ARTICLE IN PRESS
Journal of Multivariate Analysis ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Direct shrinkage estimation of large dimensional precision matrix

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

[a] Department of Mathematics, Stockholm University, Roslagsvägen 101, SE-10691 Stockholm, Sweden
[b] Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43403, USA
[c] Institute of Empirical Economics, Leibniz University of Hannover, 30167, Germany

ARTICLE INFO

Article history:
Received 9 January 2015
Available online xxxx

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

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

ABSTRACT

In this work we construct an optimal shrinkage estimator for the precision matrix in high dimensions. We consider the general asymptotics when the number of variables $p \to \infty$ and the sample size $n \to \infty$ so that $p/n \to c \in (0, +\infty)$. The precision matrix is estimated directly, without inverting the corresponding estimator for the covariance matrix. The recent results from random matrix theory allow us to find the asymptotic deterministic equivalents of the optimal shrinkage intensities and estimate them consistently. The resulting distribution-free estimator has almost surely the minimum Frobenius loss. Additionally, we prove that the Frobenius norms of the inverse and of the pseudo-inverse sample covariance matrices tend almost surely to deterministic quantities and estimate them consistently. Using this result, we construct a bona fide optimal linear shrinkage estimator for the precision matrix in case $c < 1$. At the end, a simulation is provided where the suggested estimator is compared with the estimators proposed in the literature. The optimal shrinkage estimator shows significant improvement even for non-normally distributed data.

1. Introduction

The estimation of the covariance matrix, as well as its inverse (the precision matrix), plays an important role in many disciplines from finance and genetics to wireless communications and engineering. In fact, having a suitable estimator for the precision matrix we are able to construct a good estimator for different types of optimal portfolios (see [44,19]). Similarly, in the array processing, the beamformer or the so-called minimum variance distortionless response spatial filter is defined in terms of the precision matrix (see e.g., [58]). In practice, however, the true precision matrix is unknown and a feasible estimator, constructed from data, must be used.

If the number of variables p is much smaller than the sample size n we can use the sample estimator which is biased but a consistent estimator for the precision matrix (see e.g., [7]). This case is known in the multivariate statistics as the "standard asymptotics" (see [41]). There are many findings on the estimation of the precision matrix when a particular distribution assumption is imposed. For example, the estimation of the precision matrix under the multivariate normal distribution was considered by Krishnamoorthy and Gupta [38], Gupta and Ofori-Nyarko [32–34], Kubokawa [39] and Tsukuma and Konno [57]. The results in the case of multivariate Pearson type II distribution as well as the multivariate

* Corresponding author.
E-mail address: taras.bodnar@math.su.se (T. Bodnar).

http://dx.doi.org/10.1016/j.jmva.2015.09.010
0047-259X/© 2015 Elsevier Inc. All rights reserved.

Please cite this article in press as: T. Bodnar, et al., Direct shrinkage estimation of large dimensional precision matrix, Journal of Multivariate Analysis (2015), http://dx.doi.org/10.1016/j.jmva.2015.09.010

elliptically contoured stable distribution are obtained by Sarr and Gupta [50] as well as by Bodnar and Gupta [9] and Gupta et al. [35], respectively.

Unfortunately, in practice p is often comparable in size to n or even is greater than n, i.e., we are in the situation when both the sample size n and the dimension p tend to infinity but their ratio keeps (tends to) a positive constant. This case often arises in finance when the number of assets is comparable or even greater than the number of observations for each asset. Similarly, in genetics, the data set can be huge comparable to the number of patients. Both examples illustrate the importance of the results obtained for $p, n \to \infty$.

We deal with this type of asymptotics, called the "large dimensional asymptotics" and also known as the "Kolmogorov asymptotics", in the present paper. More precisely, it is assumed that the dimension $p \equiv p(n)$ is a function of the sample size n and $p/n \to c \in (0, +\infty)$ as $n \to \infty$. This general type of asymptotics was intensively studied by several authors (see [26,27,12] etc.). In this asymptotics the usual estimators for the precision matrix perform poorly and are not consistent anymore. There are some techniques which can be used to handle the problem. Assuming that the covariance (precision) matrix has a sparse structure, significant improvements have already been achieved (see [13,14,16]). For the low-rank covariance matrices see the work of Rohde and Tsybakov [47]. An interesting nonparanormal graphic model was recently proposed by Xue and Zou [60]. Also, in order to estimate the large dimensional covariance matrix the method of block thresholding can be applied (see [15]). If the covariance matrix has a factor structure then the progress has been made by Fan et al. [20].

However, if neither the assumption about the structure of covariance (precision) matrix nor about a particular distribution is imposed, not many results are known in the literature which are based on the shrinkage estimators in high-dimensional setting (cf. [42,40,10,59]). The shrinkage estimator was first developed by Stein [55] and forms a linear combination of the sample estimator and some target. The corresponding shrinkage coefficients are often called shrinkage intensities. Ledoit and Wolf [42] proposed to shrink the sample covariance matrix to the identity matrix and showed that the resulting estimator is well-behaved in large dimensions. This estimator is called the linear shrinkage estimator because it shrinks the eigenvalues of the sample covariance matrix linearly. Recently, Bodnar et al. [10] proposed a generalization of the linear shrinkage estimator, where the shrinkage target was chosen to be an arbitrary nonrandom matrix and they showed the almost sure convergence of the derived estimator to its oracle.

The aim of our paper is to construct a feasible estimator for the precision matrix using the linear shrinkage technique and random matrix theory. In contrast to well-known procedures, we shrink the inverse of the sample covariance matrix itself instead of shrinking the sample covariance matrix and then inverting it. The direct shrinkage estimation of the precision matrix can be used in several important practical situations where the application of the inverse of the shrinkage estimator of the covariance matrix does not perform well. For instance, this could happen when the data generating process follows a factor model which is very popular in economics and finance (cf. [4,20,22,21]). In this case the largest eigenvalue of the covariance matrix is of order p and, consequently, the inverse of the linear shrinkage estimator for the covariance matrix does not work well. In the case when $c > 1$ the pseudo inverse of the sample covariance matrix is taken. The recent results from random matrix theory allow us to find the asymptotics of the optimal shrinkage intensities and estimate them consistently.

Random matrix theory is a very fast growing branch of probability theory with many applications in statistics. It studies the asymptotic behavior of the eigenvalues of the different random matrices under general asymptotics (see e.g., [1,8]). The asymptotic behavior of the functionals of the sample covariance matrices was studied by Marčenko and Pastur [43], Yin [61], Girko and Gupta [28–30], Silverstein [51], Bai et al. [5], Bai and Silverstein [8], Rubio and Mestre [48] etc.

We extend these results in the present paper by establishing the almost sure convergence of the optimal shrinkage intensities and the Frobenius norm of the inverse sample covariance matrix. Moreover, we construct a general linear shrinkage estimator for the precision matrix which has *almost surely* the smallest Frobenius loss when both the dimension p and the sample size n increase together and $p/n \to c \in (0, +\infty)$ as $n \to \infty$. Additionally, we provide a bona fide optimal linear shrinkage estimator for the precision matrix in case $c < 1$.

The suggested approach can potentially be applied in functional data analysis (cf. [46,24,37,11,17]). For instance, Ferraty et al. [23] pointed out that functional data can be seen as a special case of a high-dimensional vector. This point has been further explored by Aneiros and Vieu [2,3]. The estimation of the covariance (precision) matrix of this high-dimensional vector can be used in determining the prediction for the dependent variable as well as the corresponding predictive design points.

The rest of the paper is organized as follows. In Section 2 we present some preliminary results from random matrix theory and formulate the assumptions used throughout the paper. In Section 3 we construct the *oracle* linear shrinkage estimator for the precision matrix and verify the main asymptotic results about the shrinkage intensities and the Frobenius norm of the inverse and pseudo-inverse sample covariance matrices. Section 4 is dedicated to the *bona fide* linear shrinkage estimator for the precision matrix while Section 5 contains the results of the simulation study. Here, the performance of the derived estimator is compared with other known estimators for the large dimensional precision matrices. Section 6 includes the summary, while the proofs of the theorems are presented in the supplementary material (Section 7).

2. Assumptions and notations

The "large dimensional asymptotics" or "Kolmogorov asymptotics" include $\frac{p}{n} \to c \in (0, +\infty)$ as both the number of variables $p \equiv p(n)$ and the sample size n tend to infinity. In this case the traditional sample estimator performs poorly or very poorly. The inverse of the sample covariance matrix $\mathbf{S}_n^{-1}$ is biased, inconsistent for $\frac{p}{n} \to c > 0$ as $n \to \infty$ and it does not exist for $c > 1$. For example, under the normality assumption $\mathbf{S}_n^{-1}$ has an inverse Wishart distribution if $c < 1$ and (cf. [31])

$$E(\mathbf{S}_n^{-1}) = \frac{n}{n-p-2}\boldsymbol{\Sigma}_n^{-1}.$$

In particular, for $p = n/2 + 2$ we have that $c = 1/2$ and $E(\mathbf{S}_n^{-1}) = 2\boldsymbol{\Sigma}_n^{-1}$. In general, as c increases the sample estimator of the precision matrix becomes worse.

We use the following notations in the paper:

- $\boldsymbol{\Sigma}_n$ stands for the covariance matrix, $\mathbf{S}_n$ denotes the corresponding sample covariance matrix.[1] The population covariance matrix $\boldsymbol{\Sigma}_n$ is a nonrandom p-dimensional positive definite matrix.
- $\|\mathbf{A}\|_F^2 = \mathrm{tr}(\mathbf{A}\mathbf{A}')$ denotes the Frobenius norm of a square matrix $\mathbf{A}$, $\|\mathbf{A}\|_{tr} = \mathrm{tr}\left[\left(\mathbf{A}\mathbf{A}'\right)^{1/2}\right]$ stands for its trace norm, while $\|\mathbf{A}\|_2$ is the spectral norm.
- The pairs $(\tau_i, \boldsymbol{v}_i)$ for $i = 1, \ldots, p$ denote the collection of eigenvalues and the corresponding orthonormal eigenvectors of the covariance matrix $\boldsymbol{\Sigma}_n$.
- $H_n(t)$ is the empirical distribution function (e.d.f.) of the eigenvalues of $\boldsymbol{\Sigma}_n$, i.e.,

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

  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. The observation matrix is defined as

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

  Only the matrix $\mathbf{Y}_n$ is observable. We know neither $\mathbf{X}_n$ nor $\boldsymbol{\Sigma}_n$ itself.
- The pairs $(\lambda_i, \mathbf{u}_i)$ for $i = 1, \ldots, p$ are the eigenvalues and the corresponding orthonormal eigenvectors of the sample covariance matrix[2]

$$\mathbf{S}_n = \frac{1}{n}\mathbf{Y}_n\mathbf{Y}_n' = \frac{1}{n}\boldsymbol{\Sigma}_n^{\frac{1}{2}}\mathbf{X}_n\mathbf{X}_n'\boldsymbol{\Sigma}_n^{\frac{1}{2}}. \tag{2.3}$$

- Similarly, the (e.d.f.) of the eigenvalues of the sample covariance matrix $\mathbf{S}_n$ is defined as

$$F_n(\lambda) = \frac{1}{p}\sum_{i=1}^{p} \mathbb{1}_{\{\lambda_i \le \lambda\}} \quad \forall\, \lambda \in \mathbb{R}. \tag{2.4}$$

- In order to handle the case when $c > 1$ we introduce the dual sample covariance matrix defined as

$$\bar{\mathbf{S}}_n = \frac{1}{n}\mathbf{Y}_n'\mathbf{Y}_n = \frac{1}{n}\mathbf{X}_n'\boldsymbol{\Sigma}_n\mathbf{X}_n \tag{2.5}$$

  with the corresponding (e.d.f.) defined by

$$\bar{F}_n(\lambda) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{\lambda_i \le \lambda\}} \quad \forall\, \lambda \in \mathbb{R}. \tag{2.6}$$

  Note that the matrix $\bar{\mathbf{S}}_n$ has the same nonzero eigenvalues as $\mathbf{S}_n$, they differ only in $|p - n|$ zero eigenvalues.

  The main assumptions, which we mention throughout the paper, are as follows

(**A1**) We assume that $H_n(t)$ converges to a limit $H(t)$ at all points of continuity of H.
(**A2**) The elements of the matrix $\mathbf{X}_n$ have uniformly bounded $4 + \varepsilon,\ \varepsilon > 0$ moments.
(**A3**) For all n large enough there exists the compact interval $[h_0, h_1]$ in $(0, +\infty)$ which contains the support of H_n.

All of these assumptions are quite general and are satisfied in many practical situations. The assumption (A1) is essential to prove the Marčenko–Pastur equation (see e.g., [51]) which is used for studying the asymptotic behavior of the spectrum of general random matrices (see e.g., [8]). The fourth moment is needed for the proof of Theorems 3.2 and 3.3. The assumption

[1] Since the dimension $p \equiv p(n)$ is a function of the sample size n, the covariance matrix $\boldsymbol{\Sigma}_n$ also depends on n via $p(n)$. That is why we make use of the subscript n for all of the considered objects in order to emphasize this fact and to simplify the notation in the paper.

[2] The sample mean vector $\bar{\mathbf{x}}$ was omitted because the 1-rank matrix $\bar{\mathbf{x}}\bar{\mathbf{x}}'$ does not influence the asymptotic behavior of the spectrum of sample covariance matrix (see [8, Theorem A.44]).

(A3) ensures that both the matrix $\boldsymbol{\Sigma}_n$ and its inverse $\boldsymbol{\Sigma}_n^{-1}$ have uniformly bounded spectral norms at infinity. It means that $\boldsymbol{\Sigma}_n$ has the uniformly bounded maximum eigenvalue and its minimum eigenvalue is greater than zero. Rubio et al. [49] pointed out that (A2) and (A3) are only some technical conditions which can be further violated. Finally, it is noted that the assumption of the independence of the columns can further be weakened by controlling the growth of the number of dependent entries, while no specific distributional assumptions are imposed on the elements of $\mathbf{Y}_n$ (see [25]).

In order to investigate the (e.d.f) $F_n(\lambda)$ the Stieltjes transform is used. For nondecreasing function with bounded variation G the Stieltjes transform is defined as

$$\forall z \in \mathbb{C}^+ \quad m_G(z) = \int_{-\infty}^{+\infty} \frac{1}{\lambda - z} dG(\lambda). \tag{2.7}$$

In our notation $\mathbb{C}^+ = \{z \in \mathbb{C} : \mathbf{Im}(z) > 0\}$ is the half-plane of complex numbers with strictly positive imaginary part and any complex number is defined as $z = \mathbf{Re}(z) + i\mathbf{Im}(z)$. More about the Stieltjes transform and its properties can be found in [52].

The Stieltjes transform of the sample (e.d.f.) $F_n(\lambda)$ for all $z \in \mathbb{C}^+$ is given by

$$m_{F_n}(z) = \frac{1}{p} \sum_{i=1}^{p} \int_{-\infty}^{+\infty} \frac{1}{\lambda - z} \delta(\lambda - \lambda_i) d\lambda = \frac{1}{p} \mathrm{tr}\{(\mathbf{S}_n - z\mathbf{I})^{-1}\} \tag{2.8}$$

where $\mathbf{I}$ is a suitable identity matrix and $\delta(\cdot)$ is the Dirac delta function.

3. Optimal linear shrinkage estimator for the precision matrix

3.1. Case $c < 1$

In this section we construct an optimal linear shrinkage estimator for the precision matrix under high-dimensional asymptotics. The estimator is an *oracle* one, i.e., it depends on unknown quantities. The corresponding *bona fide* estimator is given in Section 4. We use a procedure similar to Bodnar et al. [10] where the optimal linear shrinkage estimator for the covariance matrix was constructed. The general linear shrinkage estimator of the precision matrix $\boldsymbol{\Sigma}_n^{-1}$ for $c < 1$ is given by

$$\widehat{\boldsymbol{\Pi}}_{GSE} = \alpha_n \mathbf{S}_n^{-1} + \beta_n \boldsymbol{\Pi}_0 \quad \text{with} \quad \sup_p \frac{1}{p} \|\boldsymbol{\Pi}_0\|_F^2 \le M. \tag{3.1}$$

Note that we need the condition $c < 1$ to keep the sample covariance matrix $\mathbf{S}_n$ invertible. The assumption[3] that the target matrix $\boldsymbol{\Pi}_0$ has a uniformly bounded normalized trace norm, i.e., there exists $M > 0$ such that $\sup_p \frac{1}{p}\|\boldsymbol{\Pi}_0\|_F^2 \le M$, is rather general and it is actually needed to keep the coefficient β_n bounded for large dimensions p. This condition can be replaced with an equivalent assumption on β_n. Note that the target matrix can also be random but independent of $\mathbf{Y}_n$. In practice, $\boldsymbol{\Pi}_0 = \mathbf{I}$ is used when no information about the precision matrix is available. On the other hand, the information about the structure of the precision matrix can be included into $\boldsymbol{\Pi}_0$, like, e.g., sparsity. In order to present the results in the most general case, throughout this section we assume only that $\boldsymbol{\Pi}_0$ satisfies $\sup_p \frac{1}{p}\|\boldsymbol{\Pi}_0\|_F^2 \le M$. The choice of the target matrix $\boldsymbol{\Pi}_0$ is not treated in the paper in detail and it is left for future research.

Our aim is now to find the optimal shrinkage intensities which minimize the Frobenius-norm loss for a given nonrandom target matrix $\boldsymbol{\Pi}_0$ given by

$$L_F^2 = \|\widehat{\boldsymbol{\Pi}}_{GSE} - \boldsymbol{\Sigma}_n^{-1}\|_F^2 = \|\boldsymbol{\Sigma}_n^{-1}\|_F^2 + \|\widehat{\boldsymbol{\Pi}}_{GSE}\|_F^2 - 2\mathrm{tr}\left(\widehat{\boldsymbol{\Pi}}_{GSE}\boldsymbol{\Sigma}_n^{-1}\right).$$

As a result, using (3.1) the following optimization problem has to be solved

$$\min_{\alpha_n, \beta_n} \alpha_n^2 \|\mathbf{S}_n^{-1}\|_F^2 + 2\alpha_n\beta_n \mathrm{tr}(\mathbf{S}_n^{-1}\boldsymbol{\Pi}_0) + \beta_n^2 \|\boldsymbol{\Pi}_0\|_F^2 - 2\alpha_n \mathrm{tr}(\mathbf{S}_n^{-1}\boldsymbol{\Sigma}_n^{-1}) - 2\beta_n \mathrm{tr}(\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\Pi}_0).$$

Next, taking the derivatives of L_F^2 with respect to α_n and β_n and setting them equal to zero we get

$$\frac{\partial L_F^2}{\partial \alpha_n} = \alpha_n \|\mathbf{S}_n^{-1}\|_F^2 + \beta_n \mathrm{tr}(\mathbf{S}_n^{-1}\boldsymbol{\Pi}_0) - \mathrm{tr}(\mathbf{S}_n^{-1}\boldsymbol{\Sigma}_n^{-1}) = 0, \tag{3.2}$$

$$\frac{\partial L_F^2}{\partial \beta_n} = \alpha_n \mathrm{tr}(\mathbf{S}_n^{-1}\boldsymbol{\Pi}_0) + \beta_n \|\boldsymbol{\Pi}_0\|_F^2 - \mathrm{tr}(\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\Pi}_0) = 0. \tag{3.3}$$

The Hessian of the L_F^2 has the form

$$\mathbf{H} = \begin{pmatrix} \|\mathbf{S}_n^{-1}\|_F^2 & \mathrm{tr}(\mathbf{S}_n^{-1}\boldsymbol{\Pi}_0) \\ \mathrm{tr}(\mathbf{S}_n^{-1}\boldsymbol{\Pi}_0) & \|\boldsymbol{\Pi}_0\|_F^2 \end{pmatrix} \tag{3.4}$$

[3] The similar assumption is presented by Bodnar et al. [10] but with $\sup_p \|\boldsymbol{\Pi}_0\|_F^2 \le M$. Note that the assumption $\sup_p 1/p\|\boldsymbol{\Pi}_0\|_F^2 \le M$ is more general.

Please cite this article in press as: T. Bodnar, et al., Direct shrinkage estimation of large dimensional precision matrix, Journal of Multivariate Analysis (2015), http://dx.doi.org/10.1016/j.jmva.2015.09.010

which is always positive definite, since

$$\det(\mathbf{H}) = \|\mathbf{S}_n^{-1}\|_F^2\|\boldsymbol{\Pi}_0\|_F^2 - (\mathrm{tr}(\mathbf{S}_n^{-1}\boldsymbol{\Pi}_0))^2$$
$$\geq \|\mathbf{S}_n^{-1}\|_F^2\|\boldsymbol{\Pi}_0\|_F^2 - \|\mathbf{S}_n^{-1}\|_2^2(\mathrm{tr}(\boldsymbol{\Pi}_0))^2 \overset{Jensen}{\geq} (\|\mathbf{S}_n^{-1}\|_F^2 - \|\mathbf{S}_n^{-1}\|_2^2)\|\boldsymbol{\Pi}_0\|_F^2 > 0, \tag{3.5}$$

where the last inequality in (3.5) is taken from Horn and Johnson [36].

Thus, the optimal α_n^* and β_n^* are given by

$$\alpha_n^* = \frac{\mathrm{tr}(\mathbf{S}_n^{-1}\boldsymbol{\Sigma}_n^{-1})\|\boldsymbol{\Pi}_0\|_F^2 - \mathrm{tr}(\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\Pi}_0)\mathrm{tr}(\mathbf{S}_n^{-1}\boldsymbol{\Pi}_0)}{\|\mathbf{S}_n^{-1}\|_F^2\|\boldsymbol{\Pi}_0\|_F^2 - \left(\mathrm{tr}(\mathbf{S}_n^{-1}\boldsymbol{\Pi}_0)\right)^2}, \tag{3.6}$$

$$\beta_n^* = \frac{\mathrm{tr}(\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\Pi}_0)\|\mathbf{S}_n^{-1}\|_F^2 - \mathrm{tr}(\mathbf{S}_n^{-1}\boldsymbol{\Sigma}_n^{-1})\mathrm{tr}(\mathbf{S}_n^{-1}\boldsymbol{\Pi}_0)}{\|\mathbf{S}_n^{-1}\|_F^2\|\boldsymbol{\Pi}_0\|_F^2 - \left(\mathrm{tr}(\mathbf{S}_n^{-1}\boldsymbol{\Pi}_0)\right)^2}. \tag{3.7}$$

Now, we formulate our first main result in Theorem 3.1 which states that the normalized Frobenius norm of the inverse sample covariance matrix $1/p\|\mathbf{S}_n^{-1}\|_F^2$ tends almost surely to a nonrandom quantity.

Theorem 3.1. *Assume that* (A1) *and* (A3) *hold and* $\frac{p}{n} \to c \in (0, 1)$ *for* $n \to \infty$. *Then the normalized Frobenius norm of the inverse sample covariance matrix* $\psi_n = \frac{1}{p}\|\mathbf{S}_n^{-1}\|_F^2$ *almost surely tends to a nonrandom* ψ *which is given by*

$$\psi = \frac{1}{(1-c)^2}\int_{-\infty}^{+\infty}\frac{dH(\tau)}{\tau^2} + \frac{c}{(1-c)^3}\left(\int_{-\infty}^{+\infty}\frac{dH(\tau)}{\tau}\right)^2. \tag{3.8}$$

The proof is given in the supplementary material (see Appendix A). Theorem 3.1 presents an important result which indicates that the Frobenius norm of the inverse sample covariance matrix is asymptotically nonrandom as well as it depends on H and concentration ratio c only. Moreover, Theorem 3.1 gives us an intuitive hint how to find the asymptotic equivalent representation of $\|\mathbf{S}_n^{-1}\|_F^2$. The corresponding result is presented in Theorem 3.2.

Theorem 3.2. *Let the assumptions* (A1)–(A3) *hold and* $\frac{p}{n} \to c \in (0, 1)$. *Then as* $n \to \infty$,

$$\frac{1}{p}\left|\|\mathbf{S}_n^{-1}\|_F^2 - \left(\frac{1}{(1-c)^2}\|\boldsymbol{\Sigma}_n^{-1}\|_F^2 + \frac{c}{p(1-c)^3}\|\boldsymbol{\Sigma}_n^{-1}\|_{tr}^2\right)\right| \xrightarrow[a.s.]{} 0. \tag{3.9}$$

Additionally, for the quantity $tr(\mathbf{S}_n^{-1}\boldsymbol{\Theta})$ *with a symmetric positive definite matrix* $\boldsymbol{\Theta}$ *which has uniformly bounded trace norm as* $n \to \infty$,

$$\left|tr(\mathbf{S}_n^{-1}\boldsymbol{\Theta}) - \frac{1}{1-c}tr(\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\Theta})\right| \xrightarrow[a.s.]{} 0 \quad for\ \frac{p}{n} \to c \in (0, 1). \tag{3.10}$$

The theorem is proved in the supplementary material (see Appendix A). Theorem 3.2 provides us the asymptotic behavior of the Frobenius norm of the inverse sample covariance matrix and of the functional $\mathrm{tr}(\mathbf{S}_n^{-1}\boldsymbol{\Theta})$. It shows that the consistent estimator for the Frobenius norm of the precision matrix under the general asymptotics is not equal to its sample counterpart. Using Theorem 3.2 we can easily determine the asymptotic bias of the sample estimator which consists of the two types of biases. The multiplicative bias is violated by multiplying $\|\mathbf{S}_n^{-1}\|_F^2$ by $(1-c)^2$. After that, the additive bias is dealt by subtracting $\frac{c}{p(1-c)}\|\boldsymbol{\Sigma}_n^{-1}\|_{tr}^2$. The sample estimator of the functional $\mathrm{tr}(\mathbf{S}_n^{-1}\boldsymbol{\Theta})$ is also not a consistent estimator for $\mathrm{tr}(\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\Theta})$. The consistent estimator is obtained by multiplying $\mathrm{tr}(\mathbf{S}_n^{-1}\boldsymbol{\Theta})$ by the constant $(1-c)$.

Results similar to those given in Theorems 3.1 and 3.2 are also available for the estimation of the population covariance matrix (cf. [10]). However, in the case of the covariance matrix, the sample estimator for the Frobenius norm possesses only the additive bias $\frac{c}{p}\mathrm{tr}(\|\boldsymbol{\Sigma}_n^{-1}\|_{tr})$, while $\mathrm{tr}(\mathbf{S}_n\boldsymbol{\Theta})$ is a consistent estimator for $\mathrm{tr}(\boldsymbol{\Sigma}_n\boldsymbol{\Theta})$.

Next, we show that the optimal shrinkage intensities α_n^* and β_n^* are almost surely asymptotic equivalent to nonrandom quantities α^* and β^* under the large-dimensional asymptotics $\frac{p}{n} \to c \in (0, 1)$.

Corollary 3.1. *Assume that* (A1)–(A3) *hold and* $\frac{p}{n} \to c \in (0, 1)$ *for* $n \to \infty$. *Then for the optimal shrinkage intensities* α_n^* *and* β_n^*

$$\left|\alpha_n^* - \alpha^*\right| \xrightarrow[a.s.]{} 0, \tag{3.11}$$

where

$$\alpha^* = (1-c)\frac{\|\boldsymbol{\Sigma}_n^{-1}\|_F^2\|\boldsymbol{\Pi}_0\|_F^2 - \left(tr(\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\Pi}_0)\right)^2}{\left(\|\boldsymbol{\Sigma}_n^{-1}\|_F^2 + \frac{c}{p(1-c)}\|\boldsymbol{\Sigma}_n^{-1}\|_{tr}^2\right)\|\boldsymbol{\Pi}_0\|_F^2 - \left(tr(\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\Pi}_0)\right)^2} \tag{3.12}$$

and

$$\left|\beta_n^* - \beta^*\right| \xrightarrow[a.s.]{} 0 \quad with \quad \beta^* = \frac{tr(\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\Pi}_0)}{\|\boldsymbol{\Pi}_0\|_F^2}\left(1 - \frac{\alpha^*}{1-c}\right). \tag{3.13}$$

Note that both the asymptotic optimal intensities α^* and β^* are always positive as well as $\alpha^* \in (0, 1-c)$ due to inequality (3.5) and $c \in (0, 1)$. Using these results we are immediately able to estimate α^*, β^* consistently which is shown in Section 4.

3.2. Case $c > 1$

In this subsection we deal with the problem of the estimation of the precision matrix when the dimension p is greater than the sample size n, i.e., $c > 1$. This case is very difficult to handle because of the loss of information as c becomes greater than one. Moreover, the sample covariance matrix $\mathbf{S}_n$ is not invertible and thus the estimator $\mathbf{S}_n^{-1}$ must be replaced by a suitable one. This is usually done by using the generalized inverse matrix $\mathbf{S}_n^+$ instead of $\mathbf{S}_n^{-1}$. In this case the general shrinkage estimator has the form

$$\widehat{\boldsymbol{\Pi}}_{GSE} = \tilde{\alpha}_n\mathbf{S}_n^+ + \tilde{\beta}_n\boldsymbol{\Pi}_0 \quad \text{with} \quad \sup_p \frac{1}{p}\|\boldsymbol{\Pi}_0\|_F^2 \le M. \tag{3.14}$$

The optimal shrinkage intensities $\tilde{\alpha}_n^*$ and $\tilde{\beta}_n^*$ are determined following the procedure of Section 3.1. They are given by

$$\tilde{\alpha}_n^* = \frac{\mathrm{tr}(\mathbf{S}_n^+\boldsymbol{\Sigma}_n^{-1})\|\boldsymbol{\Pi}_0\|_F^2 - \mathrm{tr}(\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\Pi}_0)\mathrm{tr}(\mathbf{S}_n^+\boldsymbol{\Pi}_0)}{\|\mathbf{S}_n^+\|_F^2\|\boldsymbol{\Pi}_0\|_F^2 - \left(\mathrm{tr}(\mathbf{S}_n^+\boldsymbol{\Pi}_0)\right)^2}, \tag{3.15}$$

$$\tilde{\beta}_n^* = \frac{\mathrm{tr}(\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\Pi}_0)\|\mathbf{S}_n^+\|_F^2 - \mathrm{tr}(\mathbf{S}_n^+\boldsymbol{\Sigma}_n^{-1})\mathrm{tr}(\mathbf{S}_n^+\boldsymbol{\Pi}_0)}{\|\mathbf{S}_n^+\|_F^2\|\boldsymbol{\Pi}_0\|_F^2 - \left(\mathrm{tr}(\mathbf{S}_n^+\boldsymbol{\Pi}_0)\right)^2}. \tag{3.16}$$

In Theorem 3.3 we derive the asymptotic properties of two quantities used in (3.15) and (3.16), namely $\mathrm{tr}(\boldsymbol{\Theta}\mathbf{S}_n^+)$ and $\|\mathbf{S}_n^+\|_F^2$.

Theorem 3.3. *Let the assumptions* (A1)–(A3) *hold and* $\frac{p}{n} \to c \in (1, +\infty)$. *Then as* $n \to \infty$,

$$\left|\frac{1}{p}\|\mathbf{S}_n^+\|_F^2 - c^{-1}x'(0)\right| \xrightarrow[a.s.]{} 0, \quad where\ x'(0) = \frac{1}{\frac{1}{x^2(0)} - \frac{c}{p}tr\left[\left(\boldsymbol{\Sigma}_n^{-1} + x(0)\mathbf{I}\right)^{-2}\right]} \tag{3.17}$$

and $x(0)$ *is the unique solution of the equation*

$$\frac{1}{x(0)} = \frac{c}{p}tr\left[\left(\boldsymbol{\Sigma}_n^{-1} + x(0)\mathbf{I}\right)^{-1}\right]. \tag{3.18}$$

Additionally, for the quantity $tr(\boldsymbol{\Theta}\mathbf{S}_n^+)$ *with a symmetric positive definite matrix* $\boldsymbol{\Theta}$ *which has uniformly bounded spectral norm, as* $n \to \infty$,

$$\left|\frac{1}{p}tr(\boldsymbol{\Theta}\mathbf{S}_n^+) - c^{-1}y(\boldsymbol{\Theta})\right| \xrightarrow[a.s.]{} 0 \quad for\ \frac{p}{n} \to c \in (1, +\infty), \tag{3.19}$$

where $y(\boldsymbol{\Theta})$ *is the solution of*

$$\frac{1}{y(\boldsymbol{\Theta})} = \frac{c}{p}tr\left[\left(\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{\Theta}\boldsymbol{\Sigma}_n^{-1/2} + y(\boldsymbol{\Theta})\mathbf{I}\right)^{-1}\right]. \tag{3.20}$$

The proof of Theorem 3.3 is given in the supplementary material (see Appendix A). The results of Theorem 3.3 show that using the generalized inverse technique it is not clear how to estimate the functionals of $\boldsymbol{\Sigma}_n^{-1}$ consistently. The asymptotic values obtained in Theorem 3.3 are far away from the desired ones. In order to correct these biases, we need to solve the non-linear equation (3.18) and (3.20), respectively, which appears to be a difficult task. Finally, we notice, that the quantities $x(0)$ and $x'(0)$, however, can be estimated consistently using Theorem 3.3.

In an important special case when the matrix $\boldsymbol{\Theta} = \boldsymbol{\xi}\boldsymbol{\eta}'$ for some $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ with bounded Euclidean norms we get the following result summarized in Proposition 3.1 which is proved in the supplementary material (see Appendix A).

Proposition 3.1. *Under the assumptions of Theorem 3.3 and $\boldsymbol{\Theta} = \boldsymbol{\xi}\boldsymbol{\eta}'$ for some $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ with bounded Euclidean norms it holds*

$$\frac{1}{p}\left|\boldsymbol{\eta}'\mathbf{S}_n^{+}\boldsymbol{\xi} - \frac{c^{-1}}{c-1}\boldsymbol{\eta}'\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\xi}\right| \xrightarrow[a.s.]{} 0 \quad \text{for } \frac{p}{n} \to c \in (1, +\infty). \tag{3.21}$$

It is remarkable to note that the results of Proposition 3.1 are very similar to those presented in Theorem 3.2 if $\boldsymbol{\Theta} = \boldsymbol{\xi}\boldsymbol{\xi}'$. Here, the constant $1 - c$ need to be replaced by $c(c-1)$.

Next we use the asymptotic results of Theorem 3.3 for finding the asymptotic equivalents to the optimal shrinkage intensities $\tilde{\alpha}_n^*$ and $\tilde{\beta}_n^*$ given in (3.15) and (3.16), respectively.

Corollary 3.2. *Assume that (A1)–(A3) hold and $\frac{p}{n} \to c \in (1, +\infty)$ for $n \to \infty$. Then for the optimal shrinkage intensities α_n^* and β_n^* from (3.15) and (3.16) holds*

$$\left|\alpha_n^* - \alpha^*\right| \xrightarrow[a.s.]{} 0 \quad \text{with} \quad \alpha^* = \frac{y(\boldsymbol{\Sigma}_n^{-1})\|\boldsymbol{\Pi}_0\|_F^2 - y(\boldsymbol{\Pi}_0)tr(\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\Pi}_0)}{x'(0)\|\boldsymbol{\Pi}_0\|_F^2 - c^{-1}y^2(\boldsymbol{\Pi}_0)} \tag{3.22}$$

and

$$\left|\beta_n^* - \beta^*\right| \xrightarrow[a.s.]{} 0 \quad \text{with} \quad \beta^* = \frac{tr(\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\Pi}_0)x'(0) - y(\boldsymbol{\Sigma}_n^{-1})y(\boldsymbol{\Pi}_0)}{cx'(0)\|\boldsymbol{\Pi}_0\|_F^2 - y^2(\boldsymbol{\Pi}_0)}. \tag{3.23}$$

Even if the target matrix $\boldsymbol{\Pi}_0$ is chosen as a one-rank matrix, i.e., $\boldsymbol{\Pi}_0 = \boldsymbol{\xi}\boldsymbol{\eta}'$, we are not able to provide consistent estimates for α^* and β^* without an additional assumption imposed on $\boldsymbol{\Sigma}_n$. One of possible assumptions for which α^* and β^* are consistently estimable is $\boldsymbol{\Sigma}_n = \sigma\mathbf{I}$ as illustrated in Corollary 3.4. If $\boldsymbol{\Sigma}_n = \sigma\mathbf{I}$, then for $\frac{1}{p}\|\mathbf{S}_n^{+}\|_F^2$ and $\frac{1}{p}\text{tr}(\mathbf{S}_n^{+})$ we get

Corollary 3.3. *Under the assumptions of Theorem 3.3 assume additionally that $\boldsymbol{\Sigma}_n = \sigma\mathbf{I}$. Then it holds as $n \to \infty$*

$$\left|\frac{1}{p}\|\mathbf{S}_n^{+}\|_F^2 - \frac{\sigma^{-2}}{(c-1)^3}\right| \xrightarrow[a.s.]{} 0. \tag{3.24}$$

Additionally, for the quantity $tr(\mathbf{S}_n^{+})$ as $n \to \infty$ the norm

$$\left|\frac{1}{p}tr(\mathbf{S}_n^{+}) - \frac{c^{-1}}{(c-1)}\sigma^{-1}\right| \xrightarrow[a.s.]{} 0 \quad \text{for } \frac{p}{n} \to c \in (1, +\infty). \tag{3.25}$$

The proof of Corollary 3.3 is based on the fact that Eq. (3.18) has the explicit solution $x(0) = \frac{\sigma^{-1}}{c-1}$ if $\boldsymbol{\Sigma}_n = \sigma\mathbf{I}$. The rest calculations are only technical ones. It is interesting to note that the result of Corollary 3.3 coincides with the corresponding one of Theorem 3.2 for $c < 1$ if $\boldsymbol{\Sigma}_n = \sigma\mathbf{I}$.

Next we apply Corollary 3.3 with $\boldsymbol{\Sigma}_n = \sigma\mathbf{I}$ and $\boldsymbol{\Pi}_0 = \mathbf{I}$ to construct the asymptotic equivalents to the optimal shrinkage intensities $\tilde{\alpha}_n^*$ and $\tilde{\beta}_n^*$ given in (3.15) and (3.16), respectively.

Corollary 3.4. *Assume that (A1)–(A3) hold, $\boldsymbol{\Sigma}_n = \sigma\mathbf{I}$, $\boldsymbol{\Pi}_0 = \mathbf{I}$ and $\frac{p}{n} \to c \in (1, +\infty)$ for $n \to \infty$. Then for the optimal shrinkage intensities α_n^* and β_n^* from (3.15) and (3.16) holds*

$$\tilde{\alpha}_n^* \xrightarrow[a.s.]{} 0 \quad \text{and} \quad \tilde{\beta}_n^* \xrightarrow[a.s.]{} \sigma^{-1}. \tag{3.26}$$

Corollary 3.4 implies that the oracle optimal shrinkage estimator for the precision matrix in the case $c > 1$ and $\boldsymbol{\Sigma}_n = \sigma\mathbf{I}$ is equal to

$$\widehat{\boldsymbol{\Pi}}_{GSE} = \boldsymbol{\Sigma}_n^{-1} = \sigma^{-1}\mathbf{I}. \tag{3.27}$$

The quantity $\sigma^{-1} = \frac{1}{p}\text{tr}(\boldsymbol{\Sigma}_n^{-1})$ can be easily estimated using the result of Corollary 3.3. Namely, the consistent estimator of σ^{-1} is given by

$$\hat{\sigma}^{-1} = p/n\frac{p/n - 1}{p}\text{tr}(\mathbf{S}_n^{+}). \tag{3.28}$$

However, in the general case when $\boldsymbol{\Sigma}_n$ and $\boldsymbol{\Pi}_0$ are arbitrary, the results of Corollaries 3.3 and 3.4 do not hold anymore.

4. Estimation of unknown parameters

In this section we present consistent estimators for the asymptotic optimal shrinkage intensities derived in Section 3. The results of Theorem 3.2 allow us to estimate consistently the functionals of type $\mathrm{tr}(\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\Theta})$ and the Frobenius norm of the precision matrix. The consistent estimator for the functional $\theta_n(\boldsymbol{\Theta}) = \mathrm{tr}(\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\Theta})$ is given by

$$\hat{\theta}_n(\boldsymbol{\Theta}) = (1 - p/n)\mathrm{tr}(\mathbf{S}_n^{-1}\boldsymbol{\Theta}) \tag{4.1}$$

which is a generalization of the so-called $G3$-estimator obtained by Girko [27]. In particular, in the case when $\boldsymbol{\Theta} = \boldsymbol{\xi}\boldsymbol{\eta}'$ for some vectors $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ with bounded Euclidean norm, Girko [27] showed that the corresponding estimator $\hat{\theta}_n(\boldsymbol{\xi}\boldsymbol{\eta}')$ tends to $\theta_n(\boldsymbol{\xi}\boldsymbol{\eta}')$ in probability. In contrast, Theorem 3.2 ensures the consistency of $\hat{\theta}_n(\boldsymbol{\Theta})$ for more general forms of $\boldsymbol{\Theta}$ which should not be of rank 1.

Again, using (4.1) and Theorem 3.2 we construct a consistent estimator for $\rho_n = 1/p\|\boldsymbol{\Sigma}_n^{-1}\|_F^2$ which is given by

$$\hat{\rho}_n = \frac{(1-p/n)^2}{p}\|\mathbf{S}_n^{-1}\|_F^2 - \frac{1-p/n}{pn}\|\mathbf{S}_n^{-1}\|_{tr}^2. \tag{4.2}$$

Note that the result (4.2) is entirely new and it was not mentioned in the literature up to now. Moreover, it is noted that for the derivation of (4.2) we do not need the existence of the 4th moment (see the assumption (A2) in Section 2).

Using both the estimators (4.1) and (4.2), we are able now to construct the optimal linear shrinkage estimator (OLSE) for the precision matrix which is given by

$$\widehat{\boldsymbol{\Pi}}_{OLSE} = \widehat{\alpha}_n^*\mathbf{S}_n^{-1} + \widehat{\beta}_n^*\boldsymbol{\Pi}_0 \quad \text{with} \quad \sup_p 1/p\|\boldsymbol{\Pi}_0\|_{tr} \le M, \tag{4.3}$$

where

$$\begin{aligned}
\widehat{\alpha}_n^* &= (1-p/n)\frac{p\hat{\rho}_n\|\boldsymbol{\Pi}_0\|_F^2 - \hat{\theta}_n^2(\boldsymbol{\Pi}_0)}{\left(p\hat{\rho}_n + \frac{p/n}{p(1-p/n)}\hat{\theta}_n^2(\mathbf{I})\right)\|\boldsymbol{\Pi}_0\|_F^2 - \hat{\theta}_n^2(\boldsymbol{\Pi}_0)} \\
&= (1-p/n)\left(1 - \frac{\frac{p/n}{p(1-p/n)}\hat{\theta}_n^2(\mathbf{I})\|\boldsymbol{\Pi}_0\|_F^2}{\left(p\hat{\rho}_n + \frac{p/n}{p(1-p/n)}\hat{\theta}_n^2(\mathbf{I})\right)\|\boldsymbol{\Pi}_0\|_F^2 - \hat{\theta}_n^2(\boldsymbol{\Pi}_0)}\right) \\
&= 1 - p/n - \frac{\frac{1}{n}\|\mathbf{S}_n^{-1}\|_{tr}^2\|\boldsymbol{\Pi}_0\|_F^2}{\|\mathbf{S}_n^{-1}\|_F^2\|\boldsymbol{\Pi}_0\|_F^2 - \left(\mathrm{tr}(\mathbf{S}_n^{-1}\boldsymbol{\Pi}_0)\right)^2}
\end{aligned} \tag{4.4}$$

and

$$\widehat{\beta}_n^* = \frac{\hat{\theta}_n(\boldsymbol{\Pi}_0)}{\|\boldsymbol{\Pi}_0\|_F^2}\left(1 - \frac{\widehat{\alpha}_n^*}{1-p/n}\right) = \frac{\mathrm{tr}(\mathbf{S}_n^{-1}\boldsymbol{\Pi}_0)}{\|\boldsymbol{\Pi}_0\|_F^2}\left(1 - p/n - \widehat{\alpha}_n^*\right). \tag{4.5}$$

The bona fide OLSE estimator (4.3) is optimal in the sense that it minimizes the Frobenius loss. It means that the estimators $\widehat{\alpha}_n^*$ and $\widehat{\beta}_n^*$ tend almost surely to their oracle asymptotic values (3.12) and (3.13) as $n \to \infty$, respectively. According to Corollary 3.1 the oracle optimal intensities α_n^* and β_n^* given in (3.6) and (3.7) behave similarly. It is a remarkable property of the OLSE estimator which indicates that the bona fide OLSE estimator tends almost surely to its oracle one. Moreover, using the inequality (3.5) it can be easily verified that the estimated optimal shrinkage intensities $\widehat{\alpha}_n^*$ and $\widehat{\beta}_n^*$ are almost surely positive and $\widehat{\alpha}_n^*$ has the support $(0, 1-p/n)$. Only in the case when $p/n \to c = 0$ as $n \to \infty$ the shrinkage intensities satisfy $\widehat{\alpha}_n^* \to 1$ and $\widehat{\beta}_n^* \to 0$. In this case the OLSE estimator coincides with the sample estimator which is consistent for the standard asymptotics.

Next we present the asymptotic Frobenius losses of the suggested optimal linear shrinkage estimator, the traditional estimator and different benchmark estimators which are frequently used in statistical literature. The first benchmark is the optimal linear shrinkage estimator (OLSE) for the covariance matrix $\boldsymbol{\Sigma}_n$, $c \in (0, +\infty)$, suggested recently by Bodnar et al. [10]. It is given by $\widehat{\boldsymbol{\Sigma}}_{OLSE}^{-1}$ with

$$\widehat{\boldsymbol{\Sigma}}_{OLSE} = \hat{\alpha}_C^*\mathbf{S}_n + \hat{\beta}_C^*\boldsymbol{\Sigma}_0, \quad \|\boldsymbol{\Sigma}_0\|_{tr} \le M, \tag{4.6}$$

where $\boldsymbol{\Sigma}_0$ is a positive definite symmetric target matrix,

$$\hat{\alpha}_C^* = 1 - \frac{\frac{1}{n}\|\mathbf{S}_n\|_{tr}^2\|\boldsymbol{\Sigma}_0\|_F^2}{\|\mathbf{S}_n\|_F^2\|\boldsymbol{\Sigma}_0\|_F^2 - \left(\mathrm{tr}(\mathbf{S}_n\boldsymbol{\Sigma}_0)\right)^2} \quad \text{and} \quad \hat{\beta}_C^* = \frac{\mathrm{tr}(\mathbf{S}_n\boldsymbol{\Sigma}_0)}{\|\boldsymbol{\Sigma}_0\|_F^2}\left(1 - \hat{\alpha}_C^*\right). \tag{4.7}$$

Bodnar et al. [10] proved that their $\widehat{\boldsymbol{\Sigma}}_{OLSE}$ estimator possesses asymptotically almost surely the smallest Frobenius loss over all linear shrinkage estimators. Moreover, they showed by simulations that if the target matrix is $\boldsymbol{\Sigma}_0 = 1/p\mathbf{I}$ then the

Please cite this article in press as: T. Bodnar, et al., Direct shrinkage estimation of large dimensional precision matrix, Journal of Multivariate Analysis (2015), http://dx.doi.org/10.1016/j.jmva.2015.09.010

estimator $\widehat{\boldsymbol{\Sigma}}_{OLSE}$ is asymptotically equivalent to the linear shrinkage estimator proposed by Ledoit and Wolf [42] for normally distributed data.

The second considered estimator for the precision matrix is the scaled standard estimator (SSE), which is presented by several authors, e.g., Stein [56], Mestre and Lagunas [45], Srivastava [53] and Kubokawa and Srivastava [40]. It is given by

$$\widehat{\boldsymbol{\Pi}}_{SSE} = \frac{n-p-2}{n-1}\mathbf{S}_n^{-1}\delta_{(p<n)} + \frac{p}{n-1}\mathbf{S}_n^{+}\delta_{(p\geq n)} \tag{4.8}$$

where $\mathbf{S}_n^{+}$ is the Moore–Penrose inverse of $\mathbf{S}_n$ and $\delta_{(\cdot)}$ is a Dirac delta function. The third benchmark proposed by Efron and Morris [18] is expressed as

$$\widehat{\boldsymbol{\Pi}}_{EM} = \frac{n-p-2}{n-1}\mathbf{S}_n^{-1} + \frac{p^2+p-2}{(n-1)\mathrm{tr}(\mathbf{S}_n)}\mathbf{I}. \tag{4.9}$$

The last one is the empirical Bayes estimator considered in [40]

$$\widehat{\boldsymbol{\Pi}}_{KS} = p\left((n-1)\mathbf{S}_n + \mathrm{tr}(\mathbf{S}_n)\mathbf{I}\right)^{-1}. \tag{4.10}$$

In order to compare the asymptotic performance of the considered estimators, we make use of the normalized Frobenius loss as a performance measure, i.e.,

$$L(\mathbf{M}) = \frac{1}{p}\|\mathbf{M} - \boldsymbol{\Sigma}_n^{-1}\|_F^2 \tag{4.11}$$

where $\mathbf{M}$ is an arbitrary estimator for the precision matrix $\boldsymbol{\Sigma}_n^{-1}$. The following corollary follows from Theorem 3.1.

Corollary 4.1. *Under the assumptions of Theorem 3.1, it holds that*

$$L(\widehat{\boldsymbol{\Pi}}_{OLSE}) \xrightarrow[a.s.]{} \frac{1}{p}\beta^{*\,2}\|\boldsymbol{\Pi}_0\|_F^2\left(\frac{\|\boldsymbol{\Sigma}_n^{-1}\|_F^2\|\boldsymbol{\Pi}_0\|_F^2}{(tr[\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\Pi}_0])^2} - 1\right) + \frac{\alpha^{*\,2}}{(1-c)^2}\frac{c}{1-c}\frac{1}{p^2}\left(tr(\boldsymbol{\Sigma}_n^{-1})\right)^2$$

$$L(\mathbf{S}_n^{-1}) \xrightarrow[a.s.]{} \frac{c^2}{(1-c)^2}\frac{1}{p}\|\boldsymbol{\Sigma}_n^{-1}\|_F^2 + \frac{c}{1-c}\frac{1}{p^2}\left(tr(\boldsymbol{\Sigma}_n^{-1})\right)^2$$

$$L(\widehat{\boldsymbol{\Pi}}_{SSE}) \xrightarrow[a.s.]{} \frac{c}{1-c}\frac{1}{p^2}\left(tr(\boldsymbol{\Sigma}_n^{-1})\right)^2$$

$$L(\widehat{\boldsymbol{\Pi}}_{EM}) \xrightarrow[a.s.]{} \left(\frac{c}{1/p\,tr(\boldsymbol{\Sigma}_n)}\right)^2 + \frac{c}{1-c}\frac{1}{p^2}\left(tr(\boldsymbol{\Sigma}_n^{-1})\right)^2$$

$$L(\widehat{\boldsymbol{\Sigma}}_{OLSE}^{-1}) \xrightarrow[a.s.]{} \frac{1}{p}\|\boldsymbol{\Sigma}_n^{-1}\|_F^2 + \frac{1}{p}tr\left[\left(\alpha_C^*(1-c)\boldsymbol{\Sigma}_n + \beta_C^*\boldsymbol{\Sigma}_0\right)^{-2}\left(\frac{c\alpha_C^*\frac{1}{p}tr(\boldsymbol{\Sigma}_n^{-1})}{1-c}\boldsymbol{\Sigma}_n + \mathbf{I}\right)\right]$$
$$-2\frac{1}{p}tr\left[\left(\alpha_C^*(1-c)\boldsymbol{\Sigma}_n + \beta_C^*\boldsymbol{\Sigma}_0\right)^{-1}\boldsymbol{\Sigma}_n^{-1}\right]$$

$$L(\widehat{\boldsymbol{\Pi}}_{KS}) \xrightarrow[a.s.]{} \frac{1}{p}\|\boldsymbol{\Sigma}_n^{-1}\|_F^2 + \frac{1}{p}tr\left[\left(\frac{1-c}{c}\boldsymbol{\Sigma}_n + \frac{tr(\boldsymbol{\Sigma}_n)}{p}\mathbf{I}\right)^{-2}\left(\frac{\frac{1}{p}tr(\boldsymbol{\Sigma}_n^{-1})}{1-c}\boldsymbol{\Sigma}_n + \mathbf{I}\right)\right]$$
$$-2\frac{1}{p}tr\left[\left(\frac{1-c}{c}\boldsymbol{\Sigma}_n + \frac{tr(\boldsymbol{\Sigma}_n)}{p}\mathbf{I}\right)^{-1}\boldsymbol{\Sigma}_n^{-1}\right]$$

for $p/n \to c \in (0,1)$ as $n \to \infty$ where α_C^* and β_C^* are the asymptotic values for $\hat{\alpha}_C^*$ and $\hat{\beta}_C^*$ given by

$$\alpha_C^* = 1 - \frac{c/p\,(tr(\boldsymbol{\Sigma}_n))^2\,\|\boldsymbol{\Sigma}_0\|_F^2}{(\|\boldsymbol{\Sigma}_n\|_F^2 + c/p\,(tr(\boldsymbol{\Sigma}_n))^2)\|\boldsymbol{\Sigma}_0\|_F^2 - (tr(\boldsymbol{\Sigma}_n\boldsymbol{\Sigma}_0))^2}, \tag{4.12}$$

$$\beta_C^* = \frac{tr(\boldsymbol{\Sigma}_n\boldsymbol{\Sigma}_0)}{\|\boldsymbol{\Sigma}_0\|_F^2}(1-\alpha_C^*). \tag{4.13}$$

The proof of Corollary 4.1 is given in the supplementary material (see Appendix A). Corollary 4.1 possesses several interesting applications. For instance, we get

$$L(\mathbf{S}_n^{-1}) - L(\widehat{\boldsymbol{\Pi}}_{SSE}) \xrightarrow[a.s.]{} \frac{c^2}{(1-c)^2}\frac{1}{p}\|\boldsymbol{\Sigma}_n^{-1}\|_F^2$$

for $p/n \to c \in (0,1)$ as $n \to \infty$, i.e., the scaled standard estimator always overperforms the traditional estimator asymptotically.

Remark 1. In the case of other estimators the results of Corollary 4.1 cannot be applied straightforwardly, since the obtained formulas of the loss functions are complicated expressions which involved the unknown precision matrix $\boldsymbol{\Sigma}_n^{-1}$ and the concentration ratio c. In a special case, when $\boldsymbol{\Pi}_0 = \boldsymbol{\Sigma}_n^{-1}$, i.e. the shrinkage target is perfectly chosen, we get that $\widehat{\boldsymbol{\Pi}}_{OLSE}$ overperforms $\mathbf{S}_n^{-1}$, $\widehat{\boldsymbol{\Pi}}_{SSE}$, and $\widehat{\boldsymbol{\Pi}}_{EM}$. This statement follows directly from (4.4) which ensures that $\alpha^* < 1-c$ and the fact that the first summand in $L(\widehat{\boldsymbol{\Pi}}_{OLSE})$ is zero in this particular case. Moreover, from the construction of the OLSE estimator and its asymptotic properties we expect similar results also for other reasonable choices of $\boldsymbol{\Pi}_0$. The reason is that the OLSE have to provide the asymptotically smallest Frobenius loss over all linear shrinkage estimators in the form (3.1). Another interesting feature of Corollary 4.1 is that the asymptotic Frobenius loss of the OLSE estimator is a linear combination of the Frobenius loss of traditional estimator and the distance between the true precision matrix $\boldsymbol{\Sigma}_n^{-1}$ and the target matrix $\boldsymbol{\Pi}_0$.

Another important result is formulated as Corollary 4.2.

Corollary 4.2. *Let assumptions of Theorem 3.1 hold. Then $\boldsymbol{\Sigma}_n$ is proportional to $\boldsymbol{\Pi}_0$ if and only if $L(\widehat{\boldsymbol{\Pi}}_{OLSE}) = 0$ almost surely for $p/n \to c \in (0, 1)$ as $n \to \infty$. Moreover, it holds that $L(\mathbf{S}_n^{-1}) > 0$, $L(\widehat{\boldsymbol{\Pi}}_{SSE}) > 0$, $L(\widehat{\boldsymbol{\Pi}}_{EM}) > 0$, $L(\widehat{\boldsymbol{\Sigma}}_{OLSE}^{-1}) > 0$, and $L(\widehat{\boldsymbol{\Pi}}_{KS}) > 0$ almost surely for $p/n \to c \in (0, 1)$ as $n \to \infty$.*

The statement of Corollary 4.2 follows from the fact that $L(\widehat{\boldsymbol{\Pi}}_{OLSE}) = 0$ if and only if $\alpha^* = 0$ and

$$\left(\frac{\|\boldsymbol{\Sigma}_n^{-1}\|_F^2 \|\boldsymbol{\Pi}_0\|_F^2}{(\mathrm{tr}[\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\Pi}_0])^2} - 1\right) = 0.$$

These two conditions are equivalent to $(\mathrm{tr}[\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\Pi}_0])^2 = \|\boldsymbol{\Sigma}_n^{-1}\|_F^2 \|\boldsymbol{\Pi}_0\|_F^2$, i.e., $\boldsymbol{\Sigma}_n$ is proportional to $\boldsymbol{\Pi}_0$. As a result, the choice of the target matrix $\boldsymbol{\Pi}_0$ has a special role in the construction of the optimal shrinkage estimator for the precision matrix. If the target matrix is closed to the true precision matrix, then a consistent estimator for the precision matrix in terms of the normalized Frobenius norm could be constructed.

Corollary 4.1 also shows that the application of $\boldsymbol{\Pi}_0 = \mathbf{I}$ would lead to a Frobenius-norm consistent estimator with the rate $p^{-(1+\epsilon)}$, $\epsilon > 0$. This rate could be improved if the covariance matrix $\boldsymbol{\Sigma}_n$ or the precision matrix $\boldsymbol{\Sigma}_n^{-1}$ is of a special structure, like a sparse matrix. An important question is the investigation of the convergence rate in such special cases. We do not deal with this problem in the present paper and leave it for future research.

In Fig. 1 we present the asymptotic non-normalized Frobenius losses of the estimators considered in Corollary 4.1. The population covariance matrix is chosen with a fixed proportions of eigenvalues in the spectrum, namely we set the first 20% of eigenvalues equal to 1, 40% equal to 3 and the rest 40% are equal to 10. The orthogonal matrix of eigenvectors is generated independently from the Haar distribution with a fixed random generator. Thus, the covariance matrix is dense positive definite with the same spectral structure for every dimension. Further, for dimension p equal to 100 we plot the asymptotic losses found in Corollary 4.1 as a function of the concentration ratio c. Remarkable, the suggested OLSE estimator dominates all considered benchmarks uniformly with respect to c. The Frobenius losses of OLSE and inverted OLSE are bounded on the interval $c \in (0, 1)$ while the loss of other estimators rises faster than linear to infinity as $c \to 1$. Losses of the SSE and EM estimators coincide and they clearly dominate the traditional estimator. It is noted that the SSE and EM losses are smaller than the loss of the inverted OLSE for $c \le 0.5$ while the inverted OLSE estimator is ranked in the second place for $c > 0.5$. The KS estimator shows very non-linear and unpredictable behavior although it dominates the traditional as well as SSE estimator for $c > 0.6$. The Frobenius loss of the inverted OLSE estimator seems to be a constant with respect to c and it is quite close to the suggested OLSE estimator for c around 1. It means that the inverted OLSE estimator is still a good alternative in the case when the sample covariance matrix is ill-conditioned.

4.1. Choice of $\boldsymbol{\Pi}_0$

The next question is the choice of the nonrandom target matrix $\boldsymbol{\Pi}_0$ which should be positive definite with uniformly bounded normalized trace norm. Unfortunately, the answer on this question depends on the underlying data because the choice of the target matrix is equivalent to the choice of the hyperparameter for the prior distribution of $\boldsymbol{\Sigma}_n^{-1}$. This problem is well-known in Bayesian statistics. The application of different priors leads to different results. So it is very important to choose the one which works fine in most cases. The naive one is $\boldsymbol{\Pi}_0 = \mathbf{I}$ where $\mathbf{I}$ is the identity matrix. Obviously, the oracle OLSE estimator has the prior matrix as the true precision matrix $\boldsymbol{\Sigma}_n^{-1}$ and is a consistent estimator for the precision matrix as shown in Proposition 4.1. Obviously, including some new information about the true covariance matrix $\boldsymbol{\Sigma}_n$ into the prior can lead to a significant increase of performance (see [10]). In our simulation study, however, we take $\boldsymbol{\Pi}_0 = \mathbf{I}$ as a reasonable prior in the case when no additional information is available.

Consider the OLSE estimator as a matrix function $\widehat{\boldsymbol{\Pi}}_{OLSE}(\boldsymbol{\Pi}_0) : M_p \to \tilde{M}_p$, where M_p is the space of p-dimensional positive definite symmetric matrices and $\tilde{M}_p$ is the corresponding space of the p-dimensional positive definite symmetric random matrices. In the following proposition we prove that the OLSE estimator is scale invariant as a function of the prior matrix $\boldsymbol{\Pi}_0$.

Proposition 4.1. *The matrix function $\widehat{\boldsymbol{\Pi}}_{OLSE}(\boldsymbol{\Pi}_0)$ is scale invariant, i.e., for arbitrary $\sigma > 0$ $\widehat{\boldsymbol{\Pi}}_{OLSE}(\sigma\boldsymbol{\Pi}_0) = \widehat{\boldsymbol{\Pi}}_{OLSE}(\boldsymbol{\Pi}_0)$.*

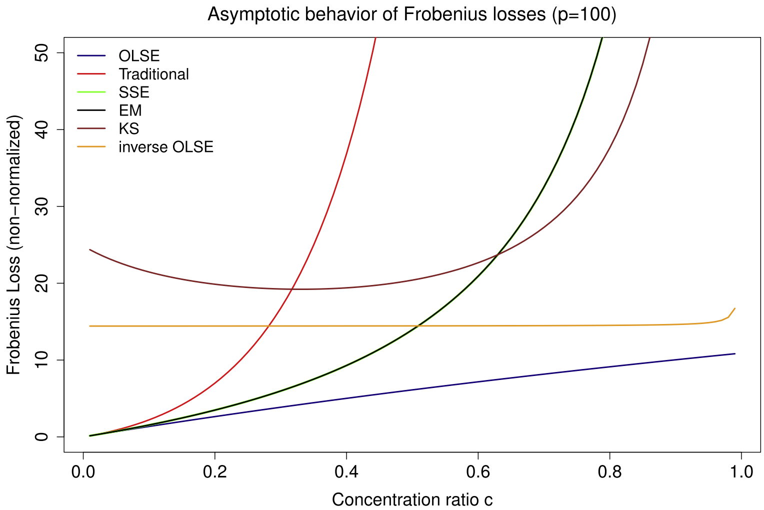

Fig. 1. Asymptotic non-normalized Frobenius loss for the suggested $\widehat{\boldsymbol{\Pi}}_{OLSE}$, the inverse of the bona fide estimator $\widehat{\boldsymbol{\Sigma}}_{OLSE}$, the scaled standard estimator (SSE), the (EM) estimator of Efron and Morris [18], the (KS) estimator of Kubokawa and Srivastava [40] and the traditional estimator plotted as a function of concentration ratio c with dimension $p = 100$.

5. Application to linear discriminant analysis

In this section we perform a Monte Carlo simulation study in order to investigate the performance of the suggested OLSE estimator for the precision matrix as well as of the existent estimators discussed in Section 4 when they are applied to linear discriminant analysis. Since many of these estimators are non-standard for the linear discriminant function, we consider its estimators when the covariance matrix is estimated by the diagonal of the sample covariance matrix $diag(\mathbf{S}_n)$ and it is replaced by the identity matrix $\mathbf{I}$ (cf. [6,54]). We observe very poor results in case of $diag(\mathbf{S}_n)$ and for this reason this approach was excluded from the simulation study.

As a performance measure, we use the normalized Euclidean distance between the linear discriminant function and the corresponding estimator given by

$$l(\mathbf{M}) = \frac{1}{p}\|\boldsymbol{\Sigma}_n^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) - \mathbf{M}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)\|, \tag{5.1}$$

where $\mathbf{M}$ is an estimator for precision matrix and $\bar{\mathbf{x}}_1$, $\bar{\mathbf{x}}_2$ are the sample counterparts of the two different population mean vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, respectively.

In each simulation run, two independent samples of equal sizes are generated from the normal distribution with same covariance matrix $\boldsymbol{\Sigma}_n$ but with different means $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$. The elements of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are simulated independently from the uniform distribution on the interval $[-1, 1]$ with a fixed `set.seed(999)`.[4] For a fixed concentration ratio $c \in \{0.5, 0.75\}$ we increase the dimension p from 15 to 300 and plot the averaged loss $p\, l(\mathbf{M})$ for each of the considered estimators. The averaging was provided via independent 1000 Monte Carlo repetitions. The structure of the population covariance matrix is the same as presented in Section 4.

In Fig. 2 we present the results of simulations under the normal distribution for $c = 1/2$ (above), whereas the results in the case of $c = 0.75$ are given below. The suggested oracle estimator is shown as a solid blue line, while the corresponding bona fide estimator is a dashed blue line. In the case $c = 1/2$ we observe a fast convergence rate of the bona fide estimator to its oracle, i.e., it converges already for the dimension as small as 50. It is remarkable that the bona fide OLSE estimator for the precision matrix $\widehat{\boldsymbol{\Pi}}_{OLSE}$ with the naive prior $\boldsymbol{\Pi}_0 = \mathbf{I}$ dominates the corresponding inverted OLSE estimator $\widehat{\boldsymbol{\Sigma}}_{OLSE}^{-1}$ with the prior $\frac{1}{p}\mathbf{I}$. The standard scaled estimator (SSE) and the EM estimator of Efron and Morris [18] lie close each to other and they are a little worse than the inverted OLSE estimator $\widehat{\boldsymbol{\Sigma}}_{OLSE}^{-1}$. Next, we rank the KS estimator of Kubokawa and Srivastava [40] followed by the estimator based on the identity matrix and the traditional one, respectively. The proposed OLSE estimator (4.3) is the best one for $c = 1/2$.

Fig. 2 shows a different picture in the case of $c = 0.75$ (below). The inverse of the OLSE estimator from Bodnar et al. [10] lies close to the one suggested in the paper for $c = 0.75$, although the later is still dominating. This finding is related to the fact that when c approaches 1, then the sample covariance matrix becomes ill-conditioned. It is noted that the bona fide OLSE estimator converges a little slower to its oracle. The KS estimator is behind the inverted OLSE estimator, whereas

[4] Using `set.seed()` function in R we are able to fix the random generator in order to provide the possibility for reproducible results.

Please cite this article in press as: T. Bodnar, et al., Direct shrinkage estimation of large dimensional precision matrix, Journal of Multivariate Analysis (2015), http://dx.doi.org/10.1016/j.jmva.2015.09.010

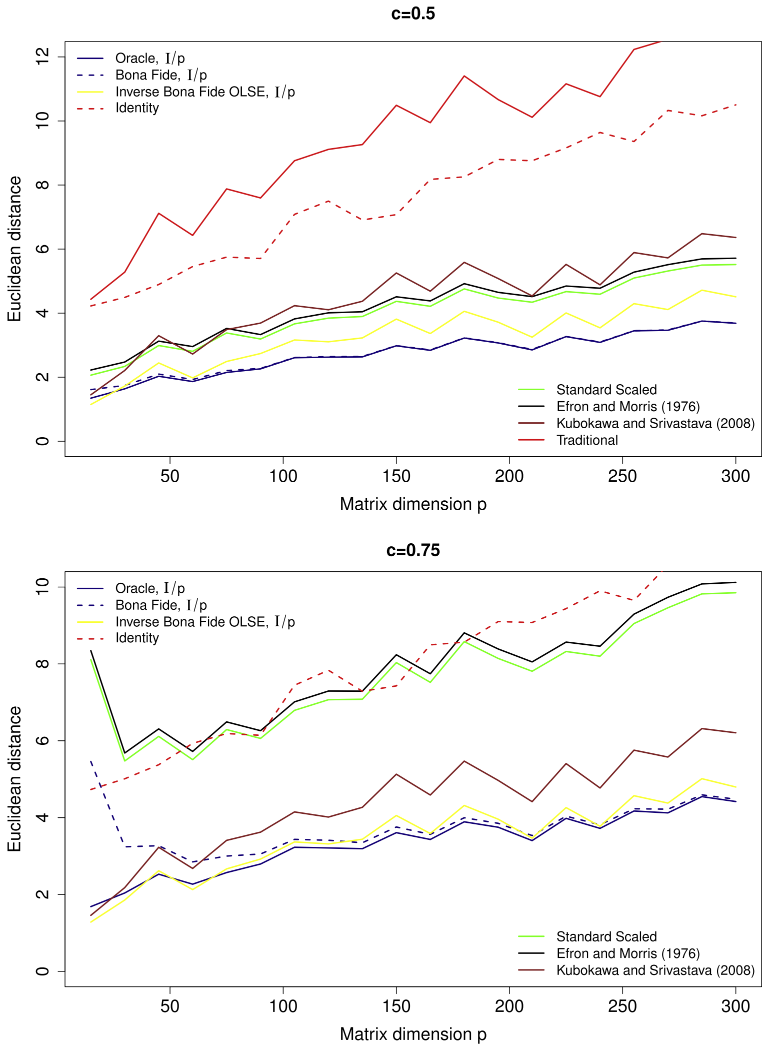

Fig. 2. Euclidean distance for the oracle and bona fide estimator $\widehat{\boldsymbol{\Pi}}_{OLSE}$ with the prior (target) matrix $\boldsymbol{\Pi}_0 = \mathbf{I}$, the inverse of the bona fide estimator $\widehat{\boldsymbol{\Sigma}}_{OLSE}$ with the prior $1/p\mathbf{I}$, the scaled standard estimator, the estimator of Efron and Morris [18], the estimator of Kubokawa and Srivastava [40], the traditional estimator, and the estimator based on the identity matrix for $p = 15k$, $k \in \{1, \ldots, 20\}$. The results are based on 1000 independent realizations.

the SSE and the EM estimators have roughly the same behavior and they are clearly worse than the other competitors. Interestingly, that the loss of estimator based on the identity matrix is close to the losses of both SSE and EM estimators, i.e. these estimators are not recommendable for the classification purposes if c close to 1. The worst behavior is detected for the traditional estimator. Its loss was around 45, that is why we have excluded it from the second part of Fig. 2.

Both the simulation results and the theoretical findings show that the OLSE estimator $\widehat{\boldsymbol{\Pi}}_{OLSE}$ is a great alternative not only to the sample estimator of the precision matrix, but also to the inverted linear shrinkage estimator proposed by Ledoit and Wolf [42] and generalized by Bodnar et al. [10] as well as to other estimators suggested in the literature. The case of $c > 1$ is even more important for the practical purposes but it seems to be more difficult to handle analytically. This can be done in an efficient way if the population covariance matrix is a multiple of identity. In general case, a good alternative would be the inverse OLSE estimator given in (4.6), but it is not optimal for the precision matrix. This point has to be treated in detail in future research.

6. Summary

In this paper, we deal with the problem of the estimation of the precision matrix for large dimensional data. Our particular interest is the case when both the dimension of the precision matrix $p \to \infty$ and the sample size $n \to \infty$ such that $p/n \to c \in (0, +\infty)$. Using the results from random matrix theory and linear shrinkage technique we develop an estimator for the precision matrix which is distribution-free (only the existence of the fourth moments is required) and possesses almost surely the smallest Frobenius loss. In particular, we prove that the Frobenius norms of the inverse and of

Please cite this article in press as: T. Bodnar, et al., Direct shrinkage estimation of large dimensional precision matrix, Journal of Multivariate Analysis (2015), http://dx.doi.org/10.1016/j.jmva.2015.09.010

the generalized inverse sample covariance matrices as well as of the optimal shrinkage intensities tend to the nonrandom quantities under high dimensional asymptotics. In order to get the optimal linear shrinkage estimator for the precision matrix we estimate the unknown quantities consistently. The performance of the suggested OLSE estimator is compared with other known estimators for the precision matrix via the simulation study.

The results of the paper can be applied and further extended in several fields of the statistics. For instance, using the suggested in the paper approach, an optimal linear shrinkage estimator for the inverse correlation matrix could be constructed. Another possible area of applications is functional data analysis, where the precision matrix is used for determining the prediction of the dependent variable and performing the variable selection of the impact points in a curve (c.f., [23,2,3]). In both cases, the estimation of the precision matrix is expected to play a special role.

Acknowledgments

The authors are grateful to the referees and the editor for their suggestions, which have improved the presentation in the paper. We would like to thank the participants of the "3rd International Workshop on Functional and Operatorial Statistics (IWFOS 2014)", the conferences "11th German Probability and Statistics Days 2014", "Random Matrix Theory: Foundations and Applications 2014" and "Statistische Woche 2014" for fruitful comments and discussions. We also thank Dr. C. Guillaume for his valuable comments on the choice of target matrix and the practical insights of the shrinkage estimation. The first author is partly supported by the German Science Foundation (DFG) via the Research Unit 1735 "Structural Inference in Statistics: Adaptation and Efficiency".

Appendix A. Supplementary material

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.jmva.2015.09.010.

References

[1] G.W. Anderson, A. Guionnet, O. Zeitouni, An Introduction to Random Matrices, Cambridge University Press, New York, USA, 2010.
[2] G. Aneiros, P. Vieu, Variable selection in infinite-dimensional problems, Statist. Probab. Lett. 94 (2014) 12–20.
[3] G. Aneiros, P. Vieu, Partial linear modelling with multi-functional covariates, Comput. Statist. 30 (2015) 647–671.
[4] J. Bai, Inferential theory for factor models of large dimensions, Econometrica 71 (2003) 135–171.
[5] Z.D. Bai, B.Q. Miao, G.M. Pan, On asymptotics of eigenvectors of large sample covariance matrix, Ann. Probab. 35 (4) (2007) 1532–1572.
[6] Z.D. Bai, H. Saranadasa, Effect of high dimension: by an example of a two sample problem, Statist. Sinica 6 (1996) 311–329.
[7] J. Bai, S. Shi, Estimating high dimensional covariance matrices and its applications, Ann. Econ. Finance 12 (2011) 199–215.
[8] Z.D. Bai, J.W. Silverstein, Spectral Analysis of Large Dimensional Random Matrices, Springer, New York, Dordrecht, Heidelberg, London, 2010.
[9] T. Bodnar, A.K. Gupta, Estimation of the precision matrix of a multivariate elliptically contoured stable distribution, Statistics 45 (2) (2011) 131–142.
[10] T. Bodnar, A.K. Gupta, N. Parolya, On the strong convergence of the optimal linear shrinkage estimator for large dimensional covariance matrix, J. Multivariate Anal. 132 (2014) 215–228.
[11] E.G. Bongiorno, E. Salinelli, A. Goia, P. Vieu, Contributions in Infinite-Dimensional Statistics and Related Topics, Società Editrice Esculapio, 2014.
[12] P. Bühlmann, S. van de Geer, Statistics for High-Dimensional Data: Methods, Theory and Applications, Springer, Heidelberg, New York, 2011.
[13] T. Cai, W. Lui, X. Luo, A constrained l_1 minimization approach to sparse precision matrix estimation, J. Amer. Statist. Assoc. 106 (2011) 594–607.
[14] T. Cai, X. Shen, High-Dimensional Data Analysis, in: Frontiers of Statistics, vol. 2, World Scientific, Singapore, 2011.
[15] T. Cai, M. Yuan, Adaptive covariance matrix estimation through block thresholding, Ann. Statist. 40 (2012) 2014–2042.
[16] T. Cai, H. Zhou, Minimax estimation of large covariance matrices under l_1 norm, Statist. Sinica 22 (2012) 1319–1378.
[17] A. Cuevas, A partial overview of the theory of statistics with functional data, J. Statist. Plann. Inference 147 (2014) 1–23.
[18] B. Efron, C. Morris, Multivariate empirical Bayes and estimation of covariance matrices, Ann. Statist. 4 (1976) 22–32.
[19] E.J. Elton, M.J. Gruber, S.J. Brown, W.N. Goetzmann, Modern Portfolio Theory and Investment Analysis, John Wiley & Sons Inc., NY, USA, 2009.
[20] J. Fan, Y. Fan, J. Lv, High dimensional covariance matrix estimation using a factor model, J. Econometrics 147 (2008) 186–197.
[21] J. Fan, Y. Liao, M. Mincheva, Large covariance estimation by thresholding principal orthogonal complements, J. R. Stat. Soc. Ser. B 75 (2013) 603–680.
[22] J. Fan, J. Zhang, K. Yu, Vast portfolio selection with gross-exposure constraints, J. Amer. Statist. Assoc. 107 (2012) 592–606.
[23] F. Ferraty, P. Hall, P. Vieu, Most-predictive design points for functional data predictors, Biometrika 97 (2006) 807–824.
[24] F. Ferraty, P. Vieu, Nonparametric Functional Data Analysis, Springer, New York, 2006.
[25] O. Friesen, M. Löwe, M. Stolz, Gaussian fluctuations for sample covariance matrices with dependent data, J. Multivariate Anal. 114 (2013) 270–287.
[26] V.L. Girko, Theory of Random Determinants, Kluwer Academic Publishers, Dordrecht, 1990.
[27] V.L. Girko, Statistical Analysis of Observations of Increasing Dimension, Kluwer Academic Publishers, Dordrecht; Boston; London, 1995.
[28] V.L. Girko, A.K. Gupta, Asymptotic behaviour of spectral function of empirical covariance matrices, Random Oper. Stoch. Equ. 2 (1994) 43–60.
[29] V.L. Girko, A.K. Gupta, Canonical equation for the resolvent of empirical covariance matrices pencil, in: A.K. Gupta, V.L. Girko (Eds.), Multidimensional Statistical Analysis and Theory of Random Matrices, VSP Publishers, Netherlands, 1996.
[30] V.L. Girko, A.K. Gupta, Multivariate elliptically contoured linear models and some aspects of the theory of random matrices, in: A.K. Gupta, V.L. Girko (Eds.), Multidimensional Statistical Analysis and Theory of Random Matrices, VSP Publishers, Netherlands, 1996.
[31] A.K. Gupta, D.K. Nagar, Matrix Variate Distributions, Chapman and Hall/CRC, Boca Raton, 2000.
[32] A.K. Gupta, S. Ofori-Nyarko, Estimation of generalized variance, precision and covariance matrices using the Pitman nearness criterion, South Afr. Statist. J. 28 (1994) 1–16.
[33] A.K. Gupta, S. Ofori-Nyarko, Improved minimax estimators of normal covariance and precision matrices, Statistics 26 (1) (1995) 19–25.
[34] A.K. Gupta, S. Ofori-Nyarko, Improved estimation of generalized variance and precision, Statistics 26 (2) (1995) 99–109.
[35] A.K. Gupta, T. Varga, T. Bodnar, Elliptically Contoured Models in Statistics and Portfolio Theory: Theory and Applications, Springer, New York, 2013.
[36] R. Horn, C. Johnson, Matrix Analysis, Cambridge University Press, Cambridge, London, NY, New Rochelle, Melbourne, Sydney, 1985.
[37] L. Horváth, P. Kokoszka, Inference for Functional Data with Applications, Springer Science & Business Media, 2012.
[38] K. Krishnamoorthy, A.K. Gupta, Improved minimax estimation of a normal precision matrix, Canad. J. Statist. 17 (1) (1989) 91–102.
[39] T. Kubokawa, A revisit to estimation of the precision matrix of the Wishart distribution, J. Statist. Res. 39 (2005) 91–114.
[40] T. Kubokawa, M. Srivastava, Estimation of the precision matrix of a singular Wishart distribution and its application in high-dimensional data, J. Multivariate Anal. 99 (2008) 1906–1928.
[41] L. Le Cam, G. Lo Yang, Asymptotics in Statistics: Some Basic Concepts, Springer, New York, 2000.

[42] O. Ledoit, M. Wolf, A well-conditioned estimator for large-dimensional covariance matrices, J. Multivariate Anal. 88 (2004) 365–411.
[43] V.A. Marčenko, L.A. Pastur, Distribution of eigenvalues for some sets of random matrices, Sb. Math. 1 (1967) 457–483.
[44] H. Markowitz, Portfolio selection, J. Finance 7 (1952) 77–91.
[45] X. Mestre, M. Lagunas, Finite sample size effect on minimum variance beamformers: Optimum diagonal loading factor for large arrays, IEEE Trans. Signal Process. 54 (2006) 69–82.
[46] J.O. Ramsay, B. Silverman, Functional Data Analysis, Springer, New York, 2005.
[47] A. Rohde, A.B. Tsybakov, Estimation of high dimensional low-rank matrices, Ann. Statist. 39 (2011) 887–930.
[48] F. Rubio, X. Mestre, Spectral convergence for a general class of random matrices, Statist. Probab. Lett. 81 (2011) 592–602.
[49] F. Rubio, X. Mestre, D. Palomar, Performance analysis and optimal selection of large minimum variance portfolios under estimation risk, IEEE J. Sel. Top. Sign. Proces. 6, 4 (2012) 337–350.
[50] A. Sarr, A.K. Gupta, Estimation of the precision matrix of multivariate Kotz type model, J. Multivariate Anal. 100 (2009) 742–752.
[51] J.W. Silverstein, Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices, J. Multivariate Anal. 55 (1995) 331–339.
[52] J.W. Silverstein, The Stieltjes transform and its role in eigenvalue behavior of large dimensional random matrices, in: Random Matrix Theory and its Applications, in: Lecture Notes Series, World Scientific, Singapore, 2009.
[53] M. Srivastava, Multivariate theory for analyzing high dimensional data, J. Japan Statist. Soc. 37 (2007) 53–86.
[54] M. Srivastava, M. Du, A test for the mean vector with fewer observations than the dimension, J. Multivariate Anal. 99 (2008) 386–402.
[55] C. Stein, Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, in: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. I, 197–206, 1956.
[56] C. Stein, Estimation of a covariance matrix, in: 39th Annual Meeting IMS Rietz Lecture, 1975.
[57] H. Tsukuma, Y. Konno, On improved estimation of normal precision matrix and discriminant coefficients, J. Multivariate Anal. 97 (2006) 1477–1500.
[58] H.L. Van Trees, Optimum Array Processing, John Wiley & Sons, NY, USA, 2002.
[59] C. Wang, G.M. Pan, T.J. Tong, L. Zhu, Shrinkage estimator of large dimensional precision matrix using random matrix theory, Statist. Sinica 25 (2015) 993–1008.
[60] L. Xue, H. Zou, Regularized rank-based estimation of high-dimensional nonparanormal graphical models, Ann. Statist. 40/5 (2012) 2541–2571.
[61] Y.Q. Yin, Limiting spectral distribution for a class of random matrices, J. Multivariate Anal. 20 (1986) 50–68.

Please cite this article in press as: T. Bodnar, et al., Direct shrinkage estimation of large dimensional precision matrix, Journal of Multivariate Analysis (2015), http://dx.doi.org/10.1016/j.jmva.2015.09.010