Exploitation de la parcimonie par la factorisation de Cholesky et son application pour la détection d'anomalies en imagerie hyperspectrale.

Ahmad W. BITAR[1], Jean-Philippe OVARLEZ[1,2], Loong-Fah CHEONG[3]

[1]SONDRA/CentraleSupélec, Plateau du Moulon, 3 rue Joliot-Curie, F-91190 Gif-sur-Yvette, France

[2]ONERA, DEMR/TSI, Chemin de la Huniere, 91120 Palaiseau, France

[3]Université Nationale de Singapour, Singapour, Singapour

ahmad.bitar@centralesupelec.fr, jeanphilippe.ovarlez@centralesupelec.fr
eleclf@nus.edu.sg

Résumé – L'estimation de la matrice de covariance est une étape essentielle pour la détection de cibles en imagerie hyperspectrale. Exploiter le fait que la matrice de covariance originale (inconnue) est parcimonieuse (*sparse* en anglais) peut potentiellement permettre l'amélioration des performances de détection, surtout pour de grandes dimensions spectrales. Dans ce papier, deux stratégies de parcimonie sont proposées et appliquées sur la matrice de covariance à travers sa matrice unitaire triangulaire inférieure $\mathbf{T}$ (ou facteur de Cholesky). La première permet d'abord d'estimer les entrées de $\mathbf{T}$ par la méthode des moindres carrés ordinaires (*Ordinary Least Squares (OLS)*), et ensuite introduire de la parcimonie en profitant de quelques techniques de seuillages. La deuxième consiste à directement estimer une version parcimonieuse de $\mathbf{T}$ en proposant une version généralisée d'une approche déjà existante et qui pénalise le maximum de vraisemblance logarithmique négatif par la norme L_1. La propriété que la matrice de covariance résultante estimée soit définie positive (et inversible) est toujours garantie. L'efficacité (en termes de détection) des méthodes proposées est évaluée sur des données artificielles pour la détection d'anomalies en imagerie hyperspectrale (HSI).

Abstract – Estimating large covariance matrices is an essential part for hyperspectral target detection. By considering that the true (unknown) covariance matrix is sparse, therefore taking advantage of the possible sparsity in the estimation can potentially improve the detection performance specially in large dimensions. In this paper, two kinds of sparsity are imposed for covariance matrices through their unit lower triangular matrix $\mathbf{T}$ (*Cholesky factor*). The first serves to estimate the entries of $\mathbf{T}$ using such a method as the Ordinary Least Squares (OLS), and then introduce sparsity by exploiting some of the generalized thresholding techniques. The second aims to directly estimate a sparse version of $\mathbf{T}$ by proposing a generalized version of the estimator already developed by penalizing the negative Normal log-likelihood with the L_1-norm penalty function. The resulting estimated covariance matrices are always guaranteed to be positive definite. The effectiveness (in terms of detection) of the proposed estimators is evaluated for artificial hyperspectral anomaly detection.

1 Introduction

L'estimation de la matrice de covariance est une étape importante dans un très grand nombre d'applications. Soit n vecteurs aléatoires $\{\mathbf{x}_i\}_{i\in[1,n]}$ indépendants et identiquement distribués caractérisés par une distribution gaussienne de moyenne nulle et de matrice de covariance inconnue $\mathbf{\Sigma}$. L'estimation traditionelle de $\mathbf{\Sigma}$ est basée sur la matrice de covariance empirique (*Sample Covariance Matrix (SCM)*) définie comme $\hat{\mathbf{\Sigma}}_{SCM} = [\hat{\sigma}_{g,l}]_{p\times p} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\mathbf{x}_i\,\mathbf{x}_i^T$. Cette dernière a deux problèmes majeurs : 1) la contrainte qu'elle soit définie positive et inversible 2) elle n'exploite pas la propriété de parcimonie (car chaque vecteur $\mathbf{x}_i$ pour $i \in [1,n]$ est selectionné de l'image hyperspecrale elle-même et donc n'est jamais parcimonieux). Afin de développer un estimateur toujours défini positif, la méthode de Pourahmadi [1] était la première servant à modéliser la matrice de covariance à partir de la régression linéaire simple. En notant $\hat{\mathbf{x}} = [\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_p]^T \in \mathbb{R}^p$, le principe est de considérer que chaque élement $\hat{x}_t$, $t \in [1,\, p]$, est la version estimée des moindres carrés linéaires de x_t en se basant sur ses $t-1$ prédécessuers $\{x_j\}_{j\in[1,\,t-1]}$. En répétant maintenant la même procédure sur les n vecteurs, $\{\mathbf{x}_i\}_{i\in[1,\,n]}$, la matrice de covariance estimée sera définie par $\hat{\mathbf{\Sigma}} = \hat{\mathbf{T}}^{-1}\,\hat{\mathbf{D}}\,\hat{\mathbf{T}}^{-T}$ qui est toujours définie positive.

Pour exploiter plus avant la parcimonie, plusieurs approches ont été proposées dans la litérature. Dans [2], Huang *et al.* ont proposé d'estimer directement une version parcimonieuse de $\mathbf{T}$ en pénalisant le logarithme du maximum de vraisemblance de la matrice $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n] \in \mathbb{R}^{p\times n}$ par la norme L_1 [3]. Dans [4], la matrice de décomposition des valeurs propres de la matrice de covariance est représentée comme une *sparse matrix transform (SMT)*. La matrice résultante estimée, notée comme $\hat{\mathbf{\Sigma}}_{SMT}$ dans ce papier, est toujours définie positive. Dans [5], Bickel *et al.* ont proposé une version *banded* de $\hat{\mathbf{\Sigma}}_{SCM}$, notée comme $B_m(\hat{\mathbf{\Sigma}}_{SCM}) = [\hat{\sigma}_{g,l}\,\mathbb{1}(|g-l| \le m)]_{g,l}$ dans ce papier, où $\mathbb{1}(.)$ est la fonction indicatrice et $0 \le m < p$. Cependant, ce type de régularization ne garantit pas la propriété que la matrice soit toujours définie positive, principalement dans les applications réelles. Dans [6], quelques opérateurs de seuillage ont été exploités sur les entrées non diagonales de $\hat{\mathbf{\Sigma}}_{SCM}$. Les opérateurs comme *Soft* et *Smoothly Clipped Absolute Deviation* [7] (*SCAD*) appliqués sur $\hat{\mathbf{\Sigma}}_{SCM}$, notés $\hat{\mathbf{\Sigma}}_{SCM}^{Soft}$ et $\hat{\mathbf{\Sigma}}_{SCM}^{SCAD}$ dans ce papier, ont l'avantage de faire conjointement du *shrinkage* et de seuillage et sont capables d'estimer les vrais zéros comme zéros avec une probabilité tendant vers 1. Cependant, ces opérateurs ne garantissent toujours pas la propriété que la matrice soit définie positive.

Le but principal de ce papier est, dans un premier temps, de supposer que la matrice de covariance inconnue est parcimonieuse et d'identifier, ensuite, cette parcimonie à travers son facteur de Cholesky $\mathbf{T}$ afin d'évaluer les performances du détecteur d'anomalies de *Kelly* [8] en imagerie hyperspectrale. Le papier présente une discussion sur deux stratégies différentes exploitant la propriété de parcimonie. La première permet d'exploiter quelques techniques de seuillage déjà existantes comme *Soft* et *SCAD* sur $\hat{\mathbf{T}}$ où les entrées de $\hat{\mathbf{T}}$ sont d'abord estimées par la technique classique *OLS* afin de rendre $\hat{\mathbf{\Sigma}} = \hat{\mathbf{T}}^{-1}\,\hat{\mathbf{D}}\,\hat{\mathbf{T}}^{-T}$ parcimonieuse. La dérnière stratégie consiste à généraliser l'algorithme proposé dans [2].

Le reste du papier est organisé comme suit. La section 2 offre un récapitulatif détaillé sur la méthode de Pourahmadi pour la modélisation de la matrice de covariance à partir de la régression linéaire. Les contributions principales du papier sont présentées dans la section 3. La section 4 montre quelques expérimentations pour une application artificielle de détection d'anomalies en imagerie hyperspectrale (HSI). La section 5 présente la conclusion.

Notation : Une lettre en italique désigne une quantité scalaire. Les caractères gras en minuscule (resp. majuscule) correspondent à des vecteurs (resp. matrices). La notation $(.)^T$ correspond à l'opérateur transposé, tandis que $|.|$, $(.)^{-1}$, $(.)'$, et $\det(.)$ caractérisent la valeur absolue, l'inverse, la dérivée et le déterminant, respectivement. Pour n'importe quelle variable z, on définie $\mathrm{sign}(z) = 1$ si $z > 0$, $\mathrm{sign}(z) = 0$ si $z = 0$ et $\mathrm{sign}(z) = -1$ si $z < 0$.

2 Estimation de la matrice de covariance à partir de la régression linéaire simple

On suppose ici que chaque élément $\hat{x}_t$, $t \in [1, p]$ est une estimation de x_t à partir de ses $t-1$ prédécesseurs $\{x_j\}_{j\in[1,\,t-1]}$ [1]. Ceci conduit à $\hat{x}_t = \sum_{j=1}^{t-1} C_{t,j}\,x_j$, pour $t \in [1,\,p]$. Soit $\epsilon_t = x_t - \hat{x}_t$ l'erreur de prédiction pour $t \in [1, p]$ de variance $E\left[\epsilon_t^2\right] = \theta_t^2$ avec les conditions initiales $\hat{x}_1 = E(x_1) = 0$ et $E\left[x_1^2\right] = \theta_1^2$. On note $\boldsymbol{\epsilon} = \mathbf{T}\,\mathbf{x}$, où $\boldsymbol{\epsilon} = (\epsilon_1, \cdots, \epsilon_p)^T \in \mathbb{R}^p$, $\mathbf{x} = (x_1, \cdots, x_p)^T \in \mathbb{R}^p$, et $\mathbf{T}$ est la matrice unitaire triangulaire inférieure d'éléments $-C_{t,j}$ à la position (t, j) pour $t \in [2,\,p]$ et $j \in [1,\ t-1]$. On obtient alors la matrice de covariance $E\left[\boldsymbol{\epsilon}\,\boldsymbol{\epsilon}^T\right] = \mathbf{T}\,\mathbf{\Sigma}\,\mathbf{T}^T = \mathbf{D}$, où $\mathbf{D}$ est une matrice diagonale composée des entrées $\left\{\theta_t^2\right\}_{t\in[1,\,p]}$. En généralisant la procédure avec une matrice $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_n] \in \mathbb{R}^{p\times n}$, on a :

$$x_{i,t} = \sum_{j=1}^{t-1} C_{t,j}\,x_{i,j} + \epsilon_{i,t}, \qquad t \in [2,\,p], \qquad i \in [1,\,n]\,. \quad (1)$$

En écrivant sous forme matricielle pour $t \in [2,\,p]$, on obtient le modèle de régression linéaire

$$\mathbf{y}_t = \mathbf{A}_{n,t}\,\boldsymbol{\beta}_t + \mathbf{e}_t \quad (2)$$

où $\mathbf{y}_t = (x_{1,t}, \cdots, x_{n,t})^T \in \mathbb{R}^n$, $\mathbf{A}_{n,t} = [x_{i,j}]_{n\times(t-1)}$, $\boldsymbol{\beta}_t = (C_{t,1}, \cdots, C_{t,t-1})^T \in \mathbb{R}^{(t-1)}$, $\mathbf{e}_t = (\epsilon_{1,t}, \cdots, \epsilon_{n,t})^T \in \mathbb{R}^n$. En supposant que $n > p$, l'estimation OLS de $\boldsymbol{\beta}_t$ et sa variance résiduelle correspondante dans $\mathbf{T}$ et $\mathbf{D}$ pour chaque $t \in [2,\,p]$, respectivement. On obtient finalement un deuxième estimateur traditionnel, noté dans ce papier comme $\hat{\mathbf{\Sigma}}_{OLS} = \hat{\mathbf{T}}_{OLS}^{-1}\,\hat{\mathbf{D}}_{OLS}\,\hat{\mathbf{T}}_{OLS}^{-T}$. Notons que $\hat{\mathbf{T}}_{OLS}$ a -$\hat{C}_{t,j}^{OLS}$ dans la position (t, j) pour $t \in [2,\,p]$ et $j \in [1,\,t-1]$.

3 Contributions principales

3.1 Seuillage basé sur le facteur de Cholesky $\mathbf{T}$

On définit l'opérateur du seuillage par $\mathrm{th}(.)$, et on note par $\mathrm{th}(\hat{\mathbf{T}}_{OLS}) = \left[\mathrm{th}\left(-\hat{C}_{t,j}^{OLS}\right)\right]_{p\times p}$ la matrice résultante après application de $\mathrm{th}(.) \in \{\text{Soft, SCAD}\}$ sur chaque élément de la matrice $\hat{\mathbf{T}}^{OLS}$ pour $t \in [2,\,p]$ et $j \in [1,\,t-1]$. On considère alors le problème de minimisation suivant :

$$\mathrm{th}(\hat{\mathbf{T}}_{OLS}) = \underset{\mathbf{T}}{\mathrm{argmin}} \sum_{t=2}^{p}\sum_{j=1}^{t-1}\left\{\frac{1}{2}\left(\hat{C}_{t,j}^{OLS} - C_{t,j}\right)^2 + p_\lambda\left\{|C_{t,j}|\right\}\right\} \quad (3)$$

où $\hat{\mathbf{T}}$ est la matrice unitaire triangulaire unitaire inférieure seuillée et où $p_\lambda \in \{p_\lambda^{L_1}, p_{\lambda,a>2}^{SCAD}\}$ avec $0 \leq \lambda \leq 1$ est défini soit par $p_\lambda^{L_1}(|C_{t,j}|) = \lambda\,|C_{t,j}|$, soit par :

$$p_{\lambda,a>2}^{SCAD}(|C_{t,j}|) = \begin{cases} \lambda\,|C_{t,j}| & |C_{t,j}| \leq \lambda \\ -\frac{|C_{t,j}^2| - 2\,a\,\lambda\,|C_{t,j}| + \lambda^2}{2\,(a-1)} & \lambda < |C_{t,j}| \leq a\,\lambda \\ \frac{(a+1)\lambda^2}{2} & |C_{t,j}| > a\,\lambda \end{cases}$$

Les solutions de (3) avec $p_\lambda^{L_1}$ et $p_{\lambda,a>2}^{SCAD}$ sont chacunes de forme analytique [6], [7]. La valeur a=3.7 recommandée par Fan et Li [7] est utilisée dans la suite de ce papier. On désigne les deux matrices seuillées obtenues par $\hat{\mathbf{T}}_{Soft}$ et $\hat{\mathbf{T}}_{SCAD}$ qui appliquent le seuillage *Soft* et *SCAD* sur $\hat{\mathbf{T}}_{OLS}$, respectivement. On obtient finalement les deux premiers estimateurs :

$$\boxed{\hat{\mathbf{\Sigma}}_{OLS}^{Soft} = \hat{\mathbf{T}}_{Soft}^{-1}\,\hat{\mathbf{D}}_{OLS}\,\hat{\mathbf{T}}_{Soft}^{-T}}$$

$$\boxed{\hat{\mathbf{\Sigma}}_{OLS}^{SCAD} = \hat{\mathbf{T}}_{SCAD}^{-1}\,\hat{\mathbf{D}}_{OLS}\hat{\mathbf{T}}_{SCAD}^{-T}}$$

3.2 Généralisation de l'estimateur dans [2]

On présente le même principe que dans [2] mais en modifiant la procédure avec laquelle les entrées de $\mathbf{T}$ ont été estimées dans [2]. La vraisemblance logarithmique normale négative de

$\mathbf{X}$ est : $\Lambda = -2\log(L(\mathbf{\Sigma}, \mathbf{x}_1, \cdots, \mathbf{x}_n)) = n\log(\det(\mathbf{\Sigma})) + \mathbf{X}^T\,\mathbf{\Sigma}^{-1}\,\mathbf{X} = n\log(\det(\mathbf{\Sigma})) + (\mathbf{T}\,\mathbf{X})^T\,\mathbf{D}^{-1}(\mathbf{T}\,\mathbf{X})$.
En rappelant que $\boldsymbol{\epsilon} = \mathbf{T}\,\mathbf{x}$, en sachant que $\det(\mathbf{T}) = 1$ et que $\mathbf{\Sigma} = \mathbf{T}^{-1}\,\mathbf{D}\,\mathbf{T}^{-T}$, il s'ensuit que $\det(\mathbf{\Sigma}) = \det(\mathbf{D}) = \prod_{t=1}^{p}\theta_t^2$, et que, de fait, $\Lambda = n\sum_{t=1}^{p}\log\theta_t^2 + \sum_{t=1}^{p}\sum_{i=1}^{n}\epsilon_{i,t}^2/\theta_t^2$. En appliquant la fonction pénalité $\sum_{t=2}^{p}\sum_{j=1}^{t-1} p_\alpha\{|C_{t,j}|\}$ à Λ, où $p_\alpha \in \{p_\alpha^{L_1},\, p_{\alpha,a>2}^{SCAD}\}$ (voir la sous section 3.1), on obtient :

$$n\log\theta_1^2 + \sum_{i=1}^{n}\frac{\epsilon_{i,1}^2}{\theta_1^2} + \sum_{t=2}^{p}\left(n\log\,\theta_t^2 + \sum_{i=1}^{n}\frac{\epsilon_{i,t}^2}{\theta_t^2} + \sum_{j=1}^{t-1}p_\alpha\{|C_{t,j}|\}\right), \tag{4}$$

avec $\alpha \in [0, \infty)$. La minimisation de (4) par rapport à θ_1^2 et θ_t^2 conduit respectivement aux solutions $\hat{\theta}_1^2 = \frac{1}{n}\sum_{i=1}^{n}\epsilon_{i,1}^2 = \frac{1}{n}\sum_{i=1}^{n}x_{i,1}^2$ et $\hat{\theta}_t^2 = \frac{1}{n}\sum_{i=1}^{n}\epsilon_{i,t}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_{i,t} - \sum_{j=1}^{t-1}C_{t,j}x_{i,j})^2$.
Il nécessite maintenant à estimer les entrées de $\mathbf{T}$. En minimisant (4) en termes de $\boldsymbol{\beta}_t$, et à partir des equations (1) et (2), on a pour chaque $t \in [2,\, p]$:

$$\begin{aligned}\hat{\boldsymbol{\beta}}_t &= \underset{\boldsymbol{\beta}_t}{\operatorname{argmin}}\sum_{i=1}^{n}\frac{\epsilon_{i,t}^2}{\theta_t^2} + \sum_{j=1}^{t-1}p_\alpha\{|C_{t,j}|\} \\ &= \underset{\boldsymbol{\beta}_t}{\operatorname{argmin}}\frac{1}{\theta_t^2}\sum_{i=1}^{n}\left(x_{i,t} - \sum_{j=1}^{t-1}C_{t,j}x_{i,j}\right)^2 + \sum_{j=1}^{t-1}p_\alpha\{|C_{t,j}|\} \qquad (5) \\ &= \underset{\boldsymbol{\beta}_t}{\operatorname{argmin}}\frac{1}{\theta_t^2}||\mathbf{y}_t - \mathbf{A}_{n,t}\boldsymbol{\beta}_t||_F^2 + \sum_{j=1}^{t-1}p_\alpha\{|C_{t,j}|\}\end{aligned}$$

En notant $l(\boldsymbol{\beta}_t) = \frac{1}{\theta_t^2}||\mathbf{y}_t - \mathbf{A}_{n,t}\boldsymbol{\beta}_t||_F^2$ et $r(\boldsymbol{\beta}_t) = \sum_{j=1}^{t-1}p_\alpha\{|C_{t,j}|\}$, on résout $\boldsymbol{\beta}_t$ itérativement en utilisant l'algorithme *General Iterative Shrinkage and Thresholding (GIST)* [9] :

$$\begin{aligned}\boldsymbol{\beta}_t^{(k+1)} &= \underset{\boldsymbol{\beta}_t}{\operatorname{argmin}}\, l\left(\boldsymbol{\beta}_t^{(k)}\right) + \frac{w^{(k)}}{2}||\boldsymbol{\beta}_t - \boldsymbol{\beta}_t^{(k)}||^2 \\ &\quad + r(\boldsymbol{\beta}_t) + \left(\nabla l(\boldsymbol{\beta}_t^{(k)})\right)^T\left(\boldsymbol{\beta}_t - \boldsymbol{\beta}_t^{(k)}\right) \\ &= \underset{\boldsymbol{\beta}_t}{\operatorname{argmin}}\,\frac{1}{2}\,||\boldsymbol{\beta}_t - \mathbf{u}_t^{(k)}||^2 + \frac{1}{w^{(k)}}r(\boldsymbol{\beta}_t)\,, \qquad (6)\end{aligned}$$

où $\mathbf{u}_t^{(k)} = \boldsymbol{\beta}_t^{(k)} - \nabla l(\boldsymbol{\beta}_t^{(k)})/w^{(k)}$, et où $w^{(k)}$ est le taille du pas (*step size*) initialisée par la règle de *Barzilai-Browein*.
En décomposant (6) en $(t-1)$ problèmes d'optimisation univariés indépendants, on obtient pour $j \in [1, t-1]$:

$$C_{t,j}^{(k+1)} = \underset{C_{t,j}}{\operatorname{argmin}}\,\frac{1}{2}\left(C_{t,j} - u_{t,j}^{(k)}\right)^2 + \frac{1}{w^{(k)}}\,r_j(C_{t,j})\,, \tag{7}$$

où $\mathbf{u}_t^{(k)} = \left(u_{t,1}^{(k)}, \cdots, u_{t,t-1}^{(k)}\right)^T$ et $r(\boldsymbol{\beta}_t) = \sum_{j=1}^{t-1}r_j(C_{t,j})$. Résoudre (7) avec la fonction $p_\alpha^{L_1}$ de pénalité L_1 conduit à la solution analytique suivante :

$$C_{t,j,(L_1)}^{(k+1)} = \operatorname{sign}\left(u_{t,j}^{(k)}\right)\,\max\left(0,\, u_{t,j}^{(k)} - \alpha/w^{(k)}\right)\,. \tag{8}$$

Pour la fonction de pénalité SCAD, $p_{\alpha,a>2}^{SCAD}$, la solution contient en fait trois parties qui correspondent à trois conditions différentes (sous section 3.1). Dans ce cas, et en décomposant le problème (7) en trois sous problèmes de minimisation pour chaque condition, et après avoir résolus, on obtient les trois sous solutions $h_{t,j}^1$, $h_{t,j}^2$, and $h_{t,j}^3$:
$h_{t,j}^1 = \operatorname{sign}\left(u_{t,j}^{(k)}\right)\min\left(\alpha, \max\left(0, |u_{t,j}^{(k)}| - \alpha/w^{(k)}\right)\right)$,
$h_{t,j}^2 = \operatorname{sign}\left(u_{t,j}^{(k)}\right)\min\left(a\,\alpha, \max\left(\alpha, \frac{w^{(k)}|u_{t,j}^{(k)}|\,(a-1) - a\,\alpha}{w^{(k)}(a-2)}\right)\right)$,
$h_{t,j}^3 = \operatorname{sign}\left(u_{t,j}^{(k)}\right)\max\left(a\,\alpha, |u_{t,j}^{(k)}|\right)$.

On obtient alors la solution analytique suivante :

$$\begin{aligned}C_{t,j,(SCAD)}^{(k+1)} &= \underset{q_{t,j}}{\operatorname{argmin}}\,\frac{1}{2}\left(q_{t,j} - u_{t,j}^{(k)}\right)^2 + \frac{1}{w^{(k)}}\,r_j(q_{t,j}) \\ &\qquad \text{s.t.}\quad q_{t,j} \in \{h_{t,j}^1,\, h_{t,j}^2,\, h_{t,j}^3\}\,.\end{aligned} \tag{9}$$

On désigne nos deux seconds estimateurs comme suit :

$$\boxed{\hat{\mathbf{\Sigma}}_{L_1} = \hat{\mathbf{T}}_{L_1}^{-1}\,\hat{\mathbf{D}}\,\hat{\mathbf{T}}_{L_1}^{-T}}$$
$$\boxed{\hat{\mathbf{\Sigma}}_{SCAD} = \hat{\mathbf{T}}_{SCAD}^{-1}\,\hat{\mathbf{D}}\,\hat{\mathbf{T}}_{SCAD}^{-T}}$$

où $\hat{\mathbf{T}}_{L_1}$ et $\hat{\mathbf{T}}_{SCAD}$ contiennent respectivement $-\hat{C}_{t,j,(L_1)}$ et $-\hat{C}_{t,j,(SCAD)}$ dans la position (t,j) pour $t \in [2,\, p]$ et $j \in [1,\, t-1]$, tandis que $\hat{\mathbf{D}}$ contient $(\hat{\theta}_1^2,\, \hat{\theta}_t^2)$ sur la diagonale. Notons que dans [2], les auteurs ont utilisé l'approximation Locale Quadratique Linéaire (LQA) de la norme L_1 afin d'obtenir une solution analytique pour $\boldsymbol{\beta}_t$ dans (5). L'algorithme proposé est maintenant plus général car il exploite l'algorithme de *GIST* pour résoudre (6). Il peut être facilement être étendu à d'autres fonctions de pénalités comme *SCAD*, *Capped-L1*, *Log Sum*, etc. et toutes possèdent des solutions analytiques [9]. Ce papier présente seulement les fonctions de pénalités L_1 et SCAD.

4 La détection d'anomalies

Une image hyperspectrale est constituée d'une série d'images de la même scène spatiale, mais prises dans plusieurs dizaines de longueurs d'onde contiguës et très étroites (10-20 nm), qui correspondent à autant de "couleurs". Lorsque la dimension spectrale est très grande, la détection de cibles devient délicate et caractérise une des applications les plus importantes de l'imagerie hyperspectrale [10]. Une information préalable sur la signature spectrale de la cible n'est souvent pas disponible pour l'utilisateur. On parle alors du problème de la détection d'anomalies, une anomalie caractérisant tout ce qui pourrait être différent du fond. Considérons maintenant le modèle du signal suivant :

$$\begin{cases} H_0 : \mathbf{x} = \mathbf{n}, & \mathbf{x}_i = \mathbf{n}_i, \quad i = 1, \cdots, n \\ H_1 : \mathbf{x} = \gamma\,\mathbf{d} + \mathbf{n}, & \mathbf{x}_i = \mathbf{n}_i, \quad i = 1, \cdots, n \end{cases} \tag{10}$$

où $\mathbf{n}_1, \cdots, \mathbf{n}_n$ sont n vecteurs i.i.d. caractérisant les données *secondaires* et distribués selon la loi normale $\mathcal{N}(\mathbf{0}_p, \mathbf{\Sigma})$. Le vecteur $\mathbf{d}$ caractérise l'information spectrale "inconnue" de l'anomalie à détecter et $\gamma > 0$ désigne son amplitude, également inconnue. Le détecteur du Maximum de Vraisemblance généralisé

HSIs	$\mathbf{\Sigma_{TRUE}}$	$\hat{\mathbf{\Sigma}}_{SCM}$	$\hat{\mathbf{\Sigma}}_{OLS}$	$\hat{\mathbf{\Sigma}}_{OLS}^{Soft}$	$\hat{\mathbf{\Sigma}}_{OLS}^{SCAD}$	$\hat{\mathbf{\Sigma}}_{L_1}$	$\hat{\mathbf{\Sigma}}_{SCAD}$	$\hat{\mathbf{\Sigma}}_{SMT}$	$B_k(\hat{\mathbf{\Sigma}}_{SCM})$	$\hat{\mathbf{\Sigma}}_{Soft}^{SCM}$	$\hat{\mathbf{\Sigma}}_{SCAD}^{SCM}$
Modèle 1	0.9541	0.7976	0.8331	0.9480	0.9480	**0.9509**	**0.9509**	0.9503	**0.9509**	**0.9509**	**0.9509**
Modèle 2	0.9540	0.7977	0.8361	0.9124	0.9124	0.9264	0.9264	0.9184	**0.9478**	0.9274	0.9270
Modèle 3	0.9541	0.7978	0.8259	0.8169	0.8257	0.8236	**0.8261**	0.7798	0.5321	0.5969	0.5781

Table 1. Une liste des valeurs AUC pour les estimateurs $\hat{\mathbf{\Sigma}}_{OLS}^{Soft}$, $\hat{\mathbf{\Sigma}}_{OLS}^{SCAD}$, $\hat{\mathbf{\Sigma}}_{L_1}$, $\hat{\mathbf{\Sigma}}_{SCAD}$ comparées avec des autres.

(GLRT) optimal pour le test d'hypothèses présenté constitue le détecteur de *Kelly* [8] et est décrit par :

$$D_{KellyAD\hat{\mathbf{\Sigma}}}(\mathbf{x}) = \mathbf{x}^T \hat{\mathbf{\Sigma}}_{SCM}^{-1} \mathbf{x} \underset{H_0}{\overset{H_1}{\gtrless}} \delta \,, \qquad (11)$$

où δ est un seuil de détection fixée par la probabilité de Fausse Alarme. La matrice $\hat{\mathbf{\Sigma}}_{SCM}$ peut être remplacée par un autre estimateur. On parle alors de *two-step* GLRT.

4.1 Application sur des données artificielles

Les expérimentations suivantes sont effectuées sur trois modèles de matrices de covariance :

— Modèle 1 : $\mathbf{\Sigma}$ = I, la matrice identité,
— Modèle 2 : le modèle autorégressif d'ordre 1, AR(1), $\mathbf{\Sigma} = [\sigma_{gl}]_{p\times p}$, où $\sigma_{gl} = c^{|g-l|}$, pour $c = 0.3$,
— Modèle 3 : $\mathbf{\Sigma} = [\sigma_{gl}]_{p\times p}$, où $\sigma_{gl} = (1 - ((|g-l|)/r))_+$, pour $r = p/2$: la matrice triangulaire.

Le Modèle 1 est très *sparse*, le Modèle 2 est relativement *sparse* tandis que le Modèle 3 avec $r = p/2$ peut-être considéré le moins *sparse* des trois modèles considérés [6].

Les performances du détecteur de Kelly construit avec les estimateurs $\hat{\mathbf{\Sigma}}_{OLS}^{Soft}$, $\hat{\mathbf{\Sigma}}_{OLS}^{SCAD}$, $\hat{\mathbf{\Sigma}}_{L_1}$ et $\hat{\mathbf{\Sigma}}_{SCAD}$ sont évaluées dans le tableau 1 par les valeurs de l'aire sous la courbe Pd-Pfa (en Anglais "*Area Under Curve (AUC) values"*) pour un rapport signal sur bruit égal à 15 dB. Les estimateurs utilisés pour la comparaison sont : $\hat{\mathbf{\Sigma}}_{SCM}$, $\hat{\mathbf{\Sigma}}_{OLS}$, $\hat{\mathbf{\Sigma}}_{SMT}$ [4], $B_k(\hat{\mathbf{\Sigma}}_{SCM})$ [5], $\hat{\mathbf{\Sigma}}_{SCM}^{Soft}$ [6], et $\hat{\mathbf{\Sigma}}_{SCM}^{SCAD}$ [6]. Une procédure de validation logarithmique croisée [2] est mise en œuvre pour fixer les paramètres λ (section 3.1) et α (section 3.2).

Toutes les valeurs dans le tableau 1 sont calculées à partir de 10^5 tirages de Monte-Carlo. On choisit $n = 80$ données secondaires pour l'estimation de la matrice de covariance sous l'hypothèse Gaussienne, et le nombre de bandes spectrales p = 60. L'anomalie artificielle considérée est un vecteur contenant des nombres pseudo-aléatoires normalement distribués (le même vecteur est utilisé pour les trois modèles).

Les plus grandes valeurs de AUC pour chaque modèle apparaissent en gras dans le tableau 1. Les résultats obtenus démontrent que pour les modèles parcimonieux (Modele 1 et 2), nos estimateurs $\{\hat{\mathbf{\Sigma}}_{OLS}^{Soft}, \hat{\mathbf{\Sigma}}_{OLS}^{SCAD}, \hat{\mathbf{\Sigma}}_{L_1}, \hat{\mathbf{\Sigma}}_{SCAD}\}$ améliorent la détection par rapport aux estimateurs classiques $\{\hat{\mathbf{\Sigma}}_{SCM}, \hat{\mathbf{\Sigma}}_{OLS}\}$ d'une manière significative, et ayant des résultats compétitifs à l'état de l'art $\{B_k(\hat{\mathbf{\Sigma}}_{SCM}), \hat{\mathbf{\Sigma}}_{SCM}^{Soft}, \hat{\mathbf{\Sigma}}_{SCM}^{SCAD}\}$.

Le résultat le plus important est que même si le modèle n'est pas parcimonieux (Modèle 3), nos estimateurs ne détériorent pas la dtection par rapport aux estimateurs traditionnels. Cependant, bien que $\hat{\mathbf{\Sigma}}_{OLS}^{Soft}$, $\hat{\mathbf{\Sigma}}_{OLS}^{SCAD}$ et $\hat{\mathbf{\Sigma}}_{L_1}$, ont une très légère détérioration de détection par rapport à $\hat{\mathbf{\Sigma}}_{OLS}$, cette détérioration est encore acceptable et donc on peut l'ignorer. D'un autre côté, on remarque que $\{B_k(\hat{\mathbf{\Sigma}}_{SCM}), \hat{\mathbf{\Sigma}}_{SCM}^{Soft}, \hat{\mathbf{\Sigma}}_{SCM}^{SCAD}\}$ présentent une très grande détérioration de détection par rapport aux estimateurs traditionnels, et donc ceci démontre le principal avantage de nos estimateurs par rapport à l'état de l'art.

5 CONCLUSION

La parcimonie est imposée sur la matrice de covariance à travers son facteur de Cholesky $\mathbf{T}$. Une application artificielle pour trois modèles de matrices démontre l'efficacité des méthodes proposées pour la détection d'anomalies en imagerie hyperspectrale.

Références

[1] M. Pourahmadi, "Joint mean-covariance models with applications to longitudinal data : unconstrained parameterisation," *Biometrika*, vol. 86, no. 3, pp. 677–690, 1999.

[2] J. Z. Huang, N. Liu, M. Pourahmadi, and L. Liu, "Covariance matrix selection and estimation via penalised normal likelihood," *Biometrika*, vol. 93, no. 1, pp. 85–98, 2006.

[3] R. Tibshirani, "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[4] G. Cao and C. Bouman, "Covariance estimation for high dimensional data vectors using the sparse matrix transform," in *Advances in Neural Information Processing Systems 21*. Curran Associates, Inc., 2009, pp. 225–232.

[5] P. J. Bickel and E. Levina, "Regularized estimation of large covariance matrices," *The Annals of Statistics*, vol. 36, no. 1, pp. 199–227, 2008.

[6] A. J. Rothman, E. Levina, and J. Zhu, "Generalized thresholding of large covariance matrices," *Journal of the American Statistical Association*, vol. 104, no. 485, pp. 177–186, 2009.

[7] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American Statistical Association*, vol. 96, no. 456, pp. 1348–1360, 2001.

[8] E. J. Kelly, "An adaptive detection algorithm," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 23, no. 1, pp. 115–127, November 1986.

[9] P. Gong, C. Zhang, Z. Lu, J. Huang, and J. Ye, "Gist : General iterative shrinkage and thresholding for non-convex sparse learning," 2013.

[10] D. Manolakis, D. Marden, and G. Shaw, "Hyperspectral image processing for automatic target detection applications," *Lincoln Laboratory Journal*, vol. 14, no. 1, pp. 79–116, 2003.