Computational Statistics and Data Analysis 83 (2015) 251–261

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Nonparametric Stein-type shrinkage covariance matrix estimators in high-dimensional settings

Anestis Touloumis *

Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, CB2 0RE, United Kingdom

ARTICLE INFO

Article history:
Received 17 April 2014
Received in revised form 17 October 2014
Accepted 17 October 2014
Available online 23 October 2014

Keywords:
Covariance matrix
High-dimensional settings
Nonparametric estimation
Shrinkage estimation

ABSTRACT

Estimating a covariance matrix is an important task in applications where the number of variables is larger than the number of observations. Shrinkage approaches for estimating a high-dimensional covariance matrix are often employed to circumvent the limitations of the sample covariance matrix. A new family of nonparametric Stein-type shrinkage covariance estimators is proposed whose members are written as a convex linear combination of the sample covariance matrix and of a predefined invertible target matrix. Under the Frobenius norm criterion, the optimal shrinkage intensity that defines the best convex linear combination depends on the unobserved covariance matrix and it must be estimated from the data. A simple but effective estimation process that produces nonparametric and consistent estimators of the optimal shrinkage intensity for three popular target matrices is introduced. In simulations, the proposed Stein-type shrinkage covariance matrix estimator based on a scaled identity matrix appeared to be up to 80% more efficient than existing ones in extreme high-dimensional settings. A colon cancer dataset was analyzed to demonstrate the utility of the proposed estimators. A rule of thumb for adhoc selection among the three commonly used target matrices is recommended.

1. Introduction

The problem of estimating large covariance matrices arises frequently in modern applications, such as in genomics, cancer research, clinical trials, signal processing, financial mathematics, pattern recognition and computational convex geometry. Formally, the goal is to estimate the covariance matrix $\mathbf{\Sigma}$ based on a sample of N independent and identically distributed (i.i.d) p-variate random vectors $\mathbf{X}_1, \ldots, \mathbf{X}_N$ with mean vector $\boldsymbol{\mu}$ in the "small N, large p" paradigm, that is when N is a lot smaller compared to p. It is a well-known fact that the sample covariance matrix

$$\mathbf{S} = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T,$$

where $\bar{\mathbf{X}} = \sum_{i=1}^{N}\mathbf{X}_i/N$ is the sample mean vector, is not performing satisfactory in high-dimensional settings. For example, $\mathbf{S}$ is singular even when $\mathbf{\Sigma}$ is a strictly positive definite matrix. Recent research in estimating high-dimensional covariance matrices includes banding, tapering, penalization and shrinkage methods. We focus on the Steinian shrinkage method (Stein, 1956) as adopted by Ledoit and Wolf (2004) because it leads to covariance matrix estimators that are: (i) non-singular, (ii) well-conditioned, (iii) invariant to permutations of the order of the p variables, (iv) consistent to departures from

* Tel.: +44 1223769682.
E-mail address: Anestis.Touloumis@cruk.cam.ac.uk.

http://dx.doi.org/10.1016/j.csda.2014.10.018
0167-9473/

a multivariate normal model, (v) not necessarily sparse, (vi) expressed in closed form and (vii) computationally cheap regardless of p.

Ledoit and Wolf (2004) proposed a Stein-type covariance matrix estimator for $\mathbf{\Sigma}$ based on

$$\mathbf{S}^{\star} = (1-\lambda)\mathbf{S} + \lambda\nu\mathbf{I}_p, \tag{1}$$

where $\mathbf{I}_p$ is the $p \times p$ identity matrix, and where λ and ν minimize the risk function $\mathrm{E}\left[\|\mathbf{S}^{\star} - \mathbf{\Sigma}\|_F^2\right]$, that is

$$\lambda = \frac{\mathrm{E}\left[\|\mathbf{S} - \mathbf{\Sigma}\|_F^2\right]}{\mathrm{E}\left[\|\mathbf{S} - \nu\mathbf{I}_p\|_F^2\right]}$$

and

$$\nu = \frac{\mathrm{tr}(\mathbf{\Sigma})}{p}.$$

The optimal shrinkage intensity parameter λ in (1) suggests how much we must shrink the eigenvalues of the sample covariance matrix $\mathbf{S}$ towards the eigenvalues of the target matrix $\nu\mathbf{I}_p$. For example, $\lambda = 0$ implies no contribution of $\nu\mathbf{I}_p$ to $\mathbf{S}^{\star}$, while $\lambda = 1$ implies no contribution of $\mathbf{S}$ to $\mathbf{S}^{\star}$. Intermediate values for λ reveal the simultaneous contribution of $\mathbf{S}$ and $\nu\mathbf{I}_p$ to $\mathbf{S}^{\star}$. Despite the attractive interpretation, $\mathbf{S}^{\star}$ is not a covariance matrix estimator because ν and λ depend on the unobservable covariance matrix $\mathbf{\Sigma}$. For this reason, Ledoit and Wolf (2004) proposed to plug-in nonparametric N-consistent estimators for ν and λ in (1) and use the resulting matrix as a shrinkage covariance matrix estimator for $\mathbf{\Sigma}$. Although ν seems to be adequately estimated by $\hat{\nu} = \mathrm{tr}(\mathbf{S})/p$, we noticed via simulations that the estimator of λ proposed by Ledoit and Wolf (2004) was biased in extreme high-dimensional settings and when $\mathbf{\Sigma} = \mathbf{I}_p$. This is counter-intuitive because $\lambda = 1$ and the plug-in estimator of $\mathbf{S}^{\star}$ is expected to be as close as possible to the target matrix $\nu\mathbf{I}_p$. In addition, this observation underlines the importance of choosing a target matrix that approximates well the true underlying dependence structure. To this direction, Fisher and Sun (2011) proposed Stein-type shrinkage covariance matrix estimators for alternative target matrices. However, they are no longer nonparametric as their construction was based on a multivariate normal model assumption.

Motivated by the above, we improve estimation of the optimal shrinkage intensity by providing a consistent estimator of λ in high-dimensional settings. To construct the estimator of λ we follow three simple steps: (i) expand the expectations in the numerator and denominator of λ assuming a multivariate normal model, (ii) prove that this ratio, say $\lambda^{\star}$, is asymptotically equivalent to λ, and (iii) replace each unknown parameter in $\lambda^{\star}$ with unbiased and consistent estimators constructed using U-statistics. The last step is essential in our proposal so as to ensure consistent and nonparametric estimation of λ. Further, we relax the normality assumption in Fisher and Sun (2011) for target matrices other than $\nu\mathbf{I}_p$ in (1) and we illustrate how to estimate consistently the corresponding optimal shrinkage intensities in high-dimensional settings. In other words, we propose a new nonparametric family of Stein-type shrinkage estimators suitable for high-dimensional settings that preserve the attractive properties mentioned in the first paragraph and can accommodate arbitrary target matrices.

The rest of this paper is organized as follows. In Section 2, we present the working framework that allows us to manage the high-dimensional setting. Section 3 contains the main results where we derive consistent and nonparametric estimators for the optimal shrinkage intensity of three different target matrices. We evaluate the performance of the proposed covariance matrix estimators via simulations in Section 4. In Section 5, we illustrate the use of the proposed estimators in a colon cancer study and we recommend a rule of thumb for selecting the target matrix. In Section 6, we summarize our findings and discuss future research. The technical details can be found in Appendix. Throughout the paper, we use $\|\mathbf{A}\|_F^2 = \mathrm{tr}(\mathbf{A}^T\mathbf{A})/p$ to denote the scaled Frobenius norm of $\mathbf{A}$, $\mathrm{tr}(\mathbf{A})$ to denote the trace of the matrix $\mathbf{A}$, $\mathbf{D}_{\mathbf{A}}$ to denote the diagonal matrix with elements the diagonal elements of $\mathbf{A}$, and $\mathbf{A} \circ \mathbf{B}$ to denote the Hadamard product of the matrices $\mathbf{A}$ and $\mathbf{B}$, i.e., the matrix whose (a, b)th element is the product of the corresponding elements of $\mathbf{A}$ and $\mathbf{B}$. In the above, it is implicit that $\mathbf{A}$ and $\mathbf{B}$ are $p \times p$ matrices.

2. Framework for high-dimensional settings

Let $\mathbf{X}_1, \ldots, \mathbf{X}_N$ be a sample of i.i.d. p-variate random vectors from the nonparametric model

$$\mathbf{X}_i = \mathbf{\Sigma}^{1/2}\mathbf{Z}_i + \boldsymbol{\mu}, \tag{2}$$

where $\boldsymbol{\mu} = \mathrm{E}[\mathbf{X}_i]$ is the p-variate mean vector, $\mathbf{\Sigma} = \mathrm{cov}[\mathbf{X}_i] = \mathbf{\Sigma}^{1/2}\mathbf{\Sigma}^{1/2}$ is the $p \times p$ covariance matrix, and $\mathbf{Z}_1, \ldots, \mathbf{Z}_N$ is a collection of i.i.d. p-variate random vectors. Instead of distributional assumptions, moments restrictions are imposed on the random variables in $\mathbf{Z}_i$. In particular, let Z_{ia} be the ath random variable in $\mathbf{Z}_i$ and suppose that $\mathrm{E}[Z_{ia}] = 0$, $\mathrm{E}[Z_{ia}^2] = 1$, $\mathrm{E}[Z_{ia}^4] = 3 + B$ with $-2 \leq B < \infty$ and for any nonnegative integers $l_1, \ldots, l_4$ such that $\sum_{\nu=1}^{4} l_\nu \leq 4$

$$\mathrm{E}[Z_{ia_1}^{l_1} Z_{ia_2}^{l_2} Z_{ia_3}^{l_3} Z_{ia_4}^{l_4}] = \mathrm{E}[Z_{ia_1}^{l_1}]\mathrm{E}[Z_{ia_2}^{l_2}]\mathrm{E}[Z_{ia_3}^{l_2}]\mathrm{E}[Z_{ia_4}^{l_4}], \tag{3}$$

where the indexes $a_1, \ldots, a_4$ are distinct. The nonparametric model (2) includes the p-variate normal distribution $\mathrm{N}_p(\boldsymbol{\mu}, \mathbf{\Sigma})$ as a special case obtained if Z_{ia} are i.i.d. $\mathrm{N}(0, 1)$ random variables. Since $B = 0$ under a multivariate normal model, B can be interpreted as a measure of departure of the fourth moment of Z_{ia} to that of a $\mathrm{N}(0, 1)$ random variable. The assumption of