THRESHOLDING APPROACH FOR LOW-RANK CORRELATION MATRIX BASED ON MM ALGORITHM

A PREPRINT

Kensuke Tanioka*
Department of Biomedical Sciences and Informatics
Doshisha University
Kyoto, Japan
ktanioka@mail.doshisha.ac.jp

Yuki Furotani
Graduate School of Life and Medical Sciences, Doshisha University
Kyoto, Japan
yfurotani@mis.doshisha.ac.jp

Satoru Hiwa
Department of Biomedical Sciences and Informatics
Doshisha University
Kyoto, Japan
shiwa@mail.doshisha.ac.jp

December 29, 2021

ABSTRACT

Background Low-rank approximation is a very useful approach for interpreting the features of a correlation matrix; however, a low-rank approximation may result in estimation far from zero even if the corresponding original value was far from zero. In this case, the results lead to misinterpretation.

Methods To overcome these problems, we propose a new approach to estimate a sparse low-rank correlation matrix based on threshold values combined with cross-validation. In the proposed approach, the MM algorithm was used to estimate the sparse low-rank correlation matrix, and a grid search was performed to select the threshold values related to sparse estimation.

Results Through numerical simulation, we found that the FPR and average relative error of the proposed method were superior to those of the tandem approach. For the application of microarray gene expression, the FPRs of the proposed approach with $d = 2,\ 3$, and 5 were 0.128, 0.139, and 0.197, respectively, while FPR of the tandem approach was 0.285.

Conclusions We propose a novel approach to estimate sparse low-rank correlation matrix. The advantage of the proposed method is that it provides results that are easy to interpret and avoid misunderstandings. We demonstrated the superiority of the proposed method through both numerical simulations and real examples.

Keywords Cross Validation · Proportional-threhold · Sparse estimation

*

1 Background

To describe the linear relationship between two variables or subjects, a correlation matrix is calculated from multivariate data. For example, in the domain of genomics, a correlation matrix between genes is used in combination with the heatmap [Wilkinson and Friendly, 2009]. In addition, to classify sample profiles based on multivariate data, a correlation matrix between these samples is also used [Wei and Li, 2010]. However, network structures of these correlation coefficients are masked because real data are observed including noise.

An approach to this problem is the use of low-rank approximation [ten Berge, 1993]. For the estimation of low-rank correlation matrices, various methods have been proposed [Pietersz and Groenen, 2004, Simon and Abell, 2010, Grubii and Pietersz, 2007, Duan et al., 2016], and there are two advantages for this estimati2on. First, it becomes easy to interpret network structure of the correlation matrix because the variation of these low-rank correlation coefficients tend to be larger. Therefore, the heat map of the estimated low-rank correlation matrix provides a readable visualization. Second, such a low-rank approximation also can describe the clustering structure [Ding and He, 2004], which helps us to improve the interpretation. Finally, the values of the estimated low-rank correlation coefficient can range from -1 to 1, in contrast to the result of singular valued decomposition (SVD). Therefore, researchers can interpret the relationship easily because these values are bounded.

It is indeed the case that low-rank correlation matrices simplify these relations; however, there are several problems with this. First, currently, the size of data tends to become larger owing to improvement in information technology. Therefore, the number of coefficients needed for interpretation is above the cognitive boundary. Second, even if a true correlation coefficient is close to zero, the corresponding estimated low-rank correlation coefficient can be far from zero.

To overcome this problem, in this study, we propose a new approach to estimate a sparse low-rank correlation matrix. The proposed approach, combined with the heat map, provides a visual interpretation of the relationships between variables. For the sparse methods of the original correlation matrix and covariance matrix [Engel et al., 2017, Lam, 2020], there are two types. The first is to add sparsity penalty to the objective functions [Bien and Tibshirani, 2011, Rothman, 2012, Xue et al., 2012, Cai et al., 2011, D'aspremont et al., 2008, Friedman et al., 2008, Rothman et al., 2008, Yuan and Lin, 2007, Cui et al., 2016]. The other type uses thresholding values to achieve a sparse structure. [Bickel and Levina, 2008a] proposed the thresholding matrix estimator and various related methods have been developed [Cai and Liu, 2011, Bickel and Levina, 2008b, El Karoui, 2008, Jiang, 2013]. In addition, to estimate sparse correlation matrix, [Rothman et al., 2009, Lam and Fan, 2009, Liu et al., 2014] used generalized thresholding operator based methods [Rothman et al., 2009], respectively. For the estimation of sparse low-rank matrices, methods based on penalty terms have also been proposed [Zhou et al., 2015, Savalle et al., 2012].

In the proposed approach, we adopt an approach that uses hard thresholding based on [Bickel and Levina, 2008a] and [Jiang, 2013] because the approach is quite simple and easy to interpret the results. Therefore, to estimate sparse low-rank correlation matrix, we combine the majorize-minimization algorithm (MM algorithm) [Pietersz and Groenen, 2004, Simon and Abell, 2010] and the hard thresholding approach. There are two advantages for the proposed approach. First, the estimated sparse low-rank correlation matrix allow for easy interpretation the correlation matrix. Second, the proposed approach can avoid misunderstanding the interpretation of the correlation matrix. When the true correlation coefficient is zero, the proposed method can correctly estimate the corresponding coefficient as zero. In addition, we focus only on positive correlation coefficients, not negative correlation coefficients. From the focus on only positive relations, it becomes easy to interpret the features of relations easily.

The rest of this paper is structured as follows. We explain the model and algorithm in section 2. In section3, we evaluate the proposed approach, and a numerical simulation is presented. In addition, the results of applying the proposed method to real data are provided in Section 4. Finally, we conclude out study and discuss the proposed approach in Section 5.

2 Method

2.1 Adaptive thresholding for sparse and low-rank correlation matrix estimation

In this section, we present the proposed approach for estimating a sparse low-rank correlation matrix. First, the estimation of the low-rank correlation matrix is introduced based on the MM algorithm [Pietersz and Groenen, 2004, Simon and Abell, 2010]. Next, to achieve the sparse low-rank correlation structure, the hard thresholding operator and proposed cross-validation function are described.

2.1.1 Optimization problem of low-rank correlation matrices

Let $\boldsymbol{R} = (r_{ij})\ r_{ij} \in [-1,\ 1]\ (i, j = 1, 2, \cdots, p)$ and $\boldsymbol{W} = (w_{ij})\ w_{ij} \in \{0, 1\}\ (i, j = 1, 2, \cdots, p)$ be correlation matrix between variables and the binary matrix, respectively, where p is the number of variables. Given the number of low dimensions $d \leq p$ and correlation matrix $\boldsymbol{R}$, binary matrix $\boldsymbol{W}$, the optimization problem of estimating a low-rank correlation matrix is defined as follows.

$$f(\boldsymbol{Y}|\ \boldsymbol{R}, \boldsymbol{W}) = \left\|\boldsymbol{R} - \boldsymbol{W} \odot \boldsymbol{Y}\boldsymbol{Y}^T\right\|_F^2 \longrightarrow \text{Min.} \tag{1}$$

subject to

$$\|\boldsymbol{y}_j\| = 1 \quad \text{for all } j = 1, 2, \cdots, p. \tag{2}$$

where $\boldsymbol{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_p)^T, \boldsymbol{y}_j = (y_{j1}, y_{j2}, \cdots, y_{jd})^T, y_{jo} \in \mathbb{R}\ (j = 1, 2, \cdots, p;\ o = 1, 2, \cdots, d)$ is coordinate matrix of variables on d dimensions, $\odot$ is the Hadamard product, and $\|\cdot\|_F$ is Frobenius norm. The objective function in Eq. (1) was explained by [Knol and Ten Berge, 2012]. From the constraint (2), $\boldsymbol{Y}^T\boldsymbol{Y}$ becomes the correlation matrix.

2.1.2 Estimation of low-rank correlation matrices based on MM algorithm

To estimate a low-rank correlation matrix, the MM algorithm proposed by [Simon and Abell, 2010] is explained. To estimate $\boldsymbol{Y}$ under the constraint (2), the quadratic optimization problem for $\boldsymbol{Y}$ must be converted to a linear optimization problem. Using the linear function, we can derive the updated formula combined with the Lagrange multiplier in closed form. Let $\boldsymbol{y}^{(t)} \in \mathbb{R}^d$ be the parameter of t step in the algorithm of the optimization problem, and $g(\boldsymbol{y}|\boldsymbol{y}^{(t)})$ be a real function $g : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}$. If $g(\boldsymbol{y}|\boldsymbol{y}^{(t)})$ satisfy the following conditions such that

$$\forall \boldsymbol{y} \in \mathbb{R}^d; \quad g(\boldsymbol{y}|\boldsymbol{y}^{(t)}) \geq f(\boldsymbol{y}) \qquad \text{and} \tag{3}$$

$$g(\boldsymbol{y}^{(t)}|\boldsymbol{y}^{(t)}) = f(\boldsymbol{y}^{(t)}), \tag{4}$$

$g(\boldsymbol{y}|\boldsymbol{y}^{(t)})$ is defined as majorizing function of $f(\boldsymbol{y})$ at the point $\boldsymbol{y}^{(t)}$, where $f : \mathbb{R}^d \mapsto \mathbb{R}$ is the original function. Simply, to estimate the parameters in MM algorithm, we minimize $g(\boldsymbol{y}|\boldsymbol{y}^{(t)})$, not $f(\boldsymbol{y})$. In many situations, $g(\boldsymbol{y}|\boldsymbol{y}^{(t)})$ is expected to minimize the value easily. For the detail of MM algorithm, see [Hunter and Lange, 2004].

Before deriving the majorizing function, the objective function (1) is redescribed as follows:

$$\begin{aligned} f(\boldsymbol{Y}|\ \boldsymbol{R}, \boldsymbol{W}) =& \left\|\boldsymbol{R} - \boldsymbol{W} \odot \boldsymbol{Y}\boldsymbol{Y}^T\right\|_F^2 \\ =& \sum_{i=1}^{p}\sum_{j\neq i}^{p}(r_{ij} - w_{ij}\boldsymbol{y}_i^T\boldsymbol{y}_j)^2 \\ =& c_1 + \sum_{i=1}^{p}\boldsymbol{y}_i^T\Big(\sum_{j\neq i} w_{ij}\boldsymbol{y}_j\boldsymbol{y}_j^T\Big)\boldsymbol{y}_i - 2\sum_{i=1}^{p}\boldsymbol{y}_i^T\Big(\sum_{j\neq i} w_{ij}r_{ij}\boldsymbol{y}_j\Big) \\ =& c_1 + \sum_{i=1}^{p}\boldsymbol{y}_i^T\boldsymbol{B}_i\boldsymbol{y}_i - 2\sum_{i=1}^{p}\boldsymbol{y}_i^T\Big(\sum_{j\neq i} w_{ij}r_{ij}\boldsymbol{y}_j\Big), \end{aligned} \tag{5}$$

where $\boldsymbol{B}_i = \sum_{j\neq i} w_{ij}\boldsymbol{y}_j\boldsymbol{y}_j^T$ and c_1 are constants. Here, the parameter estimation of $\boldsymbol{Y}$ is conducted by $\boldsymbol{y}_i$. The corresponding part of Eq.(5) and the majorizing function can be described as follows:

$$f_i(\boldsymbol{y}_i|\ \{\boldsymbol{y}_j\}_{j\neq i}) = \boldsymbol{y}_i^T\boldsymbol{B}_i\boldsymbol{y}_i - 2\boldsymbol{y}_i^T\Big(\sum_{j\neq i} w_{ij}r_{ij}\boldsymbol{y}_j\Big) \tag{6}$$

$$\leq c_2 - 2\boldsymbol{y}_i^T(\lambda_i\boldsymbol{I}_d - \boldsymbol{B}_i)\boldsymbol{y}_i^{(t-1)} - 2\boldsymbol{y}_i^T\Big(\sum_{j\neq i} w_{ij}r_{ij}\boldsymbol{y}_j\Big) \tag{7}$$

$$= g(\boldsymbol{y}_i|\ \boldsymbol{y}_i^{(t-1)},\ \{\boldsymbol{y}_j\}_{j\neq i}),$$

where $g(\boldsymbol{y}_i|\ \boldsymbol{y}_i^{(t-1)},\ \{\boldsymbol{y}_j\}_{j\neq i})$ represents the majorizing function of Eq.(5), c_2 is constant. $\boldsymbol{I}_d$ is $d \times d$ identity matrix, λ_i is the maximum eigenvalue of $\boldsymbol{B}_i$, and $\boldsymbol{x}_i^{(t-1)}$ is $\boldsymbol{y}_i$ of $(t-1)$ step in the algorithm. Here, the inequality of Eq. (7) is satisfied because $\lambda_i\boldsymbol{I}_d - \boldsymbol{B}_i$ is negative semi-definite. In fact, if $\boldsymbol{y}_i = \boldsymbol{y}_i^{(t-1)}$, Eq.(6) and Eq.(7) becomes equal.

Using Lagrange multiplier method and Eq. (7), the updated formula of $\boldsymbol{y}_i$ is derived as follows:

$$\boldsymbol{y}_i^{(t)} \leftarrow \frac{\lambda_i \boldsymbol{y}_i^{(t-1)} - \boldsymbol{B}_i \boldsymbol{y}_i^{(t-1)} + \sum_{j \neq i} w_{ij} r_{ij} \boldsymbol{y}_j}{\left\| \lambda_i \boldsymbol{y}_i^{(t-1)} - \boldsymbol{B}_i \boldsymbol{y}_i^{(t-1)} + \sum_{j \neq i} w_{ij} r_{ij} \boldsymbol{y}_j \right\|} \quad (i = 1, 2, \cdots, p). \tag{8}$$

Algorithm 1 Algorithm for estimating low-rank correlation matrix

Input: $\boldsymbol{R}$, $d(\leq p)$ and small constant $\varepsilon > 0$
Output: $\boldsymbol{Y}$

Initialisation : Set $\boldsymbol{Y}$ satisfying $\|\boldsymbol{y}_j\| = 1$ for all j
while $f(\boldsymbol{Y}^{(t-1)}) - f(\boldsymbol{Y}^{(t)}) \geq \varepsilon$ **do**
  for $i = 1$ to p
    Calculate $\boldsymbol{B}_i \leftarrow \sum_{j \neq i} w_{ij} \boldsymbol{y}_j \boldsymbol{y}_j^T$
    Calculate λ_i to be the largest eigenvalue of $\boldsymbol{B}_i$
    Update $\boldsymbol{y}_i$ based on Eq.(8).
  End for
End while
return $\boldsymbol{Y}$ with the constraint $\|\boldsymbol{y}_j\| = 1$ for all j

2.1.3 Proposed Algorithm of Cross Validation to determine hard thresholds

In the proposed approach, to estimate a sparse low-rank correlation matrix, we adopt hard thresholding. To determine the threshold values, we introduce a cross-validation function based on [Bickel and Levina, 2008a]. The purpose of this approach is quite simple, and the proposed approach can determine the threshold values related to sparse estimation by considering the corresponding rank.

Let $h(\alpha) \in (-1, 1)$ be a threshold value corresponding to the α percentile of correlations, where $\alpha \in [0, 1]$ is the percentage point. For a correlation $r_{ij} \in [-1, 1]$, the function is defined as $1 = \mathbb{1}_{h(\alpha)}[r_{ij} \geq h]$ if $r_{ij} \geq h(\alpha)$, else $0 = \mathbb{1}_{h(\alpha)}[r_{ij} \geq h(\alpha)]$. Using them, proportional threshold operator is defined as

$$T_{h(\alpha)}(\boldsymbol{R}) = (\tilde{r}_{ij}), \quad (i, j = 1, 2, \cdots, p) \tag{9}$$

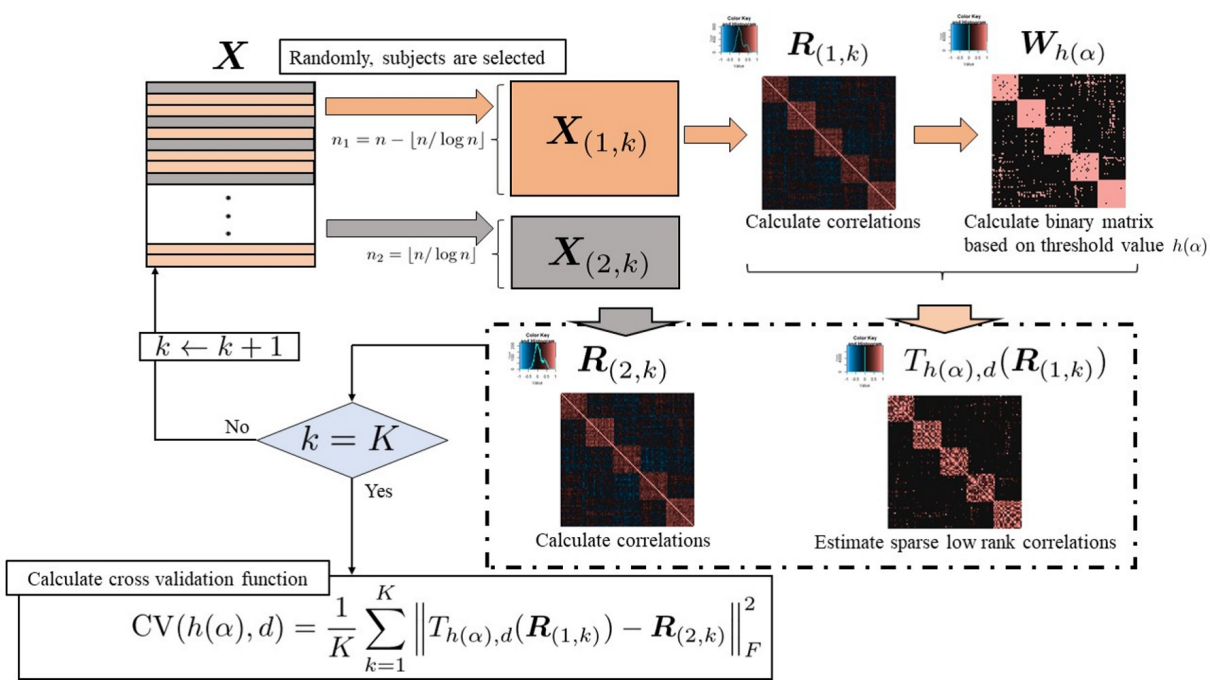

Figure 1: Framework of the proposed cross validation

where

$$\tilde{r}_{ij} = r_{ij} \cdot \mathbb{1}_{h(\alpha)}[r_{ij} \geq h(\alpha)] \quad (i, j = 1, 2, \cdots, p). \tag{10}$$

The proportional threshold operator is used in the domain of neural science [van den Heuvel et al., 2017]. Here, Eq.(9) can be described as $\boldsymbol{R}$ and the binary matrix $\boldsymbol{W}_{h(\alpha)}$ corresponding to $h(\alpha)$ such that $T_{h(\alpha)}(\boldsymbol{R}) = \boldsymbol{W}_{h(\alpha)} \odot \boldsymbol{R}$, where $\boldsymbol{W}_{h(\alpha)} = (w_{ij}^{(h(\alpha))} r_{ij}) = (\tilde{r}_{ij})$ and $w_{ij}^{(h(\alpha))} r_{ij} \in \{0,\ 1\}$. Here, Eq.(10) is modified for the original function of Bickel and Levina [2008a]. Originally, $\mathbb{1}_{h(\alpha)}[|r_{ij}| \geq h(\alpha)]$ is used, however, we only focus on higher correlations and not on negative correlation. Using modifications, it becomes easy to interpret the results.

To estimate sparse a low-rank correlation matrix, We use the proportional threshold operator in Eq.(9) because the interpretation of the proportional threshold is quite simple. For the choice of the threshold value $h(\alpha)$, a cross-validation was introduced (e.g., Bickel and Levina [2008a], Jiang [2013]). The cross-validation procedure for the estimation of $h(\alpha)$ consists of three steps, as shown in Figure1. First, original multivariate data $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ is split into two groups such as $\boldsymbol{X}_{(1,k)} \in \mathbb{R}^{n_1 \times p}$ and $\boldsymbol{X}_{(2,k)} \in \mathbb{R}^{n_2 \times p}$, respectively. where $n_1 = n - \lfloor n/\log n \rfloor$, $n_2 = \lfloor n/\log n \rfloor$ and k represents the index of the number of iterations for cross validation. For n_1 and n_2, [Bickel and Levina, 2008a] determines both n_1 and n_2 from the perspective of theory. Second, correlation matrices for both $\boldsymbol{X}_{(1,k)}$ and $\boldsymbol{X}_{(2,k)}$ are calculated as $\boldsymbol{R}_{(1,k)}$ and $\boldsymbol{R}_{(2,k)}$, respectively. Third, low-rank correlation matrix with rank d, $\boldsymbol{\Psi}_{(1,k)} = \boldsymbol{W}_{h(\alpha)} \odot \boldsymbol{Y}\boldsymbol{Y}^T$, is estimated for both $\boldsymbol{R}_{(1,k)}$ and $h(\alpha)$ based on (1) with constraint (2). Forth, for $h(\alpha)$, the procedure from the first step to the third step is repeated K times and the proposed cross-validation function is calculated as follows.

$$\mathrm{CV}(h(\alpha), d) = \frac{1}{K} \sum_{k=1}^{K} \left\| T_{h(\alpha),d}(\boldsymbol{R}_{(1,k)}) - \boldsymbol{R}_{(2,k)} \right\|_F^2 \tag{11}$$

where $T_{h(\alpha),d}(\boldsymbol{R}_{(1,k)}) = T_{h(\alpha)}(\boldsymbol{\Psi}_{(1,k)}) = \boldsymbol{W}_{h(\alpha)} \odot \boldsymbol{\Phi}_{(1,k)}$ and K is the number of iterations for the cross validation. Among the candidates of threshold values, $h(\alpha)$ is selected as it the value such that the expression in Eq.(11) is minimized. The algorithm for the cross-validation is presented in **Algorithm 2**.

Algorithm 2 Algorithm of Cross-validation for tuning proportional thresholds

Input: candidates of threshold values $\{h(\alpha)\}_\alpha$, $\boldsymbol{D}$, $d(\leq p)$ and small constant $\varepsilon > 0$
Output: $\mathrm{CV}(h(\alpha)^\dagger, d)$
Initialisation : Set $\boldsymbol{v} = (\mathrm{NA}, \mathrm{NA}, \cdots, \mathrm{NA})$ with the length of $|\{h(\alpha)\}_\alpha|$
For $i = 1$ **to** $|\{h(\alpha)\}_\alpha|$
  For $k = 1$ to K
    Split $\boldsymbol{D}$ into $\boldsymbol{D}_{(1,k)}$ and $\boldsymbol{D}_{(2,k)}$
    Calculate both $\boldsymbol{R}_{(1,k)}$ and $\boldsymbol{R}_{(2,k)}$
    Calculate $\boldsymbol{W}_{h(\alpha)}$ from corresponding $h(\alpha)$ and $\boldsymbol{R}_{(1,k)}$.
    Given $\boldsymbol{R}_{(1,k)}$ and $\boldsymbol{W}_{h(\alpha)}$, **Algorithm 1** is applied to estimate $\boldsymbol{\Psi}_{(1,k)} = \boldsymbol{W}_{h(\alpha)^\dagger} \odot \boldsymbol{Y}\boldsymbol{Y}^T$,
  End For
  calculate $\mathrm{CV}(h(\alpha), d)$
  $\boldsymbol{v}[i] \leftarrow \mathrm{CV}(h(\alpha), d)$
End For
$h(\alpha)^\dagger \leftarrow \mathrm{argmin}\{\boldsymbol{v}[1], \boldsymbol{v}[2], \cdots, \boldsymbol{v}[|\{h(\alpha)\}_\alpha|]\}$
return $h(\alpha)^\dagger$

Finally, $h(\alpha)^\dagger$, corresponding to the minimum value of Eq.(11) among the candidate threshold values is selected and $\boldsymbol{W}_{h(\alpha)^\dagger} \odot \boldsymbol{Y}\boldsymbol{Y}^T$ is estimated based on Eq.(1).

2.2 Numerical Simulation and Real example

In this section, we present a numerical simulation to evaluate the proposed approach. The numerical simulation was conducted in the same manner as that in [Cui et al., 2016]. In addition, we present a real example of applying the proposed method to a micro-array gene expression dataset from [Khan et al., 2001].

2.2.1 Simulation design of numerical simulation

In this subsection, the simulation design is presented. The framework of the numerical simulation consists of three steps. First, artificial data with true correlation matrix are generated. Second, a sparse low-rank correlation matrix is estimated

using two methods, including the proposed method. In addition, a sample correlation matrix and sparse correlation matrix based on threshold are also applied. Third, using several evaluation indices, these estimated correlation matrices are evaluated and their performances are compared.

In this simulation, three kinds of correlation models are used. Let I and J be a set of indices for the rows and columns of correlation matrices, respectively. In addition,

$$\begin{aligned} I_k =& \{i_{(k-1)20+1}, i_{(k-1)20+2}, \cdots, i_{(k-1)20+20}\} \quad \text{and} \\ J_k =& \{j_{(k-1)20+1}, j_{(k-1)20+2}, \cdots, j_{(k-1)20+20}\} \ (k = 1, 2, \cdots, 5), \end{aligned}$$

where i_k and J_k are the maximum number of indices of I_k and J_k, respectively. Using these notations, three true correlation models $\boldsymbol{R}^{(1)} = (r_{ij}^{(1)})$, $\boldsymbol{R}^{(2)} = (r_{ij}^{(2)})$, and $\boldsymbol{R}^{(3)} = (r_{ij}^{(3)})$. are set as

$$r_{ij}^{(1)} = \Big(1 - \frac{|i-j|}{10}\Big)_+ \quad (i, j = 1, 2, \cdots, p), \tag{12}$$

$$r_{ij}^{(2)} = 0.3^{|i-j|}, \quad (i, j = 1, 2, \cdots, p) \quad \text{and}, \tag{13}$$

$$r_{ij}^{(3)} = 0.6\mathbb{1}_{i=j} + 0.4\sum_{k=1}^{K} \mathbb{1}_{i\in I_k, j\in J_k} + 0.4\sum_{k=1}^{K}(\mathbb{1}_{i=i_k, j\in J_{k+1}} + \mathbb{1}_{i\in I_{k+1}, j=i_{k+1}}), \tag{14}$$

respectively. The models for(12) and (13) are called sparse models, while the modelfor (14) is called a non-sparse model by [Cui et al., 2016]. The models for(12) and (13) are used in [Bickel and Levina, 2008a], [Xue et al., 2012], and [Rothman, 2012]; for these, see Figure 2. These artificial data are generated as $\boldsymbol{x}_i \sim N(\mathbf{0}_p, \boldsymbol{R}^{(\ell)})$ $(i = 1, 2, \cdots, n;\ \ell = 1, 2, 3)$, where $\mathbf{0}_p$ is a zero vector with a length of p. In this simulation, we set $p = 100$ and the number of cross-validations $K = 5$. For the methods of sparse low-rank correlation matrix, there are $2(\text{Factor1}) \times 3(\text{Factor2}) \times 3(\text{Factor3}) \times 2(\text{Factor4 proposal and tandem}) = 36$, patterns and for the methods without low-rank approximation, there are $2(\text{Factor1}) \times 3(\text{Factor3}) \times 2(\text{Factor4 proposal and tandem}) = 12$ patterns. Simply, there are 48 patterns in this numerical simulation. In each pattern, artificial data is generated 100 times and evaluated using several indices. In addition, Both the proposed approach and tandem approach runs from a random start 50 times, and the best solution is selected. For $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$, the candidates of α are set as 0.66 to 0.86 in steps of 0.02. However, for $\boldsymbol{R}_1$, the candidates of α are set from 0.66 to 0.82 in steps of 0.02.

Next, the factors of the numerical simulation are presented. For the summary, see Table 1. Factor 1 was set to evaluate the effects of the number of subjects. If the number of subjects is smaller, the estimated sparse low-rank correlation is expected to be unstable. To evaluate the effect of rank, factor 2 was set. When a smaller rank is set, the variance between the estimated sparse low-rank correlation coefficients become larger. Therefore, it becomes easy to interpret the results, although the estimated coefficients generally tend to be far from true. Next, as explained in Eq.(12), Eq.(13) and Eq.(14), there are three levels in factor 3. Finally, in factor 4, we set four methods: the proposed approach, tandem approach, sample correlation matrix calculation, and sparse correlation matrix estimation based on threshold value [Jiang, 2013] with modifications. The purpose of both the proposed approach and the tandem approach is to estimate a sparse low-rank correlation matrix. In the tandem approach, estimation of the low-rank correlation matrix is the same as that of the proposed approach; however, the determination way of proportional threshold is different from Eq.(11). In contrast to Eq.(11), the proportional threshold is selected using the method of [Bickel and Levina, 2008a] [Jiang, 2013] with modification, which does not consider the features of the low-rank correlation matrix. Here, in the tandem approach, we use (10) as a threshold function. Therefore, in the tandem approach, given the corresponding $\boldsymbol{W}_{h(\alpha)}$ and correlation matrix $\boldsymbol{R}$, a sparse low-rank correlation matrix is estimated based on the optimization problem of Eq.(1). To estimate the sparse correlation matrix without dimensional reduction, (10) is used as the threshold function: although $r_{ij} \cdot \mathbb{1}_{h(\alpha)}[|r_{ij}| \geq h(\alpha)]$ was used in [Jiang, 2013]. In this paper, we follow the approach without dimensional reduction, Jiang (2013), with modifications.

Table 1: Factors of numerical simulation

Factor name	Levels	Description
Factor 1: The number of subjects	2	$n = 50,\ 75$
Factor 2: Rank	3	$d = 2,\ 3,\ 5$
Factor 3: True correlation model	3	Eq.(12), Eq.(13) and Eq.(14)
Factor 4: Methods	4	proposed approach, tandem approach sample correlation, and Jiang (2013) with modification

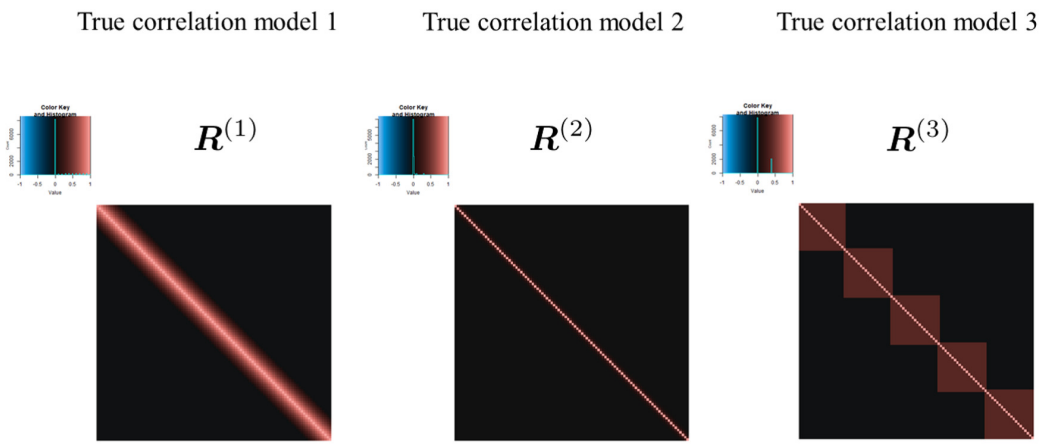

Figure 2: True correlation models

Next, in the same manner as the approach pursued in [Cui et al., 2016], we adopt four evaluation indices. To evaluate the fitting between estimated sparse low-rank correlation matrix and true correlation matrix, the average relative error of the Frobenius norm (F-norm) and of a spectral norm (S-norm) are adopted as follows:

$$\text{F-norm}(\hat{\boldsymbol{R}}) = \frac{\|\hat{\boldsymbol{R}} - \boldsymbol{R}^{(\ell)}\|_F}{\|\boldsymbol{R}^{(\ell)}\|_F} \quad \text{and} \tag{15}$$

$$\text{S-norm}(\hat{\boldsymbol{R}}) = \frac{\|\hat{\boldsymbol{R}} - \boldsymbol{R}^{(\ell)}\|_S}{\|\boldsymbol{R}^{(\ell)}\|_S}, \tag{16}$$

respectively, where $\|\cdot\|_S$ indicates spectral norm, $\hat{\boldsymbol{R}}$ is an estimator of the sparse low-rank correlation matrix, and $\boldsymbol{R}^{(\ell)}$ $(\ell = 1, 2, 3)$ is the true correlation matrix corresponding to Eq.(12), Eq.(13) and Eq.(14), respectively . In addition, to evaluate the results on sparseness, the true positive rate (TPR) and false-positive rate (FPR) are defined as follows:

$$\text{TPR} = \frac{|\{(i,j)\ \hat{r}_{ij} \neq 0,\ r_{ij}^{(\ell)} \neq 0\}|}{|\{(i,j)|\ r_{ij}^{(\ell)} \neq 0\}|} \quad \text{and} \tag{17}$$

$$\text{FPR} = \frac{|\{(i,j)\ \hat{r}_{ij} \neq 0,\ r_{ij}^{(\ell)} = 0\}|}{|\{(i,j)|\ r_{ij}^{(\ell)} = 0\}|} \tag{18}$$

where $|\cdot|$ indicates the cardinality of a set, $\hat{\boldsymbol{R}} = (\hat{r}_{ij})$, and $\boldsymbol{R}^{(\ell)} = (r_{ij}^{(\ell)})$ $(\ell = 1, 2, 3)$.

2.2.2 Application of microarray gene expression dataset

Here, we present the results of applying both the proposed approach and the tandem approach to the microarray gene expression dataset in [Khan et al., 2001]. The purpose of this real application is to evaluate the differences between two clusters of genes on the results of estimating sparse low-rank correlation matrices.

In [Rothman et al., 2009], the same dataset was used as an application of their method. Specifically, the dataset provided by the R package “MADE4" [Culhane et al., 2005] is used in this example. The dataset includes 64 training sample and 306 genes. In addition, there are four types of small round blue cell tumors of childhood (SRBCT), such as neuroblastoma (NB), rhabdomyosarcoma(RMS), Burkitt lymphoma, a subset of non-Hodgkin lymphoma (BL) and the Ewing family of tumors (EWS). Simply, there are four sample classes in this dataset. As was done in [Rothman et al., 2009], these genes are classified into two clusters: “informative" and “noninformative,” where genes belonging to “informative" have information to discriminate four classes and those belonging to “noninformative" do not.

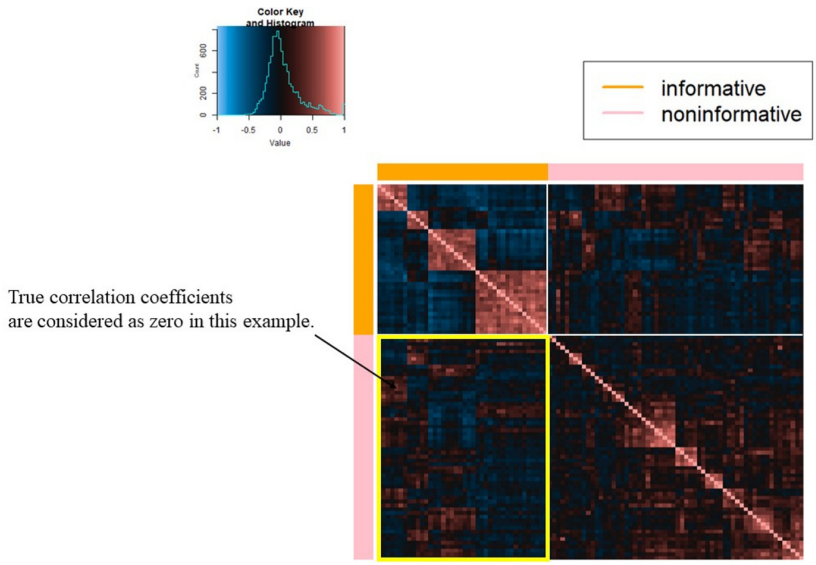

Figure 3: Sample correlation matrix among selected 100 genes

Next, to construct “informative" cluster and “noninformative" cluster, F statistics is calculated for each gene as follows:

$$F_j = \frac{(G-1)^{-1}\sum_{g=1}^{G} n_g(\bar{x}_{gj} - \bar{x}_j)^2}{(n-G)^{-1}\sum_{g=1}^{G}(n_g - 1)s_{gj}^2} \quad (j = 1, 2, \cdots, 306)$$

where G indicates the number of classes, such as NB, RMS, BL, and EWS, n_g is the number of subjects belonging to class g, and $\bar{x}_{gj}$ is mean of class g for gene j, and $\bar{x}_j$ is the mean of gene j, s_{gj} is the sample variance of the class g for gene j. Here, if F_j is relatively higher, gene j is considered as “informative" because the corresponding j tends to include information such that each class is discriminated. From the calculated F_j, the top 40 genes and bottom 60 genes are set as “informative" cluster and “noninformative" cluster, respectively. Then, the correlation matrix for 100 genes was calculated and set as input data. See Figure 3.

To compare the results of the proposed approach with those of the tandem approach, the FPR is calculated. For the tandem approach, see “2.2.1 Simulation design of numerical simulation". In this application, true correlations between gene belonging to “informative" cluster and gene belonging to “noninformative" cluster is considered as 0. Therefore, the denominator of the FPR is set to $2 \times 40 \times 60 = 4800$. For TPR, it is difficult to determine the true structure because correlations within each cluster are not necessarily non-zero. For the rank, we set 2, 3 and 5. The candidates of α for determining the threshold value are set as 0.50 to 0.83 in steps of 0.01 for both approaches, and these algorithms start from 50 different initial parameters. In addition, as was done for the numerical simulation, both the sample correlation matrix and Jiang (2013) with modifications, are also employed.

3 Results

In this section, we present the results of the numerical simulation and real application.

3.1 Simulation result

In this subsection, we present the simulation results by the true correlation models. Table 2, Table 3 and Table 4 indicate the FPRs and TPRs for applying $\boldsymbol{R}^{(1)}$, $\boldsymbol{R}^{(2)}$, and $\boldsymbol{R}^{(3)}$, respectively. Each cell indicates mean of these indices. Here, $\boldsymbol{R}^{(2)}$ is a non-sparse correlation matrix and therefore, FPR cannot be calculated, and both the TPR and FPR of the sample correlation matrix cannot be calculated because the sample correlation is not a sparse matrix. From the results of the numerical simulation, the FPRs of the proposed approach were the lowest among those of all the methods in all situations. while the TPRs of the proposed approach tended to be inferior to those of the other approaches. Simply, the proposed approach makes it a sparser low-rank correlation matrix compared to the tandem approach when a smaller rank is used.

For the relative error of $\mathrm{F\text{-}norm}$, Figure 4, Figure 5 and Figure 6 indicates the results of applying these methods to $\boldsymbol{R}^{(1)}$, $\boldsymbol{R}^{(2)}$, and $\boldsymbol{R}^{(3)}$, respectively. Hence, the median of the proposed approach was lower than that of the tandem approach for each pattern. In addition, the interquartile range of the proposed approach was smaller than that of the

Table 2: Results of FPRs and TPRs for $\boldsymbol{R}^{(1)}$ and each value indicate the mean.

p | n	Methods	d	FPR	TPR
100 | 50	Sample correlation matrix		–	–
	Jiang (2013) with modification		0.07	**0.82**
	Tandem	2	0.07	**0.82**
	Proposal		**0.02**	0.76
	Tandem	3	0.07	**0.82**
	Proposal		0.03	0.78
	Tandem	5	0.08	**0.82**
	Proposal		0.06	0.81
100 | 75	Sample correlation matrix		–	–
	Jiang (2013) with modification		0.06	**0.85**
	Tandem	2	0.06	**0.85**
	Proposal		**0.01**	0.79
	Tandem	3	0.07	**0.85**
	Proposal		0.02	0.81
	Tandem	5	0.07	**0.85**
	Proposal		0.05	0.84

Table 3: Results of FPRs and TPRs for $\boldsymbol{R}^{(2)}$ and each value indicate the mean.

p | n	Methods	d	FPR	TPR
100 | 50	Sample correlation matrix		–	–
	Jiang (2013) with modification		–	**0.17**
	Tandem	2	–	**0.17**
	Proposal		–	0.15
	Tandem	3	–	**0.17**
	Proposal		–	0.15
	Tandem	5	–	0.16
	Proposal		–	0.15
100 | 75	Sample correlation matrix		–	–
	Jiang (2013) with modification		–	**0.17**
	Tandem	2	–	**0.17**
	Proposal		–	0.15
	Tandem	3	–	**0.17**
	Proposal		–	0.15
	Tandem	5	–	**0.17**
	Proposal		–	0.15

Table 4: Results of FPRs and TPRs for $\boldsymbol{R}^{(3)}$ and each value indicate the mean.

p	n	Methods	d	FPR	TPR
100	50	Sample correlation matrix		–	–
		Jiang (2013) with modification		0.08	0.93
		Tandem	2	0.08	0.93
		Proposal		**0.02**	0.83
		Tandem	3	0.08	0.93
		Proposal		**0.02**	0.84
		Tandem	5	0.09	**0.94**
		Proposal		0.03	0.86
100	75	Sample correlation matrix		–	–
		Jiang (2013) with modification		0.07	**0.97**
		Tandem	2	0.07	**0.97**
		Proposal		**0.01**	0.86
		Tandem	3	0.07	**0.97**
		Proposal		**0.01**	0.87
		Tandem	5	0.07	**0.97**
		Proposal		0.02	0.90

tandem approach in each pattern. Therefore, we confirmed that the results of the proposed approach are effective and stable compared to those of the tandem approach. As rank is set as larger, The results of both the proposed approaches become lower and close to those of Jiang (2013) with modifications in all situations. Among these methods, the result of Jiang (2013) with modifications is the best for the relative error of $rmFmathchar` - norm$. However, it is natural things from the properties of low-rank approximation. As was done for $\mathrm{F\text{-}norm}$, those of $\mathrm{S\text{-}norm}$ for $\boldsymbol{R}^{(1)}$, $\boldsymbol{R}^{(2)}$ and $\boldsymbol{R}^{(3)}$ are shown in Figures 7, 8 and 9, respectively. The tendency of the results for $\mathrm{S\text{-}norm}$ is quite similar to that for $\mathrm{S\text{-}norm}$. From the results for $\mathrm{F\text{-}norm}$, we observe that the result of the proposed approach with rank 5 is quite close to that of Jiang (2013) with modification.

For the estimated correlation matrices, Figure 10, Figure 11 and Figure 12 correspond to true correlation model 1, true correlation model 2, and true correlation model 3 with $n = 50$, respectively. As the same way, Figure 13, Figure 14 and Figure 15 correspond to true correlation model 1, true correlation model 2, and true correlation model 3 with $n = 75$, respectively. From Figures 10, 13, 12 and 15, we found that the estimated correlation matrices of the proposed approach can estimate zero correctly compared to those of the tandem approach. In addition, rank is set larger, estimated correlation matrices tend to be close to the results of Jiang (2013) with modifications.

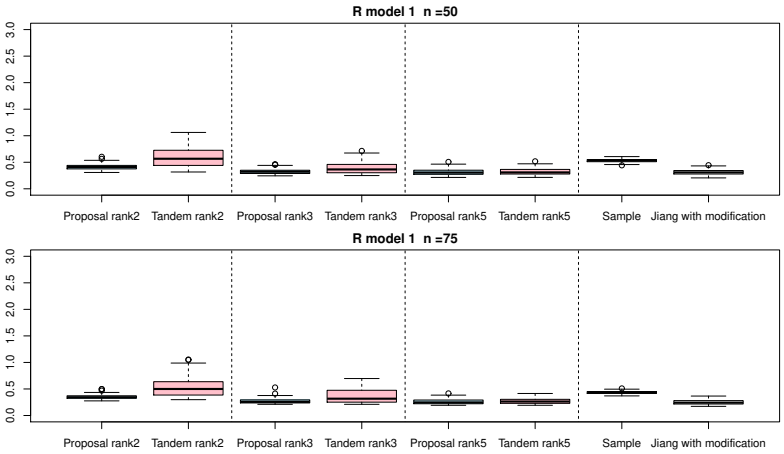

Figure 4: Relative errors of F-norm for $\boldsymbol{R}^{(1)}$ with $n = 50$ and $n = 75$; the vertical axis indicate the results of relative errors of F-norm.

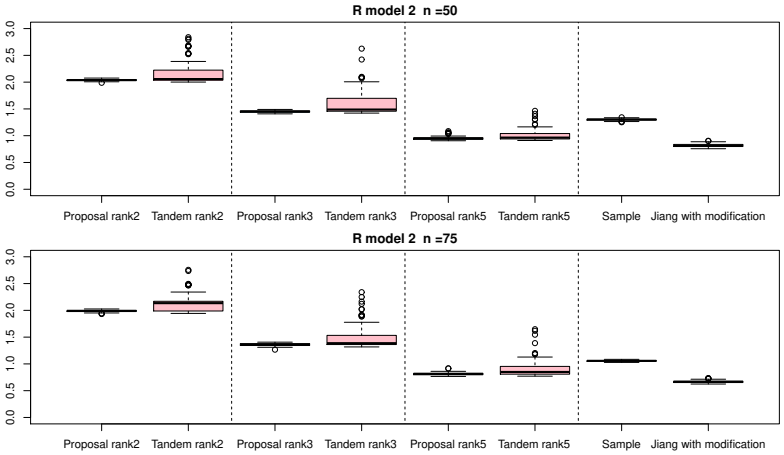

Figure 5: Relative errors of F-norm for $\boldsymbol{R}^{(2)}$ with $n = 50$ and $n = 75$; the vertical axis indicate the results of relative errors of F-norm.

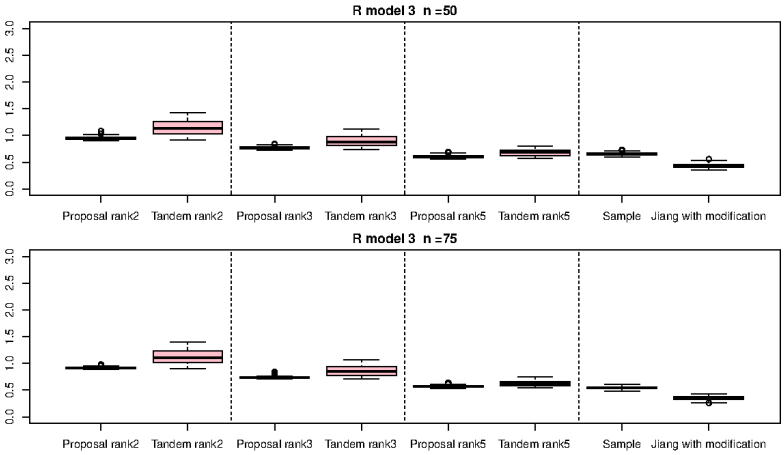

Figure 6: Relative errors of F-norm for $\boldsymbol{R}^{(3)}$ with $n = 50$ and $n = 75$; the vertical axis indicate the results of relative errors of F-norm.

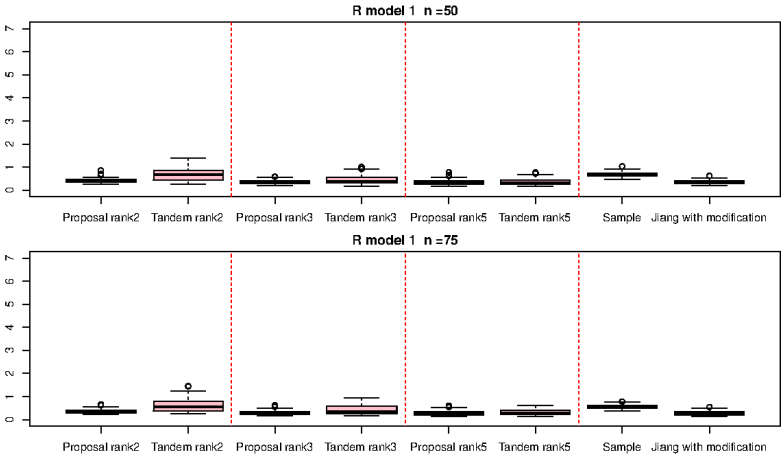

Figure 7: Relative errors of S-norm for $\boldsymbol{R}^{(1)}$ with $n = 50$ and $n = 75$; the vertical axis indicate the results of relative errors of S-norm.

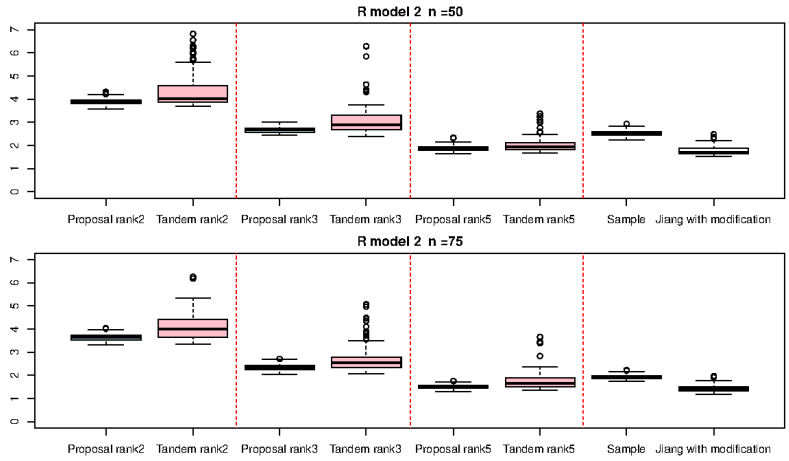

Figure 8: Relative errors of S-norm for $\boldsymbol{R}^{(2)}$ with $n = 50$ and $n = 75$; the vertical axis indicate the results of relative errors of S-norm.

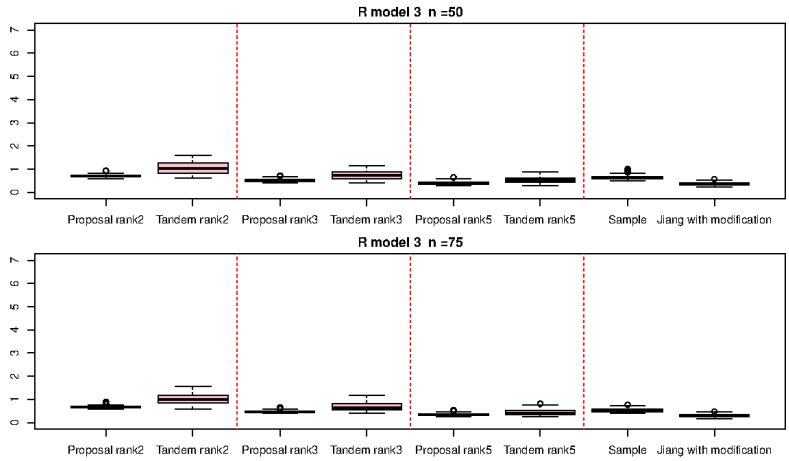

Figure 9: Relative errors of S-norm for $\boldsymbol{R}^{(3)}$ with $n = 50$ and $n = 75$; the vertical axis indicate the results of relative errors of S-norm.

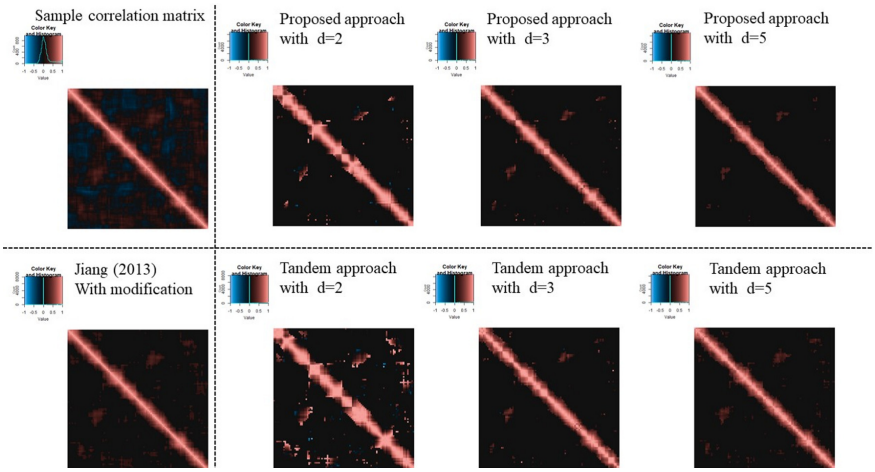

Figure 10: Examples of estimated correlation matrices for true correlation model 1 ($n = 50$)

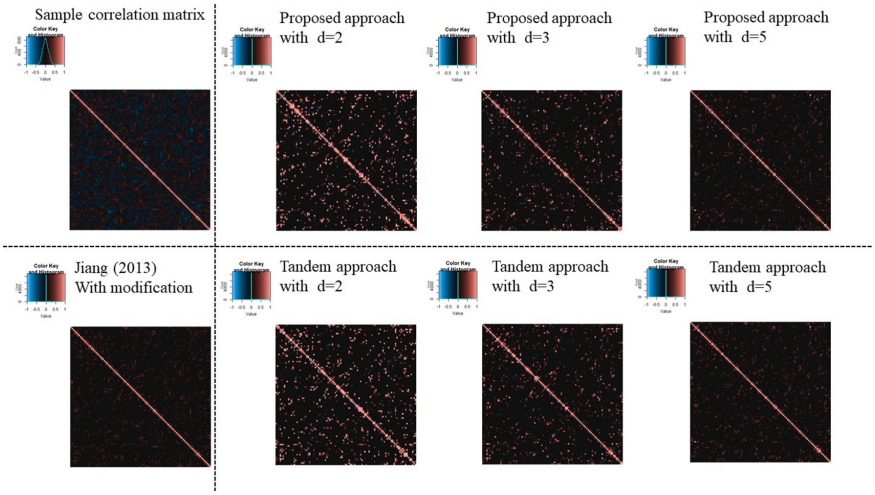

Figure 11: Examples of estimated correlation matrices for the true correlation model 2 ($n = 50$)

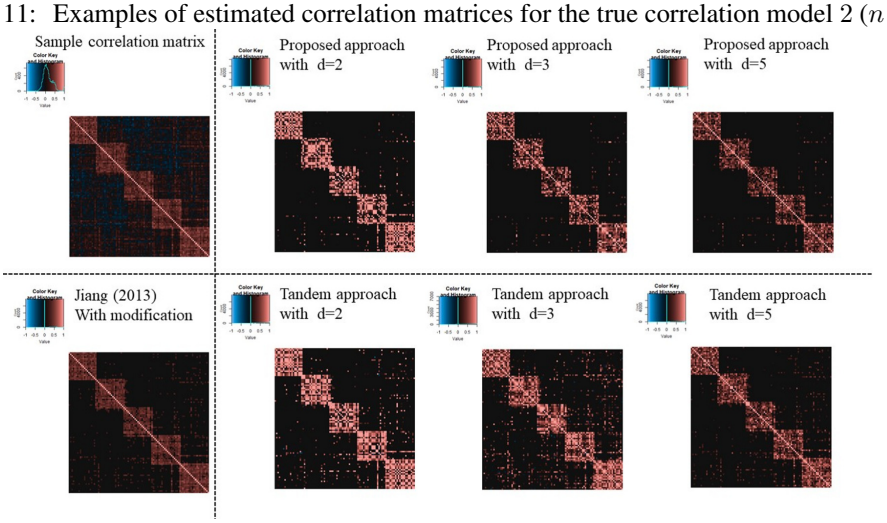

Figure 12: Examples of estimated correlation matrices for the true correlation model 3 ($n = 50$)

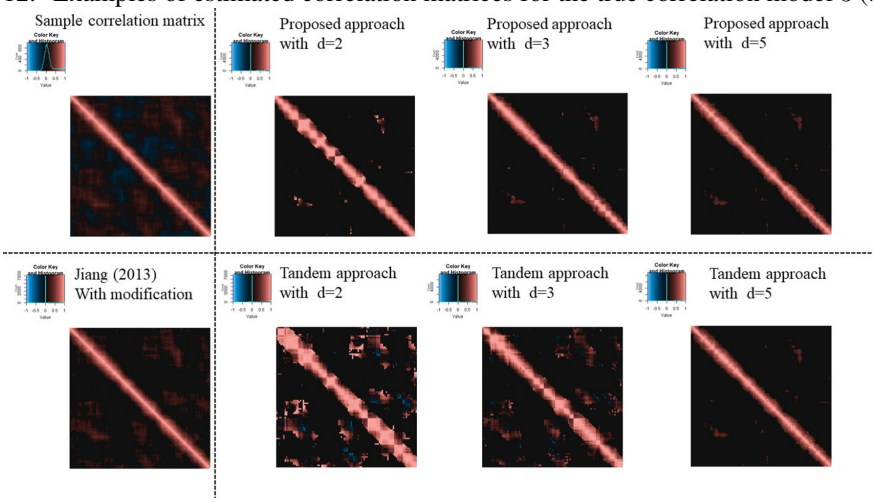

Figure 13: Examples of estimated correlation matrices for the true correlation model 1 ($n = 75$)

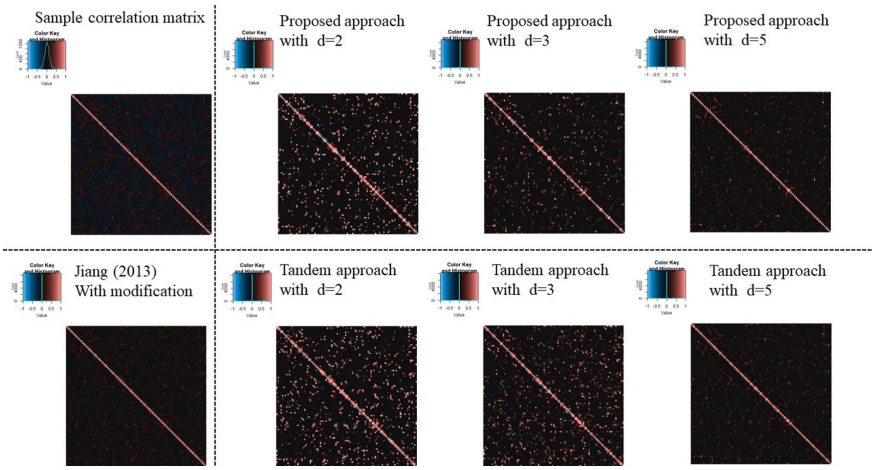

Figure 14: Examples of estimated correlation matrices for the true correlation model 2 ($n = 75$)

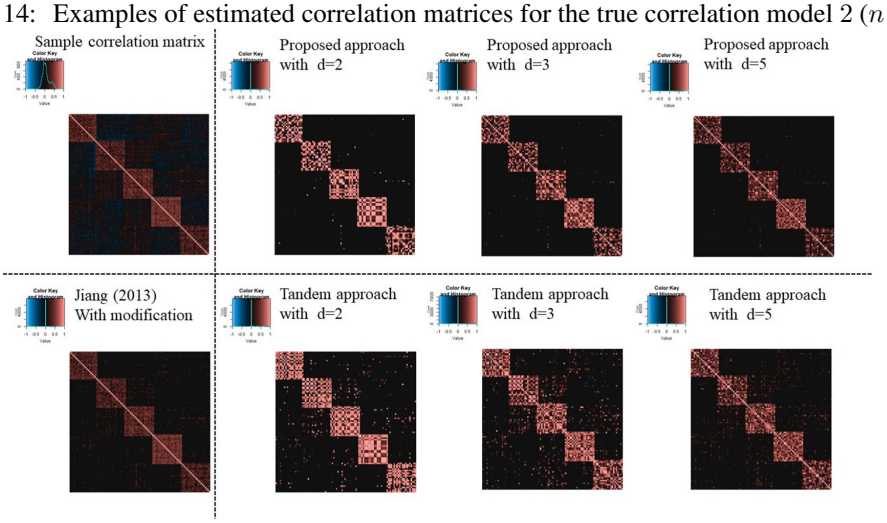

Figure 15: Examples of estimated correlation matrices for true correlation model 3 ($n = 75$)

3.2 Result of application of microarray gene expression dataset

In this subsection, the results of application of microarray gene expression dataset is shown. For the estimated original correlation matrix, Jiang (2013) with modification, the proposed approach and tandem approach, see Figure 16. Hence, the percentage points of $d = 2,\ 3$, and 5 in the proposed approach were $\alpha = 0.82,\ 0.81$, and 0.75, respectively, while the percentage points in the tandem approach and Jiang (2013) with modification were both $\alpha = 0.65$. The estimated results of Jiang (2013) with modification are as presened in Figure 16. However, FPRs were higher than those of the proposed approach. Here, the FPR is not affected by the choice of rank in the tandem approach. From these results, the estimated sparse low-rank correlation matrix tends to be sparser when the rank is set as lower. In fact, it can be confirm that in Figure 16. In addition, as the rank is set larger, the estimated correlations of the proposed approach become similar to those of the tandem approach. We also confirmed that the estimated sparse low-rank correlation matrix between genes belonging to “informative" cluster tend to be similar to the results obtained in Rothman et al. [2009] using the heatmap.

Next, Table 5 shows the results of the FPR of the proposed approach, tandem approach, and Jiang (2013) with modifications. Hence, the FPRs in the proposed method with $d = 2,\ 3$, and 5 were all lower than those of both the tandem approach and Jiang (2013) with modifications. This tendency was observed regarding the results of numerical simulations.

Table 5: FPR of application of microarray gene expression dataset

Method	d	FPR
Proposal	2	**0.128**
Proposal	3	0.139
Proposal	5	0.197
Tandem	–	0.285
Jiang (2013) with modification	–	0.285

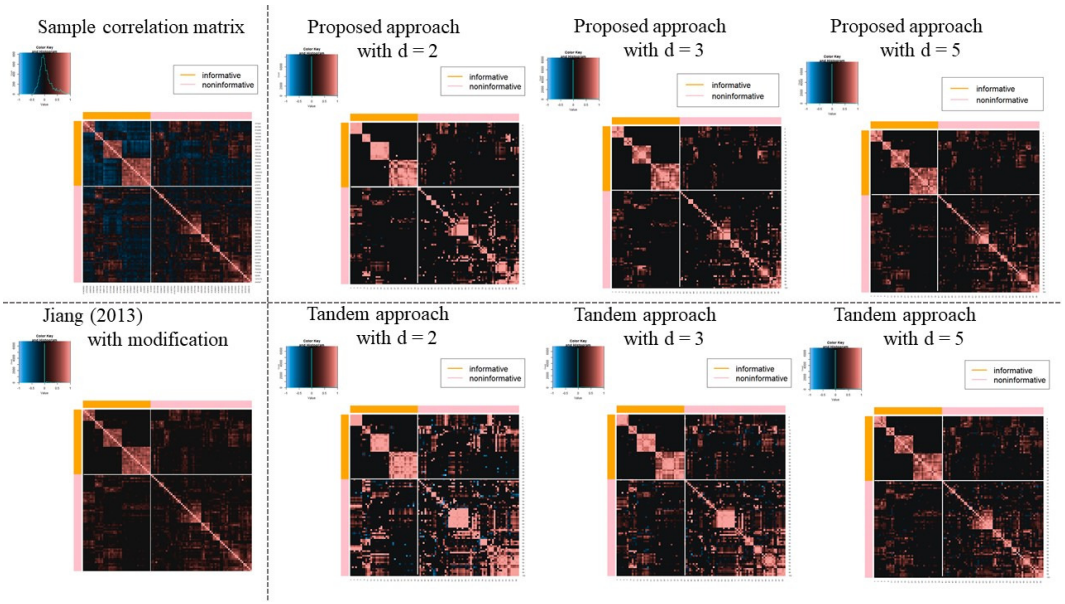

Figure 16: Estimated sparse low-rank correlation matrices with $d = 2,\ 3$ and 5, sample correlation matrix and sparse correlation matrix without rank reduction.

4 Conclusion

In this paper, we proposed a novel estimation method for sparse low-rank correlations based on the MM algorithm. The proposed approach can overcome the problem of estimating low-rank correlation matrix. Low rank approximation is a very powerful tool, and the approach provides us with an easy interpretation of the feature because the contrast of the estimated coefficients becomes larger. However, the estimation sometimes lead to misunderstanding. Simply, even if the true correlation coefficient is zero, the corresponding estimated coefficient of the low-rank approximation without sparse estimation may be greater than zero. For efficiency, we confirmed the advantages via a numerical simulation and real example. In fact, in the real example of microarray gene expression dataset, the FPR of the proposed approach with $d = 2,\ 3$, and 5 were 0.128, 0.139, and 0.197, respectively. although those of the tandem approach and Jiang (2013) with modifications were 0.285 and 0.285, respectively. Therefore, we confirmed that the FPRs of the proposed approach are the best, irrespective of the rank. In the same manner, from the numerical simulation, we confirmed that the FPRs of the proposed approach are superior to those of the tandem approach and Jiang (2013) with modifications.

References

L. Wilkinson and M. Friendly. The history of the cluster heat map. *The American Statistician*, 36:179–184, 2009.

X. Wei and K.C. Li. Exploring the within- and between-class correlation distributions for tumor classification. *Proceedings of the National Academy of Sciences of the United States of America*, 107:6737–6742, 2010.

J.M.F. ten Berge. Least squares optimization in multivariate analysis. *Keiden:DSWO Press*, 1993.

R. Pietersz and P. Groenen. Rank reduction of correlation matrices by majorization. *Quant.Finance*, 4(6):649–662, 2004.

D. Simon and J. Abell. A majorization algorithm for constrained approximation. *Linear Algebra and its Application*, 432:1152–1164, 2010.

I. Grubii and R. Pietersz. Efficient rank reduction of correlation matrices. *Linear Algebra and its Application*, 422: 629–653, 2007.

X.F. Duan, J.C. Bai, J.F. Li, and J.J. Peng. On the low rank solution of the q-weighted nearest correlation matrix problem. *Numerical linear algebra with applications*, 23:340–355, 2016.

C. Ding and X. He. k-means clustering via principle component analysis. *In Proceedings of the International Conference on Machine Learning (ICML)*, 29, 2004.

J. Engel, L. Buydens, and L. Blanchet. An overview of large-dimensional covariance and precision matrix estimators with applications in chemometrics. *Journal of chemometrics*, 31(4):e2880, 2017.

C. Lam. High-dimensional covariance matrix estimation. *Wiley Interdisciplinary reviews: computational statistics*, 12(2):e1485, 2020.

J. Bien and R.J. Tibshirani. Sparse estimation of a covariance matrix. *Biometrika*, 99:807–820, 2011.

A. Rothman. Positive definite estimators of large covariance matrices. *Biometrika*, 99:733–740, 2012.

L. Xue, S. Ma, and H. Zou. Positive definite l_1 penalized estimation of large covariance matrices. *Journal of the American Statistical Association*, 107:1480–1491, 2012.

T. Cai, W. Liu, and X. Luo. A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106:594–607, 2011.

A. D'aspremont, O. Banerjee, and L.E. Ghaoui. First-order methods for sparse covariance selection. *SIAM Journal on Matrix Analysis and Applications*, 30:56–66, 2008.

J. Friedman, T. Hastie, and R Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.

A. Rothman, P.J. Bickel, E. Levina, and J. Zhu. Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics*, 2:495–515, 2008.

M. Yuan and Y. Lin. Model selection and estimation in gaussian graphical model. *Biometrika*, 94(1):19–35, 2007.

Y. Cui, C. Leng, and D. Sun. Sparse estimation of high-dimensional correlation matrices. *Computational Statistics and Data Analysis*, 93:390–403, 2016.

P.J. Bickel and E. Levina. Covariance regularization by thresholding. *The Annals of statistics*, 36(6):2577–2604, 2008a.

T. Cai and W. Liu. Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607, 2011.

P.J. Bickel and E. Levina. Regularized estimation of large covariance matrices. *The Annals of Statistics*, 36(1):199–227, 2008b.

N. El Karoui. Operator norm consistent estimation of large-dimensional sparse covariance matrices. *The Annals of Statistics*, 36(6):2717–2756, 2008.

B. Jiang. Covariance selection by thresholding the sample correlation matrix. *Statistics and probability letters*, 83: 2492–2498, 2013.

A. Rothman, E. Levina, and J. Zhu. Generalized thresholding of large covariance matrices. *Journal of the American Statistical Association*, 104:177–186, 2009.

C. Lam and J. Fan. Sparsistency and rates of convergence in large covariance matrices estimation. *The Annals of Statistics*, 37:4254–4278, 2009.

H.L. Liu, L. Wang, and T. Zhao. Sparse covariance matrix estimation with eigenvalue constraints. *Journal of Computational and Graphical Statistics*, 37:439–459, 2014.

S.H. Zhou, N.H. Xiu, Z.Y. Luo, and L.C. Kong. Sparse and low-rank covariance matrix estimation. *Journal of the operation research society of China*, 3:231–250, 2015.

P.A. Savalle, E. Richard, and N. Vayatis. Estimation of simultaneously sparse and low rank matrices. *In International Conference on Machine Learning (ICML)*, 2012.

D.K. Knol and J.M.F. Ten Berge. Least-squares approximation of an improper correlation matrix by a proper one. *Psychometrika*, 54(1):53–61, 2012.

D.R. Hunter and K. Lange. A tutorial on mm algorithm. *The American Statistician*, 58(1):30–37, 2004.

M.P. van den Heuvel, S.C. de Lange, A. Zalesky, C. Seguin, B.T. Thomas Yeo, and R. Schmidt. Proportional thresholding in resting-state fmri functional connectivity networks and consequences for patient-control connectome studies: Issues and recommendations. *NeuroImage*, 152:437–449, 2017.

J. Khan, J.S. Wei, Saal L.H. Ringner, M., M. Ladanyi, F. Westermann, F. Berthold, M. Schwab, C.R. Antonescu, and P.S. Peterson, C. Meltzer. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nature medicine*, 7:673–679, 2001.

A.C. Culhane, J. Thioulouse, G. Perriere, and D.G. Higgins. Made4: an r package for multivariate analysis of gene expression data. *Bioinformatics*, 21(11):2789–2790, 2005.