Journal of Multivariate Analysis 149 (2016) 114–123

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Marčenko–Pastur law for Tyler's M-estimator

Teng Zhang [a,*], Xiuyuan Cheng [b], Amit Singer [c]

[a] Department of Mathematics, University of Central Florida, Orlando, FL 32816, USA

[b] Applied Mathematics Program, Yale University, New Haven, CT 06511, USA

[c] Department of Mathematics and PACM, Princeton University, Princeton, NJ 08544, USA

ARTICLE INFO

Article history:
Received 2 June 2015
Available online 12 April 2016

AMS subject classifications:
15B52
62H10

Keywords:
Covariance estimation
Robust statistics
Random matrix theory

ABSTRACT

This paper studies the limiting behavior of Tyler's M-estimator for the scatter matrix, in the regime that the number of samples n and their dimension p both go to infinity, and p/n converges to a constant y with $0 < y < 1$. We prove that when the data samples $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are identically and independently generated from the Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{I})$, the operator norm of the difference between a properly scaled Tyler's M-estimator and $\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i^\top/n$ tends to zero. As a result, the spectral distribution of Tyler's M-estimator converges weakly to the Marčenko–Pastur distribution.

1. Introduction

Many statistical estimators and signal processing algorithms require the estimation of the covariance matrix of the data samples. When the underlying distribution of the data samples $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{R}^p$ is assumed to have zero mean, a commonly used estimator is the sample covariance matrix $\boldsymbol{S}_n = \sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i^\top/n$.

However, the estimator $\boldsymbol{S}_n$ is sensitive to outliers, and performs poorly in terms of statistical efficiency (i.e., it has a large variance) for heavy-tailed distributions, e.g., when the tail decays slower than the Gaussian tail.

A popular robust covariance estimator is an M-estimator introduced by Tyler [20], denoted by $\hat{\Sigma}$, which is the unique solution to

$$\hat{\Sigma} = \frac{p}{n}\sum_{i=1}^n \frac{\boldsymbol{x}_i\boldsymbol{x}_i^\top}{\boldsymbol{x}_i^\top \hat{\Sigma}^{-1}\boldsymbol{x}_i}, \quad \operatorname{tr}(\hat{\Sigma}) = 1. \tag{1}$$

Tyler's M-estimator gives the "shape" of the covariance, but is missing its magnitude. However, for many applications the "shape" of the covariance suffices, for example, the principal components can be obtained from the "shape".

Compared with the sample covariance estimator, Tyler's M-estimator is more robust to heavy-tailed elliptical distributions. The density function of elliptical distributions in $\mathbb{R}^p$ takes the form

$$f(\boldsymbol{x}; \Sigma, \mu) = |\Sigma|^{-1/2} g\{(\boldsymbol{x}-\mu)^\top \Sigma^{-1}(\boldsymbol{x}-\mu)\},$$

where g is some nonnegative function such that $\int_0^\infty x^{p-1}g(x)\,\mathrm{d}x$ is finite. This family of distributions is a natural generalization of the Gaussian distribution by allowing heavier or lighter tails while maintaining the elliptical geometry of the equidensity contours. Elliptical distributions are considered important in portfolio theory and financial data, and we

* Corresponding author.
E-mail addresses: teng.zhang@ucf.edu (T. Zhang), xiuyuan.cheng@yale.edu (X. Cheng), amits@math.princeton.edu (A. Singer).

http://dx.doi.org/10.1016/j.jmva.2016.03.010

refer to the work by El Karoui [11, Section 4] for further discussion. Besides, elliptical distributions are used by Ollila and Tyler [19] in modeling radar data, where the empirical distributions are heavy-tailed because of outliers.

Tyler [20] showed that when a data set follows an unknown elliptical distribution (with mean zero), Tyler's M-estimator is the most robust covariance estimator in the sense of minimizing the maximum asymptotic variance. This property suggests that Tyler's M-estimator should be more accurate than the sample covariance estimator for elliptically distributed data. Empirically, it has been shown to outperform the sample covariance estimator in applications such as finance in the work by Frahm and Jaekel [13], anomaly detection in wireless sensor networks by Chen et al. [4], antenna array processing by Ollila and Koivunen [18], and radar detection by Ollila and Tyler [19].

1.1. Asymptotic analysis in a high-dimensional setting

Many scientific domains customarily deal with sets of high dimensional data samples, and therefore it is increasingly common to work with data sets where the number of variables, p, is of the same order of magnitude as the number of observations, n. Under this high-dimensional setting, the asymptotic spectral properties of $\boldsymbol{S}_n$ at the limit of infinite number of samples and infinite dimensions have been well studied by Johnstone [15]. A noticeable example is the convergence of the spectral distribution. Denoting the eigenvalues of a matrix $\boldsymbol{A}$ by $\lambda_1(\boldsymbol{A}), \dots, \lambda_n(\boldsymbol{A})$, its spectral distribution is a discrete probability measure

$$P = P(\cdot|\boldsymbol{A}) = \frac{1}{n}\sum_{i=1}^{n}\delta_{\lambda_i(\boldsymbol{A})}$$

with δ_s denoting Dirac measure at $s \in \mathbb{R}$. Marčenko and Pastur [16] showed that when the entries of $\{\boldsymbol{x}_i\}_{i=1}^n$ are Gaussian independent identically distributed random variables with mean 0 and variance 1, $p, n \to \infty$ and $p/n \to y$, where $0 < y \le 1$, the spectral distribution of the eigenvalues of $\boldsymbol{S}_n$ converges weakly to the Marčenko–Pastur distribution defined by

$$\rho_{\mathrm{MP},y}(x) = \frac{1}{2\pi}\frac{y\sqrt{(y_+ - x)(x - y_-)}}{x}\mathbf{1}_{[y_-,y_+]}, \quad \text{where } y_\pm = \left(1 \pm \sqrt{y}\right)^2. \tag{2}$$

Tyler's M-estimator is closely related to and can be considered as a special case of Maronna's M-estimator, which is defined by

$$\bar{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} u(\boldsymbol{x}_i^\top \bar{\Sigma}^{-1}\boldsymbol{x}_i)\boldsymbol{x}_i\boldsymbol{x}_i^\top \tag{3}$$

for a nonnegative function $u : [0, \infty) \to [0, \infty)$. The properties of Maronna's M-estimator in the high-dimensional regime when $p, n \to \infty$, $p/n \to y$ and $0 < y < 1$ have been analyzed in recent works by Couillet et al. [7,8], which obtained convergence results for a properly scaled Maronna's M-estimator under the assumptions that $u(x)$ is nonnegative, nonincreasing and continuous; $xu(x)$ is nondecreasing and bounded and $\sup_x xu(x) > 1$. Moreover, spiked random matrix models were also studied by Couillet [5]. However, these results do not apply to Tyler's M-estimator, although Frahm and Jaekel [13] have conjectured that the spectral distribution converges weakly to the Marčenko–Pastur distribution. Some works focused on the performance of Tyler's M-estimator for the case $p, n \to \infty$ and $p/n \to 0$: Dümbgen [10] showed that the condition number of Tyler's estimator is $1 + 4\sqrt{p/n} + o(\sqrt{p/n})$, and Frahm and Glombek [12] showed that the spectral distribution of $\sqrt{n/p}(\bar{\Sigma} - \mathbf{I})$ converges weakly to a semicircle distribution.

1.2. Main results

In this paper, we analyze Tyler's M-estimator in the high-dimensional setting. Our main results, Theorem 2.3 and Corollary 2.4, show that as $p, n \to \infty$ and $p/n \to y$, $0 < y < 1$, the spectral distribution of a properly scaled Tyler's M-estimator converges weakly to the Marčenko–Pastur distribution $\rho_{\mathrm{MP},y}(x)$. Based on the properties of Tyler's M-estimator, this paper analyzes the spectral distribution when data samples are i.i.d. drawn from other distributions, such as elliptical distributions.

When data samples are generated from elliptical distributions, the spectral distribution of the sample covariance estimator has been studied by El Karoui [11, Theorem 2]. Compared to Corollary 2.4, the limiting spectral distribution of $\boldsymbol{S}_n$ is much more complicated, and therefore our result might be more applicable in practice.

High-dimensional analysis of Maronna's M-estimator of the covariance are generally obtained by showing that the operator norm of the difference between M-estimator and a standard Wishart matrix (or sample covariance matrix) tends to 0: Dümbgen [10] proved it by a linear expansion of the M-estimator, and Couillet et al. [7,8] proved it by representing Maronna's M-estimator as a weighted sum of $\boldsymbol{x}_i\boldsymbol{x}_i^\top$ and proved the uniform convergence of the weights. We follow the same direction while giving an alternate proof for the convergence of the weights, by considering the weights as the solution to an optimization problem, which can handle Tyler's M-estimator that is not covered by the results in Couillet et al. [7,8]. We remark that this approach can also be applied to Maronna's M-estimator to prove some of the results in Couillet et al. [7,8].