Journal of Multivariate Analysis 132 (2014) 215–228

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

On the strong convergence of the optimal linear shrinkage estimator for large dimensional covariance matrix

Taras Bodnar [a], Arjun K. Gupta [b,*], Nestor Parolya [c]

[a] *Department of Mathematics, Humboldt-University of Berlin, D-10099 Berlin, Germany*

[b] *Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43403, USA*

[c] *Institute of Empirical Finance (Econometrics), Leibniz University of Hannover, 30167, Germany*

ARTICLE INFO

Article history:
Received 9 March 2013
Available online 4 September 2014

AMS 2010 subject classifications:
60B20
62H12
62G20
62G30

Keywords:
Large-dimensional asymptotics
Random matrix theory
Covariance matrix estimation

ABSTRACT

In this work we construct an optimal linear shrinkage estimator for the covariance matrix in high dimensions. The recent results from the random matrix theory allow us to find the asymptotic deterministic equivalents of the optimal shrinkage intensities and estimate them consistently. The developed distribution-free estimators obey *almost surely* the smallest Frobenius loss over all linear shrinkage estimators for the covariance matrix. The case we consider includes the number of variables $p \to \infty$ and the sample size $n \to \infty$ so that $p/n \to c \in (0, +\infty)$. Additionally, we prove that the Frobenius norm of the sample covariance matrix tends almost surely to a deterministic quantity which can be consistently estimated.

Published by Elsevier Inc.

1. Introduction

Nowadays, the estimation of the covariance matrix is one of the most important problems not only in statistics but also in finance, wireless communications, biology, etc. The traditional estimator of the covariance matrix, i.e. its sample counterpart, seems to be a good decision only when the dimension p is much smaller than the sample size n. This case is called the "standard asymptotics" (see, e.g., [21]). Here, the sample covariance matrix is proven to be an unbiased and a consistent estimator for the covariance matrix. More problems arise when p is comparable to n, i.e. both the dimension p and the sample size n tend to infinity while their ratio p/n tends to a positive constant c. It is called the "large dimensional asymptotics" or "Kolmogorov asymptotics" (see, e.g., [6,8]). This type of asymptotics have been exhaustively studied by Girko [15,16], where it was called the "general statistical analysis". There is a great amount of research done on the asymptotic behavior of functionals of the sample covariance matrix under the large dimensional asymptotics (see, e.g., [17–19,5]).

There are some significant improvements in the case when the covariance matrix has a special structure, e.g. sparse, low rank, etc. (see, [7,24,9,10], etc.). The case when the underlying random process obeys the factor structure is studied by Fan et al. [11]. In these cases the covariance matrix can be consistently estimated even in high-dimensional case. In the case when no additional information on the structure of the covariance matrix is available, the problem has not been studied in detail up to now. The exception is the paper of Ledoit and Wolf [22], where a linear shrinkage estimator was suggested which possesses the smallest Frobenius loss in quadratic mean.

* Corresponding author.
E-mail address: gupta@bgsu.edu (A.K. Gupta).

http://dx.doi.org/10.1016/j.jmva.2014.08.006
0047-259X/Published by Elsevier Inc.

Marčenko and Pastur [23], Yin [31], Silverstein [26], Bai et al. [2], Bai and Silverstein [5] used the large dimensional asymptotics to study the asymptotic behavior of the eigenvalues of general random matrices. They discovered that appropriately transformed random matrix at infinity has a nonrandom behavior and showed how to find the limiting density of its eigenvalues. In particular, Silverstein [26] proved under very general conditions that the Stieltjes transform of the sample covariance matrix tends almost surely to a nonrandom function which satisfies some equation. This equation was first derived by Marčenko and Pastur [23], who showed how the real covariance matrix and its sample estimate are connected at infinity. In our work we use this result for estimating functionals of the covariance matrix consistently.

In this work we concentrate on certain type of estimators, namely the shrinkage estimators. The shrinkage estimators were introduced by Stein [30]. They are constructed as a linear combination of the sample estimator and some known target. These estimators have remarkable property: they are biased but can significantly reduce the mean square error of the estimator. In the large as well as in the small dimensional cases it is difficult to find the consistent estimators for the so-called shrinkage intensities. In this situation Ledoit and Wolf [22] made progress when the target matrix is the identity and found a feasible linear shrinkage estimator for the covariance matrix which is optimal in the sense of the squared mean. This estimator provided a remarkable dominance over the sample estimator and other known estimators for the covariance matrix. The linear shrinkage presented by Ledoit and Wolf [22] shows its best performance in case when the eigenvalues of the covariance matrix are not dispersed and/or the concentration ratio c is large.

In this paper we extend the work of Ledoit and Wolf [22] by constructing a more general linear shrinkage estimator for a large dimensional covariance matrix. The target matrix here is considered to be an arbitrary symmetric positive definite matrix with uniformly bounded trace norm. Using random matrix theory we prove that the optimal shrinkage intensities are nonrandom at infinity, find their asymptotic deterministic equivalents and estimate them consistently. Additionally we show that the Frobenius norm of the covariance matrix tends to a deterministic quantity which can also be estimated consistently. The resulting estimator obeys *almost surely* the smallest Frobenius loss when the dimension p and the sample size n increase together and $p/n \to c \in (0, \infty)$ as $n \to \infty$.

The rest of paper is organized as follows. In Section 2 we present the preliminary results from the random matrix theory which are used in the proofs of the theorems. Section 3 contains the *oracle* linear shrinkage estimator and the main asymptotic results on the shrinkage intensities and the Frobenius norm of the sample covariance matrix. In Section 4 we present the *bona fide* linear shrinkage estimator for the covariance matrix and make a short comparison with the well-known Ledoit and Wolf [22] estimator. The results of the empirical study are provided in Section 5, while Section 6 summarizes all main results of the paper. The proofs of the theorems are moved to the Appendix.

2. Preliminary results and large dimensional asymptotics

By "large dimensional asymptotics" or "Kolmogorov asymptotics" it is understood that $\frac{p}{n} \to c \in (0, +\infty)$ where the number of variables $p \equiv p(n)$ and the sample size n both tend to infinity. In this case the traditional sample estimators perform poorly or very poorly and tend to over/underestimate the population covariance matrix.

We use the following notations in the paper:

- $\mathbf{\Sigma}_n$ stands for the covariance matrix, and $\mathbf{S}_n$ denotes the corresponding sample covariance matrix.
- The pairs $(\tau_i, \boldsymbol{v}_i)$ for $i = 1, \ldots, p$ are the collection of eigenvalues and the corresponding orthonormal eigenvectors of the covariance matrix $\mathbf{\Sigma}_n$.
- $H_n(t)$ is the empirical distribution function (e.d.f.) of the eigenvalues of $\mathbf{\Sigma}_n$, i.e.,

$$H_n(t) = \frac{1}{p}\sum_{i=1}^{p} \mathbb{1}_{\{\tau_i < t\}} \tag{2.1}$$

  where $\mathbb{1}_{\{\cdot\}}$ is the indicator function.
- Let $\mathbf{X}_n$ be a $p \times n$ matrix which consists of independent and identically distributed (i.i.d.) real random variables with zero mean and unit variance such that

$$\mathbf{Y}_n = \mathbf{\Sigma}_n^{\frac{1}{2}} \mathbf{X}_n. \tag{2.2}$$

In the derivation of the main results the following five assumptions are used.

(A1) The population covariance matrix $\mathbf{\Sigma}_n$ is a nonrandom p-dimensional positive definite matrix.
(A2) Only the matrix $\mathbf{Y}_n$ is observable. We know neither $\mathbf{X}_n$ nor $\mathbf{\Sigma}_n$ itself.
(A3) We assume that $H_n(t)$ converges to some limit $H(t)$ at all points of continuity of H.
(A4) The elements of the matrix $\mathbf{X}_n$ have uniformly bounded moments of order $4 + \varepsilon$, $\varepsilon > 0$.
(A5) The largest eigenvalue of the covariance matrix $\mathbf{\Sigma}_n$ is at most of the order $O(\sqrt{p})$. Moreover, we assume that the order of only finite number of eigenvalues could depend on p.

The assumptions (A1)–(A3) are important to prove Marčenko–Pastur equation (see, e.g., [26]) and they are standard in the large dimensional asymptotics (see, e.g., [5]). In particular, the assumption (A3) on the existence of the limiting population