LARGE DIMENSIONAL ANALYSIS OF MARONNA'S M-ESTIMATOR WITH OUTLIERS

David Morales-Jimenez**, Romain Couillet*†*, Matthew R. McKay**

* Hong Kong University of Science and Technology, ECE Department
† Supélec, Telecommunication Department

ABSTRACT

Building on recent results in the random matrix analysis of robust estimators of scatter, we show that a certain class of such estimators obtained from samples containing outliers behaves similar to a well-known random matrix model in the limiting regime where both the population and sample sizes grow to infinity at the same speed. This result allows us to understand the structure of such estimators when a certain fraction of the samples is corrupted by outliers and, in particular, to derive their asymptotic eigenvalue distributions. This analysis is a first step towards an improved usage of robust estimation methods under the presence of outliers when the number of independent observations is not too large compared to the size of the population.

Index Terms— Robust estimation, outliers, random matrix theory.

1. INTRODUCTION

The growing momentum of big data applications along with the recent advances in large random matrix theory have raised a great interest for problems in statistical inference and signal processing under the assumption of similar population and sample sizes. New source detection schemes have in particular been proposed (see, e.g., [1, 2]) based on the works on the extreme and isolated eigenvalues of large sample covariance matrices. New subspace methods in large array processing have also been derived that outperform traditional algorithms by exploiting statistical inference methods on large random matrices (see, e.g., [3]). Most of these signal processing methods fundamentally rely on the structure of the sample covariance matrix (SCM) $\frac{1}{n}\sum_{i=1}^{n}\mathbf{y}_i\mathbf{y}_i^\dagger$ formed from independent or linearly dependent samples $\mathbf{y}_1,\ldots,\mathbf{y}_n\in\mathbb{C}^N$, which are by now well understood objects. However, there are applications where, even when $n \gg N$, the SCM fails to provide a good estimate of the population covariance, hence the need for more robust methods. Robust scatter M-estimation techniques are precisely used to better approximate population covariance (or scatter) matrices whenever (i) the distribution of the $\mathbf{y}_i$'s is heavier-tailed than Gaussian (e.g., elliptical data) or (ii) the $\mathbf{y}_i$'s contain outliers [4, 5].

Given the usually quite involved implicit expression of these robust estimators, it is not obvious to study their behavior but recent works have provided some first answers for Gaussian or elliptical i.i.d. data, see e.g., [6] for Maronna's M-estimator, [7] for Tyler's estimator, or [8] for a regularized adaptation of Tyler's estimator. Robust regressors have also been investigated in [9]. These works entailed the design of improved detectors and estimators accounting for the impulsiveness of data, see e.g., [10] for an application to portfolio optimization in finance, [11] for subspace estimators in array processing, or [12] for generalized likelihood ratio tests under elliptical noise data.

Implicit to all of these works is the assumption of an outlier-free model, in which samples are drawn i.i.d. from an analytically-tractable distribution (i.e., Gaussian or elliptical typically). Thus, very little is known concerning the impact of outliers on the robust estimators, despite the fact that these estimators were originally designed by Huber [4] for the very purpose of mitigating outliers. In this work, we consider robust scatter estimators of the Maronna type (defined below) in the double asymptotic regime where $N, n \to \infty$ with $N/n \to c \in (0,1)$, and characterize their behavior when the set of data samples contains deterministic or random outliers. Our main finding is to show that, under mild assumptions, the estimator behaves for large N, n as a weighted version of the SCM with different weights for the model-fitting samples (usually considered in majority) and for the outlying samples. An analysis of these weights in the limiting case of few outliers reveals the following messages: (i) the robust estimators tend to reduce the importance of outliers with strong norm, thus precluding the problem of arbitrary large bias, and (ii-a) strong correlation in the model-fitting data induces in general stronger outlier rejection but (ii-b) in a worst case scenario, the impact of outliers may be amplified, thus necessitating a careful choice of estimators within the Maronna class, and in particular estimators originally proposed by Huber.

2. PROBLEM STATEMENT

Consider $\mathbf{Y}\in\mathbb{C}^{N\times n}$ to be a matrix composed in columns of n stacked N-dimensional data vectors, with $\varepsilon_n n$ of these

The work of D. Morales-Jimenez and M. R. McKay was supported by the Hong Kong Research Grants Council under grant number 16206914. Couillet's work is supported by the ERC MORE EC-120133.

samples being outliers and assume, without loss of generality, that the columns of $\mathbf{Y}$ are reordered as

$$\mathbf{Y} = \left[\mathbf{y}_1, \ldots, \mathbf{y}_{(1-\varepsilon_n)n}, \mathbf{a}_1, \ldots . \mathbf{a}_{\varepsilon_n n}\right] \tag{1}$$

Here $\mathbf{y}_1, \ldots, \mathbf{y}_{(1-\varepsilon_n)n} \in \mathbb{C}^N$ are random model-fitting samples with $\mathbf{y}_i = \mathbf{C}_N^{1/2}\mathbf{x}_i$, where $\mathbf{C}_N \in \mathbb{C}^{N\times N}$ is deterministic and $\mathbf{x}_1, \ldots, \mathbf{x}_{(1-\varepsilon_n)n}$ are i.i.d. random with i.i.d. zero mean and unit variance entries,[1] whereas $\mathbf{a}_1, \ldots, \mathbf{a}_{\varepsilon_n n} \in \mathbb{C}^N$ are arbitrary outlying samples. We further denote $c_n \triangleq N/n$ and shall consider the following growth regime.

Assumption 1 *As $N, n \to \infty$, $c_n \to c$ and $\varepsilon_n \to \varepsilon \in [0, 1)$ with $0 < c < 1 - \varepsilon$.*

Assumption 2 *For each N, $\mathbf{C}_N \succ 0$, $\limsup_N \|\mathbf{C}_N\| < \infty$ and $0 < \min_i \liminf_N \|\mathbf{a}_i\| \leq \max_i \limsup_N \|\mathbf{a}_i\| < \infty$.*

Define Maronna's M-estimator $\hat{\mathbf{C}}_N$ as the (almost surely unique) solution to the equation in $\mathbf{Z}$ [13]

$$\begin{aligned}\mathbf{Z} = \frac{1}{n}\sum_{i=1}^{(1-\varepsilon_n)n} u\left(\frac{1}{N}\mathbf{y}_i^\dagger \mathbf{Z}^{-1}\mathbf{y}_i\right)\mathbf{y}_i\mathbf{y}_i^\dagger \\ + \frac{1}{n}\sum_{i=1}^{\varepsilon_n n} u\left(\frac{1}{N}\mathbf{a}_i^\dagger \mathbf{Z}^{-1}\mathbf{a}_i\right)\mathbf{a}_i\mathbf{a}_i^\dagger\end{aligned} \tag{2}$$

where u is defined on $[0, \infty)$, nonnegative, continuous and non-increasing, and such that $\phi(x) = xu(x)$ is increasing and bounded with $\lim_{x\to\infty}\phi(x) \triangleq \phi_\infty$, and $1 < \phi_\infty < c^{-1}$.

Following the works [6, 14], our main objective is to find a large N, n random matrix equivalent for $\hat{\mathbf{C}}_N$ which is more tractable and prone to analysis.

3. MAIN RESULTS

We are now in position to introduce our main result, a proof sketch of which is provided in Appendix A. A complete proof will be provided in an extended version of the present article.

Theorem 1 (Asymptotic Behavior) *Let Assumptions 1-2 hold and let $\hat{\mathbf{C}}_N$ be the a.s. unique solution to (2). Then, as $N, n \to \infty$,*

$$\left\|\hat{\mathbf{C}}_N - \hat{\mathbf{S}}_N\right\| \xrightarrow{a.s.} 0 \tag{3}$$

where

$$\hat{\mathbf{S}}_N \triangleq \frac{1}{n}\sum_{i=1}^{(1-\varepsilon_n)n} v\left(\gamma_n\right)\mathbf{y}_i\mathbf{y}_i^\dagger + \frac{1}{n}\sum_{i=1}^{\varepsilon_n n} v\left(\alpha_{i,n}\right)\mathbf{a}_i\mathbf{a}_i^\dagger \tag{4}$$

with γ_n and $\alpha_{1,n}, \ldots, \alpha_{\varepsilon_n n,n}$ the unique positive solutions to the system of $\varepsilon_n n + 1$ equations ($i = 1, \ldots, \varepsilon_n n$)

$$\begin{aligned}\gamma_n &= \frac{1}{N}\operatorname{tr}\mathbf{C}_N\left(\frac{(1-\varepsilon)v(\gamma_n)}{1+cv(\gamma_n)\gamma_n}\mathbf{C}_N + \frac{1}{n}\sum_{j=1}^{\varepsilon_n n} v\left(\alpha_{j,n}\right)\mathbf{a}_j\mathbf{a}_j^\dagger\right)^{-1} \\ \alpha_{i,n} &= \frac{1}{N}\mathbf{a}_i^\dagger\left(\frac{(1-\varepsilon)v(\gamma_n)}{1+cv(\gamma_n)\gamma_n}\mathbf{C}_N + \frac{1}{n}\sum_{j\neq i}^{\varepsilon_n n} v\left(\alpha_{j,n}\right)\mathbf{a}_j\mathbf{a}_j^\dagger\right)^{-1}\mathbf{a}_i\end{aligned} \tag{5}$$

and $v(x)$ a non-increasing function with $v(0) = u(0)$ and $\lim_{x\to\infty} v(x) = 0$, defined precisely by $v(x) = u\left(g^{-1}(x)\right)$, $g(x) = x/(1 - c\phi(x))$.

This result characterizes the spectral behavior of $\hat{\mathbf{C}}_N$ for large N, n. In particular, a corollary to Theorem 1 is that $\max_i |\lambda_i(\hat{\mathbf{C}}_N) - \lambda_i(\hat{\mathbf{S}}_N)| \xrightarrow{\text{a.s.}} 0$, where $\lambda_i(\mathbf{X})$ are the ordered eigenvalues of the Hermitian matrix $\mathbf{X}$.

Observe that the approximation matrix $\hat{\mathbf{S}}_N$ consists of two terms: a normalized SCM and a weighted sum of the outlier outer products. These weights allow for an *automated balancing between model-fitting data and outliers*. To get some insight on the properties of $\hat{\mathbf{C}}_N$ induced by these weights, let us consider the single-outlier case where $\varepsilon_n = 1/n \to 0$. Regarding the model-data weights, we obtain by a rank-one perturbation argument that $\gamma_n \to \gamma$, where γ is the solution to $\gamma = (1 + cv(\gamma)\gamma)/v(\gamma)$. Moreover, using the definition of v, it can be seen that $\gamma = \phi^{-1}(1)/(1 - c)$, the result originally proved in [14] in the absence of outliers. As for the outlier's weight, we have $|\alpha_{1,n} - \bar{\alpha}_{1,n}| \to 0$ with

$$\bar{\alpha}_{1,n} = \frac{\phi^{-1}(1)}{1-c}\frac{1}{N}\mathbf{a}_1^\dagger\mathbf{C}_N^{-1}\mathbf{a}_1.$$

Hence, provided that $\liminf_n \frac{1}{N}\mathbf{a}_1^\dagger\mathbf{C}_N^{-1}\mathbf{a}_1 > 1$, $v(\alpha_{1,n}) \leq v(\gamma)$ for all large n, and thus a larger value of $\frac{1}{N}\mathbf{a}_1^\dagger\mathbf{C}_N^{-1}\mathbf{a}_1$ will generally lead to a larger attenuation of the outlier $\mathbf{a}_1$. However, if $\limsup_n \frac{1}{N}\mathbf{a}_1^\dagger\mathbf{C}_N^{-1}\mathbf{a}_1 < 1$, then the opposite occurs and, in fact, the effect of the outlier may be amplified. As such:

- to avoid boosting the effect of outliers, $v(x)$ should be set to a constant for all $x \leq \frac{\phi^{-1}(1)}{1-c}$, or equivalently $u(x)$ is constant for $x \leq \phi^{-1}(1)$. A particular example of such a choice is $u(x) = \min\{1, (1+t)/(t+x)\}$ for some $t > 0$, which is (almost) the original Huber estimator from [4].[2]

- for $\mathbf{C}_N$ close to the identity matrix, only the norm of $\mathbf{a}_1$ dictates its relative impact. Departing from the identity, a good rejection to outliers is expected if $\mathbf{a}_1$ is not

[1] We could have considered samples with elliptical-like distributions instead but, in order not to confuse messages, we only characterize here the behavior of Maronna's estimator for light-tailed data versus outliers.

[2] Huber considered a $t = 0$ and a slightly more general form, but $t = 0$ is usually not enough to ensure uniqueness of the solution to (2).

aligned to the dominant eigenvectors of $\mathbf{C}_N$ (i.e., to the least dominant eigenvectors of $\mathbf{C}_N^{-1}$). Contrarily, if $\mathbf{a}_1$ were aligned to a dominant eigenvector of $\mathbf{C}_N$, outlier rejection would be compromised.

Returning to the multi-outlier case, other considerations can be made. If $\mathbf{a}_1 = \ldots = \mathbf{a}_{\varepsilon_n n}$, we can see (through similar arguments as above) that, as $\varepsilon_n n$ grows, the outlier-rejection gain brought by the (possibly large) quadratic form $\frac{1}{N}\mathbf{a}_1^\dagger \mathbf{C}_N^{-1}\mathbf{a}_1$ is quickly overrun so that, if $\varepsilon > 0$ and $\limsup_N \frac{1}{N}\mathbf{a}_1^\dagger \mathbf{C}_N^{-1}\mathbf{a}_1 < \infty$, $\alpha_{1,n} \to 0$ as $n \to \infty$ and the outliers are not rejected but boosted instead.

Of interest also is the case where the $\mathbf{a}_i$'s are random i.i.d., not following the same distribution as $\mathbf{y}_i$. This gives in particular the following corollary.

Corollary 1 (Random Outliers) *Let Assumptions 1-2 hold and let $\mathbf{a}_1, \ldots, \mathbf{a}_{\varepsilon_n n}$ be random with $\mathbf{a}_i = \mathbf{D}_N^{1/2}\hat{\mathbf{x}}_i$, where $\mathbf{D}_N \in \mathbb{C}^{N\times N}$ is deterministic and $\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_{\varepsilon_n n}$ are i.i.d. random with i.i.d. zero mean and unit variance entries. Let us further assume that, for each N, $\mathbf{D}_N \succ 0$ and $\limsup_N \|\mathbf{D}_N\| < \infty$. Then, as $N, n \to \infty$,*

$$\left\| \hat{\mathbf{C}}_N - \hat{\mathbf{S}}_N^{\mathrm{rnd}} \right\| \xrightarrow{a.s.} 0 \tag{6}$$

where

$$\hat{\mathbf{S}}_N^{\mathrm{rnd}} \triangleq \frac{1}{n}\sum_{i=1}^{(1-\varepsilon_n)n} v\left(\tilde{\gamma}_n\right)\mathbf{y}_i\mathbf{y}_i^\dagger + \frac{1}{n}\sum_{i=1}^{\varepsilon_n n} v\left(\tilde{\alpha}_n\right)\mathbf{a}_i\mathbf{a}_i^\dagger, \tag{7}$$

with $\tilde{\gamma}_n$ and $\tilde{\alpha}_n$ the unique positive solutions to

$$\tilde{\gamma}_n = \frac{1}{N}\operatorname{tr}\mathbf{C}_N\left(\frac{(1-\varepsilon)v(\tilde{\gamma}_n)}{1+cv(\tilde{\gamma}_n)\tilde{\gamma}_n}\mathbf{C}_N + \frac{\varepsilon v(\tilde{\alpha}_n)}{1+cv(\tilde{\alpha}_n)\tilde{\alpha}_n}\mathbf{D}_N\right)^{-1}$$
$$\tilde{\alpha}_n = \frac{1}{N}\operatorname{tr}\mathbf{D}_N\left(\frac{(1-\varepsilon)v(\tilde{\gamma}_n)}{1+cv(\tilde{\gamma}_n)\tilde{\gamma}_n}\mathbf{C}_N + \frac{\varepsilon v(\tilde{\alpha}_n)}{1+cv(\tilde{\alpha}_n)\tilde{\alpha}_n}\mathbf{D}_N\right)^{-1}.$$

In this scenario, $\hat{\mathbf{C}}_N$ is equivalent to a weighted sum of two sample covariance matrices, one for the model-fitting data and the other for the outlier data. Again, it is interesting to study the regime where $\varepsilon = 0$. In this regime, we get that $\tilde{\gamma}_n = \gamma$ given as above, i.e., $\gamma = \phi^{-1}(1)/(1-c)$, and

$$\tilde{\alpha}_n = \frac{\phi^{-1}(1)}{1-c}\frac{1}{N}\operatorname{tr}\mathbf{D}_N\mathbf{C}_N^{-1} \triangleq \bar{\alpha}_n.$$

The factor of importance is now the trace $\frac{1}{N}\operatorname{tr}\mathbf{D}_N\mathbf{C}_N^{-1}$. Similar to before, the larger this value, the stronger the rejection of the outlier samples; whilst if this value becomes small, the outliers may indeed be amplified. Note that, for $\mathbf{D}_N$ and $\mathbf{C}_N$ of similar trace, it is of key importance that $\mathbf{C}_N$ be as distinct from $\mathbf{I}_N$ as possible for outlier rejection to be possible. In addition, when seen as functions of ε, $\tilde{\gamma}_n(\varepsilon) \to \gamma$ and $\tilde{\alpha}_n(\varepsilon) \to \bar{\alpha}_n$ continuously with $\varepsilon \to 0$, so that the predicted behavior for $\varepsilon = 0$ is a good approximation of the behavior for all small $\varepsilon > 0$.

4. NUMERICAL DISCUSSION

We now provide simulation results that shed further light on the conclusions drawn from Theorem 1 and Corollary 1.

Let us place ourselves first under the setting of Theorem 1. Taking $N = 100$, $n = 500$, we assume $[\mathbf{C}_N]_{ij} = .9^{|i-j|}$ and let $\varepsilon_n n = 2$ with $\mathbf{a}_1 = \mathbf{1}$, the vector of all-ones, and $\mathbf{a}_2$ a steering vector at 30°, i.e., $[\mathbf{a}_2]_k = \exp(\pi\imath k)$, $\imath = \sqrt{-1}$. In this setting, $\frac{1}{N}\mathbf{a}_1^\dagger\mathbf{C}_N^{-1}\mathbf{a}_1 \simeq 0.06$ while $\frac{1}{N}\mathbf{a}_2^\dagger\mathbf{C}_N^{-1}\mathbf{a}_2 \simeq 19$. We compare the results obtained for $u_1(x) = (1+t)/(t+x)$ against $u_2(x) = \min\{1, (1+t)/(t+x)\}$ for $t = .1$ (labeling v_1, v_2 accordingly).

Numerically, we obtain

$$v_1(\gamma_n) \simeq .992,\ v_1(\alpha_{1,n}) \simeq 6.42,\ v_1(\alpha_{2,n}) \simeq .006,$$

indicating strong attenuation of the second outlier, while strong enhancement of the first. Comparatively,

$$v_2(\gamma_n) \simeq .984,\ v_2(\alpha_{1,n}) = 1.00,\ v_2(\alpha_{2,n}) \simeq .006.$$

Thus Huber's type estimator u_2 prevents, as it should, the outlier $\mathbf{a}_1$ to be enhanced. This however induces a slight loss in the closeness of $v_2(\gamma_n)$ to one.

We now place ourselves under the setting of Corollary 1 with $[\mathbf{C}_N]_{ij} = .9^{|i-j|}$, $N = 100$, $c = .2$, while $\mathbf{D}_N = \mathbf{I}_N$, $\varepsilon = 0.05$, i.e., a 5% data pollution by outliers, and $u(x) = u_2(x)$. We wish to compare the eigenvalue distribution of the SCM $\frac{1}{n}\mathbf{Y}\mathbf{Y}^\dagger$ and that of $\hat{\mathbf{C}}_N$ with the outlier-free SCM $\frac{1}{n}\sum_{i=1}^{(1-\varepsilon_n)n}\mathbf{y}_i\mathbf{y}_i^\dagger$ in order to see the robustness of SCM and $\hat{\mathbf{C}}_N$ against outliers. To avoid lengthy and imprecise Monte Carlo simulations, we instead compare the theoretical limiting eigenvalue distributions[3] as $N, n \to \infty$ but for the limiting eigenvalue distribution of $\mathbf{C}_N$ maintained to that with $N = 100$; thus, we precisely compare the eigenvalue densities of the so-called deterministic equivalents for the various random matrices under study. This is depicted in Fig. 1, which shows a tight match between $\hat{\mathbf{C}}_N$ and the target distribution (i.e., the outlier-free SCM). The SCM, on the other hand, is strongly distorted as a consequence of the outliers.

5. CONCLUSION

We have provided a first large dimensional characterization for robust covariance estimators of the Maronna-type when the data set contains outliers. We specifically showed that, under mild assumptions, the Maronna estimator behaves as a weighted version of the sample covariance matrix, where model-fitting data versus outliers are weighted differently. Our analysis paves the way to an improved usage of robust estimators of scatter in application contexts prone to outliers. As an important outcome, we have found that M-estimators of the Huber form are preferable to estimators of the Tyler form, as these are far from optimally robust against outliers.

[3]The limiting eigenvalue density is obtained from the inverse Stieltjes transform formula for the model under study; see e.g., [15].

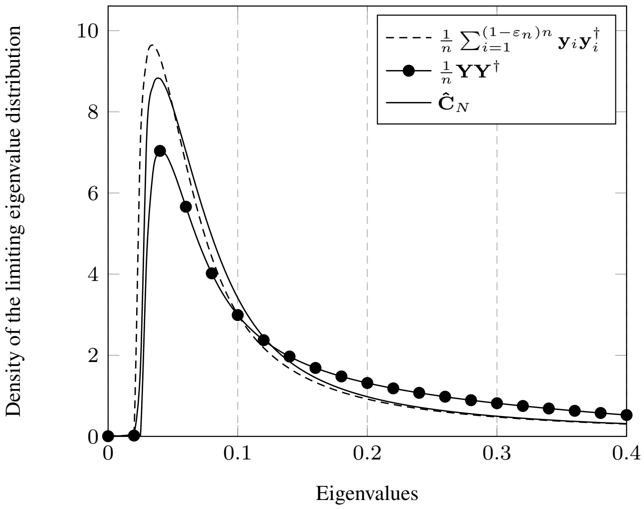

Fig. 1. Limiting eigenvalue distributions. $[\mathbf{C}_N]_{ij} = .9^{|i-j|}$, $\mathbf{D}_N = \mathbf{I}_N$, $\varepsilon = .05$.

A. INTUITIVE DERIVATION OF THE RESULTS

Both intuitive and accurate proofs follow the ideas of [6], with appropriate modifications. We provide here only the non-rigorous (although more insightful) sketch of the proof.

We start from the solution to (2), $\hat{\mathbf{C}}_N$, and define $\underline{\hat{\mathbf{C}}}_N = \mathbf{C}_N^{-1/2}\hat{\mathbf{C}}_N\mathbf{C}_N^{-1/2}$, which allows us to write

$$\underline{\hat{\mathbf{C}}}_N = \frac{1}{n}\sum_{i=1}^{(1-\varepsilon_n)n} u\left(\frac{1}{N}\mathbf{x}_i^\dagger\underline{\hat{\mathbf{C}}}_N^{-1}\mathbf{x}_i\right)\mathbf{x}_i\mathbf{x}_i^\dagger + \frac{1}{n}\sum_{i=1}^{\varepsilon_n n} u\left(\frac{1}{N}\tilde{\mathbf{a}}_i^\dagger\underline{\hat{\mathbf{C}}}_N^{-1}\tilde{\mathbf{a}}_i\right)\tilde{\mathbf{a}}_i\tilde{\mathbf{a}}_i^\dagger \tag{8}$$

where $\tilde{\mathbf{a}}_i = \mathbf{C}_N^{-1/2}\mathbf{a}_i$. The intuitive idea is to approximate the quadratic forms $\frac{1}{N}\mathbf{x}_i^\dagger\underline{\hat{\mathbf{C}}}_N^{-1}\mathbf{x}_i$ and $\frac{1}{N}\tilde{\mathbf{a}}_i^\dagger\underline{\hat{\mathbf{C}}}_N^{-1}\tilde{\mathbf{a}}_i$ by some deterministic quantities making use of standard random matrix results. To that end, the main difficulty lies in the dependence structure between $\underline{\hat{\mathbf{C}}}_N$ and the vectors $\mathbf{x}_i$. Following the same steps as in [16, III.A], this dependence can be 'weakened' by rewriting (8) as

$$\underline{\hat{\mathbf{C}}}_N = \frac{1}{n}\sum_{i=1}^{(1-\varepsilon_n)n} v(d_i)\,\mathbf{x}_i\mathbf{x}_i^\dagger + \frac{1}{n}\sum_{i=1}^{\varepsilon_n n} v(b_i)\,\tilde{\mathbf{a}}_i\tilde{\mathbf{a}}_i^\dagger \tag{9}$$

with $d_1,\ldots,d_{(1-\varepsilon_n)n}$ and $b_1,\ldots,b_{\varepsilon_n n}$ the unique solutions to the n equations

$$\begin{aligned} d_i &= \frac{1}{N}\mathbf{x}_i^\dagger\hat{\mathbf{C}}_{(x_i)}^{-1}\mathbf{x}_i, \quad i = 1,\ldots,(1-\varepsilon_n)n \\ b_i &= \frac{1}{N}\tilde{\mathbf{a}}_i^\dagger\hat{\mathbf{C}}_{(a_i)}^{-1}\tilde{\mathbf{a}}_i, \quad i = 1,\ldots,\varepsilon_n n, \end{aligned} \tag{10}$$

where $\hat{\mathbf{C}}_{(x_i)}$ and $\hat{\mathbf{C}}_{(a_i)}$ are built from $\underline{\hat{\mathbf{C}}}_N$ by removing the outer product involving $\mathbf{x}_i$ and $\mathbf{a}_i$, respectively. Note that $\hat{\mathbf{C}}_{(x_i)}$ and $\mathbf{x}_i$ are not completely independent since $\underline{\hat{\mathbf{C}}}_N^{-1}$ (in the argument of the u function for all samples) is built on $\mathbf{x}_i$. This dependence, however, seems to be 'weak' since $\mathbf{x}_i$ is only one among a growing number n of $\mathbf{x}_j$ vectors. Approximating this 'weak' dependence by independence, we can use trace and rank-one perturbation arguments (see, e.g. [17, Lemma 3.1]) which suggest that

$$d_i = \frac{1}{N}\mathbf{x}_i^\dagger\hat{\mathbf{C}}_{(x_i)}^{-1}\mathbf{x}_i \approx \frac{1}{N}\operatorname{tr}\underline{\hat{\mathbf{C}}}_N^{-1} \triangleq d. \tag{11}$$

From known large random matrix results (see, e.g., [18, 19]), we also expect d and b_i to have deterministic equivalents. Assume this is true, i.e., there exist deterministic sequences $\{\gamma_n\}_{n=1}^\infty$ and $\{\alpha_{i,n}\}_{n=1}^\infty$ such that

$$|d-\gamma_n| \xrightarrow{\text{a.s.}} 0 \tag{12}$$

$$|b_i-\alpha_{i,n}| \xrightarrow{\text{a.s.}} 0, \quad i = 1,\ldots,\varepsilon_n n. \tag{13}$$

We can then approximate

$$\underline{\hat{\mathbf{C}}}_N \approx \frac{1}{n}\sum_{i=1}^{(1-\varepsilon_n)n} v(\gamma_n)\,\mathbf{x}_i\mathbf{x}_i^\dagger + \frac{1}{n}\sum_{i=1}^{\varepsilon_n n} v(\alpha_{i,n})\,\tilde{\mathbf{a}}_i\tilde{\mathbf{a}}_i^\dagger \tag{14}$$

and, consequently,

$$d \approx \frac{1}{N}\operatorname{tr}\left(\frac{1}{n}\sum_{i=1}^{(1-\varepsilon_n)n} v(\gamma_n)\,\mathbf{x}_i\mathbf{x}_i^\dagger + \frac{1}{n}\sum_{i=1}^{\varepsilon_n n} v(\alpha_{i,n})\,\tilde{\mathbf{a}}_i\tilde{\mathbf{a}}_i^\dagger\right)^{-1} \tag{15}$$

$$b_i \approx \frac{1}{N}\tilde{\mathbf{a}}_i^\dagger\left(\frac{1}{n}\sum_{j=1}^{(1-\varepsilon_n)n} v(\gamma_n)\,\mathbf{x}_j\mathbf{x}_j^\dagger + \frac{1}{n}\sum_{j\neq i}^{\varepsilon_n n} v(\alpha_{j,n})\,\tilde{\mathbf{a}}_j\tilde{\mathbf{a}}_j^\dagger\right)^{-1}\tilde{\mathbf{a}}_i, \tag{16}$$

with $v(\gamma_n)$ now independent of $\mathbf{x}_i$, and recall that $\tilde{\mathbf{a}}_i$'s are deterministic. Then, (15) and (16) are functionals of a general class of random matrices whose deterministic equivalents are known (see, e.g., [18, 19]). From a direct application of [18, Thm. 1], we would then expect γ_n and $\alpha_{i,n}$, $i = 1,\ldots,\varepsilon_n n$, to be given by (5), the system of fixed-point equations in Theorem 1. In fact, we can prove rigorously that such γ_n and $\alpha_{i,n}$ are well-defined and satisfy $\max_{1\le i\le(1-\varepsilon_n)n}|d_i-\gamma_n| \xrightarrow{\text{a.s.}} 0$ and $\max_{1\le i\le\varepsilon_n n}|b_i-\alpha_{i,n}| \xrightarrow{\text{a.s.}} 0$. This uniform convergence ensures that $\left\|\underline{\hat{\mathbf{C}}}_N - \underline{\hat{\mathbf{S}}}_N\right\| \xrightarrow{\text{a.s.}} 0$ where

$$\underline{\hat{\mathbf{S}}}_N = \frac{1}{n}\sum_{i=1}^{(1-\varepsilon_n)n} v(\gamma_n)\,\mathbf{x}_i\mathbf{x}_i^\dagger + \frac{1}{n}\sum_{i=1}^{\varepsilon_n n} v(\alpha_{i,n})\,\tilde{\mathbf{a}}_i\tilde{\mathbf{a}}_i^\dagger. \tag{17}$$

It is then immediate to see under Assumption 2 that this, along with $\hat{\mathbf{C}}_N = \mathbf{C}_N^{1/2}\underline{\hat{\mathbf{C}}}_N\mathbf{C}_N^{1/2}$, yields the result in Theorem 1.

For the case of random outliers, the result in Corollary 1 can be derived from Theorem 1 by using the same random matrix arguments, i.e., trace and rank-one perturbation arguments along with the deterministic equivalent from [18, Thm. 1], but now focused on the random outlying vectors $\mathbf{a}_i$.

B. REFERENCES

[1] P. Bianchi, J. Najim, M. Maida, and M. Debbah, "Performance analysis of some eigen-based hypothesis tests for collaborative sensing," in *IEEE 15th Workshop on Statistical Signal Processing*, 2009.

[2] B. Nadler, "Nonparametric detection of signals by information theoretic criteria: performance analysis and an improved estimator," *IEEE Transactions on Signal Processing*, vol. 58, no. 5, pp. 2746–2756, 2010.

[3] X. Mestre and M. A. Lagunas, "Modified subspace algorithms for DoA estimation with large arrays," *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 598–614, 2008.

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

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

[6] R. Couillet, F. Pascal, and J. W. Silverstein, "The random matrix regime of Maronna's M-estimator with elliptically distributed samples," *arXiv preprint arXiv:1311.7034*, 2013.

[7] T. Zhang, X. Cheng, and A. Singer, "Marchenko-Pastur law for Tyler's and Maronna's M-estimators," *http://arxiv.org/abs/1401.3424*, 2014.

[8] R. Couillet and M. McKay, "Large dimensional analysis and optimization of robust shrinkage covariance matrix estimators," *Journal of Multivariate Analysis*, vol. 131, pp. 99–120, 2014.

[9] N El Karoui, "Asymptotic behavior of unregularized and ridge-regularized high-dimensional robust regression estimators: rigorous results," *arXiv preprint arXiv:1311.2445*, 2013.

[10] L. Yang, R. Couillet, and M. McKay, "Minimum variance portfolio optimization with robust shrinkage covariance estimation," in *Proc. IEEE Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, 2014.

[11] R. Couillet, "Robust spiked random matrices and a robust G-MUSIC estimator," *submitted to Journal of Multivariate Analysis*, 2014.

[12] R. Couillet, A. Kammoun, and F. Pascal, "Second order statistics of robust estimators of scatter. application to GLRT detection for elliptical signals," *(submitted to) Elsevier Journal of Multivariate Analysis*, 2014.

[13] J. T. Kent and D. E. Tyler, "Redescending M-estimates of multivariate location and scatter," *The Annals of Statistics*, pp. 2102–2119, 1991.

[14] R. Couillet, F. Pascal, and J. W. Silverstein, "Robust estimates of covariance matrices in the large dimensional regime," *IEEE Transactions on Information Theory*, 2013.

[15] R. Couillet, M. Debbah, and J. W. Silverstein, "A deterministic equivalent for the analysis of correlated MIMO multiple access channels," *IEEE Transactions on Information Theory*, vol. 57, no. 6, pp. 3493–3514, 2011.

[16] R. Couillet and F. Pascal, "Robust M-estimator of scatter for large elliptical samples," *IEEE Worshop on Statistical Signal Processing (SSP'14)*, Gold Coast (Australia), 2014.

[17] J. W. Silverstein and ZD 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.

[18] F. Rubio and X. Mestre, "Spectral convergence for a general class of random matrices," *Statistics and Probability Letters*, vol. 81, no. 5, pp. 592–602, 2011.

[19] S. Wagner, R. Couillet, M. Debbah, and D. T. Slock, "Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback," *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4509–4537, July 2012.