

Robust Estimation of Scatter Matrix, Random Matrix Theory and an Application to Spectrum Sensing

Thesis by
Zhedong Liu

In Partial Fulfillment of the Requirements

For the Degree of

Master of Science

King Abdullah University of Science and Technology

Thuwal, Kingdom of Saudi Arabia

April, 2019

EXAMINATION COMMITTEE PAGE

The thesis of Zhedong Liu is approved by the examination committee

Committee Chairperson: Mohamed-Slim Alouini

Committee Members: Mohamed-Slim Alouini, Abla Kammoun , Håvard Rue

©April,2019

Zhedong Liu

All Rights Reserved

ABSTRACT**Robust Estimation of Scatter Matrix, Random Matrix Theory and
Applications to Spectrum Sensing**

Zhedong Liu

The covariance estimation is one of the most critical tasks in multivariate statistical analysis. In many applications, reliable estimation of the covariance matrix, or scatter matrix in general, is required. The performance of the classical maximum likelihood method relies a great deal on the validity of the model assumption. Since the assumptions are often approximately correct, many robust statistical methods have been proposed to be robust against the deviation from the model assumptions. M-estimator is an important class of robust estimator of the scatter matrix. The properties of these robust estimators under high dimensional setting, which means the number of dimensions has the same order of magnitude as the number of observations, is desirable. To study these, random matrix theory is a very important tool. With high dimensional properties of robust estimators, we introduced a new method for blind spectrum sensing in cognitive radio networks.

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisor Prof. Mohamed Slim Alouini and Dr. Abla Kammoun for their continued advice, support, and encouragement. Without your guidance, the thesis would have been much more difficult to achieve. Dr. Abla always efforts in providing me with constant feedback when I have any questions.

I would also like to express my heartfelt appreciation to my beloved friends and colleagues at KAUST. They have always been there to provide me with spiritual support and endless encouragement. I am so blessed to study and work with you all.

Finally, I would like to extend my most profound appreciation to my family members and dedicate my thesis to them.

TABLE OF CONTENTS

Examination Committee Page	2
Copyright	3
Abstract	4
Acknowledgements	5
List of Figures	7
1 Introduction	8
2 Estimation of Scatter Matrix	12
2.1 Maximum Likelihood Estimator of Scatter Matrix	13
2.2 M-estimator of Scatter Matrix	14
2.3 Measurement of Robustness	18
2.4 A Toy Example	20
3 Random Matrix Theory and Estimators of the Scatter Matrix	24
3.1 Sample Covariance Matrix	26
3.1.1 Asymptotic Empirical Spectral Distribution	26
3.1.2 Asymptotic Distribution of the Largest Eigenvalue	27
3.2 Tyler’s M-estimator and Maronna’s M-estimator	30
3.3 Spike Model	32
4 Application on Spectrum Sensing in Cognitive Radio Networks	38
5 Conclusion	44
References	46

LIST OF FIGURES

2.1	Empirical Probability Density of the Sample Covariance Matrix . . .	22
2.2	Empirical Probability Density of the Tyler Estimator of the Covariance Matrix	23
3.1	ESD of SCM under different distributions	28
3.2	Tracy-Widom density functions for $\beta = 1, 2$	30
3.3	ESD of Tyler's M-estimator under different distributions	31
3.4	Spike Model	37
4.1	System Setting	39
4.2	ROC curves	43

Chapter 1

Introduction

The covariance estimation is one of the central parts in multivariate statistical analysis. In many applications, estimation of the covariance matrix, or the scatter matrix in general, is required, such as signal processing, wireless communication system, biology, and financial engineering.[1, 2, 3, 4]

When there is a large data set, it is desired to assume the data set is generated by a Gaussian distribution. With the normality condition, the covariance matrix is often estimated using the classical method of Maximum Likelihood Estimation (MLE), and the resulting estimator is the Sample Covariance Matrix (SCM). Many techniques in engineering field rely on the Gaussian assumption which can be justified in many situations. With Gaussian assumption, it is easy to have SCM as the optimal estimator. However, this optimality does not hold if the estimator was derived under Gaussian distribution and the data set is not Gaussian in reality. Also, many data sets are generated by an unknown or unclear model. Thus MLE is not always applicable. Nevertheless, SCM is always applied to those data sets due to invalid Gaussian assumption. It is well known that even a slight deviation from the assumed distribution could distort the performance of SCM which makes the estimator far from optimal. [5]. Many fields of studies show that the distribution of data set is non-Gaussian and is often heavy-tailed. [6, 7, 8, 9] These situations require robust estimation of the scatter matrix, which should be almost optimal under Gaussian case and also perform well when assumptions are only approximately valid since the data generation process is not always clear. [5]

A frequent way to robustify the SCM is detecting and rejecting suspicious observations or outliers. Outliers are usually observations that exceed a few standard deviations which should also be estimated robustly in the one-dimensional case. In multidimensional case, outliers are sought to be detected by either distribution- [10] or distance-based methods [11]. The first approach consists of learning the probability distributions of the data by modeling the data as a mixture of outliers and clean data from the assumed distribution. The second approach is based on a measurement of similarity for each data point to all the other data points and outliers are those data points with the lowest similarity score. All those methods are used to detect outliers which might be treated as erroneous data or identified as the most informative data.

An important class of robust estimation is M-estimation[5]. This method is understood well in terms of statistical properties, and it can resist outliers without preprocessing the data. M-estimation is a generalization of MLE. The M-estimator of scatter matrix is obtained by minimizing a loss function or equivalently solving a fixed point of a function. It should be noted that many of these estimators are MLEs for scatter matrices of certain distributions, such as the family of elliptical distributions[1] but it is not always the case.

Another part of this thesis relates to random matrix theory. Many scientific studies deal with sets of high dimensional data samples, and therefore it is common to have data samples with dimension having the same order of magnitude as the number of observations. Under this high dimensional setting, the asymptotic eigenvalue behaviors, including the spectral distribution and behavior of largest eigenvalue with a certain population covariance matrix, of SCM at the limit of infinite number of samples and dimensions with fixed ratio have been studied well [12, 13, 14]. The asymptotic properties of M-estimator under the high dimensional setting have been analyzed in [15, 16], which obtained convergence results for a scaled M-estimator. Tyler's M-estimator is a special case of M-estimator, which has been showed that the

spectral distribution of a scaled Tyler's M-estimator converges weakly to a deterministic distribution when data samples are i.i.d drawn from elliptical distributions.

With robust estimation and its asymptotic properties in high dimensional setting, we introduced a new method for blind spectrum sensing in cognitive radio networks. Cognitive networks [17] have been proposed as a promising solution to solve the problem of spectrum scarcity by making full use of the free spectrum. The users of Cognitive networks, or secondary users (SUs), have to be able to sense the free spectrum in which no signal of the licensed users, or primary users (PUs), exists. This process is called cooperative spectrum sensing when SUs combine their information to sense the signal. There are several techniques for the spectrum sensing, such as eigenvalue based spectrum sensing [18, 19, 20], energy detection, the matched filter, the cyclostationary feature detection, and self-correlated detection. Unfortunately, most of these techniques require knowledge of signal feature of PUs or noise information. Among these techniques, eigenvalue based spectrum sensing requires no information of signal nor noise, which is lacked in cognitive radio networks, and only a small sample size [20].

Most of the sensing techniques are designed for Gaussian noise. The Gaussian assumption is always justified by central limit theorem, but it is also very often to deal with the non-Gaussian (impulsive or heavy-tailed) noise environment. In practical wireless communication system, the impulsive noise occurs due to many reasons such as vehicle ignition or switching a transient in power line [21], working electrical appliance such as microwave ovens, light switch and so on[6]. Under those circumstances, sensing techniques designed for Gaussian noise are highly susceptible to have severe degradation of performance including conventional eigenvalue based spectrum sensing which uses eigenvalues of SCM.

In this thesis, to deal with the non-Gaussian noise environment, we propose a new spectrum sensing method. The new method applies robust estimators [5] of the

scatter matrix to eigenvalue based spectrum sensing. Since the robust estimators can successfully mitigate the effects of impulsive noise, the new method is robust in the sense of insensitive to noise environment. The new method is also blind in the sense of requiring no information of signals and noise.

The remainder of the article is structured as follows. Chapter 2 provides an introduction to maximum likelihood estimation of the scatter matrix and M-estimation of the scatter matrix. Chapter 3 introduces some random matrix theory results associated with those scatter matrix estimators. Chapter 4 introduced the new spectrum sensing method in cognitive radio networks. Chapter 5 concludes this thesis.

Chapter 2

Estimation of Scatter Matrix

The usual parameter estimation focuses on the development of efficient estimators depending on a given model, where the efficiency is measured by the number of observations needed to achieve a given performance. The standard procedure of estimation is to construct the likelihood function according to the proposed model. Then maximize the likelihood function to get the maximum likelihood estimators (MLE), which have been proven to be asymptotically optimal in the sense of maximum likelihood principle given a correct model. In this thesis we are talking about the scatter matrix, which is a fraction of a covariance matrix if it exists. Under the Gaussian distribution, the famous sample covariance matrix (SCM) is the MLE and thus has the smallest asymptotic variance. The optimality of the SCM, however, can be only achieved if the underlying distribution is Gaussian and the performance of the estimator deteriorates significantly, even for minor departure from the assumed model, e.g., a small portion of data points are drawn from a non-Gaussian distribution. This motivates the search for robust alternatives of SCM.

Robust estimators tradeoff some efficiency under the assumed model to obtain reliability when the assumed model is not exact. In another word, in conventional estimation, the purpose is to find estimators possessing desirable properties at an exact model, while the purpose is loosely speaking to find estimators with desired behavior in an approximate model, for example, elliptical symmetric distributions [1, 22], skew-elliptical distributions [23], contamination model[5], etc. Generally, the robust estimator should be robust against any departure from the assumed model.

For example, the model assumes the underlying distribution is Gaussian while the real data follows an asymmetric and heavy-tailed distribution. In this thesis, we only discuss estimators robust against the heavy-tailedness.

2.1 Maximum Likelihood Estimator of Scatter Matrix

Maximum likelihood estimation (MLE) is very popular in statistical analysis. The intuition behind the MLE is that a good estimate of the unknown parameter vector θ would be the value of θ that maximizes the likelihood of getting the data we observed. The following is the definition of MLE. Let $x = \{x_1, \dots, x_n\}$ be an i.i.d sample from a p -dimensional distribution with probability density function $f(x; \theta)$ and the unknown parameter vector $\theta \in \Theta$, then the likelihood function is defined as $\mathcal{L}(\theta|x) = \prod_{i=1}^n f(x_i; \theta)$. Note that the likelihood function is a function of parameter vector θ given all the observations. Then the MLE is the maximizer of the likelihood function $\mathcal{L}(\theta|x)$ or equivalently the maximizer of the log likelihood function $l(\theta|x) = \sum_{i=1}^n \log f(x_i; \theta)$.

Lets show that the SCM is the MLE of the multivariate Gaussian distribution given mean μ and covariance C_p . The log likelihood function given a Gaussian data set is

$$\begin{aligned} l(\mu, C_p|x) &= \text{const} - \frac{n}{2} \log |C_p| - \frac{1}{2} \sum_{i=1}^n (x_i - \mu)^T C_p^{-1} (x_i - \mu) \\ &= \text{const} + \frac{n}{2} \log |C_p^{-1}| - \frac{1}{2} \sum_{i=1}^n (x_i - \mu)^T C_p^{-1} (x_i - \mu). \end{aligned}$$

Take derivative with respect to C_p^{-1} on the log likelihood function (applying $\frac{\partial b^T A b}{\partial A} = b b^T$ and $\frac{\partial |A|}{\partial A} = |A| A^{-1}$),

$$\frac{\partial l}{\partial C_p^{-1}} = \frac{n}{2} C_p - \frac{1}{2} \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^T.$$

Clearly the critical point of C_p is $C_p^* = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^T$ with the same procedure we can find the critical point of μ is $\mu^* = \frac{1}{n} \sum_{i=1}^n x_i$. It can be argued that $(\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, C_{SCM} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T)$ is the MLE of multivariate Gaussian distribution and C_{SCM} is the sample covariance matrix.

Given a probability model with fixed dimension, as n grows, C_{MLE} is consistent, i.e., $C_{MLE} \rightarrow C_p$ in probability. Also, MLE processes the asymptotic normality, i.e., $n \rightarrow \infty$, $\sqrt{n}(\text{vec } C_{MLE} - \text{vec } C_p) \rightarrow N(0, I^{-1}(\theta))$ in distribution, where $I(\theta)$ is the fisher information defined by, $I(\theta) = E_{\theta}(l'(x|\theta)^2) = -E_{\theta}(l''(x|\theta))$. The most celebrated result is that $I^{-1}(\theta)$ reaches the Cramer-Rao lower bound which is the lower bound on the variance of unbiased estimators. Hence in the sense of minimum asymptotic variance, the MLE is the best asymptotic unbiased estimator. In practice, there are various popular statistical software packages that provide tools for computing MLE for many of the broadly used probability models.

The problem is that it is not always possible to have an exact model. In this case, the MLE is impossible to be implemented. Also, in many practical problems, researchers tend to use SCM due to Gaussian assumptions justified the central limit theorem. However, if the asymptotic normality of the data set does not hold or the exact distribution of the data set is far away from the asymptotic distribution, the performance of the SCM may not be optimal or even not be acceptable. Thus it is important to find estimators insensitive to the model assumption with "good" performance under a large range of model assumptions. M-estimator is a good candidate.

2.2 M-estimator of Scatter Matrix

The robustness we consider in this thesis includes only robustness to heavy-tailedness. Thus the underlying class of distributions we are interested in is elliptically symmetric (ES) families (real or complex). Elliptical symmetric distributions encompass multivariate normal distribution, multivariate t -distribution, K -distribution, generalized

Gaussian distribution, etc. as special cases. The probability density function of an elliptical distribution is in the form of,

$$f(x; \mu, C_p) = |C_p|^{-\frac{1}{2}} g\{(x - \mu)^T C_p^{-1} (x - \mu)\}, \quad (2.1)$$

where g is a positive valued function such that f integrates to one; μ is a p -dimensional location parameter, and C_p is a positive definite symmetric (PDS) matrix of dimension $p \times p$. If X is ES distributed, we can decompose this random variable by $X = \sqrt{\tau} \times C_p^{\frac{1}{2}} \times \frac{z}{\|z\|}$, where z is standard Gaussian distributed, τ is a random scalar independent of z and C_p is the scatter matrix. To remove the ambiguity, the scatter matrix is selected such that $\text{tr}C_p = p$. In the case of complex elliptically symmetric (CES) distribution [1], z is standard complex Gaussian distributed.

Here are some examples of ES distributions. In Gaussian case, $g(d) = c \exp(-d/2)$ with $c = (2\pi)^{-p/2}$. An important example of a non-Gaussian elliptical distribution is the p -variate Student-t distribution with ν degree of freedom by choosing

$$g(d) = \frac{c}{(d + \nu)^{(p+\nu)/2}}.$$

The Generalized Gaussian distribution with shape parameter s is generated by the choice of density generator function,

$$g(d) = c \exp\left(-\frac{d^s}{2m^s}\right),$$

where m and s are the scale, shape parameter and c is a normalizing constant. It is interesting to know that with $s = 1$, the distribution corresponds to the multivariate Gaussian distribution and with $s = 0.5$, the distribution corresponds to the multivariate Laplace distribution.

In this thesis, we assume the location parameter μ is known to be a zero vector.

Thus the model reduced to the form of,

$$f(x; C_p) = |C_p|^{-\frac{1}{2}} g\{x^T C_p^{-1} x\}.$$

Our objective is to estimate robustly the parameter C_p . For the whole class of elliptical distribution, C_p is generally the scatter matrix but not necessarily the covariance matrix. If the covariance matrix of a ES distribution exists, C_p is proportional to it.

The M-estimator of the scatter matrix is a generalization of the MLE for the scatter matrix of ES distributions, 2.1, introduced by Maronna [24].

Let $x = \{x_1, \dots, x_n\}$ be an i.i.d sample from an p-dimensional ES distribution. The maximum likelihood estimator of C_p depending on the sample x , denoted by $C(x)$, maximizes the likelihood function,

$$\mathcal{L}(C_p) = \frac{1}{|C_p|^{n/2}} \prod_{i=1}^n g\{x_i^T C_p^{-1} x_i\},$$

and is equivalent to the minimizer of

$$2 \log \mathcal{L}(C_p) = n \log |C_p| + \sum_{i=1}^n \rho(x_i^T C_p^{-1} x_i), \quad (2.2)$$

where $\rho(d) = -2 \log g(d)$. Critical points are then solutions to the fixed point equation

$$C(x) = \frac{1}{n} \sum_{i=1}^n u(x_i^T C(x)^{-1} x_i) x_i x_i^T. \quad (2.3)$$

M-estimators of the scatter matrix can be defined by allowing a general u function in (2.3), not necessarily derived from any g function in ES distribution. In the following, we use both $C(x)$ and C to denote estimator of the scatter matrix. When we need not emphasize the estimator is a function of data x , we use C . We may interpret $C(x)$ as a weighted SCM with weights depending on a distance measure $d_i = x_i^T C^{-1} x_i$. d_i is

square of estimated Mahalanobis distance which is a measure of the distance between a point P and its center (0 here). The weight function u is chosen to be non-negative, continuous and non-increasing. Some examples of M-estimators are given below.

SCM. In the Gaussian case, $g(d) = c \exp(-d/2)$, $\rho(d) = d$ and $u(d) = \rho'(d) = 1$, so (2.3) becomes

$$C_{SCM} = C = \frac{1}{n} \sum_{i=1}^n x_i x_i^T.$$

Tyler's M-estimator [25, 1, 26] is derived by choosing

$$\rho(d) = d \ln d \quad \text{and} \quad u(d) = \rho'(d) = \frac{p}{d}.$$

This weight function is not related to any ES distribution and can be computed via simple fixed point-iterations. Tyler's M-estimator is the "most robust" estimator of the scatter matrix for elliptical distributions with the minimized maximum asymptotic variance. [25] In practice, it has also been shown to outperform the SCM in many applications.

The existence and uniqueness of the estimator is guaranteed by $x_1 \dots x_n \in \mathbb{C}^{p \times 1}$ such that $n > p$ and $\text{span}(x_i) = \mathbb{C}^p$. Tyler's M-estimator gives the "shape" or relative magnitudes of the scatter matrix but is missing its magnitude. However, for some applications, such as spectrum sensing, the "shape" of the covariance suffices. Compared to SCM, Tyler's M estimator is more robust to heavy-tailed data as shown in the Example latter. This estimator does not have a closed form and has to be iteratively computed by using fixed-point iteration. The following algorithm converges to a unique solution for the equation (2.3) with $u(d) = \frac{p}{d}$.

Algorithm 1 Algorithm for Tyler's M-estimator

- 1: Initialize C_0 as an arbitrary positive definite matrix.
- 2: Iterate

$$\tilde{C}_{t+1} = \frac{p}{n} \sum_{i=1}^n \frac{x_i x_i^T}{x_i^T \tilde{C}_t^{-1} x_i}.$$

$$C_{t+1} = \frac{\tilde{C}_{t+1}}{\text{Tr } \tilde{C}_{t+1}}$$

until convergence.

A special case of *Maronna's M-estimator* is derived from real multivariate student-t distribution,

$$u(d) = \frac{p+k}{d+k},$$

where k is a tuning constant that controls the balance of robustness and efficiency under Gaussian distribution of the estimator. Note that for $k \rightarrow \infty$, this estimator approaches the SCM, and for $k \rightarrow 0$, the estimator approaches Tyler's M-estimator. The existence and uniqueness of the solution is guaranteed by $x_1, \dots, x_n \in \mathbb{C}^{p \times 1}$ such that $n > p$, $\text{span}(x_i) = \mathbb{C}^p$ and $k > p^2 - p$. Similar to Tyler's M estimator, this estimator misses its magnitude but only gives the relative magnitude.

Algorithm 2 Algorithm for Maronna's M-estimator

- 1: Initialize C_0 as an arbitrary positive definite matrix. Determine the value of k .
- 2: Iterate

$$C_{t+1} = \frac{k+p}{n} \sum_{i=1}^n \frac{x_i x_i^T}{x_i^T C_t^{-1} x_i + k}.$$

until convergence.

2.3 Measurement of Robustness

There are some metrics of robustness for an estimator [5], $\hat{\theta}_n$ depending on an i.i.d sample $x = \{x_1, \dots, x_n\}$ of size n with distribution F . Before introducing those metrics, we clarify some notations. For $\hat{\theta}_n$, there is a value, $\hat{\theta}_\infty = \hat{\theta}_\infty(F)$ depending on F ,

such that

$$\hat{\theta}_n \xrightarrow{p} \hat{\theta}_\infty(F),$$

with dimension of the data p fixed and the sample size $n \rightarrow \infty$. $\hat{\theta}_\infty(F)$ is the asymptotic value of the estimate at F . We define the *contamination neighborhoods*:

$$\mathcal{F}(F, \epsilon) = \{(1 - \epsilon)F + \epsilon G : G \in \mathcal{G}\},$$

where \mathcal{G} is a suitable set of all possible distributions and $0 \leq \epsilon \leq 1$ is the portion of contamination. In some case, \mathcal{G} is the set of point mass distributions, where the point mass distribution δ_{x_0} is the distribution such that $P(x = x_0) = 1$. The first metrics is the *influence function* (IF) defined as

$$\begin{aligned} \text{IF}_{\hat{\theta}}(x_0, F) &= \lim_{\epsilon \rightarrow 0^+} \frac{\hat{\theta}_\infty((1 - \epsilon)F + \epsilon\delta_{x_0}) - \hat{\theta}_\infty(F)}{\epsilon} \\ &= \frac{\partial}{\partial \epsilon} \hat{\theta}_\infty((1 - \epsilon)F + \epsilon\delta_{x_0}), \end{aligned}$$

where $\lim_{\epsilon \rightarrow 0^+}$ stands for limit from the right. If $\hat{\theta}_\infty$ is an estimator for p -dimensional parameter, IF is a p -dimensional vector. The IF may be considered as an asymptotic version of the standardized sensitivity curve,

$$\frac{\theta_{n+1}(x_1, \dots, x_n, x_0) - \theta_n(x_1, \dots, x_n)}{1/(n+1)},$$

which is a function of x_0 . The standardized sensitivity curve is the difference made by a single additional observation standardized by the portion of contamination $\epsilon = 1/(n+1)$ similar in IF.

Another measure of robustness is the *breakdown point* (BP) of an estimator, which is the largest proportion of atypical points that the data may contain such that $\hat{\theta}$ still conveys some information about the distribution of the "typical" data. The

definition of asymptotic contamination BP will be defined by the largest portion of contamination $\epsilon^* \in (0, 1)$ such that for $\epsilon < \epsilon^*$, $\hat{\theta}_\infty((1 - \epsilon)F + \epsilon G)$ as a function of G remains bounded and also bounded away from the boundary of parameter space.

There are also many other measurements for robustness, such as Finite-sample breakdown point and maximum asymptotic bias. Further detail can be found in [5].

For robust estimator of covariance matrix or scatter matrix generally, there are some desirable properties[27]. We require the estimator to be *affine equivariant*. A PDS scatter matrix estimate $C(X_n)$ is affine equivariant if

$$C(AX_n + b) = AC(X_n)A^T,$$

where X_n is a p -dimensional data set with n observations, A is a non-singular $p \times p$ matrix, b is a p -dimensional shift vector and $AX + b = \{Ax_1 + b, Ax_2 + b, \dots, Ax_n + b\}$. This property is useful in the operation related to rotation and scaling of the axis such as principle component analysis.

We also require the estimator to be consistent. A scatter matrix estimate $C(X_n)$ is consistent if

$$C(X_n) \xrightarrow{p} C_p,$$

where \xrightarrow{p} indicates convergence in probability.

2.4 A Toy Example

Let $\{x_1, \dots, x_{100}\}$ be an i.i.d sample from a multivariate student-t distribution with location parameter $\mu = (0, 0)^T$, the covariance matrix $C_p = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$, and degree of freedom $\nu = 3$. We have no information about the underlying distribution of the data set and it is not possible to learn it with a limited sample size. Let's compare the performance of SCM and Tyler's M-estimator under this circumstance. This ex-

ample shows that SCM has very low efficiency. Figure 2.1 is the empirical probability density function for each component of SCM and Figure 2.2 is the empirical probability density function for each component of Tyler's M-estimator. The empirical density function of the sample covariance matrix has a large range of estimate and high variance. We conclude that the SCM breaks down in this case in the sense of variance and it always produces unreasonable estimation. Since the SCM fails in this case, we may conclude that the data distribution is not Gaussian and we need other estimators. Although the SCM is centered at the true value, i.e., unbiased, the variance is too high compared to Tyler's M-estimator. If we apply SCM in this case, with low probability, we can have a reasonable estimation of the covariance matrix. The Tyler's M-estimator varies in a reasonable range around the true parameter, i.e., 1 for diagonal entries and 0.5 for off-diagonal entries. If we apply Tyler's M-estimator in this case, with high probability, we can have a reasonable estimation of the covariance matrix.

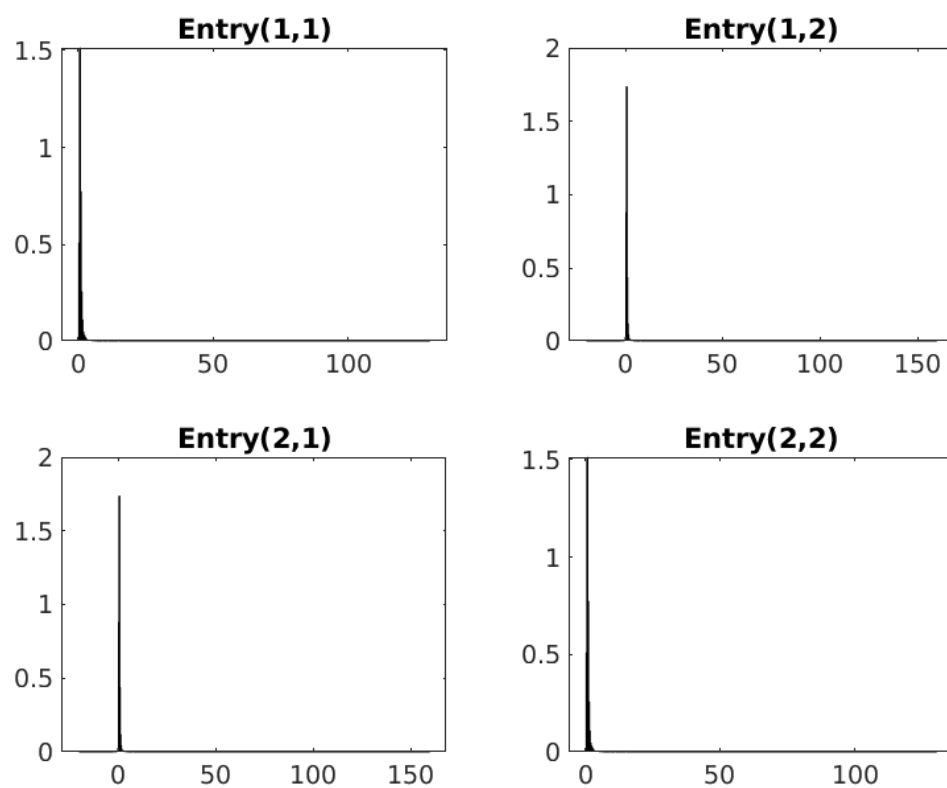


Figure 2.1: Empirical Probability Density of the Sample Covariance Matrix

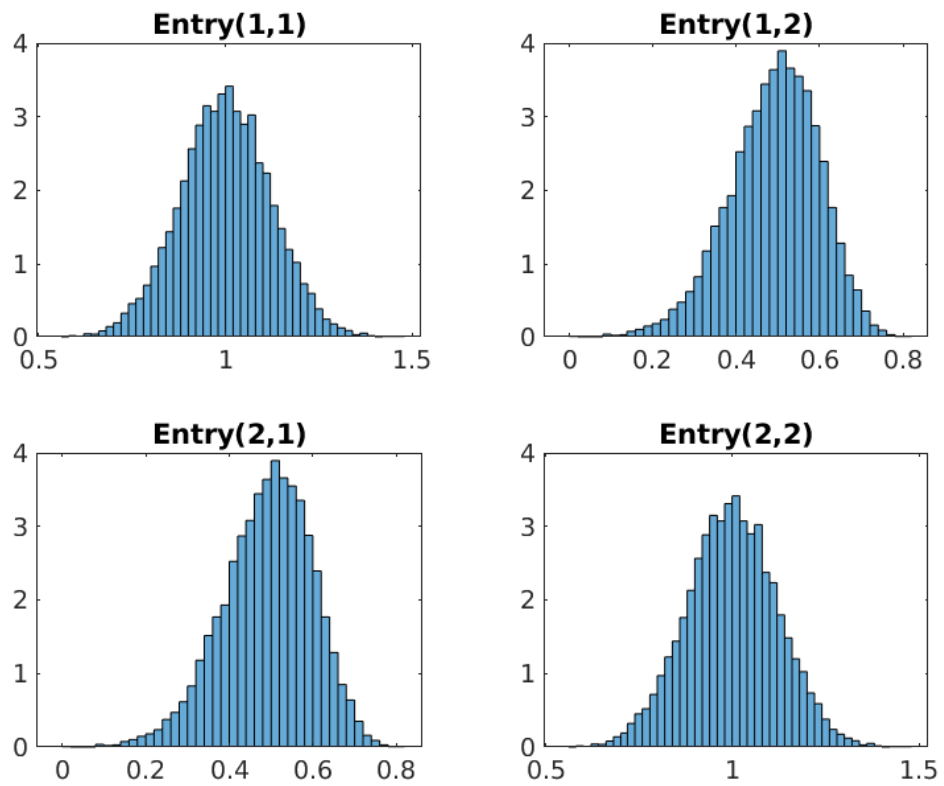


Figure 2.2: Empirical Probability Density of the Tyler Estimator of the Covariance Matrix

Chapter 3

Random Matrix Theory and Estimators of the Scatter Matrix

In this chapter, we mainly focus on several important results of random matrix theorem (RMT) related to the SCM and M-estimators. The topics in RMT that have a close relationship to spectrum sensing are the behavior of the eigenvalues of estimated scatter matrices, specifically the empirical spectral distribution and the maximum eigenvalue. Suppose that \mathbf{X} is an positive definite $N \times N$ matrix with eigenvalues $\lambda_1, \dots, \lambda_N \in \mathbb{R}$. The empirical distribution of the eigenvalues of \mathbf{X} , called *empirical spectral distribution* (ESD) of \mathbf{X} , is the function,

$$F^{\mathbf{X}}(x) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{\lambda_i \leq x\}}, \quad (3.1)$$

for $x \in \mathbb{R}$.

Suppose $x = \{x_1, \dots, x_n\}$ are generated independently by a p -dimensional probability distribution, a scatter matrix estimator, $C(x)$, calculated according to these data is a random matrix and its ESD and maximum eigenvalue are also random function and random variable. It is often very difficult or impossible to study those random objects directly in the finite case. Thus, we turned to study them in an asymptotic framework to have some useful asymptotic properties. It is often much easier to study a complex statistics in an asymptotic framework than it in finite case. Asymptotic properties are also relatively independent of the data distribution. Thus, we may always apply asymptotic results on a large class of data. The asymptotic

framework is usually assuming that the sample size n grows to infinity with fixed dimension p , while in RMT, we require not only $n \rightarrow \infty$ but also $p(n) \rightarrow \infty$ with condition,

$$\lim_{n \rightarrow \infty} \frac{p}{n} \rightarrow \alpha \in (0, \infty). \quad (3.2)$$

An application of the asymptotic trick is in the hypothesis test. To perform a hypothesis test, we need to know the distribution of the test statistic given the null or alternative hypothesis is true such that we can control probability of errors. In most of the case, it cannot be figured out exactly, because the distribution of the statistics is often analytically intractable or we only have partial information of the statistical model. Hence, an approximation of the distribution is required. Usually, we use the asymptotic distribution of the statistics as the approximation.

Before getting into the RMT asymptotic framework, let's take an example in the ordinary asymptotic framework. Suppose we have a sample of independent observations $x = \{x_1, \dots, x_n\}$ from an unknown distribution and all the elements are scalar, we wish to test on the null hypothesis $H_0 : \mu = \mu_0$ and alternative hypothesis $H_a : \mu \neq \mu_0$. The famous t -test is based on the statistics,

$$T = \frac{\sqrt{n}(\bar{x} - \mu_0)}{\hat{\sigma}}, \quad (3.3)$$

where $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ and $\hat{\sigma} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$. If the sample is generated by the Gaussian distribution, the distribution of T is t -distribution with $n - 1$ degrees of freedom exactly. However, if the data is not generated by the Gaussian distribution, the exact distribution of T is unknown. It can be proved that T is standard Gaussian distributed as $n \rightarrow \infty$ if the random samples have a finite second moment. Thus when n is fairly large, we can treat T as Gaussian distributed. The performance of the test will depend on both sample size n and the real distribution generating the data set. In the point of robustness, this statistic is not quite robust to skewness and

heavy-tailedness. When the data is generated by skewed or heavy-tailed distribution, T may need larger sample size to have approximately Gaussian distribution.

We will introduce some asymptotic results regarding ESD and maximum eigenvalue of the estimated scatter matrix in this chapter. Under certain condition and (3.2), the ESD of estimated scatter matrix converges to a deterministic distribution, and the maximum eigenvalue of the estimated scatter matrix has an asymptotic distribution. The asymptotic results are quite reliable even with very small sample size and only moments conditions are required.

3.1 Sample Covariance Matrix

Suppose $x = \{x_1, \dots, x_n\}$ are generated independently by a p -dimensional probability distribution. Now we stack all the observation horizontally to make \mathbf{X} , a $p \times n$ data matrix. Then the $p \times p$ matrix

$$C_{SCM} = \frac{1}{n} \mathbf{X} \mathbf{X}^T \quad (3.4)$$

is the (uncentered) SCM defined in the previous lecture. The distribution of unnormalized version $nS = \mathbf{X} \mathbf{X}^T$ is the *Wishart distribution* $W_p(n, C_p)$ if each x_i with $i = 1, \dots, n$ is independently distributed by $N_p(0, C_p)$. Detailed study on this distribution can be found in [28].

3.1.1 Asymptotic Empirical Spectral Distribution

It has been showed that the ESD of the SCM has a nonrandom limit depending on the limiting ratio $\alpha = \lim_{n \rightarrow \infty} \frac{p}{n}$. This result was derived by Marčenko and Pastur[29] assuming that the fourth moments of the entries of X are finite. After development by [30, 31] and [32], the conditions on the matrix entries for which the ESD has a deterministic limit has been weakened. The following theorem has the minimal

moment conditions.

Theorem 1. *Consider a $p \times n$ matrix \mathbf{X} whose entries are independently drawn from a zero-mean real(or complex) distribution with variance σ^2 . As $p, n \rightarrow \infty$ with $\frac{p}{n} \rightarrow \alpha$, the empirical distribution of $\frac{1}{n}\mathbf{X}\mathbf{X}^T$ converges almost surely to a deterministic distribution, known as Marčenko-Pastur law, with density*

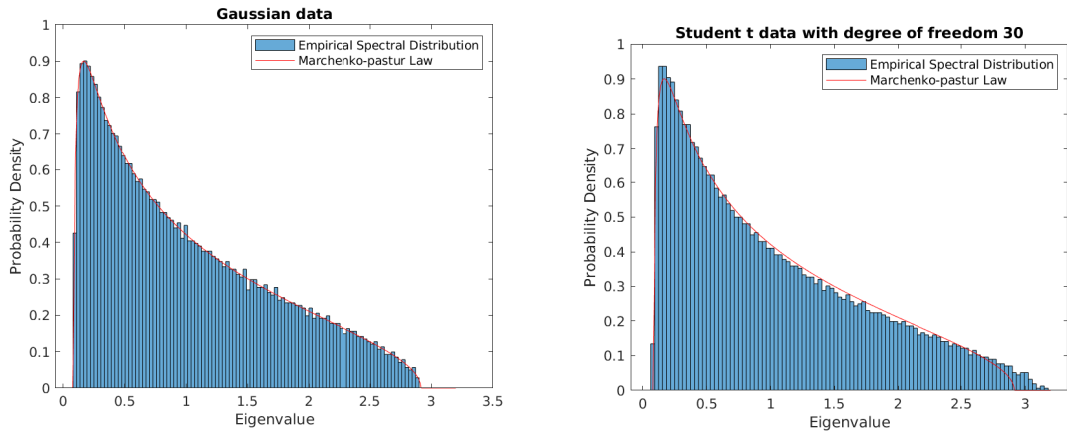
$$f(x) = \left(1 - \frac{1}{\alpha}\right)^+ \delta(x) + \frac{\sqrt{(x-a)^+(b-x)^+}}{2\pi\alpha x}$$

where $a = \sigma^2(1 - \sqrt{\alpha})^2$ and $b = \sigma^2(1 + \sqrt{\alpha})^2$.

Theorem 1 gives an approximation of ESD of SCM given the entries of the data matrix have zero mean and finite variance in closed form. We are also interested in how well this approximation is when the sample size n is finite especially when the entries of the data matrix is heavy-tailed distributed. Figure 3.1 shows the influence of heavy-tailedness on the performance of the approximation by simulation. In the most extreme case that the distribution of data matrix entries has an infinite variance or no defined variance, Theorem 1 does not hold as shown in Figure 3.1d. When the data distribution has finite variance and it is also heavy-tailed, the support of the Marčenko-Pastur law no longer bounds all the eigenvalues as shown in Figure 3.1b and Figure 3.1c.

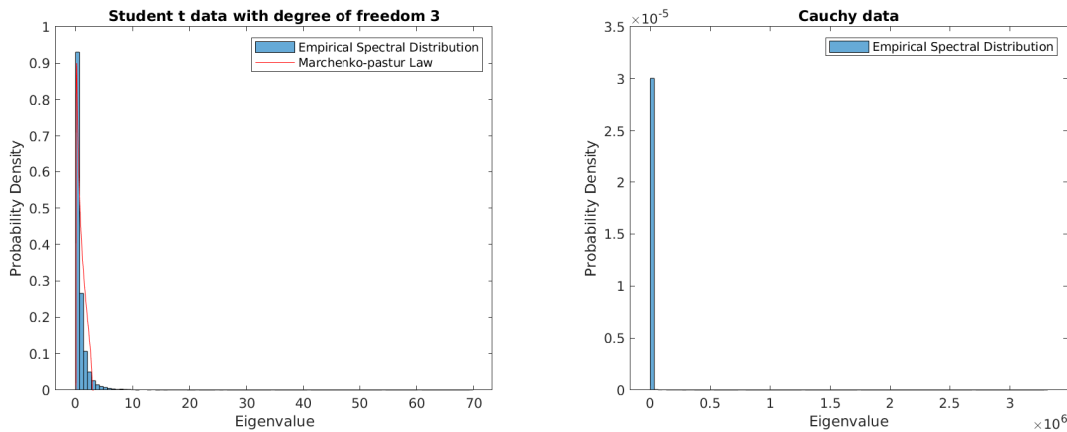
3.1.2 Asymptotic Distribution of the Largest Eigenvalue

Although we know the maximum eigenvalue of the SCM converges to the right edge of the Marčenko-Pastur law[13], we do not have the distribution of the largest eigenvalue. Tracy and Widom [33, 14] derived an equation which can be used to describe the asymptotic distribution of the largest eigenvalue of SCM. The c.d.f of the so called



(a) The approximation is very good.

(b) The approximation is not that good.



(c) The approximation is bad.

(d) The eigenvalues are unbounded.

Figure 3.1: ESD of SCM under different distributions

The data matrix is of size 10000×5000 ; data of (a) is generated by a Gaussian distribution with zero mean and unit variance; data of (b) is generated by a student t distribution with zero mean, unit variance and degree of freedom 30; (c) is generated by a student t distribution with zero mean and degree of freedom 3; data of (d) is generated by a standard Cauchy distribution with mean and variance undefined.

Tracy-Widom laws denoted by F_1 and F_2 is,

$$F_1(s) = \exp\left(-\frac{1}{2} \int_s^\infty (q(x) + (x-s)q^2(x))dx\right), \quad s \in \mathbb{R} \quad (3.5)$$

and

$$F_2(s) = \exp\left(-\int_s^\infty (x-s)q^2(x)dx\right), \quad s \in \mathbb{R} \quad (3.6)$$

where F_1 is corresponding to real entries, F_2 is corresponding to complex entries, $q(x)$ satisfies the Painlevé II differential equation $q''(x) = xq(x) + 2q^3(x)$ with the feature that $q(x) - A(x) \rightarrow 0$ as $x \rightarrow \infty$, where $A(x)$ is the Airy function with details in [34]. Here is the result in terms of these distributions [12],

Theorem 2. *Consider an $p \times n$ matrix \mathbf{X} whose entries are independently sampled from a real Gaussian distribution with zero mean and variance one. Let λ_1 denote the largest eigenvalue of $\mathbf{X}\mathbf{X}^T$. If $p, n \rightarrow \infty$ with $\frac{p}{n} \rightarrow \alpha \in (0, 1]$, then*

$$\frac{\lambda_1 - \mu}{\sigma} \xrightarrow{d} W_1$$

where $\mu = (\sqrt{n-1} + \sqrt{p})^2, \sigma = (\sqrt{n-1} + \sqrt{p})\left(\frac{1}{\sqrt{n-1}} + \frac{1}{\sqrt{p}}\right)^{1/3}$.

If the entries of \mathbf{X} is complex Gaussian distributed with zero mean and variance one and λ_1 denotes the largest eigenvalue of $\mathbf{X}\mathbf{X}^H$, then

$$\frac{\lambda_1 - \mu'}{\sigma'} \xrightarrow{d} W_2$$

where $\mu' = (\sqrt{n} + \sqrt{p})^2$ and $\sigma' = (\sqrt{n} + \sqrt{p})\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{p}}\right)^{1/3}$, here the random variable W_1 and W_2 have distributions with c.d.f. F_1 and F_2 respectively.

Theorem 2 gives the asymptotic distribution for the scaled largest eigenvalue of the SCM. This theorem requires the entries are Gaussian distributed which is not as universal as Theorem 1. Figure 3.2 depicts the Tracy-Widom density function for

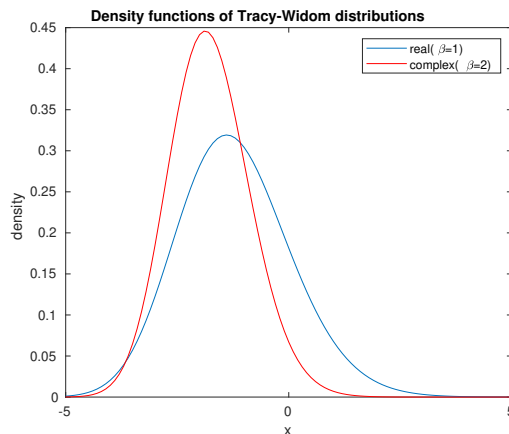


Figure 3.2: Tracy-Widom density functions for $\beta = 1, 2$.

$\beta = 1, 2$. [35] has a similar result for non-Gaussian distributed data entries.

3.2 Tyler's M-estimator and Maronna's M-estimator

Suppose $x = \{x_1, \dots, x_n\}$ are generated independently by a p -dimensional probability distribution. Now we stack all the observation horizontally to make \mathbf{X} , a $p \times n$ data matrix. Then the $p \times p$ matrix C_{TY} , defined by the following equation

$$C_{TY} = \frac{1}{n} \sum_{i=1}^n \frac{x_i x_i^T}{x_i^T C_{TY}^{-1} x_i}. \quad (3.7)$$

C_{TY} is Tyler's M-estimator defined in Chapter 2. It has been proved in [36] that when the data entries are identically Gaussian distributed with mean zero and unit variance, the operator norm of the difference between a scaled Tyler's converges to zero. As a result, the ESD of Tyler's M-estimators converges to Marčenko-Pastur law in distribution. This result can also be extended to the case that data is generated by a symmetric elliptical distribution.

Figure 3.3 shows the influence of the heavy-tailedness on the performance of the approximation by simulation. In all cases, the Marčenko-Pastur law fits well to the ESD of Tyler's M-estimator. If we apply Tyler's M-estimator to this case, with high

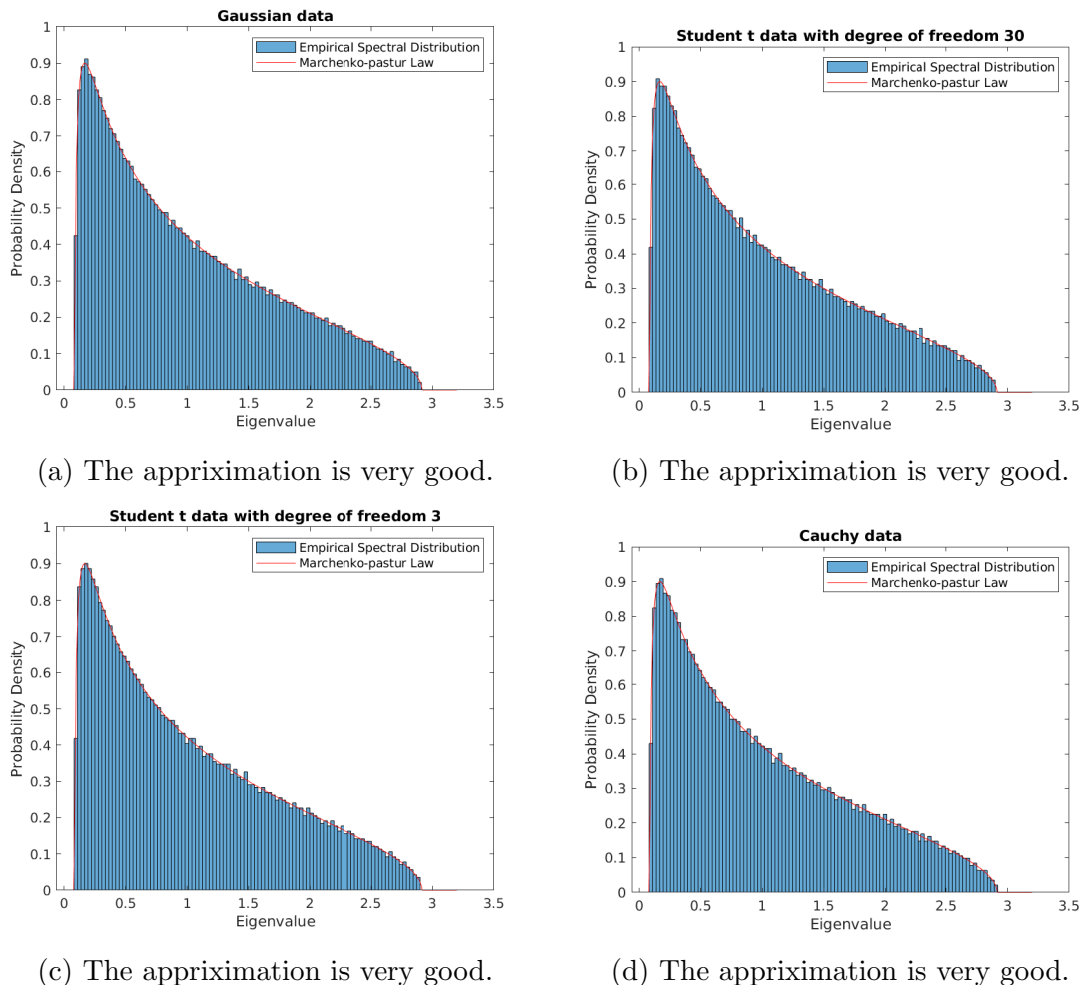


Figure 3.3: ESD of Tyler's M-estimator under different distributions

The dataset is generated in the same way as Figure 3.1

probability, we can have a reasonable estimation of the scatter matrix. Tyler's M-estimator is robust when data are elliptical symmetric distributed since the texture parameter will be canceled during the calculation.

There is no published result regarding Marčenko-Pastur law for Maronna's M-estimators. In the preprint [37], the researchers claimed the ESD of Maronna's M-estimators converges to Marčenko-Pastur law, but they deleted the result in the published version. Even though the result has not been proved rigorously, the result seems to be correct in the case of $u(d) = \frac{p+k}{d+k}$.

3.3 Spike Model

All the theory above are based on the assumption that all the columns of the data matrix are i.i.d, which means each column is drawn from a multivariate distribution with the identity matrix as the covariance matrix. It is also interesting to know the case that the data is generated by a distribution with non-identity population covariance matrix. One of the simplest non-identity matrices is the matrix with a finite rank perturbation of the identity matrix, which is called the "spiked population model"[38]. Let's take the population covariance matrix with one rank perturbation as an example. Suppose the non-unit eigenvalue of the population matrix is σ_1 . As the dimension p becomes large, one would expect that if σ_1 is around 1, σ_1 would have an ignorable effect on the sample covariance matrix. If σ_1 is much greater than 1, then even if p becomes large, σ_1 might pull one eigenvalue of SCM out of the bulk. It has been proven that if $\sigma_1 > 1 + \sqrt{\alpha}$, one eigenvalue of the SCM will be pulled out of the bulk and converges to $\sigma_1 + \frac{\alpha\sigma_1}{\sigma_1-1}$.

The spike model is specified in detail. Let C be a fixed $p \times p$ semi-definite symmetric(hermitian) matrix. Let W be a $p \times n$ data matrix, whose entries are i.i.d real(complex) random variables with

$$E(W_{ij}) = 0, \quad E(|W_{ij}|^2) = 1 \quad \text{and} \quad E(|W_{ij}|^4) < \infty,$$

and let W_j be j th column of W . Note that $X = C^{\frac{1}{2}}W_j$ will be a p -dimensional random vector with covariance matrix C , where $C^{\frac{1}{2}}$ is a Hermitian square root of C . Then, C is the population covariance matrix of the data vector. When W_{ij} is i.i.d Gaussian distributed, the sample vector becomes multivariate Gaussian distributed with population covariance matrix C . Except for Gaussian distribution, this model can produce many random vectors with desired population covariance matrix. Now

SCM of X becomes

$$C_{SCM} = \frac{1}{n} C^{\frac{1}{2}} X X^H C^{\frac{1}{2}}. \quad (3.8)$$

Let the ordered eigenvalue of C_{SCM} be $\lambda_1 \geq \lambda_2 \dots \geq \lambda_p \geq 0$. Let $\sigma_1 > \dots > \sigma_M > 0$ be fixed real numbers for a fixed $M \geq 0$. Let $k_1 \dots k_M \geq 0$ be fixed integers and set $r = k_1 + \dots k_M$. The spiked model will be defined by assuming that the first r eigenvalues of C are equal to $\sigma_1 \dots \sigma_M$ with multiplication $k_1 \dots k_M$, respectively and the rest of eigenvalues are 1. Then

$$\text{eig}(C) = (\underbrace{\sigma_1, \dots, \sigma_1}_{k_1}, \underbrace{\sigma_2, \dots, \sigma_2}_{k_2}, \dots, \underbrace{\sigma_M, \dots, \sigma_M}_{k_M}, \underbrace{1, \dots, 1}_{p-r}),$$

where $\text{eig}(\ast)$ is the operator to compute eigenvalues. Set $k_0 = 0$.

Theorem 3. (Case $\alpha = \frac{p}{n} < 1$) Assume that $n \rightarrow \infty$ and $p \rightarrow \infty$ such that

$$\frac{p}{n} \rightarrow \alpha.$$

Let M_0 be the number of j 's such that $\sigma_j > 1 + \sqrt{\alpha}$, and let M_1 be the number of j 's such that $\sigma_j > 1 - \sqrt{\alpha}$. Then the following holds:

1. For each $1 \leq j \leq M_0$,

$$\lambda_{k_1 + \dots + k_{j-1} + i} \rightarrow \sigma_j + \frac{\alpha \sigma_j}{\sigma_j - 1}, \quad 1 \leq i \leq k_j$$

almost surely.

- 2.

$$\lambda_{k_1 + \dots + k_{M_0} + 1} \rightarrow (1 + \sqrt{\alpha})^2$$

almost surely.

3.

$$\lambda_{p-r+k_1+\dots+k_{M_1}} \rightarrow (1 - \sqrt{\alpha})^2$$

almost surely.

4. For each $M_1 + 1 \leq j \leq M$,

$$\lambda_{k_1+\dots+k_{j-1}+i} \rightarrow \sigma_j + \frac{\alpha\sigma_j}{\sigma_j - 1}, \quad 1 \leq i \leq k_j$$

almost surely.

In this case, the eigenvalues of SCM are split into three groups according to the eigenvalues of the population covariance matrix. The number of eigenvalues of SCM outside the support $[(1 - \sqrt{\alpha})^2, (1 + \sqrt{\alpha})^2]$ of the Marčenko-Pastur law equal to the number of eigenvalues of population covariance matrix out side the boundary $[1 - \sqrt{\alpha}, 1 + \sqrt{\alpha}]$. What's more, those outliers converge to $\sigma + \frac{\alpha\sigma}{\sigma-1}$. The rest of the eigenvalues of SCM will be bounded by the support of Marčenko-Pastur law in the limit.

For instance, when $r=1$ and the only non-unit eigenvalues of population covariance matrix is σ_1 , the largest sample eigenvalues λ_1 satisfies

$$\lambda_1 \rightarrow \begin{cases} (1 + \sqrt{\alpha})^2 & \sigma_1 \leq 1 + \sqrt{\alpha} \\ \sigma_1 + \frac{\alpha\sigma_1}{\sigma_1-1} & \sigma_1 > 1 + \sqrt{\alpha} \end{cases}$$

almost surely.

Theorem 4. (Case $\alpha = \frac{p}{n} > 1$) Assume that $n \rightarrow \infty$ and $p \rightarrow \infty$ such that

$$\frac{p}{n} \rightarrow \alpha$$

. Let M_0 be the number of j 's such that $\sigma_j > 1 + \sqrt{\alpha}$. Then the following holds:

1. For each $1 \leq j \leq M_0$,

$$\lambda_{k_1+\dots+k_{j-1}+i} \rightarrow \sigma_j + \frac{\alpha\sigma_j}{\sigma_j - 1}, \quad 1 \leq i \leq k_j$$

almost surely.

2.

$$\lambda_{k_1+\dots+k_{M_0}+1} \rightarrow (1 + \sqrt{\alpha})^2$$

almost surely.

3.

$$\lambda_n \rightarrow (1 - \sqrt{\alpha})^2$$

almost surely.

4. For all p , $\lambda_{n+1} = \dots = \lambda_p = 0$.

In this case, small eigenvalues of C have no influence on the eigenvalues of SCM.

Theorem 5. (Case $\alpha = 1$) Assume that $n \rightarrow \infty$ and $p \rightarrow \infty$ such that

$$\frac{p}{n} \rightarrow 1$$

. Let M_0 be the number of j 's such that $\sigma_j > 2$. Then the following holds:

1. For each $1 \leq j \leq M_0$,

$$\lambda_{k_1+\dots+k_{j-1}+i} \rightarrow \sigma_j + \frac{\sigma_j}{\sigma_j - 1}, \quad 1 \leq i \leq k_j$$

almost surely.

2.

$$\lambda_{k_1+\dots+k_{M_0}+1} \rightarrow 4$$

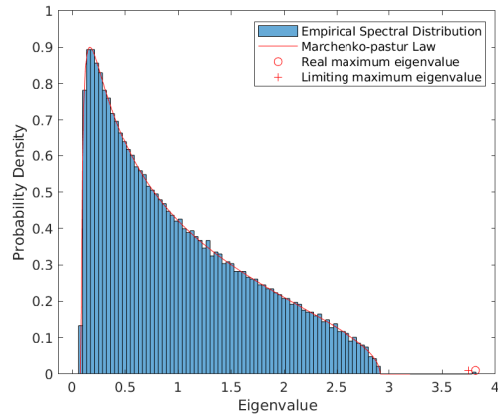
almost surely.

3.

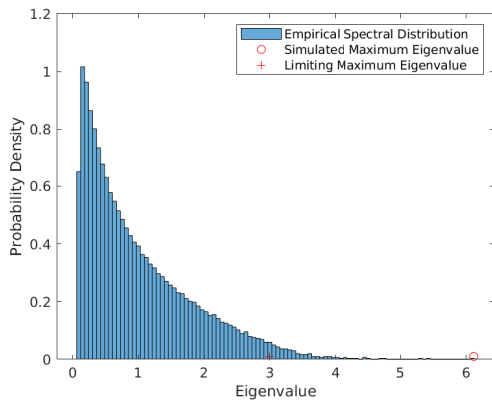
$$\lambda_{\min\{p,n\}} \rightarrow 0$$

almost surely.

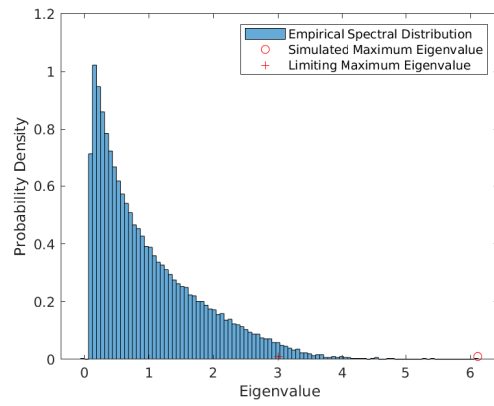
We include some plots for the case when $\alpha = 0.5$ and there is only one non-unit eigenvalue in population covariance matrix given by 2. In this case, the critical value of the eigenvalues are $1 + \sqrt{0.5}$ and $1 - \sqrt{0.5}$. We will expect that the maximum eigenvalue of the SCM $4 + \frac{2}{3} \approx 4.6667$ is outside of the interval $[(1 - \sqrt{0.5})^2, (1 + \sqrt{0.5})^2] \approx [0.08578, 2.91422]$. The histogram of Figure 3.4a is from the Gaussian samples when $p = 5000$ and $n = 10000$. We are also interested in the case when the distribution of samples having heavy tails. Figure 3.4b is from student's t distribution with degree of freedom 10 and Figure 3.4c is the same realization with the identity covariance matrix when $p = 5000$ and $n = 10000$. We can found there is no obvious difference between the two cases. However, if we replace SCM by tyler's M estimator, we can have the only one eigenvalue far away from the bulk, which is depicted in Figure 3.4d.



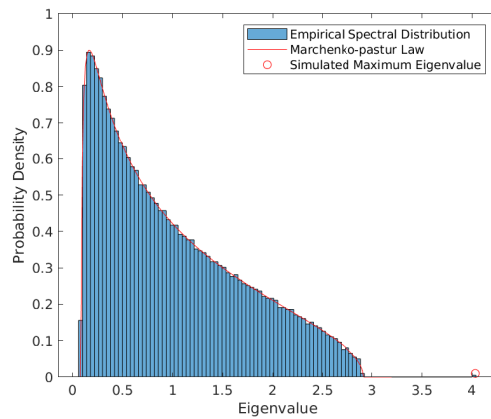
(a) SCM under Gaussian noise



(b) SCM under Student's t distribution (Spike model)



(c) SCM under Student's t distribution (Identity covariance)



(d) Tyler's M-estimator under Student's t distribution

Figure 3.4: Spike Model

Chapter 4

Application on Spectrum Sensing in Cognitive Radio Networks

The simplest version of the spectrum sensing is detection of a signal from a noisy environment. This task can be formulated by a hypothesis test, whose null hypothesis is that a signal does not exist and the alternative hypothesis is that a signal exists. The received signal samples under two hypothesis are formulated as,

$$x(i) = \begin{cases} z(i) & H_0 : \text{signal does not exist} \\ hs(i) + z(i) & H_1 : \text{signal exists,} \end{cases} \quad (4.1)$$

where $x(i)$ is the received sample vector at instant i , h represents the fading component including the effects of path loss and multipath fading, $s(i)$ is the transmitted symbol which we want to detect, and $z(i)$ is the received noise vector which is assumed to be i.i.d in time, with mean zero and variance σ^2 not necessarily Gaussian distributed. We assume channel h being constant during $i = 1..n$ transmissions.

The system [20] is depicted in Figure 4.1, in which primary users (in white) communicate to their dedicated (primary) base station. Secondary base stations, $BS_1, BS_2, BS_3, \dots, BS_p$, are cooperatively sensing the channel to identify a free spectrum and exploit it. There are some assumptions on this system:

1. The p secondary base stations are communicated by high speed and low noise medium.
2. The stations are analyzing the same portion of the spectrum.

3. Channels from each base station to the primary station are independent.

The received sample matrix generated by the system is a $p \times n$ matrix consisting of all the sample vectors from p secondary base stations:

$$X = \begin{bmatrix} x_1(1) & x_1(2) & x_1(3) & \dots & x_1(n) \\ x_2(1) & x_2(2) & x_2(3) & \dots & x_2(n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_p(1) & x_p(2) & x_p(3) & \dots & x_p(n) \end{bmatrix}.$$

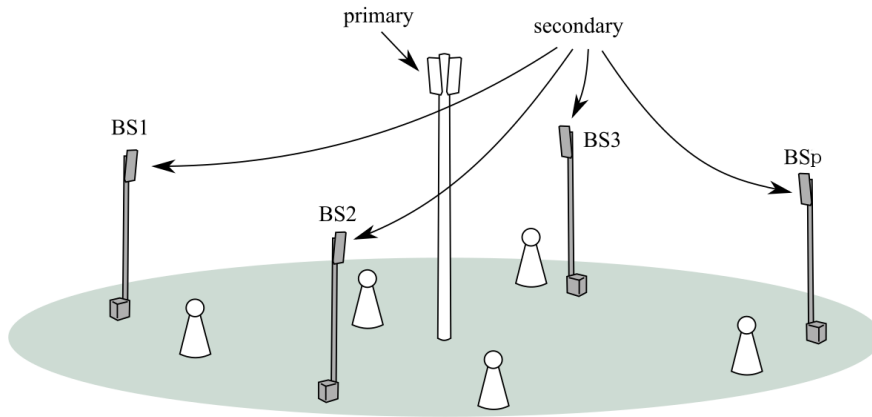


Figure 4.1: System Setting

Based on the data matrix, BSs decide on whether a portion of the spectrum is occupied or not. Now we can view each column of the data matrix as a random vector sampled from a p -dimensional distribution. If the distribution is Gaussian, under H_0 the components of a sample vector are independent whereas if the distribution is non-Gaussian, the components of a sample vector may be dependent. However, they are uncorrelated under both cases. Under H_1 , both cases are correlated. If the cognitive radio network is in the environment with impulsive noise, we can model the impulsive noise by Complex Elliptically Symmetric (CES) distribution. The texture parameter τ is used to model the impulsive behavior of the noise. And each random vector shares one texture parameter, which makes components of the random vector dependent.

In the case of no presence of signal, the received samples have identity population covariance matrix whereas in the case of presence of signal the samples have identity population covariance with one entry perturbation. Here are the population covariance matrices C_p ,

$$C_p = E[X(j)X(j)^H] = \sigma^2 I \quad (\text{under } H_0)$$

$$C_p = hh^H + \sigma^2 I \quad (\text{under } H_1),$$

where $X(j)$ is the j th column of X and h is the channel from primary base station to BSs.

We note that under H_1 , C_p has one eigenvalue equal to $\sigma^2 + h^H h$ and all the rest σ^2 . The behavior of SCM of data generated by this population covariance matrix is studied by the spiked model introduced in the chapter 3. If the quantity $h^H h > \sigma^2 \sqrt{\alpha}$ and $\alpha < 1$, where $\alpha = \frac{\rho}{n}$, we have the maximum eigenvalue of SCM, λ_{max} , converges almost surely to

$$b' = (h^H h + \sigma^2) \left(1 + \frac{\alpha}{\rho}\right),$$

where $\rho = \frac{h^H h}{\sigma^2}$ is the defined signal to noise ratio (SNR). Due to $b' > b$, where $b = \sigma^2(1 + \sqrt{\alpha})$ is the right edge of the support of the Marčenko-Pastur law, the maximum eigenvalue under the null hypothesis should be less than it under the alternative hypothesis in limit and also in finite case. Let λ_{max} be the maximum eigenvalue of scatter estimator, λ_{min} be the minimum eigenvalue of scatter estimator. Under the case of noise variance unknown, the ratio $\frac{a}{b} = \frac{(1 + \sqrt{\alpha})^2}{(1 - \sqrt{\alpha})^2}$ does not depend on the noise variance. Thus we choose the test statistics as $\frac{\lambda_{max}}{\lambda_{min}}$ which converges to $\frac{a}{b}$ under H_0 . If we can derive the asymptotic distribution of this statistics, the asymptotic distribution should be independent of the noise power, hence we can have a test without knowledge of the noise power.

However, when the noise is impulsive, the statistics derived from SCM has no

difference under H_0 and H_1 even when the sample size is huge (Figure 3.4b and Figure 3.4c). The robust estimator will be a better choice since the maximum eigenvalue of robust estimator will be pulled out under H_1 . Tyler's M-estimator and Maronna's M-estimator associated with real student's t distribution reserves the Marčenko-Pastur Law under both impulsive noise and Gaussian noise.

The performance of spectrum sensing can be primarily determined based on two metrics: the probability of detection (POD) and the probability of false alarm (POF). POD is the probability that secondary users claim that the spectrum is occupied when the spectrum is indeed occupied. POF is the probability that secondary users claim that the spectrum is occupied when the spectrum is free. POD is closely related to quality-of-service (QoS) of PUs since low POD means the communication of PUs will be interfered often by SUs. POF is closely associated with the QoS of SUs since a false alarm will reduce the spectral usage efficiency. The optimal detector for spectrum sensing usually has the maximized POD given the constraint of the POF.

In our case, we assume there is no information on neither signal nor noise available. It is impossible to derive the expression of POD or POF analytically. To characterize the performance and control either POD or POF, the nonparametric scheme is used. First, we compute the empirical cumulative distribution function (ECDF) of the statistics based on data. The ECDF is computed by, $\hat{F}_N(t) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{T_i \leq t\}}$, where T_i is the statistics computed based on data generated in null hypothesis. With $\hat{F}_N(t)$ we can easily approximate the POF.

Simulations were carried out to establish the performance of the detectors using robust estimators in comparison to the detectors using SCM in different noise environment with a different degree of impulsiveness. To measure impulsiveness of the noise, we may refer to its kurtosis which is the scaled version of the fourth central moment and excess kurtosis which is defined as kurtosis minus kurtosis of Gaussian distribution, 3. The higher the excess kurtosis is for a noise distribution, the more

impulsive will the noise be. Five secondary users, i.e., $p = 5$ were simulated along with $n = 100$ for each secondary users. SNR in this simulation was chosen to be $\rho = 0$ dB and the noise power $\sigma^2 = 1$. The noise was chosen to be Gaussian, Laplace and Generalized Gaussian with parameter $s = 0.1$ and $s = 0.2$ which are more impulsive than Laplace. The excess kurtosis of the Laplace distribution is 3, and it of the Generalized Gaussian distribution with $s = 0.1$ is around 1956.30 and it with $s = 0.2$ is around 48.95. We use the Laplace noise to show that Maronna's M-estimator of our choice may be adapted to have optimal performance among all three estimators. Figure 4.2 is the receiver operating characteristic (ROC) curve for the test described above using three different scatter estimators, among them Maronna's M estimator is computed with parameter $t = 21$ in all four cases and $t = 0.7$ which is the optimal setting in the Laplace Noise. The way we optimize the parameter is by fix the POF in different t and search for the t gives the highest POD.

From the simulation, we can found that the SCM works well only when the noise is not much impulsive. When the degree of impulsiveness of noise goes to high, the SCM breaks down which performs even worse than random guess if the impulsiveness is extremely high(see Figure 4.2d and Figure 4.2c.). Tyler's M-estimator performs well in all the four cases, especially in the case of high degree of impulsiveness. We can conclude that Tyler's M-estimator is quite robust in either impulsive or non-impulsive noise environment. The Maronna's M-estimator has a free parameter to adjust. By arbitrarily choose $t = 21$, it is outperformed by Tyler's M-estimator. In the case of the Laplace Noise, the optimized Maronna's M-estimator outperforms Tyler's M-estimator(see Figure 4.2b).

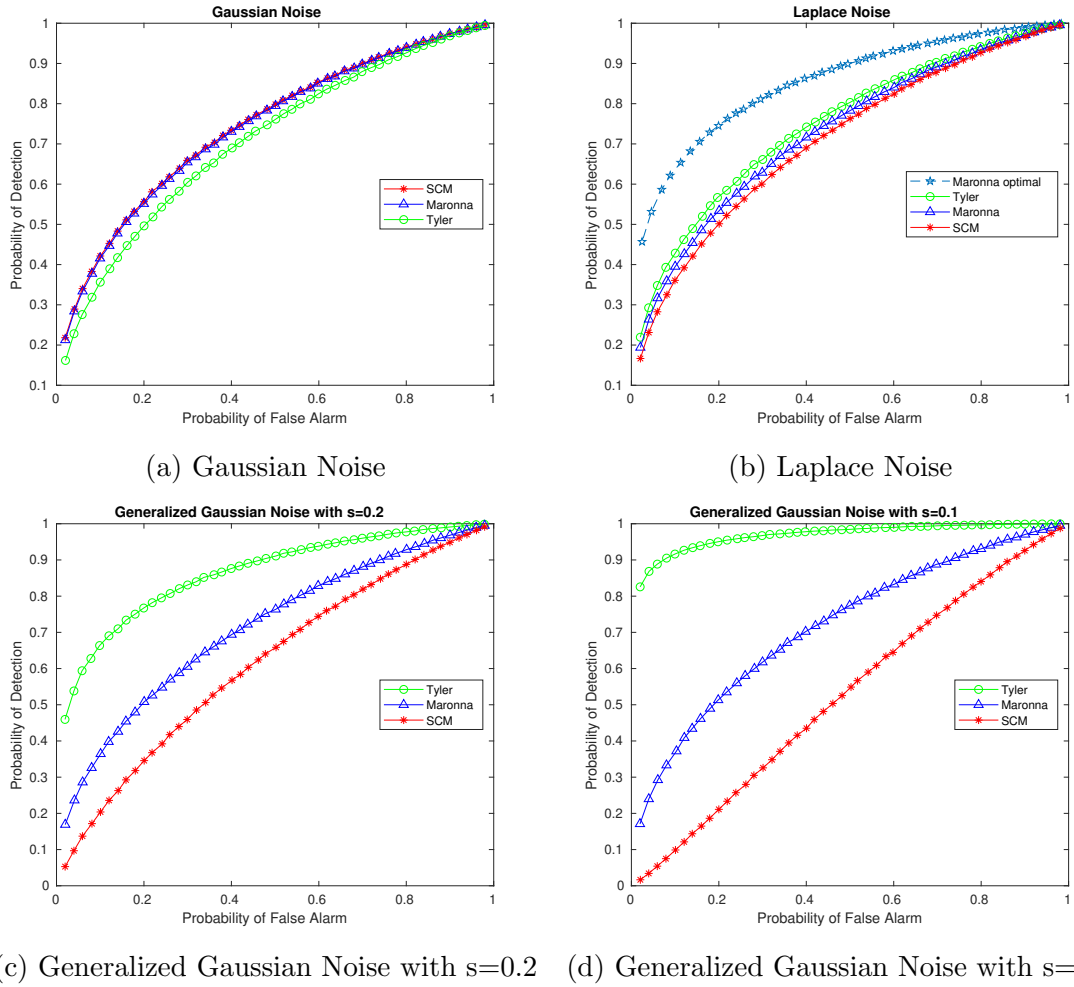


Figure 4.2: ROC curves

Chapter 5

Conclusion

In this thesis, we investigated estimation of the scatter matrix. The SCM is the standard estimator of the population covariance matrix: it is simple, fast and converges to the population matrix in the limit of infinitely many observations in many cases. The disadvantage of this estimator is that it is vulnerable to extreme observations which are common when the data set is generated by heavy-tailed distributions. The extreme value will cause the variation of this estimator to be extremely high. This estimator is also sensitive to non-Gaussian data. The estimator becomes inefficient when a small portion of the data set generated by non-Gaussian distributions. The so-called robust estimation is proposed to handle the drawbacks of the SCM. We focus on M-estimation which is a generalization of MLE. This estimator can resist extreme observations without preprocessing the data set and is easy to be derived. The computation of this estimator is by solving a fixed point of a function, and the computation is not complicated. We also introduced some high dimensional properties of SCM in the framework of random matrix theorem. Inspired by those high dimensional properties, an eigenvalue-based spectrum sensing method in the cognitive network is proposed, and by simulation, we found it outperforms the existing method using SCM. By choosing M-estimator properly, we may have a spectrum sensing method with the ability to adaptively fit the noise environment as shown in chapter 4.

Regarding the proposed spectrum sensing method, there are two questions that need to be further studied. The decision threshold controlling the POF is derived

by numerical method in the thesis. Thus, the derivation of the analytical result of the threshold is still an open question. Also how to choose an optimal estimator in different noise environment needs to be investigated.

REFERENCES

- [1] E. Ollila, D. E. 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.
- [2] 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.
- [3] P. J. Bickel, E. Levina *et al.*, “Regularized estimation of large covariance matrices,” *The Annals of Statistics*, vol. 36, no. 1, pp. 199–227, 2008.
- [4] T. Kubokawa and M. Srivastava, “Estimating the covariance matrix: a new approach,” *Journal of Multivariate Analysis*, vol. 86, no. 1, pp. 28–47, 2003.
- [5] R. A. Maronna, R. D. Martin, V. J. Yohai, and M. Salibián-Barrera, *Robust Statistics: Theory and Methods (with R)*. Wiley, 2018.
- [6] T. K. Blankenship, D. Kriztman, and T. S. Rappaport, “Measurements and simulation of radio frequency impulsive noise in hospitals and clinics,” in *Vehicular Technology Conference, 1997, IEEE 47th*, vol. 3. IEEE, 1997, pp. 1942–1946.
- [7] E. Conte, M. Longo, and M. Lops, “Modelling and simulation of non-rayleigh radar clutter,” in *IEE Proceedings F (Radar and Signal Processing)*, vol. 138, no. 2. IET, 1991, pp. 121–130.
- [8] M. Sekine, S. Ohtani, T. Musha, T. Irabu, E. Kiuchi, T. Hagsawa, and Y. Tomita, “Weibull-distributed ground clutter,” *IEEE Transactions on Aerospace and Electronic Systems*, no. 4, pp. 596–598, 1981.
- [9] M. Sekine, S. Ohatani, T. Musha, T. Irabu, E. Kiuchi, T. Hagsawa, and Y. Tomita, “Mti processing and weibull-distributed ground clutter,” *IEEE Transactions on Aerospace and Electronic Systems*, no. 6, pp. 729–730, 1982.
- [10] E. Eskin, “Anomaly detection over noisy data using learned probability distributions,” in *In Proceedings of the International Conference on Machine Learning*. Citeseer, 2000.

- [11] S. Chawla, D. Hand, and V. Dhar, “Outlier detection special issue,” *Data mining and knowledge discovery*, vol. 20, no. 2, pp. 189–190, 2010.
- [12] I. M. Johnstone, “On the distribution of the largest eigenvalue in principal components analysis,” *Annals of statistics*, pp. 295–327, 2001.
- [13] Y.-Q. Yin, Z.-D. Bai, and P. R. Krishnaiah, “On the limit of the largest eigenvalue of the large dimensional sample covariance matrix,” *Probability theory and related fields*, vol. 78, no. 4, pp. 509–521, 1988.
- [14] C. A. Tracy and H. Widom, “On orthogonal and symplectic matrix ensembles,” *Communications in Mathematical Physics*, vol. 177, no. 3, pp. 727–754, 1996.
- [15] R. Couillet, F. Pascal, and J. W. Silverstein, “Robust estimates of covariance matrices in the large dimensional regime.” *IEEE Trans. Information Theory*, vol. 60, no. 11, pp. 7269–7278, 2014.
- [16] —, “The random matrix regime of maronnas m-estimator with elliptically distributed samples,” *Journal of Multivariate Analysis*, vol. 139, pp. 56–78, 2015.
- [17] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, “Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey,” *Computer Networks*, vol. 50, no. 13, pp. 2127 – 2159, 2006. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1389128606001009>
- [18] P. Bianchi, J. Najim, G. Alfano, and M. Debbah, “Asymptotics of eigenbased collaborative sensing,” in *2009 IEEE Information Theory Workshop*, Oct 2009, pp. 515–519.
- [19] Y. Zeng and Y. . Liang, “Eigenvalue-based spectrum sensing algorithms for cognitive radio,” *IEEE Transactions on Communications*, vol. 57, no. 6, pp. 1784–1793, June 2009.
- [20] L. S. Cardoso, M. Debbah, P. Bianchi, and J. Najim, “Cooperative spectrum sensing using random matrix theory,” in *2008 3rd International Symposium on Wireless Pervasive Computing*, May 2008, pp. 334–338.
- [21] D. Middleton, “Non-gaussian noise models in signal processing for telecommunications: new methods an results for class a and class b noise models,” *IEEE Transactions on Information Theory*, vol. 45, no. 4, pp. 1129–1149, 1999.
- [22] G. Frahm, “Generalized elliptical distributions: theory and applications,” Ph.D. dissertation, Universität zu Köln, 2004.

- [23] M. G. Genton, *Skew-elliptical distributions and their applications: a journey beyond normality*. CRC Press, 2004.
- [24] R. A. Maronna, “Robust m-estimators of multivariate location and scatter,” *The annals of statistics*, pp. 51–67, 1976.
- [25] D. E. Tyler, “A distribution-free m-estimator of multivariate scatter,” *The Annals of Statistics*, pp. 234–251, 1987.
- [26] F. Pascal, Y. Chitour, J.-P. Ovarlez, P. Forster, and P. Larzabal, “Covariance structure maximum-likelihood estimates in compound gaussian noise: Existence and algorithm analysis,” *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 34–48, 2008.
- [27] D. D. Boos and L. A. Stefanski, *Essential statistical inference: theory and methods*. Springer Science & Business Media, 2013, vol. 120.
- [28] T. W. Anderson, T. W. Anderson, T. W. Anderson, T. W. Anderson, and E.-U. Mathématicien, *An introduction to multivariate statistical analysis*. Wiley New York, 1958, vol. 2.
- [29] V. A. Marčenko and L. A. Pastur, “Distribution of eigenvalues for some sets of random matrices,” *Mathematics of the USSR-Sbornik*, vol. 1, no. 4, p. 457, 1967.
- [30] K. W. Wachter *et al.*, “The strong limits of random matrix spectra for sample matrices of independent elements,” *The Annals of Probability*, vol. 6, no. 1, pp. 1–18, 1978.
- [31] Y. Q. Yin, “Limiting spectral distribution for a class of random matrices,” *Journal of multivariate analysis*, vol. 20, no. 1, pp. 50–68, 1986.
- [32] Y. Yin, Z. Bai, and P. Krishnaiah, “Limiting behavior of the eigenvalues of a multivariate f matrix,” *Journal of Multivariate Analysis*, vol. 13, no. 4, pp. 508–516, 1983.
- [33] C. A. Tracy and H. Widom, “Level-spacing distributions and the airy kernel,” *Communications in Mathematical Physics*, vol. 159, no. 1, pp. 151–174, 1994.
- [34] F. Olver, *Asymptotics and special functions*. AK Peters/CRC Press, 1997.
- [35] A. Soshnikov, “A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices,” *Journal of Statistical Physics*, vol. 108, no. 5-6, pp. 1033–1056, 2002.
- [36] T. Zhang, X. Cheng, and A. Singer, “Marčenko–pastur law for tyler’s m-estimator,” *Journal of Multivariate Analysis*, vol. 149, pp. 114–123, 2016.

- [37] —, “Marchenko-pastur law for tylers and maronnas m-estimators,” *arXiv preprint*, 2014.
- [38] J. Baik and J. W. Silverstein, “Eigenvalues of large sample covariance matrices of spiked population models,” *Journal of multivariate analysis*, vol. 97, no. 6, pp. 1382–1408, 2006.