

# On the asymptotics of Maronna's robust PCA

Gordana Drašković, *Student Member, IEEE*, Arnaud Breloy, *Member, IEEE* and Frédéric Pascal, *Senior Member, IEEE*

**Abstract**—The eigenvalue decomposition (EVD) parameters of the second order statistics are ubiquitous in statistical analysis and signal processing. Notably, the EVD of robust scatter  $M$ -estimators is a popular choice to perform robust probabilistic PCA or other dimension reduction related applications. Towards the goal of characterizing the behavior of these quantities, this paper proposes new asymptotics for the EVD parameters (i.e. eigenvalues, eigenvectors and principal subspace) of the scatter  $M$ -estimator in the context of complex elliptically symmetric distributions. First, their Gaussian asymptotic distribution is obtained by extending standard results on the sample covariance matrix in a Gaussian context. Second, their convergence rate towards the EVD parameters of a Gaussian-Core Wishart Equivalent is derived. This second result represents the main contribution in the sense that it quantifies when it is acceptable to directly plug-in well-established results on the EVD of Wishart-distributed matrix for characterizing the EVD of  $M$ -estimators. Eventually, some examples (low-rank adaptive filtering and Intrinsic bias analysis) are provided to illustrate where the obtained results can be leveraged.

## I. INTRODUCTION

The second order statistics plays a key role in signal processing and machine learning applications. Usually, this parameter is unknown and has to be estimated in order to apply a so-called adaptive process. In this scope, the  $M$ -estimators of the scatter [1], [2] have attracted a lot of interest [3]–[8] due to their robustness properties over the family of Complex Elliptically Symmetric (CES) distributions [5]. They notably offer robustness to outliers and heavy tailed samples (now common in modern datasets), where the traditional Sample Covariance Matrix (SCM) usually fails to provide an accurate estimation.

The statistical characterization of the  $M$ -estimators of the scatter is a complex issue since they are defined by fixed-point equations. While the SCM in a Gaussian setting follows a well-known Wishart distribution [9], the true distribution of the  $M$ -estimators remains unknown. Several works derived asymptotic characterization for these estimators. Their asymptotic Gaussian distribution was derived in [10] and extended to the complex case in [5], [11]. Probably approximately correct (PAC) error bounds have been studied in [12]. Their analysis in the large random matrix regime (i.e. when both the number of samples and the dimension tends to infinity at the same rate) has been established in [13], [14]. Recently, [15], [16] showed that their distribution can be very accurately approximated by a Wishart one of an equivalent Gaussian core model referred to as Gaussian Core Wishart Equivalent (GCWE).

Gordana Drašković and Frédéric Pascal are with L2S - CentraleSupélec - CNRS - Université Paris-Sud - 3 rue Joliot-Curie, F-91192 Gif-sur-Yvette Cedex, France (e-mails: gordana.draskovic@l2s.centralesupelec.fr, frederic.pascal@l2s.centralesupelec.fr), Arnaud Breloy is with LEME - EA4416, University Paris-Nanterre, France (e-mail: abreloy@parisnanterre.fr).

Additionally, the eigenvalue decomposition (EVD) of  $M$ -estimators is required in numerous processes. Indeed, the eigenvectors of the scatter matrix are involved in probabilistic PCA algorithms [17], [18], as well as in the derivation of robust counterparts of low rank filters or detectors [19], [20]. The eigenvalues of the scatter are used in model order selection [21], [22], functions of eigenvalues are involved in various applications such as regularization parameter selection [6], [23], detection [24], and classification [25]. Hence accurately characterizing the distribution of the  $M$ -estimators EVD represents an interest, both from the points of view of performance analysis and optimal process design. In this paper, we derive new asymptotic characterizations for the EVD parameters of scatter  $M$ -estimators in the general context of CES-distributed samples. For the eigenvalues, the eigenvectors and the principal subspace (i.e. the subspace spanned by the  $r$  strongest eigenvectors), we derive:

- The standard Gaussian asymptotic distribution. This result is obtained by extending the analysis of [26] (for the SCM) and perturbation analysis of [27], [28] to the complex  $M$ -estimators. This asymptotic analysis provides an extension of [17], [29] since it gives the information about the covariance between the eigenvalues of an  $M$ -estimator and provides the exact structure of the asymptotic covariance and pseudo-covariance matrix of principal subspace. Also, contrary to [17], [29], all the results in this paper are derived for complex data.
- The convergence rate towards the EVD parameters of a GCWE by extending the results of [15], [16]. This result represents the main contribution in the sense that it quantifies when it is acceptable to directly plug-in well established results on the EVD of Wishart-distributed matrices for characterizing the EVD of  $M$ -estimators [9], [30].

In the last part, we eventually give some examples where the proposed results can be leveraged. Concerning the eigenvectors and principal subspaces, we derive the performance of Low rank filters [31] build from  $M$ -estimators. Regarding the eigenvalues, we address the complex issue of characterizing the intrinsic bias [32] of  $M$ -estimators in CES distribution. This quantity has been studied in [32] for the SCM in a Gaussian context thanks to the distribution of the eigenvalues of a Wishart matrix [9]. Extending this analysis to  $M$ -estimators represents, at first sight, an intractable problem. However, our proposed GCWE allows to derive an accurate approximation of this quantity.

The rest of this paper is organized as follows. Section II introduces the CES distributions and  $M$ -estimators. Section III contains the main results about eigenvalue decomposition

of  $M$ -estimators. In Section IV, we introduce LR standard models and present main results about principal subspaces of  $M$ -estimators. In Section V Monte Carlo simulations are presented in order to validate the theoretical results. In addition, examples of applications of the results are presented. Finally, some conclusions and perspectives are drawn in Section VI.

*Notations* - Vectors (resp. matrices) are denoted by bold-faced lowercase letters (resp. uppercase letters).  $T$ ,  $*$ ,  $H$  and  $+$  respectively represent the transpose, conjugate, Hermitian operator and pseudo-inverse of a matrix. i.i.d. stands for “independent and identically distributed”, w.r.t. for “with respect to” and  $\sim$  means “is distributed as”.  $\stackrel{d}{=}$  stands for “shares the same distribution as”,  $\stackrel{d}{\rightarrow}$  denotes convergence in distribution and  $\otimes$  denotes the Kronecker product.  $\text{vec}(\cdot)$  is the operator which transforms a matrix  $p \times n$  into a vector of length  $pn$ , concatenating its  $n$  columns into a single column. Moreover,  $\mathbf{I}_p$  is the  $p \times p$  identity matrix,  $\mathbf{0}$  the matrix of zeros with appropriate dimension and  $\mathbf{K}$  is the commutation matrix (square matrix with appropriate dimensions) which transforms  $\text{vec}(\mathbf{A})$  into  $\text{vec}(\mathbf{A}^T)$ , i.e.  $\mathbf{K} \text{vec}(\mathbf{A}) = \text{vec}(\mathbf{A}^T)$ .  $\mathcal{H}_M^{++}$  is the set of Hermitian positive definite matrices. The set of semi-unitary matrices is denoted as  $\mathcal{U}_r^p = \{\mathbf{U} \in \mathbb{C}^{p \times r}, \mathbf{U}^H \mathbf{U} = \mathbf{I}_r\}$ . Finally,  $\mathcal{GCN}(\mathbf{0}, \mathbf{V}, \mathbf{W})$  denotes the zero-mean non-circular improper complex normal distribution with covariance matrix  $\mathbf{V}$  and pseudo-covariance matrix  $\mathbf{W}$  [5].

## II. BACKGROUND

### A. CES distributions

Complex Elliptically Symmetric (CES) distributions form a general family of circular multivariate distributions [5]. The probability density functions (PDFs) of a CES distribution can be written as

$$f_{\mathbf{z}}(\mathbf{z}) = C |\Sigma|^{-1} g_{\mathbf{z}}((\mathbf{z} - \boldsymbol{\mu})^H \Sigma^{-1} (\mathbf{z} - \boldsymbol{\mu})) \quad (1)$$

where  $C$  is a normalisation constant and  $g_{\mathbf{z}} : [0, \infty) \rightarrow [0, \infty)$  is any function (called the density generator) ensuring Eq. (1) defines a PDF. These CES distributions will be denoted by  $\mathcal{CES}(\boldsymbol{\mu}, \Sigma, g_{\mathbf{z}})$ . The Complex Normal (Gaussian) distribution is a particular case of CES distributions in which  $g_{\mathbf{z}}(z) = e^{-z}$  and  $C = \pi^{-p}$ . We denote this distribution as  $\mathbf{z} \sim \mathcal{CN}(\boldsymbol{\mu}, \Sigma)$  (see [5] and Section V for more examples of CES distributions).

*Gaussian-cores representation* In [16] the so-called Gaussian-core model has been used as an alternative to the classic stochastic representation [33] of CES-distributed vectors. A random vector  $\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \Sigma, g_{\mathbf{z}})$  can be represented as

$$\mathbf{z} \stackrel{d}{=} \frac{\sqrt{\mathcal{Q}}}{\|\mathbf{g}\|} \mathbf{A} \mathbf{g} \quad (2)$$

where  $\Sigma = \mathbf{A} \mathbf{A}^H$  is a factorization of  $\Sigma$  and  $\mathbf{g} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ .  $\mathcal{Q}$  is a non-negative real random variable, called the modular variate, independent of  $\mathbf{g}$  with a PDF depending only on  $g_{\mathbf{z}}$ . We refer to  $\mathbf{x} = \mathbf{A} \mathbf{g}$  as the Gaussian-core of  $\mathbf{z}$ .

### B. $M$ -estimators and SCM

Let  $(\mathbf{z}_1, \dots, \mathbf{z}_n)$  be an  $n$ -sample of  $p$ -dimensional complex i.i.d. vectors with  $\mathbf{z}_i \sim \mathcal{CES}(\mathbf{0}, \Sigma, g_{\mathbf{z}})$ . An  $M$ -estimator, denoted by  $\hat{\Sigma}$ , is defined by the solution of the following  $M$ -estimating equation

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n u(\mathbf{z}_i^H \hat{\Sigma}^{-1} \mathbf{z}_i) \mathbf{z}_i \mathbf{z}_i^H \quad (3)$$

where  $u$  is any real-valued weight function on  $[0, \infty)$  that respects Maronna’s conditions (ensuring existence and uniqueness) [1]. The theoretical scatter matrix  $M$ -functional is defined as a solution of

$$\mathbb{E}[u(\mathbf{z}^H \Sigma_{\sigma}^{-1} \mathbf{z}) \mathbf{z} \mathbf{z}^H] = \Sigma_{\sigma}. \quad (4)$$

The  $M$ -functional is proportional to the true scatter matrix parameter  $\Sigma$  as  $\Sigma_{\sigma} = \sigma^{-1} \Sigma$ , where the scalar factor  $\sigma > 0$  can be found by solving

$$\mathbb{E}[\Psi(\sigma t)] = p \quad (5)$$

with  $\Psi(\sigma t) = u(\sigma t) \sigma t$  and  $t = \mathbf{z}^H \hat{\Sigma}^{-1} \mathbf{z}$ .

The sample covariance matrix (SCM) [34] is given by

$$\hat{\Sigma}_{\text{SCM}} = \frac{1}{n} \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^H. \quad (6)$$

The SCM can be considered as a “limited case” of Eq. (3) when  $u(\mathbf{z}_i^H \hat{\Sigma}^{-1} \mathbf{z}_i) = 1$ . This estimator is usually used when the data is assumed to be Gaussian-distributed, since it is the Maximum Likelihood Estimator (MLE) in that case. Note that for the SCM, (3) becomes explicit which makes this estimator very convenient for statistical analysis. Indeed, for  $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \Sigma)$ , it follows a Wishart distribution with well-known properties [9]. However, since the SCM is not robust, it can perform very poorly in comparison to  $M$ -estimators in CES framework.

### C. Standard Asymptotic Regime

Let  $(\mathbf{z}_1, \dots, \mathbf{z}_n)$  be an  $n$ -sample of  $p$ -dimensional complex independent vectors with  $\mathbf{z}_i \sim (\mathbf{0}, \Sigma, g_{\mathbf{z}})$ ,  $i = 1, \dots, n$ . We consider the complex  $M$ -estimator  $\hat{\Sigma}$  that verifies Eq. (3) and follows Maronna’s conditions [1], and we denote  $\Sigma_{\sigma}$  the solution of Eq. (4).

**Theorem II.1** *The asymptotic distribution of  $\hat{\Sigma}$  is given by [5], [11] as*

$$\sqrt{n} \text{vec}(\hat{\Sigma} - \Sigma_{\sigma}) \stackrel{d}{\rightarrow} \mathcal{GCN}(\mathbf{0}, \mathbf{C}, \mathbf{P})$$

where the asymptotic covariance and pseudo-covariance matrices are

$$\begin{cases} \mathbf{C} = \vartheta_1 \Sigma_{\sigma}^T \otimes \Sigma_{\sigma} + \vartheta_2 \text{vec}(\Sigma_{\sigma}) \text{vec}(\Sigma_{\sigma})^H, \\ \mathbf{P} = \vartheta_1 (\Sigma_{\sigma}^T \otimes \Sigma_{\sigma}) \mathbf{K} + \vartheta_2 \text{vec}(\Sigma_{\sigma}) \text{vec}(\Sigma_{\sigma})^T. \end{cases} \quad (7)$$

The constants  $\vartheta_1 > 0$  and  $\vartheta_2 > -\vartheta_1/p$  are given by

$$\begin{aligned} \vartheta_1 &= c_M^{-2} a_M p(p+1) \\ \vartheta_2 &= (c_M - p^2)^{-2} (a_M - p^2) - c_M^{-2} a_M (p+1) \end{aligned} \quad (8)$$

where

$$\begin{aligned} a_M &= E[\Psi^2(\sigma\mathcal{Q})] \\ c_M &= E[\Psi'(\sigma\mathcal{Q})\sigma\mathcal{Q}] + p^2. \end{aligned}$$

with  $\mathcal{Q}$  defined in Eq. (2).

**Remark II.1** Note that for the SCM built with Gaussian-distributed data  $\vartheta_1 = 1$  and  $\vartheta_2 = 0$  with  $\Sigma_\sigma = \Sigma$ .

#### D. Gaussian-Core Wishart Equivalent (GCWE)

The asymptotic distribution of the difference between an  $M$ -estimator and the SCM built with Gaussian-cores of CES data (Eq. (2)), has been recently derived in [16].

*Assumed Gaussian-core model:* Let us assume  $n$  measurements  $(\mathbf{z}_1, \dots, \mathbf{z}_n)$  where  $\mathbf{z}_i = \sqrt{\mathcal{Q}_i}/\|\mathbf{g}_i\| \mathbf{A}\mathbf{g}_i$  following  $\mathcal{CES}(\mathbf{0}, \Sigma, g_{\mathbf{z}})$ ,  $i = 1, \dots, n$ , where

- $\hat{\Sigma}$  is an  $M$ -estimator built with  $(\mathbf{z}_1, \dots, \mathbf{z}_n)$  using Eq. (3),
- $\hat{\Sigma}_{\text{GCWE}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^H$  is the SCM built with Gaussian-cores  $\mathbf{x}_i = \mathbf{A}\mathbf{g}_i \sim \mathcal{CN}(\mathbf{0}, \Sigma)$  of  $\mathbf{z}_i$ ,  $i = 1, \dots, n$ , given by Eq. (2) that represent only fictive data used for theoretical purposes.

Hereafter, we always consider the same model.

**Theorem II.2** Let  $\sigma$  be the solution of Eq. (5). Then, the asymptotic distribution of  $\sigma\hat{\Sigma} - \hat{\Sigma}_{\text{GCWE}}$  is given by [16]

$$\sqrt{n} \text{vec}(\sigma\hat{\Sigma} - \hat{\Sigma}_{\text{GCWE}}) \xrightarrow{d} \mathcal{GCN}(\mathbf{0}, \tilde{\mathbf{C}}, \tilde{\mathbf{P}}) \quad (9)$$

where  $\tilde{\mathbf{C}}$  and  $\tilde{\mathbf{P}}$  are defined by

$$\begin{aligned} \tilde{\mathbf{C}} &= \sigma_1 \Sigma^T \otimes \Sigma + \sigma_2 \text{vec}(\Sigma) \text{vec}(\Sigma)^H, \\ \tilde{\mathbf{P}} &= \sigma_1 (\Sigma^T \otimes \Sigma) \mathbf{K} + \sigma_2 \text{vec}(\Sigma) \text{vec}(\Sigma)^T \end{aligned} \quad (10)$$

with  $\sigma_1$  and  $\sigma_2$  given by

$$\begin{aligned} \sigma_1 &= \frac{ap(p+1) + c(c-2b)}{c^2} \\ \sigma_2 &= \frac{a-p^2}{(c-p^2)^2} - \frac{a(p+1)}{c^2} + 2 \frac{p(c-b)}{c(c-p^2)} \end{aligned} \quad (11)$$

where

$$\begin{aligned} a &= E[\Psi^2(\sigma\mathcal{Q})] \\ b &= E[\Psi(\sigma\mathcal{Q})\|\mathbf{g}\|^2] \\ c &= E[\Psi'(\sigma\mathcal{Q})\sigma\mathcal{Q}] + p^2. \end{aligned}$$

An important note is that these factors are much smaller than the ones in the regular asymptotic regime (Eq. (8)) meaning that the behavior of an  $M$ -estimator can be accurately approximated with the behavior of the corresponding Wishart-distributed matrix.

### III. ASYMPTOTICS OF $M$ -ESTIMATORS' EIGENVALUE DECOMPOSITION

The EigenValue Decomposition (EVD) of a (scatter) matrix  $\Sigma$  is defined as

$$\Sigma \stackrel{\text{EVD}}{=} \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H \quad (12)$$

with

$$\mathbf{U}^H \mathbf{U} = \mathbf{I} \quad (13)$$

where  $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_p] \in \mathcal{U}_p^p$  and  $\mathbf{\Lambda} = \text{diag}(\boldsymbol{\lambda})$ ,  $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_p] \in \mathbb{R}^p$ . In the following we assume ordered eigenvalues  $\lambda_1 > \dots > \lambda_p > 0$ .

We define the operators  $\boldsymbol{\theta}_j$  and  $\phi_j$  returning respectively the  $j^{\text{th}}$  eigenvector and eigenvalue as

$$\begin{cases} \mathbf{u}_j = \boldsymbol{\theta}_j(\Sigma), \\ \lambda_j = \phi_j(\Sigma). \end{cases} \quad (14)$$

Let us again assume the Gaussian-core model proposed in Section II-D and

$$\begin{cases} \hat{\mathbf{u}}_j^M = \boldsymbol{\theta}_j(\hat{\Sigma}) & \hat{\mathbf{u}}_j^{\text{GCWE}} = \boldsymbol{\theta}_j(\hat{\Sigma}_{\text{GCWE}}), \\ \hat{\lambda}_j^M = \phi_j(\hat{\Sigma}) & \hat{\lambda}_j^{\text{GCWE}} = \phi_j(\hat{\Sigma}_{\text{GCWE}}). \end{cases} \quad (15)$$

with  $\phi = [\phi_1, \dots, \phi_p]$ . In the following we derive the asymptotic distribution for these quantities.

**Theorem III.1 (Standard asymptotic regime)** Let  $\hat{\Sigma}$  be a scatter  $M$ -estimator with  $\hat{\Sigma} \stackrel{\text{EVD}}{=} \hat{\mathbf{U}} \text{diag}(\hat{\boldsymbol{\lambda}}) \hat{\mathbf{U}}^H$ . The asymptotic distribution of the eigenvalues and eigenvectors is characterized by

$$\begin{cases} \sqrt{n}(\sigma\hat{\boldsymbol{\lambda}}^M - \boldsymbol{\lambda}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \vartheta_1 \mathbf{\Lambda}^2 + \vartheta_2 \boldsymbol{\lambda} \boldsymbol{\lambda}^T), \\ \sqrt{n} \mathbf{\Pi}_j^\perp \hat{\mathbf{u}}_j^M \xrightarrow{d} \mathcal{CN}(\mathbf{0}, \Xi_j). \end{cases} \quad (16)$$

where

$$\Xi_j = \vartheta_1 \lambda_j \mathbf{U} \mathbf{\Lambda} (\lambda_j \mathbf{I} - \mathbf{\Lambda})^{+2} \mathbf{U}^H \quad (17)$$

with  $\mathbf{\Pi}_j^\perp = \mathbf{I} - \mathbf{u}_j \mathbf{u}_j^H$  and  $\vartheta_1, \vartheta_2$  given by Eq. (8).

*Proof:* See Appendix A. ■

**Theorem III.2 (GCWE)** Asymptotic distribution of the difference between the eigenvalues and eigenvectors of an  $M$ -estimators and GCWE is given by

$$\begin{cases} \sqrt{n}(\sigma\hat{\boldsymbol{\lambda}}^M - \hat{\boldsymbol{\lambda}}^{\text{GCWE}}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma_1 \mathbf{\Lambda}^2 + \sigma_2 \boldsymbol{\lambda} \boldsymbol{\lambda}^T), \\ \sqrt{n} \mathbf{\Pi}_j^\perp (\hat{\mathbf{u}}_j^M - \hat{\mathbf{u}}_j^{\text{GCWE}}) \xrightarrow{d} \mathcal{CN}(\mathbf{0}, \sigma_1 / \vartheta_1 \Xi_j). \end{cases} \quad (18)$$

with  $\Xi_j$  and  $\sigma_1, \sigma_2$  given by Eqs. (17) and (11), respectively.

*Proof:* See Appendix B. ■

**Remark III.1** • The results given in Theorem III.1 are interesting since, besides the variance of each eigenvalue, they provide the correlation between them. Note that for a Wishart-distributed matrix this correlation is equal to zero, as shown in [26] for real case. Conversely, we showed that the eigenvalues of an  $M$ -estimator are asymptotically correlated, as stated in [17] (but not

explicitly characterized). This correlation depends on the second scale parameter  $\vartheta_2$ . Concerning the eigenvectors, note that the covariance depends only on  $\vartheta_1$  since  $\theta_j$  is scale invariant w.r.t. to the covariance matrix (see [11] for more details).

- Theorem III.2 characterizes the asymptotic variance of the EVD of an  $M$ -estimator compared to the one of its GCWE. It shows that their covariance structure is the same as for the standard asymptotic regime. However the scale  $\sigma_1$  is much smaller than  $\vartheta_1$ , especially when  $p$  increases. Therefore, the GCWE provides a better asymptotic characterization of the  $M$ -estimator's EVD.

#### IV. ASYMPTOTICS OF $M$ -ESTIMATORS' PRINCIPAL SUBSPACE

Consider the special case of low-rank plus identity (also referred to as factor model)

$$\Sigma = \Sigma_r + \gamma^2 \mathbf{I}_p \quad (19)$$

with

$$\Sigma_r = \mathbf{U}_r \mathbf{\Lambda}_r \mathbf{U}_r^H \quad (20)$$

with  $\mathbf{U}_r \in \mathcal{U}_r^p$  and  $\mathbf{\Lambda} \in \mathbb{C}^{r \times r}$ . The principal subspace is defined as

$$\Pi_r = \mathbf{U}_r \mathbf{U}_r^H. \quad (21)$$

Let us consider an  $M$ -estimator built with  $(\mathbf{z}_1, \dots, \mathbf{z}_n)$  where  $\mathbf{z}_i \sim \mathcal{CES}(\mathbf{0}, \Sigma_r + \gamma^2 \mathbf{I}_p)$  and the GCWE built with fictive data  $(\mathbf{x}_1, \dots, \mathbf{x}_n)$  given by Eq. (2) where  $\mathbf{x}_i \sim \mathcal{CN}(\mathbf{0}, \Sigma_r + \gamma^2 \mathbf{I}_p)$ ,  $i = 1, \dots, n$ . Assume then that  $\hat{\mathbf{U}}_r^M$  is the estimate of  $\mathbf{U}_r$  obtained with the  $M$ -estimator, i.e.,  $\hat{\mathbf{U}}_r^M = [\theta_1(\hat{\Sigma}), \dots, \theta_r(\hat{\Sigma})]$ , while  $\hat{\mathbf{U}}_r^{\text{GCWE}}$  is its estimate obtained with the GCWE, i.e.,  $\hat{\mathbf{U}}_r^{\text{GCWE}} = [\theta_1(\hat{\Sigma}_{\text{GCWE}}), \dots, \theta_r(\hat{\Sigma}_{\text{GCWE}})]$ . Therefore, one can construct the following projectors

$$\begin{cases} \hat{\Pi}_r^M = \hat{\mathbf{U}}_r^M (\hat{\mathbf{U}}_r^M)^H, \\ \hat{\Pi}_r^{\text{GCWE}} = \hat{\mathbf{U}}_r^{\text{GCWE}} (\hat{\mathbf{U}}_r^{\text{GCWE}})^H. \end{cases} \quad (22)$$

**Theorem IV.1 (Standard asymptotic regime)** Let  $\hat{\Pi}_r^M$  the estimate of the projector  $\Pi_r$  obtained using an  $M$ -estimator defined in Eq. (22). The asymptotic distribution of  $\hat{\Pi}_r^M$  is given by

$$\sqrt{n} \text{vec}(\hat{\Pi}_r^M - \Pi_r) \xrightarrow{d} \mathcal{GCN}(\mathbf{0}, \vartheta_1 \Sigma_\Pi, \vartheta_1 \Sigma_\Pi \mathbf{K}), \quad (23)$$

where

$$\Sigma_\Pi = \mathbf{A}^T \otimes \mathbf{B} + \mathbf{B}^T \otimes \mathbf{A} \quad (24)$$

with  $\mathbf{A} = \mathbf{U}_r (\gamma^2 \mathbf{\Lambda}_r^{-2} + \mathbf{\Lambda}_r^{-1}) \mathbf{U}_r^H$ ,  $\mathbf{B} = \gamma^2 \Pi_r^\perp$  and  $\vartheta_1, \vartheta_2$  given by Eq. (8).

*Proof:* See Appendix C. ■

**Theorem IV.2 (GCWE)** Let  $\hat{\Pi}_r^M$  and  $\hat{\Pi}_r^{\text{GCWE}}$  be the estimates of the projector  $\Pi_r$  defined in Eq. (22). Then, the asymptotic distribution of  $\hat{\Pi}_r^M$  is given by

$$\sqrt{n} \text{vec}(\hat{\Pi}_r^M - \hat{\Pi}_r^{\text{GCWE}}) \xrightarrow{d} \mathcal{GCN}(\mathbf{0}, \sigma_1 \Sigma_\Pi, \sigma_1 \Sigma_\Pi \mathbf{K}) \quad (25)$$

with  $\Sigma_\Pi$  and  $\sigma_1, \sigma_2$  given by Eqs. (24) and (11), respectively.

*Proof:* See Appendix D. ■

**Remark IV.1** Theorem IV.1 (resp. IV.2) extend the results of Theorem III.1 (resp. III.2) to the principal subspace of  $M$ -estimators, which is a parameter of significant interest. We can draw the same conclusions as previously, notably, that the GCWE provides a better asymptotic characterization of this parameter.

#### V. SIMULATIONS AND EXAMPLES

##### A. Parameters setup

In order to validate the theoretical results we draw zero-mean  $t$ -distributed data with  $d$  degrees of freedom (DoF) whose PDF is given by Eq. (1) with

$$g_{\mathbf{z}}(x) = \left(1 + \frac{2x}{d}\right)^{-(p+d/2)} \quad (26)$$

and  $C_t = 2^p \Gamma(p + \frac{d}{2}) / [(\pi d)^p \Gamma(\frac{d}{2})]$ . The corresponding stochastic representation is given by Eq. (2) for  $\mathcal{Q} \sim pF_{2p,d}$ .

The DoF parameter is set to 3. The dimension of the data is  $p = 20$ . The scatter matrix is Toeplitz, i.e. elements are defined by  $\Sigma_{jk} = \rho^{|j-k|}$ ,  $j, k = 1, \dots, p$ , with correlation coefficient  $\rho$  set to  $0.9(1 + \sqrt{-1})/\sqrt{2}$ .

In order to carry out the simulations we will use the Student's  $M$ -estimator that is the MLE for Student's  $t$ -distribution and can be obtained as solution of Eq. (3) for

$$u(x) = \frac{2p+d}{d+2x}. \quad (27)$$

In this context, the parameters for asymptotic distribution of Student's  $M$ -estimator are given in Table I.

Standard regime	Gaussian equivalent
$\vartheta_1 = \frac{p+d/2+1}{p+d/2}$	$\sigma_1 = \frac{1}{p+d/2}$
$\vartheta_2 = \frac{2}{d} \times \frac{p+d/2+1}{p+d/2}$	$\sigma_2 = \frac{2}{d} \times \frac{p+d/2+1}{p+d/2}$

Table I: Coefficients  $\vartheta_1, \vartheta_2, \sigma_1$  and  $\sigma_2$  for Student's  $M$ -estimator

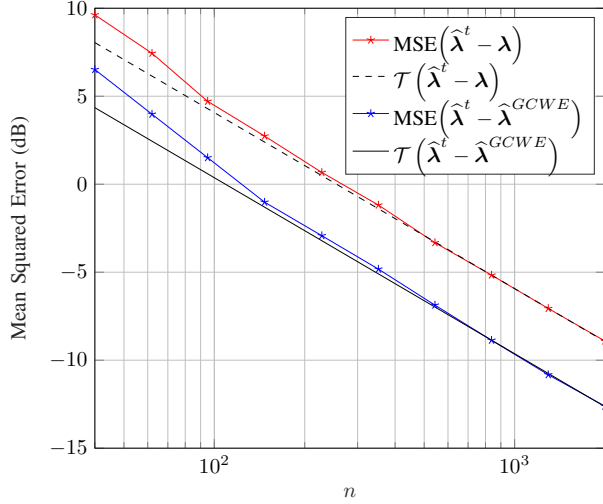


Figure 1: Empirical (MSE) and theoretical ( $\mathcal{T}$ ) mean squared error on eigenvalues: Results for the standard asymptotic regime ( $\hat{\lambda}^t - \lambda$ ) and Gaussian equivalent ( $\hat{\lambda}^t - \hat{\lambda}^{GCWE}$ );  $t$ -distributed data with  $p = 20$ ,  $d = 3$ .

### B. Experiments for eigenvalues

1) *Validation of theoretical results:* Let us consider  $\hat{\lambda}^t = \phi(\hat{\Sigma}_t)$  where  $\hat{\Sigma}_t$  is the Student's  $M$ -estimator and  $\hat{\lambda}^{GCWE} = \phi(\hat{\Sigma}_{GCWE})$ , where  $\hat{\Sigma}_{GCWE}$  is the SCM built with the Gaussian kernels of the observed data (GCWE).

Figure 1 displays the empirical mean squared error (MSE) of  $\hat{\lambda}^t$  when comparing to  $\lambda$  and  $\hat{\lambda}^{GCWE}$ , denoted as  $\text{MSE}(\hat{\lambda}^t - \lambda)$  and  $\text{MSE}(\hat{\lambda}^t - \hat{\lambda}^{GCWE})$  respectively. These quantities are compared to the corresponding asymptotic theoretical values, i.e. the traces of the asymptotic covariance matrices given in Eqs. (16) and (18), denoted as  $\mathcal{T}(\hat{\lambda}^t - \lambda) = \text{Tr}(\vartheta_1 \Lambda^2 + \vartheta_2 \lambda \lambda^T)$  and  $\mathcal{T}(\hat{\lambda}^t - \hat{\lambda}^{GCWE}) = \text{Tr}(\sigma_1 \Lambda^2 + \sigma_2 \lambda \lambda^T)$ . The horizontal scale represent the number of observations  $n$ .

First, we observe from Figure 1 that the empirical results (red and blue curves) tend to the corresponding theoretical ones (black curves) as  $n$  increases. Another remark is that the error of  $\hat{\lambda}^t$  is much smaller when comparing to  $\hat{\lambda}^{GCWE}$  than to  $\lambda$ . This support the idea that the distribution of the eigenvalues of an  $M$ -estimator (in this case Student's MLE) is better approximated with the one of the eigenvalues of the GCWE, then with the asymptotic Normal distribution based on the eigenvalues of the true scatter matrix.

2) *Application:* In [32] were derived an Intrinsic (i.e. Riemannian Manifold oriented) counterpart of the Cramér-Rao inequality. In the context of covariance matrix estimation, these results allows notably to bound the expected natural Riemannian distance (rather than the Frobenius norm):

$$d_{nat}^2(\Sigma_1, \Sigma_2) = \left\| \ln \left( \Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2} \right) \right\|_F^2 = \sum_{j=1}^p \ln^2 \lambda_j^d$$

where  $\lambda_j^d$  is the  $j^{\text{th}}$  eigenvalue of  $\Sigma_1^{-1} \Sigma_2$ . This analysis also reveals unexpected and hidden properties of estimators, such as a bias of the SCM w.r.t. the natural metric on  $\mathcal{H}_M^{++}$ . In this scope, the biased Intrinsic Cramér-Rao bound (CRLB) is established for the SCM in a Gaussian context as [32, Corollary 5]:

$$\mathbb{E} \left[ d_{nat}^2 \left( \hat{\Sigma}_{GCWE}, \Sigma \right) \right] \geq \frac{p^2}{n} + p\eta(p, n)^2. \quad (28)$$

The term  $\eta(p, n)$  in (28) is related to the intrinsic bias (IB) of the SCM given in [32, Theorem 7] by

$$\mathbb{E} \left[ \exp_{\Sigma}^{-1} \hat{\Sigma}_{GCWE} \right] = -\eta(p, n) \Sigma \quad (29)$$

with

$$\begin{aligned} \eta(p, n) &= \frac{1}{p} \{ p \ln n + p - \psi(n - p + 1) \\ &\quad + (n - p + 1) \psi(n - p + 2) \\ &\quad + \psi(n + 1) - (n + 1) \psi(n + 2) \} \end{aligned} \quad (30)$$

and where  $\psi(x) = \Gamma'(x)/\Gamma(x)$  is the digamma function.

For CES-distributed samples, the CRLB on  $d_{nat}^2$  has been derived in [35] for any unbiased estimator  $\hat{\Sigma}$  as

$$\mathbb{E} \left[ d_{nat}^2 \left( \hat{\Sigma}, \Sigma \right) \right] \geq \frac{p^2 - 1}{n\alpha} + (n(\alpha + p\beta))^{-1}. \quad (31)$$

with  $\alpha = \left( 1 - \frac{\mathbb{E}[\mathcal{Q}^2 \phi'(\mathcal{Q})]}{p(p+1)} \right)$  and  $\beta = \alpha - 1$ . Extending the Corollary 5 of [32] in this context would requires to derive the intrinsic bias of an  $M$ -estimator obtained with CES-distributed samples. The problem appears intractable since this result is mainly obtained thanks to the distribution of the eigenvalues of a Wishart-distributed matrix. However the GCWE equivalent from Theorem III.2 (as well as the previous simulation results) gives a reasonable theoretical ground for the approximation  $\mathbb{E}[\exp_{\Sigma}^{-1} \hat{\Sigma}_M] \simeq -\eta(p, n) \Sigma$  for any  $M$ -estimator consistent in scale under the CES framework. Hence, we can propose to incorporate an equivalent bias term in (31) to obtain an accurate approximation of the biased intrinsic CRLB for  $M$ -estimators build from CES-distributed samples (AB CRLB).

Figure 2 illustrates this point. Indeed, it can be seen that the empirical intrinsic bias obtained with Student's  $M$ -estimator computed with  $t$ -distributed data coincides with the intrinsic bias based on the SCM built with corresponding Gaussian-cores and the theoretical result (Eq. (30)). This once again confirms previous results and supports the proposed approximation. In addition, on Figure 3, the results for CRLB on  $d_{nat}^2$  have been plotted. Empirical mean of the natural Riemannian distance of  $\hat{\Sigma}_t$  (denoted as  $\epsilon^N(\hat{\Sigma}_t)$ ) is compared to the theoretical CRLB valid for any unbiased estimator (Eq. (31)) and recommended approximation equal to the sum of the latter and bias term from Eq. (28). As expected, one can see that by introducing the bias term AB CRLB approaches to  $\epsilon^N(\hat{\Sigma}_t)$  and gives more accurate theoretical results for CRLB.

### C. Eigenvectors and Principal Subspace

1) *Validation of theoretical results:* Figure 4 illustrates the results for the first eigenvector of  $\hat{\Sigma}_t$ ,  $\hat{\mathbf{u}}_1^t = \theta_1(\hat{\Sigma}_t)$ . It is

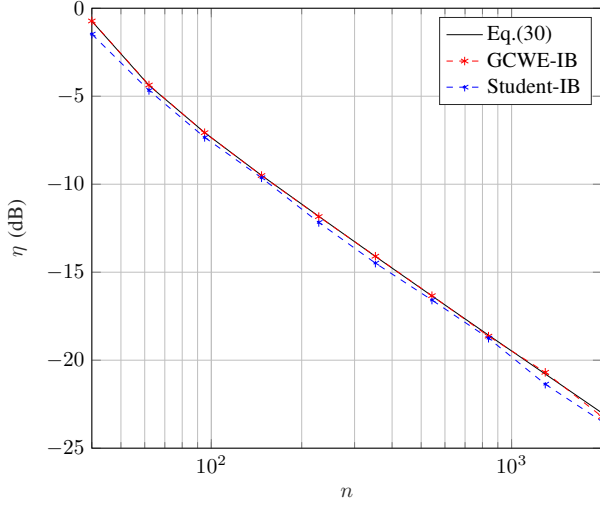


Figure 2: Empirical intrinsic bias for Student's  $M$ -estimator (Student-IB) and the Gaussian core SCM (GCWE-IB) compared to the theoretical result obtained for the GCWE (Eq. (30))

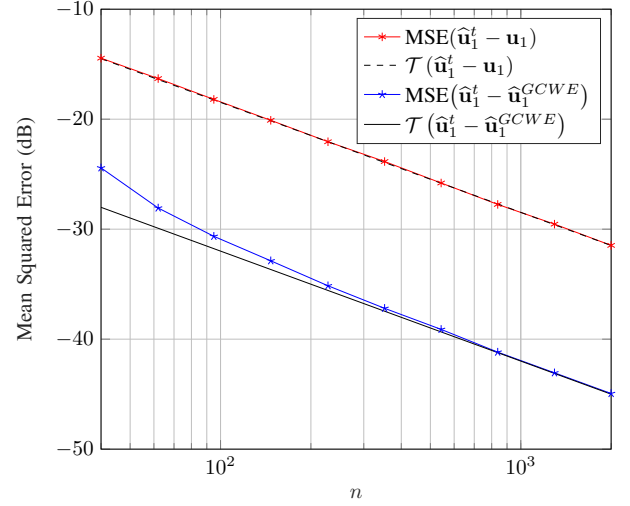


Figure 4: Empirical and theoretical mean squared error on eigenvectors: Results for the first eigenvector in the standard regime ( $\hat{\mathbf{u}}_1^t - \mathbf{u}_1$ ) and for the Gaussian equivalent ( $\hat{\mathbf{u}}_1^t - \hat{\mathbf{u}}_1^{GCWE}$ );  $t$ -distributed data with  $p = 20$ ,  $d = 3$ .

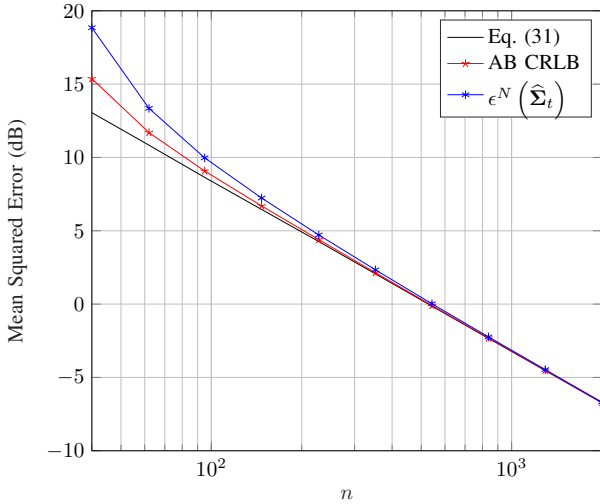


Figure 3: Empirical mean of  $d_{\text{nat}}^2(\hat{\Sigma}_t, \Sigma)$  denoted as  $\epsilon^N(\hat{\Sigma}_t)$  versus theoretical CRLB for an unbiased estimator in the CES framework (Eq. (31)) and approximated biased intrinsic CRLB (AB CRLB)

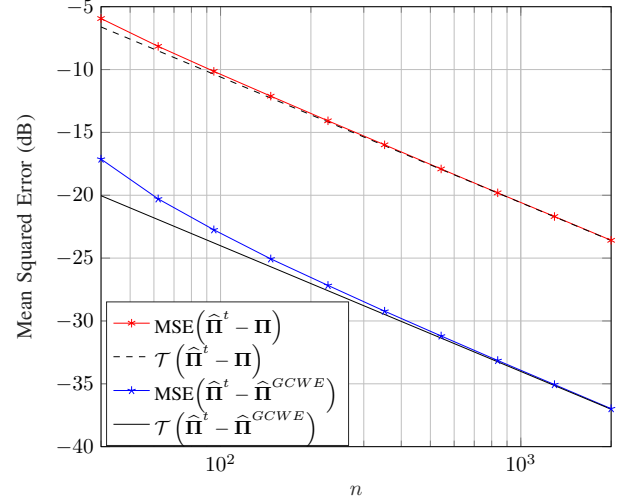


Figure 5: Empirical and theoretical mean squared error on projector: Results for the standard regime ( $\hat{\Pi}^t - \Pi$ ) and for the Gaussian equivalent ( $\hat{\Pi}^t - \hat{\Pi}^{GCWE}$ );  $t$ -distributed data with  $p = 20$ ,  $r = 5$ ,  $d = 3$ .

apparent from the plotted curves that empirical results, ones again, coincides well with the theoretical ones. Moreover, the figure shows a significant difference between the results for the standard regime and GCWE. This can be explained by the fact that the covariance matrix of the eigenvectors depends only on the first scale factor, contrary to the one of eigenvalues. As detailed in Section III, this is expected since the eigenvector are scale-invariant functions of the scatter matrix. For the Student's  $M$ -estimator the first scale factor  $\sigma_1$  is much smaller than  $\sigma_2$ , especially when the data dimension grows, and the approximation in this case is even stronger.

Figure 5 presents the MSE for the projector defined by (22).

The data dimension  $p$  is equal to 20, while the rank  $r$  of  $\Sigma_r$  is set to 5. Parameter  $\gamma^2$  is set to 1 and  $\Lambda_r$  is designed such that  $\min(\text{diag}(\Lambda_r)) \gg \gamma^2$ . The figure validates the theoretical results proposed in Theorems IV.1 and IV.2 and leads us to the same conclusions as previously.

2) *SNR Loss*: Let us consider the STAP problem with the factor model introduced previously and the optimal filter  $\mathbf{w}_{\text{opt}}$  [36] given by

$$\mathbf{w}_{\text{opt}} = \Sigma^{-1} \mathbf{p} \quad (32)$$

where  $\mathbf{p}$  is the known STAP steering vector. In the low-rank clutter case an alternative is to use the low-rank STAP filter

$\mathbf{w}_R$  [37], [38] defined as

$$\mathbf{w}_r = \Pi^\perp \mathbf{p}. \quad (33)$$

In practice, in order to use the STAP filter one has to estimate the covariance matrix  $\Sigma$  and the projector  $\Pi^\perp$  from the secondary data  $\mathbf{z}_i \sim \mathcal{CES}(\mathbf{0}, \Sigma, \mathbf{g}_z)$ , which is usually done with an  $M$ -estimator.

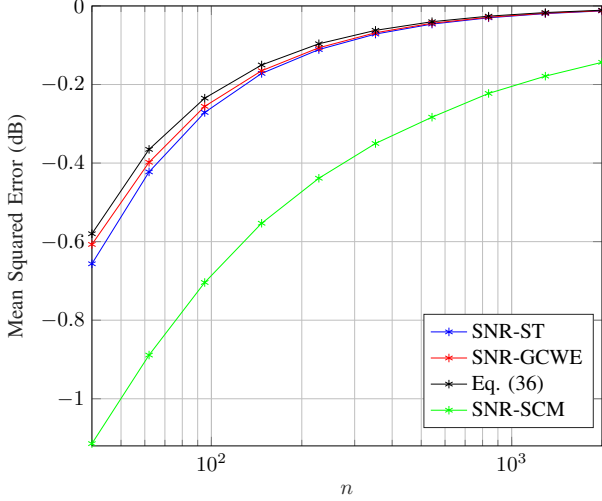


Figure 6: Empirical SNR Loss obtained with the Student's  $M$ -estimator (ST-SNR), GCWE (GCWE-SNR) and SCM (SCM-SNR) versus the theoretical result given by Eq. (36);  $t$ -distributed data with  $p = 20$ ,  $r = 5$ ,  $d = 3$ .

We are then interested in computing the SNR Loss  $\rho$  given by

$$\rho = \frac{SNR_{out}}{SNR_{max}} = \frac{|\hat{\mathbf{w}}^H \mathbf{p}|^2}{(\hat{\mathbf{w}}^H \Sigma \hat{\mathbf{w}})(\mathbf{p}^H \Sigma \mathbf{p})} \quad (34)$$

or equivalently

$$\rho = \gamma^2 \frac{(\mathbf{p}^H \hat{\Pi}^\perp \mathbf{p})^2}{\mathbf{p}^H \hat{\Pi}^\perp \Sigma \hat{\Pi}^\perp \mathbf{p}}. \quad (35)$$

In [38] it has been shown that when the data are Gaussian-distributed and the covariance matrix estimated using the SCM, the theoretical SNR Loss is given by

$$\mathbb{E}[\rho] = 1 - \frac{r}{n}. \quad (36)$$

Figure 6 draws a comparison between the values of empirical mean of SNR Loss obtained with the projector estimate based on  $\hat{\Sigma}_t$  with  $t$ -distributed secondary data (SNR-ST), empirical mean of SNR Loss computed with the corresponding GCWE (SNR-GCWE) which theoretical expectation is given in Eq. (36). One can notice that the value of SNR-ST is very close to the one SNR-GCWE, as anticipated, which supports the idea to approximate the behavior of SNR-ST with the one of SNR-GCWE [38] when necessary. The green curve presents empirical mean of SNR Loss based on the SCM computed with  $t$ -distributed data showing the importance of  $M$ -estimators in CES context.

## VI. CONCLUSION

This paper has analysed the asymptotic distribution of the EVD as well as the one of the principal subspace of scatter  $M$ -estimators. The results in the standard asymptotic regime have been derived. Then, relying on [16] the results have been extended, giving the convergence towards EVD of GCWE. The derived moments of second order appear to be much smaller in this case than in standard regime, offering a better approximation of the elements behavior. The applications of the theoretical results on SNR Loss and biased Intrinsic CLRB have been illustrated. We came up with the same conclusion that is that the behavior of EVD parameters are much better explained with the one of GCWE EVD parameters than with their standard asymptotic Normal distribution. The great benefit of these results is that one can use  $M$ -estimators to compute scatter matrix estimator and obtain a more precise estimation of EVD in CES framework, while leaning on the theoretical results obtained for the corresponding GCWE. Importantly, these results can be easily applied to a wide scope of problems.

## APPENDICES

To prove all theorems we will use the basic results obtained in the following theorem.

**Lemma .1** Let  $\{\hat{\mathbf{z}}\}$  be a sequence of complex random vectors  $\hat{\mathbf{z}}$  and  $\mathbf{z}$  a compatible fixed vector. Assume that  $\sqrt{N}(\hat{\mathbf{z}} - \mathbf{z}) \xrightarrow{d} \mathcal{GCN}(\mathbf{0}, \mathbf{V}, \mathbf{W})$ . Let  $\xi(\mathbf{y})$  be a vector function of a vector  $\mathbf{y}$  with first and a second derivatives existing in a neighbourhood of  $\mathbf{y} = \mathbf{z}$ . Then

$$\sqrt{N}(\xi(\hat{\mathbf{y}}) - \xi(\mathbf{y})) \xrightarrow{d} \mathcal{GCN}(\mathbf{0}, \mathbf{D}\mathbf{V}\mathbf{D}^H, \mathbf{D}\mathbf{W}\mathbf{D}^T) \quad (37)$$

where

$$\mathbf{D} = \left. \frac{d(\xi(\mathbf{y}))}{d\mathbf{y}} \right|_{\mathbf{y}=\mathbf{z}} \quad (38)$$

is a matrix derivative.

## APPENDIX A PROOF OF THEOREM III.1

*Proof:* To derive the derivatives of  $\phi$  and  $\theta_j$  with respect to  $\text{vec}(\mathbf{M})$  at the point  $\mathbf{M} = \Sigma$  we differentiate  $\mathbf{M}\theta_j = \phi_j\theta_j$

$$d\mathbf{M}\mathbf{u}_j + \Sigma d\theta_j = d\phi_j\mathbf{u}_j + \lambda_j d\theta_j. \quad (39)$$

Multiplying each side of the last equation by  $\mathbf{u}_j^H$ , one has

$$d\phi_j = \mathbf{u}_j^H (d\mathbf{M}) \mathbf{u}_j$$

since  $\mathbf{u}_j^H \Sigma = \lambda_j \mathbf{u}_j^H$  and  $\mathbf{u}^H \mathbf{u} = 1$ . Thus,

$$\left. \frac{d\phi_j}{d\text{vec}(\mathbf{M})} \right|_{\mathbf{M}=\Sigma} = \mathbf{u}_j^T \otimes \mathbf{u}_j^H.$$

If  $\phi = (\phi_1, \dots, \phi_p)$ , then one has

$$\left. \frac{d\phi_j}{d\text{vec}(\mathbf{M})} \right|_{\mathbf{M}=\Sigma} = \mathbf{E}^T (\mathbf{U}^T \otimes \mathbf{U}^H)$$

with  $\mathbf{E} = (\mathbf{e}_1 \otimes \mathbf{e}_1 \dots \mathbf{e}_p \otimes \mathbf{e}_p)$  where  $\mathbf{e}_j, j = 1, \dots, p$  are unit vectors. Further, combining the statement given in Lemma 1 with Eq.(7), one obtains

$$\begin{aligned} & \mathbf{E}^T (\mathbf{U}^T \otimes \mathbf{U}^H) \left( \vartheta_1 \left( \Sigma_\sigma^T \otimes \Sigma_\sigma \right) \right) (\mathbf{U}^* \otimes \mathbf{U}) \mathbf{E} \\ & + \mathbf{E}^T (\mathbf{U}^T \otimes \mathbf{U}^H) \vartheta_2 \text{vec}(\Sigma_\sigma) \text{vec}(\Sigma_\sigma)^H (\mathbf{U}^* \otimes \mathbf{U}) \mathbf{E} \\ & = \vartheta_1 \mathbf{E}^T (\Lambda^T \otimes \Lambda) \mathbf{E} + \vartheta_2 \mathbf{E}^T \left( \text{vec}(\Lambda) \text{vec}(\Lambda)^H \right) \mathbf{E} \\ & = \vartheta_1 \Lambda^2 + \vartheta_2 \Lambda \Lambda^T. \end{aligned}$$

Note that since the eigenvalues are real one obtains the same result using the expression for the pseudo-covariance matrix.

In order to obtain the results for eigenvectors, we will multiply Eq. (39) by  $\mathbf{u}_k^H, k \neq j$ . Thus, one obtains

$$\mathbf{u}_k^H (d\mathbf{M}) \mathbf{u}_j = (\lambda_j - \lambda_k) \mathbf{u}_k^H d\boldsymbol{\theta}_j$$

as  $\mathbf{u}_k^H \mathbf{u}_j = 0$ . Following the same steps as in [26] (done for the real case), it is easy to show that

$$d\boldsymbol{\theta}_j = \sum_{j \neq k} (\lambda_j - \lambda_k)^{-1} \mathbf{u}_k \mathbf{u}_k^H (d\mathbf{M}) \mathbf{u}_j + \mathbf{u}_j \mathbf{u}_j^H d\boldsymbol{\theta}_j.$$

In fact, the last element in the previous equality is omitted in the real case since  $\mathbf{u}_j^T d\boldsymbol{\theta}_j = 0$  (from  $\boldsymbol{\theta}_j^T \boldsymbol{\theta}_j = 1$ ). However, in the complex case  $\mathbf{u}_j^H d\boldsymbol{\theta}_j \neq 0$ , as from  $\boldsymbol{\theta}_j^H \boldsymbol{\theta}_j = 1$  one has  $\mathbf{u}_j^H d\boldsymbol{\theta}_j + d\boldsymbol{\theta}_j^H \mathbf{u}_j = 0$  and it is obvious that  $\mathbf{u}_j^H d\boldsymbol{\theta}_j \neq d\boldsymbol{\theta}_j^H \mathbf{u}_j$ . In some works, the authors use the different normalization for eigenvectors  $\mathbf{u}_j^H \boldsymbol{\theta}_j = 1$  which directly implies  $\mathbf{u}_j^H d\boldsymbol{\theta}_j = 0$  and in those circumstances the results correspond to the ones in the real case. In the general (more common) case, one obtains

$$(\mathbf{I} - \mathbf{u}_j \mathbf{u}_j^H) d\boldsymbol{\theta}_j = \left( \mathbf{u}_j^T \otimes \mathbf{U} (\lambda_j \mathbf{I} - \Lambda)^+ \mathbf{U}^H \right) d\mathbf{M},$$

which actually gives the projection of the derivative onto the subspace orthogonal to the one of the eigenvector. Now, employing Eq. (37) with the previous derivatives and since

$$\begin{aligned} \left( \mathbf{u}_j^T \otimes \mathbf{U} (\lambda_j \mathbf{I} - \Lambda)^+ \mathbf{U}^H \right) \mathbf{K} &= \mathbf{U} (\lambda_j \mathbf{I} - \Lambda)^+ \mathbf{U}^H \otimes \mathbf{u}_j^T \\ & \quad (\lambda_j \mathbf{I} - \Lambda)^+ \mathbf{e}_j = \mathbf{0} \\ \left[ \mathbf{u}_j^T \otimes \mathbf{U} (\lambda_j \mathbf{I} - \Lambda)^+ \mathbf{U}^H \right] \text{vec}(\Sigma) &= \mathbf{0} \end{aligned}$$

one obtains the final results. Note that  $\mathcal{GCN}$  becomes  $\mathcal{CN}$  since the pseudo-covariance matrix is equal to zero. ■

## APPENDIX B PROOF OF THEOREM III.2

*Proof:* Rewriting the left-hand side of Eq. (16)

$$\begin{aligned} \sqrt{n} \left( \sigma \hat{\boldsymbol{\lambda}}^M - \hat{\boldsymbol{\lambda}}^{\text{GCWE}} \right) &= \sqrt{n} \left( \sigma \hat{\boldsymbol{\lambda}}^M - \boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}}^{\text{GCWE}} + \boldsymbol{\lambda} \right) = \\ &= \sqrt{n} \left( \left( \sigma \hat{\boldsymbol{\lambda}}^M - \boldsymbol{\lambda} \right) - \left( \hat{\boldsymbol{\lambda}}^{\text{GCWE}} - \boldsymbol{\lambda} \right) \right). \end{aligned}$$

Then

$$\begin{aligned} & \text{var}_n \left( \sigma \hat{\boldsymbol{\lambda}}^M - \hat{\boldsymbol{\lambda}}^{\text{GCWE}} \right) = \\ & \mathbb{E} \left[ n \left( \sigma \hat{\boldsymbol{\lambda}}^M - \hat{\boldsymbol{\lambda}}^{\text{GCWE}} \right) \left( \sigma \hat{\boldsymbol{\lambda}}^M - \hat{\boldsymbol{\lambda}}^{\text{GCWE}} \right)^T \right] \\ & = \text{var}_n \left( \sigma \hat{\boldsymbol{\lambda}}^M \right) - 2 \text{cov}_n \left( \hat{\boldsymbol{\lambda}}^M, \sigma \hat{\boldsymbol{\lambda}}^{\text{GCWE}} \right) + \text{var}_n \left( \hat{\boldsymbol{\lambda}}^{\text{GCWE}} \right). \end{aligned}$$

Since from (16) one has

$$\begin{aligned} \text{var}_n \left( \sigma \hat{\boldsymbol{\lambda}}^M \right) &\xrightarrow{n \rightarrow +\infty} \vartheta_1 \Lambda^2 + \vartheta_2 \Lambda \Lambda^T \quad \text{and} \\ \text{var}_n \left( \hat{\boldsymbol{\lambda}}^{\text{GCWE}} \right) &\xrightarrow{n \rightarrow +\infty} \Lambda^2, \end{aligned}$$

it remains only to derive the expression for

$$\text{cov}_n \left( \sigma \hat{\boldsymbol{\lambda}}^M, \hat{\boldsymbol{\lambda}}^{\text{GCWE}} \right) = \mathbb{E} \left[ n \left( \sigma \hat{\boldsymbol{\lambda}}^M - \boldsymbol{\lambda} \right) \left( \hat{\boldsymbol{\lambda}}^{\text{GCWE}} - \boldsymbol{\lambda} \right)^T \right].$$

Using the Delta method, one can show that

$$\text{cov}_n \left( \sigma \hat{\boldsymbol{\lambda}}^M, \hat{\boldsymbol{\lambda}}^{\text{GCWE}} \right) \rightarrow \left. \frac{\sigma d\phi}{d\text{vec}(\mathbf{M})} \right|_{\mathbf{M}=\Sigma} \mathbf{Q} \left. \frac{d\phi}{d\text{vec}(\mathbf{M})} \right|_{\mathbf{M}=\Sigma}^T$$

where  $\mathbf{Q}$  is the asymptotic pseudo-covariance matrix of  $\hat{\Sigma}$ . This matrix is equal to

$$\mathbf{Q} = \gamma_1 \left( \Sigma_\sigma^T \otimes \Sigma_\sigma \right) \mathbf{K} + \gamma_2 \text{vec}(\Sigma_\sigma) \text{vec}(\Sigma_\sigma)^T \quad (40)$$

as given in ref. Repeating the same steps as in Eqs. (40), one shows that the right-hand side of the right-hand side of Eq. (40) becomes

$$\gamma_1 \Lambda^2 + \gamma_2 \Lambda \Lambda^T$$

which, since  $\sigma_1 = \vartheta_1 - 2\gamma_1 + 1$  and  $\sigma_2 = \vartheta_2 - 2\gamma_2$ , leads to the final results.

The results for the eigenvectors can be obtained following the same procedure as for the eigenvalues. ■

## APPENDIX C PROOF OF THEOREM IV.1

*Proof:* If we define the pseudo-inverse of  $\Sigma_r$  as

$$\Phi = \mathbf{U}_r \Lambda_r^{-1} \mathbf{U}_r^H \quad (41)$$

one has from ref that

$$\hat{\Pi}_r = \Pi_r + \delta \Pi_r + \dots + \delta^i \Pi_r + \dots$$

where

$$\begin{aligned} \delta \Pi_r &= \Pi_r^\perp \Delta \Sigma \Phi + \Phi \Delta \Sigma \Pi_r^\perp \\ \delta^i \Pi_r &= -\Pi_r^\perp (\delta^{i-1} \Pi) \Delta \Sigma \Phi + \Pi_r^\perp (\delta^{i-1} \Pi) \Delta \Sigma \Phi \end{aligned}$$

with  $\Delta \Sigma = \hat{\Sigma} - \Sigma$ .

In the asymptotic regime, when  $n \rightarrow \infty$  we can write

$$\hat{\Pi}_r = \Pi_r + \delta \Pi_r$$

since  $\Delta \Sigma$  is close to zero. Hence, taking a vec of the  $\hat{\Pi}_r - \Pi_r = \delta \Pi_r$ , one gets

$$\text{vec}(\hat{\Pi}_r - \Pi_r) = \mathbf{F} \text{vec}(\hat{\Sigma} - \Sigma)$$

with

$$\mathbf{F} = \left( \Phi^T \otimes \Pi_r^\perp + \left( \Pi_r^\perp \right)^T \otimes \Phi \right).$$



It is now obvious that the covariance (resp. pseudo-covariance) matrix of  $\sqrt{n}(\Pi_r^M - \Pi_r)$  is equal to  $\mathbf{FCF}^H$  (resp.  $\mathbf{FPF}^T$ ) where  $\mathbf{C}$  and  $\mathbf{P}$  are given in Eqs. (7). Further

$$\begin{aligned}\mathbf{FC} &= \left( \Phi^T \otimes \Pi_r^\perp + \left( \Pi_r^\perp \right)^T \otimes \Phi \right) \left( \Sigma^T \otimes \Sigma \right) \\ &+ \left( \Phi^T \otimes \Pi_r^\perp + \left( \Pi_r^\perp \right)^T \otimes \Phi \right) \text{vec}(\Sigma) \text{vec}(\Sigma)^H \\ &= \left( \Phi^T \Sigma^T \otimes \Pi_r^\perp \Sigma + \left( \Pi_r^\perp \right)^T \Sigma^T \otimes \Phi \Sigma \right)\end{aligned}$$

as  $\left( \Phi^T \otimes \Pi_r^\perp + \left( \Pi_r^\perp \right)^T \otimes \Phi \right) \text{vec}(\Sigma) = \mathbf{0}$  using  $(\mathbf{T}^T \otimes \mathbf{R}) \text{vec}(\mathbf{S}) = \text{vec}(\mathbf{RST})$  and  $\Pi_r^\perp \Sigma \Phi = \Phi \Sigma \Pi_r^\perp = \mathbf{0}$ . Finally, after the postmultiplication by  $\mathbf{F}^H$  and since

$$\begin{aligned}\Sigma &= \Sigma^H \neq \Sigma^T \\ \Phi &= \Phi^H \neq \Phi^T \\ \Pi_r^\perp &= \left( \Pi_r^\perp \right)^H \neq \left( \Pi_r^\perp \right)^T\end{aligned}$$

one obtains

$$\mathbf{FCF}^H = \left( (\Phi \Sigma \Phi)^T \otimes \Pi_r^\perp \Sigma \Pi_r^\perp + \left( \Pi_r^\perp \Sigma \Pi_r^\perp \right)^T \otimes \Phi \Sigma \Phi \right)$$

which with  $\Phi$  given by Eq. (41) and  $\Sigma = \mathbf{U}_r \Lambda_r \mathbf{U}_r^H + \gamma^2 \mathbf{I}_p$  yields the final result.

Analogously, one can derive the results for the pseudo-covariance using the equality  $\mathbf{K}(\mathbf{A} \otimes \mathbf{B}) = (\mathbf{B} \otimes \mathbf{A}) \mathbf{K}$ . ■

## APPENDIX D

### PROOF OF THEOREM IV.2

Following the steps from Appendix B and using the results of Theorem IV.1 one gets the results of the theorem.

## REFERENCES

- [1] R. A. Maronna, "Robust  $M$ -estimators of multivariate location and scatter," *Annals of Statistics*, vol. 4, no. 1, pp. 51–67, January 1976.
- [2] D. E. Tyler, "A distribution-free  $M$ -estimator of multivariate scatter," *The Annals of Statistics*, vol. 15, no. 1, pp. 234–251, 1987.
- [3] F. Gini and M. Greco, "Covariance matrix estimation for cfar detection in correlated heavy tailed clutter," *Signal Processing*, vol. 82, no. 12, pp. 1847–1859, 2002.
- [4] F. Pascal, Y. Chitour, J.-P. Ovarlez, P. Forster, and P. Larzabal, "Covariance structure maximum-likelihood estimates in Compound-Gaussian noise: existence and algorithm analysis," *Signal Processing, IEEE Transactions on*, vol. 56, no. 1, pp. 34–48, January 2008.
- [5] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, "Complex elliptically symmetric distributions: Survey, new results and applications," *Signal Processing, IEEE Transactions on*, vol. 60, no. 11, pp. 5597–5625, November 2012.
- [6] E. Ollila and D. E. Tyler, "Regularized  $M$ -estimators of scatter matrix," *Signal Processing, IEEE Transactions on*, vol. 62, no. 22, pp. 6059–6070, Nov. 2014.
- [7] Y. Sun, P. Babu, and D. P. Palomar, "Regularized Tyler's Scatter Estimator: Existence, Uniqueness and Algorithms," *Signal Processing, IEEE Transactions on*, vol. 62, no. 19, pp. 5143–5156, Oct. 2014.
- [8] A. Wiesel and T. Zhang, "Structured robust covariance estimation," *Foundations and Trends in Signal Process.*, vol. 8, no. 3, pp. 127–216, 2015. [Online]. Available: <http://dx.doi.org/10.1561/20000000053>
- [9] R. J. Muirhead, "Aspects of multivariate statistical analysis," *JOHN WILEY & SONS, INC., 605 THIRD AVE., NEW YORK, NY 10158, USA*, 1982, 656, 1982.
- [10] D. E. Tyler, "Radial estimates and the test for sphericity," *Biometrika*, vol. 69, no. 2, p. 429, 1982.
- [11] M. Mahot, F. Pascal, P. Forster, and J.-P. Ovarlez, "Asymptotic properties of robust complex covariance matrix estimates," *IEEE Transactions on Signal Processing*, vol. 61, no. 13, pp. 3348–3356, July 2013.
- [12] I. Soloveychik and A. Wiesel, "Performance analysis of Tyler's covariance estimator," *Signal Processing, IEEE Transactions on*, vol. 63, no. 2, pp. 418–426, Jan 2015.
- [13] T. Zhang, X. Cheng, and A. Singer, "Marchenko-Pastur Law for Tyler's and Maronna's  $M$ -estimators," *arXiv preprint*, 2014.
- [14] R. Couillet, F. Pascal, and J. W. Silverstein, "The Random Matrix Regime of Maronna's  $M$ -estimator with elliptically distributed samples," *Journal of Multivariate Analysis*, vol. 139, pp. 56–78, July 2015.
- [15] G. Drašković and F. Pascal, "New properties for Tyler's covariance matrix estimator," in *2016 50th Asilomar Conference on Signals, Systems and Computers*, Nov 2016, pp. 820–824.
- [16] —, "New Insights Into the Statistical Properties of  $M$ -Estimators," *IEEE Transactions on Signal Processing*, vol. 66, no. 16, pp. 4253–4263, Aug 2018.
- [17] C. Croux and G. Haesbroeck, "Principal component analysis based on robust estimators of the covariance or correlation matrix: Influence functions and efficiencies," *Biometrika*, vol. 87, pp. 603–618, 2000.
- [18] J. Zhao and Q. Jiang, "Probabilistic PCA for  $t$  distributions," *Neurocomputing*, vol. 69, no. 16–18, pp. 2217–2226, 2006.
- [19] M. Rangaswamy, "Statistical analysis of the nonhomogeneity detector for non-Gaussian interference backgrounds," *Signal Processing, IEEE Transactions on*, vol. 53, no. 6, pp. 2101–2111, June 2005.
- [20] G. Ginolhac and P. Forster, "Approximate distribution of the low-rank adaptive normalized matched filter test statistic under the null hypothesis," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 4, pp. 2016–2023, August 2016.
- [21] P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules," *IEEE Signal Processing Magazine*, vol. 21, no. 4, pp. 36–47, July 2004.
- [22] E. Terreaux, J.-P. Ovarlez, and F. Pascal, "Robust model order selection in large dimensional elliptically symmetric noise," *arXiv:1710.06735*, 2018.
- [23] A. Kammoun, R. Couillet, F. Pascal, and M. Alouini, "Optimal Design of the Adaptive Normalized Matched Filter Detector Using Regularized Tyler Estimators," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 2, pp. 755–769, April 2018.
- [24] D. Ciuonzo, V. Carotenuto, and A. De Maio, "On multiple covariance equality testing with application to SAR change detection," *IEEE Transactions on Signal Processing*, vol. 65, no. 19, pp. 5078–5091, 2017.
- [25] C. Bouveyron and C. Brunet-Saumard, "Model-based clustering of high-dimensional data: A review," *Computational Statistics & Data Analysis*, vol. 71, pp. 52–78, 2014.
- [26] T. Kollo and H. Neudecker, "Asymptotics of eigenvalues and unit-length eigenvectors of sample variance and correlation matrices," *Journal of Multivariate Analysis*, vol. 47, no. 2, pp. 283 – 300, 1993.
- [27] H. Krim, P. Forster, and J. G. Proakis, "Operator approach to performance analysis of root-music and root-min-norm," *IEEE Transactions on Signal Processing*, vol. 40, no. 7, pp. 1687–1696, Jul 1992.
- [28] H. Krim and P. Forster, "Projections on unstructured subspaces," *IEEE Transactions on Signal Processing*, vol. 44, no. 10, pp. 2634–2637, Oct 1996.
- [29] D. Tyler, "Asymptotic inference for eigenvectors," *The Annals of Statistics*, pp. 725–736, 1981.
- [30] A. Zanella, M. Chiani, and M. Z. Win, "On the marginal distribution of the eigenvalues of Wishart matrices," *IEEE Transactions on Communications*, vol. 57, no. 4, pp. 1050–1060, April 2009.
- [31] G. Ginolhac and P. Forster, "Performance analysis of a robust low-rank STAP filter in low-rank Gaussian clutter," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, March 2010, pp. 2746–2749.
- [32] S. T. Smith, "Covariance, subspace, and intrinsic Cramér-Rao bounds," *IEEE Transactions on Signal Processing*, vol. 53, no. 5, pp. 1610–1630, 2005.
- [33] K. Yao, "A representation theorem and its applications to spherically invariant random processes," *Information Theory, IEEE Transactions on*, vol. 19, no. 5, pp. 600–608, September 1973.
- [34] M. Bilodeau and D. Brenner, *Theory of Multivariate Statistics*, ser. New York, NY. USA:Springer-Verlag, 1999.
- [35] A. Breloy, G. Ginolhac, A. Renaux, and F. Bouchard, "Intrinsic Cramér-Rao Bounds for Scatter and Shape Matrices Estimation in CES Distributions," *submitted to IEEE Signal Processing Letters*, 2018.
- [36] J. Ward, "Space-Time Adaptive Processing for Airborne Radar MIT Lincoln Lab., Lexington," MA, Tech. Rep. 1015, Tech. Rep., 1994.

- [37] I. P. Kirsteins and D. W. Tufts, "Adaptive detection using low rank approximation to a data matrix," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 30, no. 1, pp. 55–67, 1994.
- [38] A. Haimovich, "Asymptotic distribution of the conditional signal-to-noise ratio in an eigenanalysis-based adaptive array," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 33, no. 3, pp. 988–997, 1997.