

Estimateur de Tyler régularisé dans le cas sous-déterminé. Application à la détection d'objets enfouis.

Quentin HOARAU¹, Arnaud BRELOY², Guillaume GINOLHAC¹, Abdourrahmane M. ATTO¹, Jean-Marie NICOLAS³

¹LISTIC, Université Savoie Mont Blanc, France

²LEME, Université Paris Nanterre, France

³LTCI, Télécom ParisTech, France

{quentin.hoarau, guillaume.ginolhac, abourrahmane.atto}@univ-smb.fr,
abreloy@u-paris10.fr, jean-marie.nicolas@telecom-paristech.fr

Résumé – Parmi les estimateurs de matrice de covariance, l'estimateur de Tyler régularisé offre des résultats indépendants de la distribution statistique des données, tout en étant robuste à la présence de données aberrantes. Cependant, la sélection de la valeur du paramètre de régularisation dépend fortement de l'application ciblée et de la configuration des données, et a une influence directe sur les performances de l'estimateur. Ainsi, proposer une règle de sélection du paramètre de régularisation qui soit générale pour toutes les configurations de données reste un problème non trivial. Une méthode de sélection du paramètre de régularisation dans le cas sous-déterminé est proposée, basée sur une approche par sous-espace. L'intérêt de la méthode est validé à l'aide de simulations, avant de l'appliquer au problème de détection adaptative d'objets enfouis.

Abstract – Among the various covariance matrix estimators, the regularised Tyler estimator performs independently from the data distribution and is robust to data outlier corruption. However, the shrinkage parameter value selection depends on the target application and data configuration, and have a direct influence on the estimator performance. Thus finding a generic rule optimal for every criterion is not straightforward. This paper proposes a new regularisation parameter selection based on a subspace approach. The performances of this method are investigated both in simulation and application to the adaptive buried objects detection problem.

1 Introduction

L'estimation de matrices de covariances est une étape majeure pour un grand nombre d'applications, à l'instar du traitement d'antennes où la précision d'estimation de ce paramètre influe directement sur les performances des détecteurs adaptatifs. À partir d'un jeu de K données $\{\mathbf{x}_k\}_{1,K} \in \mathbb{C}^N$, la *Sample Covariance Matrix* (SCM) est l'estimateur de matrice de covariance le plus utilisé. Cependant ses performances sont moindres quand le jeu de donnée est corrompu par des données aberrantes et/ou que la distribution statistique est hétérogène.

Pour faire face à cette difficulté, les M -Estimateurs (estimateurs du maximum de vraisemblance généralisés pour les distributions CES) sont prisés, du fait de leur robustesse [1]. Néanmoins, ces estimateurs ne sont pas adaptés pour les cas où les \mathbf{x}_k sont de grande taille mais plutôt en faible nombre. Plus particulièrement, les M -estimateurs ne sont pas définis dans le cas sous-déterminé ($K < N$) tandis que la règle générale suggère $K > 2N$ pour obtenir de bonnes performances d'estimation.

Des méthodes de régularisation ont été proposées pour traiter le problème [2, 3, 4, 5, 6] et les estimateurs de Tyler régularisés (RTEs), qui s'expriment :

$$\hat{\mathbf{R}}_{RTE}(\alpha) = (1 - \alpha) \frac{N}{K} \sum_{k=1}^K \frac{\mathbf{x}_k \mathbf{x}_k^H}{\mathbf{x}_k^H \hat{\mathbf{R}}_{RTE}^{-1}(\alpha) \mathbf{x}_k} + \alpha \mathbf{I}_N \quad (1)$$

sont maintenant cibles de travaux d'optimisation de leur performances [7, 8, 9, 10]. L'ajout d'une part de matrice iden-

tité géré du paramètre α assure l'existence de l'estimateur dans le cas sous-déterminé (les RTEs existent et sont uniques pour $\alpha \in (\max(0, 1 - K/N); 1]$), ainsi qu'un bon conditionnement de la matrice. Les estimateurs sont calculés à l'aide des itérations :

$$\hat{\mathbf{R}}^{(i+1)} = (1 - \alpha) \frac{N}{K} \sum_{k=1}^K \frac{\mathbf{x}_k \mathbf{x}_k^H}{\mathbf{x}_k^H \hat{\mathbf{R}}^{(i)-1} \mathbf{x}_k} + \alpha \mathbf{I}_N, \quad (2)$$

convergeant vers la solution solution $\hat{\mathbf{R}}_{RTE}(\alpha)$.

Arrive avec ce nouveau type d'estimateur la question du choix de la valeur du paramètre de régularisation α . En effet, le compromis entre biais et performance dépend fortement de l'application, ce qui empêche l'établissement d'une règle générale qui soit optimale pour tout critère et/ou distribution statistique. Plusieurs approches existent dans la littérature :

- Oracles associés à l'algorithme (2) : [5] minimise l'erreur de forme à une constante multiplicative près.
- Estimateurs basés sur la théorie des matrices aléatoires (RMT) (K et N tendent vers l'infini à ratio K/N fixe) : [7] minimise l'erreur quadratique moyenne (MSE), [9] optimise les performances du détecteur ANMF, [8] minimise le "portfolio variance" (similaire au critère SINR-Loss).
- Variantes de l'algorithme (2) (p.ex: itérations avec normalisation de la trace) : [2] oracle minimisant la MSE, [11, 12] maximise le rapport de vraisemblance.

À la lumière de ces travaux, on propose d'adapter les oracles de [2, 5] pour le cas sous-déterminé ($K < N$) à l'aide d'une approche par réduction de dimension [12] pour deux raisons principales :

- les oracles aboutissent à des formes analytiques facilitant l'implémentation,
- les oracles proposés par [2, 5] ne sont pas adaptés au cas sous-déterminé, comme expliqué et illustré dans la section simulation de cet article.

On dérivera aussi l'oracle de [2] adapté pour l'algorithme (2) dans le cas réel comme complexe. On étudiera les performances de la méthode proposée en simulation et sur données réelles, appliquée au problème de détection adaptative d'objets enfouis.

2 Contexte

2.1 Estimateur de Tyler régularisé

Soit $\{\mathbf{x}_k\}_{1,K} \in \mathbb{C}^N$ un jeu de K vecteurs aléatoires $\mathbf{x}_k \sim CE_N(\mathbf{0}, \mathbf{R}, g)$, de moyenne $\mathbf{0} \in \{0\}^N$, matrice de dispersion $\mathbf{R} \in \mathbb{C}^{N \times N}$ et génératrice g . Dans la plupart des cas, la vraie valeur de \mathbf{R} est inconnue et est estimée à partir des données. L'estimateur du maximum de vraisemblance minimise la fonction :

$$\mathcal{L}(\mathbf{R}) = \frac{1}{K} \sum_{k=1}^K \rho(\mathbf{x}_k^H \mathbf{R}^{-1} \mathbf{x}_k) - \ln |\mathbf{R}^{-1}|, \quad (3)$$

avec $\rho(t) = -\ln g(t)$. Cet estimateur peut être généralisé pour obtenir un M -Estimateur en choisissant une fonction ρ plus générale. Pour le cas particulier $\rho(t) = N \ln t$, (3) aboutit à l'estimateur de Tyler $\hat{\mathbf{R}}_{TE}$. Cependant, quand le nombre de données K est trop faible ($K < 2N$) la matrice $\hat{\mathbf{R}}_{TE}$ est mal conditionnée, et dans le cas $K < N$ elle n'existe pas. L'estimateur est alors régularisé pour contourner le problème. [13] propose un premier algorithme pour calculer l'estimateur de Tyler régularisé :

$$\begin{cases} \tilde{\mathbf{R}}_{CWH}^{(i+1)} = (1 - \alpha) \frac{N}{K} \sum_{k=1}^K \frac{\mathbf{x}_k \mathbf{x}_k^H}{\mathbf{x}_k^H \hat{\mathbf{R}}_{CWH}^{(i-1)} \mathbf{x}_k} + \alpha \mathbf{I}_N, \\ \hat{\mathbf{R}}_{CWH}^{(i+1)} = N \frac{\tilde{\mathbf{R}}_{CWH}^{(i+1)}}{\text{Tr}(\tilde{\mathbf{R}}_{CWH}^{(i+1)})}. \end{cases} \quad (4)$$

Ces itérations du point-fixe convergent $\forall \alpha \in (0; 1)$ [2], mais cette solution est heuristique car elle ne découle d'aucune minimisation de fonction de coût. L'estimateur de Tyler régularisé comme solution d'une minimisation de fonction de coût pénalisée a été donnée dans [4, 5, 6] :

$$\mathcal{L}_\alpha(\mathbf{R}) = \frac{1}{K} \sum_{k=1}^K \rho(\mathbf{x}_k^H \mathbf{R}^{-1} \mathbf{x}_k) - \ln |\mathbf{R}^{-1}| + \alpha \mathcal{P}(\mathbf{R}), \quad (5)$$

avec $\mathcal{P}(\mathbf{R}) = \text{Tr}(\mathbf{R}^{-1})$. On obtient la solution unique (1), calculée par les itérations (2), qui convergent pour $\alpha \in (\max(0, 1 - \frac{K}{N}); 1]$. On se focalisera sur cet algorithme dans la suite de cet article.

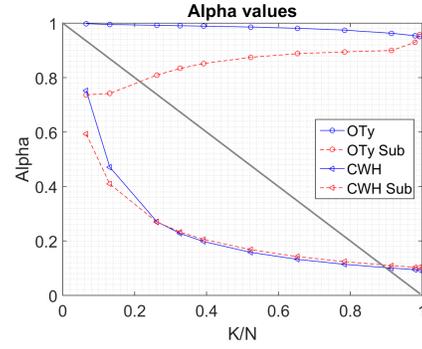


FIG. 1: Valeurs de α pour différents ratios K/N . La ligne grise représente la limite de convergence. Paramètres : $N = 153$, $\rho = 0.9$, $\nu = 0.5$, résultats moyennés sur 1000 réalisations.

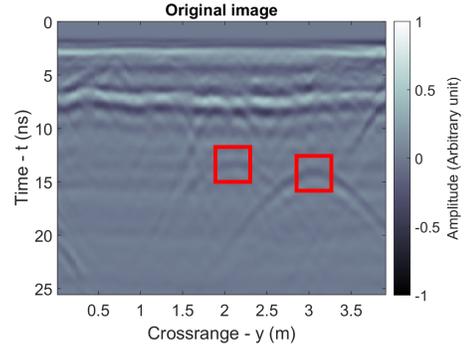


FIG. 2: Image GPR originale. La position des objets est indiquée par des carrés rouges.

2.2 Sélection du paramètre de régularisation

La valeur du paramètre α doit être choisie en fonction de l'application. [2] propose un oracle minimisant la MSE :

$$\alpha_{CWH} = \arg\max_{\alpha} E[\|\mathbf{R} - \hat{\mathbf{R}}\|^2]. \quad (6)$$

Proposition 2.1 Soit $\mathfrak{C}_1 = ((N + 1)(K + 1) - 4) \text{Tr}(\mathbf{R}^2)$ et $\mathfrak{C}_2 = (KN + K - 1) \text{Tr}(\mathbf{R}^2)$. Pour des vecteurs CES i.i.d, la solution à (6) pour l'algorithme (2) est, dans le cas réel :

$$\alpha_{CWH} = \frac{(N - 2) \text{Tr}(\mathbf{R}^2) + N \text{Tr}^2(\mathbf{R})}{\mathfrak{C}_1 + N \text{Tr}^2(\mathbf{R}) - 2K(N + 2) \text{Tr}(\mathbf{R}) + KN(N + 2)}, \quad (7)$$

et dans le cas complexe :

$$\alpha_{CWH} = \frac{N \text{Tr}^2(\mathbf{R}) - \text{Tr}(\mathbf{R}^2)}{\mathfrak{C}_2 + N \text{Tr}^2(\mathbf{R}) - 2K(N + 1) \text{Tr}(\mathbf{R}) + KN(N + 1)} \quad (8)$$

Preuve: Voir [2] avec adaptations mineures pour le cas complexe, et puisque l'on se place dans le cas de l'algorithme (2), la normalisation $\text{Tr}(\mathbf{R}) = N$ ne s'applique pas.

[5] propose un second critère d'oracle pour α , en se focalisant sur une estimation correcte de la forme de la matrice plutôt que ses valeurs :

$$\alpha_{OTy} = \arg\max_{\alpha} E[\|\mathbf{R}^{-1} \hat{\mathbf{R}} - \frac{\text{Tr}(\mathbf{R}^{-1} \hat{\mathbf{R}})}{N} \mathbf{I}_N\|^2]. \quad (9)$$

Pour cet oracle, les formules analytiques et leur démonstrations sont disponibles dans [5].

Ces valeurs de α sont appelés oracles car il nécessitent la connaissance de $\text{Tr}(\mathbf{R})$ et $\text{Tr}(\mathbf{R}^2)$. La vraie matrice de covariance étant inconnue, on peut la remplacer par un RTE comme proposé dans [5]. Dans le cas $K \simeq N$, la substitution semble assez précise, mais quand $K \ll N$ l'estimation des traces devient grandement imprécise, mettant tout le poids sur l'identité ou causant la divergence des itérations (cf. Section 4). Afin de résoudre ce problème, on propose une nouvelle méthode d'estimation des matrices de covariances dans le cas sous-déterminé, en s'inspirant du travail de [12].

3 Approche par sous-espace

Le processus d'estimation proposé consiste à (i) projeter les données sur le K -sous-espace qu'elles engendrent, (ii) estimer la matrice de covariance dans ce sous-espace à l'aide d'un RTE et (iii) faire l'expansion du résultat précédent pour obtenir un estimateur $N \times N$.

3.1 Réduction de dimension

La première étape est de récupérer le sous-espace engendré par les données $\{\mathbf{x}_k\}$ à l'aide de la SVD de la SCM, $\hat{\mathbf{R}}_{SCM} = \hat{\mathbf{U}} \hat{\mathbf{D}} \hat{\mathbf{U}}^H$, avec $\hat{\mathbf{D}}$ une matrice diagonale et $\hat{\mathbf{U}} = [\hat{\mathbf{u}}_1 \dots \hat{\mathbf{u}}_N]$. Dans le cas $K < N$, seules K valeurs propres sont non-nulles et les vecteurs propres associés $\hat{\mathbf{U}}_K = [\hat{\mathbf{u}}_1 \dots \hat{\mathbf{u}}_K]$ (K premières colonnes de $\hat{\mathbf{U}}$) engendrent le sous-espace des données. On note aussi $\hat{\mathbf{U}}_K^\perp = [\hat{\mathbf{u}}_{K+1} \dots \hat{\mathbf{u}}_N]$. À l'aide de cette base, la vraie matrice de covariance \mathbf{R} est décomposée en deux blocs :

$$\mathbf{R} = \begin{bmatrix} \hat{\mathbf{U}}_K & \hat{\mathbf{U}}_K^\perp \end{bmatrix} \begin{bmatrix} \mathbf{R}_K & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_K^\perp \end{bmatrix} \begin{bmatrix} \hat{\mathbf{U}}_K & \hat{\mathbf{U}}_K^\perp \end{bmatrix}^H. \quad (10)$$

Comme expliqué dans [12], seul \mathbf{R}_K peut être correctement estimé, ce qui est fait à l'aide des projections des vecteurs \mathbf{x}_k sur le sous-espace $\tilde{\mathbf{x}}_k = \hat{\mathbf{U}}_K^H \mathbf{x}_k$.

3.2 RTE dans le K -sous-espace

À l'aide des projections $\tilde{\mathbf{x}}_k$, le problème passe du cas $N \gg K$ au cas $N = K$ où il y a possibilité d'une bonne estimation de $\text{Tr}(\mathbf{R}_K)$ pour l'injecter dans les oracles α ainsi que la suppression des conditions de convergence sur α pour le RTE le rendant de ce fait compatible avec tous les oracles présentés à la section 2. Le RTE des projections est calculé à l'aide de l'équation (2) où on applique $N = K$, et la valeur oracle α choisie est notée $\hat{\alpha}_0$.

3.3 Expansion

Après avoir calculé $\hat{\mathbf{R}}_K(\hat{\alpha}_0)$, il reste à retourner dans le N -espace de départ et y reconstruire l'estimateur de la matrice de covariance $\hat{\mathbf{R}}$. Comme le bloc $\hat{\mathbf{R}}_K^\perp$ ne peut être estimé, on le remplace par une solution non-informative $\alpha^\perp \mathbf{I}_{N-K}$. On propose ici de choisir $\alpha^\perp = \hat{\alpha}_0$ pour ne pas dénaturer le spectre de la matrice estimée, puisque la plus faible valeur propre de $\hat{\mathbf{R}}_K$ est supérieure à $\hat{\alpha}_0$. $\hat{\mathbf{R}}$ s'exprime alors comme :

$$\hat{\mathbf{R}} = \begin{bmatrix} \hat{\mathbf{U}}_K & \hat{\mathbf{U}}_K^\perp \end{bmatrix} \begin{bmatrix} \hat{\mathbf{R}}_K(\hat{\alpha}_0) & \mathbf{0} \\ \mathbf{0} & \hat{\alpha}_0 \mathbf{I}_{N-K} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{U}}_K & \hat{\mathbf{U}}_K^\perp \end{bmatrix}^H. \quad (11)$$

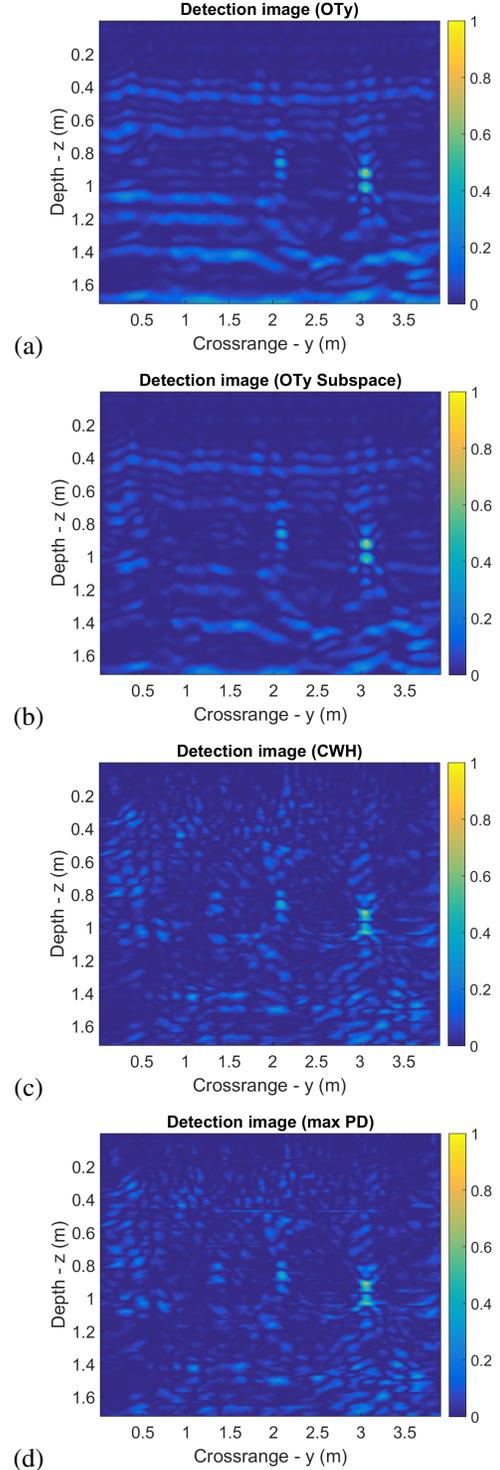


FIG. 3: Résultats pour (a) α_{OTy} sans projection ($\alpha \in [0.998; 1]$) (b) α_{OTy} avec projection ($\alpha \in [0.8; 0.95]$), (c) α_{CWH} avec projection ($\alpha \in [0.1; 0.4]$), et (d) maximum PD $\hat{\alpha}_{PD}$ ($\alpha \in [0.4; 0.9]$). Paramètres : $N = 153$, $K = 40$

4 Simulations

Cette section présente des simulations visant à illustrer la supériorité de la méthode proposée par rapport aux oracles proposés dans la section 2 dans le cas sous-déterminé. Pour se placer en ligne avec les données de la section 5, les simulations sont

faites dans le cas réel, avec $\{\mathbf{x}_k\}_{1,K}$ générés selon un modèle SIRV, $\mathbf{x}_k = \sqrt{\tau_k} \mathbf{g}_k$, avec $\mathbf{g}_k \sim \mathcal{N}_N(\mathbf{0}, \mathbf{R})$, \mathbf{R} Toeplitz $[\mathbf{R}]_{ij} = \rho^{|i-j|}$, et $\tau \sim \Gamma(\nu, 1/\nu)$.

La fig. 1 montre les valeurs de $\hat{\alpha}_0$ pour les deux critères avec et sans projection sur le sous-espace. Une première observation est que l'oracle CWH retourne des valeurs similaires dans les deux cas mais sous la limite de convergence, ce qui montre que la projection permet désormais d'utiliser cet oracle avec l'algorithme (2). Pour le critère de forme on observe une amélioration des valeurs de $\hat{\alpha}_{OTy}$, la projection sur le sous-espace permettant d'alléger le poids de l'identité et de faire meilleur usage des données.

5 Application sur données réelles

La méthode proposée est appliquée au problème de détection adaptative d'objets sur image GPR [14]. Le problème consiste à détecter un signal connu $\mathbf{p} \in \mathbb{R}^N$ (réponse d'un objet enfoui) dans une observation $\mathbf{x} \in \mathbb{R}^N$, en ayant à disposition un jeu d'observation secondaires supposées sans signal utile $\{\mathbf{x}_k\}_{1,K}$ avec $K \ll N$. Deux hypothèses sont formulées $H_0: \mathbf{x} = \mathbf{n}$ et $H_1: \mathbf{x} = a\mathbf{p} + \mathbf{n}$, qui mène après résolution au détecteur :

$$\hat{\Lambda} = \max_{\epsilon' \in \mathbb{R}^+} \frac{|\mathbf{p}^T \hat{\mathbf{R}}^{-1} \mathbf{x}|^2}{(\mathbf{p}^T \hat{\mathbf{R}}^{-1} \mathbf{p})(\mathbf{x}^T \hat{\mathbf{R}}^{-1} \mathbf{x})} \underset{H_1}{\overset{H_0}{\leq}} \eta. \quad (12)$$

Ce détecteur est appliqué à une image contenant deux objets (fig. 2). Trois estimateurs sont utilisés : le RTE avec $\hat{\alpha}_{OTy}$ avec et sans projection sur le sous-espace, le RTE avec $\hat{\alpha}_{CWH}$ et avec projection. Un quatrième estimateur est ajouté à l'étude pour comparaison : le RTE avec $\hat{\alpha}_{PD}$ qui maximise la probabilité de détection en s'appuyant sur la théorie des matrices aléatoires [9].

Les résultats de la fig. 3 montrent une amélioration de la détection pour le critère (9), réduisant le niveau de bruit dans le bas de l'image. L'estimateur avec $\hat{\alpha}_{CWH}$ renvoie un niveau de bruit encore plus faible. Ses résultats sont également comparables à ceux du paramètre $\hat{\alpha}_{PD}$ tout en étant bien plus rapide à calculer. Pour une meilleure comparaison des méthodes, on applique un seuillage à 20% du maximum de chaque image pour obtenir le taux de PFA correspondant : $PFA_{OTy} = 1.15 \cdot 10^{-2}$, $PFA_{OTySub} = 2.1 \cdot 10^{-3}$, $PFA_{CWH} = 1.4 \cdot 10^{-3}$, $PFA_{PD} = 1.2 \cdot 10^{-3}$.

6 Conclusion

Une nouvelle approche pour l'estimation des matrices de covariance dans le cas $K \ll N$ a été proposée, basée sur la projection des données sur le K-sous-espace qu'elles engendrent. Cette approche supprime les problèmes de convergence de l'estimateur de Tyler et permet une meilleure estimation des oracles $\hat{\alpha}$, comme illustré lors des simulations. L'application au problème de détection adaptative a montré des performances satisfaisantes.

References

[1] E. Ollila, D. Tyler, V. Koivunen, and H. Poor, "Complex elliptically symmetric distributions: Survey, new results

and applications," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5597–5625, 2012.

- [2] Y. Chen, A. Wiesel, and A. O. Hero, "Robust shrinkage estimation of high-dimensional covariance matrices," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4097–4107, 2011.
- [3] A. Wiesel, "Unified framework to regularized covariance estimation in scaled Gaussian models," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 29–38, 2012.
- [4] F. Pascal, Y. Chitour, and Y. Quek, "Generalized robust shrinkage estimator and its application to stap detection problem," *IEEE Transactions on Signal Processing*, vol. 62, no. 21, pp. 5640–5651, Nov 2014.
- [5] E. Ollila and D. Tyler, "Regularized M -estimators of scatter matrix," *IEEE Transactions on Signal Processing*, vol. 62, no. 22, pp. 6059–6070, Nov 2014.
- [6] Y. Sun, P. Babu, and D. Palomar, "Regularized Tyler's scatter estimator: Existence, uniqueness, and algorithms," *IEEE Transactions on Signal Processing*, vol. 62, no. 19, pp. 5143–5156, Oct 2014.
- [7] R. Couillet and M. McKay, "Second order statistics of robust estimators of scatter. application to glrt detection for elliptical signals," *Elsevier Journal of Multivariate Analysis*, vol. 131, pp. 99–120, 2014.
- [8] L. Yang, R. Couillet, and M. McKay, "A robust statistics approach to minimum variance portfolio optimization," *IEEE Transactions on Signal Processing*, vol. 63, no. 24, pp. 6684–6697, 2015.
- [9] A. Kammoun, R. Couillet, F. Pascal, and i. M. Alouin, "Optimal design of the adaptive normalized matched filter detector," (*submitted to*) *IEEE Transactions on Information Theory*, *arXiv Preprint 1504.01252*, 2016.
- [10] A. Kammoun, R. Couillet, F. Pascal, and M. Alouini, "Convergence and fluctuations of regularized tyler estimators," *IEEE Transactions on Signal Processing*, vol. 64, no. 4, pp. 1048–1060, 2016.
- [11] Y. I. Abramovich and O. Besson, "Regularized covariance matrix estimation in complex elliptically symmetric distributions using the expected likelihood approach; part 1: The over-sampled case," *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5807–5818, Dec 2013.
- [12] O. Besson and Y. I. Abramovich, "Regularized covariance matrix estimation in complex elliptically symmetric distributions using the expected likelihood approach; part 2: The under-sampled case," *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5819–5829, Dec 2013.
- [13] Y. I. Abramovich and N. K. Spencer, "Diagonally loaded normalised sample matrix inversion (Insmi) for outlier-resistant adaptive filtering," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, vol. 3, April 2007, pp. III–1105–III–1108.
- [14] Q. Hoarau, G. Ginolhac, A. Atto, and J. Nicolas, "Robust adaptive detection of buried pipes using gpr," *Signal Processing*, pp. –, 2016.