

# Etude à distance finie d'un $R$ -estimateur pour la matrice de forme des distributions symétriques elliptiques complexes

Stefano FORTUNATI<sup>1,2</sup>, Alexandre RENAUX<sup>1</sup>, Frédéric PASCAL<sup>1</sup>,

<sup>1</sup>Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des Signaux et Systèmes, 91190, Gif-sur-Yvette, France

<sup>2</sup>DR2I-IPSA, 94200, Ivry sur Seine, France

stefano.fortunati@centralesupelec.fr, alexandre.renaux@universite-paris-saclay.fr,  
frederic.pascal@centralesupelec.fr

**Résumé** – Cette communication propose une analyse, à partir de simulations, des principales propriétés d'un  $R$ -estimateur de la matrice de forme d'un vecteur symétrique elliptique complexe à distance finie. D'abord proposé par Hallin, Oja et Paindaveine pour le cas des vecteurs réels puis récemment étendu aux cas de vecteurs complexes par nos soins, cet estimateur est à la fois robuste du point de vue de la distribution et efficace du point de vue semi-paramétrique. Après une présentation théorique de cet estimateur, nous étudions ici, à distance finie et pour différentes configurations possibles, son erreur quadratique moyenne que nous comparons à la borne de Cramér-Rao semi-paramétrique contrainte.

**Abstract** – This paper aims at presenting a simulative analysis of the main properties of a new  $R$ -estimator of shape matrices in Complex Elliptically Symmetric distributed observations. First proposed by Hallin, Oja and Paindaveine for the real-valued case and then extended to the complex field in our recent work, this  $R$ -estimator has the remarkable property to be, at the same time, *distributionally robust* and *semiparametric efficient*. Here, the efficiency of different possible configurations of this  $R$ -estimator are investigated by comparing the resulting Mean Square Error (MSE) with the Constrained Semiparametric Cramér-Rao Bound.

## 1 Introduction

Le problème de l'estimation d'une matrice de covariance ou d'une matrice de forme à partir d'un jeu d'observations représente une étape cruciale dans les applications de traitement du signal ou d'apprentissage statistique telles que la détection ou le partitionnement de données (clustering). Parmi les différents modèles de distributions non-gaussiens, la famille des distributions symétriques elliptiques complexes est connue pour fournir une modélisation générale et fiable dans une grande variété de scénarios [1].

Un ensemble de  $L$  vecteurs i.i.d. distribués elliptiquement  $\mathbb{C}^N \ni \mathbf{z}_l \sim CES_N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, h_0)$ ,  $l = 1, \dots, L$ , est complètement caractérisé par un paramètre de position  $\boldsymbol{\mu}_0$ , une matrice de dispersion  $\boldsymbol{\Sigma}_0$  et une fonction dite génératrice de distribution  $h_0 : \mathbb{R}_0^+ \rightarrow \mathbb{R}^+$  qui joue généralement le rôle d'une fonction de nuisance. En effet, les procédures d'inférence pour les données elliptiques consistent généralement à estimer de manière conjointe  $\boldsymbol{\mu}_0$  et  $\boldsymbol{\Sigma}_0$  en présence d'une fonction génératrice de distribution inconnue  $h_0$ . De manière remarquable, l'ajout de ce paramètre de nuisance de dimension infinie nous place naturellement dans le cadre des modèles semi-paramétriques. Remarquons également que, du fait de l'ambiguïté d'échelle bien connue entre  $\boldsymbol{\Sigma}_0$  et  $h_0$ , seulement une version à l'échelle de la matrice de dispersion, appelée *matrice de forme*, est identifiable. Nous considérons donc ici la matrice de forme  $\mathbf{V}_{1,0} \triangleq \boldsymbol{\Sigma}_0 / [\boldsymbol{\Sigma}_0]_{1,1}$  comme paramètre d'intérêt en lieu et place de la

matrice de dispersion (non-contrainte)  $\boldsymbol{\Sigma}_0$ .

Comme récemment mentionné dans la littérature des communautés issues des statistiques [2–5] et du traitement du signal [6–8], la nature *semi-paramétrique* des distributions elliptiques a permis d'établir à la fois des bornes d'information semi-paramétriques ainsi que des algorithmes d'estimation robustes capables de traiter le manque de connaissances *a priori* sur la fonction génératrice de distribution  $h_0$ . Une famille classique d'estimateurs robustes pour les données modélisées par des distributions elliptiques est basée sur les  $M$ -estimateurs qui incluent les estimateurs de Huber et de Tyler comme cas particuliers. Les deux avantages principaux de ces  $M$ -estimateurs de la matrice de forme sont : *i*) leurs performances ne se dégradent pratiquement pas dans le cas d'un scénario non-gaussien et *ii*) ils sont  $\sqrt{L}$ -consistents quelle que soit la fonction inconnue  $h_0$ . Néanmoins, ces estimateurs souffrent d'un inconvénient majeur qui est leur manque d'efficacité du point de vue semi-paramétrique comme montré dans [6, 7].

Dans leur article fondateur [3], basé sur la théorie de la normalité asymptotique locale de Le Cam ainsi que sur les propriétés d'invariance des statistiques de rang, Hallin, Oja et Paindaveine ont montré qu'il était possible d'établir un estimateur de la matrice de forme capable de réconcilier les deux propriétés dichotomiques que sont *robustesse* et *efficacité semi-paramétrique*. Cet estimateur, établi dans [3] pour des distributions symétriques réelles appartient à la classe des  $R$ -estimateurs. Dans un travail récent [8], nous avons présenté de manière tu-

torielle le calcul de cet estimateur ainsi que son extension aux cas des vecteurs complexes.

Le but de cette communication est de valider, pour le cas complexe, les résultats théoriques obtenus dans [8] en proposant une étude des propriétés statistiques de cet estimateur. Plus particulièrement, nous analysons ces performances à distance finie pour différents scénarios et nous adressons la question cruciale de son efficacité semi-paramétrique en comparant son erreur quadratique moyenne (EQM) à la borne de Cramér-Rao semi-paramétrique calculée dans [6, 7].

*Notations* : nous utiliserons ici les notations introduites dans [8] dont nous ne rappelons que les plus importantes à propos des opérateurs et des matrices spéciales utilisés tout au long de cette communication.  $\text{vec}$  indique l'opérateur qui concatène les vecteurs colonnes d'une matrice  $N \times N$  pour en faire un vecteur colonne  $\text{vec}(\mathbf{A})$  de taille  $N^2$ . L'opérateur  $\text{vec}(\mathbf{A})$  définit un vecteur de taille  $N^2 - 1$  obtenu en supprimant le premier élément de  $\text{vec}(\mathbf{A})$ . En d'autres mots,  $\text{vec}(\mathbf{A}) \triangleq [a_{11}, \text{vec}(\mathbf{A})^T]^T$ . Une matrice  $\mathbf{A}$  telle que  $[\mathbf{A}]_{1,1} \triangleq 1$  est notée  $\mathbf{A}_1$ . Nous introduisons également les deux matrices suivantes :

$$\Pi_{\text{vec}(\mathbf{I}_N)}^\perp = \mathbf{I}_{N^2} - N^{-1} \text{vec}(\mathbf{I}_N) \text{vec}(\mathbf{I}_N)^T, \quad (1)$$

$$\mathbf{P} \triangleq [\mathbf{e}_2 | \mathbf{e}_3 | \dots | \mathbf{e}_{N^2}]^T, \quad (2)$$

où  $\mathbf{e}_i$  est le  $i$ -ème vecteur de la base canonique de  $\mathbb{R}^{N^2}$ . Pour des raisons d'interprétabilité et de consistance avec la littérature existante, toutes les simulations présentent les versions renormalisées de chaque estimateur considéré :

$$\widehat{\mathbf{V}}_\gamma^\varphi \triangleq N \widehat{\mathbf{V}}_{1,\gamma}^\varphi / \text{tr}(\widehat{\mathbf{V}}_{1,\gamma}^\varphi), \quad (3)$$

où  $\gamma$  et  $\varphi$  indiquent l'estimateur considéré. Comme indice de performance, nous utiliserons l'EQM :

$$\varsigma_\gamma^\varphi \triangleq \|E\{\text{vec}(\widehat{\mathbf{V}}_\gamma^\varphi - \mathbf{V}_0) \text{vec}(\widehat{\mathbf{V}}_\gamma^\varphi - \mathbf{V}_0)^H\}\|_F, \quad (4)$$

tandis que pour les bornes de performances nous utiliserons [6, 7] :

$$\varepsilon_{\text{CSCRB}} \triangleq \|[\text{CSCRB}(\boldsymbol{\Sigma}_0, h_0)]\|_F. \quad (5)$$

De plus,  $\mathbf{V}_0 = N \boldsymbol{\Sigma}_0 / \text{tr}(\boldsymbol{\Sigma}_0)$  et, enfin,  $\boldsymbol{\Sigma}_0$  et  $h_0$  représentent les vraies valeurs de la matrice de dispersion et de la fonction génératrice de distribution, respectivement. Dans toutes les simulations, nous utiliserons le scénario suivant :

- $\boldsymbol{\Sigma}_0$  est une matrice de Toeplitz hermitienne dont la première colonne est donnée par  $[1, \rho, \dots, \rho^{N-1}]^T$ ;  $\rho = 0.8e^{j2\pi/5}$  et  $N = 8$ .
- Les données sont générées en utilisant une  $t$ -distribution complexe à moyenne nulle  $P_Z(\mathbf{z} | \boldsymbol{\Sigma}_0, h_0)$  dont la densité de probabilité est :

$$p_Z(\mathbf{z} | \boldsymbol{\Sigma}_0, h_0) = |\boldsymbol{\Sigma}_0|^{-1} h_0 (\mathbf{z}^H \boldsymbol{\Sigma}_0^{-1} \mathbf{z}), \quad \text{et} \quad (6)$$

$$h_0(t) = \frac{\Gamma(\lambda + N)}{\pi^N \Gamma(\lambda)} \left(\frac{\lambda}{\eta}\right)^\lambda \left(\frac{\lambda}{\eta} + t\right)^{-(\lambda + N)}, \quad (7)$$

où  $\lambda \in (1, +\infty)$  est un paramètre de forme qui contrôle la queue de distribution, tandis que  $\eta$  est un paramètre d'échelle qui rend compte de la puissance des données  $\sigma^2$ . Plus précisément, sous hypothèse de moments du second ordre finis, nous avons  $\sigma^2 = \lambda / (\eta(\lambda - 1))$ . Nous avons choisi  $\sigma^2 = 4$ .

— Le nombre de Monte Carlo est de  $10^6$ .

On notera ici que le choix spécifique d'une  $t$ -distribution complexe pour les observations ne représente pas de limitation particulière puisque, du fait de la nature semi-paramétrique de notre  $R$ -estimateur, les résultats obtenus pour cette distribution resteront valides pour n'importe quelle autre distribution.

## 2 Un estimateur semi-paramétrique efficace

Dans cette section, on rappelle, d'un point de vue algorithmique, le  $R$ -estimateur introduit dans [8] pour le cas complexe. Le lecteur trouvera également dans cet article une étude détaillée de ses propriétés asymptotiques.

Soit  $\mathbf{z}_l \sim \text{CES}_N(\mathbf{0}, \mathbf{V}_{1,0}, h_0)$ ,  $l = 1, \dots, L$  un ensemble d'observations elliptiques complexes i.i.d.. Un  $R$ -estimateur robuste et efficace du point de vue semi-paramétrique de  $\mathbf{V}_{1,0}$  est donnée par :

$$\text{vec}(\widehat{\mathbf{V}}_{1,R}) = \text{vec}(\widehat{\mathbf{V}}_1^*) + L^{-1/2} \widehat{\mathbf{Y}}^{-1} \widetilde{\boldsymbol{\Delta}}_{\widehat{\mathbf{V}}_1^*}, \quad (8)$$

où  $\widehat{\mathbf{V}}_1^*$  est un estimateur préliminaire  $\sqrt{L}$ -consistant de  $\mathbf{V}_{1,0}$ . La matrice  $\widehat{\mathbf{Y}}$  et le vecteur  $\widetilde{\boldsymbol{\Delta}}_{\widehat{\mathbf{V}}_1^*}$  sont définis respectivement par :

$$\widehat{\mathbf{Y}} \triangleq \hat{\alpha} \mathbf{L}_{\widehat{\mathbf{V}}_1^*} \mathbf{L}_{\widehat{\mathbf{V}}_1^*}^H, \quad (9)$$

$$\widetilde{\boldsymbol{\Delta}}_{\widehat{\mathbf{V}}_1^*} \triangleq L^{-1/2} \mathbf{L}_{\mathbf{V}_1^*} \sum_{l=1}^L K\left(\frac{r_l^*}{L+1}\right) \text{vec}(\hat{\mathbf{u}}_l^* (\hat{\mathbf{u}}_l^*)^H) \quad (10)$$

et le scalaire  $\hat{\alpha}$  peut être obtenu par :

$$\hat{\alpha} = \|\widetilde{\boldsymbol{\Delta}}_{\widehat{\mathbf{V}}_1^* + L^{-1/2} \mathbf{H}^0} - \widetilde{\boldsymbol{\Delta}}_{\widehat{\mathbf{V}}_1^*}\| / \|\mathbf{L}_{\widehat{\mathbf{V}}_1^*} \mathbf{L}_{\widehat{\mathbf{V}}_1^*}^H \text{vec}(\mathbf{H}^0)\|, \quad (11)$$

où  $\mathbf{H}^0$  est une matrice Hermitienne de "petite perturbation", telle que  $[\mathbf{H}^0]_{1,1} = 0$ . Comme dans [8], on pose  $\mathbf{H}^0 = (\mathbf{G} + \mathbf{G}^H)/2$  où  $[\mathbf{G}]_{i,j} \sim \mathcal{CN}(0, v^2)$ ,  $[\mathbf{G}]_{1,1} = 0$  et  $v = 0.01$ .

La fonction  $K : (0, 1) \rightarrow \mathbb{R}^+$  est appelée *fonction score* dont nous précisons les propriétés à la section 4.

Les autres termes utilisés dans la définition du  $R$ -estimateur (8) sont les suivants [8] :

- $\hat{Q}_l^* \triangleq \mathbf{z}_l^H [\widehat{\mathbf{V}}_1^*]^{-1} \mathbf{z}_l$ ,
- $\hat{\mathbf{u}}_l^* \triangleq (\hat{Q}_l^*)^{-1/2} [\widehat{\mathbf{V}}_1^*]^{-1/2} \mathbf{z}_l$ ,
- $r_1^*, \dots, r_L^*$  sont les rangs des variables aléatoires réelles continues  $\hat{Q}_1^*, \dots, \hat{Q}_L^*$ ,
- $\mathbf{L}_{\widehat{\mathbf{V}}_1^*} \triangleq \mathbf{P} \left( [\widehat{\mathbf{V}}_1^*]^{-T/2} \otimes [\widehat{\mathbf{V}}_1^*]^{-1/2} \right) \Pi_{\text{vec}(\mathbf{I}_N)}^\perp$ .

On peut noter que, pour l'implémentation pratique du  $R$ -estimateur (8), seulement deux termes ont besoin d'être spécifiés : l'estimateur préliminaire  $\widehat{\mathbf{V}}_1^*$  et la fonction score  $K \in \mathcal{K}$ . Nous discutons à présent de ces choix.

## 3 Le choix de $\widehat{\mathbf{V}}_1^*$

Nous étudions ici l'impact de l'estimateur préliminaire  $\widehat{\mathbf{V}}_1^*$  sur les performances à distance finie du  $R$ -estimateur (8). En

effet, bien qu'en théorie n'importe quel choix d'estimateur préliminaire  $\sqrt{L}$ -consistant conduit à un  $R$ -estimateur *asymptotiquement* efficace du point de vue semi-paramétrique, il n'en est pas de même pour les performances à distance finie. Nous étudions ici deux estimateurs préliminaires : la matrice de covariance empirique et l'estimateur de Tyler.

### 3.1 La matrice de covariance empirique comme estimateur préliminaire

Soit  $\{\mathbf{z}_l\}_{l=1}^L \sim CES_N(\mathbf{0}, \mathbf{V}_{1,0}, h_0)$  un ensemble i.i.d. de vecteurs aléatoires elliptiquement distribués avec une fonction génératrice de distribution  $h_0$ . La matrice de covariance empirique  $\widehat{\mathbf{V}}_{1,SCM}^*$  est donnée par :

$$\widehat{\mathbf{V}}_{1,SCM}^* = \frac{\widehat{\Sigma}_{SCM}}{[\widehat{\Sigma}_{SCM}]_{1,1}}, \quad \widehat{\Sigma}_{SCM} \triangleq \frac{1}{L} \sum_{l=1}^L \mathbf{z}_l \mathbf{z}_l^H. \quad (12)$$

En supposant les moments du second ordre finis,  $\widehat{\mathbf{V}}_{1,SCM}^*$  est un estimateur  $\sqrt{L}$ -consistant de la matrice de forme  $\mathbf{V}_{1,0}$  quelle que soit la fonction génératrice de distribution, est peut donc être choisi comme estimateur préliminaire. L'estimateur de la matrice de covariance empirique est connu pour être d'une complexité calculatoire intéressante mais souffre de performances faibles en environnement non-gaussien. Dans la figure 1a, nous traçons l'EQM en fonction du nombre d'observation  $L$  pour  $\widehat{\mathbf{V}}_{SCM}^*$  issu de (12) ainsi que du  $R$ -estimateur de (8) qui utilise  $\widehat{\mathbf{V}}_{SCM}^*$  comme estimateur préliminaire. Notons que les deux estimateurs sont re-normalisés comme dans (3). Comme fonction score, nous avons utilisé le score de *van der Waerden* [8] :<sup>1</sup>

$$K_{vdW}(u) \triangleq \Phi_G^{-1}(u), \quad (13)$$

où  $\Phi_G^{-1}$  est la réciproque de la fonction de répartition d'une variable aléatoire Gamma de paramètres  $(N, 1)$ . Le  $R$ -estimateur ainsi obtenu sera noté  $\widehat{\mathbf{V}}_{R,vdW}^{SCM}$ . Comme nous pouvons le remarquer à partir de la figure 1a, le terme de correction linéaire "à une étape"  $L^{-1/2} \widehat{\mathbf{Y}}^{-1} \widehat{\Delta}_{\widehat{\mathbf{V}}_{1,SCM}^*}$  permet d'augmenter de manière significative l'efficacité de l'estimateur de la covariance empirique sans augmenter le coût calculatoire. En effet, cette correction peut être calculée analytiquement et ne requiert pas d'utiliser un algorithme itératif de point fixe comme, par exemple, pour l'utilisation d'un  $M$ -estimateur.

### 3.2 L'estimateur de Tyler comme estimateur préliminaire

Le résultat de la figure 1a a été obtenu en posant comme paramètre de forme  $\lambda = 2$  pour nos observations  $t$ -distribués. Il est donc également intéressant de vérifier l'efficacité semi-paramétrique du  $R$ -estimateur (8) en fonction de  $\lambda$ , c'est-à-dire en fonction de la non-gaussianité des données (et ceci pour un nombre fini  $L$  de données). Puisque l'estimateur de la covariance empirique se dégrade dans le cas non-gaussien, on

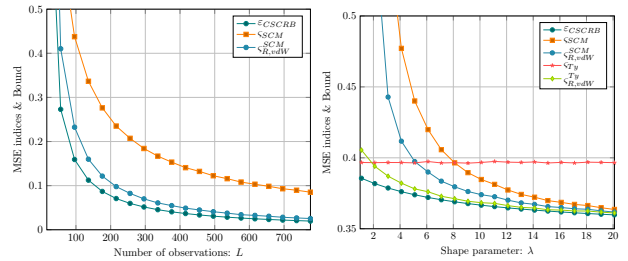
peut espérer que l'utilisation d'un  $M$ -estimateur, par exemple l'estimateur de Tyler, comme estimateur préliminaire permettra d'améliorer les performances à distance finie. L'estimateur de Tyler  $\widehat{\mathbf{V}}_{Ty}$  est obtenu comme le point de convergence de l'algorithme de point fixe suivant :

$$\widehat{\mathbf{V}}_{Ty}^{(k+1)} = \frac{N}{L} \sum_{l=1}^L \frac{\mathbf{z}_l \mathbf{z}_l^H}{\mathbf{z}_l^H [\widehat{\mathbf{V}}_{Ty}^{(k)}]^{-1} \mathbf{z}_l}, \quad (14)$$

où l'initialisation est, par exemple,  $\mathbf{V}_{Ty}^{(0)} = \mathbf{I}_N$ . Afin d'obtenir un  $R$ -estimateur normalisé dans (8), nous imposons la contrainte suivante classique sur le premier élément :  $\widehat{\mathbf{V}}_{1,Ty}^* = \widehat{\mathbf{V}}_{Ty} / [\widehat{\mathbf{V}}_{Ty}]_{1,1}$ .

Après re-normalisation dans (3), nous avons tracé dans la figure 1b l'EQM des estimateurs de la covariance empirique et de Tyler  $\widehat{\mathbf{V}}_{SCM}^*$  et  $\widehat{\mathbf{V}}_{Ty}^*$ , ainsi que l'EQM des  $R$ -estimateurs correspondant, c'est-à-dire  $\widehat{\mathbf{V}}_{R,vdW}^{SCM}$  et  $\widehat{\mathbf{V}}_{R,vdW}^{Ty}$ . Comme pour la figure précédente, nous avons utilisé le score de *van der Waerden*  $K_{vdW}$  dans (13). Notons que le nombre d'observation est fixé à  $L = 5N$  afin de garantir un régime à distance fini.

Comme espéré,  $\widehat{\mathbf{V}}_{R,vdW}^{Ty}$  donne de meilleures performances que  $\widehat{\mathbf{V}}_{R,vdW}^{SCM}$  en présence de données non-gaussiennes (c'est-à-dire pour les petites valeurs de  $\lambda$ ). Ceci est sans nul doute due aux bonnes propriétés de robustesse de l'estimateur de Tyler. Bien sûr, cette amélioration des performance induit un coût calculatoire plus important du à l'algorithme de point fixe utilisé pour obtenir  $\widehat{\mathbf{V}}_{Ty}$  dans l'équation (14). De plus on remarque que  $\widehat{\mathbf{V}}_{R,vdW}^{Ty}$  est (quasiment) efficace quelle que soit la valeur de  $\lambda$  et ce, même à distance finie. L'EQM de  $\widehat{\mathbf{V}}_{R,vdW}^{Ty}$  atteint la borne de Cramér-Rao semi-paramétrique (contrainte) à partir de  $\lambda > 6$ . Evidement, dans le régime asymptotique ( $L \rightarrow \infty$ ), cette propriété d'efficacité est atteinte pour de plus petite valeur de  $\lambda$ . Enfin, il est intéressant de remarquer que le  $R$ -estimateur  $\widehat{\mathbf{V}}_{R,vdW}^{Ty}$  obtient de meilleures performances que l'estimateur de Tyler pour presque toute les valeurs de  $\lambda$  tout en gardant de bonne propriétés de robustesse et ce, pour un coût calculatoire négligeable (par rapport à Tyler).



(a) Matrice de covariance empirique pour  $\widehat{\mathbf{V}}_1^*$  (b) Estimateur de Tyler pour  $\widehat{\mathbf{V}}_1^*$

FIGURE 1 – EQM et BCR en fonction de  $\lambda$  ( $L = 5N$ ).

1. Le choix de la fonction score est discuté dans la prochaine section.

## 4 Sur le choix de la fonction score $K$

Dans le contexte des statistiques de rang, le terme de *fonction score* indique une fonction continue  $K : (0, 1) \rightarrow \mathbb{R}^+$  qui satisfait les deux propriétés suivantes : *i*)  $K$  est de carré sommable et *ii*)  $K$  peut être exprimé comme la différence de deux fonctions monotones croissantes [2]. Un exemple classique de fonction score est l'ensemble défini par  $K_a(u) = N(a+1)u^a$  où  $u \in (0, 1)$  et où  $a \geq 0$  est un paramètre de réglage [9]. Les cas où  $(a = 1)$  (score de Wilcoxon) et  $(a = 2)$  (score de Spearman) sont les plus connus. Une autre manière de construire une fonction score est celle décrite dans [2], [3] et [8] où une approche à base de modèle "mal spécifié" est utilisée. Nous nous limiterons ici à deux exemples qui sont la fonction score de *van der Waerden* introduite dans l'équation (13) et la fonction score  $t_\nu$  donnée par :

$$K_{t_\nu}(u) = \frac{N(2N + \nu)F_{2N,\nu}^{-1}(u)}{\nu + 2NF_{2N,\nu}^{-1}(u)}, \quad u \in (0, 1), \quad (15)$$

où  $F_{2N,\nu}(u)$  est la fonction de répartition de Fisher avec  $2N$  et  $\nu \in (0, \infty)$  degrés de liberté. On notera que d'après les propriétés de la distribution de Fisher, nous avons :  $\lim_{\nu \rightarrow \infty} K_{t_\nu}(u) = K_{vdW}(u)$ . Dans la figure 2, après re-normalisation (3), l'EQM de quatre  $R$ -estimateurs utilisant les fonctions scores de *Wilcoxon* ( $\hat{V}_{R,W_i}^{Ty}$ ), de *Spearman* ( $\hat{V}_{R,S_p}^{Ty}$ ),  $t_\nu$ - ( $\hat{V}_{R,Ct_\nu}^{Ty}$ ) et de *van der Waerden* ( $\hat{V}_{R,vdW}^{Ty}$ ) sont reportées et comparées à la borne de Cramér-Rao semi-paramétrique (contrainte). Concernant le score  $t_\nu$ , un paramètre de réglage  $\nu = 5$  a été choisi. Enfin, quel que soit le  $R$ -estimateur considéré, nous avons utilisé l'estimateur de Tyler comme estimateur préliminaire. On remarque sur la figure 2 que, pour  $\lambda > 6$ , le score de *van der Waerden* conduit à l'EQM la plus basse. De plus, concernant les scores de *Wilcoxon* et de *Spearman*, l'EQM augmente lorsque le paramètre de réglage  $a$  augmente. Cependant, à notre connaissance, ce résultat est pour le moment un problème ouvert du point de vue théorique. Le score  $t_\nu$  obtient les meilleures performances pour des données hautement non-gaussienne ( $1 < \lambda < 5$ ). Pour conclure sur cette partie, nous pouvons dire que le score de *van der Waerden* est une fonction acceptable puisque qu'elle conduit à une quasi efficacité semi-paramétrique et ne dépend pas d'un paramètre de réglage supplémentaire.

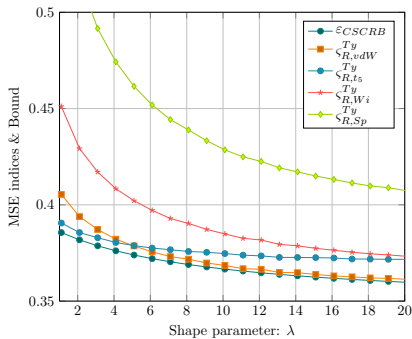


FIGURE 2 – EQM et BCR en fonction de  $\lambda$  ( $L = 5N$ ).

## 5 Conclusions

Nous avons étudié les performances, à distance finie, d'un  $R$ -estimateur de la matrice de forme dans des données elliptiquement distribuées pour différents scénarios. Cette étude numérique a montré que l'utilisation de l'estimateur de Tyler comme estimateur préliminaire ainsi que du score *van der Waerden* conduisent à de meilleures performances en terme d'EQM (par rapport aux  $M$ -estimateurs) dans pratiquement toutes les situations. Les perspectives envisagées concernent l'étude de ce  $R$ -estimateur du point de vue de la robustesse.

## Références

- [1] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, "Complex elliptically symmetric distributions : Survey, new results and applications," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5597–5625, 2012.
- [2] M. Hallin and D. Paindaveine, "Semiparametrically efficient rank-based inference for shape I. optimal rank-based tests for sphericity," *The Annals of Statistics*, vol. 34, no. 6, pp. 2707–2756, 2006.
- [3] M. Hallin, H. Oja, and D. Paindaveine, "Semiparametrically efficient rank-based inference for shape II. optimal  $R$ -estimation of shape," *The Annals of Statistics*, vol. 34, no. 6, pp. 2757–2789, 2006.
- [4] M. Hallin and D. Paindaveine, "Parametric and semiparametric inference for shape : the role of the scale functional," *Statistics & Decisions*, vol. 24, no. 3, pp. 327–350, 2009.
- [5] D. Paindaveine, "A canonical definition of shape," *Statistics & Probability Letters*, vol. 78, no. 14, pp. 2240 – 2247, 2008.
- [6] S. Fortunati, F. Gini, M. S. Greco, A. M. Zoubir, and M. Rangaswamy, "Semiparametric inference and lower bounds for real elliptically symmetric distributions," *IEEE Transactions on Signal Processing*, vol. 67, no. 1, pp. 164–177, Jan 2019.
- [7] S. Fortunati, F. Gini, M. S. Greco, A. M. Zoubir, and M. Rangaswamy, "Semiparametric CRB and Slepian-Bangs formulas for complex elliptically symmetric distributions," *IEEE Transactions on Signal Processing*, vol. 67, no. 20, pp. 5352–5364, Oct 2019.
- [8] S. Fortunati, A. Renaux, and F. Pascal, "Robust semiparametric efficient estimators in complex elliptically symmetric distributions," *IEEE Transactions on Signal Processing*, vol. 68, pp. 5003–5015, 2020.
- [9] M. Hallin, D. Paindaveine, and T. Verdebout, "Optimal rank-based testing for principal components," *The Annals of Statistics*, vol. 38, no. 6, pp. 3245–3299, 12 2010.