T$. En se basant sur la connaissance *a priori* de l'ensemble \mathcal{T} , il est alors possible d'exploiter la connaissance du sous-espace $\langle \mathbf{B}^\psi \rangle^\perp$ où $\mathbf{B}^\psi = \Psi \mathbf{B}$. La méthodologie utilisée est schématisée sur la Fig. 1-(a). Plus précisément, soit la décomposition du projecteur $\mathbf{P}^\perp = \mathbf{I} - \mathbf{B}^\psi (\mathbf{B}^\psi)^\dagger = \mathbf{U}_{\mathbf{B}^\psi} \mathbf{U}_{\mathbf{B}^\psi}^T$ où $\mathbf{U}_{\mathbf{B}^\psi}$ est une matrice de taille $(N - L_B) \times L_A$, vérifiant $\mathbf{U}_{\mathbf{B}^\psi}^T \mathbf{U}_{\mathbf{B}^\psi} = \mathbf{I}$. En conséquence et selon la Fig. 1-(b), le modèle de l'observation après acquisition comprimée puis projection est donné par

$$\bar{\mathbf{y}} = \mathbf{U}_{\mathbf{B}^\psi}^T \mathbf{y} = \mathbf{U}_{\mathbf{B}^\psi}^T \mathbf{A}^\psi \boldsymbol{\alpha} + \bar{\mathbf{n}} \quad (3)$$

où $\bar{\mathbf{n}} = \mathbf{U}_{\mathbf{B}^\psi}^T \mathbf{n}$ et $\mathbf{A}^\psi = \Psi \mathbf{A}$. L'objectif de ce travail est alors de présenter le calcul de la Borne de Cramér-Rao Moyennée pour les paramètres d'amplitude $\boldsymbol{\alpha}$, de distribution connue $p(\boldsymbol{\alpha})$, en se basant sur le modèle d'observation ci-dessus et dans un contexte asymptotique, soit pour $N \rightarrow \infty$ et $L_A, L_B \rightarrow \infty$ où $N/L_A \rightarrow \rho$ et $(N - L_B)/L_A \rightarrow \tilde{\rho}$.

3 Performance d'estimation en régime asymptotique

Pour tout estimé $\hat{\boldsymbol{\alpha}}(\bar{\mathbf{y}})$ de $\boldsymbol{\alpha}$, la Borne de Cramér-Rao Moyennée (BCRM) [3, 16] borne inférieurement l'Erreur Quadratique Moyenne Bayésienne (EQMB) selon l'inégalité suivante :

$$\frac{1}{L_A} \mathbb{E}_{\bar{\mathbf{y}}, \boldsymbol{\alpha}} [\|\hat{\boldsymbol{\alpha}}(\bar{\mathbf{y}}) - \boldsymbol{\alpha}\|^2] \geq \mathbb{E}_{\boldsymbol{\alpha}} [C(\boldsymbol{\alpha})]. \quad (4)$$

Pour le modèle considéré, il est simple de montrer grâce à l'inégalité de Jensen [16] que la BCRM coïncide avec la BCR Bayésienne proposée par Vantrees [16] pour une puissance de bruit suffisamment faible et pour tout prior sur les paramètres d'intérêt. Notons de plus, qu'il est raisonnable de normaliser la BCRM par le nombre de degré de liberté de notre problème, *i.e.*, le nombre de paramètres inconnus. En observant maintenant que $\bar{\mathbf{y}} \sim \mathcal{N}(\mathbf{U}_{\mathbf{B}^\psi}^T \mathbf{A}^\psi \boldsymbol{\alpha}, \sigma^2 \mathbf{I})$, après quelques calculs (omis ici), la BCRM s'exprime selon

$$\mathbb{E}_{\boldsymbol{\alpha}} [C(\boldsymbol{\alpha})] = \frac{\sigma^2}{L_A} \text{Trace} \left\{ \left(\mathbf{A}^{\psi T} \mathbf{P}_{\mathbf{B}^\psi}^\perp \mathbf{A}^\psi \right)^{-1} \right\}. \quad (5)$$

L'intérêt de la BCR a été démontré dans les références [4, 15]. Il s'agit maintenant d'exprimer (5) dans le contexte asymptotique. Ce régime asymptotique est celui défini par l'étude des matrices aléatoires de grandes dimensions [1, 10, 13]. Nous proposons par conséquent le lemme suivant.

Lemme 1. Soit $\mathbf{Z} = \mathbf{U}_{\mathbf{B}^\psi}^T \mathbf{A}^\psi \in \mathbb{R}^{(N-L_B) \times L_A}$ dont les éléments $\{z_{ij}\}_{i,j=1 \dots N-L_B, L_A}$ sont *i.i.d.* de moyenne nulle et de variance $\frac{1}{N}$. Pour $N \rightarrow \infty$, $L_A, L_B \rightarrow \infty$ et $N/L_A \rightarrow \rho > 1$, $(N - L_B)/L_A \rightarrow \tilde{\rho} > 1$, alors

$$\frac{1}{L_A} \text{Trace} \left\{ \left(\mathbf{Z}^T \mathbf{Z} \right)^{-1} \right\} \xrightarrow{p.s.} \frac{\rho}{\tilde{\rho} - 1} \quad (6)$$

où la convergence est presque sûre.

Démonstration. D'après les hypothèses du Lemme 1, la distribution empirique de $\frac{N}{L_A} \mathbf{Z}^T \mathbf{Z}$ converge presque sûrement vers une distribution déterministe dont la densité est celle de Marchenko-Pastur [1]. La transformée de Stieltjes $S_\mu(z)$ de cette distribution vérifie l'équation quadratique suivante [10] : $S_\mu(z) = \frac{-1}{z} + \frac{\tilde{\rho}}{z} \frac{S_\mu(z)}{(1+S_\mu(z))}$ et converge vers le résolvant $\frac{1}{L_A} \text{Trace} \left\{ \left(\frac{N}{L_A} \mathbf{Z}^T \mathbf{Z} - z \right)^{-1} \right\}$ [10, 13]. Notons maintenant que la transformée de Stieltjes pondérée par $\frac{L_A}{N}$ et prise en $z = 0$ correspond à $\frac{1}{L_A} \text{Trace} \left\{ \left(\mathbf{Z}^T \mathbf{Z} \right)^{-1} \right\}$. Le lemme 1 est ainsi démontré en exprimant $S_\mu(0)$. \square

Le lemme précédent nous permet d'énoncer le résultat fondamental suivant.

Résultat 1. L'Erreur Quadratique Moyenne Bayésienne Normalisée, en régime asymptotique définie selon les hypothèses du Lemme 1, est bornée inférieurement par la

BCRM selon l'expression

$$\lim_{N \rightarrow \infty} \frac{\mathbb{E}_{\bar{\mathbf{y}}, \boldsymbol{\alpha}} [\|\hat{\boldsymbol{\alpha}}(\bar{\mathbf{y}}) - \boldsymbol{\alpha}\|^2]}{L_A} \geq \lim_{N \rightarrow \infty} \mathbb{E}_{\boldsymbol{\alpha}} [C(\boldsymbol{\alpha})] = \sigma^2 \frac{\rho}{\tilde{\rho} - 1} \quad (7)$$

Démonstration. Ce résultat s'obtient en exploitant l'expression (5) et en appliquant le résultat du Lemme 1. \square

4 Simulations et Résultats

Dans cette section, nous comparons plusieurs estimateurs parcimonieux standards pour l'estimation du vecteur $\boldsymbol{\alpha}$ de distribution Laplacienne. Afin d'intégrer dans ces estimateurs, la connaissance partielle de l'observation, nous considérons le dictionnaire projeté sur l'espace $\langle \mathbf{B}^\psi \rangle^\perp$. Les Erreurs Quadratiques Moyennes (EQM) de ces estimateurs sont évaluées grâce à des simulations de type Monte-Carlo moyennées sur 500 réalisations portant sur le bruit et les amplitudes. La BCRM, donnée en (5), coïncidant avec la BCR Bayésienne pour une variance de bruit suffisamment petite, permet de caractériser l'efficacité statistique d'un estimateur. Selon la Fig. 2-(a), les méthodes BPDN [9] et Cosamp [14] n'atteignent pas la BCRM même pour des RSB élevés. Toutefois l'algorithme Cosamp présente de meilleures performances que l'algorithme BPDN. Parmi les trois algorithmes comparés, la méthode OMP est efficace sur un intervalle de RSB étendu.

Cette partie de l'analyse se concentre sur l'expression exacte de la BCRM donnée par (5) et son expression en régime asymptotique donnée en (7). Leur courbes, affichées en Fig. 2-(b), se confondent, ce qui confirme la validité de (7) même pour des valeurs L_A petites. Outre la validation de la BCRM asymptotique, la Fig. 2-(b) apporte également une information importante sur le comportement des méthodes d'estimation des amplitudes dans le contexte de l'acquisition comprimée. Nous remarquons au travers de cette simulation que plus le vecteur des amplitudes est parcimonieux, plus les performances d'estimation sont bonnes. En effet, la meilleure performance, *i.e.* la plus faible EQM, est obtenue pour $L_A = 4$ et $L_B = 40$, les performances se dégradent nettement par la suite lorsque soit L_A soit L_B soit les deux, augmentent.

5 Conclusion

Dans ce travail, nous avons proposé de considérer les performances minimales d'estimation pour un modèle d'observation bruitée mais partiellement connu dans le contexte de l'acquisition comprimée. Dans ce cadre, nous avons adapté certains estimateurs parcimonieux standards des amplitudes en intégrant la prise en compte de la connaissance partielle d'un signal d'interférence. Ce contexte de travail est souvent adapté à des applications réalistes telles que par exemple le contrôle non destructif de défauts.

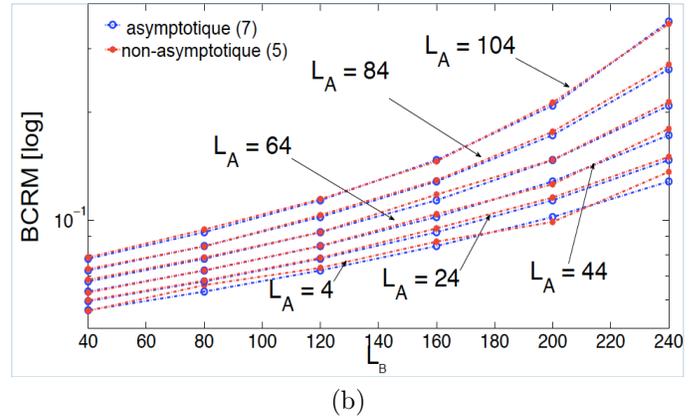
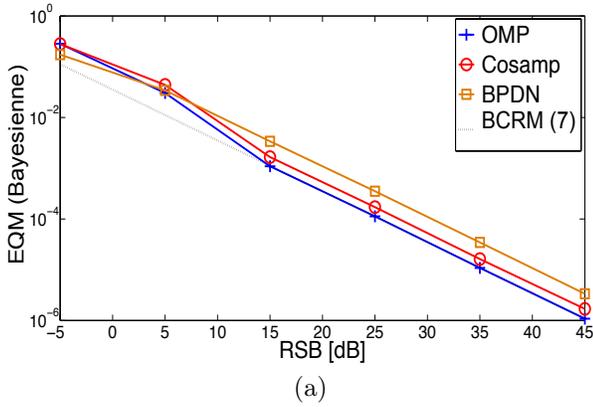


FIGURE 2 – (a) EQM (Bayesienne) Vs. RSB in dB. La BCRM considérée est donnée dans le Résultat 1. Le nombre de mesures est $N = 200$ avec $L_A = 5$ amplitudes d'intérêt parmi $K = 400$ dont $L_B = 60$ amplitudes interférentes. (b) Comparisons en fonction de L_B des expressions analytiques et de la BCRM asymptotique et non-asymptotique données en (5) et (7), respectivement. Nombre de mesures $N = 400$ et $K = 2N$ pour un RSB de 15 dB. Les dimensions de $\langle \mathbf{B}^\psi \rangle$ et $\langle \mathbf{A}^\psi \rangle$ varient en fonction de L_B et L_A respectivement.

Nous dérivons alors la Borne de Cramér-Rao Moyennée (BCRM) associée au modèle d'observation dans un contexte de l'estimation Bayésienne d'un vecteur parcimonieux d'amplitudes. En se basant, sur les propriétés structurelles de la BCRM, il nous a été possible d'exploiter certains résultats puissants issus de la théorie des matrices aléatoires de grandes dimensions. Par ce biais, nous obtenons une expression asymptotique de la BCRM particulièrement compacte et facile à interpréter. Enfin, grâce à des simulations numériques de type Monte-Carlo, nous illustrons et corroborons tout l'intérêt des expressions obtenus de la BCRM au regard de certains estimateur parcimonieux populaires.

Références

- [1] D. Bai and J.W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices", *Annals of Probability*, 26(1) :316–345, 1998
- [2] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices", *Construct. Approx.*, vol. 28, no. 3, pp. 253-263, 2008.
- [3] Z. Ben-Haim and Y.C. Eldar, "A Lower Bound on the Bayesian MSE Based on the Optimal Bias Function", *IEEE Transactions on Information Theory*, vol.55, no.11, pp.5179,5196, Nov. 2009
- [4] Z. Ben-Haim and Y. C. Eldar, "The Cramér-Rao Bound for Estimating a Sparse Parameter Vector", *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3384-3389, June 2010.
- [5] G. Bouleux, F. Guillet, R. Boyer and M. El Badaoui, "A subspace-based rejection method for detecting bearing fault in asynchronous motor", *IEEE International Conference on Condition Monitoring and Diagnosis (CMD 08)*, 2008, Beijing.
- [6] G. Bouleux, "Oblique Projection pre-Processing and TLS Application for diagnosing rotor bar defects by improving power spectrum estimation", *Mechanical Systems and Signal Processing*, Vol. 41, No. 1-2, Dec. 2013.
- [7] Boyer, R. and Bouleux, G., *Oblique projections for direction-of-arrival estimation with prior knowledge*, *IEEE Trans. on Signal Processing*, Volume 56, No. 4, 2008, pp. 1374-1387.
- [8] E. Candes and T. Tao, "Decoding by linear programming", *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4203-4215, 2005.
- [9] S.S. Chen, D.L. Donoho and M.A. Saunders, "Atomic decomposition by basis pursuit", *SIAM Journal on Scientific Computing*, vol. 20, pp. 33-61, 1998.
- [10] R. Couillet and M. Debbah, "Random matrix methods for wireless communications", Cambridge University Press
- [11] Y. Cui, K. Liu, and J. Wang, "Direction-of-arrival estimation for coherent GPS signals based on oblique projection". *Signal Processing*, 92(1), 294-299, 2012.
- [12] J.P. Goy, F. Vincent, et J.Y. Tourneret, "Clutter rejection for MTI radar using a single antenna and a long integration time", *CAMSAP 2011*
- [13] W. Hachem, P. Loubaton, J. Najim, X. Mestre, P. Vallet, "Large information plus noise random matrix models and consistent subspace estimation in large sensor networks", *RMTA*, vol. 1, no. 2, 2012
- [14] D. Needella and J.A. Tropp, "CoSaMP : Iterative signal recovery from incomplete and inaccurate samples", *Applied and Computational Harmonic Analysis*, vol. 26, No 3, pp. 301-321, May 2009.
- [15] R. Niazadeh, M. Babaie-Zadeh and C. Jutten, "On the achievability of Cramér-Rao bound in noisy compressed sensing", *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp.518-526, 2012
- [16] H.L. Van Trees, "Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking", Sept. 2007, Wiley-IEEE Press