ROBUST M-ESTIMATOR OF SCATTER FOR LARGE ELLIPTICAL SAMPLES

Romain Couillet1 and Frédéric Pascal2

1 Telecommunication department, Supélec, Gif sur Yvette, France.
2 SONDRA Laboratory, Supélec, Gif sur Yvette, France.

ABSTRACT

It is shown that a certain family of robust scatter estimators of elliptical samples behaves similar to a well-known random matrix model in the limiting regime where both the population $N$ and sample $n$ sizes grow to infinity at the same speed. This result allows us to understand the structure of such estimators and in particular to derive their limiting eigenvalue distributions. This analysis is a first step towards an improved usage of robust estimation methods when the number of independent observations is not too large compared to the size of the population.

Index Terms—random matrix theory, robust estimation.

I. INTRODUCTION AND PROBLEM STATEMENT

The recent advances in the spectral analysis of large dimensional random matrices, and particularly of matrices of the sample covariance type, have triggered a new wave of interest for (sometimes old) problems in statistical inference and signal processing, understanding the structure of such estimators and in particular to derive improved methods for source detection and parameter estimation for sample covariance matrix-based estimators. Adaptations (and improvements) of these results to robust estimation are currently under investigation. The results presented in this paper are an excerpt of the complete version [5] where more general results and all proofs can be found.

We now introduce our main notations and assumptions. Let $x_1, \ldots, x_n \in \mathbb{C}^N$ be $n$ random vectors with $x_i = \sqrt{n_i} A_N y_i$, where $\tau_1, \ldots, \tau_n \in \mathbb{R}^+$ and $y_1, \ldots, y_n \in \mathbb{C}^N$ are random and $A_N \in \mathbb{C}^{N \times N}$ is deterministic. We denote $c_N \triangleq N/n$ and $\bar{c}_N \triangleq N/N$ and shall consider the following growing regime.

Assumption 1: For each $N$, $c_N < 1$, $\bar{c}_N \geq 1$ and

$$0 < c_- \leq \lim inf \frac{c_N}{n} \leq \lim sup \frac{c_N}{n} < c_+ < 1.$$  

We define Maronna’s M-estimator $\hat{C}_N$, when it exists, as a (possibly unique) solution to the equation in $Z \in \mathbb{C}^{N \times N}$

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

(1)

where $u$ satisfies the following properties:

(i) $u : [0, \infty) \to (0, \infty)$ is nonnegative continuous and non-increasing

(ii) $\phi : x \mapsto x u(x)$ is increasing and bounded with $\lim_{x \to 0} \phi(x) \triangleq \phi_0 > 1$

(iii) $\phi_0 < c_+^{-1}$.

These assumptions are minor variations of Maronna’s original assumptions [3, p. 53]. Next we detail the conditions on the $x_i$’s.

Assumption 2: The vectors $x_i = \sqrt{n_i} A_N y_i$ satisfy:

1) $n_i = \frac{1}{N} \sum_{i=1}^{n} \delta_i x_i$ is such that $\int x u_n(dx) \to 1$

2) there exist $\varepsilon < 1 - \phi^{-1}_0 < 1 - c_+$ and $m > 0$ such that, for all large $n$ a.s. $\nu_n([0,m]) < \varepsilon$

3) defining $C_N \triangleq A_N A_N^*$, $C_N \sim 0$ and $\lim sup_N \|C_N\| < \infty$

4) $y_1, \ldots, y_n \in \mathbb{C}^N$ are independent unitarily invariant complex zero-mean vectors with, for each $i$, $\|y_i\|^2 = N$, and are independent of $\tau_1, \ldots, \tau_n$.

These conditions are met in particular if the $\tau_i$ are independent and identically distributed (i.i.d.) with unit mean distribution $\nu$ (then $\int x u_n(dx) \to 1$ by the strong law of large numbers) such that $\nu([0,1]) = 0$. If in addition $N \to \infty$, then $x_1, \ldots, x_n$ are i.i.d. zero-mean complex (or real) elliptically distributed with full rank [6, Theorem 3]. In particular, if $\tau_1$ is Rayleigh distributed, $x_1$ is complex zero mean Gaussian. If $1/\tau_1$ is chi-squared distributed, $x_1$
is instead zero mean complex Student distributed, etc. (see [6] for further discussions and recent results on elliptical distributions).

**Assumption 3:** For each $a > b > 0$, a.s.
\[
\limsup_{t \to \infty} \sup_{n \leq N} \nu_n((t, \infty)) = 0.
\]

Assumption 3 controls the relative speed of the tail of $\nu_n$ versus the flattening speed of $\phi(x)$ as $x \to \infty$. Practical examples satisfying Assumption 3 are:

- There exists $M > 0$ such that, for all $n$, $\max_{1 \leq i \leq n} \tau_i < M$ a.s. In this case, $\nu_n((t, \infty)) = 0$ a.s. for $t > M$ while $\phi(at) - \phi(bt) \neq 0$ since $\phi$ is increasing.
- For $u(t) = (1 + \alpha)/(\alpha + t)$ for some $\alpha > 0$ and $\tau_i$ i.i.d. with distribution $\nu$, by Markov inequality, it suffices that $\int x^{1+\epsilon} \nu(dx) < \infty$ for some $\epsilon > 0$.

This article provides two results: (i) existence and uniqueness of a solution $\hat{C}_N$ to (1) or chosen arbitrarily if not. Then

\[
\hat{C}_N = \lim_{t \to \infty} Z(t)^{t+1} = \frac{1}{n} \sum_{i=1}^{n} a \left( \frac{1}{N} x_i \right) \left( \frac{1}{Z(t)} \right)^{x_i} x_i \hat{x}_i^*.
\]

Having defined $\hat{C}_N$, the main result of the article provides a random matrix equivalent to $\hat{C}_N$, much easier to study than $\hat{C}_N$ itself.

**Theorem 1 (Uniqueness):** Let Assumptions 1 and 2 hold, with $\limsup_N \|C_N\|$ non necessarily bounded. Then, for all large $n$ a.s., (1) has a unique solution $\hat{C}_N$ given by

\[
\hat{C}_N = \lim_{t \to \infty} Z(t)^{(t)}
\]

where $Z^{(t)} > 0$ is arbitrary and, for $t \in \mathbb{N}$,

\[
Z^{(t+1)} = \frac{1}{n} \sum_{i=1}^{n} a \left( \frac{1}{N} x_i \right) \left( \frac{1}{Z(t)} \right)^{x_i} x_i \hat{x}_i^*.
\]

The fact that $\hat{C}_N$ is well approximated by $\hat{S}_N$, which follows a random matrix model studied extensively in [8], [9], has important consequences. From a purely mathematical standpoint, this provides a full characterization of the spectral behavior of $\hat{C}_N$ for large $N$, $n$. For application purposes, this first enables the performance analysis in the large $N$, $n$ horizon of standard signal processing methods already relying on $\hat{C}_N$ (these methods were so far analyzed solely in the fixed $N$ large $n$ regime). A second, more important, consequence for signal processing applications is the possibility to fully exploit the structure of $\hat{C}_N$ for large $N$, $n$ to improve existing robust schemes. Deriving such improved methods is not the subject of the current article but should be directly accessible from Theorem 2, while performance analysis of these methods may demand supplementary treatment, such as central limit theorems for functionals of $\hat{C}_N$.

An immediate corollary of Theorem 2 along with classical arguments from [9], [10] is when the $\tau_i$’s are i.i.d., leading to elliptical distributions for $x_i$, for which $\hat{C}_N$ has an (almost) limiting eigenvalue spectrum.

**Corollary 1 (Elliptical case):** Let Assumptions 1–3 hold and in addition, let $\tau_i$ be i.i.d. with law $\nu$ and let $cn \to c$. Then

\[
\hat{C}_N \sim \frac{1}{n} \sum_{i=1}^{n} \nu(\tau_i \gamma^*) x_i x_i^*
\]

\[
\text{where } \gamma^* \text{ is the unique positive solution to the equation in } \gamma
\]

\[
1 = \int \frac{\psi(s \gamma^*)}{1 + c \psi(s \gamma^*)} \nu(dt)
\]

with $\psi = \lim_{cn \to c} \psi$. Moreover, if $\frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(C_N)} \to \nu^C$ weakly, then

\[
\frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(\hat{C}_N)} \weakly \to \mu
\]

weakly with $\mu$ a probability measure with continuous density of bounded support $\delta$, the Stieltjes transform $m(z)$ of which is given for $z \in \mathbb{C}^+$ by

\[
m(z) = -\frac{1}{z} \int \frac{1}{1 + \delta(z) \nu^C(dt)}
\]

where $\delta(z)$ is the unique solution in $\mathbb{C}^+$ of the equations in $\hat{\delta}$

\[
\hat{\delta} = -\frac{1}{z} \int \frac{\psi(s \gamma^*)}{1 + c \psi(s \gamma^*)} \delta(dt)
\]

Finally, for every closed set $A \subset \mathbb{R}$ with $A \cap \mathcal{S} = \emptyset$,

\[
\left\{ \lambda_i(\hat{C}_N) \right\}_{i=1}^{N} \weakly \to \emptyset.
\]

Figure 1 depicts the empirical histograms of the eigenvalues of $\hat{C}_N$, for $N = 500$ and $n = 2500$ with $u(t) = (1 + \alpha)/(t + \alpha)$, $\alpha = 0.1$, $C_N = \text{diag}(I_{252}, 3I_{252}, 10I_{252})$, and $\tau_1, \ldots, \tau_n$ i.i.d. with $\Gamma(5,2)$ distribution. In thick line is also depicted the density of $\mu$ in Corollary 1 which shows an accurate match to the empirical spectrum. As a comparison, Figure 2 shows the empirical histogram

\[\text{This function } u(t) \text{ is often met in robust statistics as it is such that } \hat{C}_N \text{ corresponds to the maximum-likelihood estimate of the scale parameter of independent and identically distributed multivariate Student-t vectors.} \]
of the eigenvalues of the sample covariance matrix $\frac{1}{n} \sum_{i=1}^{n} x_i x_i^*$ under the same parametrization against the deterministic equivalent density for this model in thick line [7]. This graph presents an unbounded limiting eigenvalue spectrum support which is expected since $\tau_1$ has unbounded support. Also note the gain of separability in the spectrum of $\hat{C}_N$ which exhibits clearly three compacts subsets of eigenvalues, reminiscent of the three masses in the eigenvalue distribution of $C_N$, while $\frac{1}{n} \sum_{i=1}^{n} x_i x_i^*$ exhibits a single compact set of eigenvalues. This has important consequences from detection and estimation purposes in signal processing application of robust estimation.

The proof of Theorem 1 follows from a similar approach as in [4] and will therefore not be detailed here. We instead concentrate on the more fundamental Theorem 2.

### III. INTUITIVE DERIVATION OF THE MAIN RESULT

The proof of our main result, Theorem 2, is thoroughly detailed in [5]. Here we only provide an intuitive approach to this result (the rigorous proof follows a quite different approach). First note that we can assume $C_N = I_N$ by studying $C_N^{1/2} \hat{C}_N C_N^{-1/2}$ in place of $\hat{C}_N$ (see (1)). Therefore, from now on, we assume $C_N = A_N A_N^* = I_N$.

From there, the main difficulty to tackle lies in the dependence structure between the rank-one matrices $u \left( \frac{1}{\sqrt{n}} x_i^{1/2} \right)^* x_i x_i^*$ that compose $\hat{C}_N$: this structure disrupts from the standard random matrix assumptions, which rely on an explicit dependence of these rank-one matrices. At least intuitively, we can however weaken the dependence structure by rewriting the fundamental equation (1). This rewriting is performed in Section III-A below. Approximating weak dependence by independence, we then provide the final result. This is performed in Section III-B.

#### III-A. Rewriting (1)

We first introduce some new notations. Write $x_i = \sqrt{n} A_N y_i \triangleq \sqrt{n} z_i$ and recall that $C_N = I_N$ (in particular, $\|z_i\|$ is of order $\sqrt{n}$ for most $z_i$). Assuming $\hat{C}_N$ is well-defined, we denote $\hat{C}_N \triangleq \hat{C}_N - \frac{1}{n} u \left( \frac{1}{\sqrt{n}} x_i^{1/2} \right)^* x_i x_i^*$. Note that $\hat{C}_N$ depends on $x_i$ only through the terms $u \left( \frac{1}{\sqrt{n}} x_i^{1/2} \right)^* x_i x_i^*$, $j \neq i$, since $\hat{C}_N$ is built on $x_i$. But since $x_i$ is only one among a growing number $n$ of $x_j$ vectors, this dependence structure looks intuitively “weak”. This informal weak dependence between $x_i$ and $\hat{C}_N$, along with classical random matrix theory considerations, suggests that $\frac{1}{\sqrt{n}} z_i^{1/2} \hat{C}_N z_i^{1/2}$, $i = 1, \ldots, n$, are all well approximated by $\frac{1}{\sqrt{n}} \text{tr} \hat{C}_N$ (see e.g. [7, Lemma 3.1]).

With this in mind, we rewrite $\hat{C}_N$ as a function of $\frac{1}{\sqrt{n}} z_i^{1/2} \hat{C}_N z_i^{1/2}$ instead of $\frac{1}{\sqrt{n}} x_i^{1/2} \hat{C}_N x_i$, $i = 1, \ldots, n$. For this, let $Z \in \mathbb{C}^{N \times N}$ be positive definite such that for each $i$, $Z(i, i) \triangleq Z - \frac{1}{\sqrt{n}} u \left( \frac{1}{\sqrt{n}} z_i^{1/2} \right)^* \tau_i z_i z_i^*$ is positive definite. Using the identity $(A + \tau z z^*)^{-1} z = A^{-1} z / (1 + \tau z^* A^{-1} z)$ for invertible $A$, vector $z$, and positive scalar $\tau$, we have

$$1 \sqrt{n} z_i^{1/2} Z^{-1} z_i = \frac{1}{\sqrt{n} z_i^{1/2} Z^{-1} z_i}$$

so that

$$\frac{1}{\sqrt{n}} z_i^{1/2} Z^{-1} z_i \left( 1 - c_N \tau_i u \left( \frac{1}{\sqrt{n}} z_i^{1/2} Z^{-1} z_i \right) \frac{1}{\sqrt{n}} z_i^{1/2} Z^{-1} z_i \right) = \frac{1}{\sqrt{n}} z_i^{1/2} Z^{-1} z_i$$

which, by the definition of $\phi$, is

$$\frac{1}{\sqrt{n}} z_i^{1/2} Z^{-1} z_i \left( 1 - c_N \phi \left( \frac{1}{\sqrt{n}} z_i^{1/2} Z^{-1} z_i \right) \right) = \frac{1}{\sqrt{n}} z_i^{1/2} Z^{-1} z_i.$$

Using Assumption 1 and $\phi_\infty < c_N^{-1}$, taking $n$ large enough to have $\phi(x) \leq \phi_\infty < 1/c_N$, this can be rewritten

$$\frac{1}{\sqrt{n}} z_i^{1/2} Z^{-1} z_i = \frac{1}{1 - c_N \phi \left( \frac{1}{\sqrt{n}} z_i^{1/2} Z^{-1} z_i \right)}.$$

(2)
Now, since \( \phi \) is increasing, \( g : [0, \infty) \to [0, \infty) \), \( x \mapsto x/(1 - c_N \phi(x)) \) is increasing, nonnegative, and maps \([0, \infty)\) onto \([0, \infty)\). Thus, \( g \) is invertible with inverse denoted \( g^{-1} \). Thus, from (2),

\[
\tau_i \frac{1}{N} z_i^* Z^{-1} z_i = g^{-1} \left( \tau_i \frac{1}{N} z_i^* Z^{(i)} z_i \right).
\]

Call now \( v : [0, \infty) \to [0, \infty), x \mapsto u \circ g^{-1} \). Since \( g \) is increasing and nonnegative and \( u \) is non-increasing, \( v \) is non-increasing and positive. Moreover, \( \psi : x \mapsto xv(x) \) satisfies:

\[
\psi(x) = xv(g^{-1}(x)) = g(g^{-1}(x))u(g^{-1}(x)) = \phi(g^{-1}(x)) \frac{\phi(g^{-1}(x))}{1 - c_N \phi(g^{-1}(x))}
\]

which is increasing, nonnegative, with limit \( \psi^\infty = \phi^\infty / (1 - c_N \phi^\infty) \) as \( x \to \infty \). Hence, \( v \) and \( \psi \) keep the same properties as \( u \) and \( \phi \), respectively.

With these notations, any positive definite solution \( Z \) to (1) is equivalently a solution to

\[
Z = \frac{1}{n} \sum_{i=1}^n \tau_i v \left( \tau_i \frac{1}{N} z_i^* Z^{(i)} z_i \right) z_i z_i^*
\]

which is easily proved to be also characterized as the matrix \( Z = \frac{1}{n} \sum_{i=1}^n \tau_i \psi(\tau_i d_i) z_i z_i^* \) where \( d_1, \ldots, d_n \geq 0 \) are the only solutions to the \( n \) equations:

\[
d_j = \frac{1}{N} z_j^* \left( \frac{1}{n} \sum_{i \neq j} \tau_i \psi(\tau_i d_i) z_i z_i^* \right)^{-1} z_j, \quad 1 \leq j \leq n. \tag{3}
\]

### III-B. Hint on the main result

Since we have assumed that \( \hat{C}_N \) is well defined as the unique solution to (1), the \( d_i \) above are also unique and well defined (let us say, at least for all large \( n \) a.s.).

From the discussion in Section III-A, we may expect the terms \( d_i \) to be all close to \( \frac{1}{n} \text{tr} \hat{C}_N^{-1} \) for \( N, n \) large enough. From random matrix intuition, we may also expect \( \frac{1}{n} \text{tr} \hat{C}_N^{-1} \) to have a deterministic equivalent \( \gamma_N \), i.e. there should exist a deterministic sequence \( \{\gamma_N\}_{N=1}^\infty \) such that \( \frac{1}{n} \text{tr} \hat{C}_N^{-1} \rightarrow \gamma_N \) a.s. Let us say that all this is true. Since \( \frac{1}{n} \text{tr} \hat{C}_N^{-1} \) is the Stieltjes transform \( \frac{1}{n} \text{tr} (\hat{C}_N - z I_N)^{-1} \) of the empirical spectral distribution of \( \hat{C}_N \) at point \( z = 0 \), and since \( \hat{C}_N \) is expected to be close to \( \frac{1}{n} \sum_{i=1}^n \tau_i \psi(\tau_i \gamma_N) z_i z_i^* \) with now \( \psi(\tau_i \gamma_N) \) independent of \( z_1, \ldots, z_n \), from classical random matrix works, e.g. [7], we would expect that one such \( \gamma_N \) be given by (recall that \( \hat{C}_N = I_N \))

\[
\gamma_N = \left( \frac{1}{n} \sum_{i=1}^n \psi(\tau_i \gamma_N) \right)^{-1} \frac{\psi(\tau_i \gamma_N)}{1 + c_N \psi(\tau_i \gamma_N) \gamma_N}
\]

if this fixed-point equation makes sense at all. This can be equivalently written as

\[
1 = \frac{1}{n} \sum_{i=1}^n \frac{\psi(\tau_i \gamma_N)}{1 + c_N \psi(\tau_i \gamma_N) \gamma_N}, \tag{4}
\]

We in fact prove in [5] that such a positive \( \gamma_N \) is well defined, unique, and satisfies \( \max_{1 \leq i \leq n} |d_i - \gamma_N| \overset{a.s.}{\to} 0 \) (under correct assumptions). Showing this result is the main difficulty of the proof and is in particular this part of the proof that fully exploits Assumption 3. This convergence along with classical random matrix arguments shall then ensure that for all large \( n \) a.s.

\[
\|\hat{C}_N - \hat{S}_N\| \overset{a.s.}{\to} 0
\]

where

\[
\hat{S}_N = \frac{1}{n} \sum_{i=1}^n \psi(\tau_i \gamma_N) \tau_i z_i z_i^*
\]

with \( \gamma_N \) the unique positive solution to (4). It is then immediate under Assumption 2–3 to see that the result holds true also for \( C_N \neq I_N \). This therefore gives the expected result.

### IV. CONCLUSION

We have provided a large dimensional analysis for robust estimators of scatter matrices of the Maronna-type for elliptical samples. We specifically showed that, under mild assumptions, the Maronna estimator behaves similar to a classical sample covariance matrix model. This opens new roads in the analysis of signal processing methods based on robust scatter matrix estimation. In a similar manner as in [3, Theorem 6], it is believed that second order statistics for well behaved functionals of \( \hat{C}_N \) can be further analyzed, which would provide more information on the asymptotic fluctuations of \( \hat{C}_N - \hat{S}_N \).

### V. REFERENCES


