IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received November 8, 2017, accepted December 4, 2017, date of publication December 11, 2017, date of current version February 14, 2018.

Digital Object Identifier 10.1109/ACCESS.2017.2782208

Estimation of Large Covariance Matrices by Shrinking to Structured Target in Normal and Non-Normal Distributions

JIANBO LI, JIE ZHOU, AND BIN ZHANG
College of Mathematics, Sichuan University, Chengdu 610064, China
Corresponding author: Jie Zhou (jzhou@scu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61374027 and in part by the Program for Changjiang Scholars and Innovative Research Team in University under Grant IRT_16R53 from the Chinese Education Ministry.

ABSTRACT This paper addresses the estimation of large-dimensional covariance matrices under both normal and nonnormal distributions. The shrinkage estimators are constructed by convexly combining the sample covariance matrix and a structured target matrix. The optimal oracle shrinkage intensity is obtained analytically for any prespecified target in a class of matrices which includes various structured matrices such as banding, thresholding, diagonal, and block diagonal matrices. After deriving the unbiased and consistent estimates of some quantities in the oracle intensity involving unknown population covariance matrix, two classes of available optimal intensities are proposed under normality and nonnormality, respectively, by plug-in technique. For the target matrix with unknown parameter such as bandwidth in banded target, an analytic estimate of unknown parameter is provided. Both the numerical simulations and applications to signal processing and discriminant analysis show the comparable performance of the proposed estimators for large-dimensional data.

INDEX TERMS Covariance matrix, structured target matrix, large dimension, shrinkage estimation.

I. INTRODUCTION

Estimation of population covariance matrix from a random sample has attracted a lot of attentions because of the fundamental role of covariance matrices in many science and technology areas such as statistical inference [1], signal processing [2], [3], communication [4], biometrics [5] and financial economics [6]. Various realistic applications usually require a covariance matrix estimator to be not only invertible but also well-conditioned. Let $(x_1, \ldots, x_n)$ be an independent and identically distributed (i.i.d.) sample of size n from a p-dimensional population with mean zero and covariance Σ, the sample covariance matrix (SCM)

$$S = (s_{ij})_{p\times p} = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T$$

is widely adopted as a classical estimator of Σ, which has some important statistical properties provided in the classical statistics for large sample cases. As the high- or large-dimensional data are collected in various ways in the practical applications, the dimensionality becomes non-negligible, but the traditional methods directly utilizing the SCM as the estimate of the population covariance matrix often perform very poorly. In fact, the SCM S can not anymore be considered as a good estimate of the true covariance matrix Σ when the dimension p is large compared to the sample size n, and even singular when $p > n$. For large-dimensional covariance matrices, even if $n > p$, the SCM S is typically not well-conditioned. Some researches, e.g., [7]–[10] reveal some important features of random matrices such as the SCM using random matrix theory. Particularly, the famous Marčenko–Pastur law describes the asymptotic behavior, which depends on the ratio of the dimension p and the sample size n, of eigenvalues of S as large random matrix. However, the high- or large-dimensional covariance estimation is known to still be a difficult problem, especially in the "large p small n" setting. In recent years, a lot of attentions have been devoted to looking for alternative estimators of the SCM in high- or large-dimensional background and many improved estimators of covariance matrix are proposed. One can refer to [11]–[29] and references therein.

The shrinkage estimation is a popular and efficient approach which reduces the mean squared error (MSE) of the estimate by convexly combining the SCM S and a prespecified target matrix. In [13], the identity matrix I_p multiplied by a scalar $p^{-1}\text{tr}(S)$ is introduced as target matrix and a distribution-free consistent estimator (named LW estimator) is proposed to approximate the oracle estimate. The LW estimator shrinks the eigenvalues of the SCM towards their grand mean, and is asymptotically optimal in the sense of minimizing the expected quadratic loss function as both the sample size n and the dimension p go to infinity together but the ratio p/n remaining bounded. For normal cases, in [17], the oracle approximating shrinkage (OAS) estimator and the Rao-Blackwell Ledoit-Wolf (RBLW) estimator (named owing to apply the well-known Rao–Blackwell theorem to the LW method) are proposed in which the latter improves the LW estimator. In [21], the diagonal target matrix $D_S = \text{diag}(s_{11}, \ldots, s_{pp})$ taking the diagonal elements of S is explored and a corresponding shrinkage estimator under normality is suggested. Correspondingly, the shrinkage estimate with such diagonal target is extended to the non-normal cases in [28]. Instead of the diagonal D_S, in [22], a tapering matrix, proposed in [30], is selected as the target, and then a shrinkage-to-tapering oracle approximating (STOA) estimator is suggested which inherits the advantages of both tapering and shrinkage estimators. Note that the target matrix D_S can be thought as a special case of tapering target with bandwidth 0. More recently, some researchers consider two targets as candidates of shrinkage targets and proposed double shrinkage estimators as [28]. A multi-target shrinkage estimation by optimization under non-distribution assumption is proposed in [31]. In [32], an iterative method similar to the one in [22] is proposed to approximate the oracle estimator, especially the convergency is proved and a closed-form limit is obtained. Also, there are other shrinkage strategies and their applications in [25]–[27], [33], and [34].

The main contributions of this paper are as follows:

1) Considering the variousness and importance of target matrices in shrinkage covariance matrix estimation, we introduce a class of targets which includes a number of structured matrices such as banding, thresholding, diagonal and block diagonal matrices. This matrix class is first considered in [32] only for normal distribution, however, in this paper we devote to tackling with covariance estimation for not only normal but also non-normal population distributions. We derive analytically a class of optimal oracle shrinkage estimators in the sense of minimizing the expected quadratic loss under both normality and non-normality.
2) The plug-in strategy is adopted to estimate the oracle shrinkage intensity and the population covariance matrix for any given target in the aforementioned matrix class. This approach is different from the iterative approaches proposed in [17], [22], and [32] that are difficult to derive the limit of iterations and even if prove their convergence for general distributions. In this paper, we provide the unbiased and consistent estimators of some functions of unknown population covariance matrix in the optimal oracle shrinkage intensity under normality and non-normality. The well-known estimators in [35] and [36] are the special cases of our proposed estimators when the population mean is zero. Moreover, by plugging the above unbiased and consistent estimators into the related quantities in oracle shrinkage intensity, we obtain a set of shrinkage estimators for a class of structured target matrices in both normal and non-normal distributions.
3) For the banded target with unknown bandwidth, which emerges in many applications such as signal processing and time series analysis, we suggest a parametric method and derive analytically the expression of estimation risk for choosing the optimal bandwidth. Therefore, some non-parametric methods such as cross validation suffering from significantly higher computational complexity are avoided.
4) We evaluate the proposed covariance matrix estimation approaches via some simulation experiments and two applications to the adaptive beamforming and the discriminant analysis of real gene expression data. The results show that the comparable performance of the proposed estimators with the existing estimators for large-dimension cases under both normality and non-normality.

This paper is organized as follows. Section II introduces the class of target matrices and formulates the shrinkage estimation for large covariance matrices. The optimal oracle intensity and the related available estimators under normal and non-normal distributions are proposed in Section III and Section IV respectively. After deriving an analytical risk for banded target, in Section V, a parametric method is proposed to obtain the optimal bandwidth for banded target with unknown bandwidth. In Section VI, the performance of the proposed shrinkage estimators is investigated via some numerical simulations and two applications. Some conclusions and discussions are given in Section VII. Proofs of some mathematical results are provided in the appendix.

NOTATIONS

The transpose and Hermitian transpose of a vector or matrix are indicated by the superscripts T and H respectively. For a matrix A, $\text{tr}(A)$ and $\|A\|$ represent its trace and Frobenius norm respectively. $A \circ B$ means the Hadamard (element-wise) product of two matrices A and B. The notation $\text{sign}(\cdot)$ denotes the sign of a quantity with $\text{sign}(0) = 0$. The symbols 1 and I denote the vector having all entries 1 and the identity matrix with appropriate dimension respectively. The operation $\mathbb{E}(\cdot)$ denotes the mathematical expectation of random variable and $\to^p$ means convergence in probability.

II. FORMULATION AND PRELIMINARIES

Let S be the SCM of n independent p-dimensional random vectors from population with mean zero and covariance matrix Σ. We consider a linear shrinkage estimator