



**HAL**  
open science

# New insights into the statistical properties of M-estimators

Gordana Draskovic, Frédéric Pascal

► **To cite this version:**

Gordana Draskovic, Frédéric Pascal. New insights into the statistical properties of M-estimators. IEEE Transactions on Signal Processing, 2018, 66 (16), pp.4253-4263. 10.1109/TSP.2018.2841892 . hal-01816084

**HAL Id: hal-01816084**

**<https://hal.science/hal-01816084>**

Submitted on 26 Feb 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# New insights into the statistical properties of $M$ -estimators

Gordana Drašković, *Student Member, IEEE*, Frédéric Pascal, *Senior Member, IEEE*

**Abstract**—This paper proposes an original approach to better understanding the behavior of robust scatter matrix  $M$ -estimators. Scatter matrices are of particular interest for many signal processing applications since the resulting performance strongly relies on the quality of the matrix estimation. In this context,  $M$ -estimators appear as very interesting candidates, mainly due to their flexibility to the statistical model and their robustness to outliers and/or missing data. However, the behavior of such estimators still remains unclear and not well understood since they are described by fixed-point equations that make their statistical analysis very difficult. To fill this gap, the main contribution of this work is to prove that these estimators distribution is more accurately described by a Wishart distribution than by the classical asymptotical Gaussian approximation. To that end, we propose a new “Gaussian-core” representation for Complex Elliptically Symmetric (CES) distributions and we analyze the proximity between  $M$ -estimators and a Gaussian-based Sample Covariance Matrix (SCM), unobservable in practice and playing only a theoretical role. To confirm our claims we also provide results for a widely used function of  $M$ -estimators, the Mahalanobis distance. Finally, Monte Carlo simulations for various scenarios are presented to validate theoretical results.

**Index Terms**— $M$ -estimators, Complex Elliptical Symmetric distributions, robust estimation, Wishart distribution, Mahalanobis distance.

## I. INTRODUCTION

In signal processing applications, the knowledge of scatter matrix is of crucial importance. It arises in diverse applications such as filtering, detection, estimation or classification. In recent years, there has been growing interest in covariance matrix estimation in a vast amount of literature on this topic (see e.g., [1]–[8] and references therein). Generally, in most of signal processing methods the data can be locally modelled by a multivariate zero-mean circular Gaussian stochastic process, which is completely determined by its covariance matrix. Complex multivariate Gaussian, also called complex normal (CN), distribution plays a vital role in the theory of statistical analysis [9]. Very often the multivariate observations are approximately normally distributed. This approximation is (asymptotically) valid even when the original data is not multivariate normal, due to the central limit theorem. In that case, the classical covariance matrix estimator is the sample covariance matrix (SCM) whose behavior is perfectly known. Indeed, it follows the Wishart distribution [10] which is the multivariate extension of the gamma distribution. Thanks to

its explicit form, the SCM is easy to manipulate and therefore widely used in the signal processing community.

Nevertheless, the complex normality sometimes presents a poor approximation of underlying physics. Noise and interference can be spiky and impulsive i.e., have heavier tails than the Gaussian distribution. An alternative has been proposed by introducing elliptical distributions [11], namely the Complex Elliptically Symmetric (CES) distributions. These distributions present an important property which states that their higher order moment matrices are scalars multiple of their correspondent normal distribution. This presents a starting point for the analysis that is done in this paper. These distributions have been frequently employed for non-Gaussian modeling (see e.g., for radar applications [12]–[16]).

Although Huber introduced robust  $M$ -estimators in [17] for the scalar case, Maronna provided the detailed analysis of the corresponding scatter matrix estimators in the multivariate real case in his seminal work [18].  $M$ -estimators correspond to a generalization of the well-known Maximum Likelihood estimators (MLE), that have been widely studied in the statistics literature [19], [20]. In contrast to  $ML$ -estimators where the estimating equation depends on the probability density function (PDF) of a particular CES distribution, the weight function in the  $M$ -estimating equation can be completely independent of the data distribution. Consequently,  $M$ -estimators presents a wide class of scatter matrix estimators, including the  $ML$ -estimators, robust to the data model. In [18], it is shown that, under some mild assumptions, the estimator is defined as the unique solution of a fixed-point equation and that the robust estimator converges almost surely (a.s.) to a deterministic matrix, equal to the scatter matrix up to a scale quantity (depending on the true statistical model). Their asymptotical properties have been studied by Tyler in the real case [21]. This has been recently extended to the complex case, more useful for signal processing applications, in [1], [6].

In most of the papers, three main  $M$ -estimators are studied and used in practice: the Student’s  $M$ -estimator that is MLE for  $t$ -distribution, the Huber’s  $M$ -estimator and the Tyler’s  $M$ -estimator [22], also known as Fixed Point (FP) estimator [2]. Student  $t$ -distribution is widely employed for non-Gaussian data modeling since it offers flexibility thanks to an additive parameter, namely the Degree of Freedom (DoF). As a consequence, Student’s  $M$ -estimator is often used for scatter matrix estimation. Huber’s  $M$ -estimator, especially its complex multivariate extension, has received a lot of attention since proven to be very robust to outliers. Tyler’s  $M$ -estimator is not exactly an  $M$ -estimator<sup>1</sup> but it is very useful because

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). “This paper has supplementary downloadable material available at <http://ieeexplore.ieee.org>, provided by the author. The material includes the results for the real case. This material is 149KB in size.”

<sup>1</sup>especially because it does not respect all Maronna conditions [18]

of rare property that any CES distribution with the same scatter matrix leads to the same result (hence “distribution-free”). Asymptotical properties of this estimator have been analyzed in [1], [23]. Recently, it has been shown that the behavior of Tyler’s estimator can be better approximated by a Wishart distribution [24]. In this work, one aims at providing more general results that can be applied to all  $M$ -estimators and one wants to analyze the gain of this approach on the robust Mahalanobis distance [25], [26], very useful in various problems such as detection, clustering etc.

The contributions of this work are multiple. First, the originality of the results comes from a new CES representation introducing “Gaussian cores”. This representation is a modified stochastic representation given in [1] and is crucial to understand the proposed method. Second, in this paper,  $M$ -estimators are, for the first time, analyzed thanks to a comparison with a very simple estimator, the SCM. Indeed, the direct statistical analysis of these estimators is difficult because they are defined as the solution of an implicit equation and have been analyzed only in classical asymptotic regimes. Here, we propose a different approach to overcome this difficulty. More precisely, a sort of distance between  $M$ -estimators and the SCM is computed in order to propagate SCM non-asymptotic properties towards  $M$ -estimators. Third, the paper gives new insights into the correlation between  $M$ -estimators and the corresponding SCM in the Gaussian context which is the central part of our approach. Finally, we present a practical interest of the results, specifically the application to the Mahalanobis distance. Note that all the results are provided in the complex case. For completeness purposes, a supplemental material containing analogous results in the real case is provided together with this article.

The rest of this paper is organized as follows. Section II introduces the considered CES-models based on Gaussian cores as well as the  $M$ -estimators and Mahalanobis distance. Section III contains the main contribution of the paper with discussions and further explanations. Moreover, closed-form expressions are derived for some particular cases of  $M$ -estimators and the application to the Mahalanobis distance is presented. In Section IV, Monte Carlo simulations are presented in order to validate the theoretical results. Finally, some conclusions and perspectives are drawn in Section V.

*Notations* - Vectors (resp. matrices) are denoted by bold-faced lowercase letters (resp. uppercase letters).  $T$ ,  $*$   $H$  respectively represent the transpose, conjugate and the Hermitian operator.  $Re(\cdot)$  and  $Im(\cdot)$  denote respectively the real and the imaginary part of a complex quantity, i.i.d. stands for “independent and identically distributed” while  $\sim$  means “is distributed as”.  $=_d$  stands for “shares the same distribution as”,  $\xrightarrow{d}$  denotes convergence in distribution and  $\otimes$  denotes the Kronecker product.  $\mathbb{1}$  is indicator function and  $\text{vec}(\cdot)$  is the operator which transforms a matrix  $m \times n$  into a vector of length  $mn$ , concatenating its  $n$  columns into a single column. Moreover,  $\mathbf{I}_m$  is the  $m \times m$  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.  $\mathbf{K} \text{vec}(\mathbf{A}) = \text{vec}(\mathbf{A}^T)$ .

## II. PROBLEM FORMULATION

### A. Complex distributions

Let  $\mathbf{z} = Re(\mathbf{z}) + jIm(\mathbf{z})$  be an  $m$ -dimensional complex random vector which consists of a pair of real random vectors  $Re(\mathbf{z})$  and  $Im(\mathbf{z})$ . The distribution of  $\mathbf{z}$  on  $\mathbb{C}^m$  determines the joint real  $2m$ -variate distribution of  $Re(\mathbf{z})$  and  $Im(\mathbf{z})$  on  $\mathbb{R}^{2m}$  and conversely. To completely define the second-order moments of  $Re(\mathbf{z})$  and  $Im(\mathbf{z})$ ,  $\mathbf{z}$  is given by its covariance matrix  $\mathbf{C} = E[(\mathbf{z} - \boldsymbol{\mu})(\mathbf{z} - \boldsymbol{\mu})^H]$  and pseudo-covariance matrix  $\mathbf{P} = E[(\mathbf{z} - \boldsymbol{\mu})(\mathbf{z} - \boldsymbol{\mu})^T]$ . If the complex vector is *circular* (see [1] for details), the pseudo-covariance vanishes, i.e.  $\mathbf{P} = \mathbf{0}$ .

1) *Generalized Complex Normal distribution*: An  $m$ -dimensional random vector has the generalized normal distribution  $\mathbf{z} \sim \mathcal{CN}(\boldsymbol{\mu}, \mathbf{C}, \mathbf{P})$  if its probability density function (PDF) can be written as

$$h_{\mathbf{z}}(\mathbf{z}) = \frac{\exp \left\{ -\frac{1}{2} [(\mathbf{z} - \boldsymbol{\mu})^H \quad (\mathbf{z} - \boldsymbol{\mu})^T] \mathbf{V}^{-1} \begin{bmatrix} \mathbf{z} - \boldsymbol{\mu} \\ \mathbf{z}^* - \boldsymbol{\mu}^* \end{bmatrix} \right\}}{\pi^m \sqrt{|\mathbf{V}|}} \quad (1)$$

where  $\boldsymbol{\mu}$  is the statistical mean and  $\mathbf{V} = \begin{bmatrix} \mathbf{C} & \mathbf{P} \\ \mathbf{P}^* & \mathbf{C}^* \end{bmatrix}$ . If  $\mathbf{z}$  is circular CN-distributed the pseudo-covariance will be omitted in the notation, i.e.  $\mathbf{z} \sim \mathcal{CN}(\boldsymbol{\mu}, \mathbf{C})$ .

### B. Complex Elliptically Symmetric distributions

An important class of circular distributions are the CES distributions. An  $m$ -dimensional random vector has a CES distribution if its probability density function (PDF) can be written as

$$h_{\mathbf{z}}(\mathbf{z}) = C |\mathbf{M}|^{-1} g_{\mathbf{z}} \left( (\mathbf{z} - \boldsymbol{\mu})^H \mathbf{M}^{-1} (\mathbf{z} - \boldsymbol{\mu}) \right) \quad (2)$$

where  $C$  is a constant,  $g_{\mathbf{z}} : [0, \infty) \rightarrow [0, \infty)$  is any function (called the density generator) such that (2) defines a PDF and  $\mathbf{M}$  is the scatter matrix. The matrix  $\mathbf{M}$  reflects the structure of the covariance matrix  $\mathbf{C}$  of  $\mathbf{z}$ , i.e., the covariance matrix is equal to  $\mathbf{M}$  up to a scale factor  $\mathbf{C} = \xi \mathbf{M}^2$ . This CES distribution will be denoted by  $\mathcal{CES}(\boldsymbol{\mu}, \mathbf{M}, g_{\mathbf{z}})$ . In this paper, we will assume that  $\boldsymbol{\mu} = \mathbf{0}$ , as it is generally the case for many signal processing applications.

1) *Stochastic Representation Theorem*: A zero mean random vector  $\mathbf{z} \sim \mathcal{CES}_m(\mathbf{0}, \mathbf{M}, g_{\mathbf{z}})$  if and only if it admits the following stochastic representation [27]

$$\mathbf{z} \stackrel{d}{=} \sqrt{Q} \mathbf{A} \mathbf{u}, \quad (3)$$

where the non-negative real random variable  $Q$ , called the modular variate, is independent of the random vector  $\mathbf{u}$  that is uniformly distributed on the unit complex  $m$ -sphere and  $\mathbf{M} = \mathbf{A} \mathbf{A}^H$  is a factorization of  $\mathbf{M}$ .

2) *Circular Complex Normal Distribution*: Complex normal (Gaussian) distribution is a particular case of CES distributions in which  $g_{\mathbf{z}}(z) = e^{-z}$  and  $C = \pi^{-m}$ . Thus, the PDF of  $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \mathbf{M})$  is given by

$$h_{\mathbf{z}}(\mathbf{z}) = \frac{\exp \{ -\mathbf{z}^H \mathbf{M}^{-1} \mathbf{z} \}}{\pi^m |\mathbf{M}|}. \quad (4)$$

<sup>2</sup>if the random vector has a finite second-order moment (see [1] for details)

Note that for  $\boldsymbol{\mu} = \mathbf{0}$  the PDF (1) reduces to the PDF above since the scatter matrix is equal to the covariance matrix  $\mathbf{C} = \mathbf{M}$  (i.e., the scale factor  $\xi$  is equal to 1). Regarding the previous stochastic representation theorem, for CN-distributed vector  $\mathbf{z}$  the random variable  $Q$  has a scaled chi-squared distribution  $Q \sim (1/2)\chi_{2m}^2$ .

3) *Gaussian-core representation of CES*: In order to better explain the context of this work, we will rewrite the stochastic representation using the fact that  $\mathbf{u} \stackrel{d}{=} \mathbf{n}/\|\mathbf{n}\|$ , where  $\mathbf{n} \sim \mathbb{CN}(\mathbf{0}, \mathbf{I})$ . Hence, a random vector  $\mathbf{z} \sim \text{CES}(\mathbf{0}, \mathbf{M}, g_{\mathbf{z}})$  can be represented as

$$\mathbf{z} \stackrel{d}{=} \frac{\sqrt{Q}}{\|\mathbf{n}\|} \mathbf{A} \mathbf{n} \quad (5)$$

with  $Q$  and  $\mathbf{A}$  defined as in Eq. (3). If  $\sqrt{Q}/\|\mathbf{n}\|$  is independent of  $\mathbf{n}$ , the vector  $\mathbf{z}$  is said to have a compound-Gaussian distribution and it can be represented as  $\mathbf{z} \stackrel{d}{=} \sqrt{\tau} \mathbf{x}$ , where the non-negative real random variable  $\tau$ , generally called the texture, is independent of the vector  $\mathbf{x} \sim \mathbb{CN}(\mathbf{0}, \mathbf{M})$ .

4) *Student  $t$ -distribution*: A zero-mean random vector  $\mathbf{z}$  follows a complex multivariate  $t$ -distribution with  $\nu$  ( $0 < \nu < \infty$ ) degrees of freedom if the corresponding stochastic representation admits  $Q \sim mF_{2m, \nu}$ . This distribution belongs to the compound-Gaussian distributions where  $\tau \sim \text{IG}(\nu/2, \nu/2)$ , with IG denoting the inverse Gamma distribution. Note that the case  $\nu \rightarrow \infty$  leads to the Gaussian distribution. The multivariate  $t$ -distributions, besides the Gaussian distribution, encompass also multivariate Laplace distribution (for  $\nu = 1/2$ ) and the multivariate Cauchy distribution (for  $\nu = 1$ ) which are heavy-tailed alternatives to the Gaussian distribution. The complex multivariate  $t$ -distributions are thus useful for studying robustness of multivariate statistics as a decrement of  $\nu$  yields to distributions with an increased heaviness of the tails. We shall write  $\mathbb{C}t_{\nu}(\mathbf{0}, \mathbf{M})$  to denote this case.

### C. Wishart distribution

The complex Wishart distribution is the distribution of  $\sum_{k=1}^K \mathbf{x}_k \mathbf{x}_k^H$ , when  $\mathbf{x}_k$  are  $m$ -dimensional complex circular i.i.d. zero-mean Gaussian vectors with covariance matrix  $\mathbf{M}$ . Let

$$\widehat{\mathbf{M}}_{SCM} = \frac{1}{K} \sum_{k=1}^K \mathbf{x}_k \mathbf{x}_k^H$$

be the related sample covariance matrix (SCM) which will be also referred to as a Wishart matrix. Its asymptotic distribution [10], is given by

$$\sqrt{K} \text{vec} \left( \widehat{\mathbf{M}}_{SCM} - \mathbf{M} \right) \xrightarrow{d} \mathbb{CN}(\mathbf{0}, \boldsymbol{\Sigma}_{SCM}, \boldsymbol{\Omega}_{SCM}) \quad (6)$$

where the asymptotic covariance and pseudo-covariance matrices are

$$\begin{cases} \boldsymbol{\Sigma}_{SCM} = \mathbf{M}^T \otimes \mathbf{M} \\ \boldsymbol{\Omega}_{SCM} = (\mathbf{M}^T \otimes \mathbf{M}) \mathbf{K}. \end{cases} \quad (7)$$

### D. $M$ -estimators

Let  $(\mathbf{z}_1, \dots, \mathbf{z}_K)$  be a  $K$ -sample of  $m$ -dimensional complex i.i.d. vectors with  $\mathbf{z}_k \sim \text{CES}(\mathbf{0}, \mathbf{M}, g_{\mathbf{z}})$ . An  $M$ -estimator,

denoted by  $\widehat{\mathbf{M}}$ , is defined by the solution of the following  $M$ -estimating equation

$$\widehat{\mathbf{M}} = \frac{1}{K} \sum_{k=1}^K \varphi \left( \mathbf{z}_k^H \widehat{\mathbf{M}}^{-1} \mathbf{z}_k \right) \mathbf{z}_k \mathbf{z}_k^H \quad (8)$$

where  $\varphi$  is any real-valued weight function on  $[0, \infty)$  that respects Maronna's conditions<sup>3</sup> [18]. The theoretical (population) scatter matrix  $M$ -functional is defined as a solution of

$$\mathbb{E} [\varphi (\mathbf{z}^H \mathbf{M}_{\sigma}^{-1} \mathbf{z}) \mathbf{z} \mathbf{z}^H] = \mathbf{M}_{\sigma}$$

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

$$\mathbb{E} [\psi(\sigma t)] = m \quad (9)$$

with  $\psi(\sigma t) = \varphi(\sigma t) \sigma t$  and  $t = \mathbf{z}^H \widehat{\mathbf{M}}^{-1} \mathbf{z}$ .

**Theorem II.1** *Let  $\widehat{\mathbf{M}}$  be a complex  $M$ -estimator following Maronna's conditions [18]. Then*

$$\sqrt{K} \text{vec} \left( \widehat{\mathbf{M}} - \mathbf{M}_{\sigma} \right) \xrightarrow{d} \mathbb{CN}(\mathbf{0}, \boldsymbol{\Sigma}_M, \boldsymbol{\Omega}_M)$$

where the asymptotic covariance and pseudo-covariance matrices are

$$\begin{cases} \boldsymbol{\Sigma}_M = \vartheta_1 \mathbf{M}_{\sigma}^T \otimes \mathbf{M}_{\sigma} + \vartheta_2 \text{vec}(\mathbf{M}_{\sigma}) \text{vec}(\mathbf{M}_{\sigma})^H, \\ \boldsymbol{\Omega}_M = \vartheta_1 (\mathbf{M}_{\sigma}^T \otimes \mathbf{M}_{\sigma}) \mathbf{K} + \vartheta_2 \text{vec}(\mathbf{M}_{\sigma}) \text{vec}(\mathbf{M}_{\sigma})^T. \end{cases} \quad (10)$$

The constants  $\vartheta_1 > 0$  and  $\vartheta_2 > -\vartheta_1/m$  are given in [1], [6].

1) *Tyler's estimator*: For the particular function  $\varphi(t) = m/t$ , Tyler's estimator  $\widehat{\mathbf{M}}_{FP}$  is the solution (up to a scale factor) of the following equation

$$\widehat{\mathbf{M}}_{FP} = \frac{m}{K} \sum_{k=1}^K \frac{\mathbf{z}_k \mathbf{z}_k^H}{\mathbf{z}_k^H \widehat{\mathbf{M}}_{FP}^{-1} \mathbf{z}_k}. \quad (11)$$

It should be noted that for Tyler's estimator,  $\boldsymbol{\Sigma}_T$  and  $\boldsymbol{\Omega}_T$  are also defined as in Eq. (10) (see [23]) for  $\sigma = 1$  and

$$\begin{cases} \vartheta_1 = m^{-1}(m+1) \\ \vartheta_2 = -m^{-2}(m+1) \end{cases} \quad (12)$$

2) *Huber's  $M$ -estimator*: The complex extension of Huber's  $M$ -estimator is defined by

$$\begin{aligned} \widehat{\mathbf{M}}_{Hub} = & \frac{1}{K\beta} \sum_{k=1}^K \left[ \mathbf{z}_k \mathbf{z}_k^H \mathbb{1}_{\mathbf{z}_k^H \widehat{\mathbf{M}}_{Hub}^{-1} \mathbf{z}_k \leq p^2} \right] \\ & + \frac{1}{K\beta} p^2 \sum_{k=1}^K \left[ \frac{\mathbf{z}_k \mathbf{z}_k^H}{\mathbf{z}_k^H \widehat{\mathbf{M}}_{Hub}^{-1} \mathbf{z}_k} \mathbb{1}_{\mathbf{z}_k^H \widehat{\mathbf{M}}_{Hub}^{-1} \mathbf{z}_k > p^2} \right], \end{aligned} \quad (13)$$

where  $p^2$  and  $\beta$  depend on a single parameter  $0 < q < 1$ , according to  $q = F_{2m}(2p^2)$  and  $\beta = F_{2m+2}(2p^2) + p^2 \frac{1-q}{m}$  where  $F_m(\cdot)$  is the cumulative distribution function of a  $\chi^2$  distribution with  $m$  degrees of freedom. Note that the Huber's  $M$ -estimator can be interpreted as a weighted combination between the SCM and Tyler's estimator.

<sup>3</sup>The weight function  $\varphi$  does not need to be related to the PDF of any particular CES distribution, and hence  $M$ -estimators constitute a wide class of scatter matrix estimators.

3) *Student's M-estimator*: The MLE for the Student  $t$ -distribution, denoted  $\widehat{\mathbf{M}}_t$ , is obtained as a solution of the following equation

$$\widehat{\mathbf{M}}_t = \frac{m + \nu/2}{K} \sum_{k=1}^K \frac{\mathbf{z}_k \mathbf{z}_k^H}{\nu/2 + \mathbf{z}_k^H \widehat{\mathbf{M}}_t^{-1} \mathbf{z}_k}. \quad (14)$$

The motivation to analyze this estimator arises from the fact that it presents a trade-off between the SCM and Tyler's estimator, but in a different way as for the Huber's  $M$ -estimator. Indeed,  $\nu \rightarrow \infty$  leads to the Gaussian distribution and the resulting MLE is the SCM ( $\varphi(t) \rightarrow 1$ ) while  $\nu \rightarrow 0$  yields Tyler's estimator ( $\varphi(t) \rightarrow m/t$ ). Finally,  $\widehat{\mathbf{M}}_t$  is widely used both in theory (as a benchmark) and in practice which presents strong motivation for understanding its behavior. Note also that as others  $M$ -estimators, it is not always used as a MLE for the  $t$  distribution.

### E. Mahalanobis distance

Mahalanobis distance [25], [26] is one of the most common measures in multivariate statistics and signal processing. It is based on the correlation between variables thanks to which different models can be identified and analyzed. The Mahalanobis distance of  $\mathbf{z}$  from  $\boldsymbol{\mu}$  is given by  $\Delta(\boldsymbol{\mu}, \mathbf{M})$  where

$$\Delta^2(\boldsymbol{\mu}, \mathbf{M}) = (\mathbf{z} - \boldsymbol{\mu})^H \mathbf{M}^{-1} (\mathbf{z} - \boldsymbol{\mu}). \quad (15)$$

where  $\boldsymbol{\mu}$  is population mean and  $\mathbf{M}$  is common scatter matrix. Since we work with zero mean vectors, we will analyze  $\Delta^2(\mathbf{M}) = \mathbf{z}^H \mathbf{M}^{-1} \mathbf{z}$ , without loss of generality. If the data are normal distributed,  $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \mathbf{M})$ , and the distance is based on the true scatter matrix  $\mathbf{M}$ , then it follows a scaled chi-squared distribution

$$\Delta^2(\mathbf{M}) \sim (1/2) \chi_{2m}^2. \quad (16)$$

Since the scatter matrix is usually unknown, the distance is computed with its estimate. If the SCM is plugged in instead of the true scatter matrix, the distance becomes  $\beta'$ -distributed<sup>4</sup> with an asymptotic chi-squared distribution

$$\Delta^2(\widehat{\mathbf{M}}_{SCM}) \sim K \beta'(m, K - m + 1) \quad (17)$$

where  $\beta'(a, b)$  denotes a Beta prime distribution with real shape parameters  $a$  and  $b$ .

Beside testing if an observed random sample is from a multivariate normal distribution (detecting outliers) [28], [29], the Mahalanobis distance is also a useful way to determine similarities between sets of known and unknown data. Thus, it is widely used in classification problems [30], [31], feature selection problems [32], anomaly detection in hyperspectral images [33], [34], etc.

The object of our study is to analyze the robust Mahalanobis distance, i.e. the distance computed with  $M$ -estimators, comparing it to the one based on the SCM, in order to better understand its behavior.

<sup>4</sup>Beta prime distribution corresponds to a scaled F-distribution.

## III. MAIN CONTRIBUTION

This section is devoted to the main contribution of the paper. First, the results for the asymptotic distribution of the difference between any  $M$ -estimator and the corresponding SCM in a Gaussian context are derived. Then, the results for particular  $M$ -estimators and the application to Mahalanobis distance are presented. Finally, discussion and some explanations are provided to emphasize the significance of the theoretical results.

### A. $M$ -estimators

Basing on the previously introduced Gaussian-core model, let us assume that  $K$  measurements are defined as follows:

- $\mathbf{z}_k = \sqrt{Q_k} / \|\mathbf{n}_k\| \mathbf{A} \mathbf{n}_k \sim \text{CES}(\mathbf{0}, \mathbf{M})$ ,  $k = 1, \dots, K$  with
  - $(\mathbf{n}_1, \dots, \mathbf{n}_K)$  be a  $K$ -sample of  $m$ -dimensional complex i.i.d. vectors with  $\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$
  - $Q_1, \dots, Q_K$  a  $K$ -sample of non-negative real i.i.d. random variables independent of the  $\mathbf{n}_k$ 's
  - $\mathbf{M} = \mathbf{A} \mathbf{A}^H$  is a factorization of  $\mathbf{M}$

$(\mathbf{z}_1, \dots, \mathbf{z}_K)$  corresponds to observed data, without more specifications on their distribution, and are used to design an  $M$ -estimator  $\widehat{\mathbf{M}}$ .

Let us also consider some "fictive" data (non observable) given by:

- $\mathbf{x}_k = \mathbf{A} \mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{M})$ ,  $k = 1, \dots, K$

and consider the SCM  $\widehat{\mathbf{M}}_{SCM}$  built with  $(\mathbf{x}_1, \dots, \mathbf{x}_K)$ . Hereafter, we always consider the same model unless it is stressed differently.

**Theorem III.1** *Let  $\widehat{\mathbf{M}}$  be defined by Eq. (8) and  $\sigma$  is the solution of Eq. (9). The asymptotic distribution of  $\sigma \widehat{\mathbf{M}} - \widehat{\mathbf{M}}_{SCM}$  is given by*

$$\sqrt{K} \text{vec} \left( \sigma \widehat{\mathbf{M}} - \widehat{\mathbf{M}}_{SCM} \right) \xrightarrow{d} \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}) \quad (18)$$

where  $\boldsymbol{\Sigma}$  and  $\boldsymbol{\Omega}$  are defined by

$$\begin{aligned} \boldsymbol{\Sigma} &= \sigma_1 \mathbf{M}^T \otimes \mathbf{M} + \sigma_2 \text{vec}(\mathbf{M}) \text{vec}(\mathbf{M})^H, \\ \boldsymbol{\Omega} &= \sigma_1 (\mathbf{M}^T \otimes \mathbf{M}) \mathbf{K} + \sigma_2 \text{vec}(\mathbf{M}) \text{vec}(\mathbf{M})^T \end{aligned} \quad (19)$$

with

$$\begin{aligned} \sigma_1 &= \frac{am(m+1) + c(c-2b)}{c^2} \\ \sigma_2 &= \frac{a - m^2}{(c - m^2)^2} - \frac{a(m+1)}{c^2} + 2 \frac{m(c-b)}{c(c-m^2)} \end{aligned} \quad (20)$$

where  $a = E[\psi^2(\sigma t_1)]$ ,  $b = E[\psi(\sigma t_1)t_2]$  and  $c = E[\psi'(\sigma t_1)\sigma t_1] + m^2$ .

**Remark III.1** *Notice that the structure of the asymptotic covariance matrix  $\boldsymbol{\Sigma}$  is the same as in classical asymptotic results (Eqs. (7) and (10)) but the coefficients are different. In the case of the identity matrix as covariance matrix, this very particular structure involves only three non-null elements  $d_1, d_2$  and  $d_3$  at the positions  $(i, j)$  and equal to:*

- $d_1 = \sigma_1 + \sigma_2$  for  $i = j = p + m(p-1)$  with  $p = 1, \dots, m$ ,

- $d_2 = \sigma_1$  for  $i = j = p + m(q - 1)$  with  $p \neq q$  and  $p, q = 1, \dots, m$ ,
- $d_3 = \sigma_2$  for  $i = p + m(p - 1), j = q + m(q - 1)$  with  $p \neq q$  and  $p, q = 1, \dots, m$ .

Similar comment with slight modifications is valid for the pseudo-covariance matrix.

*Proof sketch:* We provide only a sketch of the proof, while the detailed proof of Theorem III.1 is given in Appendix A. The main idea is to represent the matrix  $\Sigma$  as

$$\Sigma = \Sigma_1(\mathbf{M}) - 2\Sigma_2(\mathbf{M}) + \Sigma_3(\mathbf{M})$$

where  $\Sigma_1(\mathbf{M})$  and  $\Sigma_3(\mathbf{M})$  are given by Eq. (10) and Eq. (6), respectively and the matrix  $\Sigma_2(\mathbf{M})$  is the correlation matrix between an  $M$ -estimator and the corresponding SCM in a Gaussian context. The second important step relies on a decomposition  $\Sigma_2(\mathbf{M})$ :

$$\Sigma_2(\mathbf{M}) = \mathbf{D}_1^{-1}(\mathbf{M})\mathbf{B}_2(\mathbf{M})(\mathbf{D}_2^{-1}(\mathbf{M}))^H$$

where  $\mathbf{D}_1(\mathbf{M}) = E[d\{\text{vec}\Psi_1(\mathbf{M})\}/d\{\text{vec}(\mathbf{M})\}]$ ,  $\mathbf{B}_2(\mathbf{M}) = \text{cov}(\text{vec}\Psi_1(\mathbf{M}), \text{vec}\Psi_2(\mathbf{M}))$  and  $\mathbf{D}_2(\mathbf{M}) = E[d\{\text{vec}\Psi_2(\mathbf{M})\}/d\{\text{vec}(\mathbf{M})\}]$  with  $\Psi_1(\mathbf{M}) = \sigma\varphi(\mathbf{z}^H(\sigma^{-1}\mathbf{M})^{-1}\mathbf{z})\mathbf{z}\mathbf{z}^H - \mathbf{M}$  and  $\Psi_2(\mathbf{M}) = \mathbf{x}\mathbf{x}^H - \mathbf{M}$ , which is a generalization of a result derived in [18]. Finally, using the dependence between the practical and fictive data one can derive elements of the matrix  $\Sigma_2(\mathbf{M})$  and obtain the final result. ■

**Remark III.2** In this paper, we consider only complex  $M$ -estimators since they are used in signal processing applications. The results for the real case are given in the supplemental material. In the proof we provide only the steps that differ from the ones obtained in the complex case. It should be noted that the results of Theorem III.1 can also be derived using the results for the real case and vector/matrix complex-to-real mapping [1]. This is briefly discussed at the end of the additional document.

### B. Particular cases

1) *Tyler's estimator:* Hereafter, the results derived in [24], are presented. The first scale factor in the result can be (roughly speaking) obtained from the Theorem III.1<sup>5</sup> while the derivation of the second one requires a different approach.

**Theorem III.2** Let  $\widehat{\mathbf{M}}_{FP}$  be defined by Eq. (11). The asymptotic distribution of  $\widehat{\mathbf{M}}_{FP} - \widehat{\mathbf{M}}_{SCM}$  is given by

$$\sqrt{K}\text{vec}\left(\widehat{\mathbf{M}}_{FP} - \widehat{\mathbf{M}}_{SCM}\right) \xrightarrow{d} \mathbb{C}\mathcal{N}(\mathbf{0}, \Sigma_{FP}, \Omega_{FP})$$

where  $\Sigma_{FP}$  and  $\Omega_{FP}$  are defined by

$$\begin{aligned} \Sigma_{FP} &= \frac{1}{m}\mathbf{M}^T \otimes \mathbf{M} + \frac{m-1}{m^2}\text{vec}(\mathbf{M})\text{vec}(\mathbf{M})^H, \\ \Omega_{FP} &= \frac{1}{m}(\mathbf{M}^T \otimes \mathbf{M})\mathbf{K} + \frac{m-1}{m^2}\text{vec}(\mathbf{M})\text{vec}(\mathbf{M})^T. \end{aligned}$$

<sup>5</sup>by considering the function  $\psi(x) = m$ .

**Remark III.3** In the case of Tyler's estimator  $\psi(\sigma t_1) = m$  which leads to  $a = m^2$ ,  $b = m^2$  and  $c = m^2$ . Substituting these values in the expression of  $\sigma_1$ , one obtains the previous result. This is in agreement with the results obtained in [24], [35].

2) *Student's  $M$ -estimator:* In this subsection, one gives the results for the Student's  $M$ -estimator and  $t$ -distributed data. Let

- $\mathbf{x}_k \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \mathbf{M})$ ,  $k = 1, \dots, K$
- $\tau_k \sim \text{IG}(\nu/2, \nu/2)$ ,  $k = 1, \dots, K$
- $\mathbf{z}_k = \sqrt{\tau_k}\mathbf{x}_k \sim \mathbb{C}t_\nu(\mathbf{0}, \mathbf{M})$ ,  $k = 1, \dots, K$

where  $\mathbf{M} = \mathbf{A}\mathbf{A}^H$  is a factorization of  $\mathbf{M}$ . Consider the SCM  $\widehat{\mathbf{M}}_{SCM}$  built with  $(\mathbf{x}_1, \dots, \mathbf{x}_K)$  and the Student's  $M$ -estimator  $\widehat{\mathbf{M}}_t$  built with  $(\mathbf{z}_1, \dots, \mathbf{z}_K)$ .

**Corollary III.1** Let  $\widehat{\mathbf{M}}_t$  be defined by Eq. (14). The asymptotic distribution of  $\widehat{\mathbf{M}}_t - \widehat{\mathbf{M}}_{SCM}$  is given by (19) with  $\sigma_1 = (m + \nu/2)^{-1}$  and  $\sigma_2 = 2/\nu(m + 1 + \nu/2)(m + \nu/2)^{-1}$ .

*Proof:* See Appendix B. ■

3) *Huber's  $M$ -estimator:* The theoretical derivation of the asymptotical distribution for Huber's  $M$ -estimator is impossible, since the function  $\psi(t_1)$  is not differentiable in each point. However, we will present empirical results for this estimator in the next section.

### C. Application to Mahalanobis distance

In this subsection we provide results for the robust Mahalanobis distance which shows the main interest of our contribution.

**Theorem III.3** Let  $\widehat{\mathbf{M}}$  be defined by Eq. (8) and  $\sigma$  is the solution of (9). For the Mahalanobis distance based on  $\sigma\widehat{\mathbf{M}}$  one has, conditionally to the distribution of  $\mathbf{z}$ , the following asymptotic distribution

$$\sqrt{K} \left( \frac{\mathbf{z}^H(\sigma\widehat{\mathbf{M}})^{-1}\mathbf{z} - \mathbf{z}^H\widehat{\mathbf{M}}_{SCM}^{-1}\mathbf{z}}{\mathbf{z}^H\mathbf{M}^{-1}\mathbf{z}} \right)_{\mathbf{z}} \xrightarrow{d} N(0, \phi) \quad (21)$$

where

$$\phi = \sigma_1 + \sigma_2 \quad (22)$$

with  $\sigma_1$  and  $\sigma_2$  given by Eq. (20) and where the notation  $(\cdot)_{\mathbf{z}}$  stresses the conditional distribution, conditional to  $\mathbf{z}$ .

*Proof:* See Appendix C. ■

**Remark III.4** The asymptotic variance of the robust Mahalanobis distance when centering around Wishart-based distance is smaller than the one when centering around the distance based on the true scatter matrix since  $\sigma_1 + \sigma_2 < \vartheta_1 + \vartheta_2$ . The results are accurate even when  $K$  is small which will be demonstrated in the simulation section. These findings reveal that the distribution of the robust (squared) Mahalanobis distance is better approximated with a scaled Beta prime distribution than with a scaled chi-squared distribution.

#### D. Discussion

Here are some general comments on the proposed results as well as their great interest in practice.

- 1) First, to examine the values of the scale factors in Eq. (20), we discuss the values of  $E[\psi^2(\sigma t_1)]$ ,  $E[\psi(\sigma t_1)t_2]$  and  $E[\psi'(\sigma t_1)\sigma t_1] + m^2$ . Since  $0 < \psi(\sigma t_1) < M$  and  $E(\psi(\sigma t_1)) = m$ , using Bhatia-Davis inequality [36], one has that  $\text{var}(\psi(\sigma t_1)) < (M - m)m$  and thus  $E(\psi(\sigma t_1)^2) < Mm$ . Since  $M$  is of same magnitude as  $m$  and  $M > m$ , one obtains that  $E(\psi(\sigma t_1)^2)$  is of same magnitude as  $m^2$  (for Tyler's estimator  $E(\psi(\sigma t_1)^2) = m^2$ , for Student  $M$ -estimator  $E(\psi(\sigma t_1)^2) = \frac{m(m+1)(m+\frac{\nu}{2})}{m+1+\frac{\nu}{2}}$ , for SCM  $E(\psi(\sigma t_1)^2) = m^2 + m\dots$ ). From this, it follows that  $b$  is also of the same magnitude as  $m^2$  since  $m^2 = \frac{E(\psi(\sigma t_1))E(t_2)}{\sqrt{E(\psi(\sigma t_1)^2)}\sqrt{E(t_2^2)}} < \frac{E(\psi(\sigma t_1)t_2)}{\sqrt{E(\psi(\sigma t_1)^2)}\sqrt{E(t_2^2)}} < \frac{E(\psi(\sigma t_1)t_2)}{\sqrt{(m^2 + m)}\sqrt{E(\psi(\sigma t_1)^2)}}$ . It is obvious that  $c$  is also of the same magnitude as  $m^2$ . Generally, for all widely used  $M$ -estimators, one obtains that  $a, b, c = m^2 + \alpha m$ ,  $\alpha > 0$  which leads to  $\sigma_1$  inversely proportional to  $m$ . For  $\sigma_2$ , one can not provide precise information about its value, but it turns out that it is either smaller (e.g., Tyler's estimator) or unchanged (e.g., Student's  $M$ -estimator) comparing to the scale factor given in Eq. (10). This ensures the strong "proximity" between  $M$ -estimators and SCM, justifying the approximation of  $M$ -estimators behavior thanks to a Wishart distribution.
- 2) The results derived in this paper show that all  $M$ -estimators are asymptotically closer to the SCM than to the true covariance matrix. By "close", we mean that the asymptotic variance when centering about the Gaussian-based SCM is much smaller than the one when centering about true scatter matrix. Also, this difference is more obvious when the dimension  $m$  increases. This remark is of course also obvious for Tyler's estimator.
- 3) An important consequence of the previous remark is that any  $M$ -estimators (including Tyler's one) behavior can be approximated by the SCM one (built with Gaussian random vectors), namely by the Wishart distribution. This is of great interest in practice since all the analytical performance of functionals of robust scatter estimators can be derived based on its equivalent for the simplest Wishart distribution, while keeping the inherent robustness brought by  $M$ -estimators (contrary to the SCM). To summarize, robust estimators are better approximated by Wishart distribution than by the asymptotic Gaussian distribution with the true scatter matrix as mean.
- 4) Another comment is that, roughly speaking, one has the following result for any robust scatter matrix estimator  $\hat{\mathbf{M}}$ :

$$\sqrt{mK} \left( \hat{\mathbf{M}} - \hat{\mathbf{M}}_{SCM} \right) \xrightarrow{K \rightarrow \infty} \mathcal{CN}(\mathbf{0}, \mathbf{D}, \mathbf{Q})$$

where  $\mathbf{D}$  and  $\mathbf{Q}$  are "fixed". Thus, one has a gain in terms of convergence of  $m$ . This is agreement with the results obtained in [37] for a different convergence regime ( $m, K \rightarrow \infty$  with  $m/K$  tending to a positive constant).

- 5) Finally, it should be pointed out that the results can be applied to various signal processing problems. One can note that the scaled variance of the robust Mahalanobis distance when centering around the one based on the SCM in a Gaussian context depends only on the scale factors given by Eq. (20). This directly leads to the conclusion that the distribution of the robust distance can be better approximated with the one of the SCM-based distance, than with the asymptotical chi-squared distribution. These results can be extended to various problems such as detection or classification problems (see e.g., [35]).

## IV. SIMULATIONS

### A. Validation of the theoretical results

In this section we first present some simulations that validate the theoretical results of Theorem III.1. Figure 1 presents the empirical mean<sup>6</sup> norm of the difference between the empirical covariance matrix of  $\sqrt{K}(\sigma\hat{\mathbf{M}} - \hat{\mathbf{M}}_{SCM})$  (Eq. (18)), denoted as  $\Sigma^{(K)}$  and the theoretical results obtained in Theorem III.1. The plotted results are obtained from  $t$ -distributed data with a DoF  $\nu$  set to 2 and using the Student's  $M$ -estimator (for which theoretical results are explicitly given in Corollary III.1). The

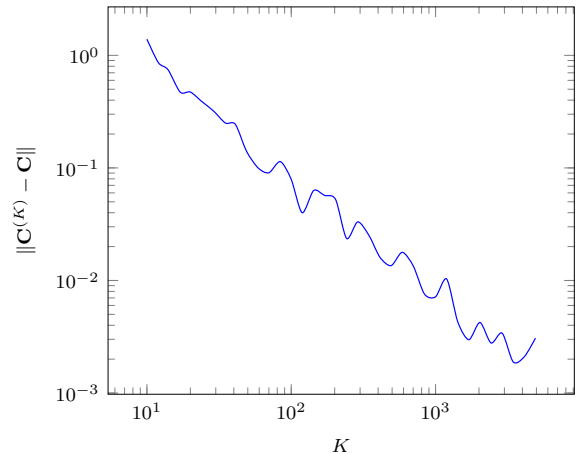
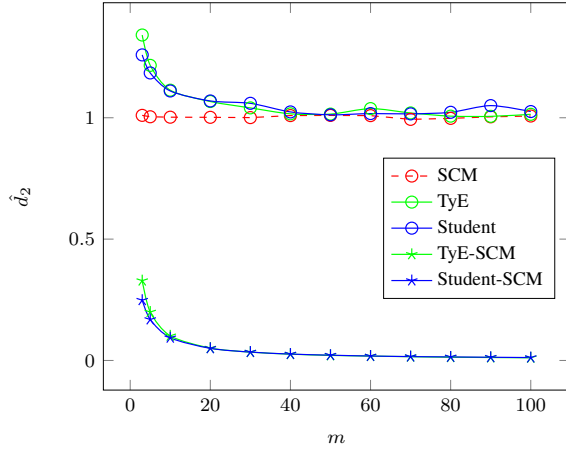


Figure 1: Euclidean norm of the difference between the empirical covariance matrix of Eq. (18) and the theoretical result Eq. (19) with  $\sigma_1$  and  $\sigma_2$  of Corollary III.1.  $m = 5$

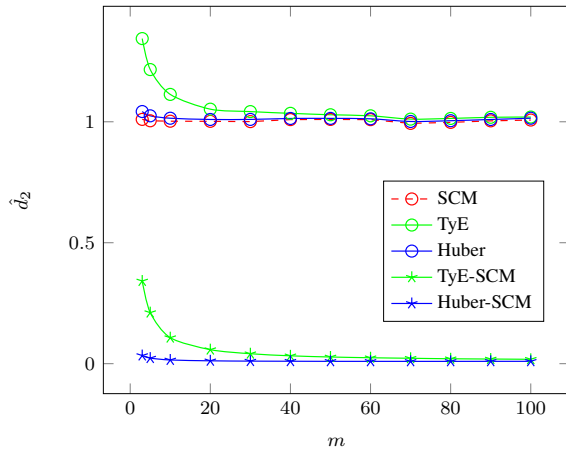
scatter matrix  $\mathbf{M}$  is defined by  $M_{i,j} = \rho^{|i-j|}$ ,  $i = 1, \dots, m$ . The correlation coefficient  $\rho$  is set to 0, i.e. the scatter matrix is equal to the identity matrix. One can notice that Figure 1 validates results obtained in Theorem III.1 since the quantity tends to zero when the number  $K$  of samples tends to infinity.

Recall that following Remark III.1 when the scatter matrix is equal to identity, the matrices  $\Sigma$  and  $\Omega$  contain only three

<sup>6</sup>obtained as the empirical mean of the quantities obtained from  $I$  Monte Carlo runs ( $I = 10K$ )



(a) Student-t distribution



(b) Gaussian plus outliers

Figure 2: Empirical second diagonal element in the asymptotic covariance matrices versus the dimension  $m$  for  $K = 1000$ .

- SCM corresponds to the empirical version of 1 in Eq. (7)
- TyE: empirical version of  $\vartheta_1 = m^{-1}(m+1)$  in Eq. (12),
- Huber: empirical version of  $\vartheta_1$  in Eq. (10) for Huber  $M$ -estimator,
- Student: empirical version of  $\vartheta_1$  in Eq. (10) for the Student  $M$ -estimator,
- TyE-SCM: empirical version of  $1/m$  of Theorem III.2,
- Huber-SCM: empirical version of  $\sigma_1$  of Theorem III.1 computed with Huber  $M$ -estimator,
- Student-SCM: empirical version of  $\sigma_1 = (m + \nu/2)^{-1}$  of Corollary III.1

different non-null elements:  $\sigma_1 + \sigma_2$ ,  $\sigma_1$  and  $\sigma_2$ . Here, we will compare the empirical value of  $\sigma_1$  to the empirical value of  $\vartheta_1$  (first scale factor of the empirical covariance matrix of  $\sqrt{K}(\sigma\hat{\mathbf{M}} - \mathbf{M})$ ) in distinct non-Gaussian environments. Results are similar for other coefficients and will be omitted.

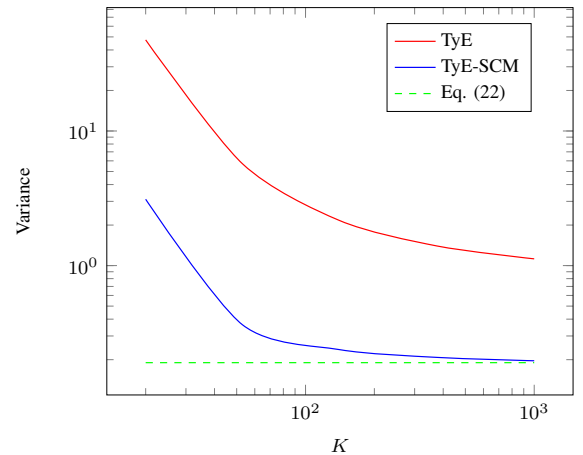
Figure 2 presents results in various non-Gaussian cases. On Figure 2(a), the results obtained for complex  $t$ -distributed data ( $\nu = 2$ ) are presented. The second diagonal element for Tyler's and Student's  $M$ -estimator are plotted. The horizontal scale presents the dimension of the data. The number of samples  $K$  is set to 1000. One can notice that the second diagonal element for  $M$ -estimators vanishes when  $m$  increases, as expected.

Indeed, if we look at the results from Theorem III.2 and Corollary III.1, the first scale factor is inversely proportional to the dimension  $m$ .

On Figure 2(b), we present the results for Tyler's and Huber's  $M$ -estimators when the data are corrupted by some outliers. The parameter  $q$  for Huber's  $M$ -estimator is set to 0.95, which means that 95% of the data are considered to be Gaussian distributed while the remaining 5% are treated as outliers. As it can be noted, the results are the same as on Figure 2(a), showing the robustness of these two estimators and validating the theoretical results. Generally speaking, these tests show that  $M$ -estimators are better characterized by a Wishart distribution than a Gaussian distribution centered on the true matrix  $\mathbf{M}$ .

### B. Application to Mahalanobis distance

We now present results for the robust Mahalanobis distance. On Figure 3(a), the results for Tyler's  $M$ -estimator are presented when data follow a complex  $t$ -distribution with  $\nu = 2$ .



(a) Tyler's estimator

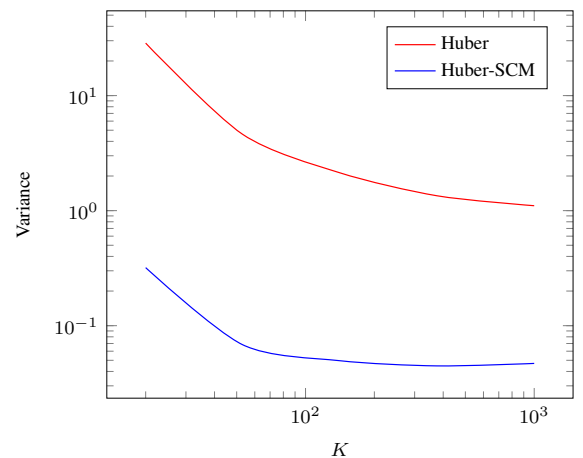
(b) Huber's  $M$ -estimator

Figure 3: Scaled empirical variance of the robust Mahalanobis distance compared to the one when centering around SCM-based distance (with the result of Theorem III.3) versus  $K$ ,  $m = 10$



The empirical variance of the robust distance and the one of the difference between the robust distance and the distance computed with the SCM in a Gaussian context (compared to the theoretical result of Theorem 21 - Eq. (22)) are plotted. On Figure 3(b) the results for Huber's  $M$ -estimator are plotted. 95% of the data follow a Gaussian distribution, while the outliers (remaining 5% of the data) are modelled with  $t$ -distribution ( $\nu = 2$ ). One can notice that the value of the robust distance is much closer to the one based on the SCM, than to the distance computed with the true scatter matrix, which once again justifies the statement that the behavior of  $M$ -estimators can be approximated by a Wishart distribution. This also implies that the distribution of robust distances can be better approximated with a theoretical distribution of the SCM-based distance in the Gaussian framework than with the asymptotic distribution based on the true scatter matrix.

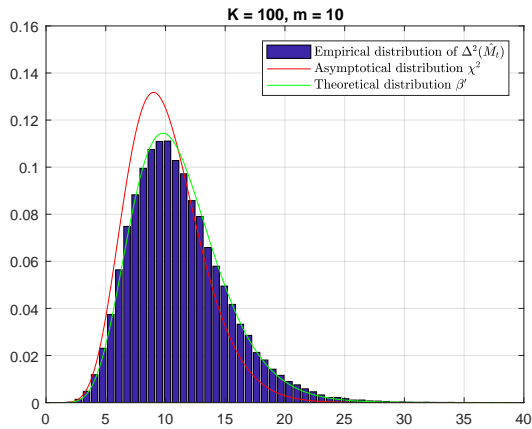


Figure 4: Histogram distribution of the robust Mahalanobis distance based on Student's  $M$ -estimator  $\Delta^2(\widehat{\mathbf{M}}_t)$  versus asymptotic distribution (Eq. (15)) in red and versus theoretical approximative distribution (Eq. (16)) in green where  $m = 10$ ,  $K = 100$ ,  $\mathbf{z} \sim \mathcal{C}t_\nu$  with  $\nu = 2$ .

Figure 4 presents the empirical distribution of the robust Mahalanobis distance built with the Student's  $M$ -estimator and two corresponding distributions proposed in Eqs. (15) and (16) for  $t$ -distributed data with  $\nu = 2$  ( $K = 100, m = 10$ ). We observe that the empirical distribution matches significantly better the scaled Beta prime than the scaled chi-squared distribution. The essential advantage of these findings in, for instance, outliers detection is that they support the idea to use the robust  $M$ -estimators to estimate the scatter matrix and to rely on the theoretical distribution of the Wishart-based distance when computing the detection threshold.

## V. CONCLUSIONS

This paper investigated the statistical properties of  $M$ -estimators. To that end, a new "Gaussian-core" model has been introduced for CES distributions. We have proposed a new approach that consists in comparing  $M$ -estimators to the well-known Gaussian-based SCM in order to derive new properties. In other words, the approach can be summarized as

follows: explaining the behavior of an "intractable" estimator  $\hat{\theta}$  by analyzing its proximity with a well-known estimator  $\hat{\theta}_1$ . It has been shown that the second order statistics of  $M$ -estimators when centering around a Wishart distributed matrix are much smaller than the ones when centering around the true scatter matrix. It has also been revealed that this difference is even more meaningful for high-dimensional data. It should be stressed that these results provide a better approximation of  $M$ -estimators properties than any other analyses in the literature. In our view, these results represent an excellent initial step toward the better understanding of the behavior of  $M$ -estimators applied in various problems. In this work, we have presented the application to the widely used Mahalanobis distance. This approach could also be applied to the adaptive detection problems and thus, be very helpful in the improvement of the detection performances. Moreover, one potential application of our findings can be found in polarimetric SAR images restoration, clustering and/or target detection. To conclude, we are confident that the results of this work are very promising and that can be applied to a wide range of signal processing problems.

## APPENDIX A PROOF OF THEOREM III.1

To prove the statement let us rewrite the right hand side of equation (19) as follows:

$$\begin{aligned} & \sqrt{K} \left( \text{vec}(\sigma\widehat{\mathbf{M}} - \widehat{\mathbf{M}}_{SCM}) \right) = \\ & \sqrt{K} \left( \text{vec}(\sigma\widehat{\mathbf{M}} - \mathbf{M} - \widehat{\mathbf{M}}_{SCM} + \mathbf{M}) \right) \\ & = [\mathbf{1}, -\mathbf{1}] \begin{bmatrix} \sqrt{K} \text{vec}(\sigma\widehat{\mathbf{M}} - \mathbf{M}) \\ \sqrt{K} \text{vec}(\widehat{\mathbf{M}}_{SCM} - \mathbf{M}) \end{bmatrix} \end{aligned}$$

Therefore one has  $\Sigma^{(K)} = \Sigma_1^{(K)} - 2\Sigma_2^{(K)} + \Sigma_3^{(K)}$  with

$$\begin{aligned} \Sigma_1^{(K)} &= KE \left[ \text{vec}(\sigma\widehat{\mathbf{M}} - \mathbf{M}) \text{vec}(\sigma\widehat{\mathbf{M}} - \mathbf{M})^H \right] \\ \Sigma_2^{(K)} &= KE \left[ \text{vec}(\sigma\widehat{\mathbf{M}} - \mathbf{M}) \text{vec}(\widehat{\mathbf{M}}_{SCM} - \mathbf{M})^H \right] \\ \Sigma_3^{(K)} &= KE \left[ \text{vec}(\widehat{\mathbf{M}}_{SCM} - \mathbf{M}) \text{vec}(\widehat{\mathbf{M}}_{SCM} - \mathbf{M})^H \right]. \end{aligned}$$

One has now

$$\Sigma^{(K)} \xrightarrow{K \rightarrow +\infty} \Sigma = \Sigma_1(\mathbf{M}) - 2\Sigma_2(\mathbf{M}) + \Sigma_3(\mathbf{M}) \quad (23)$$

where the matrices  $\Sigma_1(\mathbf{M})$  and  $\Sigma_3(\mathbf{M})$  are given by (10) and (7), respectively.

Following the similar ideas used in [18], [22], we provide a more general result that allows to compute a correlation between two estimators

$$\Sigma_2^{(K)} \xrightarrow{K \rightarrow +\infty} \Sigma_2(\mathbf{M}) = \mathbf{D}_1^{-1}(\mathbf{M})\mathbf{B}(\mathbf{M})\mathbf{D}_2^{-1}(\mathbf{M})$$

$$\begin{aligned} \mathbf{D}_1(\mathbf{M}) &= E [d\{\text{vec}\Psi_1(\mathbf{M})\}/d\{\text{vec}(\mathbf{M})\}], \\ \mathbf{B}(\mathbf{M}) &= \text{cov}(\text{vec}\Psi_1(\mathbf{M}), \text{vec}\Psi_2(\mathbf{M})) \quad \text{and} \\ \mathbf{D}_2(\mathbf{M}) &= E [d\{\text{vec}\Psi_2(\mathbf{M})\}/d\{\text{vec}(\mathbf{M})\}] \quad \text{with} \\ \Psi_1(\mathbf{M}) &= \sigma\varphi(\mathbf{z}^H(\sigma^{-1}\mathbf{M})^{-1}\mathbf{z})\mathbf{z}\mathbf{z}^H - \mathbf{M} \quad \text{and} \\ \Psi_2(\mathbf{M}) &= \mathbf{x}\mathbf{x}^H - \mathbf{M}. \end{aligned}$$

Without loss of generality, we will assume that  $\mathbf{M} = \mathbf{I}$ . Indeed, one has that

$$\Sigma_2(\mathbf{M}) = \left( \mathbf{M}^{T/2} \otimes \mathbf{M}^{1/2} \right) \Sigma_2(\mathbf{I}) \left( \mathbf{M}^{T/2} \otimes \mathbf{M}^{1/2} \right)^H.$$

In order to determine the final result, we will derive the expression for  $\Sigma_2(\mathbf{I})$ . One can show that

$$\mathbf{D}_1^{-1}(\mathbf{I}) = \alpha_1 \mathbf{I} + \alpha_2 \text{vec}(\mathbf{I}) \text{vec}(\mathbf{I})^T$$

where  $\alpha_1 = -\frac{m(m+1)}{c}$  and  $\alpha_2 = \frac{m(c^2 - m^2 - m)}{c(c-m^2)}$  with  $c = E[\sigma t_1 \psi'(\sigma t_1)] + m^2$ . Moreover, it is simple to show that  $\mathbf{D}_2(\mathbf{I})^{-1} = -\mathbf{I}$ .

Then, basing on Theorem 2 from [22], one can derive more general result

$$\mathbf{B}_2(\mathbf{I}) = \beta_1 \mathbf{I} + \beta_2 \text{vec}(\mathbf{I}) \text{vec}(\mathbf{I})^T$$

where

$$\begin{aligned} \beta_1 &= \text{cov}[\Psi_1(\mathbf{I})_{jk} \Psi_2(\mathbf{I})_{jk}] \\ &= E[\varphi(\sigma t_1) t_2 u_j^2 u_k^2] \\ &= \frac{E[\psi(\sigma t_1) t_2]}{m(m+1)} = \frac{b}{m(m+1)} \end{aligned}$$

and

$$\begin{aligned} \beta_2 &= \text{cov}[\Psi_1(\mathbf{I})_{jj} \Psi_2(\mathbf{I})_{kk}] \\ &= \beta_1 - E[\psi(\sigma t_1)] E[t_2] / m^2 \\ &= \beta_1 - 1 \end{aligned}$$

since  $u_i^2 \sim \text{Beta}(1, m-1)$ ,  $E[u_j^2] = 1/m$  and  $E[u_j^2 u_k^2] = 1/(m(m+1))$ . After some mathematical manipulations, one obtains

$$\Sigma_2(\mathbf{I}) = \gamma_1 \mathbf{I} + \gamma_2 \text{vec}(\mathbf{I}) \text{vec}(\mathbf{I})^T$$

with

$$\begin{aligned} \gamma_1 &= -\alpha_1 \beta_1 = \frac{b}{c} \\ \gamma_2 &= -(\alpha_1 \beta_2 + 2\alpha_2 \beta_1 + 2m\alpha_2 \beta_2) = \frac{m(b-c)}{c(c-m^2)}. \end{aligned} \quad (24)$$

This leads to the final expression of  $\Sigma_2(\mathbf{M})$ :

$$\Sigma_2(\mathbf{M}) = \gamma_1 \mathbf{M}^T \otimes \mathbf{M} + \gamma_2 \text{vec}(\mathbf{M}) \text{vec}(\mathbf{M})^H \quad (25)$$

Combining Eq. (25) together with Eqs. (10) and (7) in Eq. (23), one obtains the coefficients  $\sigma_1$  and  $\sigma_2$  as follows

$$\begin{aligned} \sigma_1 &= \vartheta_1 - 2\gamma_1 + 1 \\ &= \frac{am(m+2) + c(c-2b)}{c^2} \end{aligned}$$

and

$$\begin{aligned} \sigma_2 &= \vartheta_2 - 2\gamma_2 \\ &= \frac{a-m^2}{(c-m^2)^2} - \frac{a(m+2)}{c^2} + 2 \frac{m(c-b)}{c(c-m^2)}. \end{aligned}$$

Finally, one can easily prove that  $\Omega = \Sigma \mathbf{K}$  [6], which leads to the final results and concludes the proof.

## APPENDIX B PROOF OF COROLLARY III.1

For Student's  $t$ -distribution  $t_1 \sim mF_{2m, \nu}$  yields

$$f(t_1) = C_m t_1^{m-1} \left( 1 + \frac{2}{\nu} t_1 \right)^{-\frac{2m+\nu}{2}}$$

with

$$C_m = \left( \frac{2}{\nu} \right)^m \frac{\Gamma(m + \frac{\nu}{2})}{\Gamma(m) \Gamma(\frac{\nu}{2})}$$

where  $\Gamma(\cdot)$  is the Gamma function. Since  $\sigma = 1$  for every  $ML$ -estimator [1], one has

$$\psi(\sigma t_1) = \psi(t_1) = \frac{2m + \nu}{\nu + 2t_1} t_1$$

Now one obtains

$$\begin{aligned} E[\psi^2(t_1)] &= E\left[ \frac{(2m + \nu)^2 t_1^2}{(\nu + 2t_1)^2} \right] \\ &= C_m \int_0^{+\infty} \frac{(2m + \nu)^2 t_1^2}{\nu^2 (1 + \frac{2}{\nu} t_1)^2} t_1^{m-1} \left( 1 + \frac{2t_1}{\nu} \right)^{-\frac{2m+\nu}{2}} dt_1 \\ &= C_m \frac{(2m + \nu)^2}{\nu^2} \frac{1}{C_{m+2}} \\ &= \frac{m(m+1)(m + \frac{\nu}{2})}{m+1 + \frac{\nu}{2}} \end{aligned}$$

and

$$\begin{aligned} E[t_1 \psi'(t_1)] &= E\left[ \frac{(2m + \nu)\nu}{(\nu + 2t_1)^2} t_1 \right] \\ &= C_m \int_0^{+\infty} \frac{(2m + \nu)\nu t_1}{\nu^2 (1 + \frac{2}{\nu} t_1)^2} t_1^{m-1} \left( 1 + \frac{2t_1}{\nu} \right)^{-\frac{2m+\nu}{2}} dt_1 \\ &= C_m \frac{\nu(2m + \nu)}{\nu^2} \int_0^{+\infty} t_1^{-1} t_1^{m+1} \left( 1 + \frac{2t_1}{\nu} \right)^{-\frac{2m+4+\nu}{2}} dt_1 \\ &= C_m \frac{2m + \nu}{\nu} \frac{1}{C_{m+2}} E[t_1^{-1}] \end{aligned}$$

where now  $t/(m+2) \sim F_{2m+4, \nu}$  or equivalently  $(m+2)/t \sim F_{\nu, 2m+4}$  which gives

$$E[t_1^{-1}] = \frac{1}{m+2} \frac{2m+4}{2m+4-2} = \frac{1}{m+1}$$

and finally

$$E[t \psi'(t_1)] = \frac{\nu}{2} \frac{m}{(m+1 + \frac{\nu}{2})}.$$

To compute  $E[\psi(t_1) t_2]$  let us remind that  $t_1 = \tau t_2$  where  $\tau$  and  $t_2$  are independent,  $\tau \sim \text{IG}(\nu/2, \nu/2)$  and  $t_2 \sim (1/2)\chi_{2m}^2$ . Thus, one can write

$$I = E[\psi(t_1) t_2] = C \iint_{\mathbb{R}_+^2} \frac{1}{1 + \frac{2\tau}{\nu} t_2} \tau^{-\frac{\nu}{2}} e^{-\frac{\nu}{2\tau}} t_2^{m+1} e^{-t_2} d\tau dt_2$$

where  $C = (\frac{2m}{\nu} + 1)^{\frac{\nu}{2}} / (\Gamma(\frac{\nu}{2}) \Gamma(m))$ . The change of variable  $u = \frac{2\tau}{\nu} t_2$  gives  $du = \frac{2\tau}{\nu} dt_2$  and hence

$$I = C \int_0^\infty \frac{\nu}{2\tau} \tau^{-\frac{\nu}{2}} e^{-\frac{\nu}{2\tau}} \left( \frac{\nu}{2\tau} \right)^{m+1} \int_0^\infty \frac{1}{1+u} u^{m+1} e^{-\frac{\nu}{2\tau} u} du d\tau.$$

Then, using the equality

$$\int_0^\infty \frac{1}{1+u} u^{m+1} e^{-\frac{\nu}{2\tau}u} du = e^{\frac{\nu}{2\tau}} (m+1)! \Gamma\left(-1-m; \frac{\nu}{2\tau}\right)$$

where  $\Gamma(\cdot; \cdot)$  stands for the upper incomplete Gamma function, one obtains

$$I = C' \int_0^\infty \tau^{-\frac{\nu}{2}-m-2} \Gamma\left(-1-m; \frac{\nu}{2\tau}\right) d\tau$$

where  $C' = C(m+1)! \left(\frac{\nu}{2}\right)^{m+2}$ . Since

$$\Gamma\left(-1-m; \frac{\nu}{2\tau}\right) = \left(\frac{\nu}{2\tau}\right)^{-m-1} E_{m+2}\left(\frac{\nu}{2\tau}\right)$$

where  $E_{m+2}$  is the generalized exponential integral, one has

$$I = C'' \int_0^\infty \tau^{-\frac{\nu}{2}-1} E_{m+2}\left(\frac{\nu}{2\tau}\right) d\tau$$

where  $C'' = C' \left(\frac{\nu}{2}\right)^{m+1}$  which leads to

$$\begin{aligned} I &= C'' \int_0^\infty \tau^{-\frac{\nu}{2}-1} \int_1^\infty e^{-\frac{\nu}{2\tau}t} t^{-m-2} dt d\tau \\ &= C'' \int_1^\infty t^{-m-2} \int_0^\infty \tau^{-\frac{\nu}{2}-1} e^{-\frac{\nu}{2\tau}t} dt d\tau \\ &= C''' \int_1^\infty t^{-m-2-\frac{\nu}{2}} dt = \frac{C'''}{m+1+\frac{\nu}{2}} \end{aligned}$$

where  $C''' = C'' \left(\frac{\nu}{2}\right)^{-\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)$  and finally

$$E[\psi(t_1)t_2] = \frac{(m+\frac{\nu}{2})m(m+1)}{m+1+\frac{\nu}{2}}.$$

This leads to the following values for  $a, b, c$

$$\begin{aligned} a &= \frac{m(m+1)(m+\frac{\nu}{2})}{m+1+\frac{\nu}{2}}, \\ b &= \frac{m(m+1)(m+\frac{\nu}{2})}{m+1+\frac{\nu}{2}}, \\ c &= \frac{\nu}{2} \frac{m}{(m+1+\frac{\nu}{2})} + m^2 = \frac{m(m+1)(m+\frac{\nu}{2})}{m+1+\frac{\nu}{2}}. \end{aligned}$$

Substituting previous results in Eq. (20), one finally obtains

$$\sigma_1 = \frac{1}{m+\nu/2} \quad \text{and} \quad \sigma_2 = \frac{2(m+1+\nu/2)}{\nu(m+\nu/2)}.$$

### APPENDIX C

#### PROOF OF THEOREM III.3

To prove the statement of Theorem III.3, we will rewrite  $\phi$  as

$$\phi = (\phi_M - 2\phi_{corr} + \phi_{SCM}) / (\mathbf{y}^H \mathbf{M}^{-1} \mathbf{y})^2 \quad (26)$$

where

$$\begin{aligned} \phi_M &= E \left[ \left( f(\sigma \widehat{\mathbf{M}}) - f(\mathbf{M}) \right)^2 \right] \\ \phi_{corr} &= E \left[ \left( f(\sigma \widehat{\mathbf{M}}) - f(\mathbf{M}) \right) \left( f(\widehat{\mathbf{M}}_{SCM}) - f(\mathbf{M}) \right) \right] \\ \phi_{SCM} &= E \left[ \left( f(\widehat{\mathbf{M}}_{SCM}) - f(\mathbf{M}) \right)^2 \right] \end{aligned}$$

with  $f(\mathbf{M}) = \mathbf{y}^H \mathbf{M}^{-1} \mathbf{y}$ . Using the Delta method one can obtain  $\phi_M = f'(\mathbf{M}) \Sigma_M f'(\mathbf{M})^H$  [17] where  $f'(\mathbf{M})$

is the first derivative with respect to  $\mathbf{M}$  and  $\Sigma_M = E \left[ \left( \sigma \widehat{\mathbf{M}} - \mathbf{M} \right) \left( \sigma \widehat{\mathbf{M}} - \mathbf{M} \right)^H \right]$  is defined by Eq. (10) (with  $\mathbf{M} = \sigma \mathbf{M}_\sigma$ ). Moreover, one has a more general result  $\phi_{corr} = f'(\mathbf{M}) \Sigma_2 f'(\mathbf{M})^H$  where  $\Sigma_2 = E \left[ \left( \sigma \widehat{\mathbf{M}} - \mathbf{M} \right) \left( \widehat{\mathbf{M}}_{SCM} - \mathbf{M} \right)^H \right] = \gamma_1 (\mathbf{M}^T \otimes \mathbf{M}) + \gamma_2 \text{vec}(\mathbf{M}) \text{vec}(\mathbf{M})^H$  with  $\gamma_1$  and  $\gamma_2$  the complex versions of Eq. (24). In [?] it has been shown that  $f'(\mathbf{M}) = \text{vec}^H(\mathbf{y}\mathbf{y}^H) (\mathbf{M}^T \otimes \mathbf{M})^{-1}$ . Since  $(\mathbf{M}^T \otimes \mathbf{M})^{-1} \text{vec}(\mathbf{M}) = \text{vec}(\mathbf{M}^{-1})$  one has

$$\begin{aligned} \phi_M &= f'(\mathbf{M}) \Sigma f'(\mathbf{M})^H \\ &= \vartheta_1 \text{vec}^H(\mathbf{y}\mathbf{y}^H) (\mathbf{M}^T \otimes \mathbf{M})^{-1} \text{vec}(\mathbf{y}\mathbf{y}^H) \\ &\quad + \vartheta_2 \text{vec}^H(\mathbf{y}\mathbf{y}^H) \text{vec}(\mathbf{M}^{-1}) \text{vec}(\mathbf{M}^{-1})^H \text{vec}(\mathbf{y}\mathbf{y}^H) \\ &= \vartheta_1 \text{vec}^H(\mathbf{y}\mathbf{y}^H) \text{vec}(\mathbf{M}^{-1} \mathbf{y}\mathbf{y}^H \mathbf{M}^{-1}) \\ &\quad + \vartheta_2 \text{Tr}(\mathbf{y}\mathbf{y}^H \mathbf{M}^{-1}) \text{Tr}(\mathbf{M}^{-1} \mathbf{y}\mathbf{y}^H) \\ &= \vartheta_1 \text{Tr}(\mathbf{y}^H \mathbf{M}^{-1} \mathbf{y}\mathbf{y}^H \mathbf{M}^{-1} \mathbf{y}) + \vartheta_2 \text{Tr}(\mathbf{y}^H \mathbf{M}^{-1} \mathbf{y})^2 \\ &= (\vartheta_1 + \vartheta_2) (\mathbf{y}^H \mathbf{M}^{-1} \mathbf{y})^2. \end{aligned}$$

It is now clear that  $\phi_{SCM} = (\mathbf{y}^H \mathbf{M}^{-1} \mathbf{y})^2$  and  $\phi_{corr} = f'(\mathbf{M}) \Sigma_2 f'(\mathbf{M})^H = (\gamma_1 + \gamma_2) (\mathbf{y}^H \mathbf{M}^{-1} \mathbf{y})^2$  which leads to the final result.

### REFERENCES

- [1] 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.
- [2] 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.
- [3] Y. Chen, A. Wiesel, and A. O. Hero, "Robust shrinkage estimation of high-dimensional covariance matrices," *Signal Processing, IEEE Transactions on*, vol. 59, no. 9, pp. 4097–4107, 2011.
- [4] A. Wiesel, "Unified framework to regularized covariance estimation in scaled Gaussian models," *Signal Processing, IEEE Transactions on*, vol. 60, no. 1, pp. 29–38, 2012.
- [5] F. Pascal, L. Bombrun, J.-Y. Tourneret, and Y. Berthoumieu, "Parameter estimation for multivariate generalized Gaussian distributions," *Signal Processing, IEEE Transactions on*, vol. 61, no. 23, pp. 5960–5971, December 2013.
- [6] 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.
- [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] 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.
- [9] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*. Chapman & Hall/CRC, 2000.
- [10] M. Bilodeau and D. Brenner, *Theory of Multivariate Statistics*, ser. New York, NY. USA: Springer-Verlag, 1999.
- [11] D. Kelker, "Distribution theory of spherical distributions and a location-scale parameter generalization," *Sankhyā: The Indian Journal of Statistics, Series A*, vol. 32, no. 4, pp. 419–430, December 1970.
- [12] F. Gini and M. S. Greco, "Sub-optimum approach to adaptive coherent radar detection in Compound-Gaussian clutter," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 35, no. 3, pp. 1095–1103, July 1999.
- [13] F. Gini, M. S. Greco, M. Diani, and L. Verrazzani, "Performance analysis of two adaptive radar detectors against non-Gaussian real sea clutter data," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 36, no. 4, pp. 1429–1439, October 2000.

- [14] F. Gini and M. S. Greco, "Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter," *Signal Processing, special section on SP with Heavy Tailed Distributions*, vol. 82, no. 12, pp. 1847–1859, December 2002.
- [15] E. Conte and M. Longo, "Characterization of radar clutter as a spherically invariant random process," *IEE Proceeding, Part. F*, vol. 134, no. 2, pp. 191–197, April 1987.
- [16] E. Conte, A. De Maio, and G. Ricci, "Covariance matrix estimation for adaptive cfar detection in Compound-Gaussian clutter," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 38, no. 2, pp. 415–426, April 2002.
- [17] P. J. Huber, "Robust estimation of a location parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, January 1964.
- [18] R. A. Maronna, "Robust  $M$ -estimators of multivariate location and scatter," *Annals of Statistics*, vol. 4, no. 1, pp. 51–67, January 1976.
- [19] J. T. Kent and D. E. Tyler, "Maximum likelihood estimation for the wrapped Cauchy distribution," *Journal of Applied Statistics*, vol. 15, no. 2, pp. 247–254, 1988.
- [20] A. Balleri, A. Nehorai, and J. Wang, "Maximum likelihood estimation for Compound-Gaussian clutter with inverse gamma texture," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 43, no. 2, pp. 775–779, April 2007.
- [21] D. E. Tyler, "Radial estimates and the test for sphericity," *Biometrika*, vol. 69, no. 2, p. 429, 1982.
- [22] —, "A distribution-free  $M$ -estimator of multivariate scatter," *The Annals of Statistics*, vol. 15, no. 1, pp. 234–251, 1987.
- [23] F. Pascal, P. Forster, J.-P. Ovarlez, and P. Larzabal, "Performance analysis of covariance matrix estimates in impulsive noise," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2206–2217, June 2008.
- [24] G. Drašković and F. Pascal, "New properties for Tyler's covariance matrix estimator," in *2016 50th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, Nov 2016, pp. 820–824.
- [25] P. C. Mahalanobis, "On the generalized distance in statistics," *Proceedings of the National Institute of Sciences (Calcutta)*, vol. 2, pp. 49–55, 1936.
- [26] R. D. Maesschalck, D. Jouan-Rimbaud, and D. Massart, "The Mahalanobis distance," *Chemometrics and Intelligent Laboratory Systems*, vol. 50, no. 1, pp. 1 – 18, 2000. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0169743999000477>
- [27] 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.
- [28] P. J. Rousseeuw and B. C. Van Zomeren, "Unmasking multivariate outliers and leverage points," *Journal of the American Statistical Association*, vol. 85, no. 411, pp. 633–639, 1990.
- [29] A. S. Hadi, "Identifying multiple outliers in multivariate data," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 761–771, 1992.
- [30] S. Xiang, F. Nie, and C. Zhang, "Learning a Mahalanobis distance metric for data clustering and classification," *Pattern Recognition*, vol. 41, no. 12, pp. 3600 – 3612, 2008. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0031320308002057>
- [31] K. Q. Weinberger, J. Blitzer, and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," in *Advances in neural information processing systems*, 2006, pp. 1473–1480.
- [32] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern recognition letters*, vol. 15, no. 11, pp. 1119–1125, 1994.
- [33] C.-I. Chang and S.-S. Chiang, "Anomaly detection and classification for hyperspectral imagery," *IEEE transactions on geoscience and remote sensing*, vol. 40, no. 6, pp. 1314–1325, 2002.
- [34] J. Frontera-Pons, M. A. Veganzones, S. Velasco-Forero, F. Pascal, J. P. Ovarlez, and J. Chanussot, "Robust anomaly detection in hyperspectral imaging," in *2014 IEEE Geoscience and Remote Sensing Symposium*, July 2014, pp. 4604–4607.
- [35] G. Drašković, F. Pascal, A. Breloy, and J.-Y. Tournet, "New asymptotic properties for the robust ANMF," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-17*, New Orleans, USA, March 2017.
- [36] R. Bhatia and C. Davis, "A better bound on the variance," *The American Mathematical Monthly*, vol. 107, no. 4, pp. 353–357, 2000.
- [37] 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.
- [38] F. Pascal and J.-P. Ovarlez, "Asymptotic Properties of the Robust ANMF," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-15*, Brisbane, Australia, April 2015.