Higher-order nonlinear shrinkage estimator of large-dimensional precision matrix

Bodnar, Taras; Parolya, N.

DOI
10.1090/tpms/1239

Publication date
2025

Document Version
Accepted author manuscript

Published in
Theory of Probability and Mathematical Statistics

Citation (APA)
Bodnar, T., & Parolya, N. (2025). Higher-order nonlinear shrinkage estimator of large-dimensional precision matrix. *Theory of Probability and Mathematical Statistics*, *113*(113), 23-38. https://doi.org/10.1090/tpms/1239

Theor. Probability and Math. Statist.
No. 0, XXXX, Pages 000–000

HIGHER-ORDER NONLINEAR SHRINKAGE ESTIMATOR OF LARGE-DIMENSIONAL PRECISION MATRIX

TARAS BODNAR AND NESTOR PAROLYA

Abstract. This paper introduces a new type of nonlinear shrinkage estimators for the precision matrix in high-dimensional settings, where the dimension of the data-generating process exceeds the sample size. The proposed estimators incorporate the Moore-Penrose inverse and the ridge-type inverse of the sample covariance matrix, and they include linear shrinkage estimators as special cases. Recursive formulae of these higher-order nonlinear shrinkage estimators are derived using partial exponential Bell polynomials. Through simulation studies, the new methods are compared with the oracle nonlinear shrinkage estimator of the precision matrix for which no analytical expression is available.

1. Introduction

Shrinkage estimation has become a widely used approach in both theoretical and applied statistics, with numerous applications across various scientific fields, particularly in economics and finance (cf., [26], [25], [31], [16], [5],[17]). In point estimation theory, the shrinkage approach is employed to reduce the estimation error present in conventional estimators of model parameters. It was introduced in the seminal paper of Stein (see [42]) as an improved estimator for the mean vector of a normal distribution and was further extended in [23]. The core idea of this approach is to shrink the sample mean vector, which is the classical estimator of the population mean vector, towards a deterministic vector. Although this procedure introduces bias in an estimator, it can significantly reduce the estimation error. Other shrinkage-based estimators for the mean vector have been proposed in [20] and [12], among others.

Recently, Stein's idea has been successfully applied to derive improved point estimators for other multivariate quantities, such as the covariance matrix ([10], [43], [32]), the precision matrix ([44], [11], [34]), and expressions involving both the mean vector and the covariance matrix, such as optimal portfolio weights in financial applications ([25], [14]). Most of these papers introduce shrinkage estimators in high-dimensional settings, where the model dimension is of the same order as the sample size, and demonstrate that these estimators outperform conventional sample estimators.

The double asymptotic regime, also known as the high-dimensional asymptotic regime, refers to the scenario where both the model dimension and the sample size tend to infinity, with their ratio (the concentration ratio) approaching a finite number as the sample size increases (see [2], [19], [46], [7]). Unlike the classical asymptotic regime, where the dimension of the parameter space is assumed to be fixed and significantly smaller than the sample size, the high-dimensional asymptotic regime addresses the problem of

2020 *Mathematics Subject Classification.* Primary 60B20, 15A09, 62R07; Secondary 62H12, 62F12.
Key words and phrases. Moore-Penrose inverse, ridge-type inverse, nonlinear shrinkage, Bell polynomials, sample covariance matrix, random matrix theory, high-dimensional asymptotics.

dimensionality. In the derivation of the theoretical results, methods from random matrix theory are usually used.

Random matrix theory is a rapidly growing branch of probability theory. It studies the behavior of the eigenvalues of random matrices under high-dimensional asymptotic settings (see [35], [40], [2], among others). Researchers have discovered that appropriately transformed random matrices exhibit nonrandom behavior at infinity and have shown how to determine the limiting density of their eigenvalues. In particular, it was proved under very general conditions in [40] that the Stieltjes transform of the sample covariance matrix almost surely converges to a nonrandom function that satisfies a deterministic equation. This equation was first derived in [35], who demonstrated the connection between the population covariance matrix and its sample estimator at infinity. Utilizing results from random matrix theory, the asymptotic distributions of linear spectral statistics were derived in [1], [47], [37] in high-dimensional settings and they were applied to hypothesis testing in [9], [8].

The situation becomes more challenging when the model dimension exceeds the sample size. In this case, the sample covariance matrix is singular and cannot be inverted to estimate the population precision matrix. A possible solution is to use a generalized inverse of the sample covariance matrix, with the Moore-Penrose inverse and ridge-type inverse being the most popular methods (see [30], [44], [28]). However, the asymptotic properties of these generalized inverses have not been extensively studied in the statistical literature. Under the assumption of normality, the sample covariance matrix follows a singular Wishart distribution ([41]), while its Moore-Penrose inverse has a generalized inverse Wishart distribution ([13]). The upper and lower limits for the mean matrix and covariance matrix of the Moore-Penrose inverse of the sample covariance matrix were derived in [28] for the general case. Additionally, the exact mean matrix and covariance matrix are presented in [21] for the very restrictive special case where the true covariance matrix is proportional to the identity matrix. Recently, the asymptotic behavior of the weighted sample traced moments of the Moore-Penrose and ridge-type inverses of the sample covariance matrix have been derived in [15]. We apply these theoretical findings in the current paper to derive higher-order nonlinear shrinkage estimators of the precision matrix.

The rest of the paper is structured as follows. In Section 2, we present the main results of the paper. Section 2.1 provides a higher-order nonlinear shrinkage estimator based on the Moore-Penrose inverse, while Section 2.2 develops a higher-order nonlinear shrinkage estimator based on the ridge inverse. The results of the numerical study are given in Section 3, and Section 4 summarizes the findings.

2. Higher-order nonlinear shrinkage estimator

Let $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n$ be an independent and identically distributed sample from a p-dimensional distribution with $\mathbb{E}(\mathbf{y}_i) = \boldsymbol{\mu}$ and $\mathbb{C}ov(\mathbf{y}_i) = \mathbf{\Sigma}$ for $i \in 1, ..., n$ and let $\mathbf{Y}_n = (\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n)$ be the $p \times n$ observation matrix. In this paper, we derive a nonlinear shrinkage estimator of the inverse of the covariance matrix $\mathbf{\Sigma}$, known as the precision matrix in the literature. The findings are deduced in the large-dimensional case, i.e., when $p > n$.

Using the observation matrix $\mathbf{Y}_n$, the sample estimator of the population covariance matrix $\mathbf{\Sigma}$ is defined by

$$\mathbf{S}_n = \frac{1}{n}\mathbf{Y}_n\mathbf{Y}_n^\top - \bar{\mathbf{y}}_n\bar{\mathbf{y}}_n^\top \quad \text{with} \quad \bar{\mathbf{y}}_n = \frac{1}{n}\mathbf{Y}_n\mathbf{1}_n. \tag{2.1}$$

The sample estimator $\mathbf{S}_n$ presents the starting point of the new nonlinear shrinkage estimator for the precision matrix, derived in the paper.

No specific distributional assumption is imposed in the derivation of the theoretical results. Following the literature in random matrix theory (see [2], [46], among others), we assume that there exists a $p \times n$ random matrix $\mathbf{X}_n = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n)$ which consists of independent and identically distributed random variables with zero mean and unit variance such that

$$\mathbf{Y}_n \stackrel{d}{=} \boldsymbol{\mu}\mathbf{1}_n^\top + \boldsymbol{\Sigma}^{1/2}\mathbf{X}_n, \tag{2.2}$$

where $\mathbf{1}_n$ is the n-dimensional vector of ones and the symbol $\stackrel{d}{=}$ denotes the equality in distribution. It is further assumed that

(A1) $\boldsymbol{\Sigma}$ is a nonrandom positive definite matrix with $\sup_p \lambda_{max}(\boldsymbol{\Sigma}) < \infty$ and $\inf_p \lambda_{min}(\boldsymbol{\Sigma}) > 0$.

(A2) The elements of $\mathbf{X}_n$ have bounded $4+\varepsilon$ moments for some $\varepsilon > 0$.

Assumption **(A1)** presents a classical technical assumption in random matrix theory (see, e.g., [38], [32]) that ensures that the smallest and largest eigenvalues of the population covariance matrix $\boldsymbol{\Sigma}$ are uniformly bounded in p away from zero and infinity. Uniform boundedness over p is important since $\boldsymbol{\Sigma}$ implicitly depends on n through p as it is assumed throughout the paper that $p/n \to c > 1$ for $(p, n) \to \infty$. Thus, the dimension $p \equiv p(n)$ is implicitly a function of the sample size n. Assumption (A1) means in particular that the only source of the singularity of the sample covariance matrix $\mathbf{S}_n$ is the lack of data, i.e., the sample size n is smaller than the dimension of the data-generating model p. Assumption **(A2)** imposes no distributional assumptions and presents another commonly used condition in random matrix theory (cf., [2], [44], [33]).

In many applications, the precision matrix needs to be estimated from the available data. For example, the weights of optimal portfolios are functions of the precision matrix and for the practical implementation of the trading strategy, the population precision matrix needs to be replaced by its estimator (see, e.g., [16], [5], [14], [17]). Similarly, the precision matrix is present in prediction theory (cf., [18]). Finally, the precision matrix is also present in the expression of the minimum variance filter in signal processing (see [24]).

In the high-dimensional case $n < p$, the most commonly used estimators of the precision matrix are the Moore-Penrose inverse (see, e.g., [21], [6], [28], [15]) and the ridge-type inverse (see, e.g., [30] [44], [15]) of the sample covariance matrix, which are defined by

- The **Moore-Penrose inverse** of the sample covariance matrix $\mathbf{S}_n$ is the matrix $\mathbf{S}_n^+$ that fulfills the following four conditions:
  (i) $\mathbf{S}_n^+\mathbf{S}_n\mathbf{S}_n^+ = \mathbf{S}_n^+$,
  (ii) $\mathbf{S}_n\mathbf{S}_n^+\mathbf{S}_n = \mathbf{S}_n$,
  (iii) $(\mathbf{S}_n^+\mathbf{S}_n)^\top = \mathbf{S}_n^+\mathbf{S}_n$,
  (iv) $(\mathbf{S}_n\mathbf{S}_n^+)^\top = \mathbf{S}_n\mathbf{S}_n^+$.
- The **ridge-type inverse** of $\mathbf{S}_n$ is defined as the matrix $\mathbf{S}_n^-(t)$ given by
$$\mathbf{S}_n^-(t) = (\mathbf{S}_n + t\mathbf{I}_p)^{-1},$$
  where $\mathbf{I}_p$ is the p-dimensional identity matrix and $t > 0$ is a tuning parameter.

Both the estimators $\mathbf{S}_n^+$ and $\mathbf{S}_n^-(t)$ of the large-dimensional precision matrix $\boldsymbol{\Sigma}^{-1}$ suffer from large amounts of estimation error. To reduce the variability present in $\mathbf{S}_n^+$ and $\mathbf{S}_n^-(t)$, linear shrinkage estimators of the precision matrix were developed in [15] which complement the Moore-Penrose inverse and the ridge-type inverse of the sample covariance matrix. In this section, we extend the results of [15] by deriving a new type of shrinkage estimator for the precision matrix, the so-called higher-order nonlinear shrinkage estimator.

Let $\mathbf{S}_n^{\#}(t)$ denote a generalized inverse of the sample covariance matrix, either the Moore-Penrose inverse $\mathbf{S}_n^{+}$ or the ridge-type inverse $\mathbf{S}_n^{-}(t)$, that is $\mathbf{S}_n^{\#}(t) \in \{\mathbf{S}_n^{+}, \mathbf{S}_n^{-}(t)\}$. Let $\mathbf{S}_n^{\#}(t) = \mathbf{H}\mathbf{D}\mathbf{H}^\top$ be the eigenvalue decomposition of $\mathbf{S}_n^{\#}(t)$ where $\mathbf{D} = \text{diag}(d_1, ..., d_p)$ is the diagonal matrix with eigenvalues. Then for any analytical function $f(.)$, it holds that

$$f(\mathbf{S}_n^{\#}(t)) = \mathbf{H}f(\mathbf{D})\mathbf{H}^\top = \mathbf{H}\text{diag}(f(d_1), ..., f(d_p))\mathbf{H}^\top. \tag{2.3}$$

We consider the Taylor series expansion of $f(\cdot)$ at some point $d_0 > 0$ expressed as

$$f(d) = \sum_{j=0}^{\infty} \frac{f^{(j)}(d_0)}{j!}(d - d_0)^j,$$

where $f^{(j)}(d_0)$ denotes the j-th partial derivative of $f(.)$ computed at point d_0 with $f^{(0)}(d_0) = f(d_0)$. Then, the m-order approximation of $f(d)$ is given by

$$f(d) \approx f_m(d) = \sum_{j=0}^{m} \frac{f^{(j)}(d_0)}{j!}(d - d_0)^j,$$

which yields

$$\begin{aligned} f(\mathbf{S}_n^{\#}(t)) &\approx \mathbf{H}f_m(\mathbf{D})\mathbf{H}^\top = \mathbf{H}\text{diag}(f_m(d_1), ..., f_m(d_p))\mathbf{H}^\top \\ &= \sum_{j=0}^{m} \frac{f^{(j)}(d_0)}{j!}(\mathbf{S}_n^{\#}(t) - d_0\mathbf{I}_p)^j = \alpha_0\mathbf{I}_p + \sum_{j=1}^{m} \alpha_j(\mathbf{S}_n^{\#}(t))^j = f_m(\mathbf{S}_n^{\#}(t)), \end{aligned} \tag{2.4}$$

where α_j, $j = 0, 1, \ldots, m$, depend on d_0 and $f^{(j)}(d_0)$, $j = 0, 1, \ldots, m$.

The shrinkage intensities α_j, $j = 0, 1, \ldots, m$ are unknown in practice. Let $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_m)$ be the vector of unknown shrinkage intensities. Then, for given m, $\boldsymbol{\alpha}$ is chosen by minimizing the loss function expressed as [see 27, 29, 45, 44, 11, 34, 15]

$$L_n(\boldsymbol{\alpha}) = \frac{1}{p}\|f_m(\mathbf{S}_n^{\#}(t))\mathbf{\Sigma} - \mathbf{I}_p\|_F^2 \tag{2.5}$$

where $\|\cdot\|_F$ is the Frobenious norm which for a matrix $\boldsymbol{A}$ is defined by $\|\boldsymbol{A}\|_F = \sqrt{\text{tr}(\boldsymbol{A}^\top\boldsymbol{A})}$. The loss function (2.5) is slightly different than the one considered in [11]. This difference is motivated in the discussion of Section 5 in [34], who noted that the population precision matrix $\mathbf{\Sigma}^{-1}$ should be avoided in the definition of the loss function in the large-dimensional case due to potential numerical instabilities.

The optimal shrinkage intensities $\boldsymbol{\alpha}_n^*$ is obtained by minimizing the loss (2.5) that can be rewritten in the following way

$$\begin{aligned} L_n(\boldsymbol{\alpha}) &= \frac{1}{p}\left\|\left(\alpha_0\mathbf{I}_p + \sum_{j=1}^{m}\alpha_j(\mathbf{S}_n^{\#}(t))^j\right)\mathbf{\Sigma} - \mathbf{I}_p\right\|_F^2 \\ &= \frac{1}{p}\text{tr}\left[\left(\left(\alpha_0\mathbf{I}_p + \sum_{j=1}^{m}\alpha_j(\mathbf{S}_n^{\#}(t))^j\right)\mathbf{\Sigma} - \mathbf{I}_p\right)^\top\left(\left(\alpha_0\mathbf{I}_p + \sum_{j=1}^{m}\alpha_j(\mathbf{S}_n^{\#}(t))^j\right)\mathbf{\Sigma} - \mathbf{I}_p\right)\right] \\ &= \boldsymbol{\alpha}^\top\mathbf{M}(m,t)\boldsymbol{\alpha} - 2\mathbf{m}(m,t)^\top\boldsymbol{\alpha} + 1, \end{aligned}$$

with

$$\mathbf{m}(m,t) = \begin{pmatrix} \frac{1}{p}\text{tr}[\mathbf{\Sigma}] \\ \frac{1}{p}\text{tr}[\mathbf{S}_n^{\#}(t)\mathbf{\Sigma}] \\ \vdots \\ \frac{1}{p}\text{tr}[(\mathbf{S}_n^{\#}(t))^m\mathbf{\Sigma}] \end{pmatrix}$$

and

$$\mathbf{M}(m,t)=\begin{pmatrix}\frac{1}{p}\mathrm{tr}[\boldsymbol{\Sigma}^2] & \frac{1}{p}\mathrm{tr}[\mathbf{S}_n^{\#}(t)\boldsymbol{\Sigma}^2] & \dots & \frac{1}{p}\mathrm{tr}[(\mathbf{S}_n^{\#}(t))^m\boldsymbol{\Sigma}^2]\\ \frac{1}{p}\mathrm{tr}[\mathbf{S}_n^{\#}(t)\boldsymbol{\Sigma}^2] & \frac{1}{p}\mathrm{tr}[(\mathbf{S}_n^{\#}(t))^2\boldsymbol{\Sigma}^2] & \dots & \frac{1}{p}\mathrm{tr}[(\mathbf{S}_n^{\#}(t))^{m+1}\boldsymbol{\Sigma}^2]\\ \vdots & \vdots & \ddots & \vdots\\ \frac{1}{p}\mathrm{tr}[(\mathbf{S}_n^{\#}(t))^m\boldsymbol{\Sigma}^2] & \frac{1}{p}\mathrm{tr}[(\mathbf{S}_n^{\#}(t))^{m+1}\boldsymbol{\Sigma}^2] & \dots & \frac{1}{p}\mathrm{tr}[(\mathbf{S}_n^{\#}(t))^{2m}\boldsymbol{\Sigma}^2]\end{pmatrix}. \tag{2.6}$$

It is important to note that the matrix $\mathbf{M}(m,t)$ is a Hankel matrix. Therefore, to ensure the unique minimum of $L_n(\boldsymbol{\alpha})$, we must verify that it remains positive definite. This result is proved in Theorem 2.1.

Theorem 2.1. *Under Assumptions **(A1)** and **(A2)**, the matrix $\mathbf{M}(m,t)$ defined in* (2.6) *is positive definite.*

Proof of Theorem 2.1: Since the elements of $\mathbf{M}(m,t)$ are all nonnegative, the application of the trace inequality (see [39]) applied entrywise to $\mathbf{M}(m,t)$ yields

$$\lambda^2_{min}(\boldsymbol{\Sigma})\mathbf{M}_0(m,t)\le\mathbf{M}(m,t)\le\lambda^2_{max}(\boldsymbol{\Sigma})\mathbf{M}_0(m,t)\,, \tag{2.7}$$

where $\lambda_{min}(\boldsymbol{\Sigma})$ and $\lambda_{max}(\boldsymbol{\Sigma})$ denote the minimum and maximum eigenvalues of $\boldsymbol{\Sigma}$, respectively. The matrix $\mathbf{M}_0(m,t)$ is obtained from $\mathbf{M}(m,t)$ by setting $\boldsymbol{\Sigma}^2=\mathbf{I}$ under the traces $\mathrm{tr}((\mathbf{S}_n^{\#}(t))^{i+j-2}\boldsymbol{\Sigma}^2)$ for $i,j=1,\dots,m+1$.

Consequently, the positive definiteness of $\mathbf{M}(m,t)$ depends entirely on the positive definiteness of $\mathbf{M}_0(m,t)$. The latter property is straightforward to establish by recognizing that $\mathbf{M}_0(m,t)$ is a Gram matrix associated with a specific inner product:

$$\{\mathbf{M}_0(m,t)\}_{ij}=\frac{1}{p}\mathrm{tr}((\mathbf{S}_n^{\#}(t))^{i+j-2})=\frac{1}{p}\sum_{k=1}^{p}\lambda_k^{i+j-2}=\frac{1}{p}\sum_{k=1}^{p}\lambda_k^{i-1}\lambda_k^{j-1}$$

for $i,j=1,\dots,m+1$. Since the functions λ_k^i are linearly independent for all i, the Gram matrix $\mathbf{M}_0(m,t)$ is positive definite. This property, combined with inequality (2.7) and Assumption **(A1)** that the minimum and maximum eigenvalues of $\boldsymbol{\Sigma}$ are uniformly bounded away from zero and infinity, respectively, confirms that $\mathbf{M}(m,t)$ is also positive definite. □

Hence, it holds that

$$L_n(\boldsymbol{\alpha})=\left(\boldsymbol{\alpha}-\mathbf{M}(m,t)^{-1}\mathbf{m}(m,t)\right)^\top\mathbf{M}(m,t)\left(\boldsymbol{\alpha}-\mathbf{M}(m,t)^{-1}\mathbf{m}(m,t)\right)+L_{n;2}(m,t) \tag{2.8}$$

with

$$L_{n;2}(m,t)=1-\mathbf{m}(m,t)^\top\mathbf{M}(m,t)^{-1}\mathbf{m}(m,t). \tag{2.9}$$

As such, we get the following result:

Theorem 2.2. *The vector of optimal shrinkage intensities which minimizes the loss function* (2.5) *is given by*

$$\boldsymbol{\alpha}_n^*(m,t)=\mathbf{M}(m,t)^{-1}\mathbf{m}(m,t), \tag{2.10}$$

where the tuning parameter t is obtained by minimizing $L_{n;2}(m,t)$ in (2.9).

Although the closed-form expression of optimal shrinkage intensities is derived in Theorem 2.2, the formula (2.10) cannot be applied directly in practice since both $\mathbf{m}(m,t)$ and $\mathbf{M}(m,t)$ depend on the unknown population covariance matrix $\boldsymbol{\Sigma}$. To derive the practically relevant, i.e., completely data-driven formula of the optimal shrinkage intensities, we proceed in two steps: (i) First, the large-dimensional asymptotic limit of $\boldsymbol{\alpha}_n^*(m,t)$ will be deduced; (ii) Second, the limit value of $\boldsymbol{\alpha}_n^*(m,t)$ will be consistently estimated. The corresponding results are derived for the shrinkage estimator based on the Moore-Penrose inverse in Section 2.1 and on the ridge-type inverse in Section 2.2.

Finally, we note that the second summand in (2.8) has an important interpretation. The summand $L_{n;2}(m,t)$ is not only used to determine the tuning parameters m and t, but it also measures the quality of the derived higher-order nonlinear shrinkage estimator of the precision matrix. Similarly to $\boldsymbol{\alpha}_n^*(m,t)$, the term $L_{n;2}(m,t)$ depends on the unknown population covariance matrix. For the computation of the optimal values of the tuning parameter t, we also proceed in two steps: first, we find the limiting value of $L_{n;2}(m,t)$ under the large-dimensional asymptotic regime and second, estimate this limiting value consistently. It is important to note that the proposed approach is not based on cross-validation whose properties are not studied in high-dimensional settings to the best of our knowledge.

The partial exponential Bell polynomials (see [3, 4]) are used to formulate the theoretical results of the paper. They are defined by

(2.11)
$$B_{m,k}(v_1,v_2,...,v_{m-k+1}) = \sum \frac{m!}{j_1!j_2!...j_{m-k+1}!}\left(\frac{x_1}{1!}\right)^{j_1}\left(\frac{x_2}{2!}\right)^{j_2}...\left(\frac{x_{m-k+1}}{(m-k+1)!}\right)^{j_{m-k+1}},$$

where the sum is taken over all sequences $j_1,...,j_{m-k+1}$ of non-negative integers such that $\sum_{l=1}^{m-k+1} j_l = k$ and $\sum_{l=1}^{m-k+1} l j_l = m$. In practice, the Bell polynomials can be easily computed in the R-package *kStatistics*, see also [22].

2.1. **Higher-order nonlinear shrinkage estimator based on the Moore-Penrose inverse.** In the derivation of the bona fide higher-order nonlinear shrinkage estimator based on the Moore-Penrose inverse, we use Theorem 2.1 in [15] whose application leads to the following result

Theorem 2.3. *Let $\mathbf{Y}_n$ fulfill the stochastic representation* (2.2)*. Then, under Assumptions **(A1)** and **(A2)**, it holds for $l=1,2$ that*

$$\left|\frac{1}{p}\mathrm{tr}\left[(\mathbf{S}_n^+)^j\boldsymbol{\Sigma}^l\right] - s_{j,l}\right| \stackrel{a.s.}{\to} 0 \quad for \quad p/n \to c \in (1,\infty) \quad as \quad n\to\infty, \tag{2.12}$$

where

$$s_{j,l} = \sum_{k=1}^{j}\frac{(-1)^{j+k+1}k!}{j!}d_{k,l}B_{j,k}\left(v'(0),...,v^{(j-k+1)}(0)\right), \tag{2.13}$$

where $v(0)$ is the unique solution of the equation

$$\frac{1}{p}\mathrm{tr}\left[(v(0)\boldsymbol{\Sigma}+\mathbf{I}_p)^{-1}\right] = \frac{c_n-1}{c_n} \quad with \quad c_n = \frac{p}{n}, \tag{2.14}$$

$$v'(0) = -\frac{1}{\frac{1}{v(0)^2} - c_n\frac{1}{p}\mathrm{tr}\left\{\left[\boldsymbol{\Sigma}\left(v(0)\boldsymbol{\Sigma}+\mathbf{I}_p\right)^{-1}\right]^2\right\}}, \tag{2.15}$$

and $v''(0)$,...,$v^{(j)}(0)$ are computed recursively by

$$v^{(j)}(0) = -v'(0)\sum_{k=2}^{j}(-1)^k k! h_{k+1}B_{j,k}\left(v'(0),...,v^{(j-k+1)}(0)\right) \tag{2.16}$$

with

$$h_k = \frac{1}{[v(0)]^k} - c_n\frac{1}{p}\mathrm{tr}\left\{\left[\boldsymbol{\Sigma}\left(v(0)\boldsymbol{\Sigma}+\mathbf{I}_p\right)^{-1}\right]^k\right\}, \quad k=1,2,..., \tag{2.17}$$

and

$$
\begin{aligned}
d_{k,1} &= \frac{1}{p}\mathrm{tr}\left\{\left[\mathbf{\Sigma}\left(v(0)\mathbf{\Sigma}+\mathbf{I}_p\right)^{-1}\right]^{k+1}\right\} = \frac{1}{c_n[v(0)]^{k+1}} - \frac{h_{k+1}}{c_n}, \quad k=1,2,...,\\
d_{0,2} &= \frac{1}{v(0)}\left(\frac{1}{p}\mathrm{tr}\left\{\mathbf{\Sigma}\right\} - \frac{1}{p}\mathrm{tr}\left\{\mathbf{\Sigma}\left(v(0)\mathbf{\Sigma}+\mathbf{I}_p\right)^{-1}\right\}\right) = \frac{1}{v(0)}\left(\frac{1}{p}\mathrm{tr}\left\{\mathbf{\Sigma}\right\} - \frac{1}{c_n v(0)}\right),\\
d_{k,2} &= \frac{1}{p}\mathrm{tr}\left\{\left[\mathbf{\Sigma}\left(v(0)\mathbf{\Sigma}+\mathbf{I}_p\right)^{-1}\right]^{k+1}\mathbf{\Sigma}\right\}\\
&= \frac{1}{v(0)}\left(\frac{1}{p}\mathrm{tr}\left\{\left[\mathbf{\Sigma}\left(v(0)\mathbf{\Sigma}+\mathbf{I}_p\right)^{-1}\right]^{k}\mathbf{\Sigma}\right\} - \frac{1}{p}\mathrm{tr}\left\{\left[\mathbf{\Sigma}\left(v(0)\mathbf{\Sigma}+\mathbf{I}_p\right)^{-1}\right]^{k+1}\right\}\right)\\
&= \frac{1}{v(0)}(d_{k-1,2}-d_{k,1}), \quad k=1,2,....
\end{aligned}
$$

It is interesting to note that $h_1 = 0$ by (2.17). In the case $\mathbf{S}^{\#}(t) = \mathbf{S}_n^{+}$, we have that $\boldsymbol{\alpha}_n^*(m,t)$, $L_{n;2}(m,t)$, $\mathbf{M}(m,t)$, and $\mathbf{m}(m,t)$ do not depend on t. To indicate this observation and that the Moore-Penrose inverse is used in the construction of the higher-order nonlinear shrinkage estimator, we use the notations without t and indexed with the sign $+$. As a direct corollary of Theorem 2.3, the limiting behavior of $\boldsymbol{\alpha}_n^{+}(m)$ and $L_{n;2}^{+}(m)$ is deduced, and it is presented in Theorem 2.4

Theorem 2.4. *Let* $\mathbf{Y}_n$ *fulfill the stochastic representation* (2.2). *Then, under Assumptions* ***(A1)*** *and* ***(A2)****, it holds that*

$$\left\|\boldsymbol{\alpha}_n^{+}(m) - \boldsymbol{\alpha}^{+}(m)\right\| \stackrel{a.s.}{\to} 0 \quad \textit{with} \quad \boldsymbol{\alpha}^{+}(m) = \mathbf{M}^{+}(m)^{-1}\mathbf{m}^{+}(m) \tag{2.18}$$

and

$$\left\|L_{n;2}^{+}(m) - L_{2}^{+}(m)\right\| \stackrel{a.s.}{\to} 0 \quad \textit{with} \quad L_{2}^{+}(m) = 1 - \mathbf{m}^{+}(m)^{\top}\mathbf{M}^{+}(m)^{-1}\mathbf{m}^{+}(m) \tag{2.19}$$

for $p/n \to c \in (1,\infty)$ *as* $n\to\infty$*, where* $\|\mathbf{a}\|$ *denotes the Euclidean norm of the vector* $\mathbf{a}$ *and*

$$\mathbf{m}^{+}(m) = \begin{pmatrix} \frac{1}{p}tr[\mathbf{\Sigma}] \\ s_{1,1} \\ \vdots \\ s_{m,1} \end{pmatrix} \quad \textit{and} \quad \mathbf{M}^{+}(m) = \begin{pmatrix} \frac{1}{p}tr[\mathbf{\Sigma}^2] & s_{1,2} & \cdots & s_{m,2} \\ s_{1,2} & s_{2,2} & \cdots & s_{m+1,2} \\ \vdots & \vdots & \ddots & \vdots \\ s_{m,2} & s_{m+1,2} & \cdots & s_{2m,2} \end{pmatrix}. \tag{2.20}$$

For the specification of the bona fide nonlinear shrinkage estimator, one needs to estimate the elements of $\mathbf{m}^{+}(m)$ and $\mathbf{M}^{+}(m)$ consistently. Since the latter objects still depend on n through p, under consistency here we will always understand the concept of deterministic equivalents, i.e., the Euclidean distance between the random sequence and the nonrandom one converges to zero almost surely. Having this in mind, the consistent estimators of $\frac{1}{p}\mathrm{tr}[\mathbf{\Sigma}]$ and $\frac{1}{p}\mathrm{tr}[\mathbf{\Sigma}^2]$ are derived in [15] and they are given by

$$\hat{q}_1 = \frac{1}{p}\mathrm{tr}\left[\mathbf{S}_n\right] \quad \text{and} \quad \hat{q}_2 = \frac{1}{p}\mathrm{tr}\left[\mathbf{S}_n^2\right] - c_n\left[\frac{1}{p}\mathrm{tr}\left[\mathbf{S}_n\right]\right]^2, \tag{2.21}$$

respectively. Furthermore, consistent estimators of $v^{(j)}(0)$ for $j = 0,1,...$ are expressed as (see [15])

$$\hat{v}^{(j)}(0) = (-1)^j j! c_n \frac{1}{p}\mathrm{tr}\left[(\mathbf{S}_n^{+})^{j+1}\right]. \tag{2.22}$$

Finally, the application of (2.16) and (2.17) leads to the recursive computation of consistent estimators of h_{j+1} given by

(2.23)
$$\hat{h}_{j+1} = \frac{\hat{v}^{(j)}(0) + \hat{v}'(0)\sum_{k=2}^{j-1}(-1)^k k!\hat{h}_{k+1}B_{j,k}\left(\hat{v}'(0),...,\hat{v}^{(j-k+1)}(0)\right)}{[\hat{v}'(0)]^{j+1}(-1)^{j+1}j!} \quad \text{for} \quad j=2,3,\ldots,m$$

with

$$\hat{h}_2 = -\frac{1}{\hat{v}'(0)}. \tag{2.24}$$

Using the above results, we get consistent estimators of $\boldsymbol{\alpha}^+(m)$ and $L_2^+(m)$ summarized in Theorem 2.5.

Theorem 2.5. *Let $\mathbf{Y}_n$ fulfill the stochastic representation* (2.2)*. Then, under Assumptions **(A1)** and **(A2)**, it holds that*

$$\left\|\hat{\boldsymbol{\alpha}}^+(m) - \boldsymbol{\alpha}^+(m)\right\| \stackrel{a.s.}{\to} 0 \quad \text{with} \quad \hat{\boldsymbol{\alpha}}^+(m) = \widehat{\mathbf{M}}^+(m)^{-1}\hat{\mathbf{m}}^+(m) \tag{2.25}$$

and

$$\left\|\hat{L}_2^+(m) - L_2^+(m)\right\| \stackrel{a.s.}{\to} 0 \quad \text{with} \quad \hat{L}_2^+(m) = 1 - \hat{\mathbf{m}}^+(m)^\top\widehat{\mathbf{M}}^+(m)^{-1}\hat{\mathbf{m}}^+(m) \tag{2.26}$$

for $p/n \to c \in (1,\infty)$ as $n \to \infty$ with

$$\hat{\mathbf{m}}^+(m) = \begin{pmatrix} \hat{q}_1 \\ \hat{s}_{1,1} \\ \vdots \\ \hat{s}_{m,1} \end{pmatrix} \quad \text{and} \quad \widehat{\mathbf{M}}^+(m) = \begin{pmatrix} \hat{q}_2 & \hat{s}_{1,2} & \dots & \hat{s}_{m,2} \\ \hat{s}_{1,2} & \hat{s}_{2,2} & \dots & \hat{s}_{m+1,2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{s}_{m,2} & \hat{s}_{m+1,2} & \dots & \hat{s}_{2m,2} \end{pmatrix}, \tag{2.27}$$

where

$$\hat{s}_{j,l} = \sum_{k=1}^{j}\frac{(-1)^{j+k+1}k!}{j!}\hat{d}_{k,l}B_{j,k}\left(\hat{v}'(0),...,\hat{v}^{(j-k+1)}(0)\right), \tag{2.28}$$

with

$$\begin{aligned} \hat{d}_{k,1} &= \frac{1}{c_n[\hat{v}(0)]^{k+1}} - \frac{\hat{h}_{k+1}}{c_n}, \quad k=1,2,..., \\ d_{0,2} &= \frac{1}{\hat{v}(0)}\left(\hat{q}_1 - \frac{1}{c_n\hat{v}(0)}\right), \ \hat{d}_{k,2} = \frac{1}{\hat{v}(0)}(\hat{d}_{k-1,2} - \hat{d}_{k,1}), \quad k=1,2,.... \end{aligned} \tag{2.29}$$

where $\hat{q}_1$, $\hat{q}_2$, $\hat{v}(0)$, $\hat{v}'(0),\ldots,\hat{v}^{(m)}(0)$, $\hat{h}_2$, and $\hat{h}_{m+1}$ are given in (2.21)*,* (2.22)*,* (2.23)*, and* (2.24)*.*

The findings of Theorem 2.5 lead to the bona fide higher-order nonlinear shrinkage estimator of the precision matrix given by

$$\mathbf{S}^+_{n;HOS} = \hat{\alpha}_0^+\mathbf{I}_p + \sum_{j=1}^{m}\hat{\alpha}_j^+(\mathbf{S}_n^+)^j, \tag{2.30}$$

where $\hat{\boldsymbol{\alpha}}^+(m) = (\hat{\alpha}_0^+, \hat{\alpha}_1^+, \ldots, \hat{\alpha}_m^+)^\top$ as in (2.25). Finally, we note that the matrices $\mathbf{M}^+(m)$ and $\widehat{\mathbf{M}}^+(m)$ defined in (2.20) and (2.27), respectively, are positive definite, since they both almost surely converge to the matrix $\mathbf{M}(m,t)$ in (2.6) with $\mathbf{S}^{\#}(t)$ replaced by $\mathbf{S}_n^+$, which is an interior point of the convex cone of positive definite matrices.

2.2. **Higher-order nonlinear shrinkage estimator based on the ridge-type inverse.** When the ridge-type inverse is used in the construction of a nonlinear shrinkage estimator for the precision matrix, i.e., $\mathbf{S}_n^{\#}(t)=\mathbf{S}_n^{-}(t)$, the application of Theorem 2.5 in [15] yields

Theorem 2.6. *Let $\mathbf{Y}_n$ fulfill the stochastic representation (2.2). Then, under Assumptions **(A1)** and **(A2)**, for any $t>0$ and $j=0,1,2,\ldots$, $l\in\{1,2\}$, it holds that*

$$\left|\frac{1}{p}\mathrm{tr}((\mathbf{S}_n^{-}(t))^{j+1}\mathbf{\Sigma}^l)-\tilde{s}_{j+1,l}(t)\right|\overset{a.s.}{\to}0 \quad \text{for} \quad p/n\to c\in(1,\infty) \quad \text{as} \quad n\to\infty \tag{2.31}$$

where

$$\begin{aligned}\tilde{s}_{j+1,l}(t) &= \sum_{l=1}^{j}t^{-(j-l)-1}\sum_{k=1}^{l}\frac{(-1)^{l+k}k!}{l!}d_{k,l}(t)B_{l,k}\left(v'(t),v''(t),...,v^{(l-k+1)}(t)\right)\\ &+ t^{-j-1}d_{0,l}(t)\end{aligned}\tag{2.32}$$

where $v(t)$ is the unique solution of the equation

$$\frac{1}{p}\mathrm{tr}\left[(v(t)\mathbf{\Sigma}+\mathbf{I}_p)^{-1}\right]=\frac{c_n-1+tv(t)}{c_n}, \tag{2.33}$$

$$v'(t)=-\frac{1}{v(t)^{-2}-c_n\frac{1}{p}\mathrm{tr}\left\{\left[\mathbf{\Sigma}\left(v(t)\mathbf{\Sigma}+\mathbf{I}_p\right)^{-1}\right]^2\right\}}, \tag{2.34}$$

and $v''(t)$,...,$v^{(j)}(t)$ are computed recursively by

$$v^{(j)}(t)=-v'(t)\sum_{k=2}^{j}(-1)^k k! h_{k+1}(t)B_{j,k}\left(v'(t),...,v^{(j-k+1)}(t)\right), \tag{2.35}$$

with

$$h_k(t)=[v(t)]^{-k}-c_n\frac{1}{p}\mathrm{tr}\left\{\left[\mathbf{\Sigma}\left(v(t)\mathbf{\Sigma}+\mathbf{I}_p\right)^{-1}\right]^k\right\}, \quad k=1,2,..., \tag{2.36}$$

and

$$\begin{aligned}d_{0,1}(t) &= \frac{1}{p}\mathrm{tr}\left\{\mathbf{\Sigma}\left(v(t)\mathbf{\Sigma}+\mathbf{I}_p\right)^{-1}\right\}=\frac{1}{c_nv(t)}-\frac{t}{c_n},\\ d_{k,1}(t) &= \frac{1}{p}\mathrm{tr}\left\{\left[\mathbf{\Sigma}\left(v(t)\mathbf{\Sigma}+\mathbf{I}_p\right)^{-1}\right]^{k+1}\right\}=\frac{1}{c_n[v(t)]^{k+1}}-\frac{1}{c_nh_{k+1}(t)},\ k=1,2,...,\\ d_{0,2}(t) &= \frac{1}{v(t)}\left(\frac{1}{p}\mathrm{tr}\left\{\mathbf{\Sigma}\right\}-d_{0,1}(t)\right),\\ d_{k,2}(t) &= \frac{1}{p}\mathrm{tr}\left\{\left[\mathbf{\Sigma}\left(v(t)\mathbf{\Sigma}+\mathbf{I}_p\right)^{-1}\right]^{k+1}\mathbf{\Sigma}\right\}\\ &= \frac{1}{v(t)}\left(\frac{1}{p}\mathrm{tr}\left\{\left[\mathbf{\Sigma}\left(v(t)\mathbf{\Sigma}+\mathbf{I}_p\right)^{-1}\right]^{k}\mathbf{\Sigma}\right\}-\frac{1}{p}\mathrm{tr}\left\{\left[\mathbf{\Sigma}\left(v(t)\mathbf{\Sigma}+\mathbf{I}_p\right)^{-1}\right]^{k+1}\right\}\right)\\ &= \frac{1}{v(t)}(d_{k-1,2}(t)-d_{k,1}(t)), \quad k=1,2,....\end{aligned}$$

When $\mathbf{S}^{\#}(t)=\mathbf{S}_n^{-}(t)$, we use the notations $\boldsymbol{\alpha}_n^{-}(m,t)$, $L_{n;2}^{-}(m,t)$, $\mathbf{M}^{-}(m,t)$, and $\mathbf{m}^{-}(m,t)$. Form Theorem 2.6, we the limiting behavior of $\boldsymbol{\alpha}_n^{-}(m,t)$ and $L_{n;2}^{-}(m,t)$ is derived in Theorem 2.7

Theorem 2.7. *Let $\mathbf{Y}_n$ fulfill the stochastic representation* (2.2)*. Then, under Assumptions **(A1)** and **(A2)**, it holds that*

$$\left\|\boldsymbol{\alpha}_n^-(m,t)-\boldsymbol{\alpha}^-(m,t)\right\| \stackrel{a.s.}{\to} 0 \quad \text{with} \quad \boldsymbol{\alpha}^-(m,t)=\mathbf{M}^-(m,t)^{-1}\mathbf{m}^-(m,t) \tag{2.37}$$

and

$$\left\|L_{n;2}^-(m,t)-L_2^-(m,t)\right\| \stackrel{a.s.}{\to} 0 \quad \text{with} \quad L_2^-(m,t)=1-\mathbf{m}^-(m,t)^\top\mathbf{M}^-(m,t)^{-1}\mathbf{m}^-(m,t) \tag{2.38}$$

for $p/n \to c \in (1,\infty)$ as $n\to\infty$, where

$$\mathbf{m}^-(m,t)=\begin{pmatrix}\frac{1}{p}tr[\boldsymbol{\Sigma}]\\ \tilde{s}_{1,1}(t)\\ \vdots\\ \tilde{s}_{m,1}(t)\end{pmatrix} \quad \text{and} \quad \mathbf{M}^-(m,t)=\begin{pmatrix}\frac{1}{p}tr[\boldsymbol{\Sigma}^2] & \tilde{s}_{1,2}(t) & \dots & \tilde{s}_{m,2}(t)\\ \tilde{s}_{1,2}(t) & \tilde{s}_{2,2}(t) & \dots & \tilde{s}_{m+1,2}(t)\\ \vdots & \vdots & \ddots & \vdots\\ \tilde{s}_{m,2}(t) & \tilde{s}_{m+1,2}(t) & \dots & \tilde{s}_{2m,2}(t)\end{pmatrix}. \tag{2.39}$$

To estimate the elements of $\mathbf{m}^-(m,t)$ and $\mathbf{M}^-(m,t)$ consistently, we use (2.21) and the results of [15] who proved that consistent estimators of $v^{(j)}(t)$ for $j=0,1,...$ and $t>0$ are given by

$$\hat{v}^{(j)}(t)=(-1)^j j! c_n\left(\frac{1}{p}\mathrm{tr}\left[(\mathbf{S}_n^-(t))^{j+1}\right]-t^{-(j+1)}\frac{c_n-1}{c_n}\right). \tag{2.40}$$

Finally, the application of (2.35) and (2.36) leads to the recursive computation of consistent estimators of $h_{j+1}(t)$ expressed as

$$\hat{h}_{j+1}(t)=\frac{\hat{v}^{(j)}(t)+\hat{v}'(t)\sum_{k=2}^{j-1}(-1)^k k!\hat{h}_{k+1}(t)B_{j,k}\left(\hat{v}'(t),...,\hat{v}^{(j-k+1)}(t)\right)}{[\hat{v}'(t)]^{j+1}(-1)^{j+1}j!} \quad \text{for} \quad j=2,3,\dots,m \tag{2.41}$$

with

$$\hat{h}_2(t)=-\frac{1}{\hat{v}'(t)}. \tag{2.42}$$

Hence, consistent estimators of $\boldsymbol{\alpha}^-(m,t)$ and $L_2^-(m,t)$ are obtained and presented in Theorem 2.8.

Theorem 2.8. *Let $\mathbf{Y}_n$ fulfill the stochastic representation* (2.2)*. Then, under Assumptions **(A1)** and **(A2)**, it holds that*

$$\left\|\hat{\boldsymbol{\alpha}}^-(m,t)-\boldsymbol{\alpha}^-(m,t)\right\| \stackrel{a.s.}{\to} 0 \quad \text{with} \quad \hat{\boldsymbol{\alpha}}^-(m,t)=\widehat{\mathbf{M}}^-(m,t)^{-1}\hat{\mathbf{m}}^-(m,t) \tag{2.43}$$

and

$$\left\|\hat{L}_2^-(m,t)-L_2^-(m,t)\right\| \stackrel{a.s.}{\to} 0 \quad \text{with} \quad \hat{L}_2^-(m,t)=1-\hat{\mathbf{m}}^-(m,t)^\top\widehat{\mathbf{M}}^-(m,t)^{-1}\hat{\mathbf{m}}^-(m,t) \tag{2.44}$$

for $p/n \to c \in (1,\infty)$ as $n\to\infty$ with

$$\hat{\mathbf{m}}^-(m,t)=\begin{pmatrix}\hat{q}_1\\ \hat{\tilde{s}}_{1,1}(t)\\ \vdots\\ \hat{\tilde{s}}_{m,1}(t)\end{pmatrix} \quad \text{and} \quad \widehat{\mathbf{M}}^-(m,t)=\begin{pmatrix}\hat{q}_2 & \hat{\tilde{s}}_{1,2}(t) & \dots & \hat{\tilde{s}}_{m,2}(t)\\ \hat{\tilde{s}}_{1,2}(t) & \hat{\tilde{s}}_2(t) & \dots & \hat{\tilde{s}}_{m+1,2}(t)\\ \vdots & \vdots & \ddots & \vdots\\ \hat{\tilde{s}}_{m,2}(t) & \hat{\tilde{s}}_{m+1,2}(t) & \dots & \hat{\tilde{s}}_{2m,2}(t)\end{pmatrix}. \tag{2.45}$$

where

$$\begin{aligned}\hat{\hat{s}}_{j+1,l}(t) &= \sum_{l=1}^{j} t^{-(j-l)-1} \sum_{k=1}^{l} \frac{(-1)^{l+k}k!}{l!} \hat{d}_{k,j}(t) B_{l,k}\left(\hat{v}'(t), \hat{v}''(t), ..., \hat{v}^{(l-k+1)}(t)\right) \\ &+ t^{-j-1}\hat{d}_{0,l}(t) \qquad (2.46)\end{aligned}$$

with

$$\hat{d}_{0,1}(t) = \frac{1}{c_n \hat{v}(t)} - \frac{t}{c_n}, \; \hat{d}_{k,1}(t) = \frac{1}{c_n[\hat{v}(t)]^{k+1}} - \frac{\hat{h}_{k+1}(t)}{c_n}, \quad k = 1, 2, ..., \tag{2.47}$$

$$\hat{d}_{0,2}(t) = \frac{1}{\hat{v}(t)}\left(\hat{q}_1 - \hat{d}_{0,1}(t)\right), \; \hat{d}_{k,2}(t) = \frac{1}{\hat{v}(t)}(\hat{d}_{k-1,2}(t) - \hat{d}_{k,1}(t)), \quad k = 1, 2, ...,$$

where $\hat{q}_1$, $\hat{q}_2$, $\hat{v}(t)$, $\hat{v}'(t), \ldots, \hat{v}^{(m)}(t)$, $\hat{h}_2(t)$, *and* $\hat{h}_{m+1}(t)$ *are given in* (2.21), (2.40), (2.41), *and* (2.42).

The findings of Theorem 2.8 yield the bona fide higher-order nonlinear shrinkage estimator of the precision matrix expressed as

$$\mathbf{S}^{-}_{n;HOS}(t) = \hat{\alpha}^{-}_0(t)\mathbf{I}_p + \sum_{j=1}^{m} \hat{\alpha}^{-}_j(t)(\mathbf{S}^{-}_n(t))^j, \tag{2.48}$$

where $\hat{\boldsymbol{\alpha}}^{-}(m,t) = (\hat{\alpha}^{-}_0(t), \hat{\alpha}^{-}_1(t), \ldots, \hat{\alpha}^{-}_m(t))^\top$ given in (2.43) and the optimal value of t is found by minimizing $\hat{L}^{-}_2(m,t)$ in (2.44). Furthermore, both the matrices $\mathbf{M}^{-}(m,t)$ and $\widehat{\mathbf{M}}^{-}(m,t)$ in (2.39) and (2.45) are positive definite due to the facts that they both almost surely converge to the matrix $\mathbf{M}(m,t)$ in (2.6) with $\mathbf{S}^{\#}(t)$ replaced by $\mathbf{S}^{-}_n(t)$ and $\mathbf{M}(m,t)$ is an interior point of the convex cone of positive definite matrices.

3. Finite sample performance

In this section, we compare the finite sample performance of the proposed higher-order nonlinear shrinkage estimators of the precision matrix computed for $m \in \{1, 2, 3, 4, 5\}$ with the state-of-the-art benchmark. The comparison is performed for $n \in \{150, 250\}$ and $c \in [1.5, 5]$. Two scenarios of the data-generating model are considered:

(i) **Normal distribution**: Elements of $\mathbf{X}_n$ are generated from the standard normal distribution.
(ii) **Scaled t-Distribution**: Elements of $\mathbf{X}_n$ are drawn from the scaled standard t-distribution with 5 degrees of freedom and a scale factor of $\sqrt{3/5}$ to ensure that the variances of the elements of $\mathbf{X}_n$ are all one.

The mean vector $\boldsymbol{\mu}$ is set to zero. The eigenvectors of the population covariance matrix $\boldsymbol{\Sigma}$ are simulated from the Haar distribution (see, e.g., [36]), and its eigenvalues are chosen as follows: 20% of the eigenvalues are equal to one, 40% are equal to three, and the remaining 40% are equal to ten.

As a performance measure, we use the Percentage Relative Improvement in Average Loss (PRIAL) defined by

$$\text{PRIAL}(\widehat{\boldsymbol{\Pi}}) = \left(1 - \frac{\mathbb{E}||\widehat{\boldsymbol{\Pi}}\boldsymbol{\Sigma} - \mathbf{I}_p||_F^2}{\mathbb{E}||\mathbf{S}^{+}_n\boldsymbol{\Sigma} - \mathbf{I}_p||_F^2}\right) \cdot 100\% \,, \tag{3.1}$$

where $\widehat{\boldsymbol{\Pi}}$ is an estimator of the precision matrix $\boldsymbol{\Sigma}^{-1}$. By definition, the PRIAL provides the percentage improvement of each strategy in comparison to the one based on the Moore-Penrose inverse. Larger values of the PRIAL indicate better performance. Moreover, it holds that PRIAL($\boldsymbol{\Sigma}^{-1}$)= 100% and PRIAL($\mathbf{S}^{+}_n$)= 0%.

As a benchmark, we use the inverse of *oracle nonlinear shrinkage estimator of the covariance matrix* which is derived for the loss function considered in [33] and it is given by

$$\mathbf{S}_{oNLSh} = \mathbf{U}\text{diag}(\tilde{d}_1^{or}, ..., \tilde{d}_p^{or})\mathbf{U}^\top, \quad \tilde{d}_i^{or} = \frac{\mathbf{u}_i'\mathbf{\Sigma}^2\mathbf{u}_i}{\mathbf{u}_i'\mathbf{\Sigma}\mathbf{u}_i} \tag{3.2}$$

where $\mathbf{U} = (\mathbf{u}_1, ...\mathbf{u}_p)$ is the matrix with the eigenvectors of $\mathbf{S}_n$. Its inverse is expressed as

$$\widehat{\mathbf{\Pi}}_{oNLSh} = \mathbf{U}\text{diag}(1/\tilde{d}_1^{or}, ..., 1/\tilde{d}_p^{or})\mathbf{U}^\top, \tag{3.3}$$

with $\tilde{d}_1^{or}, ..., \tilde{d}_p^{or}$ given in (3.2). Theoretically, this unattainable benchmark (since it depends on $\mathbf{\Sigma}$) represents the best possible performance in the given setting, since it is finite-sample optimal. This raises the question: how well can our purely data-driven higher-order nonlinear shrinkage estimators approximate it? The answer to this question is provided in Figures 1 and 2, where the Moore-Penrose inverse was used in the computation of higher-order shrinkage estimators.

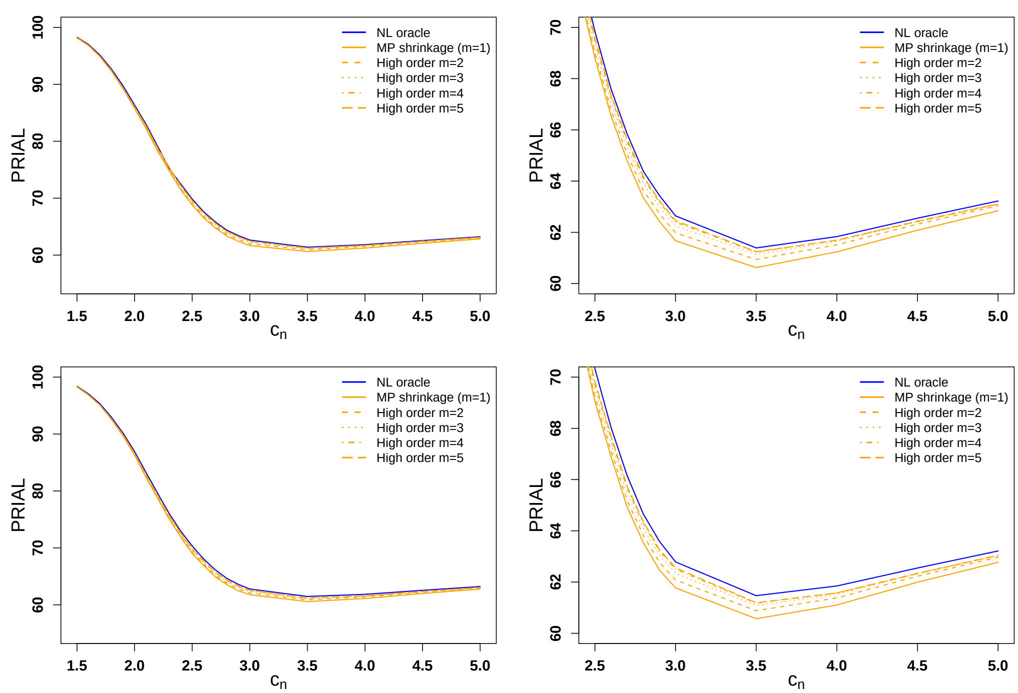

FIGURE 1. PRIAL for $c_n \in [1.5, 5]$, $n = 150$ when the elements of $\mathbf{X}_n$ are drawn from the normal distribution (first row) and scale t-distribution (second row). Second column shows the corresponding zoomed figures.

Figure 1 presents the results for the sample size $n = 150$. The first row corresponds to the normal distribution, while the second row represents the scaled t-distribution with 5 degrees of freedom. All proposed higher-order shrinkage estimators closely approximate the oracle nonlinear shrinkage. We consider polynomial orders up to $m = 5$, as any further increase provides only marginal improvement and only increases the computational time. The leftmost figures show that even with $m = 1$, the approximation to the oracle is quite good, improving rapidly as m increases. The results are similar for both the normal distribution and the t-distribution, with the latter performing slightly worse. However, this difference is barely noticeable, even in the zoomed-in versions of the figures.

Figure 2, which corresponds to $n = 250$, exhibits the same patterns, demonstrating the robustness of the proposed higher-order shrinkage estimators. Based on this simulation, $m = 4$ appears sufficient to achieve reasonable precision and fast computational time.

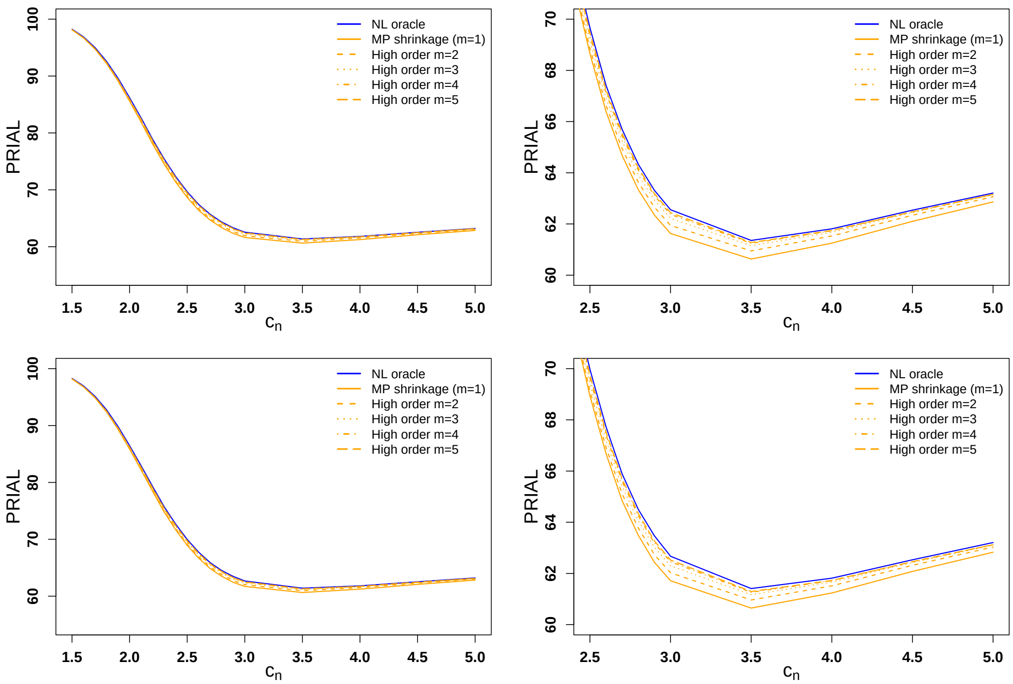

FIGURE 2. PRIAL for $c_n \in [1.5, 5]$, $n = 250$ when the elements of $\mathbf{X}_n$ are drawn from the normal distribution (first row) and scale t-distribution (second row). Second column shows the corresponding zoomed figures.

4. Summary

Estimating the precision matrix is a challenging task in multivariate statistics, especially in high-dimensional settings where the dimension of the data-generating model exceeds the sample size. In such cases, the sample covariance matrix becomes singular, and its inverse does not exist. To address this, generalized inverses like the Moore-Penrose inverse or the ridge-type inverse are typically employed. However, the impact of these generalized inverses is not well studied in statistics and probability theory, with only a few papers examining their properties under very restrictive assumptions.

The recent paper by [15] sheds light on the asymptotic properties of the Moore-Penrose inverse and the ridge-type inverse of the sample covariance matrix. Additionally, two linear shrinkage estimators of the precision matrix are suggested in [15] by utilizing the asymptotic properties of these generalized inverses.

In this paper, we extend the approach of [15] by proposing higher-order nonlinear shrinkage estimators of the precision matrix. These new approaches leverage the properties of the Moore-Penrose inverse and the ridge-type inverse of the sample covariance matrix, including the two linear shrinkage estimators of [15] as special cases. Furthermore, they complement the existing nonlinear shrinkage estimators of [32] and [34] by providing analytical expressions for new nonlinear shrinkage estimators that can be easily implemented in practice. Finally, the finite-sample performance of the proposed approaches is evaluated through a simulation study and compared to the finite-sample oracle nonlinear shrinkage estimator.

Acknowledgement

The authors would like to thank Professor Yu. S. Mishura and an anonymous Reviewer for their constructive comments that improved the quality of this paper.

Bibliography

1. Z.D. Bai and J. W. Silverstein, *CLT for linear spectral statistics of large-dimensional sample covariance matrix*, The Annals of Probability **32** (2004), no. 1A, 553–605.
2. Zhidong Bai and Jack W Silverstein, *Spectral analysis of large dimensional random matrices*, Springer, 2010.
3. Eric Temple Bell, *Partition polynomials*, Annals of Mathematics **29** (1927-1928), no. 1/4, 38–46.
4. ______, *Exponential polynomials*, Annals of Mathematics **35** (1934), no. 2, 258–277.
5. Olha Bodnar, Taras Bodnar, and Nestor Parolya, *Recent advances in shrinkage-based high-dimensional inference*, Journal of Multivariate Analysis (2021), 104826.
6. Taras Bodnar, Holger Dette, and Nestor Parolya, *Spectral analysis of the moore–penrose inverse of a large dimensional sample covariance matrix*, Journal of Multivariate Analysis **148** (2016), 160 – 172.
7. ______, *Testing for independence of large dimensional vectors*, The Annals of Statistics **47** (2019), no. 5, 2977–3008.
8. Taras Bodnar, Solomiia Dmytriv, Yarema Okhrin, Nestor Parolya, and Wolfgang Schmid, *Statistical inference for the expected utility portfolio in high dimensions*, IEEE Transactions on Signal Processing **69** (2021), 1–14.
9. Taras Bodnar, Solomiia Dmytriv, Nestor Parolya, and Wolfgang Schmid, *Tests for the weights of the global minimum variance portfolio in a high-dimensional setting*, IEEE Transactions on Signal Processing **67** (2019), 4479–4493.
10. Taras Bodnar, Arjun K Gupta, and Nestor Parolya, *On the strong convergence of the optimal linear shrinkage estimator for large dimensional covariance matrix*, Journal of Multivariate Analysis **132** (2014), 215–228.
11. ______, *Direct shrinkage estimation of large dimensional precision matrix*, Journal of Multivariate Analysis **146** (2016), 223–236.
12. Taras Bodnar, Ostap Okhrin, and Nestor Parolya, *Optimal shrinkage estimator for high-dimensional mean vector*, Journal of Multivariate Analysis **170** (2019), 63–79.
13. Taras Bodnar and Yarema Okhrin, *Properties of the singular, inverse and generalized inverse partitioned Wishart distributions*, Journal of Multivariate Analysis **99** (2008), no. 10, 2389–2405.
14. Taras Bodnar, Yarema Okhrin, and Nestor Parolya, *Optimal shrinkage-based portfolio selection in high dimensions*, Journal of Business & Economic Statistics **41** (2023), no. 1, 140–156.
15. Taras Bodnar and Nestor Parolya, *Reviving pseudo-inverses: Asymptotic properties of large dimensional Moore-Penrose and ridge-type inverses with applications*, 2024, arXiv preprint arXiv:2403.15792.

16. Taras Bodnar, Nestor Parolya, and Wolfgang Schmid, *Estimation of the global minimum variance portfolio in high dimensions*, European Journal of Operational Research **266** (2018), no. 1, 371–390.
17. Taras Bodnar, Nestor Parolya, and Erik Thorsén, *Two is better than one: regularized shrinkage of large minimum variance portfolios*, Journal of Machine Learning Research **25** (2024), no. 173, 1–32.
18. Peter J Brockwell and Richard A Davis, *Introduction to time series and forecasting*, Springer, 2002.
19. Peter Bühlmann and Sara Van De Geer, *Statistics for high-dimensional data: Methods, theory and applications*, Springer, Berlin, Heidelberg, 2011.
20. Didier Chételat and Martin T. Wells, *Improved multivariate normal mean estimation with unknown covariance when p is greater than n*, The Annals of Statistics **40** (2012), 3137–3160.
21. R Dennis Cook and Liliana Forzani, *On the mean and variance of the generalized inverse of a singular Wishart matrix*, Electronic Journal of Statistics **5** (2011), 146–158.
22. Elvira Di Nardo, Giuseppe Guarino, and Domenico Senato, *A unifying framework for k-statistics, polykays and their multivariate generalizations*, Bernoulli **14** (2008), 440–468.
23. Bradley Efron and Carl Morris, *Families of minimax estimators of the mean of a multivariate normal distribution*, Annals of Statistics **4** (1976), no. 1, 11–21.
24. Yiyong Feng and Daniel P. Palomar, *A signal processing perspective on financial engineering*, now Publishers Inc., Boston and Delft, 2016.
25. G. Frahm and C. Memmel, *Dominating estimators for minimum-variance portfolios*, Journal of Econometrics **159** (2010), 289–302.
26. Vasyl Golosnoy and Yarema Okhrin, *Multivariate shrinkage for optimal portfolio weights*, The European Journal of Finance **13** (2007), no. 5, 441–458.
27. L.R Haff, *An identity for the Wishart distribution with applications*, Journal of Multivariate Analysis **9** (1979), no. 4, 531–544.
28. Shinpei Imori and Dietrich von Rosen, *On the mean and dispersion of the Moore-Penrose generalized inverse of a Wishart matrix*, The Electronic Journal of Linear Algebra **36** (2020), 124–133.
29. K. Krishnamoorthy and A. K. Gupta, *Improved minimax estimation of a normal precision matrix*, The Canadian Journal of Statistics / La Revue Canadienne de Statistique **17** (1989), no. 1, 91–102.
30. Tatsuya Kubokawa and Muni S Srivastava, *Estimation of the precision matrix of a singular Wishart distribution and its application in high-dimensional data*, Journal of Multivariate Analysis **99** (2008), no. 9, 1906–1928.
31. Olivier Ledoit and Michael Wolf, *Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks*, The Review of Financial Studies **30** (2017), no. 12, 4349–4388.
32. Olivier Ledoit and Michael Wolf, *Analytical nonlinear shrinkage of large-dimensional covariance matrices*, The Annals of Statistics **48** (2020), no. 5, 3043 – 3065.
33. Olivier Ledoit and Michael Wolf, *Shrinkage estimation of large covariance matrices: Keep it simple, statistician?*, Journal of Multivariate Analysis **186** (2021), 104796.
34. Olivier Ledoit and Michael Wolf, *Quadratic shrinkage for large covariance matrices*, Bernoulli **28** (2022), no. 3, 1519 – 1547.
35. Vladimir A Marčenko and Leonid Andreevich Pastur, *Distribution of eigenvalues for some sets of random matrices*, Mathematics of the USSR-Sbornik **1** (1967), no. 4, 457.

36. Robb Muirhead, *Aspects of multivariate statistical theory*, New York: Wiley, 1982.
37. Jamal Najim and Jianfeng Yao, *Gaussian fluctuations for linear spectral statistics of large random covariance matrices*, Annals of Applied Probability **26** (2016), no. 3, 1837–1887.
38. Guangming Pan, *Comparison between two types of large sample covariance matrices*, Annales de l'IHP Probabilités et Statistiques **50** (2014), no. 2, 655–677.
39. Axel Ruhe, *Perturbation bounds for means of eigenvalues and invariant subspaces*, BIT Numerical Mathematics **10** (1970), 343–354.
40. Jack W Silverstein, *Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices*, Journal of Multivariate Analysis **55** (1995), no. 2, 331–339.
41. Muni S Srivastava, *Singular Wishart and multivariate beta distributions*, The Annals of Statistics **31** (2003), no. 5, 1537–1560.
42. Charles Stein, *Inadmissibility of the usual estimator for the mean of a multivariate normal distribution*, Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics (Berkeley, Calif.), University of California Press, 1956, pp. 197–206.
43. Jun Tong, Rui Hu, Jiangtao Xi, Zhitao Xiao, Qinghua Guo, and Yanguang Yu, *Linear shrinkage estimation of covariance matrices using low-complexity cross-validation*, Signal Processing **148** (2018), 223–233.
44. Cheng Wang, Guangming Pan, Tiejun Tong, and Lixing Zhu, *Shrinkage estimation of large dimensional precision matrix using random matrix theory*, Statistica Sinica **25** (2015), 993–1008.
45. Ruoyong Yang and James O. Berger, *Estimation of a covariance matrix using the reference prior*, The Annals of Statistics **22** (1994), no. 3, 1195 – 1211.
46. Jianfeng Yao, Shurong Zheng, and ZD Bai, *Sample covariance matrices and high-dimensional data analysis*, Cambridge University Press Cambridge, 2015.
47. Shurong Zheng, Zhidong Bai, and Jianfeng Yao, *Substitution principle for CLT of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing*, Annals of Statistics **43** (2015), no. 2, 546–591.

Department of Management and Engineering, Linköping University, SE 581 83 Linköping, Sweden

Email address: taras.bodnar@liu.se

Delft Institute of Applied Mathematics, Delft University of Technology, 2628CD Delft, The Netherlands

Email address: N.Parolya@tudelft.nl

Received 28/FEB/2025