

Estimation de l’Ordre de Modèle pour un Bruit Elliptique Symétrique Complexe en Grande Dimension

Eugénie TERREAUX¹, Jean-Philippe OVARLEZ^{1,2}, Frédéric PASCAL³

¹CentraleSupélec-SONDRA
3 rue Joliot Curie, 91190 Gif-Sur-Yvette, France

²ONERA, DEMR/TSI
Chemin de la Humière, 91120 Palaiseau, France

³L2S/CentraleSupélec-CNRS-Université Paris-Sud
3 rue Joliot Curie, 91190 Gif-Sur-Yvette, France

eugenie.terreaux@centralesupelec.fr, jean-philippe.ovarlez@onera.fr
frederic.pascal@centralesupelec.fr

Résumé – Cet article présente une nouvelle méthode de sélection d’ordre de modèle pour des applications en traitement du signal dans le cadre de grandes dimensions (cadre de la Théorie de Matrices Aléatoires (Random Matrix Theory (RMT))), lorsque le bruit, corrélé et supposé non-gaussien, suit un modèle symétrique elliptique complexe (CES) général de matrice de dispersion inconnue. La méthode proposée consiste à estimer la matrice de covariance du bruit grâce à sa structure de Toeplitz, puis, après un processus de blanchiment, les valeurs propres de n’importe quel M -estimateur sont exploitées pour estimer l’ordre du modèle. Cette méthodologie est appliquée sur des signaux issus de simulation et sur des images hyperspectrales et montre son efficacité en comparaison avec d’autres techniques classiques.

Abstract – For signal processing application in large dimensional regime (large Random Matrix Theory), this paper presents a new model order selection technique when the noise is assumed to be Complex Elliptically Symmetric (CES) distributed, with an unknown scatter matrix. The proposed method consists first in estimating the background covariance matrix using its Toeplitz structure. Then, after a whitening process, the eigenvalues distribution of any Maronna’s M -estimators are exploited, leading to the order selection. Simulations on hyperspectral images demonstrate the superiority of the proposed technique compared to other presented methods.

1 Introduction

Les matrices de covariance de signaux constitués de sources et d’un bruit additif, peuvent permettre l’estimation d’ordre de modèle, un problème récurrent en traitement du signal. Cette matrice doit généralement être estimée, et s’il s’agit de N capteurs détectant des signaux de taille m , cette estimation est non seulement dégradée par la présence du bruit mais aussi lorsque N et m grandissent avec un rapport constant [1]. De plus, pour beaucoup d’applications, le bruit est supposé être spatialement et temporellement aléatoire, blanc et gaussien. La plupart des méthodes classiques, telles que celles qui reposent sur le critère d’information théorique (e.g, AIC [2]) échouent dans la sélection d’ordre dès que le bruit ne respecte plus ces hypothèses. De même, lorsque les dimensions de ces observations grandissent, ces méthodes produisent également de mauvais résultats [3]. La théorie des matrices aléatoires (Random Matrix Theory (RMT)), récemment développée en traitement du signal, offre une

alternative intéressante [4].

Cet article présente une nouvelle méthode pour estimer la matrice de covariance lorsque le bruit est corrélé et non-gaussien (pour N et m grands). Cet estimateur est ici développé dans le cas d’un bruit complexe symétrique elliptique (Complex Elliptical Symmetric (CES)) [5], qui caractérise souvent mieux le bruit dans de nombreuses applications. La méthode proposée se décompose en deux parties : “toeplitzification” de la matrice estimée afin de blanchir les observations puis estimation robuste à partir des données blanchies. La partie 2 présente le modèle, ainsi que la “toeplitzification” de la matrice de covariance empirique (la Sample Covariance Matrix (SCM)) et le blanchiment des observations. Ensuite, nous proposons l’utilisation d’un M -estimateur (de type Maronna [6]) pour la matrice de covariance. Comme dans [7], cet estimateur est une alternative aux méthodes AIC et MDL. Enfin, la méthode d’estimation d’ordre du signal est développée : nous prouvons que toutes les valeurs propres supérieures à un seuil sont liées aux sources présentes dans le signal.

La partie 3 illustre la pertinence et les performances de la méthode proposée sur des données simulées ainsi que sur des images hyperspectrales réelles.

Notations : Les vecteurs sont en gras, les matrices sont en gras et en lettres capitales. Soit \mathbf{A} une matrice, \mathbf{A}^T et \mathbf{A}^H sont respectivement la transposée et la transposée hermitienne de \mathbf{A} . $\{\lambda_i(\mathbf{A})\}_{i=1,\dots,m}$ sont les valeurs propres de la matrice carrée \mathbf{A} (taille $m \times m$) classées dans l'ordre décroissant. $\xrightarrow{p.s.}$ représente la convergence presque sûre. a^* est le complexe conjugué d'un scalaire a . $d_1(\cdot)$ est la distance associée à la norme l_1 . La distribution δ symbolise la mesure de dirac, $\text{supp}(\cdot)$ le support d'une mesure et $\|\cdot\|$ la norme spectrale matricielle. L'opérateur de Toeplitz est défini ainsi : $\forall \mathbf{x}, \mathcal{T} : \mathbf{x} \rightarrow \mathcal{T}(\mathbf{x})$ avec $([\mathcal{T}(\mathbf{x})]_{i,j})_{i \leq j} = x_{i-j}$ et $([\mathcal{T}(\mathbf{x})]_{i,j})_{i > j} = x_{i-j}^*$.

2 Contributions théoriques

2.1 Modèle et hypothèses

Cet article se place dans le cas du régime RMT, *i.e.* $N \rightarrow \infty$, $m \rightarrow \infty$ avec $c_N = \frac{m}{N} \rightarrow c$, $c > 0$. Le modèle choisi est celui de p sources¹ orthogonales altérées par la présence d'un bruit additif CES. Il peut s'écrire pour N observations $\{\mathbf{y}_i\}_{i \in [1,N]}$ de dimension m de la façon suivante :

$$\mathbf{y}_i = \sum_{j=1}^p s_{i,j} \mathbf{m}_j + \sqrt{\tau_i} \mathbf{C}^{1/2} \mathbf{x}_i, \quad i \in [1, N], \quad (1)$$

les \mathbf{m}_j , vecteurs orthogonaux, représentant, pour chaque observation i , le m -vecteur inconnu de la j -ième source déterministe de puissance $s_{i,j}$, les \mathbf{x}_i étant des m -vecteurs aléatoires centrés indépendants, identiquement distribués de loi uniforme sur la m -sphère unité, les τ_1, \dots, τ_N représentant des textures aléatoires (scalaires positifs), et la matrice \mathbf{C} une matrice de covariance hermitienne de Toeplitz définie telle que : $\mathbf{C} = \mathcal{T}((c_0, \dots, c_{m-1})^T)$. Les hypothèses suivantes, notées (H), sont nécessaires pour la preuve du Théorème 1 :

- $\mu_N = \frac{1}{N} \sum_{i=1}^N \delta_{\tau_i}$ vérifie $\int \tau \mu_N(d\tau) \xrightarrow{p.s.} 1$,
- $\frac{1}{N} \sum \delta_{\lambda_i(\mathbf{C})} \xrightarrow{p.s.} \nu$,
- $\max_i d_1(\lambda_i(\mathbf{C}), \text{supp}(\nu)) \rightarrow 0$,
- $\{c_k\}_{k \in [0, m-1]}$ sont des coefficients absolument sommables, et $c_0 \neq 0$.

Soient $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_N]$ la matrice $m \times N$ contenant toutes les observations, $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]$ la matrice $m \times N$ contenant le bruit blanc, \mathbf{T} la matrice $N \times N$ contenant les $\{\tau_i\}_{i \in [1,N]}$ sur sa diagonale et zéro ailleurs, $\mathbf{M} = [\mathbf{m}_1, \dots, \mathbf{m}_p]$ la matrice de mélange de taille $m \times p$, et $(\mathbf{S}^T)_{i,j} = s_{i,j}$ la matrice $N \times p$ représentant la puissance

des sources. Alors, $\mathbf{Y} = \mathbf{M}\mathbf{S} + \mathbf{R}$ avec $\mathbf{R} = \mathbf{C}^{1/2} \mathbf{X} \mathbf{T}^{1/2}$ représentant le bruit additif CES corrélé.

2.2 Blanchiment

Contrairement au cas classique, la matrice \mathbf{C} n'est pas supposée égale à l'identité. Nous proposons dans cette partie d'utiliser un estimateur consistant de la matrice de covariance \mathbf{C} pour blanchir le signal. Dans un premier temps, nous supposons avoir accès à des observations "bruit seul" ne contenant pas les sources². Soit $\check{\mathbf{c}} = (\check{c}_0, \dots, \check{c}_{m-1})^T$ le vecteur constitué des moyennes sur chaque diagonale de la SCM $\mathbf{Y} \mathbf{Y}^H / N = \mathbf{R} \mathbf{R}^H / N$:

$$\check{c}_k = \frac{1}{mN} \sum_{i=1}^m \sum_{j=1}^N y_{i,j} y_{i+k,j}^* \mathbb{1}_{1 \leq i+k \leq m}, \quad (2)$$

où $k \in [0, m-1]$ et $\mathbb{1}_P$ représente la fonction indicatrice sur P .

Théorème 1 (Estimateur consistant de \mathbf{C})

Sous les hypothèses (H), on a

$$\|\mathcal{T}(\check{\mathbf{c}}) - \mathbb{E}[\tau] \mathbf{C}\| \xrightarrow{p.s.} 0. \quad (3)$$

$\check{\mathbf{C}} = \mathcal{T}(\check{\mathbf{c}})$ est, au facteur $\mathbb{E}[\tau]$ près, une estimation consistante de \mathbf{C} .

Preuve la preuve suit celle de [7] et ne sera pas faite ici par manque de place, elle repose sur le Lemme 4.1 de [8] qui établit une inégalité entre la norme l_2 d'une matrice et la transformée en série de Fourier des coefficients de corrélation. La preuve est ensuite menée en découpant en deux parties le terme de gauche de cette inégalité, pour prouver ensuite que chaque partie tend vers zéro. Une fois dans l'espace de Fourier, les étapes de la preuve sont les mêmes que [7], à un facteur près. ■

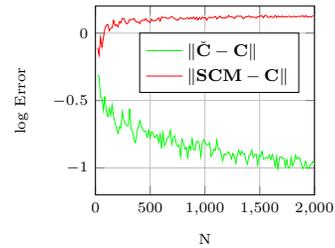


FIGURE 1 – Consistance de $\check{\mathbf{C}} - c = 0.2$ et $\rho = 0.7$ (échelle logarithmique en ordonnée).

La Figure 1 montre la consistance de cet estimateur. Le signal utilisé est un bruit blanc Gaussien \mathbf{X} corrélé avec une matrice de Toeplitz $\mathbf{C} = \mathcal{T}((\rho^0, \rho^1, \dots, \rho^{m-1})^T)$ avec $\rho = 0.7$. La texture $\{\tau_i\}_{i \in [1,N]}$ est constituée d'observations d'une loi inverse gamma de moyenne 1. Le bruit

2. Cette hypothèse peut être relaxée avec certaines hypothèses entre \mathbf{M} et \mathbf{R}

1. p est fixé

résultant suit une loi de Student. Sur cette figure sont tracées les normes spectrales (échelle logarithmique) des différences entre \mathbf{C} et respectivement, l'estimateur proposé $\check{\mathbf{C}}$ en vert et la SCM (habituellement utilisée) en rouge (sur 20 tirages Monte Carlo). La SCM est non consistante. L'estimateur proposé converge lentement vers la matrice de covariance réelle.

2.3 Estimation du rang du sous espace source

Dans cette partie, les observations \mathbf{Y} blanchies par l'estimateur $\check{\mathbf{C}}$, notées \mathbf{Y}_w , sont définies par :

$$\mathbf{Y}_w = \check{\mathbf{C}}^{-1/2} \mathbf{Y} = \check{\mathbf{C}}^{-1/2} \mathbf{M} \mathbf{S} + \mathbf{R}_w, \quad (4)$$

avec $\mathbf{R}_w = \check{\mathbf{C}}^{-1/2} \mathbf{C}^{1/2} \mathbf{X} \mathbf{T}^{1/2}$. Pour estimer le nombre de sources, il est possible d'utiliser la distribution des valeurs propres de n'importe quel M -estimateur appliqué aux observations \mathbf{Y}_w . Ainsi, l'estimateur robuste choisi ici est défini comme l'unique solution, si elle existe, de l'équation suivante :

$$\check{\Sigma} = \frac{1}{N} \sum_{i=1}^N u \left(\frac{1}{m} \mathbf{y}_{wi}^H \check{\Sigma}^{-1} \mathbf{y}_{wi} \right) \mathbf{y}_{wi} \mathbf{y}_{wi}^H, \quad (5)$$

avec $u : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ continue et décroissante [9].

A cause de l'équation implicite précédente, l'étude de la matrice $\check{\Sigma}$ n'est pas aisée quand $N, m \rightarrow \infty$. Il est cependant possible de relier à cette matrice, une matrice $\hat{\mathbf{S}}$ dont les propriétés sont facilement exploitables.

Théorème 2 (Convergence de $\check{\Sigma}$)

$$\left\| \check{\Sigma} - \hat{\mathbf{S}} \right\| \xrightarrow{p.s.} 0, \quad (6)$$

avec $\hat{\mathbf{S}}$ la matrice définie dans le théorème 1 de [10].

Preuve Soit $\check{\Sigma}$ l'unique solution de

$$\check{\Sigma} = \frac{1}{N} \sum_{i=1}^N u \left(\frac{1}{m} \tilde{\mathbf{y}}_i^H \check{\Sigma}^{-1} \tilde{\mathbf{y}}_i \right) \tilde{\mathbf{y}}_i \tilde{\mathbf{y}}_i^H. \quad (7)$$

avec $\tilde{\mathbf{y}}_i = \mathbf{C}^{-1/2} \mathbf{y}_i$ Il peut être facilement montré que

$$\check{\Sigma} = \check{\mathbf{C}}^{-1/2} \mathbf{C}^{1/2} \tilde{\Sigma} \mathbf{C}^{1/2} \check{\mathbf{C}}^{-1/2}. \quad (8)$$

L'équation (6) peut être réécrite de la façon suivante :

$$\left\| \check{\Sigma} - \hat{\mathbf{S}} \right\| \leq \left\| \check{\Sigma} - \tilde{\Sigma} \right\| + \left\| \tilde{\Sigma} - \hat{\mathbf{S}} \right\|. \quad (9)$$

L'article [10] prouve que $\hat{\mathbf{S}}$ est telle que

$$\left\| \tilde{\Sigma} - \hat{\mathbf{S}} \right\| \xrightarrow{p.s.} 0 \quad (10)$$

Avec (8), le premier terme de la partie de droite de (9) peut s'écrire :

$$\left\| \check{\Sigma} - \tilde{\Sigma} \right\| \leq \left\| \check{\mathbf{C}}^{-1/2} \mathbf{C}^{1/2} \tilde{\Sigma} \mathbf{C}^{1/2} \check{\mathbf{C}}^{-1/2} - \tilde{\Sigma} \mathbf{C}^{1/2} \check{\mathbf{C}}^{-1/2} \right\| + \left\| \tilde{\Sigma} \mathbf{C}^{1/2} \check{\mathbf{C}}^{-1/2} - \tilde{\Sigma} \right\|. \quad (11)$$

$$\left\| \check{\Sigma} - \tilde{\Sigma} \right\| \leq \left\| \check{\mathbf{C}}^{-1/2} \mathbf{C}^{1/2} - \mathbf{I}_m \right\| \left\| \tilde{\Sigma} \right\| \left(\left\| \mathbf{C}^{1/2} \check{\mathbf{C}}^{-1/2} \right\| + 1 \right).$$

Comme $\|\mathbf{C}\|$ a un support borné, d'après le Théorème 1, $\|\check{\mathbf{C}}\|$ a un support borné et $\|\mathbf{C} - \check{\mathbf{C}}\| \xrightarrow{p.s.} 0$ puisque $\mathbb{E}[\tau] \xrightarrow{p.s.} 1$ par hypothèse. De plus, $\hat{\mathbf{S}}$ est à support borné (car le nombre de sources p est fixé), l'équation (10) implique que $\tilde{\Sigma}$ est aussi à support borné. Ceci termine la preuve. ■

L'article [10] fixe un seuil t assurant que toutes les valeurs propres de la matrice $\hat{\mathbf{S}}$ au delà de t correspondent à des sources. Grâce au Théorème 2, le même seuil peut être appliqué aux valeurs propres de la matrice $\check{\Sigma}$ afin d'estimer le nombre de sources. Soient $\{\lambda_i(\check{\Sigma})\}_{i=[1,N]}$ les valeurs propres de $\check{\Sigma}$. Comme les sources sont supposées indépendantes, le nombre de sources estimé \hat{p} , c'est à dire le rang du sous-espace des sources, est donné par : $\hat{p} = \min_k (\lambda_k > t)$.

3 Résultats et Simulations

Cette partie présente des expérimentations qui illustrent les résultats précédents sur données simulées et réelles.

3.1 Distribution des valeurs propres

Pour les observations ne contenant pas de sources, le seuil t correspond donc à une borne supérieure pour les valeurs propres de $\check{\Sigma}$. La Figure 2 illustre l'intérêt du Théorème 1. Les observations ne contiennent que du bruit et sont simulées selon une loi de Student centrée, avec une matrice de dispersion de Toeplitz, de paramètres $\rho = 0.7$ et pour $m = 900$, $N = 2000$.

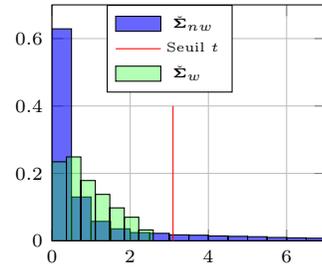


FIGURE 2 – Distribution des valeurs propres de $\check{\Sigma}_{nw}$ pour $m = 900$, $N = 2000$, loi de Marchenko-Pastur et seuil.

L'histogramme bleu représente les valeurs propres de la matrice estimée à partir de l'équation (5) mais sans avoir blanchi les observations. Cette matrice est notée $\check{\Sigma}_{nw}$ (pour *non whitened*). L'histogramme vert représente les valeurs propres de la matrice après blanchissement, notée $\check{\Sigma}_w$. La fonction u est fixée en prenant une distribution inverse gamma d'espérance unité pour les $\tau : u : x \mapsto \frac{1 + \nu}{\nu + x}$ avec $\nu = 0.1$. Le seuil t est alors égal à $\frac{(1 + \nu)(1 + \sqrt{c})^2}{\gamma_m(1 - c(1 + \nu))}$,

γ_m l'unique solution d'une équation au point fixe [10]. On peut voir qu'aucune valeur propre de Σ_w ne s'échappe au delà du seuil t , ce qui n'est pas le cas pour $\hat{\Sigma}_{nw}$, illustrant l'intérêt de l'approche proposée : sans le processus de blanchiment, le Théorème 2 n'est plus respecté.

3.2 Estimation du nombre de sources sur données réelles et simulées

Le bruit est simulé comme précédemment, et le calcul de $\hat{\Sigma}$ se fait sur les données complètes, c'est à dire sur le signal contenant bruit et sources. Sur la Figure 3, le nombre de sources \hat{p} est estimé par trois méthodes : la méthode AIC, M2 la méthode proposée ci-dessus sans l'étape de blanchiment, et M1 la méthode proposée. \hat{p} est la moyenne sur 4 tirages pour chaque SNR. La méthode proposée trouve le bon nombre de sources à partir d'un SNR de 10 dB, alors que les autres méthodes le surestiment.

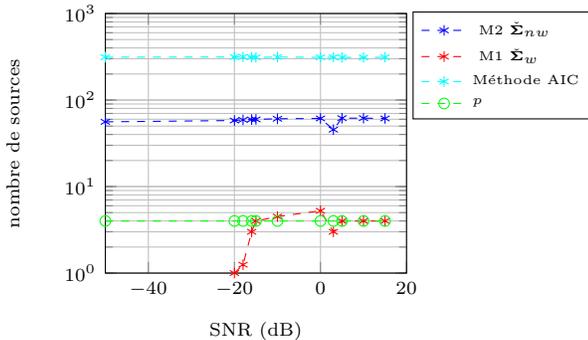


FIGURE 3 – Estimation de p (4 tirages), $m = 400$, $c = 0.2$, $p = 4$ sources, en fonction du SNR.

Enfin, sur des images hyperspectrales en accès libre : *Indian Pines*, *SalinasA* de la banque de données AVIRIS, *PaviaU* de la base de données ROSIS, nous avons testé les trois mêmes différentes méthodes. La fonction $u(\cdot)$ choisie est la même que dans les simulations précédentes. Comme nous n'avons pas accès à la distribution du bruit, nous estimons γ_m de la même façon que dans [10]. Les résultats obtenus sont résumés dans le tableau 1. Sur chaque image, le résultat de la méthode proposée est bien meilleur.

Images	Indian Pines	SalinasA	PaviaU	Cars
p	16	9	9	6
\hat{p} M1	11	9	1	3
\hat{p} M2	220	204	103	1
\hat{p} AIC	219	203	102	143

TABLE 1 – p estimé sur des images hyperspectrales.

4 Conclusion

Pour la sélection d'ordre de modèle, nous avons proposé une méthode d'estimation en deux étapes dans le cas de sources plus bruit corrélé et hétérogène. L'intérêt de

cette approche a été illustrée sur données simulées et a permis d'obtenir des résultats prometteurs sur des images hyperspectrales réelles. Cette méthode peut se généraliser pour d'autres problèmes de sélection d'ordre par exemple dans le domaine du radar. De plus, les résultats théoriques peuvent être étendus au cas d'un blanchiment robuste, i.e. à partir d'un M -estimateur, et sans données d'apprentissage (i.e. des données ne contenant pas de sources).

Remerciement

Les auteurs remercient la DGA pour son support financier.

Références

- [1] V. Marcenko and L. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Mathematics of the USSR-Sbornik*, vol. 1, no. 4, pp. 507–536, 1967.
- [2] H. Akaike, "A new look at the statistical model identification," *Automatic Control, IEEE Transactions on*, vol. 19, no. 6, pp. 716–723, Dec 1974.
- [3] E. Terreaux, J. P. Ovarlez, and F. Pascal, "Anomaly detection and estimation in hyperspectral imaging using random matrix theory tools," in *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2015 IEEE 6th International Workshop on*, Dec 2015, pp. 169–172.
- [4] Z. Bai and J. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices," *The Annals of Probability*, vol. 26, no. 1, pp. 316–345, 1998.
- [5] D. Kelker, "Distribution theory of spherical distributions and a location-scale parameter generalization," *The Indian Journal of Statistics*, pp. 419–430, 1970.
- [6] R. Maronna and V. Yohai, *Robust Estimation of Multivariate Location and Scatter*. John Wiley & Sons, Inc., 2004.
- [7] J. Vinogradova, R. Couillet, and W. Hachem, "Estimation of Toeplitz covariance matrices in large dimensional regime with application to source detection," *Signal Processing, IEEE Transactions on*, vol. 63, no. 18, pp. 4903–4913, Sept 2015.
- [8] R. M. Gray, "Toeplitz and circulant matrices : A review," *Foundations and Trends in Communications and Information Theory*, vol. 2, no. 3, pp. 155–239, 2006.
- [9] M. Mahot, F. Pascal, P. Forster, and J. P. Ovarlez, "Asymptotic properties of robust covariance matrix estimates," *Signal Processing, IEEE Transactions on*, vol. 61, no. 13, pp. 3348–3356, 2013.
- [10] R. Couillet, "Robust spiked random matrices and a robust G-MUSIC estimator," *Journal of Multivariate Analysis*, vol. 140, pp. 139–161, 2015.