

High-dimensional covariance matrix regularization using informative targets

Atiq Ur Rehman* Muhammad Farooq and Insha Ullah

Abstract. The non-invertibility of the sample covariance matrix in high-dimensional settings renders some of the classical multivariate statistical techniques inapplicable. Various regularization methods have been proposed to make the sample covariance matrix not only invertible, but also more stable. Most of these methods impose a known structure on the sample covariance matrix in the form of an informative target matrix. The importance of a suitably chosen target matrix is well recognized in the literature. However, due to its simplicity a popular choice of the target matrix remains a diagonal matrix. The intent of this article is to facilitate the use of more informative targets to regularize the sample covariance matrix. These informative targets may depend on an unknown correlation parameter, which we estimate using the maximum likelihood estimation method. The proposed method is analytically straightforward, computationally less expensive and more accurate than some of the recently proposed regularization methods when the targets are correctly specified. The usefulness of the proposed method is shown through an extensive simulation study and a real application.

Key words and phrases: Autoregressive covariance structure, Exchangeable covariance structure, Penalty parameter, MANOVA, Shrinkage estimation.

1. INTRODUCTION

High-dimensional datasets, where the number of variables, p , is greater than the sample size, n , are becoming increasingly common in many fields, including genetics (for example, see Eisen et al., 1998; Tamayo et al., 1999; Beerenwinkel et al., 2007), brain imaging, and climate data. Publicly available gene expression data in most cases contain 5000-10000 genes and less than 100 samples and both genes and samples are expected to grow overtime (Dudoit et al., 2002). For these types of high-dimensional datasets the sample covariance matrix—maximum likelihood estimator or its unbiased version—performs poorly and is not a good approximation to the true covariance matrix (even if p is comparable to n). This is because the sample covariance matrix contains estimation error and their eigenvalues tend to be overdispersed; that is, the larger eigenvalues will contain an amount of positive estimation errors (overestimated) and smaller eigenvalues will contain an amount of negative estimation errors (underestimated). In addition, the inverse covariance matrix is fundamental to classical multivariate methods such as regression, linear or quadratic discriminant analysis, Gaussian graphical models, and Mahalanobis distance. The sample covariance matrix loses its full rank and is not invertible if p exceeds n . Consequently, the non-invertibility of the sample covariance matrix precludes the use of aforementioned multivariate methods.

Previously, a number of methods have been proposed to regularize the estimate of a covariance matrix and/or approximate the inverse of a covariance matrix in a high-

*To whom correspondence should be addressed.

dimensional setting. Moore-Penrose generalized inverse (Penrose, 1955) is among the initial attempts that have been made in this line. It is based on the singular value decomposition (SVD) and takes into account only the non-zero singular values and their corresponding left and right singular vectors to approximate the inverse covariance matrix. Shrinkage estimation (Ledoit & Wolf, 2003, 2004) considers the convex combination of the sample covariance matrix and a positive definite target matrix, where the combination proportion (also referred to as shrinkage intensity) is estimated via minimizing the Frobenius-norm. Schäfer & Strimmer (2005) extend this idea and calculate the shrinkage intensity parameter using six different target matrices including the identity matrix. Note that the popular ridge regularization in ridge regression (Hoerl & Kennard, 1970) is a special case of shrinkage estimation. Another procedure for the regularization of sample covariance matrices is proposed by Warton (2008), which is derived through penalized likelihood function and the penalty parameter is chosen using a cross-validation technique. Other procedures include LASSO regularization that induce sparsity in the inverse covariance matrix by penalizing the Gaussian likelihood function (Friedman et al., 2008). Pourahmadi (2013) reviews some of the well-known regularization methods of the sample covariance matrix.

The target matrix plays a critical role in improving the accuracy of a shrinkage estimate (for example, see Schäfer & Strimmer, 2005). Substantial improvement in the accuracy may be achieved if the target matrix is correctly specified. Ledoit & Wolf (2004) have emphasized two requirements of a good target: first, it should involve a small number of parameters that need to be estimated; second, it should mimic some important characteristics the true covariance matrix. For example, identity matrix is a parameter free target and is ideal to fulfil the first requirement. However, it may not fulfil the second requirement in some cases. In order to fulfil the second requirement one would need a domain knowledge or a good target may be learned through data exploration.

In some high-dimensional data analysis problems, the objective has been to make things possible rather than to improve the results. The most commonly used target matrix is an identity matrix due to its simplicity and under the assumption of sparsity, which is a valid assumption in most of the high-dimensional settings. For example, the R package “corpcor” (Schaefer et al., 2013) specifies the identity matrix as a target. The main objective of this paper is to show that in certain situations using the identity matrix as a target may be inappropriate and using more informative targets may improve the results. Another objective is to propose a new method to estimate the penalty parameter by simply using the Gaussian likelihood function. The same likelihood function has been used to estimate the correlation parameter, which is the essential element of both the auto-regressive covariance structure and exchangeable covariance structure.

The paper is organized as follows. Our method is explained in Sections 2 and 3 and applied to simulated data and real world data in Section 4 and 5, respectively. Conclusions are presented in Section 6.

2. ESTIMATION OF REGULARIZATION PARAMETER USING NORMAL LIKELIHOOD

Let the real-valued observations are contained in $X = (x_1, \dots, x_n)$, where x_i is a p -dimensional sample realization made independently over n objects. Let μ denote the mean vector and Σ the covariance matrix. Assume also that the data in X is centred so

that $\mu = 0$, and thus the Gaussian log-likelihood function is

$$(2.1) \quad \log L(X; \Sigma) = \text{constant} - \frac{n}{2} \log |\Sigma| - \frac{1}{2} X' \Sigma^{-1} X,$$

where the first term on the right hand side does not depend on Σ . The maximum likelihood estimate of Σ is

$$S = \frac{1}{n} X' X$$

and its unbiased version is given by

$$\hat{\Sigma}_u = \frac{1}{n-1} X' X.$$

In high-dimensional applications the sample estimator (maximum likelihood estimate or its unbiased version) of the covariance matrix is not invertible. Regularization procedures make the estimated covariance invertible and more stable through incorporating additional information about the structure of the covariance matrix in the form of a target. For example, it may be reasonable to shrink the elements of a covariance matrix towards a diagonal matrix assuming sparsity, which is a sensible assumption in high-dimensional setting. The regularized estimate (a.k.a shrinkage estimate) takes the following form:

$$(2.2) \quad \hat{\Sigma}_\kappa = S + \kappa T,$$

where T is a positive definite (possibly informative) target matrix and $\kappa > 0$ is the regularization parameter. The expression in 2.2 incorporates additional information we might have about the structure of the covariance matrix through the target T . It is important for the parameter κ to have the following properties:

- For fixed p as we increase n , $\kappa \rightarrow 0$. Consequently, $\hat{\Sigma}_\kappa \rightarrow S$.
- For fixed n as we increase p , $\kappa \rightarrow \infty$. Consequently, $\hat{\Sigma}_\kappa \rightarrow T$.

Here we make use of the second property and assume that n is very small compared to p . We also assume that T is informative (structurally correctly specified). In this situation $\hat{\Sigma}_\kappa$ becomes similar (if not equal) to T , that is

$$T \approx S + \kappa T$$

or

$$(2.3) \quad \kappa T \approx S - T,$$

which can be exploited to find the value of κ as we do later in this section.

The expression in 2.3 can also be obtained by assuming that some scaled version of T is the true covariance and replacing Σ by $(1 + \kappa)T$ in equation 2.1, which then becomes

$$(2.4) \quad \log L(X; \Sigma_\kappa) = \text{Const} - \frac{n}{2} \log |T + \kappa T| - \frac{1}{2} X' \frac{1}{1 + \kappa} T^{-1} X.$$

Differentiating 2.4 with respect to κ we get

$$(2.5) \quad \frac{\partial}{\partial \kappa} \log L(X; \Sigma_\kappa) = -\frac{n}{2} (T + \kappa T)^{-1} T + \frac{1}{2} (T + \kappa T)^{-1} X' X (T + \kappa T)^{-1},$$

where $\frac{\partial}{\partial \kappa} \log |A| = \frac{1}{|A|} |A| A^{-t}$ and $\frac{\partial}{\partial \kappa} A^{-1} = -A^{-1} \frac{\partial A}{\partial \kappa} A^{-1}$. Equating equation 2.5 to zero leads to

$$(2.6) \quad \kappa T = S - T,$$

which is similar to the expression in 2.3. This leads to

$$(2.7) \quad \|\kappa T\|_1 = \|S - T\|_1,$$

where $\|\cdot\|_1$ denotes the sum of the absolute values of all entries in the matrix. If target T is the identity matrix then the value of κ becomes

$$(2.8) \quad \kappa = \frac{\|S - I\|_1}{p}.$$

Using AR(1) covariance structure as a target the expression for estimating κ becomes

$$(2.9) \quad \kappa = \frac{\|S - T\|_1}{p + 2 \sum_{k=1}^{p-1} kt^{p-k}},$$

and if T is of exchangeable structure then

$$(2.10) \quad \kappa = \frac{\|S - T\|_1}{p + p(p-1)t}.$$

It is important to highlight that for $t = 0$, 2.8 is a special case of 2.9 and 2.10. In all three cases the value of $\hat{\kappa}$ is the weighted average of the element-wise absolute error. Under a correctly specified target, the weighted average of element-wise absolute error tends to decrease as we increase n for any fixed value of p , since S is a consistent estimator of Σ . Therefore, the larger error leads to larger values of $\hat{\kappa}$ (hence more reliance on T), and vice versa if the error is small.

In order to restrict the range of the regularization parameter between zero and one, it is more appropriate to use the correlation scale rather than the covariance scale in equation 2.7. The correlation matrix can be obtained as

$$\hat{R} = S_d^{-\frac{1}{2}} S S_d^{-\frac{1}{2}},$$

where S_d is the diagonal matrix with corresponding diagonal elements of S on the main diagonal. To get the regularization parameter at correlation scale we follow Warton (2008) who regularized the correlation matrix (not the covariance matrix) and rescaled κ as

$$(2.11) \quad \gamma = \frac{1}{1 + \kappa},$$

where $\gamma \in [0, 1]$. The corresponding regularized estimator of the correlation matrix can be obtained as

$$(2.12) \quad \hat{R}_\gamma = \gamma \hat{R} + (1 - \gamma) T.$$

3. ESTIMATION OF CORRELATION PARAMETER

Unlike the identity matrix, the AR(1) and exchangeable targets rely on one additional parameter, which needs to be estimated. The AR(1) covariance is given by

$$(3.1) \quad \sigma_{ij} = t^{|i-j|} \quad \text{for } 1 \leq i, j \leq p,$$

and the exchangeable covariance structure is given by

$$(3.2) \quad \sigma_{ij} = \begin{cases} 1 & \text{when } i = j \\ t & \text{when } i \neq j \end{cases} \quad \text{for } 1 \leq i, j \leq p,$$

where t is an unknown parameter and can be estimated from the data (Wang & Carey, 2003). We denote the covariance structures by Σ_t to make its reliance on t explicit. The estimate of t is obtained by simply maximizing the log-likelihood function of the multivariate normal distribution, that is

$$(3.3) \quad \frac{\partial}{\partial t} \log L(X; \Sigma_t) = -\frac{n}{2} \frac{\partial}{\partial t} \log |\Sigma_t| - \frac{1}{2} \text{tr} \left(X^t \frac{\partial}{\partial t} \Sigma_t^{-1} X \right).$$

Now if Σ_t has AR(1) covariance structure then the determinant of Σ_t is $|\Sigma_t| = (1-t^2)^{p-1}$. Differentiating $\log |\Sigma_t|$ with respect to t we have $\frac{\partial}{\partial t} \log |\Sigma_t| = \frac{-2(p-1)t(1-t^2)^{p-2}}{(1-t^2)^{p-1}}$. The inverse of Σ_t is given by

$$\Sigma_t^{-1} = \frac{1}{(1-t^2)} \begin{pmatrix} 1 & -t & 0 & \dots & 0 & 0 \\ -t & 1+t^2 & -t & \dots & 0 & 0 \\ 0 & -t & 1+t^2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1+t^2 & -t \\ 0 & 0 & 0 & \dots & -t & 1 \end{pmatrix}.$$

Differentiating Σ_t^{-1} with respect to t is obtain

$$\frac{\partial}{\partial t} \Sigma_t^{-1} = \frac{1}{(1-t^2)^2} \begin{pmatrix} 2t & -(1+t^2) & 0 & \dots & 0 & 0 \\ -(1+t^2) & 4t & -(1+t^2) & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 4t & -(1+t^2) \\ 0 & 0 & 0 & \dots & -(1+t^2) & 2t \end{pmatrix}.$$

To find $\text{tr} \left(X^t \frac{\partial}{\partial t} \Sigma_t^{-1} X \right)$ in equation 3.3, we can write $X^t X = n\Sigma$, where Σ is true covariance matrix with entries σ_{ij} , $1 \leq i, j \leq p$. then,

$$(3.4) \quad \begin{aligned} n \text{tr} \left(\frac{\partial}{\partial t} \Sigma_t^{-1} \Sigma \right) &= \frac{1}{(1-t^2)^2} [2t(\sigma_{11} + 2\sigma_{22} + 2\sigma_{33} + \dots + 2\sigma_{(p-1)(p-1)} + \sigma_{pp})] \\ &\quad - t^2(\sigma_{12} + \sigma_{21} + \sigma_{23} + \dots + \sigma_{(p-1)(p)} + \sigma_{p(p-1)}) \\ &\quad - (\sigma_{12} + \sigma_{21} + \sigma_{23} + \dots + \sigma_{(p-1)(p)} + \sigma_{p(p-1)}). \end{aligned}$$

Since Σ is a symmetric matrix (that is, $\sigma_{ij} = \sigma_{ji}$) and its diagonal entries are all unity, we can write equation 3.4 as

$$(3.5) \quad n \operatorname{tr} \left(\frac{\partial}{\partial t} \Sigma_t^{-1} \Sigma \right) = \frac{2}{(1+t^2)^2} [t(p-1) - (t^2+1) \sum_{i=2}^p \sigma_{(i-1)i}].$$

Substituting equation $\frac{\partial}{\partial t} \log |\Sigma_t|$ and $n \operatorname{tr} \left(\frac{\partial}{\partial t} \Sigma_t^{-1} \Sigma \right)$ in equation 3.3 and equating it to zero leads to

$$(3.6) \quad t = \frac{\sum_{i=2}^p \sigma_{(i-1)i}}{p-1}.$$

Similarly, if Σ_t has exchangeable structure then $|\Sigma_t| = (1-t)^{p-1} \{1 + (p-1)t\}$. Differentiating $\log |\Sigma_t|$ with respect to t gives

$$(3.7) \quad \frac{\partial}{\partial t} \log |\Sigma_t| = \frac{(1-t)^{p-1}(p-1) + \{1 + (p-1)t\}(p-1)(1-t)^{p-2}}{(1-t)^{p-1} \{1 + (p-1)t\}}.$$

In this case the inverse of Σ_t is

$$\begin{aligned} \Sigma_t^{-1} &= \frac{1}{(1-t)\{1 + (p-1)t\}} \begin{pmatrix} 1+t & -t & \dots & -t \\ -t & 1+t & \dots & -t \\ \vdots & \vdots & \ddots & \vdots \\ -t & -t & \dots & 1+t \end{pmatrix} \\ &= \frac{1}{(1-t)} \left[\mathbf{I} - \frac{t}{\{1 + (p-1)t\}} \mathbf{J} \right], \end{aligned}$$

where \mathbf{I} is the identity matrix and \mathbf{J} is the unit matrix. Differentiating Σ_t^{-1} with respect to t we get

$$(3.8) \quad \frac{\partial}{\partial t} \Sigma_t^{-1} = \frac{1}{(1-t)^2} \mathbf{I} - \left[\frac{1}{\{1 + (p-1)t\}^2 (1-t)} + \frac{t}{\{1 + (p-1)t\} (1-t)^2} \right] \mathbf{J}.$$

Then,

$$(3.9) \quad \begin{aligned} \operatorname{tr} \left(\frac{\partial}{\partial t} \Sigma_t^{-1} \Sigma \right) &= \frac{p}{(1-t)^2} - p \left[\frac{1}{\{1 + (p-1)t\}^2 (1-t)} + \frac{t}{\{1 + (p-1)t\} (1-t)^2} \right] \\ &\quad - \left[\frac{1}{\{1 + (p-1)t\}^2 (1-t)} + \frac{t}{\{1 + (p-1)t\} (1-t)^2} \right] \sum_{i \neq j}^p \sigma_{ij}. \end{aligned}$$

Substituting the values of $\frac{\partial}{\partial t} \log |\Sigma_t|$ and $n \operatorname{tr} \left(\frac{\partial}{\partial t} \Sigma_t^{-1} \Sigma \right)$, respectively, from 3.7 and 3.9 into equation 3.3 and equating it to zero leads to

$$(3.10) \quad t = \frac{\sum_{i \neq j}^p \sigma_{ij}}{p(p-1)}.$$

4. SIMULATION STUDY

In this section, we present results of an extensive simulation study conducted to numerically demonstrate the performance of the proposed method in comparison with the shrinkage method proposed by Schäfer & Strimmer (2005) and implemented in R package “corpcor” (Schaefer et al., 2013). To examine the behaviour of the proposed method under different simulation settings, we generate various datasets taking into account a

number of different factors. These factors include varying sample sizes, numbers of variables, and covariance structures.

We draw a sample of size n from a p -variate normal distribution with mean vector, $\mu = 0$, and covariance matrix, Σ . In order to evaluate the performance in a variety of situations, we consider four different covariance structures for Σ : AR(1), exchangeable, random covariance structure (see the algorithm in [Schäfer & Strimmer, 2004](#)) and an identity matrix.

4.1 Empirical properties of $\hat{\gamma}$

Here we examine the properties of $\hat{\gamma}$ under different simulation settings. The values of $\hat{\gamma}$ exhibit some interesting properties in simulations as summarized in Figure 1. First, for a fixed p , as n increases, $\hat{\gamma}$ on the average increases, which indicates that as more data become available the sample covariance is less penalized. In other words the regularized estimator, Σ_{γ} , is a consistent estimator of the true covariance matrix. Second, for a fixed n , as p increases, $\hat{\gamma}$ on the average decreases. This translates to increased reliance of Σ_{γ} on the information provided (in the form of a target matrix) with increasing instability in S .

The $\hat{\gamma}$ values are not consistent across different covariance structures considered here. This is mostly because of the targets specified, which we specify correctly in all three cases; by “correctly”, we mean when both true covariance matrix and target have same structure. In case of the identity matrix as a target no parameter is needed to be estimated from the data. Thus with increasing n , the sample covariance becomes more accurate, which results in decreased reliance on the target matrix. In case of AR(1) and exchangeable structures, the targets also depend on the data through the estimate of the parameter t . Estimating the parameter t accurately translates to accurate estimation of the entire target matrix. Therefore, with increasing n both the sample covariance and the targets become closer to the true covariance. This is more obvious in the case of exchangeable covariance.

Note also that the variability in $\hat{\gamma}$ tend to decrease as n increases, for a fixed value of p .

4.2 Comparative performance of $\hat{\Sigma}_{\gamma}$

We used the sum of absolute errors in eigenvalues to compare the accuracy of the proposed method with the maximum likelihood estimate and the shrinkage method of [Schäfer & Strimmer \(2005\)](#). We showed the results for $n = 50$ and $p \in \{30, 50, 100\}$ to demonstrate the effect of increasing number of variables for a fixed value of n . For AR(1) and exchangeable covariance structures, we shrunk the estimated covariance matrix towards the correct targets that are, respectively, AR(1) and exchangeable. We also examined the performance in the cases where target matrix is incorrectly specified as identity matrix while the true covariance matrix is AR(1) or exchangeable. However, for random covariance structure we used only identity matrix as a target, which although incorrect, has been used extensively as a shrinkage target to regularize the covariance matrix (see, for example, [Ledoit & Wolf, 2003](#); [Schäfer & Strimmer, 2005](#)). Note that although we have conducted the experiments for a range of values of t , we show here the results only for $t = 0.5$ for both AR(1) and exchangeable structures (due to consistency of the results for different values of t). Similarly, for random covariance structure we showed results for a covariance matrix whose inverse contains 30% of the off-diagonal positions as being non-zero.

The covariance matrix is estimated using the proposed method and shrinkage method of [Schäfer & Strimmer \(2005\)](#). For the proposed method we estimate t using Gaussian estimating equations as described in Section 3. The sum of absolute errors in estimated eigenvalues averaged over 1000 simulated datasets are presented in Figure 2. The eigenvalues of sample covariance are also plotted for comparison purpose.

From the simulation results, it is clear that whenever the target is correctly specified in case of both AR(1) and exchangeable covariance structures, the proposed method performs better than the shrinkage method as it maintains the smallest estimation error and is also much precise than the competing methods. Moreover, as p increases the estimates obtained by using the proposed method become more accurate and precise. Its performance can become slightly weaker than the shrinkage method if the target is incorrectly specified as the identity matrix while the true covariance matrix is AR(1) or exchangeable. In case of random covariance structure when we use the identity matrix as a target, our proposed method also outperforms the shrinkage method in terms of accuracy and precision.

5. PRACTICAL APPLICATION

In this section, we apply the proposed method to real world data of soil compaction profiles previously analysed by [Ullah & Jones, 2015](#)). The data are from seven different ridges in the eastern Qilian Mountains, China, a rangeland habitat. Along each sunny ridge slope, samples are taken at three different heights. The heights are categorized as low, medium and high, so there are total 21 locations. At each of these 21 locations, measurements are taken at 18 consecutive depths (depths are our variables) from 2.5cm to 45cm with an interval of 2.5cm. So in summary, we have a total of 18 variables (depths) and 21 observations each of them can be categorised into one of the three height levels that is low, medium and high.

To test the height effect (low, medium and high), we used multivariate analysis of variance (MANOVA) ([Anderson et al., 1958](#)). The group-centred observations for consecutive depths are highly correlated and the correlation decreases with increasing distance between depths suggesting correlation of an AR(1) structure (see Figure 3 which is presented in [Ullah & Jones \(2015\)](#) and reproduced here for convenience). We obtain the estimate of the parameter t using the method described in Section 3 assuming an AR(1) structure, $\hat{t} = 0.9162$. Figure 4 shows the eigenvalues of the sample correlation matrix of the group-centred observations in comparison to the eigenvalues from the estimated AR(1) structure with $\hat{t} = 0.9162$, which further supports the perceived AR(1) structure.

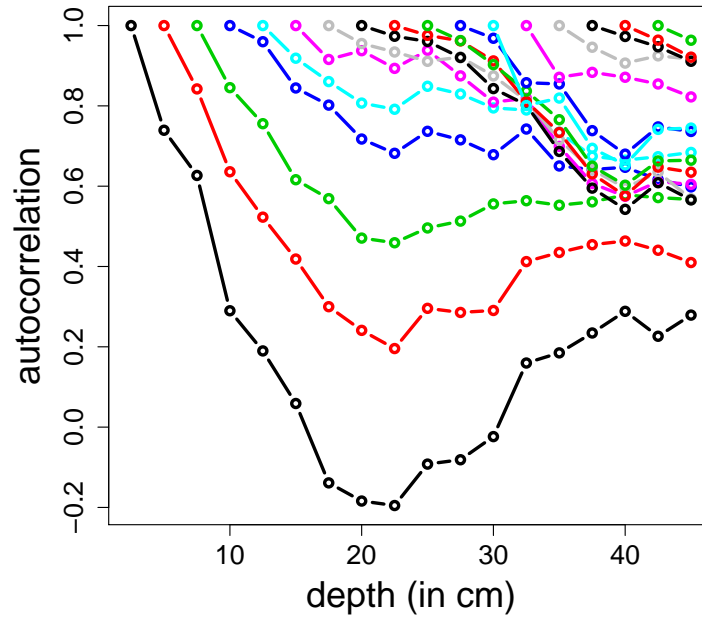


FIG 3. Each sequence of connected points shows the correlation of measurements at a particular depth with those at greater depths. For each of the 18 depths (variables) we have 21 observations. The correlations are high for proximate depths, but decreases with increasing distance.

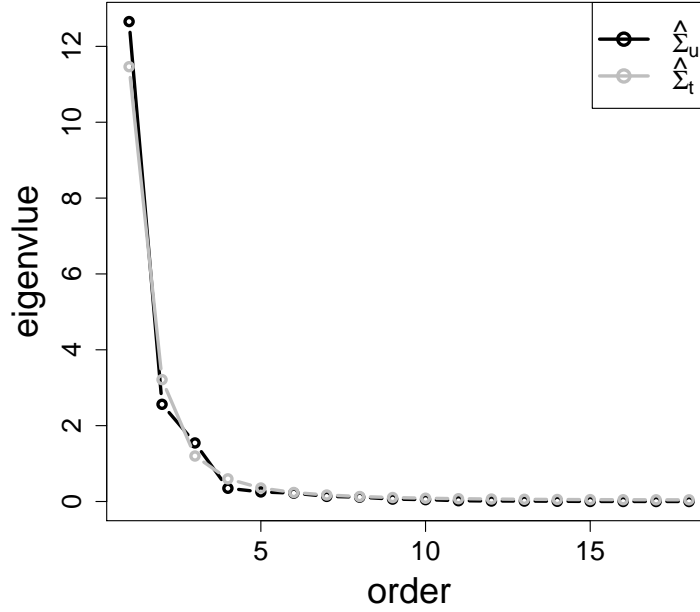


FIG 4. Scree plots based on eigenvalues of $\hat{\Sigma}_u$ and $\hat{\Sigma}_t$.

To compare the performance of the two methods in the three situations $n > p$, $n = p$ and $n < p$, we took random samples of different sizes (21, 18, 15, and 12) from 21 observations. The first set of 1000 samples each of size 21 was selected in such a way so that each of the three groups contribute seven observations with replacement, the second set of 1000 samples each of size 18 was selected in such a way so that each of the three groups contribute six observations with replacement, and so on. We fitted MANOVA model to each set of 1000 samples and summarized the significant results (at 0.05 level of significance) of fitted MANOVA models in Table 1.

The MANOVA model based on our proposed method and shrinkage method performed equally well in the $n > p$ case, that is, the mean difference was significant under both methods for all 1000 samples. However, the performance of the shrinkage method started to deteriorate sooner than the proposed method as we reduced the sample sizes. In the situation of $n = p$, the proposed method maintained the significance rate at 100% while the significance rate for shrinkage method dropped to 97.8%. Finally, in the situation of $n < p$, the significance rate for the proposed method (99.9% and 92.5%, respectively, under the sample size 15 and 12) remained much higher than significance rate of the shrinkage method (92.5% and 61.1%, respectively, under the sample size 15 and 12).

TABLE 1

Proportion of p -values less than 0.05. 1000 p -values are calculated by randomly selecting 7, 6, 5 and 4 samples (ridges) from each group using two different competing procedures.

No. of observations	Proposed method	Shrinkage
21	1.000	1.000
18	1.000	0.978
15	0.999	0.883
12	0.925	0.611

6. CONCLUSION AND DISCUSSION

In the high-dimensional setting, sample covariance exhibits some unfavourable characteristics. This is generally overcome by applying some regularization to the sample covariance, incorporating some additional information via shrinking the sample covariance towards a target. The most commonly used target is the identity matrix, probably because it satisfies some of the fundamental requirements of a good target. However, in some applications other structural targets may be preferred. In this paper, we have proposed to shrink the sample covariance matrix towards informative targets, namely the AR(1) and exchangeable covariance structures. These targets depend on an additional parameter that needs to be estimated, which we estimate by maximizing the log-likelihood function of the multivariate normal distribution. The regularized estimator depends on the penalty parameter, whose value needs to be chosen from within a range of possible values. We propose to choose the value of the penalty parameter by maximizing the log-likelihood function of the multivariate normal distribution. Being analytically simpler and computationally inexpensive, the proposed estimator is not only invertible but also well-conditioned.

To evaluate the empirical performance of the proposed estimator, we conducted a simulation study. The simulation experiments showed the improved performance of the proposed estimator over the shrinkage estimator whenever the target is correctly specified in case of AR(1) and exchangeable covariance structures. However, its performance became slightly weaker than the shrinkage estimator when the target was incorrectly specified as AR(1) or exchangeable while the true covariance matrix was the identity matrix. In the case of random covariance structure our proposed estimator performed better than the shrinkage estimator. The proposed method was also tested on a real world dataset and we found that the proposed method maintained the significance rate (when embedded in MANOVA model) comparatively much better than the shrinkage method when the sample size was reduced.

REFERENCES

- Anderson, T. W., Anderson, T. W., Anderson, T. W., Anderson, T. W., & Mathématicien, E.-U. (1958). *An introduction to multivariate statistical analysis* (Vol. 2). Wiley New York.
- Beerenwinkel, N., Antal, T., Dingli, D., Traulsen, A., Kinzler, K. W., Velculescu, V. E., . . . Nowak, M. A. (2007). Genetic progression and the waiting time to cancer. *PLoS computational biology*, 3(11), e225.
- Dudoit, S., Fridlyand, J., & Speed, T. P. (2002). Comparison of discrimination methods for the classification of tumors using gene expression data. *Journal of the American statistical association*, 97(457), 77–87.
- Eisen, M. B., Spellman, P. T., Brown, P. O., & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences*, 95(25), 14863–14868.
- Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3), 432–441.
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67.
- Ledoit, O., & Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of empirical finance*, 10(5), 603–621.
- Ledoit, O., & Wolf, M. (2004). Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4), 110–119.
- Penrose, R. (1955). A generalized inverse for matrices. In *Mathematical proceedings of the cambridge philosophical society* (Vol. 51, pp. 406–413).
- Pourahmadi, M. (2013). *High-dimensional covariance estimation: with high-dimensional data*. John Wiley & Sons.
- Schaefer, J., Opgen-Rhein, R., Zuber, V., Ahdesmäki, M., Duarte Silva, A., & Strimmer, K. (2013). corpcor: Efficient estimation of covariance and (partial) correlation. *R package version*, 1(6).
- Schäfer, J., & Strimmer, K. (2004). An empirical bayes approach to inferring large-scale gene association networks. *Bioinformatics*, 21(6), 754–764.
- Schäfer, J., & Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical applications in genetics and molecular biology*, 4(1).
- Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., . . . Golub, T. R. (1999). Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proceedings of the National Academy of Sciences*, 96(6), 2907–2912.
- Ullah, I., & Jones, B. (2015). Regularised manova for high-dimensional data. *Australian & New Zealand Journal of Statistics*, 57(3), 377–389.
- Wang, Y.-G., & Carey, V. (2003). Working correlation structure misspecification, estimation and covariate design: implications for generalised estimating equations performance. *Biometrika*, 90(1), 29–41.
- Warton, D. I. (2008). Penalized normal likelihood and ridge regularization of correlation and covariance matrices. *Journal of the American Statistical Association*, 103(481), 340–349.

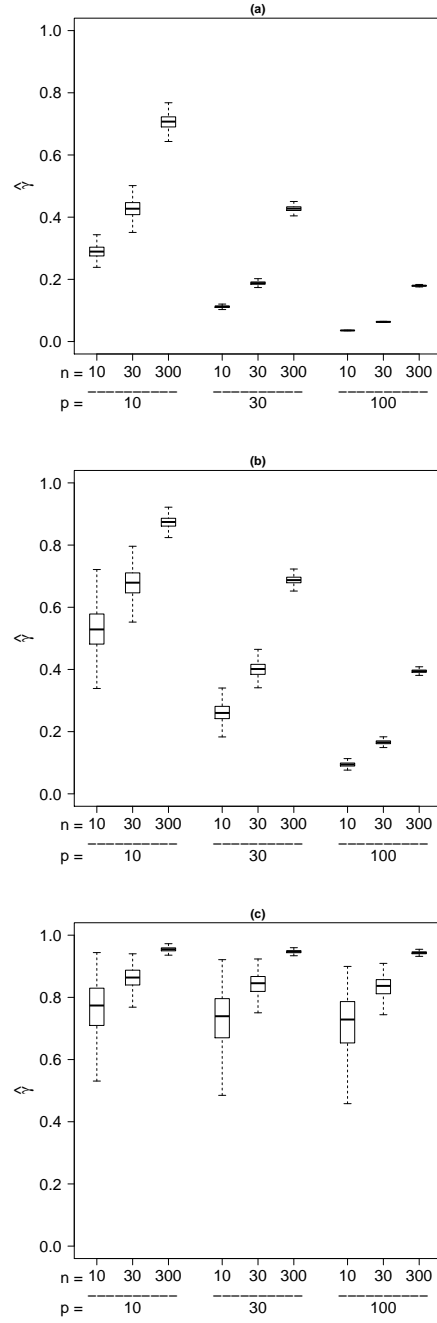


FIG 1. Distribution of $\hat{\gamma}$ values obtained in 1000 simulations from multivariate normal distribution with $\mu = 0$ and covariance matrix: (a) identity (b) AR(1) with $t = .5$ (c) exchangeable with $t = .5$. The target matrices are correctly specified in all three cases: (a) identity (b) AR(1) and (c) exchangeable. The results are shown for $n \in \{10, 30, 300\}$ and $p \in \{10, 30, 100\}$.

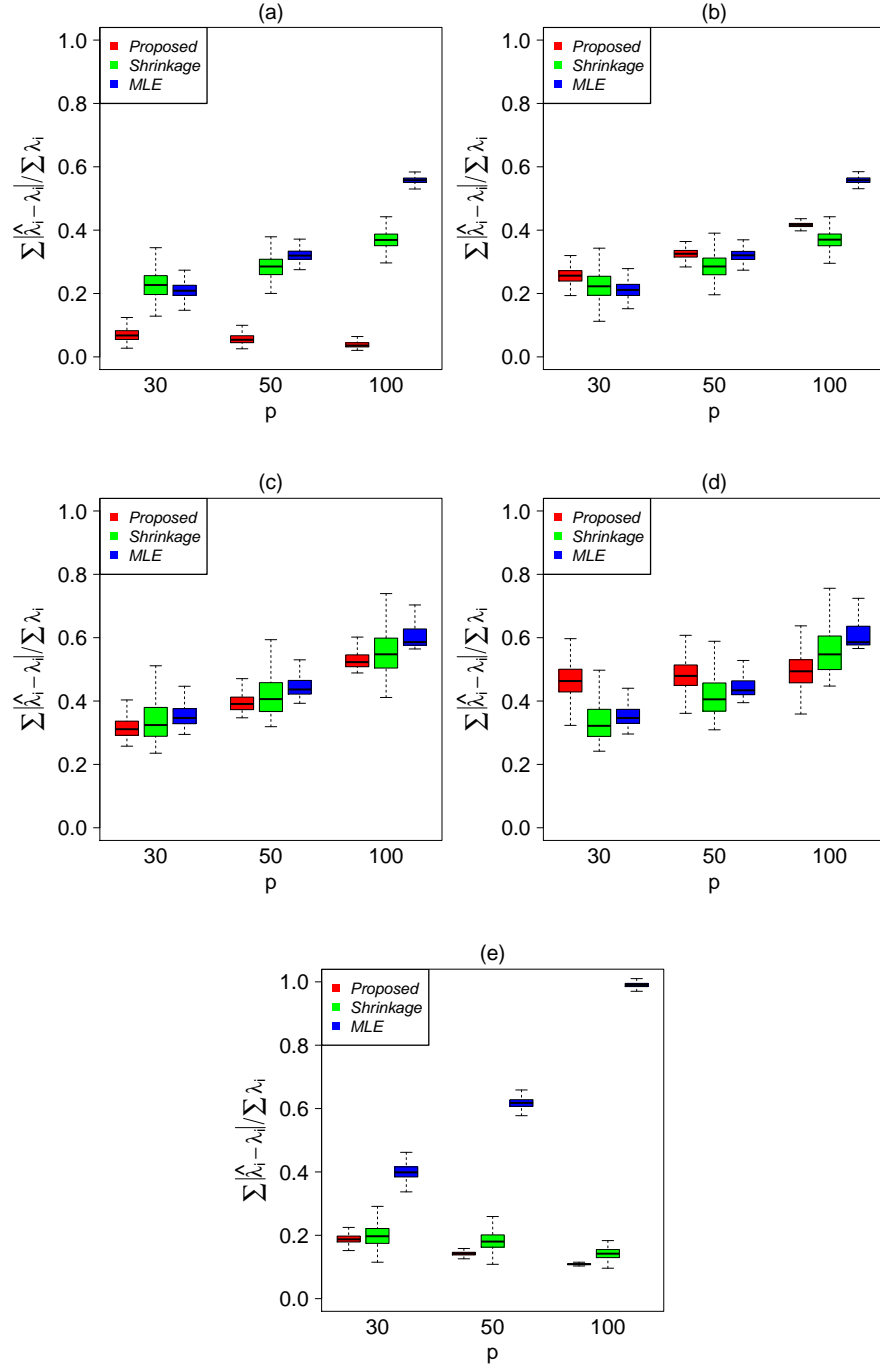


FIG 2. Distribution of the sum of absolute errors in estimated eigenvalues over 1000 simulations using maximum likelihood estimate, proposed method, and shrinkage estimate of Schafer & Strimmer (2005) under different choices of covariance matrices: (a) and (b) true covariance is AR(1) with $t = 0.5$, (c) and (d) true covariance is exchangeable with $t = 0.5$ and (e) true covariance is random of structure with 30% off-diagonal entries as non-zero. For the proposed method the target is: (a) AR(1), (b) misspecified as identity matrix, (c) exchangeable, (d) misspecified as identity matrix and (e) identity matrix. The data are generated from multivariate normal distribution with $n = 50$ and $p \in \{30, 50, 100\}$.