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 the M-estimators of the scatter matrix is a popular choice to perform robust probabilistic PCA or other dimension reduction related applications. Towards the goal of characterizing this process, this paper derives new asymptotics for the EVD parameters (i.e. eigenvalues, eigenvectors, and principal subspace) of M-estimators 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 the Gaussian context. Second, their convergence 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 rely on well-established results on the EVD of Wishart-distributed matrix for characterizing the EVD of M-estimators. Finally, some examples (intrinsic bias analysis, rank estimation, and low-rank adaptive filtering) illustrate where the obtained results can be leveraged.**

I. Introduction

SECOND order statistics play a key role in signal processing and machine learning applications. In the context of elliptical distributions, these are characterized by the scatter matrix, which describes the correlations between the entries of the samples (and is proportional to the covariance matrix, when the latter exists). Usually, this parameter is unknown and must be estimated in order to apply a so-called adaptive process. In this scope, the M-estimators of the scatter matrix, introduced in [1], have motivated research [2]–[8] due to their robustness properties over the large 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 matrix is a complex issue because 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 characterizations for these estimators. Their asymptotic Gaussian distribution is 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, in [15], [16] it has been showed that their distribution can be very accurately approximated by a Wishart one of an equivalent theoretical Gaussian 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-Nanterrre, 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 matrix 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 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 the results obtained in [17], [29], [30] 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 the analyis done in [17], [29], [30], all the results in this paper are derived for complex data.
- The Gaussian asymptotic distribution in the GCWE regime by extending the results of [15], [16]. To do so, a central limit theorem is established to show that the EVD parameters of M-estimators are asymptotically concentrated around their GCWE counterparts with a variance that is significantly lower than the one of the standard asymptotic regime (derived around the true expected values). Thus, this result represents the main contribution in the sense that it quantifies when it is acceptable to directly rely on well established results on the EVD of Wishart-distributed matrices [9], [31] for characterizing the EVD of M-estimators.

In the last part, we eventually give three examples where the proposed results can be leveraged:

1) We address the complex issue of characterizing the intrinsic bias [32] of M-estimators in the CES context. This quantity has been studied in [32] for the SCM in the Gaussian context thanks to the distribution of the eigenvalues of a Wishart matrix [9]. Extending this analysis to M-estimators in the general CES context represents, at first sight, an intractable problem because of their unknown exact distribution. However, the established convergence of the eigenvalues of an M-estimator toward their GCWE counterpart allows to derive an accurate approximation of this intrinsic bias;
2) In the context of model order selection (i.e., rank estimation) from non-Gaussian samples, we show that the use of M-estimators (rather than the SCM) in theoretic criteria derived for Gaussian models [33], [34] yields the same results as the one obtained with the theoretical GCWE. Again, this justifies a plug-in approach (using M-estimators in processes derived under the Wishart assumption), instead of a complete re-derivation that would require to assume an exact CES distribution;
3) The performance of low rank filters [35] built from M-estimators are derived in the same way (i.e., approached by the one of their GCWE) to illustrate that the approach also holds for adaptive processes based on the eigenvectors.

The body of this paper is organized as follows. Section II introduces the CES distributions and M-estimators. Section III contains the main results about the EVD of M-estimators. Section IV introduces LR models and presents the main results about principal subspaces of M-estimators. In Section V Monte Carlo simulations are presented in order to validate the theoretical results. Examples of applications of the results are presented in Section VI. Conclusions are drawn in Section VII.

Notation: Vectors (resp. matrices) are denoted by bold-faced lowercase letters (resp. uppercase letters). $\mathbf{A}^T$, $\mathbf{A}^*$, $\mathbf{A}^H$ and $\mathbf{A}^+$ respectively represent the transpose, conjugate, Hermitian operator and pseudo-inverse of the matrix $\mathbf{A}$. The acronyms i.i.d. and w.r.t. stand respectively for "independent and identically distributed," and "with respect to." The notation $\sim$ means "is distributed as," while $\overset{d}{=}$ stands for "shares the same distribution as," and $\overset{d}{\to}$ denotes convergence in distribution. The operator $\otimes$ denotes the Kronecker product, while $\text{vec}(\cdot)$ is the operator that transforms a matrix $p \times n$ into a vector of length pn (concatenating its n columns into a single one). Moreover, $\mathbf{I}_p$ is the $p \times p$ identity matrix, $\mathbf{0}$ is 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. $\mathbf{K}\,\text{vec}(\mathbf{A}) = \text{vec}(\mathbf{A}^T)$. The operator $\text{diag}(\cdot)$ transforms a vector in a diagonal matrix $\mathbf{A} = \text{diag}(\mathbf{a})$, with $[\mathbf{A}]_{i,i} = \mathbf{a}_i$. The set of Hermitian positive definite matrices is denoted $\mathcal{H}_M^{++}$. The Stiefel manifold (set of semi-unitary matrices) is denoted as $\mathcal{U}_r^p = \left\{\mathbf{U} \in \mathbb{C}^{p\times r}, \mathbf{U}^H\mathbf{U} = \mathbf{I}_r\right\}$. Finally, $\mathcal{GCN}\left(\mathbf{0}, \mathbf{V}, \mathbf{W}\right)$ 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], parameterized by a mean vector $\boldsymbol{\mu}$ and a scatter matrix $\boldsymbol{\Sigma}$, which describes the correlations between the entries. In the absolute continuous case, the probability density function (PDF) of a CES distribution can be written as

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

where C is a normalization constant and $g_{\mathbf{z}} : [0,\infty) \to [0,\infty)$ is any function (called the density generator) ensuring Eq. (1) defines a PDF. The Complex Normal (Gaussian) distribution $\mathbf{z} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a particular case of CES distributions in which $g_{\mathbf{z}}(z) = e^{-z}$ and $C = \pi^{-p}$. The density generator $g_{\mathbf{z}}$ allows to describe heavier or lighter tailed distributions (see [5] and Section V for more examples of CES distributions). These CES distributions will be denoted as $\mathcal{CES}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g_{\mathbf{z}}\right)$. CES-distributed vectors have the following stochastic representation

$$\mathbf{z} \overset{d}{=} \sqrt{\mathcal{Q}}\mathbf{A}\mathbf{u} + \boldsymbol{\mu} \tag{2}$$

where $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^H$, $\mathbf{u}$ is uniformly distributed on the complex sphere $\mathcal{U}_1^p$, and $\mathcal{Q}$ is a non-negative real random variable, called the modular variate, independent of $\mathbf{u}$ with a PDF depending only on $g_{\mathbf{z}}$. For the rest of the paper, we will focus on the case of known mean, which allows to set $\boldsymbol{\mu} = \mathbf{0}$ without loss of generality.

B. M-estimators of the scatter matrix

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

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

where u is any real-valued weight function on $[0,\infty)$ that respects Maronna's conditions, ensuring existence and uniqueness of Eq. (3) [1]. When $u(t) = -g_{\mathbf{z}}'(t)/g_{\mathbf{z}}(t)$, Eq. (3) corresponds to the MLE of the scatter matrix parameter for $\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \boldsymbol{\Sigma}, g_{\mathbf{z}})$. However, u may not be related to $g_{\mathbf{z}}$ in practice. Despite this potential mismatch, M-estimators ensure good performance in terms of estimation accuracy on the whole CES family (formally characterized in the following sections). Additionally, M-estimators present robustness to contamination by outliers, which is why they are also usually referred to as robust estimators. Below are listed some examples of M-estimators that will be used through the paper.

Example II.1 (SCM [36]) *The sample covariance matrix (SCM) is given by*

$$\widehat{\boldsymbol{\Sigma}}_{\text{SCM}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i\mathbf{z}_i^H. \tag{4}$$

The SCM can be considered as a "limit case" of Eq. (3) *when* $u(\mathbf{z}_i^H\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{z}_i) = 1$. *This estimator corresponds to the MLE in the Gaussian case. Note that for the SCM, Eq.* (3) *becomes explicit which makes this estimator very convenient for statistical analysis. Indeed, for* $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$, *the SCM follows a Wishart distribution with well-known properties [9]. However, the SCM is not robust and can perform poorly in comparison to* M*-estimators in the CES framework or in the context of contaminated data.*

Example II.2 (Tyler's M-estimator [2]) *Tyler's* M*-estimator is given as the solution of the following equation*

$$\widehat{\boldsymbol{\Sigma}}_T = \frac{p}{n}\sum_{i=1}^{n} \frac{\mathbf{z}_i\mathbf{z}_i^H}{\mathbf{z}_i^H\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{z}_i}. \tag{5}$$

In order to provide a unique solution, the trace of this equation is usually normalized giving the estimation of so-called shape matrix.

Example II.3 (Student's M-estimator) *Student's* M*-estimator is an MLE for Student's* t*-distribution. It is given as the solution of the following equation*

$$\widehat{\boldsymbol{\Sigma}}_t = \frac{p+d/2}{n}\sum_{i=1}^{n} \frac{\mathbf{z}_i\mathbf{z}_i^H}{\mathbf{z}_i^H\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{z}_i + d/2}, \tag{6}$$

where d *is number of degrees of freedom (DoF). When* $d \to \infty$ *the Student's* t*-distribution yields the Gaussian distribution and the Student's* M*-estimator tends to the SCM* ($u(t) \to 1$). *On the other hand, for* $d = 0$ *Student's* M*-estimator is equivalent to Tyler's one.*

C. Standard Asymptotic Regime

For $\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \boldsymbol{\Sigma}, g_{\mathbf{z}})$, denote $\boldsymbol{\Sigma}_\sigma$ the theoretical scatter matrix M-functional, which is defined as a solution of

$$\mathbb{E}[u(\mathbf{z}^H\boldsymbol{\Sigma}_\sigma^{-1}\mathbf{z})\mathbf{z}\mathbf{z}^H] = \boldsymbol{\Sigma}_\sigma. \tag{7}$$

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

$$\mathbb{E}[\Psi(\sigma t)] = p \tag{8}$$

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

Theorem II.1 (Standard asymptotic [5], [11]) *Let* $\widehat{\boldsymbol{\Sigma}}$ *be an* M*-estimator as in Eq.* (3) *built from* n *samples drawn as* $\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \boldsymbol{\Sigma}, g_{\mathbf{z}})$. *The asymptotic distribution of* $\widehat{\boldsymbol{\Sigma}}$ *is given by as*

$$\sqrt{n}\,\text{vec}\left(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_\sigma\right) \xrightarrow{d} \mathcal{GCN}(\mathbf{0}, \mathbf{C}, \mathbf{P}),$$

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

$$\begin{cases} \mathbf{C} = \vartheta_1\boldsymbol{\Sigma}_\sigma^T \otimes \boldsymbol{\Sigma}_\sigma + \vartheta_2 \text{vec}(\boldsymbol{\Sigma}_\sigma)\,\text{vec}(\boldsymbol{\Sigma}_\sigma)^H, \\ \mathbf{P} = \vartheta_1\left(\boldsymbol{\Sigma}_\sigma^T \otimes \boldsymbol{\Sigma}_\sigma\right)\mathbf{K} + \vartheta_2 \text{vec}(\boldsymbol{\Sigma}_\sigma)\,\text{vec}(\boldsymbol{\Sigma}_\sigma)^T. \end{cases} \tag{9}$$

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

$$\begin{array}{l} \vartheta_1 = c_M^{-2} a_M p(p+1), \\ \vartheta_2 = (c_M - p^2)^{-2}(a_M - p^2) - c_M^{-2} a_M (p+1), \end{array} \tag{10}$$

where $a_M = E[\Psi^2(\sigma\mathcal{Q})]$ *and* $c_M = E[\Psi'(\sigma\mathcal{Q})\sigma\mathcal{Q}] + p^2$.

D. Gaussian-Core Wishart Equivalent (GCWE)

First, let us define two quantities related to the hidden Gaussian cores of CES vectors.

Definition II.1 (Gaussian cores [16]) *Let* $\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \boldsymbol{\Sigma}, g_{\mathbf{z}})$. *This vector has a representation analogous to Eq.* (2), *given as*

$$\mathbf{z} \overset{d}{=} \sqrt{\mathcal{Q}}\mathbf{A}\mathbf{g}/\|\mathbf{g}\|, \tag{11}$$

where $\mathbf{g} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. *The vector* $\mathbf{x} = \mathbf{A}\mathbf{g}$ *is referred to as the Gaussian-core of* $\mathbf{z}$.

Definition II.2 (GCWE [16]) *Let* n *measurements* $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ *be drawn as* $\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \boldsymbol{\Sigma}, g_{\mathbf{z}})$ *and denote* $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ *their Gaussian cores as* $\mathbf{z}_i = \sqrt{\mathcal{Q}_i}/\|\mathbf{x}_i\|\mathbf{A}\mathbf{x}_i$ *(cf. Definition II.1). Let* $\widehat{\boldsymbol{\Sigma}}$ *be an* M*-estimator built with* $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ *using Eq.* (3). *The SCM built from the Gaussian cores, i.e.*

$$\widehat{\boldsymbol{\Sigma}}_{\text{GCWE}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i\mathbf{x}_i^H \tag{12}$$

is referred to as Gaussian Core Wishart Equivalent (GCWE) of $\widehat{\boldsymbol{\Sigma}}$.

Note that the GCWE can not be computed in practice, but represents a theoretical, Wishart distributed, quantity. The asymptotic distribution of the difference between an M-estimator and its GCWE is derived in [16].

Theorem II.2 (Asymptotic GCWE [16]) *Let* $\widehat{\boldsymbol{\Sigma}}$ *be an* M*-estimator as in Eq.* (3) *built from* n *samples drawn as* $\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \boldsymbol{\Sigma}, g_{\mathbf{z}})$, $\widehat{\boldsymbol{\Sigma}}_{\text{GCWE}}$ *be its GCWE from Definition II.2 and* σ *be the solution of Eq.* (8). *Then, the asymptotic distribution of* $\sigma\widehat{\boldsymbol{\Sigma}} - \widehat{\boldsymbol{\Sigma}}_{\text{GCWE}}$ *is given by [16]*

$$\sqrt{n}\,\text{vec}\left(\sigma\widehat{\boldsymbol{\Sigma}} - \widehat{\boldsymbol{\Sigma}}_{\text{GCWE}}\right) \xrightarrow{d} \mathcal{GCN}\left(\mathbf{0}, \widetilde{\mathbf{C}}, \widetilde{\mathbf{P}}\right), \tag{13}$$

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

$$\begin{cases} \widetilde{\mathbf{C}} = \sigma_1\boldsymbol{\Sigma}^T \otimes \boldsymbol{\Sigma} + \sigma_2\text{vec}(\boldsymbol{\Sigma})\text{vec}(\boldsymbol{\Sigma})^H, \\ \widetilde{\mathbf{P}} = \sigma_1\left(\boldsymbol{\Sigma}^T \otimes \boldsymbol{\Sigma}\right)\mathbf{K} + \sigma_2\text{vec}(\boldsymbol{\Sigma})\text{vec}(\boldsymbol{\Sigma})^T, \end{cases} \tag{14}$$

with σ_1 *and* σ_2 *given by*

$$\begin{array}{rcl} \sigma_1 & = & \left(a_M p(p+1) + c(c - 2b_M)\right)/c_M^2, \\ \sigma_2 & = & \vartheta_2 - 2p(b_M - c_M)/c_M/(c_M - p^2), \end{array} \tag{15}$$

where a_M *and* c_M *are defined in Theorem II.1, and* $b_M = E[\Psi(\sigma\mathcal{Q})\|\mathbf{g}\|^2]$.

III. Asymptotics of M-estimators' eigenvalue decomposition

The EigenValue Decomposition (EVD) of the scatter matrix $\mathbf{\Sigma}$ is denoted as

$$\mathbf{\Sigma} \overset{\text{EVD}}{=} \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H \quad \text{with} \quad \begin{array}{l} \mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_p] \in \mathcal{U}_p^p, \\ \mathbf{\Lambda} = \operatorname{diag}(\boldsymbol{\lambda}), \\ \boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_p]. \end{array} \tag{16}$$

In order to avoid ambiguity in this definition, we assume ordered eigenvalues as $\lambda_1 > \ldots > \lambda_p > 0$, and an element of each $\mathbf{u}_j$ (e.g., the first entry) for $j = 1, \ldots, n$, can be assumed to be real positive. Similarly, we define matching notations for the EVD of an M-estimator $\widehat{\mathbf{\Sigma}}$ and its GCWE $\widehat{\mathbf{\Sigma}}_{\text{GCWE}}$ as

$$\begin{aligned} &\widehat{\mathbf{\Sigma}} \overset{\text{EVD}}{=} \widehat{\mathbf{U}}^M \widehat{\mathbf{\Lambda}}^M \left(\widehat{\mathbf{U}}^M\right)^H, \\ &\widehat{\mathbf{\Sigma}}_{\text{GCWE}} \overset{\text{EVD}}{=} \widehat{\mathbf{U}}^{\text{GCWE}} \widehat{\mathbf{\Lambda}}^{\text{GCWE}} \left(\widehat{\mathbf{U}}^{\text{GCWE}}\right)^H. \end{aligned} \tag{17}$$

In the following we derive the asymptotic distributions for the quantities $\widehat{\mathbf{U}}^M$ and $\widehat{\mathbf{\Lambda}}^M$, both under standard and GCWE regimes.

Theorem III.1 (Standard asymptotic) *Let $\widehat{\mathbf{\Sigma}}$ be an M-estimator as in Eq.* (3) *built from n samples drawn as $\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, g_{\mathbf{z}})$ and σ be the solution of Eq.* (8)*. The asymptotic distribution of the EVD of $\widehat{\mathbf{\Sigma}}$ (Eq.* (17)*) is given by*

$$\begin{cases} \sqrt{n}\left(\sigma\widehat{\boldsymbol{\lambda}}^M - \boldsymbol{\lambda}\right) \overset{d}{\to} \mathcal{N}\left(\mathbf{0}, \vartheta_1 \mathbf{\Lambda}^2 + \vartheta_2 \boldsymbol{\lambda}\boldsymbol{\lambda}^T\right), \\ \sqrt{n}\mathbf{\Pi}_j^{\perp}\widehat{\mathbf{u}}_j^M \overset{d}{\to} \mathcal{CN}\left(\mathbf{0}, \mathbf{\Xi}_j\right). \end{cases} \tag{18}$$

where

$$\mathbf{\Xi}_j = \vartheta_1 \lambda_j \left(\mathbf{U}\mathbf{\Lambda}(\lambda_j \mathbf{I} - \mathbf{\Lambda})^+\right)^2 \mathbf{U}^H \tag{19}$$

with $\mathbf{\Pi}_j^{\perp} = \mathbf{I} - \mathbf{u}_j\mathbf{u}_j^H$ and ϑ_1, ϑ_2 given by Eq. (10).

Proof: See Appendix A. ∎

Theorem III.2 (Asymptotic GCWE) *Let $\widehat{\mathbf{\Sigma}}$ be an M-estimator as in Eq.* (3) *built from n samples drawn as $\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, g_{\mathbf{z}})$, $\widehat{\mathbf{\Sigma}}_{\text{GCWE}}$ be its GCWE (Definition II.2) and σ be the solution of Eq.* (8)*. The asymptotic distribution of the difference between the EVD parameters of $\widehat{\mathbf{\Sigma}}$ and $\widehat{\mathbf{\Sigma}}_{\text{GCWE}}$ is given by*

$$\begin{cases} \sqrt{n}\left(\sigma\widehat{\boldsymbol{\lambda}}^M - \widehat{\boldsymbol{\lambda}}^{\text{GCWE}}\right) \overset{d}{\to} \mathcal{N}\left(\mathbf{0}, \sigma_1 \mathbf{\Lambda}^2 + \sigma_2 \boldsymbol{\lambda}\boldsymbol{\lambda}^T\right), \\ \sqrt{n}\mathbf{\Pi}_j^{\perp}\left(\widehat{\mathbf{u}}_j^M - \widehat{\mathbf{u}}_j^{\text{GCWE}}\right) \overset{d}{\to} \mathcal{CN}\left(\mathbf{0}, \sigma_1/\vartheta_1 \mathbf{\Xi}_j\right). \end{cases} \tag{20}$$

with $\mathbf{\Xi}_j$ and σ_1, σ_2 given by Eqs. (19) *and* (15)*, 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, Theorem III.1 shows 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 ϑ_2. Concerning the eigenvectors, note that the covariance depends only on ϑ_1 since $\mathbf{u}_j$ is scale invariant w.r.t. to the covariance matrix (see [11] for more details).*

Remark III.2 *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 the one in the standard asymptotic regime, and differs only through the variance scales (σ_1, σ_2) (instead of $(\vartheta_1, \vartheta_2)$). As noted in [16], the total variance captured by the GCWE factors is much smaller than the standard one. For example, the factors σ_1 and ϑ_1 given in Table I for the Student t-distribution differ by an order $1/p$. This result supports the idea that an underlying Wishart distribution can offer a better approximation for characterizing the distribution of the M-estimator's EVD. This approximation allows us to rely on well established results [9], [31], and offers a thinner analysis compared to the asymptotic Gaussian results. Some applicative examples illustrate this point in Section VI.*

IV. Asymptotics of M-estimators' principal subspace

Consider the case of a low-rank plus identity scatter matrix (also referred to as factor model), that is commonly used in signal processing to account for low dimensional signals embedded in white noise

$$\mathbf{\Sigma} = \mathbf{\Sigma}_r + \gamma^2 \mathbf{I}_p \quad \overset{\text{EVD}}{=} \quad [\mathbf{U}_r | \mathbf{U}_r^{\perp}]\mathbf{\Lambda}[\mathbf{U}_r | \mathbf{U}_r^{\perp}]^H, \tag{21}$$

with the rank r matrix $\mathbf{\Sigma}_r = \mathbf{U}_r \mathbf{\Lambda}_r \mathbf{U}_r^H$, with $\mathbf{U}_r \in \mathcal{U}_r^p$ and $\mathbf{\Lambda}_r \in \mathbb{R}^{r \times r}$.

This section focuses on the estimation of the orthogonal projector onto the range space spanned by $\mathbf{U}_r$, the r strongest eigenvectors of $\mathbf{\Sigma}$ (referred to as "principal subspace" in the following). We define $\mathcal{R}_r\{.\}$ the operator that extracts this principal subspace from a given matrix, i.e.,

$$\begin{array}{llll} \mathcal{R}_r: & \mathcal{H}_M^+ & \longrightarrow & \mathcal{G}_r^n, \\ & \mathbf{\Sigma} \overset{\text{EVD}}{=} [\mathbf{U}_r | \mathbf{U}_r^{\perp}]\mathbf{\Lambda}[\mathbf{U}_r | \mathbf{U}_r^{\perp}]^H & \longmapsto & \mathbf{U}_r \mathbf{U}_r^H, \end{array} \tag{22}$$

where $\mathcal{G}_r^n$ is the set of rank r orthogonal projectors of $\mathbb{C}^{n \times n}$.

Let us consider an M-estimator $\widehat{\mathbf{\Sigma}}$ built with an n samples drawn as $\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma})$ where $\mathbf{\Sigma}$ is low-rank structured as in Eq. (21), and let $\widehat{\mathbf{\Sigma}}_{\text{GCWE}}$ be its GCWE from Definition II.2. We have the corresponding principal subspaces as

$$\begin{aligned} \mathbf{\Pi}_r &= \mathcal{R}_r\{\mathbf{\Sigma}\}, \\ \widehat{\mathbf{\Pi}}_r^M &= \mathcal{R}_r\{\widehat{\mathbf{\Sigma}}\}, \\ \widehat{\mathbf{\Pi}}_r^{\text{GCWE}} &= \mathcal{R}_r\{\widehat{\mathbf{\Sigma}}_{\text{GCWE}}\}. \end{aligned} \tag{23}$$

In the following we derive the asymptotic distributions for the quantities $\widehat{\mathbf{\Pi}}^M$, both under standard and GCWE regimes.

Theorem IV.1 (Standard asymptotic) *Let $\widehat{\mathbf{\Pi}}_r^M$ be the estimator of the projector $\mathbf{\Pi}_r$ obtained from an M-estimator (Eq. (23)). The asymptotic distribution of $\widehat{\mathbf{\Pi}}_r^M$ is given by*

$$\sqrt{n}\text{vec}\left(\widehat{\mathbf{\Pi}}_r^M - \mathbf{\Pi}_r\right) \xrightarrow{d} \mathcal{GCN}\left(\mathbf{0}, \vartheta_1 \mathbf{\Sigma}_{\mathbf{\Pi}}, \vartheta_1 \mathbf{\Sigma}_{\mathbf{\Pi}} \mathbf{K}\right), \quad (24)$$

where

$$\mathbf{\Sigma}_{\mathbf{\Pi}} = \mathbf{A}^T \otimes \mathbf{B} + \mathbf{B}^T \otimes \mathbf{A} \quad (25)$$

with $\mathbf{A} = \mathbf{U}_r\left(\gamma^2 \mathbf{\Lambda}_r^{-2} + \mathbf{\Lambda}_r^{-1}\right)\mathbf{U}_r^H$, $\mathbf{B} = \gamma^2 \mathbf{\Pi}_r^{\perp}$ and ϑ_1, ϑ_2 given by Eq. (10).

Proof: See Appendix C. ∎

Theorem IV.2 (Asymptotic GCWE) *Let $\widehat{\mathbf{\Pi}}_r^M$ and $\widehat{\mathbf{\Pi}}_r^{\text{GCWE}}$ be the estimators of the projector $\mathbf{\Pi}_r$ defined in Eq.* (23)*. The asymptotic distribution of $\widehat{\mathbf{\Pi}}_r^M$ is given by*

$$\sqrt{n}\text{vec}\left(\widehat{\mathbf{\Pi}}_r^M - \widehat{\mathbf{\Pi}}_r^{\text{GCWE}}\right) \xrightarrow{d} \mathcal{GCN}\left(\mathbf{0}, \sigma_1 \mathbf{\Sigma}_{\mathbf{\Pi}}, \sigma_1 \mathbf{\Sigma}_{\mathbf{\Pi}} \mathbf{K}\right) \quad (26)$$

with $\mathbf{\Sigma}_{\mathbf{\Pi}}$ and σ_1, σ_2 given by Eqs. (25) *and* (15)*, respectively.*

Proof: See Appendix D. ∎

	Student's M-estimator	Tyler's M-estimator
SA	$\vartheta_1 = \frac{p + d/2 + 1}{p + d/2}$	$\vartheta_1 = \frac{p+1}{p}$
SA	$\vartheta_2 = \frac{2}{d} \times \frac{p + d/2 + 1}{p + d/2}$	$\vartheta_2 = -\frac{p+1}{p^2}$
GCWE	$\sigma_1 = \frac{1}{p + d/2}$	$\sigma_1 = \frac{1}{p}$
GCWE	$\sigma_2 = \frac{2}{d} \times \frac{p + d/2 + 1}{p + d/2}$	$\sigma_2 = \frac{p-1}{p^2}$

Table I: Coefficients ϑ_1, ϑ_2, σ_1 and σ_2 for Student's and Tyler's M-estimator with t-distributed data. SA stands for Standard asymptotic while GCWE refers to as GCWE asymptotic

Remark IV.1 *Theorem IV.1 (*resp. *IV.2) extends the results of Theorem III.1 (*resp. *III.2) to the principal subspace of M-estimators. We can draw the same conclusions as in Remark III.2, notably, that an underlying Wishart equivalent offers a more accurate asymptotic equivalent than the standard Gaussian one.*

V. Validation simulations

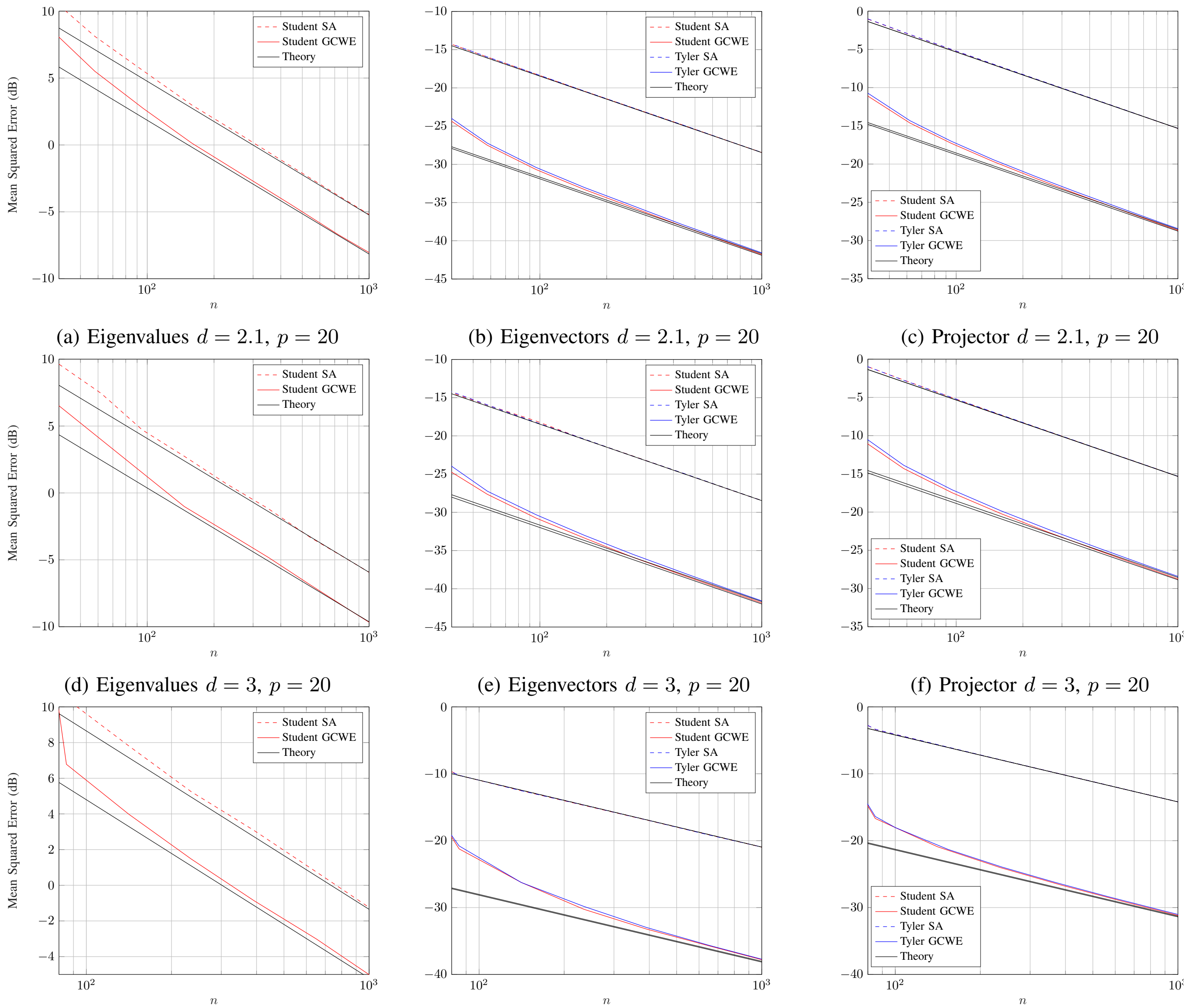
The theoretical results are validated through Monte-Carlo simulations for Student t-distributed data with the DoF parameter d, whose PDF is given by Eq. (1) with

$$g_{\mathbf{z}}(x) = (1 + 2x/d)^{-(p+d/2)} \quad (27)$$

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. (12) for $\mathcal{Q} \sim pF_{2p,d}$ (a Fisher distribution with DoFs $2p$ and d). This distribution has finite 2^{nd}-order moments for $d > 2$. If not specified, the scatter matrix is Toeplitz with entries $[\mathbf{\Sigma}]_{i,j} = \rho^{(j-i)}$ for $i \leq j$ and $[\mathbf{\Sigma}]_{i,j} = [\mathbf{\Sigma}]_{j,i}^*$ for $i > j$, $i,j = 1, \ldots, p$, with $\rho = 0.9(1 + \sqrt{-1})/\sqrt{2}$. When referring to a low rank model, the scatter matrix is constructed as $\mathbf{\Sigma} = \mathbf{\Sigma}_r + \mathbf{I}$ with $\mathbf{\Sigma}_r = \mathbf{U}_r \mathbf{\Lambda}_r \mathbf{U}^H$ where only the 5 first eigenvalues and eigenvectors of $\mathbf{\Sigma}$ are kept, and scaled so that $\text{Tr}(\mathbf{\Sigma}_r) = 100$. The results are illustrated for the Student's M-estimator (Eq. (6)) and Tyler's M-estimator (Eq. (5)). The parameters for the asymptotic distributions are given in Table I.

Figure 1 illustrates the validation of theoretical results. The first (resp. second) row displays the results for Student t-distributed data with $d = 2.1$ (resp. $d = 3$) and $p = 20$. The third row shows the results for $d = 3$ and a larger dimension $p = 50$. The values of d were chosen in that way to ensure the existence of finite 2^{nd}-order moments and to show the behavior of the data when the parameter d increases.

• The first column displays the empirical mean squared error (MSE) of $\widehat{\boldsymbol{\lambda}}^t - \boldsymbol{\lambda}$ and $\widehat{\boldsymbol{\lambda}}^t - \widehat{\boldsymbol{\lambda}}^{\text{GCWE}}$ as well as their corresponding and its theoretical values, i.e., $\text{Tr}(\vartheta_1 \mathbf{\Lambda}^2 + \vartheta_2 \boldsymbol{\lambda}\boldsymbol{\lambda}^{\text{T}})/\text{n}$ (Eq. (18)) and $\text{Tr}(\sigma_1 \mathbf{\Lambda}^2 + \sigma_2 \boldsymbol{\lambda}\boldsymbol{\lambda}^{\text{T}})/\text{n}$ (Eq. (20)). Note that the results are not presented for Tyler's M-estimator due to its inherent scaling ambiguity. First, we note that the empirical results tend to the corresponding theoretical ones as n increases. We clearly observe that the GCWE equivalent has lower variance than the standard one. Thus, these results support the idea that the distribution of the eigenvalues of an M-estimator (in this case Student's M-estimator) can be well-approximated with the one of their GCWE. In addition, one can note that this difference is slightly larger for $d = 3$, as expected. We additionally note that the results are also valid for higher data dimension.
• The second column displays the corresponding quantities for the first eigenvector for both Student's and Tyler's M-estimator. We observe that the empirical MSEs coincide well with their theoretical counterparts in both regimes. Moreover, the figure shows a significant difference between the results for the standard regime and GCWE. The eigenvectors are scale-invariant functions of the scatter matrix, so their asymptotic variance only involve factors of σ_1 (resp. ϑ_1) and not σ_2 (resp. ϑ_2). For the presented M-estimator, we have $\sigma_1 \ll \vartheta_1$ (especially when the data dimension grows) which explains these results. Moreover, one can notice that, as expected, Tyler's M-estimator is closer to Student's M-estimator for a smaller degree of freedom d and the same value of p (figure 1.(b) versus figure 1.(e)).
• Finally, the last column presents the MSE for the projector defined by Eq. (23) in the structured model. Again, the figures validate the theoretical results of Theorems IV.1 and IV.2 and leads us to the same conclusions as previously.



(a) Eigenvalues $d = 2.1$, $p = 20$
(b) Eigenvectors $d = 2.1$, $p = 20$
(c) Projector $d = 2.1$, $p = 20$
(d) Eigenvalues $d = 3$, $p = 20$
(e) Eigenvectors $d = 3$, $p = 20$
(f) Projector $d = 3$, $p = 20$
(g) Eigenvalues $d = 3$, $p = 50$
(h) Eigenvectors $d = 3$, $p = 50$
(i) Projector $d = 3$, $p = 50$

Figure 1: Validation of theoretical results for Student's and Tyler's M-estimator built with Student t-distributed data with DoF d. From left to right: results for eigenvalues, eigenvectors and projector. From top to bottom: results for $d = 2.1$ and $p = 20$, $d = 3$ and $p = 20$, $d = 3$ and $p = 50$.

VI. Applications

In this section, we give some examples where the derived results can be leveraged. The main result, consistent with other sections, is that the distribution of an M-estimator's EVD can be accurately approached by the one of the underlying Wishart model. This approximation allows us to obtain theoretical derivations in the general CES/M-estimator's framework, where true distributions are, a priori, not tractable (e.g. in the intrinsic bias analysis in Section VI-A). On a second note, this also reinforces the standard "plug-in" approach of M-estimators in criteria/processes derived using the EVD parameters under the Wishart assumption. Such type of application is illustrated for the eigenvalues in Section VI-B (rank estimation) and the principal subspace in Section VI-C (low-rank filtering).

A. *Intrinsic bias analysis*

In [32] were derived Intrinsic (i.e. Riemannian Manifold oriented) counterparts of the Cramér-Rao inequality (ICRLB). In the context of covariance matrix estimation, these results allows notably to account for the natural Riemmanian metric on $\mathcal{H}_M^{++}$, and bound the expected Riemannian distance (rather than the Euclidean one)

$$d^2_{nat}\left(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2\right) = \left\| \ln\left(\boldsymbol{\Sigma}_1^{-1/2} \boldsymbol{\Sigma}_2 \boldsymbol{\Sigma}_1^{-1/2} \right) \right\|_F^2 . \quad (28)$$

This analysis also reveals unexpected and hidden properties of estimators, such as a bias of the SCM w.r.t. the natural metric on $\mathcal{H}_M^{++}$ (that does not exists w.r.t. the Euclidean metric). In this scope, the biased ICRLB (B-ICRLB) is established for the SCM in a Gaussian context in [32, Theorem 7 and Corollary 5], and reads as follows.

Theorem VI.1 (B-ICRLB for SCM [32]) *Let* $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ *be an* n*-sample distributed as* $\mathbf{z}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma})$ *and* $\widehat{\mathbf{\Sigma}}_{\text{SCM}}$ *be the SCM as in Eq.* (4)*. The bias w.r.t. the natural metric on* $\mathcal{H}_M^{++}$ *of* $\widehat{\mathbf{\Sigma}}_{\text{SCM}}$ *is*

$$\mathbb{E}\left[\exp_{\mathbf{\Sigma}}^{-1} \widehat{\mathbf{\Sigma}}_{\text{SCM}}\right] = -\eta(p,n)\mathbf{\Sigma} \tag{29}$$

with $\exp_{\mathbf{\Sigma}}^{-1} \widehat{\mathbf{\Sigma}}_{\text{SCM}} = \mathbf{\Sigma}^{1/2} \log\left(\mathbf{\Sigma}^{-1/2}\widehat{\mathbf{\Sigma}}_{\text{SCM}}\mathbf{\Sigma}^{-1/2}\right)\mathbf{\Sigma}^{1/2}$, *and*

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

where $\psi(x) = \Gamma'(x)/\Gamma(x)$ *is the digamma function. Moreover, the natural distance Eq.* (28) *between* $\widehat{\mathbf{\Sigma}}_{\text{SCM}}$ *and* $\mathbf{\Sigma}$ *satisfies the following biased-ICRLB inequality:*

$$\mathbb{E}\left[d_{nat}^2\left(\hat{\mathbf{\Sigma}}_{\text{SCM}}, \mathbf{\Sigma}\right)\right] \geq \frac{p^2}{n} + p\eta\left(p,n\right)^2. \tag{31}$$

For CES-distributed samples, the ICRLB on d_{nat}^2 is derived in [37] as follows:

Theorem VI.2 (IRCLB for CES [37]) *Let* $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ *be an* n*-sample distributed as* $\mathbf{z}_i \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, g_{\mathbf{z}})$*. Any unbiased estimator* $\widehat{\mathbf{\Sigma}}$ *of* $\mathbf{\Sigma}$ *satisfies the inequality*

$$\mathbb{E}\left[d_{nat}^2\left(\widehat{\mathbf{\Sigma}}, \mathbf{\Sigma}\right)\right] \geq \frac{p^2-1}{n\alpha} + (n(\alpha + p\beta))^{-1}, \tag{32}$$

with $\alpha = \left(1 + \frac{\mathbb{E}\left[\mathcal{Q}^2 u'(\mathcal{Q})\right]}{p(p+1)}\right)$ *(where* $\mathcal{Q}$ *is the modular variate as in Eq.* (2)*) and* $\beta = \alpha - 1$.

Characterizing a bias term in Theorem VI.2 (similarly to the one in Theorem VI.1) would require us 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 exact distribution of the eigenvalues of a Wishart-distributed matrix, and cannot be recovered through a Delta method using the standard asymptotic of Theorem III.1. However the strong proximity of the eigenvalues of an M-estimator towards their GCWE described in Theorem III.2 (also exhibited by the previous simulation results) gives a reasonable theoretical ground for the following approximation:

Approximation VI.1 (Intrinsic bias of M-estimators) *Let* $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ *be an* n*-sample distributed as* $\mathbf{z}_i \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, g_{\mathbf{z}})$*. Let* $\widehat{\mathbf{\Sigma}}$ *be an* M*-estimator of* $\mathbf{\Sigma}$ *that is consistent in scale (i.e.,* $\sigma = 1$ *in Eq.* (8)*) and* $\widehat{\mathbf{\Sigma}}_{\text{GCWE}}$ *its GCWE (Definition II.2). The matrix* $\widehat{\mathbf{\Sigma}}_{\text{GCWE}}$ *is Wishart distributed with center* $\mathbf{\Sigma}$*, so the derivations of Theorem 7 of [32] directly apply to its intrinsic bias. Theorem III.2 then supports the approximation*

$$\begin{aligned} \mathbb{E}[\exp_{\mathbf{\Sigma}}^{-1} \widehat{\mathbf{\Sigma}}] &\simeq \mathbb{E}[\exp_{\mathbf{\Sigma}}^{-1} \widehat{\mathbf{\Sigma}}_{\text{GCWE}}] \\ &= -\eta(p,n)\mathbf{\Sigma}. \end{aligned} \tag{33}$$

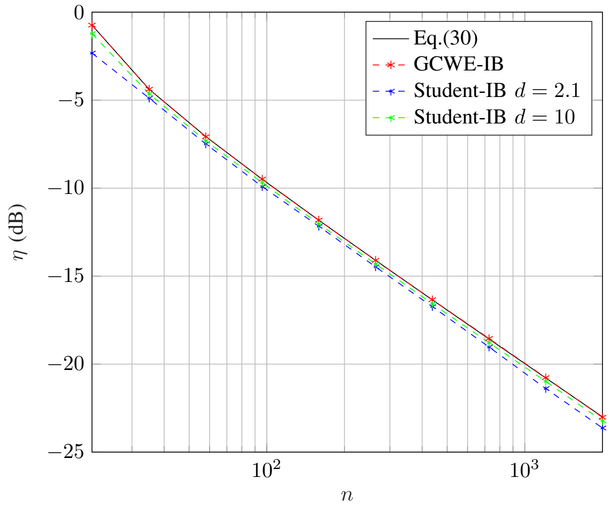


Figure 2: Empirical intrinsic bias for Student's M-estimator (Student-IB) and the Gaussian core SCM (GCWE-IB) compared to the theoretical result obtained for the GCWE (Eq. (30))

Figure 2 confirms the previous results and supports the proposed approximation. Indeed, it can be seen that the empirical intrinsic bias obtained with the Student's M-estimator computed with t-distributed data coincides with the intrinsic bias based on the GCWE and the theoretical result (Eq. (30)).

Finally, we propose to incorporate an equivalent bias term in Eq. (32) to obtain an accurate approximation of the B-ICRLB of M-estimators build form CES-distributed samples.

Approximation VI.2 (B-ICRLB for CES) *Let* $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ *be an* n*-sample distributed as* $\mathbf{z}_i \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, g_{\mathbf{z}})$*. Let* $\widehat{\mathbf{\Sigma}}$ *be an* M*-estimator of* $\mathbf{\Sigma}$ *that is consistent in scale (i.e.,* $\sigma = 1$ *in Eq.* (8)*). We have the following approached B-ICRLB*

$$\mathbb{E}\left[d_{nat}^2\left(\widehat{\mathbf{\Sigma}}, \mathbf{\Sigma}\right)\right] \geq \frac{p^2-1}{n\alpha} + (n(\alpha + p\beta))^{-1} + p\eta\left(p,n\right)^2, \tag{34}$$

with α *and* β *defined in VI.2 and* $\eta\left(p,n\right)$ *from Eq.* (30)*).*

Figure 3 illustrates this approximation. The empirical mean of the natural Riemannian distance of the Student's M-estimator $\widehat{\mathbf{\Sigma}}_t$ (denoted as $\epsilon^N\left(\widehat{\mathbf{\Sigma}}_t\right)$) is compared to the theoretical ICRLB in Eq. (32)) and the approached B-ICRLB in Eq. (34). As expected, one can see that the approached

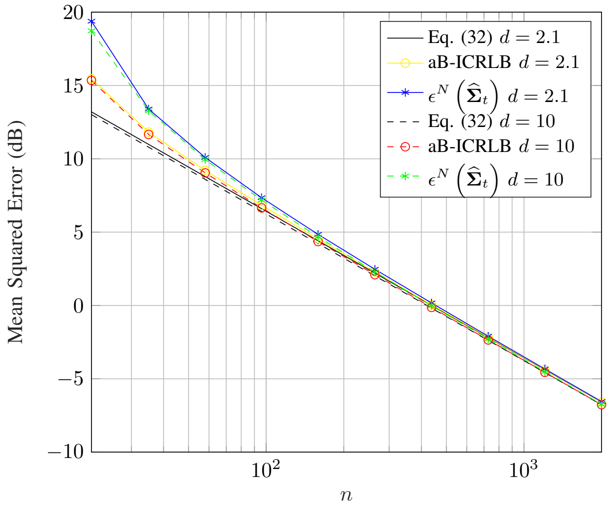


Figure 3: Empirical mean of $d^2_{\text{nat}}\left(\widehat{\boldsymbol{\Sigma}}_t, \boldsymbol{\Sigma}\right)$ denoted as $\epsilon^N\left(\widehat{\boldsymbol{\Sigma}}_t\right)$ versus theoretical CRLB for an unbiased estimator in the CES framework (Eq. (32)) and approximated biased instrinsic CRLB (aB-ICRLB)

bias term in B-ICRLB offers a more accurate theoretical approximation for bounding the expected natural distance.

B. Rank estimation

The rank estimation, or more generally the model order selection, is an important problem in data analysis. It consists in determining r when the covariance of the data is low-rank structured as in Eq. (21). In the context of Gaussian-distributed samples, several rank estimators have been proposed as functions of the eigenvalues of the SCM $\widehat{\boldsymbol{\lambda}}^{\text{SCM}} = [\widehat{\lambda}_1, \ldots, \widehat{\lambda}_p]$ [21]. Notably, we can cite two of the most commonly used:

1) The Akaike Information Criterion (AIC) [33], that minimizes the following criterion

$$\widehat{r}_{\text{AIC}} = \underset{k \in [0, p-1]}{\text{argmin}} \left[nl \times ic(\widehat{\boldsymbol{\lambda}}^{\text{SCM}}) + k(l+p) \right] \tag{35}$$

with $l = p - k$ and

$$ic(\widehat{\boldsymbol{\lambda}}^{\text{SCM}}) = \ln \left(\prod_{i=k}^{p-1} \widehat{\lambda}_i^{\frac{1}{l-1}} / \tfrac{1}{l} \sum_{i=k}^{p-1} \widehat{\lambda}_i \right). \tag{36}$$

2) The Minimum Description Length (MDL) [34] (also referred to as Bayesian Information Criterion (BIC)), that minimizes the following criterion

$$\widehat{r}_{\text{MDL}} = \underset{k \in [0, p-1]}{\text{argmin}} \left[2nl \times ic(\widehat{\boldsymbol{\lambda}}^{\text{SCM}}) + k(l+p) \ln n \right]. \tag{37}$$

In the context of CES-distributed samples, a plug-in approach using $ic(\widehat{\boldsymbol{\lambda}}^{\text{M}})$ (computed from the EVD of an M-estimator) in Eq. (35) or Eq. (37) can be envisioned rather that re-deriving information criterions assuming non-Gaussian samples. This approach is motivated by the fact that $\widehat{\boldsymbol{\lambda}}^{M}$ quickly converges to eigenvalues of the equivalent Wishart model (cf. Theorem III.2 and Remark III.2), i.e. the problem can be processed as if the data were initially Gaussian.

To illustrate this point, we consider the problem of determining how many sources are observed by an array of sensors. We assume that a planar array of p sensors observes signals produced by r sources that are centered around a known frequency ω, affecting the array from locations $\theta_1, \ldots, \theta_{n_s}$. The sources-plus-white-noise signal received by the array of sensors can be expressed as

$$\mathbf{z} = \mathbf{A}\mathbf{s} + \mathbf{n} \tag{38}$$

with $\mathbf{s}$ the signal vector, $\mathbf{n}$ the additive white noise and $\mathbf{A} = [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_{n_s})]$ and

$$\mathbf{a}(\theta) = \left[a_1(\theta) e^{-j\omega\tau_1(\theta)}, \ldots, a_p(\theta) e^{-j\omega\tau_p(\theta)} \right]^T, \tag{39}$$

where a_j the amplitude response of the j^{th} sensor towards direction θ and $\tau_j(\theta)$ propagation delay between the reference point and the j^{th} sensor. The total covariance matrix has thus a low rank structure as in Eq. (21), and is given by

$$\boldsymbol{\Sigma} = \mathbf{A}\mathbf{S}\mathbf{A}^H + \gamma^2 \mathbf{I}_p, \tag{40}$$

where $\mathbb{E}\left[\mathbf{s}\mathbf{s}^H\right] = \mathbf{S}$ and $\mathbb{E}\left[\mathbf{n}\mathbf{n}^H\right] = \gamma^2 \mathbf{I}_p$.

In the considered problem, the received data $\mathbf{z}$ is CES-distributed with the scatter matrix $\boldsymbol{\Sigma}$ given by Eq. (40). We resort to the proposed plug-in approach to estimate r. Figure 4 shows the values of the AIC and MLE criteria computed with different M-estimators. The data is t-distributed with $d = 2.1$. The number of sensors is set to 20, while the number of sources to estimate is equal to $r = 4$. The number of samples is $n = 200$. A circle indicates the minimum value of each criterion in order to highlight the estimated number r of sources. We observe that Student's and Tyler's M-estimator give an accurate estimation of r. Conversely, the result for the SCM is, as expected, not accurate due to the non-Gaussianity of the observations. Interestingly, the values of the criteria for Student's and Tyler's M-estimator are almost identical to the ones computed with the (theoretical) GCWE, which validates the proposed approach.

C. SNR Loss

In the context of STAP, the covariance of the clutter plus noise $\boldsymbol{\Sigma}$ is low rank structured as in Eq. (21), where the rank r can be evaluated thanks to the Brennan rule [38]. The optimal filter $\mathbf{w}_{opt}$ [39] given by

$$\mathbf{w}_{opt} = \boldsymbol{\Sigma}^{-1}\mathbf{p}, \tag{41}$$

where $\mathbf{p}$ is a known steering vector. In the low-rank clutter case an alternative is to use the low-rank STAP filter $\mathbf{w}_r$ [40], [41] defined as

$$\mathbf{w}_r = \boldsymbol{\Pi}_r^{\perp}\mathbf{p} \tag{42}$$

with $\boldsymbol{\Pi}_r = \mathcal{R}_r\{\boldsymbol{\Sigma}\}$ (cf. Eq. (22)). In practice, adaptive STAP filters are built with an estimate of the covariance matrix

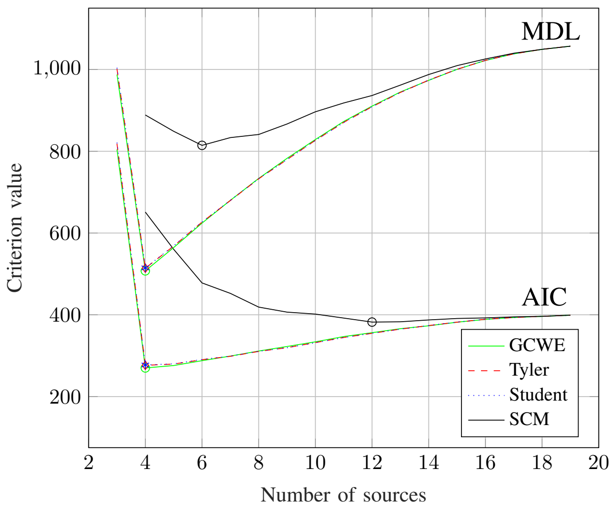


Figure 4: Number of sources estimation with AIC and MDL: Results obtained with Eqs. (35) and (37) for the SCM, Student's M-estimator and Tyler's M-estimator compared to the theoretical GCWE; Student t-distributed data with $d = 2.1$; $p = 20$, $r = 4$.

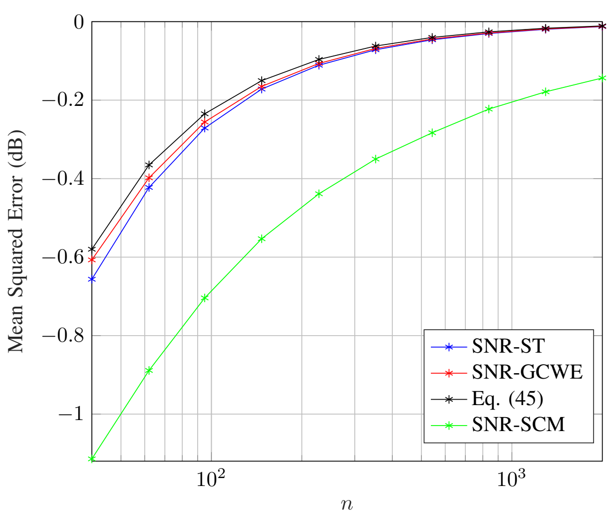


Figure 5: Empirical SNR Loss obtained with the Student's M-estimator (SNR-ST), GCWE (SNR-GCWE) and SCM (SNR-SCM) versus the theoretical result given by Eq. (45); t-distributed data with $p = 20$, $r = 5$, $d = 3$.

$\mathbf{\Sigma}$ or of the projector $\mathbf{\Pi}_r^{\perp}$ computed with secondary data $\mathbf{z}_i \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, \mathbf{g}_{\mathbf{z}})$.

The SNR Loss ρ of an adaptive filter $\widehat{\mathbf{w}}$ is given by

$$\rho = \frac{SNR_{out}}{SNR_{max}} = \frac{|\widehat{\mathbf{w}}^H \mathbf{p}|^2}{(\widehat{\mathbf{w}}^H \mathbf{\Sigma} \widehat{\mathbf{w}})\,(\mathbf{p}^H \mathbf{\Sigma} \mathbf{p})}. \tag{43}$$

For an adaptive low rank filter $\widehat{\mathbf{w}}_r$, this expression becomes

$$\rho = \gamma^2 \frac{\left(\mathbf{p}^H \widehat{\mathbf{\Pi}}_r^{\perp} \mathbf{p}\right)^2}{\mathbf{p}^H \widehat{\mathbf{\Pi}}_r^{\perp} \mathbf{\Sigma} \widehat{\mathbf{\Pi}}_r^{\perp} \mathbf{p}} \tag{44}$$

with a given estimator $\widehat{\mathbf{\Pi}}_r$. In [41] it has been shown that when the data are Gaussian-distributed and $\widehat{\mathbf{\Pi}}_r = \mathcal{R}_r\{\widehat{\mathbf{\Sigma}}_{\text{SCM}}\}$ the expected SNR Loss is given by

$$\mathbb{E}\left[\rho\right] = 1 - r/n. \tag{45}$$

This result can directly provide a good approximation of the expected SNR Loss of adaptive low rank filters build from M-estimators in the general context of CES-distributed samples. Indeed, this approach is again motivated by the fact the estimate $\widehat{\mathbf{\Pi}}_r^M$ is close to $\widehat{\mathbf{\Pi}}_r^{\text{GCWE}}$, i.e., to the principal subspace of the equivalent Wishart model (cf. Theorem IV.2). To illustrate this point, Figure 5 draws a comparison between the SNR Losses of various low rank filters built from t-distributed data. The low rank covariance matrix is build as in Section V. One can notice that the value of SNR-ST is, as expected, very close to the one SNR-GCWE, which supports the idea to approximate the behavior of SNR-ST with the one of SNR-GCWE [41].

VII. Conclusion

This paper has analyzed the asymptotic distribution of the EVD parameters of the M-estimators of the scatter matrix. In the general context of CES-distributed samples, we derived these asymptotics for both the standard asymptotic and GCWE [16] regime. Interestingly, we have shown that the behavior of the EVD parameters is more accurately characterized by an equivalent Wishart model than by their standard asymptotic Gaussian distribution. This result represents the main contribution, as it shows that one can leverage results established for the EVD of Wishart-distributed matrices for directly characterizing the EVD of M-estimators. Some examples (intrinsic bias analysis, rank estimation, and low-rank adaptive filtering) illustrated the interest of this approach.

Acknowledgements

Gordana Draskovic's work is supported by the Labex DigiCosme/Digiteo. Arnaud Breloy and Frederic Pascal works are partially supported by ANR-ASTRID MARGARITA project (ANR-17-ASTR-0015).

Appendices

To prove all theorems we will use the basic results, based on a Delta-method, obtained in the following theorem [36][Proposition 6.2] (and e.g. [11] for the formulation in the complex case) recalled in the following lemma:

Lemma .1 *Let $\{\widehat{\mathbf{z}}\}$ be a sequence of complex random vectors $\widehat{\mathbf{z}}$ and $\mathbf{z}$ a compatible fixed vector. Assume that $\sqrt{n}\,(\widehat{\mathbf{z}} - \mathbf{z}) \overset{d}{\to} \mathcal{GCN}\,(\mathbf{0}, \mathbf{V}, \mathbf{W})$. 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*

$$\sqrt{n}\,(\xi\,(\widehat{\mathbf{z}}) - \xi\,(\mathbf{z})) \overset{d}{\to} \mathcal{GCN}\left(\mathbf{0}, \mathbf{D}\mathbf{V}\mathbf{D}^H, \mathbf{D}\mathbf{W}\mathbf{D}^T\right) \tag{46}$$

where

$$\mathbf{D} = \left.\frac{\partial \xi(\mathbf{y})}{\partial \mathbf{y}}\right|_{\mathbf{y}=\mathbf{z}} \tag{47}$$

is a matrix derivative.

For the sake of simplicity, we will write (47) as $\mathbf{D} = \partial\xi(\mathbf{z})/\partial\mathbf{z}$ from now on.

Appendix A
Proof of Theorem III.1

Proof: To derive the derivatives of $\boldsymbol{\lambda}$ and $\mathbf{u}_j$ with respect to $\text{vec}(\boldsymbol{\Sigma})$, we differentiate $\boldsymbol{\Sigma}\mathbf{u}_j = \lambda_j \mathbf{u}_j$

$$d\boldsymbol{\Sigma}\mathbf{u}_j + \boldsymbol{\Sigma} d\mathbf{u}_j = d\lambda_j \mathbf{u}_j + \lambda_j d\mathbf{u}_j. \tag{48}$$

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

$$d\lambda_j = \mathbf{u}_j^H (d\boldsymbol{\Sigma}) \mathbf{u}_j$$

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

$$\frac{\partial \lambda_j}{\partial \text{vec}(\boldsymbol{\Sigma})} = \mathbf{u}_j^T \otimes \mathbf{u}_j^H.$$

If $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_p)$, then one has

$$\frac{\partial \boldsymbol{\lambda}}{\partial \text{vec}(\boldsymbol{\Sigma})} = \mathbf{E}^T \left(\mathbf{U}^T \otimes \mathbf{U}^H\right)$$

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. (9), one obtains

$$\begin{aligned}
&\mathbf{E}^T \left(\mathbf{U}^T \otimes \mathbf{U}^H\right) \left(\vartheta_1 \left(\boldsymbol{\Sigma}^T \otimes \boldsymbol{\Sigma}\right)\right) \left(\mathbf{U}^* \otimes \mathbf{U}\right) \mathbf{E} \\
&+ \mathbf{E}^T \left(\mathbf{U}^T \otimes \mathbf{U}^H\right) \vartheta_2 \text{vec}(\boldsymbol{\Sigma}) \text{vec}(\boldsymbol{\Sigma})^H \left(\mathbf{U}^* \otimes \mathbf{U}\right) \mathbf{E} \\
&= \vartheta_1 \mathbf{E}^T \left(\boldsymbol{\Lambda}^T \otimes \boldsymbol{\Lambda}\right) \mathbf{E} + \vartheta_2 \mathbf{E}^T \left(\text{vec}(\boldsymbol{\Lambda}) \text{vec}(\boldsymbol{\Lambda})^H\right) \mathbf{E} \\
&= \vartheta_1 \boldsymbol{\Lambda}^2 + \vartheta_2 \boldsymbol{\lambda}\boldsymbol{\lambda}^T.
\end{aligned} \tag{49}$$

Note that in this equality $\boldsymbol{\Sigma}$ figures instead $\boldsymbol{\Sigma}_\sigma$, since we analyze the distribution of $\sigma\widehat{\lambda}^M$ instead of $\widehat{\lambda}^M$. Note also 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. (48) by $\mathbf{u}_k^H$, $k \neq j$. Thus, one obtains

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

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

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

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

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

which actually gives the projection of the derivative onto the subspace orthogonal to the one of the eigenvector. Now, employing Eq. (46) with the previous derivatives and since

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

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. (18), one obtains

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

Then

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

Since from Eq. (18) one has

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

it remains only to derive the expression for

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

Using the Delta method [36], one can show that

$$\text{cov}_n\left(\sigma\widehat{\boldsymbol{\lambda}}^M, \widehat{\boldsymbol{\lambda}}^{\text{GCWE}}\right) \to \frac{\sigma\partial\boldsymbol{\lambda}}{\partial\text{vec}(\boldsymbol{\Sigma})} \mathbf{Q} \left(\frac{\partial\boldsymbol{\lambda}}{\partial\text{vec}(\boldsymbol{\Sigma})}\right)^T,$$

where $\mathbf{Q}$ is the asymptotic covariance matrix of $\sigma\widehat{\boldsymbol{\Sigma}}$. This matrix is equal to

$$\mathbf{Q} = \gamma_1 \left(\boldsymbol{\Sigma}^T \otimes \boldsymbol{\Sigma}\right) + \gamma_2 \text{vec}(\boldsymbol{\Sigma}) \text{vec}(\boldsymbol{\Sigma})^H \tag{50}$$

with γ_1 and γ_2 given in [16]. Repeating the same steps as in Eqs. (49), one shows that the right-hand side of Eq. (50) becomes

$$\gamma_1 \boldsymbol{\Lambda}^2 + \gamma_2 \boldsymbol{\lambda}\boldsymbol{\lambda}^T.$$

In [16] it has been shown that $\sigma_1 = \vartheta_1 - 2\gamma_1 + 1$ and $\sigma_2 = \vartheta_2 - 2\gamma_2$, which 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 $\mathbf{\Sigma}_r$ as

$$\mathbf{\Phi} = \mathbf{U}_r \mathbf{\Lambda}_r^{-1} \mathbf{U}_r^H, \quad (51)$$

one has from [27] that

$$\widehat{\mathbf{\Pi}}_r = \mathbf{\Pi}_r + \delta\mathbf{\Pi}_r + \ldots + \delta^i \mathbf{\Pi}_r + \ldots$$

where

$$\begin{aligned} \delta\mathbf{\Pi}_r &= \mathbf{\Pi}_r^\perp \Delta\mathbf{\Sigma}\mathbf{\Phi} + \mathbf{\Phi}\Delta\mathbf{\Sigma}\mathbf{\Pi}_r^\perp, \\ \delta^i\mathbf{\Pi}_r &= -\mathbf{\Pi}_r^\perp \left(\delta^{i-1}\mathbf{\Pi}\right)\Delta\mathbf{\Sigma}\mathbf{\Phi} + \mathbf{\Pi}_r^\perp \left(\delta^{i-1}\mathbf{\Pi}\right)\Delta\mathbf{\Sigma}\mathbf{\Phi}, \end{aligned}$$

with $\Delta\mathbf{\Sigma} = \widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}$.

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

$$\widehat{\mathbf{\Pi}}_r = \mathbf{\Pi}_r + \delta\mathbf{\Pi}_r$$

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

$$\text{vec}\left(\widehat{\mathbf{\Pi}}_r - \mathbf{\Pi}_r\right) = \mathbf{F}\text{vec}\left(\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\right)$$

with

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

It is now obvious that the covariance (resp. pseudo-covariance) matrix of $\sqrt{n}\left(\mathbf{\Pi}_r^M - \mathbf{\Pi}_r\right)$ is equal to $\mathbf{FCF}^H$ (resp. $\mathbf{FPF}^T$) where $\mathbf{C}$ and $\mathbf{P}$ are given in Eqs. (9). Further

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

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

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

one obtains

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

which with $\mathbf{\Phi}$ given by Eq. (51) and $\mathbf{\Sigma} = \mathbf{U}_r\mathbf{\Lambda}_r\mathbf{U}_r^H + \gamma^2\mathbf{I}_p$ yields the final result.

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

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," *Founds. and Trends in Signal Process.*, vol. 8, no. 3, pp. 127–216, 2015. [Online]. Available: http://dx.doi.org/10.1561/2000000053

[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] G. Boente, “Asymptotic theory for robust principal components,” *Journal of Multivariate Analysis*, vol. 21, no. 1, pp. 67 – 78, 1987.

[31] 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.

[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] H. Akaike, “A new look at the statistical model identification,” *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, December 1974.

[34] M. Wax and I. Ziskind, “Detection of the number of coherent signals by the mdl principle,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 8, pp. 1190–1196, Aug 1989.

[35] 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.

[36] M. Bilodeau and D. Brenner, *Theory of Multivariate Statistics*, ser. New York, NY. USA:Springer-Verlag, 1999.

[37] 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.

[38] L. E. Brennan and F. Staudaher, “Subclutter visibility demonstration,” Tech. Rep., 1992.

[39] J. Ward, “Space-Time Adaptive Processing for Airborne Radar MIT Lincoln Lab., Lexington,” MA, Tech. Rep. 1015, Tech. Rep., 1994.

[40] 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.

[41] 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.