ROBUST M-ESTIMATOR OF SCATTER FOR LARGE ELLIPTICAL SAMPLES

Romain Couillet[1] *and Frédéric Pascal*[2]

[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, under the assumption of similar population and sample sizes. For instance, new source detection schemes have been proposed based on the works on the extreme eigenvalues of large Wishart matrices. New subspace methods in large array processing have also been derived that outperform the original MUSIC algorithm by exploiting statistical inference methods on large random matrices. Most of these signal processing methods fundamentally rely on the structure of the sample covariance matrix $\frac{1}{n}\sum_{i=1}^{n} x_i x_i^*$ formed from independent or linearly dependent zero mean samples $x_1, \ldots, x_n \in \mathbb{C}^N$, which are well understood objects, see e.g. [1]. Signal processing covariance estimation methods however go beyond sample covariance matrices. In particular robust scatter M-estimation techniques are used to better approximate population covariance (or scatter) matrices whenever (i) the distribution of the x_i's is heavier-tailed than Gaussian or (ii) the x_i's contain outliers [2], [3].

Robust scatter matrix estimators are often more complex than sample covariance matrices which makes them inappropriate to standard random matrix analysis. In this work, we specifically consider robust scatter estimator of the Maronna type, proposed in [3]. In the regime where $n \to \infty$ and N fixed, [3] shows that under some conditions the estimator is well-defined as the unique solution of a fixed-point equation and that it almost surely (a.s.) converges to a deterministic matrix. We instead treat here the case where $N, n \to \infty$ and N/n remains away from zero, following the same approach as in [4] but for the more involved case of elliptical vectors x_i.

Couillet's work is supported by ANR-12-MONU-OOO3 DIONISOS project.

The study of robust scatter estimators in the large N, n regime has important consequences in understanding many signal processing algorithms exploiting these estimators. It also allows one 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 $x_i = \sqrt{\tau_i} A_N y_i$, where $\tau_1, \ldots, \tau_n \in \mathbb{R}^+$ and $y_1, \ldots, y_n \in \mathbb{C}^{\bar{N}}$ are random and $A_N \in \mathbb{C}^{N \times \bar{N}}$ is deterministic. We denote $c_N \triangleq N/n$ and $\bar{c}_N \triangleq \bar{N}/N$ and shall consider the following growth regime.

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

$$0 < c_- < \liminf_n c_N \leq \limsup_n c_N < c_+ < 1.$$

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

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

where u satisfies the following properties:

(i) $u : [0, \infty) \to (0, \infty)$ is nonnegative continuous and non-increasing
(ii) $\phi : x \mapsto xu(x)$ is increasing and bounded with $\lim_{x\to\infty} \phi(x) \triangleq \phi_\infty > 1$
(iii) $\phi_\infty < c_+^{-1}$.

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 $x_i = \sqrt{\tau_i} A_N y_i$ satisfy:

1) $\nu_n = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\delta}_{\tau_i}$ is such that $\int x \nu_n(dx) \xrightarrow{\text{a.s.}} 1$
2) there exist $\varepsilon < 1 - \phi_\infty^{-1} < 1 - c_+$ and $m > 0$ such that, for all large n a.s. $\nu_n([0, m)) < \varepsilon$
3) defining $C_N \triangleq A_N A_N^*$, $C_N \succ 0$ and $\limsup_N \|C_N\| < \infty$
4) $y_1, \ldots, y_n \in \mathbb{C}^{\bar{N}}$ are independent unitarily invariant complex zero-mean vectors with, for each i, $\|y_i\|^2 = \bar{N}$, and are independent of $\tau_1, \ldots, \tau_n$.

These conditions are met in particular if the τ_i are independent and identically distributed (i.i.d.) with unit mean distribution ν (then $\int x \nu_n(dx) \xrightarrow{\text{a.s.}} 1$ by the strong law of large numbers) such that $\nu(\{0\}) = 0$. If in addition $N = \bar{N}$, 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 τ_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.

$$\limsup_{t\to\infty} \frac{\limsup_n \nu_n((t,\infty))}{\phi(at)-\phi(bt)} = 0.$$

Assumption 3 controls the relative speed of the tail of ν_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 ϕ is increasing.
- For $u(t) = (1+\alpha)/(\alpha+t)$ for some $\alpha > 0$ and τ_i i.i.d. with distribution ν, by Markov inequality, it suffices that $\int x^{1+\varepsilon}\nu(dx) < \infty$ for some $\varepsilon > 0$.

This article provides two results: (i) existence and uniqueness of $\hat{C}_N$ as a solution to (1) is shown (Theorem 1) and (ii) the limiting spectral behavior of $\hat{C}_N$ as $N, n \to \infty$ is derived (Theorem 2). Theorem 1 somehow extends [3, Theorem 1] which is insufficient for our current needs as it imposes the strong constraint $\phi_\infty > 1/(1-c_-)$. As for Theorem 2, it is very different from [3, Theorem 5] which proves the convergence of $\hat{C}_N$ as $n \to \infty$; here instead we show that there exists a matrix $\hat{S}_N$ such that $\|\hat{C}_n - \hat{S}_N\| \xrightarrow{\text{a.s.}} 0$ in spectral norm, where $\hat{S}_N$ follows a standard random matrix model studied e.g. in [7].

II. Main Results

The first result ensures the existence and uniqueness of a solution $\hat{C}_N$ to (1) for n large enough.

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

$$\hat{C}_N = \lim_{\substack{t\to\infty \\ t\in\mathbb{N}}} Z^{(t)}$$

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

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

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 2 (Asymptotic Behavior): Let Assumptions 1–3 hold, and let $\hat{C}_N$ be given by Theorem 1 when uniquely defined as the solution of (1) or chosen arbitrarily if not. Then

$$\left\|\hat{C}_N - \hat{S}_N\right\| \xrightarrow{\text{a.s.}} 0$$

where

$$\hat{S}_N \triangleq \frac{1}{n}\sum_{i=1}^n v(\tau_i\gamma_N)x_ix_i^*$$

and γ_N is the unique positive solution of the equation in γ

$$1 = \frac{1}{n}\sum_{i=1}^n \frac{\psi(\tau_i\gamma)}{1+c_N\psi(\tau_i\gamma)}$$

with the functions $v : x \mapsto (u\circ g^{-1})(x)$, $\psi : x \mapsto xv(x)$, and $g : \mathbb{R}^+ \to \mathbb{R}^+,\ x \mapsto x/(1-c_N\phi(x))$.

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 τ_i's are i.i.d., leading to elliptical distributions for x_i, for which $\hat{C}_N$ has an (almost sure) limiting eigenvalue spectrum.

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

$$\left\|\hat{C}_N - \frac{1}{n}\sum_{i=1}^n v(\tau_i\gamma^\infty)x_ix_i^*\right\| \xrightarrow{\text{a.s.}} 0$$

where γ^∞ is the unique positive solution to the equation in γ

$$1 = \int \frac{\psi_c(t\gamma)}{1+c\psi_c(t\gamma)}\nu(dt)$$

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

$$\frac{1}{n}\sum_{i=1}^n \boldsymbol{\delta}_{\lambda_i(\hat{C}_N)} \xrightarrow{\text{a.s.}} \mu$$

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

$$m(z) = -\frac{1}{z}\int \frac{1}{1+\tilde{\delta}(z)t}\nu^C(dt)$$

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

$$\tilde{\delta} = -\frac{1}{z}\int \frac{\psi_c(t\gamma^\infty)}{\gamma^\infty + \psi_c(t\gamma^\infty)\delta}\nu(dt)$$
$$\delta = -\frac{c}{z}\int \frac{t}{1+t\tilde{\delta}}\nu^C(dt).$$

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

$$\left|\left\{\lambda_i(\hat{C}_N)\right\}_{i=1}^N \cap \mathcal{A}\right| \xrightarrow{\text{a.s.}} 0.$$

Figure 1 depicts the empirical histogram of the eigenvalues of $\hat{C}_N$, for $N = 500$ and $n = 2500$ with $u(t) = (1+\alpha)/(t+\alpha)$,[1] $\alpha = 0.1$, $C_N = \mathrm{diag}(I_{125}, 3I_{125}, 10I_{250})$, and $\tau_1, \ldots, \tau_n$ i.i.d. with $\Gamma(.5, 2)$ distribution. In thick line is also depicted the density of μ in Corollary 1 which shows an accurate match to the empirical spectrum. As a comparison, Figure 2 shows the empirical histogram

[1]This function $u(t)$ is often met in robust statistics as it is such that $\hat{C}_N$ corresponds to the maximum-likelihood estimate of the scale parameter of independent and identically distributed multivariate Student-t vectors.

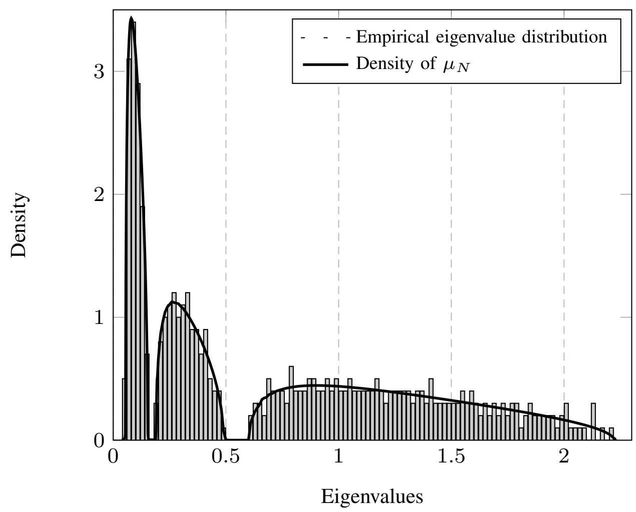

Fig. 1. Histogram of the eigenvalues of $\hat{C}_N$ for $n = 2500$, $N = 500$, $C_N = \operatorname{diag}(I_{125}, 3I_{125}, 10I_{250})$, τ_1 with $\Gamma(.5, 2)$-distribution.

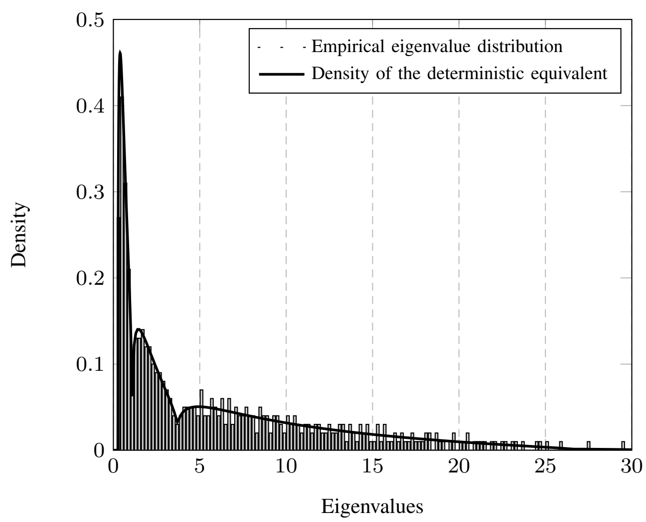

Fig. 2. Histogram of the eigenvalues of $\frac{1}{n}\sum_{i=1}^n x_i x_i^*$ for $n = 2500$, $N = 500$, $C_N = \operatorname{diag}(I_{125}, 3I_{125}, 10I_{250})$, τ_1 with $\Gamma(.5, 2)$-distribution.

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 τ_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 $C_N^{-\frac{1}{2}}\hat{C}_N C_N^{-\frac{1}{2}}$ 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 $u(\frac{1}{N}x_i^*\hat{C}_N^{-1}x_i)x_i x_i^*$ 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 $x_i = \sqrt{\tau_i}A_N y_i \triangleq \sqrt{\tau_i}z_i$ 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 $\hat{C}_{(i)} \triangleq \hat{C}_N - \frac{1}{n}u(\frac{1}{N}x_i^*\hat{C}_N^{-1}x_i)x_i x_i^*$. Note that $\hat{C}_{(i)}$ depends on x_i only through the terms $u(\frac{1}{N}x_j^*\hat{C}_N^{-1}x_j)$, $j \neq i$, 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 $\hat{C}_{(i)}$, along with classical random matrix theory considerations, suggests that $\frac{1}{N}z_i^*\hat{C}_{(i)}^{-1}z_i$, $i = 1, \ldots, n$, are all well approximated by $\frac{1}{N}\operatorname{tr}\hat{C}_N^{-1}$ (see e.g. [7, Lemma 3.1]).

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

$$\frac{1}{N}z_i^*Z^{-1}z_i = \frac{\frac{1}{N}z_i^*Z_{(i)}^{-1}z_i}{1 + \tau_i u\left(\tau_i\frac{1}{N}z_i^*Z^{-1}z_i\right)\frac{1}{n}z_i^*Z_{(i)}^{-1}z_i}$$

so that

$$\frac{1}{N}z_i^*Z_{(i)}^{-1}z_i\left(1 - c_N\tau_i u\left(\tau_i\frac{1}{N}z_i^*Z^{-1}z_i\right)\frac{1}{N}z_i^*Z^{-1}z_i\right) = \frac{1}{N}z_i^*Z^{-1}z_i$$

which, by the definition of ϕ, is

$$\frac{1}{N}z_i^*Z_{(i)}^{-1}z_i\left(1 - c_N\phi\left(\tau_i\frac{1}{N}z_i^*Z^{-1}z_i\right)\right) = \frac{1}{N}z_i^*Z^{-1}z_i.$$

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

$$\frac{1}{N}z_i^*Z_{(i)}^{-1}z_i = \frac{\frac{1}{N}z_i^*Z^{-1}z_i}{1 - c_N\phi\left(\tau_i\frac{1}{N}z_i^*Z^{-1}z_i\right)}. \tag{2}$$

Now, since ϕ 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),

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

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:

$$\psi(x) = xu(g^{-1}(x)) = g(g^{-1}(x))u(g^{-1}(x)) = \frac{\phi(g^{-1}(x))}{1 - c_N\phi(g^{-1}(x))}$$

which is increasing, nonnegative, with limit $\psi_\infty^N \triangleq \phi_\infty/(1 - c_N\phi_\infty)$ as $x \to \infty$. Hence, v and ψ keep the same properties as u and ϕ, respectively.

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

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

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

$$d_j = \frac{1}{N} z_j^* \left(\frac{1}{n}\sum_{i\neq j} \tau_i v\left(\tau_i d_i\right) z_i z_i^*\right)^{-1} z_j,\ 1 \leq j \leq n. \qquad (3)$$

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

$$\gamma_N = \left(\frac{1}{n}\sum_{i=1}^n \frac{\tau_i v(\tau_i\gamma_N)}{1 + c_N\tau_i v(\tau_i\gamma_N)\gamma_N}\right)^{-1}$$

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

$$1 = \frac{1}{n}\sum_{i=1}^n \frac{\psi(\tau_i\gamma_N)}{1 + c_N\psi(\tau_i\gamma_N)}. \qquad (4)$$

We in fact prove in [5] that such a positive γ_N is well defined, unique, and satisfies $\max_{1\leq i\leq n}|d_i - \gamma_N| \xrightarrow{\text{a.s.}} 0$ (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.

$$\left\|\hat{C}_N - \hat{S}_N\right\| \xrightarrow{\text{a.s.}} 0$$

where

$$\hat{S}_N = \frac{1}{n}\sum_{i=1}^n v\left(\tau_i\gamma_N\right)\tau_i z_i z_i^*$$

with γ_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

[1] J. W. Silverstein and S. Choi, "Analysis of the limiting spectral distribution of large dimensional random matrices," *Journal of Multivariate Analysis*, vol. 54, no. 2, pp. 295–309, 1995.

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

[3] R. A. Maronna, "Robust M-estimators of multivariate location and scatter," *The annals of statistics*, pp. 51–67, 1976.

[4] R. Couillet, F. Pascal, and J. W. Silverstein, "Robust M-estimation for Array Processing: a random matrix approach," *IEEE Transactions on Information Theory*, 2013. [Online]. Available: http://arxiv.org/abs/1204.5320

[5] ——, "The random matrix regime of Maronna's M-estimator with elliptically distributed samples," *Journal of Multivariate Analysis*, 2013. [Online]. Available: http://arxiv.org/abs/1311.7034

[6] E. Ollila, D. Tyler, V. Koivunen, and H. V. Poor, "Complex elliptically symmetric distributions: survey, new results and applications," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5597–5625, 2012.

[7] J. W. Silverstein and Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *Journal of Multivariate Analysis*, vol. 54, no. 2, pp. 175–192, 1995.

[8] D. Paul and J. W. Silverstein, "No eigenvalues outside the support of the limiting empirical spectral distribution of a separable covariance matrix," *Journal of Multivariate Analysis*, vol. 100, no. 1, pp. 37–57, 2009.

[9] R. Couillet and W. Hachem, "Analysis of the limit spectral measure of large random matrices of the separable covariance type," *Journal of Multivariate Analysis*, 2013.

[10] Z. D. Bai and J. W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices," *The Annals of Probability*, vol. 26, no. 1, pp. 316–345, 1998.