

ROBUST SHRINKAGE M-ESTIMATORS OF LARGE COVARIANCE MATRICES

Nicolas Auguin*, David Morales-Jimenez*, Matthew McKay*, Romain Couillet†

*Hong Kong University of Science and Technology, ECE Department, Hong Kong

†CentraleSupélec, Université Paris-Saclay, Gif-sur-Yvette, France

ABSTRACT

Robust high dimensional covariance estimators are considered, comprising regularized (linear shrinkage) modifications of Maronna’s classical M-estimators. Such estimators aim to provide robustness to outliers, while simultaneously giving well-defined solutions under high dimensional scenarios where the number of samples does not exceed the number of variables. By applying tools from random matrix theory, we characterize the asymptotic performance of such estimators when the number of samples and variables grow large together. In particular, our results show that, when outliers are absent, many estimators of the shrinkage-Maronna type share the same asymptotic performance, and for such estimators we present a data-driven method for choosing the asymptotically optimal shrinkage parameter. Although our results assume an outlier-free scenario, simulations suggest that certain estimators perform substantially better than others when subjected to outlier samples.

Index Terms— Robust estimation, covariance matrices, random matrix theory

1. INTRODUCTION

The estimation of covariance or scatter matrices under large dimensions is a fundamental problem in statistical signal processing [1, 2], with applications ranging from financial engineering [3] to biology [4]. Among such estimates, the sample covariance matrix (SCM) $\frac{1}{n} \sum_{i=1}^n \mathbf{y}_i \mathbf{y}_i^\dagger$, where $\mathbf{y}_1, \dots, \mathbf{y}_n \in \mathbb{C}^N$ are data samples, is a particularly appealing choice since its structure is simple and it converges (when the \mathbf{y}_i are i.i.d.) to the population covariance matrix in the asymptotic regime $n \rightarrow \infty$, with N fixed. Nonetheless, the SCM is known to suffer from three major drawbacks: (i) it is a poor estimate of the true covariance matrix whenever the number of samples n and the number of variables N are of similar order, (ii) it is not invertible for $n < N$, and (iii) it is not robust to outliers. Several methods have been proposed to address the first issue, the most popular being linear shrinkage [5] which consists of a linear combination of the SCM and a shrinkage target, such as the identity matrix.

This results in an estimator of the form

$$(1 - \mu) \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i \mathbf{y}_i^\dagger + \mu \nu \mathbf{I}_N, \quad (1)$$

where μ is a regularization parameter and ν is a scaling factor. Hereafter, this estimator will be referred to as the Ledoit-Wolf estimator [5]. Albeit powerful to deal with data scarcity, linear shrinkage does not alleviate the problem of the sensitivity to the presence of outliers. In order to provide a better inference of the true covariance matrix when outlying samples corrupt the data (or when the distribution of \mathbf{y}_i is heavy-tailed, as when dealing with elliptical data), robust covariance M-estimation methods were designed [6, 7, 8] with the purpose of harnessing these presumed outlying samples. Such estimators are however only defined when the number of samples exceeds the number of variables; thus, regularized versions have been proposed to account for possibly scarce data [1, 2, 9]. These works typically focus on a particular type of M-estimator, known as Tyler’s estimator. Despite the fact that the behavior of Tyler’s estimator is now well understood (at least in the large N, n regime [10]), it suffers from two shortcomings: (i) if there is a certain number of samples that lie in a given subspace of the data, it may not be defined [11], and (ii) in the presence of outliers, it may even favor outliers over actual data [12].

It then becomes important to consider a wider class of robust estimators that can alleviate the aforementioned two issues. In particular, Maronna’s M-estimators encompass a wide range of estimators whose particular form has been shown essential when the data is corrupted by outliers [12]. In this work, we focus on a regularized version of these estimators [11], thus well-defined when the number of variables exceeds that of the samples. Building upon recent works [10, 11], we first aim at understanding the asymptotic behavior of these estimators in the large N, n regime. Upon properly choosing the shrinkage parameter, it is shown that, in the absence of outliers, the performance of a wide class of M-estimators can be optimized with respect to (w.r.t.) the Frobenius norm of the matrix approximation error. A simple data-driven approach is proposed to estimate the optimal parameter, the efficiency of which is validated through simulations.

Notations: $\|\mathbf{A}\|$ and $\|\mathbf{A}\|_F$ denote the spectral norm and the Frobenius norm of the matrix \mathbf{A} , respectively.

N. Auguin, D. Morales-Jimenez and M. R. McKay were supported by the Hong Kong Research Grants Council under grant number 16203315. R. Couillet was supported by the ERC MORE EC – 120133.

2. PROBLEM STATEMENT AND MOTIVATION

Let $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_n] \in \mathbb{C}^{N \times n}$ be a matrix whose columns are N -dimensional data samples, drawn from $\mathbf{y}_i = \mathbf{C}_N^{1/2} \mathbf{x}_i$, where $\mathbf{C}_N \in \mathbb{C}^{N \times N} \succeq 0$ is deterministic and $\mathbf{x}_1, \dots, \mathbf{x}_n$ are i.i.d. random vectors the entries of which are i.i.d. with zero mean, unit variance and finite $(8 + \sigma)$ -th order moment (for $\sigma > 0$).

Assumption 1. The growth regime is such that $c_N \triangleq N/n \rightarrow c \in (0, \infty)$ as $N, n \rightarrow \infty$. We further assume that for all N , $\text{Tr } \mathbf{C}_N = kN$, with $k \geq 0$ and that $\limsup_N \|\mathbf{C}_N\| < \infty$.

Let us define u as a non-negative, continuous, bounded and non-increasing function on $[0, \infty)$, and such that $\phi(x) \triangleq xu(x)$ is non-decreasing with $\phi_\infty \triangleq \lim_{x \rightarrow \infty} \phi(x) < \infty$. For each $\alpha \geq 0$ and $\beta > 0$, we define the shrinkage Maronna's M-estimator $\hat{\mathbf{C}}_N(\alpha, \beta)$ associated with the function u as the unique solution to the equation in \mathbf{Z} :

$$\mathbf{Z} = \alpha \frac{1}{n} \sum_{i=1}^n u \left(\frac{1}{N} \mathbf{y}_i^\dagger \mathbf{Z}^{-1} \mathbf{y}_i \right) \mathbf{y}_i \mathbf{y}_i^\dagger + \beta \mathbf{I}_N. \quad (2)$$

Note that for an unbounded u function, (2) does not necessarily have a solution for all $\alpha \geq 0, \beta > 0$. In particular, for $u(x) = 1/x$, the equation in \mathbf{Z}

$$\mathbf{Z} = \alpha \frac{1}{n} \sum_{i=1}^n \frac{\mathbf{y}_i \mathbf{y}_i^\dagger}{\frac{1}{N} \mathbf{y}_i^\dagger \mathbf{Z}^{-1} \mathbf{y}_i} + \beta \mathbf{I}_N, \quad (3)$$

admits a solution when $n < N$ if and only if $0 \leq \alpha < n/N$ and $\beta > 0$ [2, 11]. In the case where $(\alpha, \beta) = (1 - \rho, \rho)$, with corresponding restriction on ρ , it was shown in [10] that Tyler's estimator constitutes a particularly appealing choice when dealing with a broad class of heavy-tailed distributions, e.g. elliptical data [13]. Nevertheless, a recently proposed large-dimensional analysis of robust M-estimators in the presence of outliers [12] showed that such Tyler's estimators fail to handle certain types of outliers, hence providing motivation for studying the wider class of shrinkage estimators which are solutions of (2). As the form of the estimator in (2) is rather involved, we first aim at characterizing its behavior in the large n, N regime.

3. ASYMPTOTIC BEHAVIOR

Assumption 2. $\alpha c \phi_\infty < 1$.

Definition. Under Assumptions 1-2, let $v : [0, \infty) \rightarrow (0, u(0)]$ be defined as $v(x) = u(g^{-1}(x))$ where g^{-1} denotes the inverse function of $g(x) = \frac{x}{1 - \alpha c x u(x)}$, mapping $[0, \infty)$ onto $[0, \infty)$. The function v thus defined is continuous, non-increasing and onto. Also define \mathcal{I} as the set of compacts included in $[0, 1/c\phi_\infty) \times (0, \infty]$.

Theorem 1. Let $\hat{\mathbf{C}}_N(\alpha, \beta)$ be the unique solution to (2). Then, as $N, n \rightarrow \infty$, under Assumptions 1-2,

$$\forall I \subset \mathcal{I}, \sup_{(\alpha, \beta) \in I} \left\| \hat{\mathbf{C}}_N(\alpha, \beta) - \hat{\mathbf{S}}_N(\alpha, \beta) \right\| \xrightarrow{\text{a.s.}} 0$$

where

$$\hat{\mathbf{S}}_N(\alpha, \beta) \triangleq \alpha v(\gamma) \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i \mathbf{y}_i^\dagger + \beta \mathbf{I}_N, \quad (4)$$

with γ the unique positive solution to the equation

$$\gamma = \frac{1}{N} \text{Tr} \left[\mathbf{C}_N \left(\alpha \frac{v(\gamma)}{1 + c\alpha v(\gamma)} \mathbf{C}_N + \beta \mathbf{I}_N \right)^{-1} \right].$$

Furthermore, when seen as a function of β , $\beta \mapsto \gamma(\beta)$ is bounded, continuous on $(0, \infty]$ and away from zero.

Proof. Follows along similar lines to [10, Theorem 1], adapted to account for a general u function and parameters α and β . Details will be provided in an extended version. \square

Remark 1. Theorem 1 shows that the proposed estimator asymptotically behaves (uniformly on $(\alpha, \beta) \in I$) as a particular version of the Ledoit-Wolf estimator, with weights $(\alpha v(\gamma), \beta)$ in lieu of the parameters $(1 - \mu, \mu v)$ in (1).

Notice that the matrix $\hat{\mathbf{S}}_N(\alpha, \beta)$ is only defined when Assumption 2 is verified, i.e. $\alpha < 1/c\phi_\infty$. For $\alpha \geq 1/c\phi_\infty$, although a solution to (2) does exist, it is not clear whether it can be characterized under double asymptotics. To alleviate this issue, we can resort to the following assumption (which is a particular case of Assumption 2):

Assumption 3. $c\phi_\infty \leq 1$.

Under Assumption 3, $\hat{\mathbf{S}}_N(\alpha, \beta)$ is now defined for all $\alpha < 1$. In the following, with this assumption, we show that the two-parameter formulation (α, β) can be reduced to a single parameter $(1 - \rho, \rho)$ formulation.

Proposition 1. Under Assumption 3, for all $(\alpha, \beta) \in [0, 1) \times (0, \infty)$, there exists $\rho \in (0, 1]$ such that

$$\frac{\hat{\mathbf{S}}_N(\alpha, \beta)}{\alpha k v(\gamma(\alpha, \beta)) + \beta} = \frac{\hat{\mathbf{S}}_N(1 - \rho, \rho)}{(1 - \rho) k v(\gamma(1 - \rho, \rho)) + \rho}.$$

Proof. Defining $L_N \triangleq \alpha k v(\gamma(\alpha, \beta)) + \beta$, we have $L_N - \frac{1}{N} \text{Tr } \hat{\mathbf{S}}_N(\alpha, \beta) \xrightarrow{\text{a.s.}} 0$. Then, using (4), we have:

$$\frac{\hat{\mathbf{S}}_N(\alpha, \beta)}{L_N} = \left(1 - \frac{\beta}{L_N} \right) \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i \mathbf{y}_i^\dagger + \frac{\beta}{L_N} \mathbf{I}_N.$$

If $\alpha k \neq 0$, as v is bounded and away from 0, $\frac{\beta}{\alpha k v(\gamma) + \beta} \rightarrow 0$ as $\beta \rightarrow 0$ and $\frac{\beta}{\alpha k v(\gamma) + \beta} \rightarrow 1$ as $\beta \rightarrow \infty$. As $\beta \mapsto v(\gamma(\beta))$ is continuous on $(0, \infty)$, we conclude that there exists $\rho \in (0, 1]$ such that the asymptotic equivalent $\hat{\mathbf{S}}_N(\alpha, \beta)$ is equal (up to a trace-normalization factor) to $\hat{\mathbf{S}}_N(1 - \rho, \rho)$. If $\alpha k = 0$, the result holds for $\rho = 1$. \square

Proposition 1 means that, under Assumption 3, it is equivalent in the large N, n regime to consider the two-parameter equation (2) or the one-parameter equation:

$$\mathbf{Z} = (1 - \rho) \frac{1}{n} \sum_{i=1}^n u \left(\frac{1}{N} \mathbf{y}_i^\dagger \mathbf{Z}^{-1} \mathbf{y}_i \right) \mathbf{y}_i \mathbf{y}_i^\dagger + \rho \mathbf{I}_N. \quad (5)$$

Thus, we may directly work with the solution $\hat{\mathbf{C}}_N(\rho) \triangleq \hat{\mathbf{C}}_N(1-\rho, \rho)$ of (5) for $\rho \in (0, 1]$ and its equivalent $\hat{\mathbf{S}}_N(\rho) \triangleq \hat{\mathbf{S}}_N(1-\rho, \rho)$, rather than the more complex model (2). With this simplification, we have the following useful result:

Proposition 2. For each $\underline{\rho} \in (0, 1]$, there exists a solution $\rho \in (0, 1]$ to the equation

$$\frac{\rho}{(1-\rho)kv(\gamma) + \rho} = \underline{\rho}$$

for which we have

$$\frac{\hat{\mathbf{S}}_N(\rho)}{(1-\rho)kv(\gamma) + \rho} = (1-\rho) \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i \mathbf{y}_i^\dagger + \underline{\rho} \mathbf{I}_N.$$

Proof. Details will be provided in an extended version. \square

Proposition 2 implies that any scaled¹ Ledoit-Wolf estimator (1) is equal to a trace-normalized estimator $\hat{\mathbf{S}}_N(\rho)$ for a particular ρ and therefore, thanks to Theorem 1, converges a.s. to the corresponding normalized estimator $\hat{\mathbf{C}}_N(\rho)$ solving for (5). The advantage of Proposition 2 lies in the fact that if we are able to characterize a parameter $\underline{\rho}$ such that the associated Ledoit-Wolf estimator is optimal w.r.t. a given loss function (e.g. the Frobenius loss), then, upon finding a parameter ρ that maps to $\underline{\rho}$, the corresponding shrinkage Maronna estimator will be optimal w.r.t. that same loss function. This aspect is further developed in the following section.

4. OPTIMAL SHRINKAGE

Define, for a given estimator $\hat{\mathbf{B}}_N$ of \mathbf{C}_N , the squared Frobenius loss of the difference of the trace-normalized estimator $\frac{\hat{\mathbf{B}}_N}{\frac{1}{N} \text{Tr} \hat{\mathbf{B}}_N}$ to the trace-normalized population matrix \mathbf{C}_N as

$$D_N(\hat{\mathbf{B}}_N) \triangleq \frac{1}{N} \left\| \frac{\hat{\mathbf{B}}_N}{\frac{1}{N} \text{Tr} \hat{\mathbf{B}}_N} - \frac{\mathbf{C}_N}{\frac{1}{N} \text{Tr} \mathbf{C}_N} \right\|_F^2.$$

Assumption 4. $\frac{1}{N} \text{Tr} \mathbf{C}_N^2$ converges as $N \rightarrow \infty$ to a limit noted M .

Denote $\hat{\mathbf{C}}_N(\rho)$ the solution to (5) for a given $\rho \in (0, 1]$. Then we have the following proposition:

Proposition 3. (Optimal shrinkage.) We have

$$\inf_{\rho \in (0, 1]} D_N(\hat{\mathbf{C}}_N(\rho)) \xrightarrow{\text{a.s.}} D^* \triangleq c \frac{M+1-2k}{c+M+1-2k}.$$

Furthermore, denote $\hat{\rho}^* \in (0, 1]$ a solution to

$$\frac{\hat{\rho}^*}{(1-\hat{\rho}^*)kv(\gamma) + \hat{\rho}^*} = \rho^* \triangleq \frac{c}{c+M+1-2k},$$

where ρ^* is the shrinkage parameter for which the associated Ledoit-Wolf estimator (1) would be optimal w.r.t. the Frobenius norm.

¹Here, $\underline{\rho} = \mu\nu/(1-\mu+\mu\nu)$ in (1).

Then we have $D_N(\hat{\mathbf{C}}_N(\hat{\rho}^*)) \xrightarrow{\text{a.s.}} D^*$.

(Optimal shrinkage parameter estimate.) Denote $\hat{\rho}_N \in (0, 1]$ a solution to

$$\frac{\hat{\rho}_N}{\frac{1}{N} \text{Tr} \hat{\mathbf{C}}_N(\hat{\rho}_N)} = \frac{c_N}{\frac{1}{N} \text{Tr} \left[\left(\frac{1}{n} \sum_{i=1}^n \frac{\mathbf{y}_i \mathbf{y}_i^\dagger}{\frac{1}{N} \|\mathbf{y}_i\|^2} \right)^2 \right] - 1}.$$

Then, under Assumption 3, we have $\hat{\rho}_N \xrightarrow{\text{a.s.}} \hat{\rho}^*$ and $D_N(\hat{\mathbf{C}}_N(\hat{\rho}_N)) \xrightarrow{\text{a.s.}} D^*$.

Proof. Follows along the lines of [10, Propositions 1 and 2]. Details will be provided in an extended version. \square

In words, Proposition 3 implies that, for any function u satisfying Assumption 3, there exists a shrinkage parameter $\hat{\rho}^*$ for which the corresponding Maronna shrinkage estimator is optimal (w.r.t. the Frobenius norm) in the large N, n regime. Though this shrinkage parameter $\hat{\rho}^*$ is an oracle (in that it depends on the unknown population matrix \mathbf{C}_N , through its dependence on γ and M), Proposition 3 provides a data-driven way to estimate it, and the associated estimator has asymptotically optimal performance.

Experiment 1. Let us take $N = 150$, $n = 100$ ($c_N = 3/2 > 1$, such that shrinkage becomes necessary when using Maronna's estimator), and let $[\mathbf{C}_N]_{ij} = .9^{|i-j|}$. We consider two different u functions: $u_1(x) = \min\{1, \frac{1+t}{t+x}\}$ and $u_2(x) = \frac{1}{c} u_1(x)$, with $t = .01$ (these functions are referred to as Huber-type functions in the literature). The interest of such functions will be made clear in the following section). Note that u_1 does not verify Assumption 3, whereas its scaled counterpart u_2 does. In Fig. 1, for each u function and $\rho \in (0, 1]$ we plot the expected Frobenius loss D_N associated with the solution of (5), as well as that of the corresponding equivalent $\hat{\mathbf{S}}_N(\rho)$ (when defined). The minimal value D^* is also indicated for reference.

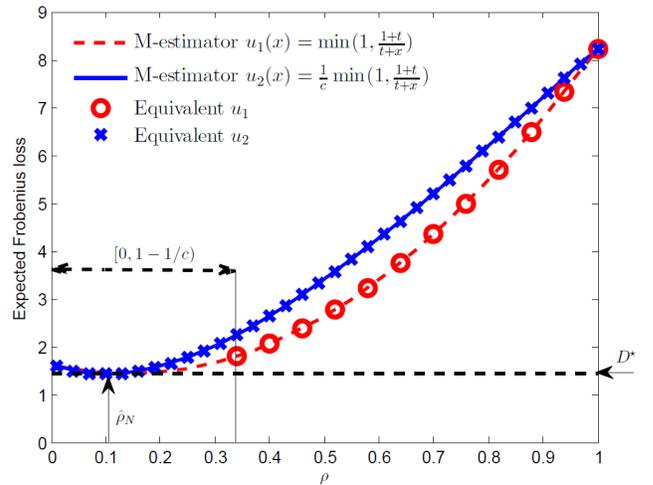


Fig. 1: Expected Frobenius loss of the estimator for different u functions as the shrinkage parameter ρ varies.

First, we observe that for all u functions and all $\rho > 1 - 1/c$, there is a close match between $\hat{\mathbf{C}}_N(\rho)$ and the cor-

responding, asymptotically-equivalent matrix $\hat{S}_N(\rho)$, which validates Theorem 1. Second, notice that for u_1 , as Assumption 3 is not verified, we do not have existence of the equivalent $\hat{S}_N(\rho)$ for $\rho \in (0, 1 - 1/c]$. However, for u_2 , which is a scaled version of u_1 such that Assumption 3 is verified, the matrix $\hat{S}_N(\rho)$ is well defined for all $\rho > 0$. Following Proposition 3, we can therefore find a shrinkage parameter such that the associated estimator has a performance close to D^* . With Proposition 3, this optimal shrinkage parameter can be estimated ($\hat{\rho}_N$, as indicated on Fig. 1 for u_2), and the corresponding estimator is guaranteed to have optimal performance.

5. PERFORMANCE WITH OUTLIERS

Proposition 3 shows that, as long as the data is not corrupted by outliers, choosing one particular u function over another will not affect the (asymptotic) accuracy of the associated estimator (upon optimizing the shrinkage parameter ρ). However, in general this is not the case when the data contains outlying samples, in which case different u functions lead to estimators which may behave very differently. The following experiment aims at shedding light on this.

Experiment 2. Here we choose $N = 150$ and $[C_N]_{ij} = .9^{|i-j|}$ as before. We assume that $n - 2$ samples are drawn from the distribution associated with C_N and that there are 2 outlying samples in the data, which are both equal to $1/\sqrt{N}\mathbf{1}_N$, where $\mathbf{1}_N$ denotes the vector of all ones. We consider the functions $u_2(x) = \frac{1}{c} \min\{1, \frac{1+t}{t+x}\}$ and $u_3(x) = \frac{1}{c} \frac{1+t}{t+x}$, properly scaled such that Assumption 3 is verified. We let n vary from 10 to 150 (therefore, c_N varies from $1/15$ to 1), and estimate for each n and each u function the optimal shrinkage parameter according to Proposition 3. The performance of the estimators associated with these estimates is reported in Fig. 2.

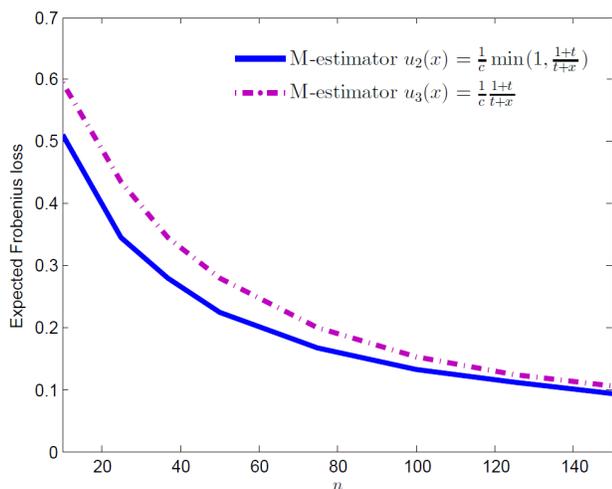


Fig. 2: Expected optimal Frobenius loss of the estimator for different u functions, as n varies.

Though the estimator associated with u_3 is well defined for all n , it performs poorly, in comparison to the estimator associated with u_2 . The reason behind this phenomenon lies in the fact that estimators corresponding to Tyler-type functions ($x \mapsto 1/x$ or $x \mapsto \frac{1+t}{t+x}$ for t small) tend to enhance certain types of outliers, as explained in [12], while the estimator associated with u_2 (Huber-type) can control such outlying samples. This experiment thus motivates the need for considering a large class of regularized robust M-estimators. Notice that, even in the presence of outliers, the proposed analysis and the estimation of an optimal shrinkage parameter ρ (though assuming clean data) are still useful, since in general, no prior knowledge about potential outliers is available. A more thorough study on the impact of different types of outliers on these estimators is the subject of ongoing work.

6. REFERENCES

- [1] Y. Abramovich, N. K. Spencer, et al., "Diagonally loaded normalised sample matrix inversion (LNSMI) for outlier-resistant adaptive filtering," in *IEEE ICASSP*, 2007, vol. 3, pp. III-1105.
- [2] F. Pascal, Y. Chitour, and Y. Quek, "Generalized robust shrinkage estimator and its application to STAP detection problem," *IEEE Trans. Signal Processing*, vol. 62, no. 21, pp. 5640-5651, 2014.
- [3] L. Yang, R. Couillet, and M. R. McKay, "A robust statistics approach to minimum variance portfolio optimization," *IEEE Trans. Signal Processing*, vol. 63, no. 24, pp. 6684-6697, 2015.
- [4] J. Schäfer and K. Strimmer, "A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics," *Statistical Applications in Genetics and Molecular Biology*, vol. 4, no. 1, 2005.
- [5] 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, 2004.
- [6] P. J. Huber, *Robust Statistics*, Springer, 2011.
- [7] R. A. Maronna and V. J. Yohai, "Robust estimation of multivariate location and scatter," *Encyclopedia of Statistical Sciences*, 1998.
- [8] R. Couillet, F. Pascal, and J. W. Silverstein, "The random matrix regime of Maronna's M-estimator with elliptically distributed samples," *Journal of Multivariate Analysis*, vol. 139, pp. 56-78, 2015.
- [9] Y. Chen, A. Wiesel, and A. O. Hero, "Robust shrinkage estimation of high-dimensional covariance matrices," *IEEE Trans. Signal Processing*, vol. 59, no. 9, pp. 4097-4107, 2011.
- [10] R. Couillet and M. R. McKay, "Large dimensional analysis and optimization of robust shrinkage covariance matrix estimators," *Journal of Multivariate Analysis*, vol. 131, pp. 99-120, 2014.
- [11] E. Ollila and D. E. Tyler, "Regularized-estimators of scatter matrix," *IEEE Trans. Signal Processing*, vol. 62, no. 22, pp. 6059-6070, 2014.
- [12] D. Morales-Jimenez, R. Couillet, and M. R. McKay, "Large dimensional analysis of robust M-estimators of covariance with outliers," *IEEE Trans. Signal Processing*, vol. 63, no. 21, pp. 5784-5797, 2015.
- [13] E. Ollila, D. E. Tyler, Koivunen, and H. V. Poor, "Complex elliptically symmetric distributions: Survey, new results and applications," *IEEE Trans. Signal Processing*, vol. 60, no. 11, pp. 5597-5625, 2012.