Convergence and Fluctuations of Regularized Tyler Estimators

Abla Kammoun, Romain Couillet, Frédéric Pascal, Mohamed-Slim Alouini

Convergence and Fluctuations of Regularized Tyler Estimators

Abla Kammoun, Romain Couillet, Frédéric Pascal, Mohamed-Slim Alouini

***Abstract*—This article studies the behavior of regularized Tyler estimators (RTEs) of scatter matrices. The key advantages of these estimators are twofold. First, they guarantee by construction a good conditioning of the estimate and second, being a derivative of robust Tyler estimators, they inherit their robustness properties, notably their resilience to the presence of outliers. Nevertheless, one major problem that poses the use of RTEs in practice is represented by the question of setting the regularization parameter ρ. While a high value of ρ is likely to push all the eigenvalues away from zero, it comes at the cost of a larger bias with respect to the population covariance matrix. A deep understanding of the statistics of RTEs is essential to come up with appropriate choices for the regularization parameter. This is not an easy task and might be out of reach, unless one considers asymptotic regimes wherein the number of observations n and/or their size N increase together. First asymptotic results have recently been obtained under the assumption that N and n are large and commensurable. Interestingly, no results concerning the regime of n going to infinity with N fixed exist, even though the investigation of this assumption has usually predated the analysis of the most difficult N and n large case. This motivates our work. In particular, we prove in the present paper that the RTEs converge to a deterministic matrix when $n \to \infty$ with N fixed, which is expressed as a function of the theoretical covariance matrix. We also derive the fluctuations of the RTEs around this deterministic matrix and establish that these fluctuations converge in distribution to a multivariate Gaussian distribution with zero mean and a covariance depending on the population covariance and the parameter ρ.**

I. Introduction

The estimation of covariance matrices is at the heart of many applications in signal processing and wireless communications. The frequently used estimator is the well-known sample covariance matrix (SCM). Its popularity owes to its low complexity and in general to a good understanding of its behavior. However, the use of the SCM in practice is hurdled by its poor performance when samples contain outliers or have an impulsive nature. This is especially the case of radar detection applications in which the noise is often modeled by heavy-tailed distributions [1]–[4]. One of the reasons why the SCM performs poorly in such scenarios is that, as opposed to the case of Gaussian observations, the SCM is not the maximum likelihood estimator (MLE) of the covariance matrix. This is for instance the case of complex elliptical distributions, originally introduced by Kelker [5] and widely used in radar applications, for which the MLE takes a strikingly different form.

In order to achieve better robustness against outliers, a class of covariance estimators termed robust estimators of scatter have been proposed by Huber, Hampel and Maronna [6]–[8], and extended more recently to the complex case [9]–[11]. This class of estimators can be viewed as a generalization of MLEs, in that they are derived from the optimization of a meaningful cost function [12], [13]. Aside from robustness to the presence of outliers, a second feature whose importance should not be underestimated, is the conditioning of the covariance matrix estimate. This feature becomes all the more central when the quantity of interest coincides with the inverse of the population covariance matrix. In order to guarantee an acceptable conditioning, regularized robust-estimators, which find their roots in the diagonal loading technique due to Abramovich and Carlson [14], [15], were proposed in [12]. The idea is to force by construction all the eigenvalues of the robust-scatter estimator to be greater than a regularization coefficient ρ.

The most popular regularized estimators that are today receiving increasing interest, are the regularized Tyler estimators (RTE), which correspond to regularized versions of the robust Tyler estimator [16]. In addition to achieving the desired robustness property, RTEs present the advantage of being well-suited to scenarios where the number of observations is insufficient or the population covariance matrix is ill-conditioned, while their non-regularized counterparts are ill-conditioned or even undefined if the number of observations n is less than their sizes N. Motivated by these interesting features, several works have recently considered the use of RTEs in radar detection applications [12], [17]–[20]. While existence and uniqueness of the robust-scatter estimator seem to be sufficiently studied [12], [18], the impact of the regularization parameter on the behavior of the RTE has remained less understood. Answering this question is essential in order to come up with appropriate designs of the RTE in practice. It poses, however, major technical challenges, mainly because it necessitates a profound analysis of the behavior of the RTE estimator, which is far from being an easy task. As a matter of fact, the main difficulty towards studying the behavior of the RTE fundamentally lies in its non-linear relation to the observations, thus rendering the analysis for fixed n and N likely out of reach. In light of this observation, recent works have considered asymptotic regimes where n and/or N are allowed to grow to infinity. Two regimes can be distinguished:

A. Kammoun and M.S. Alouini are with the Computer, Electrical, and Mathematical Sciences and Engineering (CEMSE) Division, KAUST, Thuwal, Makkah Province, Saudi Arabia (e-mail: abla.kammoun@kaust.edu.sa, slim.alouini@kaust.edu.sa)

R. Couillet and F. Pascal are with Laboratoire des Signaux et Systèmes (L2S, UMR CNRS 8506) CentraleSupélec-CNRS-Université Paris-Sud, 91192 Gif-sur-Yvette, France (e-mail: romain.couillet@centralesupelec.fr, frederic.pascal@centralesupelec.fr)

Kammoun and Alouini's work was supported by a CRG2 grant from the office of competitive research funding at KAUST. Couillet's work was supported by the ERC MORE EC–120133. Pascal's work has been partially supported by the ICODE institute, research project of the Idex Paris-Saclay.

the regime of fixed N with n growing to infinity and the regime of n and N growing large simultaneously. While the former regime, coined the large-n regime, is standard in that it was by far the most considered in the literature, the second one, which we will refer to as large-n, N regime, is very recent and is particularly driven by the recent advances in the spectral analysis of large dimensional random matrices. Interestingly, contrary to what one would imagine, very little on the behavior of RTE seems to be known in the standard regime, whereas very recent results regarding the behavior of RTE for the large-n, N regime have recently been obtained in [20], [21]. One major advantage of the large-n, N regime is that, although requiring the use of advanced tools from random matrix theory, it often leads to less involved results that let themselves to simple interpretation. This interesting feature fundamentally inheres in the double averaging effect that leads to more compact results in which only prevailing quantities remain. However, when N is not so large, the same averaging effect is no longer valid and thus cannot be leveraged. A priori, assuming that N is fixed entails major changes on the behavior of RTEs that have not thus far been grasped. Understanding what really happens in the large-n regime, besides its own theoretical interest, should lead to alternative results that might be more accurate for not so large-N scenarios. A second motivation behind working under the large-n regime is that covariance matrix estimators usually converge in this case to deterministic matrices, which opens up possibilities for easier handling of the RTE. Encouraged by these interesting practical and theoretical aspects, we study in this paper the asymptotic behavior of the RTE in the large-n regime. One of our major findings is that our results allowing derivation of central limit theorems (CLT) of functionals with kernel the robust scatter estimator can be more accurate than those of the large-n, N regime even for relatively small values of n. This becomes all the more interesting given that apart from the specific case of quadratic forms, derivation of CLTs under the large-n, N regime can be quite involved.

The remainder of the paper is organized as follows. In section II, we prove that the RTE converges to a deterministic matrix which depends on the theoretical covariance matrix and the regularization parameter before presenting its fluctuations around this asymptotic limit in section III. Numerical results are finally provided in order to support the accuracy of the derived results.

Notation. In this paper, the following notations are used. Vectors are defined as column vectors and designated with bold lower case, while matrices are given in bold upper case. The norm notation $\|.\|$ refers to the spectral norm for matrices and Euclidean norm for vectors while the norm $\|.\|_{\rm Fro}$ refers to the Frobenius norm of matrices. Notations $(.)^{\rm T}$ $(.)^*$, $\overline{(.)}$ denotes respectively transpose, Hermitian (i.e. complex conjugate transpose) and pointwise conjugate. Besides, $\mathbf{I}_N$ denotes the $N\times N$ identity matrix, for a matrix $\mathbf{A}$, $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ denote respectively the smallest and largest eigenvalues of $\mathbf{A}$, while notation $\mathrm{vec}(\mathbf{A})$ refers to the vector obtained by stacking the columns of $\mathbf{A}$. For $\mathbf{A}$, $\mathbf{B}$ two positive semi-definite matrices, $\mathbf{A} \preceq \mathbf{B}$ means that $\mathbf{B}-\mathbf{A}$ is positive semi-definite. $X_n = o_p(1)$ implies the convergence in probability to zero of X_n as n goes to infinity and $X_n = \mathcal{O}_p(1)$ implies that X_n is bounded in probability. The arrow "$\xrightarrow{\text{a.s.}}$" designates almost sure convergence while the arrow"$\xrightarrow{\mathcal{D}}$" refers to convergence in distribution.

II. Convergence of the regularized M-estimator of scatter matrix

Consider $\mathbf{x}_1, \cdots, \mathbf{x}_n$, n observations of size N defined as:

$$\mathbf{x}_i = \mathbf{\Sigma}_N^{\frac{1}{2}} \mathbf{w}_i,$$

where $\mathbf{w}_i \in \mathbb{C}^N$ are Gaussian zero-mean random vectors with covariance $\mathbf{I}_N$ and $\mathbf{\Sigma}_N \succeq 0$ is the population covariance matrix. The regularized robust scatter estimator that will be considered in this work is that defined in [18] as the unique solution $\hat{\mathbf{C}}_N(\rho)$ to:

$$\hat{\mathbf{C}}_N(\rho) = (1-\rho)\frac{1}{n}\sum_{i=1}^{n} \frac{\mathbf{x}_i\mathbf{x}_i^*}{\frac{1}{N}\mathbf{x}_i^*\hat{\mathbf{C}}_N^{-1}(\rho)\mathbf{x}_i} + \rho\mathbf{I}_N. \qquad (1)$$

with $\rho \in \left(\max(0, 1-\frac{n}{N}), 1\right]$. Such an estimator can be thought of as a hybrid robust-shrinkage estimator reminding Tyler's M-estimator of scale [16] and Ledoit-Wolf's shrinkage estimator [22]. It will be coined thus Regularized-Tyler estimator (RTE), and defines a class of regularized-robust scatter estimators indexed by the regularization parameter ρ. When $n > N$, by varying ρ from 0 to 1, one can move from the unbiased Tyler-estimator [23] to the identity matrix ($\rho = 1$) which corresponds to a trivial estimate of the unknown covariance matrix $\mathbf{\Sigma}_N$. Finally, it is worth mentioning that other variants of the RTE estimator have recently been proposed [13], [17]. We can cite for instance the RTE estimator of Chen et al [17] which is more involved and will thus not be considered in this work [1].

A. Review of the results obtained in the large-n, N regime

Letting $c_N = \frac{N}{n}$, the large-n, N regime will refer in the sequel to the one where $n \to \infty$ and $N \to \infty$ with $c_N \to c \in (0, \infty)$.

As mentioned earlier, unless considering particular assumptions on $\mathbf{\Sigma}_N$, the RTE cannot be proven to converge (in any usual matrix norm) to some deterministic matrix in the large-n, N regime. Failing that, the approach pursued in [20] consists in determining a random equivalent for the RTE, that corresponds to a standard matrix model. This finding is of utmost importance, since it allows one to replace the RTE, whose direct analysis is overly difficult, by another random object, for which an important load of results is available. The meaning of the equivalence between the RTE and the new object will be specified below.

[1] Another concurrent RTE is that of Chen {*et al* [17] which is given as the unique solution of

$$\check{\mathbf{C}}_N(\rho) = \check{\mathbf{B}}_N(\rho)\frac{1}{N}\,\mathrm{tr}\,\check{\mathbf{B}}_N(\rho)$$

where

$$\check{\mathbf{B}}_N(\rho) = (1-\rho)\frac{1}{n}\sum_{i=1}^{n} \frac{\mathbf{x}_i\mathbf{x}_i^*}{\frac{1}{N}\mathbf{x}_i^*\check{\mathbf{C}}_N(\rho)^{-1}\mathbf{x}_i} + \rho\mathbf{I}_N.$$

Prior to presenting the results of [20], we shall, for the reader convenience, gather all the observations' properties in the following assumption:

Assumption A-1. *For $i \in \{1, \cdots, n\}$, $\mathbf{x}_i = \mathbf{\Sigma}_N^{\frac{1}{2}} \mathbf{w}_i$, with:*

- $\mathbf{w}_1, \cdots, \mathbf{w}_n$ *are* $N \times 1$ *independent Gaussian random vectors with zero mean and covariance* $\mathbf{I}_N$,
- $\mathbf{\Sigma}_N \in \mathbb{C}^{N \times N} \succeq 0$ *is such that* $\frac{1}{N} \operatorname{tr} \mathbf{\Sigma}_N = 1$.

Note that our work is also valid for observations drawn from elliptical distributions [9] for which the use of robust scatter estimators can be thought of as more relevant. This is because such observations can be written as the product of a scalar and a Gaussian random vector [9]. The scalar has no impact on the RTE estimators due to their invariance to observations scaling.

It is also worth noticing that the normalization $\frac{1}{N} \operatorname{tr} \mathbf{\Sigma}_N = 1$ is considered for ease of exposition and is not limiting since the RTE is invariant to any scaling of $\mathbf{\Sigma}_N$. Denote by $\hat{\mathbf{S}}_N(\rho)$ the matrix given by:

$$\hat{\mathbf{S}}_N(\rho) = \frac{1}{\gamma_N(\rho)} \frac{1-\rho}{1-(1-\rho)c_N} \frac{1}{n} \sum_{i=1}^{n} \mathbf{w}_i \mathbf{w}_i^* + \rho I_N,$$

where $\gamma_N(\rho)$ is the unique positive solution to:

$$1 = \frac{1}{N} \operatorname{tr} \mathbf{\Sigma}_N \left(\rho \gamma_N(\rho) + (1-\rho) \mathbf{\Sigma}_N \right)^{-1}$$

then $\hat{\mathbf{S}}_N(\rho)$ is equivalent to the RTE $\hat{\mathbf{C}}_N(\rho)$ in the sense of the following theorem,

Theorem 1. *For any $\kappa > 0$ small, define $\mathcal{R}_\kappa \triangleq \left[\kappa + \max(0, 1-c^{-1}), 1\right]$. Then, as $N, n \to \infty$ with $\frac{N}{n} \to c \in (0, \infty)$ and assuming $\limsup \|\mathbf{\Sigma}_N\| < \infty$, we have:*

$$\sup_{\rho \in \mathcal{R}_\kappa} \left\| \hat{\mathbf{C}}_N(\rho) - \hat{\mathbf{S}}_N \right\| \xrightarrow{\text{a.s.}} 0.$$

B. Convergence of the RTE in the large-n regime

In this section, we will consider the regime wherein N is fixed and n tends to infinity. An illustrative tool that is frequently used to handle this regime is the strong law of large numbers (SLLN) which suggests replacing the average of independent and identically distributed random variables by their expected value. This result should particularly serve to treat the term

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\mathbf{x}_i \mathbf{x}_i^*}{\mathbf{x}_i^* \hat{\mathbf{C}}_N^{-1}(\rho) \mathbf{x}_i}$$

in the expression of the RTE. Nevertheless, because of the dependence of $\hat{\mathbf{C}}_N(\rho)$ on the observations $\mathbf{x}_i$, the SLLN cannot be directly applied to handle the above quantity. As we expect $\hat{\mathbf{C}}_N(\rho)$ to converge to some deterministic matrix, say $\mathbf{\Sigma}_0(\rho)$, it is sensible to substitute $\frac{1}{n} \sum_{i=1}^{n} \frac{\mathbf{x}_i \mathbf{x}_i^*}{\mathbf{x}_i^* \hat{\mathbf{C}}_N^{-1}(\rho) \mathbf{x}_i}$ by $\frac{1}{n} \sum_{i=1}^{n} \frac{\mathbf{x}_i \mathbf{x}_i^*}{\mathbf{x}_i^* \mathbf{\Sigma}_0^{-1}(\rho) \mathbf{x}_i}$. The latter quantity is in turn equivalent to $\mathbb{E}\left[\frac{\mathbf{x}\mathbf{x}^*}{\mathbf{x}^* \mathbf{\Sigma}_0^{-1}(\rho) \mathbf{x}} \right]$ from the SLLN where the expectation is taken over the distribution of the random vectors $\mathbf{x}_i$. Based on these heuristic arguments, a plausible guess is that $\hat{\mathbf{C}}_N(\rho)$ converges to $\mathbf{\Sigma}_0(\rho)$, the solution to the following equation:

$$\mathbf{\Sigma}_0(\rho) = N(1-\rho) \mathbb{E}\left[\frac{\mathbf{x}\mathbf{x}^*}{\mathbf{x}^* \mathbf{\Sigma}_0^{-1}(\rho) \mathbf{x}} \right] + \rho \mathbf{I}_N. \tag{2}$$

The main goal of this section is to establish the convergence of $\hat{\mathbf{C}}_N(\rho)$ to $\mathbf{\Sigma}_0(\rho)$. We will assume that $\mathbf{\Sigma}_0(\rho)$ exists for each $\rho \in (0, 1]$. The existence and uniqueness of $\mathbf{\Sigma}_0(\rho)$ will be discussed later on in this section. Similar to the large-n, N regime, we need to introduce a random equivalent for $\hat{\mathbf{C}}_N(\rho)$ that is easier to handle. Naturally, an intuitive random equivalent is obtained by replacing, in the right-hand side of (1), $\hat{\mathbf{C}}_N(\rho)$ by $\mathbf{\Sigma}_0(\rho)$, thus yielding:

$$\tilde{\mathbf{\Sigma}}(\rho) = N(1-\rho) \frac{1}{n} \sum_{i=1}^{n} \frac{\mathbf{x}_i \mathbf{x}_i^*}{\mathbf{x}_i^* \mathbf{\Sigma}_0^{-1}(\rho) \mathbf{x}_i} + \rho \mathbf{I}_N. \tag{3}$$

Unlike $\hat{\mathbf{C}}_N(\rho)$, $\tilde{\mathbf{\Sigma}}(\rho)$ is more tractable, being an explicit function of the observations' vectors. By the SLLN, $\tilde{\mathbf{\Sigma}}(\rho)$ is an unbiased estimate of $\mathbf{\Sigma}_0(\rho)$ that satisfies:

$$\mathbf{\Sigma}_0(\rho) = \tilde{\mathbf{\Sigma}}(\rho) + \boldsymbol{\epsilon}_n(\rho),$$

where $\boldsymbol{\epsilon}_n(\rho)$ is an $N \times N$ matrix whose elements converge almost surely to zero and are bounded in probability at the rate $\frac{1}{n}$, i.e,

$$\left[\boldsymbol{\epsilon}_n(\rho)\right]_{i,j} = \mathcal{O}_p\left(\frac{1}{n}\right).$$

For the above convergence to hold uniformly in ρ, one needs to check that the first absolute second moment of the entries of $\frac{\mathbf{x}\mathbf{x}^*}{\mathbf{x}^* \mathbf{\Sigma}_0^{-1}(\rho) \mathbf{x}}$ is uniformly bounded in ρ. To this end we shall additionally assume that:

Assumption A-2.
Matrix $\mathbf{\Sigma}_N$ is non-singular, i.e., the smallest eigenvalue of $\mathbf{\Sigma}_N$, $\lambda_{\min}(\mathbf{\Sigma}_N)$ satisfies:

$$\lambda_{\min}(\mathbf{\Sigma}_N) > 0.$$

Under Assumption 2, the spectral norm of $\mathbf{\Sigma}_0(\rho)$ can be bounded as:

Lemma 2. *Let $\mathbf{\Sigma}_0$ be the solution to* (2)*, whenever it exists. Then,*

$$\sup_{\rho \in [\kappa, 1]} \|\mathbf{\Sigma}_0(\rho)\| \leq \frac{\|\mathbf{\Sigma}_N\|}{\lambda_{\min}(\mathbf{\Sigma}_N)}$$

where $\kappa > 0$ is some positive scalar.

Proof: See Appendix A ∎

Equipped with the bound provided by Lemma 2, we can claim that:

$$\sup_{\rho \in [\kappa, 1]} \left| \left[\boldsymbol{\epsilon}_n(\rho)\right]_{i,j} \right| = \mathcal{O}_p\left(\frac{1}{n}\right)$$

or equivalently:

$$\sup_{\rho \in [\kappa, 1]} \left\| \tilde{\mathbf{\Sigma}}(\rho) - \mathbf{\Sigma}_0(\rho) \right\| = \mathcal{O}_p\left(\frac{1}{n}\right).$$

Characterizing the rate of convergence of $\tilde{\mathbf{\Sigma}}(\rho)$ to $\mathbf{\Sigma}_0(\rho)$ is of fundamental importance and would later help in the derivation of the second-order statistics for $\tilde{\mathbf{\Sigma}}(\rho)$ and then for $\hat{\mathbf{C}}_N(\rho)$.

Before stating our first main result, we would like to particularly stress the fact that Assumption 2 is not limiting. To see that, consider $\mathbf{\Sigma}_N = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^*$ the eigenvalue decomposition of $\mathbf{\Sigma}_N$ wherein the diagonal elements of $\mathbf{\Lambda}$, $\lambda_1, \cdots, \lambda_N$ correspond to the eigenvalues of $\mathbf{\Sigma}_N$ arranged in the decreasing order, i.e., $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_N$. Denoting by r the rank of $\mathbf{\Sigma}_N$, then, $\lambda_{r+1} = \cdots = \lambda_N = 0$. Write $\mathbf{U}$ as $\mathbf{U} = [\mathbf{U}_r, \mathbf{U}_{N-r}]$, $\mathbf{U}_r \in \mathbb{C}^{N\times r}$. Then, it is easy to see that:

$$\hat{\mathbf{C}}_N(\rho)\mathbf{U}_{N-r} = \rho\mathbf{U}_{N-r}$$

while:

$$\mathbf{U}_r^*\hat{\mathbf{C}}_N(\rho)\mathbf{U}_r = (1-\rho)\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{\Lambda_r}^{\frac{1}{2}}\tilde{\mathbf{w}}_i\tilde{\mathbf{w}}_i^*\mathbf{\Lambda}_r^{\frac{1}{2}}}{\frac{1}{N}\tilde{\mathbf{w}}_i^*\mathbf{\Lambda}_r^{\frac{1}{2}}\mathbf{U}_r^*\hat{\mathbf{C}}_N^{-1}(\rho)\mathbf{U}_r\mathbf{\Lambda}_r^{\frac{1}{2}}\tilde{\mathbf{w}}_i} + \rho\mathbf{I}_N, \tag{4}$$

where $\tilde{\mathbf{w}}_i = \mathbf{U}_r^*\mathbf{x}_i \in \mathbb{C}^r$ is a standard Gaussian vector. Since $\left(\mathbf{U}_r^*\hat{\mathbf{C}}_N(\rho)\mathbf{U}_r\right)^{-1} = \mathbf{U}_r^*\hat{\mathbf{C}}_N^{-1}(\rho)\mathbf{U}_r$, instead of using $\hat{\mathbf{C}}_N(\rho)$, it thus suffices to work with $\mathbf{U}_r^*\hat{\mathbf{C}}_N(\rho)\mathbf{U}_r$, for which Assumption 2 can be used.

The following theorem establishes the convergence of $\hat{\mathbf{C}}_N(\rho)$ to $\mathbf{\Sigma}_0(\rho)$, the hypothetical solution to (2),

Theorem 3. *Assume that there exists a unique solution $\mathbf{\Sigma}_0(\rho)$ to (2). Let $\kappa > 0$ be some small positive real scalar. Then, assuming that Assumptions 1 and 2 hold true, one has under the large-n regime:*

$$\sup_{\rho\in[\kappa,1]}\left\|\hat{\mathbf{C}}_N(\rho) - \mathbf{\Sigma}_0(\rho)\right\| \xrightarrow{\text{a.s.}} 0.$$

Moreover,

$$\sup_{\rho\in[\kappa,1]}\left\|\hat{\mathbf{C}}_N(\rho) - \mathbf{\Sigma}_0(\rho)\right\| = \mathcal{O}_p\left(\frac{1}{n}\right).$$

Proof: See Appendix B ∎

In Theorem 3, we establish the convergence of $\hat{\mathbf{C}}_N(\rho)$ to some limiting matrix $\mathbf{\Sigma}_0(\rho)$ that solves the fixed point equation (2). While (2) seems to fully characterize $\mathbf{\Sigma}_0(\rho)$, it does not clearly unveil its relationship with the observations' covariance matrix $\mathbf{\Sigma}_N$. The major intricacy stems from the expectation operator in the term $\mathbb{E}\left[\frac{\mathbf{x}\mathbf{x}^*}{\mathbf{x}^*\mathbf{\Sigma}_0^{-1}(\rho)\mathbf{x}}\right]$. A close look to this quantity reveals that it can be further developed by leveraging some interesting features of Gaussian distributed vectors. Note first that (2) is also equivalent to:

$$N(1-\rho)\mathbb{E}\left[\frac{\mathbf{w}\mathbf{w}^*}{\mathbf{w}^*\mathbf{\Sigma}_N^{\frac{1}{2}}\mathbf{\Sigma}_0^{-1}(\rho)\mathbf{\Sigma}_N^{\frac{1}{2}}\mathbf{w}}\right] + \rho\mathbf{\Sigma}_N^{-1} = \mathbf{\Sigma}_N^{-\frac{1}{2}}\mathbf{\Sigma}_0(\rho)\mathbf{\Sigma}_N^{-\frac{1}{2}}, \tag{5}$$

where $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$. Let $\mathbf{\Sigma}_N^{\frac{1}{2}}\mathbf{\Sigma}_0^{-1}(\rho)\mathbf{\Sigma}_N^{\frac{1}{2}} = \mathbf{V}\mathbf{D}\mathbf{V}^*$ be an eigenvalue decomposition of $\mathbf{\Sigma}_N^{\frac{1}{2}}\mathbf{\Sigma}_0^{-1}(\rho)\mathbf{\Sigma}_N^{\frac{1}{2}}$, where $\mathbf{D}$ is a diagonal matrix with diagonal elements $d_1, d_2, \cdots, d_N$. Notice that, of course the d_i's depend on ρ. However, for simplicity purposes, the notation with (ρ) is omitted. Since the Gaussian distribution is invariant under unitary transformation, (5) is also equivalent to:

$$N(1-\rho)\mathbb{E}\left[\frac{\mathbf{w}\mathbf{w}^*}{\mathbf{w}^*\mathbf{D}\mathbf{w}}\right] + \rho\mathbf{V}^*\mathbf{\Sigma}_N^{-1}\mathbf{V} = \mathbf{D}^{-1}. \tag{6}$$

It is not difficult to see that the off-diagonal elements of $\mathbb{E}\left[\frac{\mathbf{w}\mathbf{w}^*}{\mathbf{w}^*\mathbf{D}\mathbf{w}}\right]$ are equal to zero. In effect for $i \neq j$, writing w_i as $r_i e^{j\theta_i}$ with r_i Rayleigh distributed and θ_i independent of r_i and uniformly distributed over $[-\pi, \pi]$, one has $\mathbb{E}\left[\left[\frac{\mathbf{w}\mathbf{w}^*}{\mathbf{w}^*\mathbf{D}\mathbf{w}}\right]_{i,j}\right] = \mathbb{E}\left[\frac{r_i r_j^* e^{j(\theta_i-\theta_j)}}{\sum_{i=1}^N d_i|r_i|^2}\right]$ which can be shown to be zero by taking the expectation over the difference of phase $\theta_i - \theta_j$. Therefore, $\mathbb{E}\left[\frac{\mathbf{w}\mathbf{w}^*}{\mathbf{w}^*\mathbf{D}\mathbf{w}}\right]$ is diagonal, with diagonal elements $(\alpha_i)_{i=1,\cdots,N}$ given by:

$$\alpha_i(\mathbf{D}) = \mathbb{E}\left[\frac{|w_i|^2}{\mathbf{w}^*\mathbf{D}\mathbf{w}}\right].$$

Hence, $\mathbf{V}^*\mathbf{\Sigma}_N^{-1}\mathbf{V}$ is also diagonal, thus implying that $\mathbf{\Sigma}_N$ and $\mathbf{\Sigma}_0(\rho)$ share the same eigenvector matrix $\mathbf{U}$. In order to prove the existence of $\mathbf{\Sigma}_0(\rho)$, it suffices to check that $d_1, \cdots, d_N$ are solutions to the following equation:

$$N(1-\rho)\alpha_i(\mathbf{D}) + \frac{\rho}{\lambda_i} = \frac{1}{d_i}. \tag{7}$$

To this end, consider

$$h : \mathbb{R}_+^N \to \mathbb{R}_+^N$$
$$(x_1, \cdots, x_N) \mapsto N(1-\rho)\left(\mathbb{E}\left[\frac{|w_1|^2}{\sum_{j=1}^N \frac{1}{x_j}|w_j|^2}\right] + \frac{\rho}{\lambda_1}, \cdots,\right.$$
$$\left.\mathbb{E}\left[\frac{|w_N|^2}{\sum_{j=1}^N \frac{1}{x_j}|w_j|^2}\right] + \frac{\rho}{\lambda_N}\right).$$

Proving that $d_1, \cdots, d_N$ are the unique solutions of (7) is equivalent to showing that:

$$\mathbf{x} = h(x_1, \cdots, x_N) \tag{8}$$

admits a unique positive solution. For this, we show that h satisfies the following properties:

- Nonnegativity: For each $x_1, \cdots, x_N \geq 0$, vector $h(x_1, \cdots, x_N)$has positive elements.
- Monotonicity: For each $x_1 \geq x_1^{'}, \cdots, x_N \geq x_N^{'}$, $h(x_1, \cdots, x_N) \geq h(x_1^{'}, \cdots, x_N^{'})$ where $\geq$ holds element-wise.
- Scalability: For each $\alpha > 1$, $\alpha h(x_1, \cdots, x_N) > h(\alpha x_1, \cdots, \alpha x_N)$.

The first item is trivial. The second one follows from the fact that h is an increasing function of each x_i. As for the last item, it follows by noticing that as $\rho > 0$,

$$\mathbb{E}\left[\frac{|w_i|^2}{\sum_{j=1}^N \frac{1}{\alpha x_j}|w_j|^2}\right] + \frac{\rho}{\lambda_j} < \alpha\left(\mathbb{E}\left[\frac{|w_i|^2}{\sum_{j=1}^N \frac{1}{x_j}|w_j|^2}\right] + \frac{\rho}{\lambda_j}\right)$$

According to [24], h is a standard interference function, and if there exists $q_1, \cdots, q_N$ such that $\mathbf{q} > h(q_1, \cdots, q_N)$ where $>$ holds element-wise, then there is a unique $\mathbf{x}_\infty = (x_{1,\infty}, \cdots, x_{N,\infty})$ such that:

$$\mathbf{x}_\infty = h(x_{1,\infty}, \cdots, x_{N,\infty}).$$

Moreover, $\mathbf{x}_\infty = \lim_{t\to\infty} \mathbf{x}^{(t)}$ with $\mathbf{x}^{(0)} > 0$ arbitrary and for $t \geq 0$, $\mathbf{x}^{(t+1)} = h(x_1^{(t)}, \cdots, x_N^{(t)})$. To prove the feasibility condition, take $\mathbf{q} = (q, \cdots, q)$. Then, $h(q, \cdots, q) = (1-\rho)q + \frac{\rho}{\lambda_i}$. Setting $q \geq \frac{1}{\lambda_{\min}}$, we get that $h(q, \cdots, q) < \mathbf{q}$, thereby establishing the desired inequality.

The interest of the framework of Yates [24] is that in addition to being a useful tool for proving existence and uniqueness of the fixed-point of a standard interference function, it shows that the solution can be numerically approximated by computing iteratively $\mathbf{x}^{(t+1)} = h(x_1^t, \cdots, x_N^t)$. However, in order to implement this algorithm, one needs to further develop the terms $\alpha_i(\mathbf{D})$. This is in particular the goal of the following lemma, the proof of which is deferred to Appendix C.

Lemma 4. *Let $\mathbf{w} = [w_1, \cdots, w_N]^T$ be a standard complex Gaussian vector and $\mathbf{D} = (d_1, \cdots, d_N)$ be a diagonal matrix with positive diagonal elements. Consider $\alpha_1, \cdots, \alpha_N$, the set of scalars given by:*

$$\alpha_i(\mathbf{D}) = \mathbb{E}\left[\frac{|w_i|^2}{\sum_{i=1}^N d_i |w_i|^2}\right].$$

Then

$$\alpha_i(\mathbf{D}) = \frac{1}{2^N N} \frac{1}{d_i \prod_{j=1}^N d_j} \times F_D^{(N)}\left(N, 1, \cdots, \underset{\substack{\uparrow \\ i\text{-th} \\ \text{position}}}{2}, 1, \cdots, 1, N+1, \frac{d_1 - \frac{1}{2}}{d_1}, \cdots, \frac{d_N - \frac{1}{2}}{d_N}\right),$$

where $F_D^{(N)}$ is the Lauricella's type D hypergeometric function. [2]

Equipped with the result of Lemma 4, we will now show how one can in practice approximate $\mathbf{\Sigma}_0(\rho)$. First, one needs to approximate the solution of (8). Let $\mathbf{d}^0 = \left[d_1^{(0)}, \cdots, d_N^{(0)}\right]^{\mathrm{T}}$ be an arbitrary vector with positive elements. We set $\mathbf{d}^{(t)} = \left[d_1^{(t)}, \cdots, d_N^{(t)}\right]$ as:

$$d_i^{(t+1)} = \frac{1}{\frac{\rho}{\lambda_i} + N(1-\rho)\alpha_i(\mathrm{diag}(\mathbf{d}^{(t)}))}$$

where the expression of $\alpha_i(\mathrm{diag}(\mathbf{d}^{(t)}))$ is given by Lemma 4. As $t \to \infty$, $\mathbf{d}^{(t)}$ tends to $\mathbf{d}$, the vector of eigenvalues of $\mathbf{\Sigma}_N^{\frac{1}{2}} \mathbf{\Sigma}_0^{-1}(\rho) \mathbf{\Sigma}_N^{\frac{1}{2}}$ which is the solution of (8). Since $\mathbf{\Sigma}_N$ and $\mathbf{\Sigma}_0(\rho)$ share the same eigenvectors, the eigenvalues $s_{1,\infty}, \cdots, s_{N,\infty}$ of $\mathbf{\Sigma}_0(\rho)$ are given by $s_{i,\infty} = \frac{\lambda_i}{d_i}$. The matrix $\mathbf{\Sigma}_0(\rho)$ is finally given by:

$$\mathbf{\Sigma}_0(\rho) = \mathbf{U}\,\mathrm{diag}([s_{1,\infty}, \cdots, s_{N,\infty}])\mathbf{U}^*.$$

While the above characterization of $\mathbf{\Sigma}_0(\rho)$ seems to provide few insights in most cases, it shows that except for the particular case of $\mathbf{\Sigma}_N = \mathbf{I}_N$, the RTE $\hat{\mathbf{C}}_N(\rho)$ is biased for $\rho \in [\kappa, 1)$ in that:

$$\mathbf{\Sigma}_0(\rho) \neq \mathbf{\Sigma}_N.$$

To see that, notice that $\mathbf{\Sigma}_0(\rho) = \mathbf{\Sigma}_N$ implies that $\mathbf{D} = \mathbf{I}_N$. Replacing $\mathbf{D}$ by the identity matrix in (5) and using the fact that $\mathbb{E}\left[\frac{\mathbf{w}\mathbf{w}^*}{\mathbf{w}^*\mathbf{w}}\right] = \frac{1}{N}\mathbf{I}_N$ shows that only $\mathbf{\Sigma}_N = \mathbf{I}_N$ satisfies a null bias. Hence, it appears that improving the conditioning of the RTE by using a non-zero regularization coefficient comes in general at the cost of a higher bias.

III. Second order statistics in the large-n regime

The previous section establishes the convergence of the RTE to the limiting deterministic matrix $\mathbf{\Sigma}_0(\rho)$. In the following, for readability purposes, $\mathbf{\Sigma}_0(\rho)$ will be simply replaced by $\mathbf{\Sigma}_0$. The convergence holds in the almost sure sense, and can help infer the asymptotic limit of any functional of the RTE. More formally, for any functional f continuous around $\mathbf{\Sigma}_0$, $f(\hat{\mathbf{C}}_N)$ converges almost surely to $f(\mathbf{\Sigma}_0)$. While this result can be used to understand the convergence of inference methods using RTEs, it becomes of little help when one is required to deeply understand their fluctuations, a prerequisite that essentially arises in many detection applications. This motivates the present section which aims at establishing a Central Limit Theorem (CLT) for the RTE.

It is worth noticing that the scope of applicability of the results obtained in the large-n regime is much wider than that of the n, N large regime. As a matter of fact, using the Delta Method [25], our result can help obtain the CLT for any continuous functional of the RTE. We deeply believe that this can facilitate the design of inference methods using RTEs.

Although treatments of both regimes seem to take different directions, they have thus far presented the common denominator of relying on an intermediate random equivalent for $\hat{\mathbf{C}}_N(\rho)$, be it $\tilde{\mathbf{\Sigma}}(\rho)$ or $\hat{\mathbf{S}}_N(\rho)$ (See Theorem 1). It is thus easy to convince oneself that in order to derive the CLT for $\hat{\mathbf{C}}_N(\rho)$, a CLT for $\tilde{\mathbf{\Sigma}}(\rho)$ is required.

We denote in the sequel by $\boldsymbol{\delta}$ and $\tilde{\boldsymbol{\delta}}$ the quantities: $\boldsymbol{\delta} = \mathrm{vec}(\hat{\mathbf{C}}_N(\rho)) - \mathrm{vec}(\mathbf{\Sigma}_0)$ and $\tilde{\boldsymbol{\delta}} = \mathrm{vec}(\tilde{\mathbf{\Sigma}}(\rho)) - \mathrm{vec}(\mathbf{\Sigma}_0)$ and consider the derivation of the CLT for vectors $\boldsymbol{\delta}$ and then for $\tilde{\boldsymbol{\delta}}$. We will particularly prove that $\boldsymbol{\delta}$ and $\tilde{\boldsymbol{\delta}}$ behave in the large-n regime as Gaussian random variables that can be fully characterized by their covariance matrices $\mathbb{E}[\boldsymbol{\delta}\boldsymbol{\delta}^*]$ and $\mathbb{E}[\tilde{\boldsymbol{\delta}}\tilde{\boldsymbol{\delta}}^*]$. Starting with the observation that in many signal processing applications, the focus might be put on the second-order statistics of the real and imaginary parts of $\boldsymbol{\delta}$ and $\tilde{\boldsymbol{\delta}}$, we additionally provide expressions for the pseudo-covariance matrices $\mathbb{E}[\boldsymbol{\delta}\boldsymbol{\delta}^{\mathrm{T}}]$ and $\mathbb{E}[\tilde{\boldsymbol{\delta}}\tilde{\boldsymbol{\delta}}^{\mathrm{T}}]$ of $\boldsymbol{\delta}$ and $\tilde{\boldsymbol{\delta}}$ which, coupled with that of covariance matrices, suffice to fully characterize fluctuations of the vectors $[\Re\boldsymbol{\delta}^{\mathrm{T}}, \Im\boldsymbol{\delta}^{\mathrm{T}}]^{\mathrm{T}}$ and $[\Re\tilde{\boldsymbol{\delta}}^{\mathrm{T}}, \Im\tilde{\boldsymbol{\delta}}^{\mathrm{T}}]^{\mathrm{T}}$.

We will start by handling the fluctuations of $\tilde{\boldsymbol{\delta}}$. To this end, we need first to work out the expression of $\tilde{\mathbf{\Sigma}}(\rho)$. Recall that $\tilde{\mathbf{\Sigma}}(\rho)$ is given by:

$$\tilde{\mathbf{\Sigma}}(\rho) = \frac{N(1-\rho)}{n} \sum_{i=1}^n \frac{\mathbf{x}_i \mathbf{x}_i^*}{\mathbf{x}_i^* \mathbf{\Sigma}_0^{-1} \mathbf{x}_i} + \rho \mathbf{I}_N.$$

[2] The evaluation of the Lauricella's type D hypergeometric function is performed numerically using its integral representation

$$F_D^{(N)}(a, b_1, \cdots, b_n, c; x_1, \cdots, x_n) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(a-c)} \int_0^1 t^{a-1}(1-t)^{c-a-1} \prod_{i=1}^n (1-x_i t)^{-b_i} dt.\ \Re c > \Re a > 0.$$

Therefore,

$$\boldsymbol{\Sigma}_0^{-\frac{1}{2}}\tilde{\boldsymbol{\Sigma}}(\rho)\boldsymbol{\Sigma}_0^{-\frac{1}{2}}-\mathbf{I}_N=\frac{N(1-\rho)}{n}\sum_{i=1}^{n}\frac{\boldsymbol{\Sigma}_0^{-\frac{1}{2}}\boldsymbol{\Sigma}_N^{\frac{1}{2}}\mathbf{w}_i\mathbf{w}_i^*\boldsymbol{\Sigma}_N^{\frac{1}{2}}\boldsymbol{\Sigma}_0^{-\frac{1}{2}}}{\mathbf{w}_i^*\boldsymbol{\Sigma}_N^{\frac{1}{2}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_N^{\frac{1}{2}}\mathbf{w}_i}+\rho\boldsymbol{\Sigma}_0^{-1}-\mathbf{I}_N$$

Using the eigenvalue decomposition of $\boldsymbol{\Sigma}_N^{\frac{1}{2}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_N^{\frac{1}{2}}=\mathbf{U}\mathbf{D}\mathbf{U}^*$ and denoting $\tilde{\mathbf{w}}_i=\mathbf{U}^*\mathbf{w}_i$, we thus obtain:

$$\mathbf{U}^*\boldsymbol{\Sigma}_0^{-\frac{1}{2}}\tilde{\boldsymbol{\Sigma}}(\rho)\boldsymbol{\Sigma}_0^{-\frac{1}{2}}\mathbf{U}-\mathbf{I}_N=\frac{N(1-\rho)}{n}\sum_{i=1}^{n}\frac{\mathbf{D}^{\frac{1}{2}}\tilde{\mathbf{w}}_i\tilde{\mathbf{w}}_i^*\mathbf{D}^{\frac{1}{2}}}{\tilde{\mathbf{w}}_i^*\mathbf{D}\tilde{\mathbf{w}}_i}+\rho\mathbf{U}^*\boldsymbol{\Sigma}_0^{-1}\mathbf{U}-\mathbf{I}_N.$$

From the characterization of $\boldsymbol{\Sigma}_0$ provided in the previous section, we can easily check that:

$$N(1-\rho)\mathbb{E}\left[\frac{\mathbf{D}^{\frac{1}{2}}\tilde{\mathbf{w}}\tilde{\mathbf{w}}^*\mathbf{D}^{\frac{1}{2}}}{\tilde{\mathbf{w}}^*\mathbf{D}\tilde{\mathbf{w}}}\right]=\mathbf{I}_N-\rho\mathbf{U}^*\boldsymbol{\Sigma}_0^{-1}\mathbf{U}$$

Therefore,

$$\mathbf{U}^*\boldsymbol{\Sigma}_0^{-\frac{1}{2}}\tilde{\boldsymbol{\Sigma}}(\rho)\boldsymbol{\Sigma}_0^{-\frac{1}{2}}\mathbf{U}-\mathbf{I}_N \tag{9}$$

$$=\frac{N(1-\rho)}{n}\sum_{i=1}^{n}\left[\frac{\mathbf{D}^{\frac{1}{2}}\tilde{\mathbf{w}}_i\tilde{\mathbf{w}}_i^*\mathbf{D}^{\frac{1}{2}}}{\tilde{\mathbf{w}}_i^*\mathbf{D}\tilde{\mathbf{w}}_i}-\mathbb{E}\left[\frac{\mathbf{D}^{\frac{1}{2}}\tilde{\mathbf{w}}\tilde{\mathbf{w}}^*\mathbf{D}^{\frac{1}{2}}}{\tilde{\mathbf{w}}^*\mathbf{D}\tilde{\mathbf{w}}}\right]\right]. \tag{10}$$

From (10), it appears that the asymptotic distribution of $[\Re\tilde{\boldsymbol{\delta}}^{\mathrm{T}},\Im\tilde{\boldsymbol{\delta}}^{\mathrm{T}}]^{\mathrm{T}}$ is Gaussian and thus can be fully characterized by its asymptotic covariance and pseudo-covariance matrices. Using (10), it is easy to show that we need for that the pseudo-covariance and covariance matrices of:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{vec}(\tilde{\mathbf{w}}_i\tilde{\mathbf{w}}_i^*)}{\tilde{\mathbf{w}}_i^*\mathbf{D}\tilde{\mathbf{w}}_i}-\mathbb{E}\left[\frac{\mathrm{vec}(\tilde{\mathbf{w}}\tilde{\mathbf{w}}^*)}{\tilde{\mathbf{w}}^*\mathbf{D}\tilde{\mathbf{w}}}\right].$$

These quantities involve the following set of scalars,

$$\beta_{i,j}=\mathbb{E}\left[\frac{|w_i|^2|w_j|^2}{\left(\mathbf{w}^*\mathbf{D}\mathbf{w}\right)^2}\right]\quad i,j=1,\cdots,N$$

for which closed-form expressions need to be derived. This is the objective of the following technical lemma, which is of independent interest:

Lemma 5. *Let $\mathbf{w}=[w_1,\cdots,w_N]^T$ be a standard complex Gaussian vector and $\mathbf{D}=\mathrm{diag}(d_1,\cdots,d_N)$ be a diagonal matrix with positive diagonal elements. Consider $\beta_{i,j}$ as above. Then $\beta_{i,j}$ are given for $i=j$ and $i\neq j$ by the expressions in* (11)*,* (12) *and* (13) *at the top of next page.*

With this result at hand, the next Lemma follows immediately:

Lemma 6. *Let $\mathbf{D}$ be $N\times N$ diagonal matrix with positive diagonal elements. Consider $\tilde{\mathbf{w}}_1,\cdots,\tilde{\mathbf{w}}_n$ n independent complex Gaussian random vectors with zero-mean and covariance $\mathbf{I}_N$. Then, $\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{vec}(\tilde{\mathbf{w}}_i\tilde{\mathbf{w}}_i^*)}{\tilde{\mathbf{w}}_i^*\mathbf{D}\tilde{\mathbf{w}}_i}-\mathbb{E}\left[\frac{\mathrm{vec}(\tilde{\mathbf{w}}\tilde{\mathbf{w}}^*)}{\tilde{\mathbf{w}}^*\mathbf{D}\tilde{\mathbf{w}}}\right]\right)$ converges to a multivariate Gaussian distribution with covariance $\mathbf{B}(\mathbf{D})$ and pseudo-covariance $\mathbf{G}(\mathbf{D})$ given by:*

$$\mathbf{B}(\mathbf{D})=\tilde{\mathbf{B}}(\mathbf{D})-\mathrm{vec}(\boldsymbol{\Xi})\mathrm{vec}(\boldsymbol{\Xi})^T \tag{14}$$

$$\mathbf{G}(\mathbf{D})=\tilde{\mathbf{G}}(\mathbf{D})-\mathrm{vec}(\boldsymbol{\Xi})\mathrm{vec}(\boldsymbol{\Xi})^T \tag{15}$$

where

$$\tilde{\mathbf{B}}(\mathbf{D})=\mathbb{E}\left[\frac{\mathrm{vec}(\tilde{\mathbf{w}}\tilde{\mathbf{w}}^*)\left(\mathrm{vec}(\tilde{\mathbf{w}}\tilde{\mathbf{w}}^*)\right)^*}{(\tilde{\mathbf{w}}^*\mathbf{D}\tilde{\mathbf{w}})^2}\right]$$

$$\tilde{\mathbf{G}}(\mathbf{D})=\mathbb{E}\left[\frac{\mathrm{vec}(\tilde{\mathbf{w}}\tilde{\mathbf{w}}^*)\left(\mathrm{vec}(\tilde{\mathbf{w}}\tilde{\mathbf{w}}^*)\right)^T}{(\tilde{\mathbf{w}}^*\mathbf{D}\tilde{\mathbf{w}})^2}\right]$$

$$\boldsymbol{\Xi}(\mathbf{D})=\mathrm{diag}\left(\alpha_1(\mathbf{D}),\cdots,\alpha_N(\mathbf{D})\right)$$

Furthermore, $\tilde{\mathbf{B}}$ and $\tilde{\mathbf{G}}$ are composed of N^2 block of $N\times N$ matrices, i.e, $\tilde{\mathbf{B}}(\mathbf{D})=\begin{bmatrix}\tilde{\mathbf{B}}_{1,1}&\cdots&\tilde{\mathbf{B}}_{1,N}\\&\ddots&\\\tilde{\mathbf{B}}_{N,1}&\cdots&\tilde{\mathbf{B}}_{N,N}\end{bmatrix}$, $\tilde{\mathbf{G}}(\mathbf{D})=\begin{bmatrix}\tilde{\mathbf{G}}_{1,1}&\cdots&\tilde{\mathbf{G}}_{1,N}\\&\ddots&\\\tilde{\mathbf{G}}_{N,1}&\cdots&\tilde{\mathbf{G}}_{N,N}\end{bmatrix}$ where:

$$\tilde{\mathbf{B}}_{i,i}=\mathrm{diag}\left(\beta_{i,1}\cdots,\beta_{i,N}\right)$$

$$\left[\tilde{\mathbf{B}}_{i,j}\right]_{k,\ell}=1_{\{k=i,\ell=j\}}\beta_{i,j},\ i\neq j$$

$$\left[\tilde{\mathbf{G}}_{i,i}\right]_{k,\ell}=1_{\{k=i,\ell=j\}}\beta_{i,i}$$

$$\left[\tilde{\mathbf{G}}_{i,j}\right]_{k,\ell}=1_{\{k=i,\ell=j\}}\beta_{i,j}+1_{\{k=j,\ell=i\}}\beta_{i,j},\ i\neq j.$$

As we shall see next, it can be useful to rewrite matrices $\mathbf{G}$ and $\mathbf{B}$ using the canonical basis of the space of $N\times N$ matrices. Denote by $\mathbf{E}_{i,j}$ the $N\times N$ matrix of all zeros except for 1 at position (i,j). Then, matrices $\mathbf{B}$ and $\mathbf{G}$ write:

$$\mathbf{B}=\sum_{\substack{i,j=1\\i\neq j}}^{N}\left(\beta_{i,j}-\alpha_i\alpha_j\right)\left(\mathbf{E}_{i,j}\otimes\mathbf{E}_{i,j}\right)-\sum_{i=1}^{N}\alpha_i^2\left(\mathbf{E}_{i,i}\otimes\mathbf{E}_{i,i}\right)$$
$$+\sum_{i=1}^{N}\sum_{j=1}^{N}\beta_{i,j}\left(\mathbf{E}_{i,i}\otimes\mathbf{E}_{j,j}\right)$$

$$\mathbf{G}=\sum_{\substack{i,j=1\\i\neq j}}^{N}\left(\beta_{i,j}-\alpha_i\alpha_j\right)\left(\mathbf{E}_{i,j}\otimes\mathbf{E}_{i,j}\right)+\sum_{\substack{i,j=1\\i\neq j}}^{N}\beta_{i,j}\left(\mathbf{E}_{i,j}\otimes\mathbf{E}_{i,j}^{\mathrm{T}}\right)$$
$$+\sum_{i=1}^{N}(\beta_{i,i}-\alpha_i^2)\left(\mathbf{E}_{i,i}\otimes\mathbf{E}_{i,i}\right).$$

In particular, in the specific case of $\mathbf{D}=\mathbf{I}_N$, occuring if $\boldsymbol{\Sigma}_N=\mathbf{I}_N$ or $\rho=0$, matrices $\mathbf{B}$ and $\mathbf{G}$ become:

$$\mathbf{B}=\frac{1}{N(N+1)}\mathbf{I}_{N^2}-\frac{1}{N^2(N+1)}\mathrm{vec}(\mathbf{I}_N)\left(\mathrm{vec}(\mathbf{I}_N)\right)^{\mathrm{T}}$$

$$\mathbf{G}=\frac{1}{N(N+1)}\mathbf{K}-\frac{1}{N^2(N+1)}\mathrm{vec}(\mathbf{I}_N)\left(\mathrm{vec}(\mathbf{I}_N)\right)^{\mathrm{T}}.$$

where $\mathbf{K}$ denotes the commutation matrix.

Equipped with Lemma 6, we are now in position to state the CLT for $\tilde{\boldsymbol{\Sigma}}(\rho)$, whose proof is omitted being a direct consequence of Lemma 6:

Theorem 7. *Let $\tilde{\boldsymbol{\Sigma}}(\rho)$ be given by* (3) *wherein observations $\mathbf{x}_1,\cdots,\mathbf{x}_n$ are drawn according to Assumption 1. Consider $\boldsymbol{\Sigma}_N=\mathbf{U}\boldsymbol{\Lambda}_N\mathbf{U}^*$ the eigenvalue decomposition of $\boldsymbol{\Sigma}_N$. Denote*

$$\beta_{i,i} = \frac{1}{2^{N-1}N(N+1)} \frac{1}{d_i^2 \prod_{k=1}^N d_k} F_D^N \left(N, 1 \cdots, 1, \underset{\substack{\uparrow \\ i\text{-th} \\ \text{position}}}{3}, 1, \cdots, 1, N+2, \frac{d_1 - \frac{1}{2}}{d_1}, \cdots, \frac{d_N - \frac{1}{2}}{d_N} \right) \tag{11}$$

$$\beta_{i,j} = \frac{1}{2^{N}N(N+1)} \frac{1}{d_i d_j \prod_{k=1}^N d_k} F_D^N \left(N, 1 \cdots, 1, \underset{\substack{\uparrow \\ i\text{-th} \\ \text{position}}}{2}, 1, \cdots, 1, \underset{\substack{\uparrow \\ j\text{-th} \\ \text{position}}}{2}, 1 \cdots, 1, N+2, \frac{d_1 - \frac{1}{2}}{d_1}, \cdots, \frac{d_N - \frac{1}{2}}{d_N} \right), i < j \tag{12}$$

$$\beta_{i,j} = \beta_{j,i},\ i > j \tag{13}$$

by $\mathbf{D}$ *the diagonal matrix whose diagonal elements are solutions to the system of equations* (7)*. Then, in the asymptotic large-*n *regime,* $\sqrt{n}\tilde{\boldsymbol{\delta}} = \sqrt{n}\left(\mathrm{vec}(\tilde{\boldsymbol{\Sigma}}(\rho)) - \mathrm{vec}(\boldsymbol{\Sigma}_0)\right)$ *behaves as a zero-mean Gaussian distributed vector with covariance:*

$$\tilde{\mathbf{M}}_1 = N^2(1-\rho)^2 \left(\overline{\mathbf{U}}\boldsymbol{\Lambda}_N^{\frac{1}{2}} \otimes \mathbf{U}\boldsymbol{\Lambda}_N^{\frac{1}{2}}\right) \mathbf{B}(\mathbf{D}) \left(\boldsymbol{\Lambda}_N^{\frac{1}{2}}\mathbf{U}^T \otimes \boldsymbol{\Lambda}_N^{\frac{1}{2}}\mathbf{U}^*\right)$$

and pseudo-covariance:

$$\tilde{\mathbf{M}}_2 = N^2(1-\rho)^2 \left(\overline{\mathbf{U}}\boldsymbol{\Lambda}_N^{\frac{1}{2}} \otimes \mathbf{U}\boldsymbol{\Lambda}_N^{\frac{1}{2}}\right) \mathbf{G}(\mathbf{D}) \left(\boldsymbol{\Lambda}_N^{\frac{1}{2}}\mathbf{U}^* \otimes \boldsymbol{\Lambda}_N^{\frac{1}{2}}\mathbf{U}^T\right).$$

where $\mathbf{B}(\mathbf{D})$ *and* $\mathbf{G}(\mathbf{D})$ *are given by* (14) *and* (15) *of Lemma 6.*

In the same way as above, we discuss the results of Theorem 7 when $\mathbf{D} = \mathbf{I}_N$. We can prove that in this case matrices $\tilde{\mathbf{M}}_1$ and $\tilde{\mathbf{M}}_2$ write as:

$$\tilde{\mathbf{M}}_1 = \frac{N(1-\rho)^2}{N+1} \left(\boldsymbol{\Sigma}_N^{\mathrm{T}} \otimes \overline{\boldsymbol{\Sigma}}_N\right) - \frac{(1-\rho)^2}{N+1} \mathrm{vec}(\boldsymbol{\Sigma}_N)\mathrm{vec}(\boldsymbol{\Sigma}_N)^{\mathrm{H}}$$

$$\tilde{\mathbf{M}}_2 = \frac{(1-\rho)^2}{N+1} \left[N\mathbf{K}\left(\boldsymbol{\Sigma}_N \otimes \boldsymbol{\Sigma}_N^{\mathrm{T}}\right) - \mathrm{vec}(\boldsymbol{\Sigma}_N)\mathrm{vec}(\boldsymbol{\Sigma}_N)^{\mathrm{T}}\right]$$

Now that the fluctuations of $\tilde{\boldsymbol{\Sigma}}(\rho)$ have been determined, we are in position to derive the asymptotic distribution of $\mathrm{vec}(\hat{\mathbf{C}}_N(\rho))$. The very recent results in [20] establishing equality between the fluctuations of the bilinear-forms of $\hat{\mathbf{C}}_N(\rho)$ and those of its random equivalent $\hat{\mathbf{S}}_N(\rho)$ in the large-n, N regime might lead us to expect similar results to hold in the large-n regime. As we will show in the following theorem, contrary to these first intuitions, the asymptotic distribution of $\mathrm{vec}(\tilde{\boldsymbol{\Sigma}}(\rho))$ is different from that of $\mathrm{vec}(\hat{\mathbf{C}}_N(\rho))$, even though it plays a central role in facilitating its analytical derivation.

Theorem 8. *Under the same setting of Theorem 7, define* $\tilde{\mathbf{F}}$ *the* $N^2 \times N^2$ *matrix:*

$$\tilde{\mathbf{F}} = N(1-\rho) \left(\overline{\mathbf{U}}\mathbf{D}^{\frac{1}{2}} \otimes \mathbf{U}\mathbf{D}^{\frac{1}{2}}\right) \tilde{\mathbf{B}}(\mathbf{D}) \left(\mathbf{D}^{\frac{1}{2}}\mathbf{U}^T \otimes \mathbf{D}^{\frac{1}{2}}\mathbf{U}^*\right)$$

with $\tilde{\mathbf{B}}(\mathbf{D})$ *defined in Lemma 6. Consider* $\hat{\mathbf{C}}_N(\rho)$ *the robust scatter estimator in* (1)*. Then, in the large-*n *asymptotic regime,* $\sqrt{n}\boldsymbol{\delta} = \sqrt{n}\left(\mathrm{vec}(\hat{\mathbf{C}}_N(\rho)) - \mathrm{vec}(\boldsymbol{\Sigma}_0)\right)$ *behaves as a zero-mean Gaussian-distributed vector with covariance:*

$$\mathbf{M}_1 = \left(\left(\boldsymbol{\Sigma}_0^{\frac{1}{2}}\right)^T \otimes \boldsymbol{\Sigma}_0^{\frac{1}{2}}\right) (\mathbf{I}_{N^2} - \tilde{\mathbf{F}})^{-1} \left(\left(\boldsymbol{\Sigma}_0^{-\frac{1}{2}}\right)^T \otimes \boldsymbol{\Sigma}_0^{-\frac{1}{2}}\right) \tilde{\mathbf{M}}_1 \times \left(\left(\boldsymbol{\Sigma}_0^{-\frac{1}{2}}\right)^T \otimes \boldsymbol{\Sigma}_0^{-\frac{1}{2}}\right) (\mathbf{I}_{N^2} - \tilde{\mathbf{F}})^{-1} \left(\left(\boldsymbol{\Sigma}_0^{\frac{1}{2}}\right)^T \otimes \boldsymbol{\Sigma}_0^{\frac{1}{2}}\right)$$

and pseudo-covariance:

$$\mathbf{M}_2 = \left(\left(\boldsymbol{\Sigma}_0^{\frac{1}{2}}\right)^T \otimes \boldsymbol{\Sigma}_0^{\frac{1}{2}}\right) (\mathbf{I}_{N^2} - \tilde{\mathbf{F}})^{-1} \left(\left(\boldsymbol{\Sigma}_0^{-\frac{1}{2}}\right)^T \otimes \boldsymbol{\Sigma}_0^{-\frac{1}{2}}\right) \tilde{\mathbf{M}}_2 \times \left(\boldsymbol{\Sigma}_0^{-\frac{1}{2}} \otimes \left(\boldsymbol{\Sigma}_0^{-\frac{1}{2}}\right)^T\right) (\mathbf{I}_{N^2} - \tilde{\mathbf{F}}^T)^{-1} \left(\boldsymbol{\Sigma}_0^{\frac{1}{2}} \otimes \left(\boldsymbol{\Sigma}_0^{\frac{1}{2}}\right)^T\right).$$

Proof: The proof is deferred to Appendix E ∎

If $\mathbf{D} = \mathbf{I}_N$, the expressions for M_1 and M_2 reduce to:

$$\mathbf{M}_1 = \frac{N(1-\rho)^2(N+1)}{(N+\rho)^2} \left[\left(\boldsymbol{\Sigma}_N^{\mathrm{T}} \otimes \boldsymbol{\Sigma}_N\right) - \frac{1}{N}\mathrm{vec}(\boldsymbol{\Sigma}_N)\mathrm{vec}(\boldsymbol{\Sigma}_N)^{\mathrm{H}}\right]$$

$$\mathbf{M}_2 = \frac{N(1-\rho)^2(N+1)}{(N+\rho)^2} \left[\left(\boldsymbol{\Sigma}_N^{\mathrm{T}} \otimes \boldsymbol{\Sigma}_N\right)\mathbf{K} - \frac{1}{N}\mathrm{vec}(\boldsymbol{\Sigma}_N)\mathrm{vec}(\boldsymbol{\Sigma}_N)^{\mathrm{T}}\right]$$

Note that the covariance and pseudo-covariance corresponding to the Tyler estimator derived in [23] can be retrieved by setting ρ to 0.

IV. Numerical results

In all our simulations, we consider the case where $\mathbf{x}_1, \cdots, \mathbf{x}_n$ are independent zero-mean Gaussian random vectors with covariance matrix $\boldsymbol{\Sigma}_N$ of Toeplitz form:

$$[\mathbf{C}_N]_{i,j} = \begin{cases} b^{j-i} & i \leq j \\ \left(b^{i-j}\right)^* & i > j \end{cases}, \qquad |b| \in (0,1), \tag{16}$$

A. Which regime is expected to be more accurate

In order to study the behavior of RTE, assumptions letting the number of observations and/or their sizes increase to infinity are essential for tractability. The behavior of RTE is studied under both concurrent asymptotic regimes, namely the large-n regime, which underlies all the derivations of this paper, and the n, N-large regime recently considered in [20]. Given that the scope of the results derived in the large-n, N regime, has thus far been limited to the handling of bilinear forms, practitioners might wonder to know whether, for their specific scenario, further investigation of this regime would produce more accurate results. In this first experiment, we attempt to answer to this open question by noticing that both regimes have the common denominator of producing random matrices that act as equivalents to the robust-scatter estimator. The accuracy of each regime is thus evaluated by measuring the closeness of the robust-scatter estimator to its

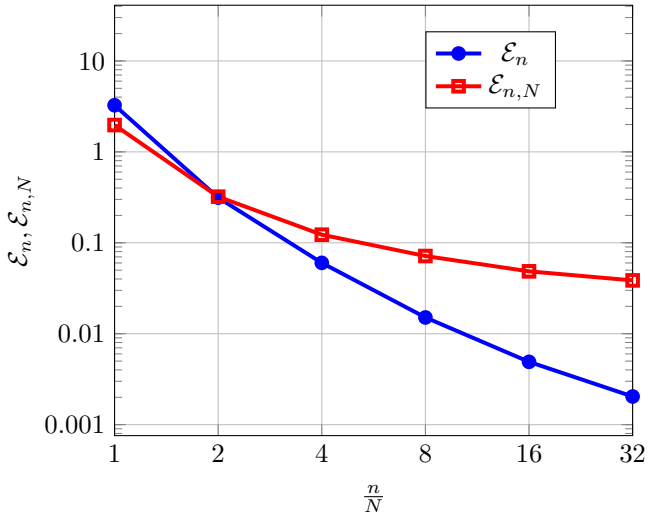

Fig. 1. Accuracy of the random equivalent when $N = 4$

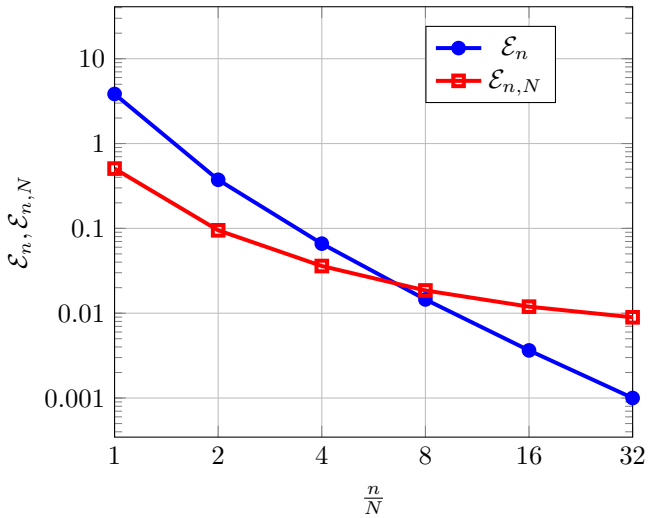

Fig. 2. Accuracy of the random equivalent when $N = 16$

random equivalent proposed by each regime. This closeness is measured using the following metrics:

$$\mathcal{E}_n \triangleq \frac{1}{N}\mathbb{E}\left\|\hat{\mathbf{C}}(\rho) - \tilde{\mathbf{\Sigma}}(\rho)\right\|_{\text{Fro}}^2$$

and

$$\mathcal{E}_{n,N} \triangleq \frac{1}{N}\mathbb{E}\left\|\hat{\mathbf{C}}(\rho) - \hat{\mathbf{S}}_N(\rho)\right\|_{\text{Fro}}^2.$$

Figures 1, 2 and 3 represent these metrics with respect to the ratio $\frac{n}{N}$ when $N = 4, 16, 32$, $b = 0.7$ and ρ set to 0.5. The region over which the use of the large-n regime is recommended corresponds to the values of $\frac{n}{N}$ for which the $\mathcal{E}_n$ curve is below the $\mathcal{E}_{n,N}$ one.

From these figures, it appears that, as N increases, the region over which results derived under the large-n regime are more accurate, corresponds to larger values of the ratio $\frac{n}{N}$.

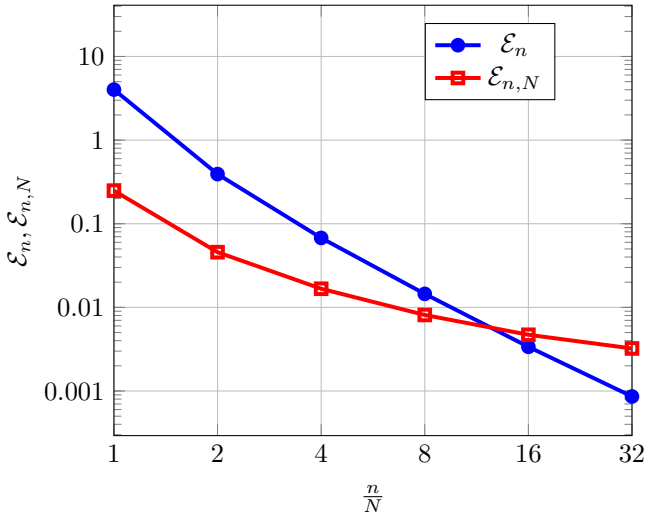

Fig. 3. Accuracy of the random equivalent when $N = 32$

B. Asymptotic bias

In this section, we assess the bias of the RTE with respect to the population covariance matrix. Since in many applications in radar detection, we only need to estimate the covariance matrix up to a scale factor, we define the bias as:

$$\text{Bias} = \frac{1}{N}\left\|\mathbb{E}\left[\frac{N}{\text{tr}\left(\mathbf{\Sigma}_N^{-1}\hat{\mathbf{C}}_N\right)}\mathbf{\Sigma}_N^{-1}\hat{\mathbf{C}}_N\right] - \mathbf{I}_N\right\|_{\text{Fro}}^2.$$

Since $\frac{N}{\text{tr}\left(\mathbf{\Sigma}_N^{-1}\hat{\mathbf{C}}_N\right)}\mathbf{\Sigma}_N^{-1}\hat{\mathbf{C}}_N$ has a bounded spectral norm, the dominated convergence theorem implies that:

$$\text{Bias} \xrightarrow[n\to+\infty]{} \frac{1}{N}\left\|\left[\frac{N}{\text{tr}\left(\mathbf{\Sigma}_N^{-1}\mathbf{\Sigma}_0\right)}\mathbf{\Sigma}_N^{-1}\mathbf{\Sigma}_0\right] - \mathbf{I}_N\right\|_{\text{Fro}}^2.$$

Figure 4 displays the asymptotic and empirical bias with respect to the Toeplitz coefficient b and for $\rho = 0.2, 0.5, 0.9$. We note that the bias is an increasing function of b. This is expected since for small values of b, the covariance matrix becomes close to the identity matrix. The RTE, viewed as a shrunk version of the Tyler to the identity matrix will thus produce small values of bias.

C. Central Limit Theorem

The central limit theorem provided in this paper can help determine fluctuations of any continuous functional of $\text{vec}(\hat{\mathbf{C}}_N)$. As an application, we consider in this section the quadratic form of type $\frac{1}{N}\mathbf{p}^*\hat{\mathbf{C}}_N^{-1}(\rho)\mathbf{p}$ with $\|\mathbf{p}\| = 1$ (used for instance for detection in array processing problems [26]), for which the large-n and the large-n, N regimes predict different kind of fluctuations. As a matter of fact, applying the Delta Method [25], one can easily prove that under the large-n,

$$T_n \triangleq \frac{\sqrt{n}\left(\frac{1}{N}\mathbf{p}^*\hat{\mathbf{C}}_N^{-1}(\rho)\mathbf{p} - \frac{1}{N}\mathbf{p}^*\mathbf{\Sigma}_0^{-1}(\rho)\mathbf{p}\right)}{\frac{1}{N}\sqrt{\mathbf{a}^{\text{T}}\mathbf{M}\mathbf{a}}}$$

$$\xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

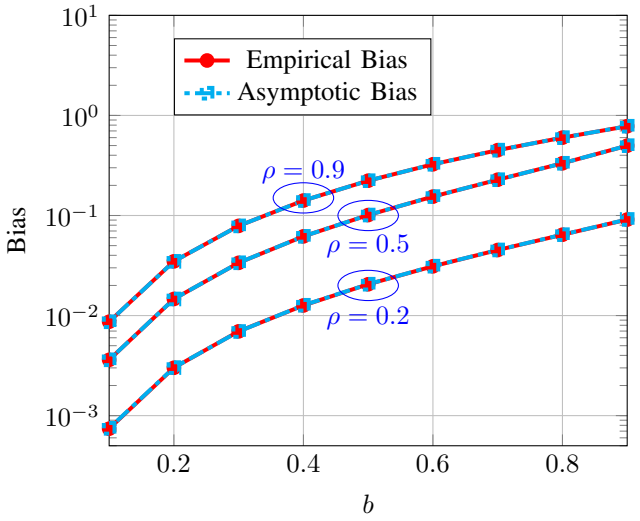

Fig. 4. Analysis of the bias for $n = 1000$ and $N = 2$ with respect to b and for different values of ρ.

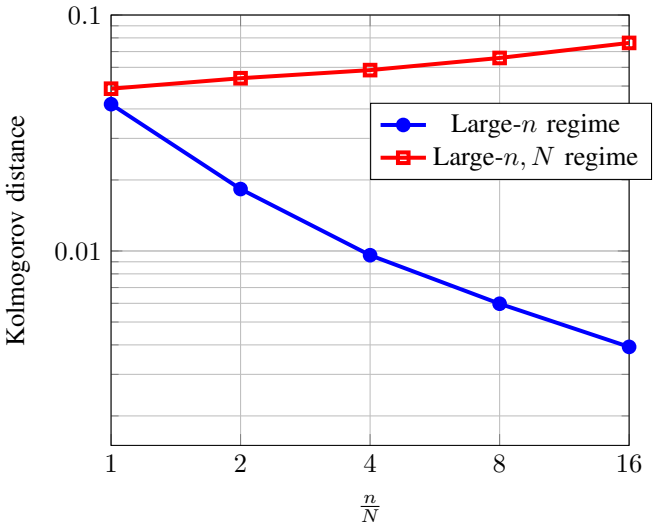

Fig. 5. Analysis of the accuracy of the CLT results for $N = 4$

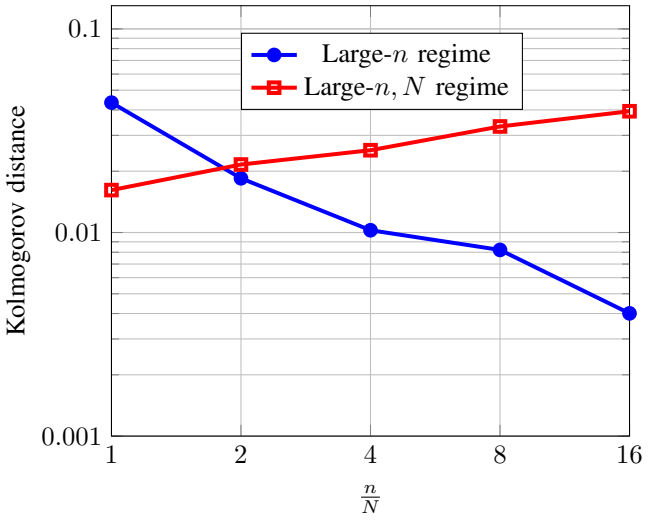

Fig. 6. Analysis of the accuracy of the CLT results for $N = 16$

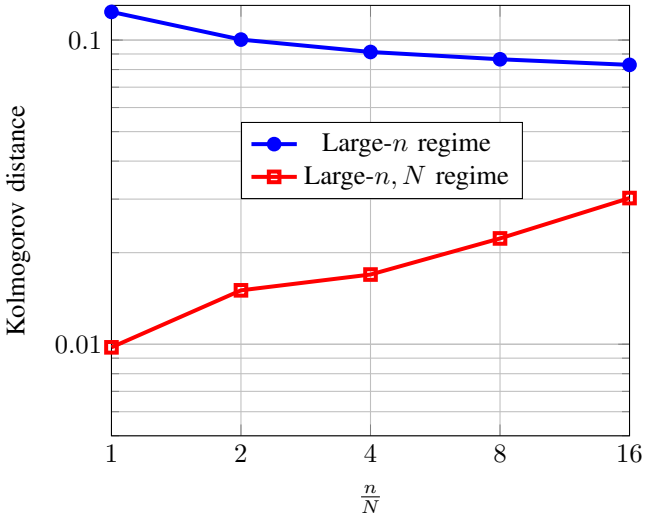

Fig. 7. Analysis of the accuracy of the CLT results for $N = 32$

where $\mathbf{a} = [\mathbf{a}_r^{\mathrm{T}} \quad \mathbf{a}_i^{\mathrm{T}}]^{\mathrm{T}}$ with

$$\mathbf{a}_r = \Re\left(\left(\left(\boldsymbol{\Sigma}_0^{-1}\right)\right)^{\mathrm{T}} \overline{\mathbf{p}} \otimes \boldsymbol{\Sigma}_0^{-1}\mathbf{p}\right)$$
$$\mathbf{a}_i = \Im\left(\left(\left(\boldsymbol{\Sigma}_0^{-1}\right)\right)^{\mathrm{T}} \overline{\mathbf{p}} \otimes \boldsymbol{\Sigma}_0^{-1}\mathbf{p}\right)$$

and $\mathbf{M} = \frac{1}{2}\begin{bmatrix} \Re\mathbf{M}_1 + \Re\mathbf{M}_2 & -\Im(\mathbf{M}_1) + \Im(\mathbf{M}_2) \\ \Im(\mathbf{M}_1) + \Im(\mathbf{M}_2) & \Re\mathbf{M}_1 - \Re\mathbf{M}_2 \end{bmatrix}$.

On the other hand, using results from [20], one can prove that under the large-n, N regime, $\frac{\sqrt{n}}{N}\mathbf{p}^*\hat{\mathbf{C}}_N^{-1}(\rho)\mathbf{p}$ satisfies:

$$T_{n,N} \triangleq \sqrt{\frac{n}{\sigma_N^2}}\left(\frac{1}{N}\mathbf{p}^*\hat{\mathbf{C}}_N^{-1}(\rho)\mathbf{p} - \frac{1}{N}\mathbf{p}^*\mathbf{Q}_N(\underline{\rho})\mathbf{p}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

where:

$$\sigma_N^2 = \frac{m(-\underline{\rho})^2(1-\underline{\rho})^2\left(\frac{1}{N}\mathbf{p}^*\boldsymbol{\Sigma}_N\mathbf{Q}_N^2\mathbf{p}\right)^2}{\rho^2(1-cm(-\underline{\rho})^2(1-\underline{\rho})^2\frac{1}{N}\boldsymbol{\Sigma}_N^2\mathbf{Q}_N^2(\underline{\rho}))}$$

with $\underline{\rho}$, $m(-\underline{\rho})$ and $\mathbf{Q}(\underline{\rho})$ have the same expressions as in [20] when $\mathbf{C}_N$ in [20] is replaced by $\boldsymbol{\Sigma}_N$. A natural question that arises is which of the two competing results is the most reliable for a particular set of values N and n. To answer this question, we plot in Figures 5, 6 and 7 the Kolmogorov-Smirnov distance between the standard normal distribution and the empirical distribution function of T_n and that of $T_{n,N}$, obtained over 100 000 realizations versus the ratio $\frac{n}{N}$ when $b = 0.7j, \rho = 0.5, \mathbf{p} = [1, \cdots, 1]$ and for $N = 4, 16, 32$. We note that for values of N up to 16, results derived under the large-n regime are more accurate for a large range of n while the use of the results from the large-n, N regime seems to be recommended for $N = 32$.

V. Conclusions

This paper addresses the study of the statistics of the RTE under the classical large$-n$ regime in which the observations' dimension is fixed while their number tends to infinity. It is worth noticing that despite the popularity of the RTE, characterizing its statistical properties has remained unclear

until the work in [20] shedding light on its behavior when the large-n, N regime is considered (the number of observations n and their size N growing simultaneously to infinity.). Interestingly, no results were provided for the standard large-n regime in which N is fixed while n goes to infinity. This has motivated our work. In particular, we established in the present work that the RTE converges, under the large-n regime, to a deterministic matrix which differs as expected from the true population covariance matrix. An important feature of this results is that it allows for the computation of the asymptotic bias incurred by the use of the RTE. Additionally, we analyzed the fluctuations of the RTE around its limit and proved that they converge to a multivariate Gaussian distribution with zero mean and a covariance matrix depending on the true population covariance matrix and the regularization parameter. Interestingly, simulations have shown the superiority, for a wide range of N and n, of the accuracy of the second-order results associated with quadratic forms of the robust scatter estimator, derived in the large-n regime, over those obtained in the large$-n, N$ regime. This opens up possibilities of using our results in order to get a more refined analysis of RTE based estimators.

Appendix A
Proof of Lemma 2

In the following appendices, for readability purposes, the notation $\boldsymbol{\Sigma}_0(\rho)$ (resp. $\tilde{\boldsymbol{\Sigma}}(\rho)$) is simply replaced by $\boldsymbol{\Sigma}_0$ (resp. $\tilde{\boldsymbol{\Sigma}}$). Of course, the dependence of $\boldsymbol{\Sigma}_0$ to ρ is not omitted.

Multiplying both sides of (2) by $\boldsymbol{\Sigma}_N^{-1}$, we show that $\boldsymbol{\Sigma}_0$ satisfies:

$$(1-\rho)\mathbb{E}\left[\frac{\mathbf{w}\mathbf{w}^*}{\frac{1}{N}\mathbf{w}^*\boldsymbol{\Sigma}_N^{\frac{1}{2}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_N^{\frac{1}{2}}\mathbf{w}}\right]+\rho\boldsymbol{\Sigma}_N^{-1}=\boldsymbol{\Sigma}_N^{-\frac{1}{2}}\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_N^{-\frac{1}{2}},$$

where $\mathbf{w}$ is zero-mean distributed with covariance matrix $\mathbf{I}_N$. Define $\mathbf{A}=\boldsymbol{\Sigma}_N^{-\frac{1}{2}}\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_N^{-\frac{1}{2}}$. Then,

$$\mathbf{A}=(1-\rho)\mathbb{E}\left[\frac{\mathbf{w}\mathbf{w}^*}{\frac{1}{N}\mathbf{w}^*\mathbf{A}^{-1}\mathbf{w}}\right]+\rho\boldsymbol{\Sigma}^{-1}$$

which yields the following bound for $\|\mathbf{A}\|$,

$$\|\mathbf{A}\|\leq(1-\rho)\|\mathbf{A}\|+\frac{\rho}{\lambda_{\min}(\boldsymbol{\Sigma}_N)}.$$

Hence,

$$\|\mathbf{A}\|\leq\frac{1}{\lambda_{\min}(\boldsymbol{\Sigma}_N)}. \tag{17}$$

Now, $\|\mathbf{A}\|$ can be lower-bounded by:

$$\begin{aligned}\|\mathbf{A}\|&=\max_{\|\mathbf{x}\|=1}\mathbf{x}^*\boldsymbol{\Sigma}_N^{-\frac{1}{2}}\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_N^{-\frac{1}{2}}\mathbf{x}\\&\overset{(a)}{\geq}\|\boldsymbol{\Sigma}_0\|\max_{\|\mathbf{x}\|=1}\mathbf{x}^*\boldsymbol{\Sigma}_N^{-\frac{1}{2}}\mathbf{u}\mathbf{u}^*\boldsymbol{\Sigma}_N^{-\frac{1}{2}}\mathbf{x}\\&\geq\|\boldsymbol{\Sigma}_0\|\mathbf{u}^*\boldsymbol{\Sigma}_N^{-\frac{1}{2}}\mathbf{u}\mathbf{u}^*\boldsymbol{\Sigma}_N^{-\frac{1}{2}}\mathbf{u}\\&\geq\frac{\|\boldsymbol{\Sigma}_0\|}{\|\boldsymbol{\Sigma}_N\|},\end{aligned} \tag{18}$$

where in (a) $\mathbf{u}$ is the eigenvector corresponding to the maximum eigenvalue of $\boldsymbol{\Sigma}_0$. Combining (17) and (18), we thus obtain:

$$\|\boldsymbol{\Sigma}_0\|\leq\frac{\|\boldsymbol{\Sigma}_N\|}{\lambda_{\min}(\boldsymbol{\Sigma}_N)}.$$

Appendix B
Proof of Theorem 3

The proof is based on controlling the random elements $d_i(\rho)$ given by:

$$d_i(\rho)=\frac{\mathbf{x}_i^*\hat{\mathbf{C}}_N^{-1}(\rho)\mathbf{x}_i-\mathbf{x}_i^*\boldsymbol{\Sigma}_0^{-1}\mathbf{x}_i}{\sqrt{\mathbf{x}_i^*\boldsymbol{\Sigma}_0^{-1}\mathbf{x}_i}\sqrt{\mathbf{x}_i^*\hat{\mathbf{C}}_N^{-1}(\rho)\mathbf{x}_i}}.$$

Recall that, by the SLLN, under the large-n regime, $\boldsymbol{\Sigma}_0$ satisfies:

$$\boldsymbol{\Sigma}_0=N(1-\rho)\frac{1}{n}\sum_{i=1}^n\frac{\mathbf{x}_i\mathbf{x}_i^*}{\mathbf{x}_i^*\boldsymbol{\Sigma}_0^{-1}\mathbf{x}_i}+\rho\mathbf{I}_N+\boldsymbol{\epsilon}_n(\rho),$$

where $\boldsymbol{\epsilon}_n$ is a $N\times N$ matrix whose elements converge almost surely to zero and satisfy $[\boldsymbol{\epsilon}_n(\rho)]_{i,j}=\mathcal{O}_p(\frac{1}{n})$.

In the sequel, we prove that for any $\kappa>0$,

$$\sup_{\rho\in[\kappa,1]}\max_{1\leq i\leq n}|d_i(\rho)|\xrightarrow{\text{a.s.}}0.$$

For that, we need to work out the differences $\mathbf{x}_i^*\hat{\mathbf{C}}_N^{-1}(\rho)\mathbf{x}_i-\mathbf{x}_i^*\boldsymbol{\Sigma}_0^{-1}\mathbf{x}_i$ for $i=1,\cdots,n$. Using the resolvent identity $\mathbf{A}^{-1}-\mathbf{B}^{-1}=\mathbf{A}^{-1}(\mathbf{B}-\mathbf{A})\mathbf{B}^{-1}$ for any $N\times N$ invertible matrices, we obtain:

$$\begin{aligned}&\mathbf{x}_j^*\hat{\mathbf{C}}_N^{-1}(\rho)\mathbf{x}_j-\mathbf{x}_j^*\boldsymbol{\Sigma}_0^{-1}\mathbf{x}_j\\&=\mathbf{x}_j^*\hat{\mathbf{C}}_N^{-1}\left[\frac{1-\rho}{n}\sum_{i=1}^n\frac{\mathbf{x}_i\mathbf{x}_i^*\left(\frac{1}{N}\mathbf{x}_i^*\boldsymbol{\Sigma}_0^{-1}\mathbf{x}_i-\frac{1}{N}\mathbf{x}_i^*\hat{\mathbf{C}}_N^{-1}(\rho)\mathbf{x}_i\right)}{\frac{1}{N}\mathbf{x}_i^*\boldsymbol{\Sigma}_0^{-1}\mathbf{x}_i\frac{1}{N}\mathbf{x}_i^*\hat{\mathbf{C}}_N^{-1}(\rho)\mathbf{x}_i}\right.\\&\left.+\boldsymbol{\epsilon}_n\right]\boldsymbol{\Sigma}_0^{-1}\mathbf{x}_j\\&=\frac{1-\rho}{n}\sum_{i=1}^n\frac{-\mathbf{x}_j^*\hat{\mathbf{C}}_N^{-1}(\rho)\mathbf{x}_i\mathbf{x}_i^*\boldsymbol{\Sigma}_0^{-1}\mathbf{x}_jd_i(\rho)}{\sqrt{\frac{1}{N}\mathbf{x}_i^*\hat{\mathbf{C}}_N^{-1}(\rho)\mathbf{x}_i\frac{1}{N}\mathbf{x}_i^*\boldsymbol{\Sigma}_0^{-1}\mathbf{x}_i}}\\&+\mathbf{x}_j^*\hat{\mathbf{C}}_N^{-1}(\rho)\boldsymbol{\epsilon}_n\boldsymbol{\Sigma}_0^{-1}\mathbf{x}_j.\end{aligned}$$

Hence,

$$\begin{aligned}d_j(\rho)=&\frac{\frac{1-\rho}{n}\sum_{i=1}^n\frac{-\mathbf{x}_j^*\hat{\mathbf{C}}_N^{-1}(\rho)\mathbf{x}_i\mathbf{x}_i^*\boldsymbol{\Sigma}_0^{-1}\mathbf{x}_jd_i(\rho)}{\sqrt{\frac{1}{N}\mathbf{x}_i^*\boldsymbol{\Sigma}_0^{-1}\mathbf{x}_i\frac{1}{N}\mathbf{x}_i^*\hat{\mathbf{C}}_N^{-1}(\rho)\mathbf{x}_i}}}{\sqrt{\mathbf{x}_j^*\hat{\mathbf{C}}_N^{-1}(\rho)\mathbf{x}_j\mathbf{x}_j^*\boldsymbol{\Sigma}_0^{-1}\mathbf{x}_j}}\\&+\frac{\mathbf{x}_j^*\hat{\mathbf{C}}_N^{-1}(\rho)\boldsymbol{\epsilon}_n\boldsymbol{\Sigma}_0^{-1}\mathbf{x}_j}{\sqrt{\mathbf{x}_j^*\hat{\mathbf{C}}_N^{-1}(\rho)\mathbf{x}_j\mathbf{x}_j^*\boldsymbol{\Sigma}_0^{-1}\mathbf{x}_j}}.\end{aligned}$$

Let $d_{\max}(\rho)=\max_{1\leq j\leq n}|d_j(\rho)|$. By the Cauchy-Schwartz

inequality, we thus obtain:

$$\begin{aligned} d_{\max}(\rho) \leq & \frac{d_{\max}(\rho)}{\sqrt{\mathbf{x}_j^* \hat{\mathbf{C}}_N^{-1}(\rho)\mathbf{x}_j \mathbf{x}_j^* \boldsymbol{\Sigma}_0^{-1}\mathbf{x}_j}} \\ & \times \sqrt{\frac{1-\rho}{n}\sum_{i=1}^n \frac{\mathbf{x}_j^* \hat{\mathbf{C}}_N^{-1}(\rho)\mathbf{x}_i \mathbf{x}_i^* \hat{\mathbf{C}}_N^{-1}(\rho)\mathbf{x}_j}{\frac{1}{N}\mathbf{x}_i^* \hat{\mathbf{C}}_N^{-1}(\rho)\mathbf{x}_i}} \\ & \times \sqrt{\frac{1-\rho}{n}\sum_{i=1}^n \frac{\mathbf{x}_j^* \boldsymbol{\Sigma}_0^{-1}\mathbf{x}_i \mathbf{x}_i^* \boldsymbol{\Sigma}_0^{-1}\mathbf{x}_j}{\frac{1}{N}\mathbf{x}_i^* \boldsymbol{\Sigma}_0^{-1}\mathbf{x}_i}} \\ & + \| \hat{\mathbf{C}}_N^{-\frac{1}{2}}(\rho)\boldsymbol{\epsilon}_n \boldsymbol{\Sigma}_0^{-\frac{1}{2}} \|. \end{aligned}$$

Therefore,

$$\begin{aligned} d_{\max}(\rho) \leq & \frac{d_{\max}(\rho)}{\sqrt{\mathbf{x}_j^* \hat{\mathbf{C}}_N^{-1}(\rho)\mathbf{x}_j \mathbf{x}_j^* \boldsymbol{\Sigma}_0^{-1}\mathbf{x}_j}} \\ & \times \sqrt{\mathbf{x}_j^* \hat{\mathbf{C}}_N^{-\frac{1}{2}} \left(\mathbf{I}_N - \rho \hat{\mathbf{C}}_N^{-1}(\rho) \right) \hat{\mathbf{C}}_N^{-\frac{1}{2}} \mathbf{x}_j} \\ & \times \sqrt{\mathbf{x}_j^* \boldsymbol{\Sigma}_0^{-\frac{1}{2}} \left(\mathbf{I}_N - \rho \boldsymbol{\Sigma}_0^{-1} \right) \boldsymbol{\Sigma}_0^{-\frac{1}{2}} \mathbf{x}_j - \mathbf{x}_j^* \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\epsilon}_n \boldsymbol{\Sigma}_0^{-1} \mathbf{x}_j} \\ & + \| \hat{\mathbf{C}}_N^{-\frac{1}{2}}(\rho)\boldsymbol{\epsilon}_n \boldsymbol{\Sigma}_0^{-\frac{1}{2}} \|. \end{aligned}$$

Using the relation $|\mathbf{x}^* \mathbf{A} \mathbf{y}| \leq \|\mathbf{x}\| \|\mathbf{A}\| \|\mathbf{y}\|$, we thus obtain:

$$\begin{aligned} d_{\max}(\rho) \leq & d_{\max}(\rho) \sqrt{\| \mathbf{I}_N - \rho \hat{\mathbf{C}}_N^{-1}(\rho) \|} \\ & \left(\| \mathbf{I}_N - \rho \boldsymbol{\Sigma}_0^{-1} \| - \frac{\mathbf{x}_j^* \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\epsilon}_n \boldsymbol{\Sigma}_0^{-1} \mathbf{x}_j}{\mathbf{x}_j^* \boldsymbol{\Sigma}_0^{-1} \mathbf{x}_j} \right)^{\frac{1}{2}} + \| \hat{\mathbf{C}}_N^{-\frac{1}{2}}(\rho) \boldsymbol{\epsilon}_n \boldsymbol{\Sigma}_0^{-\frac{1}{2}} \|. \end{aligned}$$

Since $\sup_{\rho \in [\kappa,1)} \| \hat{\mathbf{C}}_N^{-\frac{1}{2}} \boldsymbol{\epsilon}_n(\rho) \boldsymbol{\Sigma}_0^{-\frac{1}{2}} \| \leq \frac{1}{\kappa} \sup_{\rho \in [\kappa,1)} \| \boldsymbol{\epsilon}_n(\rho) \|$ and using the fact that $\| \mathbf{I}_N - \rho \hat{\mathbf{C}}_N^{-1}(\rho) \| \leq 1$, we get:

$$\begin{aligned} d_{\max}(\rho) \leq & d_{\max}(\rho) \left(\sqrt{\| \mathbf{I}_N - \rho \boldsymbol{\Sigma}_0^{-1} \|} \right. \\ & \left. + \sqrt{\| \boldsymbol{\Sigma}_0^{-\frac{1}{2}} \boldsymbol{\epsilon}_n \boldsymbol{\Sigma}_0^{-\frac{1}{2}} \|} \right) + \frac{1}{\kappa} \| \boldsymbol{\epsilon}_n \|. \end{aligned}$$

Again, as $\| \boldsymbol{\Sigma}_0^{-\frac{1}{2}} \boldsymbol{\epsilon}_n \boldsymbol{\Sigma}_0^{-\frac{1}{2}} \| \leq \frac{\| \boldsymbol{\epsilon}_n \|}{\kappa}$, we have:

$$d_{\max}(\rho) \left(1 - \sqrt{\| \mathbf{I}_N - \rho \boldsymbol{\Sigma}_0^{-1} \|} - \sqrt{\frac{1}{\kappa} \| \boldsymbol{\epsilon}_n \|} \right) \leq \frac{1}{\kappa} \| \boldsymbol{\epsilon}_n \|.$$

From Lemma 2, $\| \boldsymbol{\Sigma}_0 \| \leq \frac{\| \boldsymbol{\Sigma}_N \|}{\lambda_{\min}(\boldsymbol{\Sigma}_N)}$. Therefore, for n large enough (say large enough for the left-hand parenthesis to be greater than zero),

$$d_{\max}(\rho) \leq \frac{\frac{1}{\kappa} \| \boldsymbol{\epsilon}_n \|}{1 - \sqrt{1 - \rho \frac{\lambda_{\min}(\boldsymbol{\Sigma}_N)}{\| \boldsymbol{\Sigma}_N \|}} - \sqrt{\frac{1}{\kappa} \| \boldsymbol{\epsilon}_n \|}}.$$

Taking the supremum over $\rho \in [\kappa, 1)$, we finally obtain:

$$\sup_{\rho \in [\kappa,1)} d_{\max}(\rho) \leq \frac{\frac{1}{\kappa} \| \boldsymbol{\epsilon}_n \|}{1 - \sqrt{1 - \kappa \frac{\lambda_{\min}(\boldsymbol{\Sigma}_N)}{\| \boldsymbol{\Sigma}_N \|}} - \sqrt{\frac{1}{\kappa} \| \boldsymbol{\epsilon}_n \|}}.$$

thereby showing that $d_{\max}(\rho) \xrightarrow{\text{a.s.}} 0$ and $d_{\max}(\rho) = \mathcal{O}_p\left(\frac{1}{n}\right)$ Now, that the control of $d_{\max}(\rho)$ is performed, we are in position to handle the difference $\hat{\mathbf{C}}_N(\rho) - \boldsymbol{\Sigma}_0$. We have:

$$\begin{aligned} \hat{\mathbf{C}}_N(\rho) - \boldsymbol{\Sigma}_0 = & \frac{1-\rho}{n} \sum_{i=1}^n \frac{\mathbf{x}_i \mathbf{x}_i^* \left(\mathbf{x}_i^* \boldsymbol{\Sigma}_0^{-1} \mathbf{x}_i - \mathbf{x}_i^* \hat{\mathbf{C}}_N^{-1}(\rho) \mathbf{x}_i \right)}{\mathbf{x}_i^* \hat{\mathbf{C}}_N^{-1}(\rho) \mathbf{x}_i \frac{1}{N} \mathbf{x}_i^* \boldsymbol{\Sigma}_0^{-1} \mathbf{x}_i} \\ & - \boldsymbol{\epsilon}_n(\rho) \\ = & \frac{1-\rho}{n} \sum_{i=1}^n \frac{-\mathbf{x}_i \mathbf{x}_i^* d_i(\rho)}{\sqrt{\frac{1}{N} \mathbf{x}_i^* \hat{\mathbf{C}}_N^{-1}(\rho) \mathbf{x}_i} \sqrt{\frac{1}{N} \mathbf{x}_i^* \boldsymbol{\Sigma}_0^{-1} \mathbf{x}_i}} \\ & - \boldsymbol{\epsilon}_n(\rho). \end{aligned}$$

Therefore,

$$\begin{aligned} & \| \hat{\mathbf{C}}_N(\rho) - \boldsymbol{\Sigma}_0 \| \\ & \leq d_{\max}(\rho) \left\| \frac{1-\rho}{n} \sum_{i=1}^n \frac{\mathbf{x}_i \mathbf{x}_i^*}{\sqrt{\frac{1}{N} \mathbf{x}_i^* \hat{\mathbf{C}}_N^{-1}(\rho) \mathbf{x}_i} \sqrt{\frac{1}{N} \mathbf{x}_i^* \boldsymbol{\Sigma}_0^{-1} \mathbf{x}_i}} \right\| \\ & + \| \boldsymbol{\epsilon}_n(\rho) \|. \end{aligned}$$

By the Cauchy-Schwartz inequality, we get:

$$\begin{aligned} \| \hat{\mathbf{C}}_N(\rho) - \boldsymbol{\Sigma}_0 \| \leq & d_{\max}(\rho) \left\| \frac{1-\rho}{n} \sum_{i=1}^n \frac{\mathbf{x}_i \mathbf{x}_i^*}{\frac{1}{N} \mathbf{x}_i^* \hat{\mathbf{C}}_N^{-1}(\rho) \mathbf{x}_i} \right\|^{\frac{1}{2}} \\ & \times \left\| \frac{1-\rho}{n} \sum_{i=1}^n \frac{\mathbf{x}_i \mathbf{x}_i^*}{\frac{1}{N} \mathbf{x}_i^* \boldsymbol{\Sigma}_0^{-1} \mathbf{x}_i} \right\|^{\frac{1}{2}} + \| \boldsymbol{\epsilon}_n(\rho) \| \end{aligned}$$

or equivalently:

$$\begin{aligned} \| \hat{\mathbf{C}}_N(\rho) - \boldsymbol{\Sigma}_0 \| \leq & d_{\max}(\rho) \left\| \hat{\mathbf{C}}_N - \rho \mathbf{I}_N \right\|^{\frac{1}{2}} \| \boldsymbol{\Sigma}_0 - \rho \mathbf{I}_N - \boldsymbol{\epsilon}_n \|^{\frac{1}{2}} \\ & + \| \boldsymbol{\epsilon}_n(\rho) \|. \end{aligned}$$

Since $d_{\max}(\rho) \xrightarrow{\text{a.s.}} 0$, to conclude, we need to check that the spectral norm of $\hat{\mathbf{C}}_N$ is almost surely bounded. The proof is almost the same as that proposed in Lemma 2 to control the spectral norm of $\boldsymbol{\Sigma}_0$ with the slight difference that the expectation operator is replaced by the empirical average, and using additionally the fact that $\frac{1}{n} \sum_{i=1}^n \frac{\mathbf{w}_i \mathbf{w}_i^*}{\mathbf{w}_i^* \mathbf{w}_i} \xrightarrow{\text{a.s.}} \frac{1}{N} \mathbf{I}_N$. Details are thus omitted.

Appendix C
Proof of Lemma 4

The proof of Lemma 4 is based on the same technique as in [27]. Using the relation $\frac{1}{\alpha} = \int_0^{+\infty} e^{-\alpha t} dt$, we write $\mathbb{E}\left[\frac{|w_i|^2}{\mathbf{w}^* \mathbf{D} \mathbf{w}} \right]$ as:

$$\begin{aligned} & \mathbb{E}\left[\frac{|w_i|^2}{\mathbf{w}^* \mathbf{D} \mathbf{w}} \right] = \mathbb{E}\left[|w_i|^2 \int_0^{+\infty} e^{-t\left(d_i |w_i|^2 + \sum_{j=1, j \neq i}^N |w_j|^2 d_j \right)} \right] \\ & = \int_0^{+\infty} \int_0^{+\infty} \frac{1}{2^N} e^{-t d_i u} u \exp(-u/2) \int_0^{+\infty} \cdots \int_0^{+\infty} \\ & \times \exp\left(-t \sum_{j=1, j \neq i} u_j d_j \right) \prod_{j=1, j \neq i}^N e^{-u_j/2} du_1 \cdots du_{N-1} du dt \\ & = \int_0^{\infty} \frac{1}{2^N} \frac{1}{(\frac{1}{2} + t d_i)} \prod_{j=1}^N \frac{1}{\frac{1}{2} + t d_j} dt. \end{aligned}$$

Conducting the change of variable $t=\frac{1}{v}-1$, we eventually obtain:

$$\mathbb{E}\left[\frac{|w_i|^2}{\mathbf{w}^*\mathbf{D}\mathbf{w}}\right]=\int_0^1 \frac{1}{2^N}\frac{v^{N-1}}{d_i\prod_{j=1}^N d_j(1-v\frac{d_i-\frac{1}{2}}{d_i})}\prod_{j=1}^N\frac{1}{1-v\frac{d_j-\frac{1}{2}}{d_j}}dv.$$

We finally end the proof by using the integral representation of the Lauricella's type D hypergeometric function.

Appendix D
Proof of Lemma 5

Again the proof of the results in Lemma 5 follows the same lines as in Appendix C. We will only detail the derivations for the expressions of $\beta_{i,i}, i=1,\cdots,N$. The same kind of calculations can be used to derive that of $\beta_{i,j}, i\neq j$. Using the relation $\frac{1}{\alpha^2}=\int_0^\infty te^{-\alpha t}dt$, we write $\beta_{i,i}=\mathbb{E}\left[\frac{|w_i|^4}{(\mathbf{w}^*\mathbf{D}\mathbf{w})^2}\right]$ as:

$$\begin{aligned}\beta_{i,i}&=\mathbb{E}\left[|w_i|^4\int_0^\infty te^{-t|w|_i^2+\sum_{j=1,j\neq i}|w_j|^2d_j}\right]\\&=\int_0^\infty\int_0^\infty\frac{t}{2^N}u^2e^{-td_iu}u\exp(-u/2)\int_0^\infty\cdots\int_0^\infty\\&\times\exp\left(-t\sum_{j=1,j\neq i}^N u_jd_j\right)\prod_{j=1,j\neq i}^N e^{-u_j/2}du_1\cdots du_{N-1}dudt\\&=\frac{1}{2^{N-1}}\int_0^\infty\frac{t}{\left(\frac{1}{2}+td_i\right)^2}\prod_{k=1}^N\frac{1}{\frac{1}{2}+td_k}dt.\end{aligned}$$

Conducting the change of variable $t=\frac{1}{v}-1$, we obtain:

$$\beta_{i,i}=\frac{1}{2^{N-1}}\int_0^1\frac{(1-v)v^{N-1}dv}{d_i^2\prod_{k=1}^N d_k\left(1-\frac{v(d_i-\frac{1}{2})}{d_i}\right)^2\prod_{k=1}^N(\frac{v(\frac{1}{2}-d_k)}{d_k}+1)}.$$

Appendix E
Proof of Theorem 8

Our approach is based on a perturbation analysis of $\mathrm{vec}(\hat{\mathbf{C}}_N(\rho))$ in the vicinity of the asymptotic limit $\mathbf{\Sigma}_0$ coupled with the use of the Slutsky Theorem [25] which allows us to discard terms converging to zero in probability.

Set $\mathbf{\Delta}=\mathbf{\Sigma}_0^{-\frac{1}{2}}\left(\hat{\mathbf{C}}_N(\rho)-\mathbf{\Sigma}_0\right)\mathbf{\Sigma}_0^{-\frac{1}{2}}$. Then,

$$\mathbf{\Delta}=\frac{N(1-\rho)}{n}\sum_{i=1}^n\frac{\mathbf{\Sigma}_0^{-\frac{1}{2}}\mathbf{x}_i\mathbf{x}_i^*\mathbf{\Sigma}_0^{-\frac{1}{2}}}{\mathbf{x}_i^*\hat{\mathbf{C}}_N^{-1}(\rho)\mathbf{x}_i}+\rho\mathbf{\Sigma}_0^{-1}-\mathbf{I}_N.$$

Writing $\hat{\mathbf{C}}_N^{-1}$ as:

$$\begin{aligned}\hat{\mathbf{C}}_N^{-1}&=\left(\hat{\mathbf{C}}_N-\mathbf{\Sigma}_0+\mathbf{\Sigma}_0\right)^{-1}\\&=\mathbf{\Sigma}_0^{-\frac{1}{2}}\left(\mathbf{I}_N+\mathbf{\Delta}\right)^{-1}\mathbf{\Sigma}_0^{-\frac{1}{2}}\\&=\mathbf{\Sigma}_0^{-1}-\mathbf{\Sigma}_0^{-\frac{1}{2}}\mathbf{\Delta}\mathbf{\Sigma}_0^{-\frac{1}{2}}+o_p(\|\Delta\|)\end{aligned}$$

we obtain:

$$\begin{aligned}\mathbf{\Delta}=&\frac{N(1-\rho)}{n}\sum_{i=1}^n\frac{\mathbf{\Sigma}_0^{-\frac{1}{2}}\mathbf{x}_i\mathbf{x}_i^*\mathbf{\Sigma}_0^{-\frac{1}{2}}}{\mathbf{x}_i^*\mathbf{\Sigma}_0^{-1}\mathbf{x}_i-\mathbf{x}_i^*\mathbf{\Sigma}_0^{-\frac{1}{2}}\mathbf{\Delta}\mathbf{\Sigma}_0^{-\frac{1}{2}}\mathbf{x}_i+o_p(\|\mathbf{\Delta}\|)}\\&+\rho\mathbf{\Sigma}_0^{-1}-\mathbf{I}_N.\end{aligned}$$

From [25, Lemma 2.12], $\mathbf{\Delta}$ writes finally as:

$$\begin{aligned}\mathbf{\Delta}=&\frac{N(1-\rho)}{n}\sum_{i=1}^n\frac{\mathbf{\Sigma}_0^{-\frac{1}{2}}\mathbf{x}_i\mathbf{x}_i^*\mathbf{\Sigma}_0^{-\frac{1}{2}}}{\mathbf{x}_i^*\mathbf{\Sigma}_0^{-1}\mathbf{x}_i}\left(1+\frac{\mathbf{x}_i^*\mathbf{\Sigma}_0^{-\frac{1}{2}}\mathbf{\Delta}\mathbf{\Sigma}_0^{-\frac{1}{2}}\mathbf{x}_i}{\mathbf{x}_i^*\mathbf{\Sigma}_0^{-1}\mathbf{x}_i}\right)\\&+\rho\mathbf{\Sigma}_0^{-1}-\mathbf{I}_N+o_p(\|\mathbf{\Delta}\|)\\=&\mathbf{\Sigma}_0^{-\frac{1}{2}}\tilde{\mathbf{\Sigma}}\mathbf{\Sigma}_0^{-\frac{1}{2}}-\mathbf{I}_N\\&+\frac{N(1-\rho)}{n}\sum_{i=1}^n\frac{\mathbf{\Sigma}_0^{-\frac{1}{2}}\mathbf{x}_i\mathbf{x}_i^*\mathbf{\Sigma}_0^{-\frac{1}{2}}\mathbf{x}_i^*\mathbf{\Sigma}_0^{-\frac{1}{2}}\mathbf{\Delta}\mathbf{\Sigma}_0^{-\frac{1}{2}}\mathbf{x}_i}{\left(\mathbf{x}_i^*\mathbf{\Sigma}_0^{-1}\mathbf{x}_i\right)^2}+o_p(\|\mathbf{\Delta}\|)\\=&\mathbf{\Sigma}_0^{-\frac{1}{2}}\tilde{\mathbf{\Sigma}}\mathbf{\Sigma}_0^{-\frac{1}{2}}-\mathbf{I}_N\\&+\frac{N(1-\rho)}{n}\sum_{i=1}^n\frac{\mathbf{\Sigma}_0^{-\frac{1}{2}}\mathbf{x}_i\mathbf{x}_i^*\mathbf{\Sigma}_0^{-\frac{1}{2}}\left(\mathbf{x}_i^T(\mathbf{\Sigma}_0^{-\frac{1}{2}})^T\otimes\mathbf{x}_i^*\mathbf{\Sigma}_0^{-\frac{1}{2}}\right)\mathrm{vec}(\mathbf{\Delta})}{\left(\mathbf{x}_i^*\mathbf{\Sigma}_0^{-1}\mathbf{x}_i\right)^2}\\&+o_p(\|\mathbf{\Delta}\|).\end{aligned}$$

Let $\mathbf{F}$ be the $N^2\times N^2$ matrix given by:

$$\mathbf{F}=\frac{N(1-\rho)}{n}\sum_{i=1}^n\frac{\mathrm{vec}\left(\mathbf{\Sigma}_0^{-\frac{1}{2}}\mathbf{x}_i\mathbf{x}_i^*\mathbf{\Sigma}_0^{-\frac{1}{2}}\right)\left(\mathbf{x}_i^T(\mathbf{\Sigma}_0^{-\frac{1}{2}})^T\otimes\mathbf{x}_i^*\mathbf{\Sigma}_0^{-\frac{1}{2}}\right)}{\left(\mathbf{x}_i^*\mathbf{\Sigma}_0^{-1}\mathbf{x}_i\right)^2}.$$

Then, $\mathrm{vec}(\mathbf{\Delta})$ satisfies the following system of equations:

$$\begin{aligned}\mathrm{vec}(\mathbf{\Delta})=&\mathrm{vec}\left(\mathbf{\Sigma}_0^{-\frac{1}{2}}\tilde{\mathbf{\Sigma}}\mathbf{\Sigma}_0^{-\frac{1}{2}}-\mathbf{I}_N\right)+\mathbb{E}\left(\mathbf{F}\right)\mathrm{vec}(\mathbf{\Delta})\\&+\left(\mathbf{F}-\mathbb{E}(\mathbf{F})\right)\boldsymbol{\delta}+o_p(\|\boldsymbol{\delta}\|).\end{aligned}\tag{19}$$

Given that the two last terms in the right-hand side of (19) converges to zero at a rate faster than $\frac{1}{\sqrt{n}}$, we have:

$$\begin{aligned}\sqrt{n}\mathrm{vec}(\mathbf{\Delta})=&\sqrt{n}\left(\left(\mathbf{\Sigma}_0^{-\frac{1}{2}}\right)^{\mathrm{T}}\otimes\mathbf{\Sigma}_0^{-\frac{1}{2}}\right)\tilde{\boldsymbol{\delta}}+\sqrt{n}\mathbb{E}(\mathbf{F})\mathrm{vec}(\mathbf{\Delta})\\&+o_p(1).\end{aligned}\tag{20}$$

It remains thus to compute $\mathbb{E}(\mathbf{F})$ and to check that its spectral norm is less than 1. We will start by controlling the spectral norm of $\mathbb{E}(\mathbf{F})$. Recall that $\mathbb{E}(\mathbf{F})$ is given by:

$$\begin{aligned}\mathbb{E}(\mathbf{F})=&N(1-\rho)\\&\times\mathbb{E}\left[\frac{\mathrm{vec}\left(\mathbf{\Sigma}_0^{-\frac{1}{2}}\mathbf{x}\mathbf{x}^*\mathbf{\Sigma}_0^{-\frac{1}{2}}\right)\left(\mathbf{x}^{\mathrm{T}}(\mathbf{\Sigma}_0^{-\frac{1}{2}})^{\mathrm{T}}\otimes\mathbf{x}^*\mathbf{\Sigma}_0^{-\frac{1}{2}}\right)}{(\mathbf{x}^*\mathbf{\Sigma}_0^{-1}\mathbf{x})^2}\right]\\=&N(1-\rho)\mathbb{E}\left[\frac{\left((\mathbf{\Sigma}_0^{-\frac{1}{2}})^{\mathrm{T}}\overline{\mathbf{x}}\otimes\mathbf{\Sigma}_0^{-\frac{1}{2}}\mathbf{x}\right)\left(\mathbf{x}^{\mathrm{T}}(\mathbf{\Sigma}_0^{-\frac{1}{2}})^{\mathrm{T}}\otimes\mathbf{x}^*\mathbf{\Sigma}_0^{-\frac{1}{2}}\right)}{\left(\mathbf{x}^*\mathbf{\Sigma}_0^{-1}\mathbf{x}\right)^2}\right]\\=&N(1-\rho)\mathbb{E}\left[\frac{\left((\mathbf{\Sigma}_0^{-\frac{1}{2}})^{\mathrm{T}}\overline{\mathbf{x}}\mathbf{x}^{\mathrm{T}}(\mathbf{\Sigma}_0^{-\frac{1}{2}})^{\mathrm{T}}\right)\otimes\left(\mathbf{\Sigma}_0^{-\frac{1}{2}}\mathbf{x}\mathbf{x}^*\mathbf{\Sigma}_0^{-\frac{1}{2}}\right)}{\left(\mathbf{x}^*\mathbf{\Sigma}_0^{-1}\mathbf{x}\right)^2}\right].\end{aligned}$$

It can be easily noticed that: $\frac{(\mathbf{\Sigma}_0^{-\frac{1}{2}})^{\mathrm{T}}\overline{\mathbf{x}}\mathbf{x}^{\mathrm{T}}(\mathbf{\Sigma}_0^{-\frac{1}{2}})^{\mathrm{T}}}{\mathbf{x}^*\mathbf{\Sigma}_0^{-1}\mathbf{x}}\preceq\mathbf{I}_N$. Therefore,

$$\begin{aligned}\mathbb{E}(\mathbf{F})&\preceq N(1-\rho)\mathbf{I}_N\otimes\mathbb{E}\left[\frac{\mathbf{\Sigma}_0^{-\frac{1}{2}}\mathbf{x}\mathbf{x}^*\mathbf{\Sigma}_0^{-\frac{1}{2}}}{\mathbf{x}^*\mathbf{\Sigma}_0^{-1}\mathbf{x}}\right]\\&=\mathbf{I}_N\otimes\left(\mathbf{I}_N-\rho\mathbf{\Sigma}_0^{-1}\right)\end{aligned}$$

thus implying

$$\|\mathbb{E}(\mathbf{F})\|\leq\left\|\mathbf{I}_N-\rho\mathbf{\Sigma}_0^{-1}\right\|<1.$$

We will now provide a closed-form expression for $\mathbb{E}(\mathbf{F})$. To this end, we will use the eigenvalue decomposition of $\boldsymbol{\Sigma}_0^{-\frac{1}{2}}\boldsymbol{\Sigma}_N^{\frac{1}{2}} = \mathbf{U}\mathbf{D}^{\frac{1}{2}}\mathbf{U}^*$. Then, letting $\tilde{\mathbf{w}} = \mathbf{U}^*\mathbf{w}$ with $\mathbf{w} = \boldsymbol{\Sigma}_N^{-\frac{1}{2}}\mathbf{x}$, we obtain:

$$\mathbb{E}(\mathbf{F}) = \mathbb{E}\left[\frac{N(1-\rho)\overline{\mathbf{U}}\mathbf{D}^{\frac{1}{2}}(\overline{\tilde{\mathbf{w}}})\tilde{\mathbf{w}}^{\mathrm{T}}\mathbf{D}^{\frac{1}{2}}\mathbf{U}^{\mathrm{T}}\otimes\mathbf{U}\mathbf{D}^{\frac{1}{2}}\tilde{\mathbf{w}}\tilde{\mathbf{w}}^*\mathbf{D}^{\frac{1}{2}}\mathbf{U}^*}{\left(\tilde{\mathbf{w}}^*\mathbf{D}\tilde{\mathbf{w}}\right)^2}\right].$$

Therefore,

$$\begin{aligned}
&\left(\mathbf{D}^{-\frac{1}{2}}\mathbf{U}^{\mathrm{T}}\otimes\mathbf{D}^{-\frac{1}{2}}\mathbf{U}^*\right)\mathbb{E}(\mathbf{F})\left(\overline{\mathbf{U}}\mathbf{D}^{-\frac{1}{2}}\otimes\mathbf{U}\mathbf{D}^{-\frac{1}{2}}\right)\\
&= N(1-\rho)\mathbb{E}\left[\frac{(\overline{\tilde{\mathbf{w}}})\tilde{\mathbf{w}}^{\mathrm{T}}\otimes\tilde{\mathbf{w}}\tilde{\mathbf{w}}^*}{\left(\tilde{\mathbf{w}}^*\mathbf{D}\tilde{\mathbf{w}}\right)^2}\right]\\
&= N(1-\rho)\mathbb{E}\left[\frac{\left((\overline{\tilde{\mathbf{w}}})\otimes\tilde{\mathbf{w}}\right)(\tilde{\mathbf{w}}\otimes\tilde{\mathbf{w}}^*)}{\left(\tilde{\mathbf{w}}^*\mathbf{D}\tilde{\mathbf{w}}\right)^2}\right]\\
&= N(1-\rho)\mathbb{E}\left[\frac{\mathrm{vec}(\tilde{\mathbf{w}}\tilde{\mathbf{w}}^*)\left(\mathrm{vec}(\tilde{\mathbf{w}}\tilde{\mathbf{w}}^*)\right)^*}{\left(\tilde{\mathbf{w}}^*\mathbf{D}\tilde{\mathbf{w}}\right)^2}\right]\\
&= N(1-\rho)\tilde{\mathbf{B}}(\mathbf{D}),
\end{aligned}$$

where $\tilde{\mathbf{B}}(\mathbf{D})$ is provided by Lemma 4. A closed-form expression for $\tilde{\mathbf{F}} \triangleq \mathbb{E}(\mathbf{F})$ is thus given by:

$$\tilde{\mathbf{F}} = N(1-\rho)\left(\overline{\mathbf{U}}\mathbf{D}^{\frac{1}{2}}\otimes\mathbf{U}\mathbf{D}^{\frac{1}{2}}\right)\tilde{\mathbf{B}}(\mathbf{D})\left(\mathbf{D}^{\frac{1}{2}}\mathbf{U}^{\mathrm{T}}\otimes\mathbf{D}^{\frac{1}{2}}\mathbf{U}^*\right).$$

The linear system of equations in (20) thus becomes:

$$\sqrt{n}\mathrm{vec}(\boldsymbol{\Delta}) = \sqrt{n}(\mathbf{I}_N-\tilde{\mathbf{F}})^{-1}\left(\left(\boldsymbol{\Sigma}_0^{-\frac{1}{2}}\right)^{\mathrm{T}}\otimes\boldsymbol{\Sigma}_0^{-\frac{1}{2}}\right)\tilde{\boldsymbol{\delta}}+o_p(1).$$

Writing $\mathrm{vec}(\boldsymbol{\Delta}) = \left(\left(\boldsymbol{\Sigma}_0^{-\frac{1}{2}}\right)^{\mathrm{T}}\otimes\boldsymbol{\Sigma}_0^{-\frac{1}{2}}\right)\boldsymbol{\delta}$, we finally obtain:

$$\begin{aligned}
\sqrt{n}\boldsymbol{\delta} = &\left(\left(\boldsymbol{\Sigma}_0^{\frac{1}{2}}\right)^{\mathrm{T}}\otimes\boldsymbol{\Sigma}_0^{\frac{1}{2}}\right)(\mathbf{I}_{N^2}-\tilde{\mathbf{F}})^{-1}\left(\left(\boldsymbol{\Sigma}_0^{-\frac{1}{2}}\right)^{\mathrm{T}}\otimes\boldsymbol{\Sigma}_0^{-\frac{1}{2}}\right)\sqrt{n}\tilde{\boldsymbol{\delta}}\\
&+o_p(1).
\end{aligned}$$

Thus, $\sqrt{n}\boldsymbol{\delta}$ behaves as a zero-mean Gaussian distributed vector with covariance:

$$\begin{aligned}
\mathbf{M}_1 = &\left(\left(\boldsymbol{\Sigma}_0^{\frac{1}{2}}\right)^{\mathrm{T}}\otimes\boldsymbol{\Sigma}_0^{\frac{1}{2}}\right)(\mathbf{I}_{N^2}-\tilde{\mathbf{F}})^{-1}\left(\left(\boldsymbol{\Sigma}_0^{-\frac{1}{2}}\right)^{\mathrm{T}}\otimes\boldsymbol{\Sigma}_0^{-\frac{1}{2}}\right)\tilde{\mathbf{M}}_1\\
&\times\left(\left(\boldsymbol{\Sigma}_0^{-\frac{1}{2}}\right)^{\mathrm{T}}\otimes\boldsymbol{\Sigma}_0^{-\frac{1}{2}}\right)(\mathbf{I}_{N^2}-\tilde{\mathbf{F}})^{-1}\left(\left(\boldsymbol{\Sigma}_0^{\frac{1}{2}}\right)^{\mathrm{T}}\otimes\boldsymbol{\Sigma}_0^{\frac{1}{2}}\right)
\end{aligned}$$

and pseudo-covariance:

$$\begin{aligned}
\mathbf{M}_2 = &\left(\left(\boldsymbol{\Sigma}_0^{\frac{1}{2}}\right)^{\mathrm{T}}\otimes\boldsymbol{\Sigma}_0^{\frac{1}{2}}\right)(\mathbf{I}_{N^2}-\tilde{\mathbf{F}})^{-1}\left(\left(\boldsymbol{\Sigma}_0^{-\frac{1}{2}}\right)^{\mathrm{T}}\otimes\boldsymbol{\Sigma}_0^{-\frac{1}{2}}\right)\tilde{\mathbf{M}}_2\\
&\times\left(\boldsymbol{\Sigma}_0^{-\frac{1}{2}}\otimes\left(\boldsymbol{\Sigma}_0^{-\frac{1}{2}}\right)^{\mathrm{T}}\right)(\mathbf{I}_{N^2}-\tilde{\mathbf{F}}^{\mathrm{T}})^{-1}\left(\boldsymbol{\Sigma}_0^{\frac{1}{2}}\otimes\left(\boldsymbol{\Sigma}_0^{\frac{1}{2}}\right)^{\mathrm{T}}\right)
\end{aligned}$$

This completes the proof.

REFERENCES

[1] K. D. Ward, "Compound representation of high resolution sea clutter," *Electronics Letters*, vol. 17, no. 16, pp. 561–563, August 1981.

[2] S. Watts, "Radar detection prediction in sea clutter using the compound k-distribution model," *IEE Proceeding, Part. F*, vol. 132, no. 7, pp. 613–620, December 1985.

[3] T. Nohara and S. Haykin, "Canada east coast trials and the k-distribution," *IEE Proceeding, Part. F*, vol. 138, no. 2, pp. 82–88, 1991.

[4] J. B. Billingsley, "Ground clutter measurements for surface-sited radar," Tech. Rep. 780, MIT, February 1993.

[5] 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, Dec. 1970.

[6] P. J. Huber, "Robust estimation of a location parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964.

[7] P. J. Huber, "The 1972 wald lecture robust statistics : A review," *Annals of Mathematical Statistics*, vol. 43, no. 4, pp. 1041–1067, August 1972.

[8] R. A. Maronna, "Robust M-estimators of multivariate location and scatter," *Annals of Statistics*, vol. 4, no. 1, pp. 51–67, January 1976.

[9] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, "Complex elliptically symmetric distributions: survey, new Results and applications," *IEEE Trans. Signal Process.*, vol. 60, no. 11, Nov. 2012.

[10] M. Mahot, F. Pascal, P. Forster, and J. P. Ovalez, "Asymptotic properties of robust complex covariance matrix estimates," *IEEE Trans. Signal Process.*, vol. 61, no. 13, pp. 3348–3356, July 2013.

[11] 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," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 34–48, Jan. 2008.

[12] E. Ollila and D. E. Tyler, "Regularized M-estimators of scatter matrix," *IEEE Trans. Signal Process.*, vol. 62, no. 22, Nov. 2014.

[13] Y. Sun, P. Babu, and D. P. Palomar, "Regularized Tyler's scatter estimator: existence, uniqueness and algorithms," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 5143–5156, Oct. 2014.

[14] Y. Abramovich, "Controlled method for adaptive optimization of filters using the criterion of maximum SNR," *Radio Eng. Electron. Phys*, vol. 26, no. 3, pp. 87–95, Mar. 1981.

[15] B. D. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 4, pp. 397–401, July 1988.

[16] D. E. Tyler, "A distribution free M-estimator of multivariate scatter," *The Annals of Statistics*, vol. 15, no. 1, pp. 234–251, Mar. 1987.

[17] Y. Chen, A. Wiesel, and O. A. Hero, "Robust shrinkage estimation of high dimensional covariance matrices," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4907–4107, Apr. 2011.

[18] F. Pascal, Y. Chitour, and Y. Quek, "Generalized robust shrinkage esitmator and its application to STAP detection problem ," *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp. 5640–5651, Nov. 2014.

[19] A. Kammoun, R. Couillet, F. Pascal, and M. S. Alouini, "Optimal design of the adaptive normalized matched filter detector," *Submitted to IEEE Transactions on Signal Processing.*, 2015.

[20] R. Couillet, A. Kammoun, and F. Pascal, "Second order statistics of robust estimators of scatter. Application to GLRT detection for elliptical signals," *Submitted to Journal of Multivariate Analysis*, 2014, http://arxiv.org/abs/1410.0817.

[21] R. Couillet and M. McKay, "Large dimensional analysis and optimization of robust shrinkage covariance matrix estimators," *Journal of Multivariate Analysis*, vol. 31, pp. 99–120, 2014.

[22] O. Ledoit and M. Wolf, "A Well-conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis*, vol. 88, no. 2, pp. 365–411, Feb. 2004.

[23] F. Pascal and P. Forster and J. P. Ovarlez and P. Larzabal, "Performance analysis of covariance matrix estimate in impulsive noise," *IEEE Trans. Signal Process.*, vol. 56, no. 6, June 2008.

[24] R. D. Yates, "A Framework for uplink power control in cellular radio systems," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 7, pp. 1341–1347, Sept. 1995.

[25] A. W. van der Vaart, *Asymptotic Statistics*, Cambridge University Press, Cambridge, 1998.

[26] Harry L Van Trees, *Detection, Estimation, and Modulation Theory, Optimum Array Processing*, John Wiley & Sons, 2002.

[27] S. B. Provost and E. M. Rudiuk, "The exact density function of the ratio of two dependent linear combinations of chi-square variables," *Ann, Inst. Statist. Math*, vol. 46, no. 3, pp. 557–571, 1994.

Abla Kammoun was born in Sfax, Tunisia. She received the engineering degree in signal and systems from the Tunisia Polytechnic School, La Marsa, and the Master's degree and the Ph.D. degree in digital communications from Telecom Paris Tech [then Ecole Nationale Suprieure des Tlcommunications (ENST)]. From June 2010 to April 2012, she has been a Postdoctoral Researcher in the TSI Department, Telecom Paris Tech. Then she has been

at Suplec at the Alcatel-Lucent Chair on Flexible Radio until December 2013. Currently, she is a research scientist at KAUST university. Her research interests include performance analysis, random matrix theory, and semi-blind channel estimation. She is the recepient of the SAM 2014 second Prize Best paper award.

Romain Couillet received his MSc in Mobile Communications at the Eurecom Institute and his MSc in Communication Systems in Telecom ParisTech, France in 2007. From 2007 to 2010, he worked with ST-Ericsson as an Algorithm Development Engineer on the Long Term Evolution Advanced project, where he prepared his PhD with Supelec, France, which he graduated in November 2010. He is currently an assistant professor in the Telecommunication department of CentraleSuplec, France. His research topics are in random matrix theory applied to wireless communications, signal processing, and statistics. In 2015, he received the HDR title from University ParisSud. He is the recipient of the 2013 CNRS Bronze Medal in the section "science of information and its interactions", of the 2013 IEEE ComSoc Outstanding Young Researcher Award (EMEA Region), of the 2011 EEA/GdR ISIS/GRETSI best PhD thesis award, and of the Valuetools 2008 best student paper award.

Frédéric Pascal received the Master's degree ("Probabilities, Statistics Applications: Signal, Image et Networks") with merit, in Applied Statistics fromUniversity Paris VII - Jussieu, Paris, France, in 2003. Then, he received the Ph.D. degree of Signal Processing, from University Paris X - Nanterre, advised by Pr. Philippe Forster: "Detection and Estimation in Compound Gaussian Noise" in 2006. This Ph.D thesis was in collaboration with the French Aerospace Lab (ONERA), Palaiseau, France. From November 2006 to February 2008, he made a post-doctoral position in the Signal Processing and Information team of the laboratory SATIE, CNRS, ENS Cachan, France. Between March 2008 and December 2011 (resp. Jan. 2012 Dec. 2013), he was an Assistant Professor (resp. Associate Professor) in SONDRA, the French-Singaporean laboratory of CentraleSupelec. In 2012, he obtained a Research Directorship Habilitation (HDR) thesis in Signal Processing from the University of Paris-Sud. From January 2014, Frederic Pascal is a full Professor in the L2S laboratory at CentraleSupelec. From January 2015, Frederic Pascal is an elected senior member of the IEEE Signal Processing Society SAM technical committee and he is an Associate Editor for the IEEE Transactions on Signal Processing. His research interests contain estimation, detection and classification in statistical signal processing and radar processing.

Mohamed-Sim Alouini was born in Tunis, Tunisia. He received the Ph.D. degree in Electrical Engineering from the California Institute of Technology (Caltech), Pasadena, CA, USA, in 1998. He served as a faculty member in the University of Minnesota, Minneapolis, MN, USA, then in the Texas A&M University at Qatar, Education City, Doha, Qatar before joining King Abdullah University of Science and Technology (KAUST), Thuwal, Makkah Province, Saudi Arabia as a Professor of Electrical Engineering in 2009. His current research interests include the modeling, design, and performance analysis of wireless communication systems.