Journal of Multivariate Analysis 143 (2016) 249–274

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Second order statistics of robust estimators of scatter. Application to GLRT detection for elliptical signals[☆]

Romain Couillet[a,*], Abla Kammoun[b], Frédéric Pascal[a]

[a] CentraleSupélec, Université Paris Saclay, Gif-sur-Yvette, France

[b] King Abdullah's University of Science and Technology, Saudi Arabia

ARTICLE INFO

Article history:
Received 1 October 2014
Available online 14 September 2015

AMS subject classifications:
62H12
60F05
62G35

Keywords:
Random matrix theory
Robust estimation
Central limit theorem
GLRT

ABSTRACT

A central limit theorem for bilinear forms of the type $a^*\hat{C}_N(\rho)^{-1}b$, where $a, b \in \mathbb{C}^N$ are unit norm deterministic vectors and $\hat{C}_N(\rho)$ a robust-shrinkage estimator of scatter parametrized by ρ and built upon n independent elliptical vector observations, is presented. The fluctuations of $a^*\hat{C}_N(\rho)^{-1}b$ are found to be of order $N^{-\frac{1}{2}}$ and to be the same as those of $a^*\hat{S}_N(\rho)^{-1}b$ for $\hat{S}_N(\rho)$ a matrix of a theoretical tractable form. This result is exploited in a classical signal detection problem to provide an improved detector which is both robust to elliptical data observations (e.g., impulsive noise) and optimized across the shrinkage parameter ρ.

1. Introduction

As an aftermath of the growing interest for large dimensional data analysis in machine learning, in a recent series of articles [14,15,13,32,17], several estimators from the field of robust statistics (dating back to the seventies) started to be explored under the assumption of commensurably large sample (n) and population (N) dimensions. Robust estimators were originally designed to turn classical estimators into outlier- and impulsive noise-resilient estimators, which are of considerable importance in the recent big data paradigm. Among these estimation methods, robust regression was studied in [17] which reveals that, in the large N, n regime, the difference in norm between estimated and true regression vectors (of size N) tends almost surely to a positive constant which depends on the nature of the data and of the robust regressor. In parallel, and of more interest to the present work, Couillet et al. [14,15], Couillet and McKay [13], Zhang et al. [32] studied the limiting behavior of several classes of robust estimators $\hat{C}_N$ of scatter (or covariance) matrices C_N based on independent zero-mean elliptical observations $x_1, \ldots, x_n \in \mathbb{C}^N$. Precisely, Couillet et al. [14] show that, letting $N/n < 1$ and $\hat{C}_N$ be the (almost sure) unique solution to

$$\hat{C}_N = \frac{1}{n}\sum_{i=1}^{n} u\left(\frac{1}{N}x_i^*\hat{C}_N^{-1}x_i\right)x_ix_i^*$$

☆ Couillet's work is supported by the ERC MORE EC–120133. Pascal's work is supported by the DGA grant no. 2013.60.0011.00.470.75.01.

* Corresponding author.
E-mail addresses: romain.couillet@supelec.fr (R. Couillet), abla.kammoun@kaust.edu.sa (A. Kammoun), frederic.pascal@supelec.fr (F. Pascal).

http://dx.doi.org/10.1016/j.jmva.2015.08.021

under some appropriate conditions over the nonnegative function u (corresponding to Maronna's M-estimator [24]), $\|\hat{C}_N - \hat{S}_N\| \xrightarrow{\text{a.s.}} 0$ in spectral norm as $N, n \to \infty$ with $N/n \to c \in (0,1)$, where $\hat{S}_N$ follows a standard random matrix model (such as studied in [29,12]). In [32], the important scenario where $u(x) = 1/x$ (referred to as Tyler's M-estimator) is treated. It is in particular shown for this model that for identity scatter matrices the spectrum of $\hat{C}_N$ converges weakly to the Marčenko–Pastur law [23] in the large N, n regime. Finally, for $N/n \to c \in (0, \infty)$, Couillet and McKay [13] studied yet another robust estimation model defined, for each $\rho \in (\max\{0, 1 - n/N\}, 1]$, by $\hat{C}_N = \hat{C}_N(\rho)$, unique solution to

$$\hat{C}_N(\rho) = (1-\rho)\frac{1}{n}\sum_{i=1}^{n}\frac{x_i x_i^*}{\frac{1}{N}x_i^*\hat{C}_N^{-1}(\rho)x_i} + \rho I_N. \tag{1}$$

This estimator, proposed in [27], corresponds to a hybrid robust-shrinkage estimator reminding Tyler's M-estimator of scale [30] and Ledoit–Wolf's shrinkage estimator [22]. This estimator is particularly suited to scenarios where N/n is not small, for which other estimators are badly conditioned if not undefined. For this model, it is shown in [13] that $\sup_\rho \|\hat{C}_N(\rho) - \hat{S}_N(\rho)\| \xrightarrow{\text{a.s.}} 0$ where $\hat{S}_N(\rho)$ also follows a classical random matrix model.

The aforementioned approximations $\hat{S}_N$ of the estimators $\hat{C}_N$, the structure of which is well understood (as opposed to $\hat{C}_N$ which is only defined implicitly), allow for both a good apprehension of the limiting behavior of $\hat{C}_N$ and more importantly for a better usage of $\hat{C}_N$ as an appropriate substitute for sample covariance matrices in various estimation problems in the large N, n regime. The convergence in norm $\|\hat{C}_N - \hat{S}_N\| \xrightarrow{\text{a.s.}} 0$ is indeed sufficient in many cases to produce new consistent estimation methods based on $\hat{C}_N$ by simply replacing $\hat{C}_N$ by $\hat{S}_N$ in the problem defining equations. For example, the results of Couillet et al. [15] led to the introduction of novel consistent estimators based on functionals of $\hat{C}_N$ (of the Maronna type) for power and direction-of-arrival estimation in array processing in the presence of impulsive noise or rare outliers [11]. Similarly, in [13], empirical methods were designed to estimate the parameter ρ which minimizes the expected Frobenius norm $\operatorname{tr}[(\hat{C}_N(\rho) - C_N)^2]$, of interest for various outlier-prone applications dealing with non-small ratios N/n.[1]

Nonetheless, when replacing $\hat{C}_N$ for $\hat{S}_N$ in deriving consistent estimates, if the convergence $\|\hat{C}_N - \hat{S}_N\| \xrightarrow{\text{a.s.}} 0$ helps in producing novel consistent estimates, this convergence (which comes with no particular speed) is in general not sufficient to assess the performance of the estimator for large but finite N, n. Indeed, when second order results such as central limit theorems need to be established, say at rate $N^{-\frac{1}{2}}$, to proceed similarly to the replacement of $\hat{C}_N$ by $\hat{S}_N$ in the analysis, one would ideally demand that $\|\hat{C}_N - \hat{S}_N\| = o(N^{-\frac{1}{2}})$; but such a result, we believe, unfortunately does not hold. This constitutes a severe limitation in the exploitation of robust estimators as their performance as well as optimal fine-tuning often rely on second order performance. Concretely, for practical purposes in the array processing application of Couillet [11], one may naturally ask which choice of the u function is optimal to minimize the variance of (consistent) power and angle estimates. This question remains unanswered to this point for lack of better theoretical results.

The main purpose of the article is twofold. From a technical aspect, taking the robust shrinkage estimator $\hat{C}_N(\rho)$ defined by (1) as an example, we first show that, although the convergence $\|\hat{C}_N(\rho) - \hat{S}_N(\rho)\| \xrightarrow{\text{a.s.}} 0$ (from [13, Theorem 1]) may not be extensible to a rate $O(N^{1-\varepsilon})$, one has the bilinear form convergence $N^{1-\varepsilon}a^*(\hat{C}_N^k(\rho) - \hat{S}_N^k(\rho))b \xrightarrow{\text{a.s.}} 0$ for each $\varepsilon > 0$, each $a, b \in \mathbb{C}^N$ of unit norm, and each $k \in \mathbb{Z}$. This result implies that, if $\sqrt{N}a^*\hat{S}_N^k(\rho)b$ satisfies a central limit theorem, then so does $\sqrt{N}a^*\hat{C}_N^k(\rho)b$ with the same limiting variance. This result is of fundamental importance to any statistical application based on such quadratic forms. Our second contribution is to exploit this result for the specific problem of signal detection in impulsive noise environments via the generalized likelihood-ratio test, particularly suited for radar signals detection under elliptical noise [10,27]. In this context, we determine the shrinkage parameter ρ which minimizes the probability of false detections and provide an empirical consistent estimate for this parameter, thus improving significantly over traditional sample covariance matrix-based estimators.

The remainder of the article introduces our main results in Section 2 which are proved in Section 3. Technical elements of proof are provided in the Appendix.

Notations: In the remainder of the article, we shall denote $\lambda_1(X), \ldots, \lambda_n(X)$ the real eigenvalues of $n \times n$ Hermitian matrices X. The norm notation $\|\cdot\|$ being considered is the spectral norm for matrices and Euclidean norm for vectors. The symbol $\imath$ is the complex $\sqrt{-1}$.

2. Main results

Let $N, n \in \mathbb{N}$, $c_N \triangleq N/n$, and $\rho \in (\max\{0, 1 - c_N^{-1}\}, 1]$. Let also $x_1, \ldots, x_n \in \mathbb{C}^N$ be n independent random vectors defined by the following assumptions.

---

1 Other metrics may also be considered as in e.g. [31] with ρ chosen to minimize the return variance in a portfolio optimization problem.