Article

Thresholding Approach for Low-Rank Correlation Matrix Based on MM Algorithm

Kensuke Tanioka [1,*] , Yuki Furotani [2] and Satoru Hiwa [1]

1 Department of Biomedical Sciences and Informatics, Doshisha University, Kyoto 610-0394, Japan; shiwa@mail.doshisha.ac.jp
2 Graduate School of Life and Medical Sciences, Doshisha University, Kyoto 610-0394, Japan; yfurotani@mis.doshisha.ac.jp
* Correspondence: ktanioka@mail.doshisha.ac.jp

Abstract: Background: Low-rank approximation is used to interpret the features of a correlation matrix using visualization tools; however, a low-rank approximation may result in an estimation that is far from zero, even if the corresponding original value is zero. In such a case, the results lead to a misinterpretation. Methods: To overcome this, we propose a novel approach to estimate a sparse low-rank correlation matrix based on threshold values. We introduce a new cross-validation function to tune the corresponding threshold values. To calculate the value of a function, the MM algorithm is used to estimate the sparse low-rank correlation matrix, and a grid search was performed to select the threshold values. Results: Through numerical simulation, we found that the false positive rate (FPR), interpretability, 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 the FPR of the tandem approach was 0.285. Conclusions: We propose a novel approach to estimate sparse low-rank correlation matrices. The advantage of the proposed method is that it provides results that are interpretable using a heatmap, thereby avoiding result misinterpretations. We demonstrated the superiority of the proposed method through both numerical simulations and real examples.

Keywords: cross-validation; proportional threshold; sparse estimation

1. Background

It is essential in applications to compute the correlation matrix and present it with a heatmap [1] to understand the relationship between variables or subjects. However, when the sample size is small or contains noise, the correlation matrix may be challenging to interpret regarding the relationship between variables. Here, a correlation matrix that is easy to interpret is referred to as one that satisfies the following two properties: (a) the correlation coefficients between variables that are related are high and those that are not zero or close to zero, and (b) the number of correlation coefficients to be interpreted is small. In this paper, we deal with the problem of estimating such an easily interpretable correlation matrix.

A low-rank approximation address problem (a) [2]. Various methods have been proposed to estimate low-rank correlation matrices [3–6], and there are several advantages associated with this estimation. As one of these advantages, low-rank approximations can also effectively describe the clustering structure [7], which results in improved interpretations. These specific relations between variables are emphasized from the property, and therefore problem (a) can be solved.

While the low-rank correlation matrix has the above advantages, (i) the number of correlation coefficients to be interpreted becomes problematic when the number of variables is large, making interpretation difficult, and (ii) according to the Eckart–Young–Mirsky

theorem [8], the error between the low-rank correlation matrix and the original correlation matrix becomes large, and even if the true correlation coefficient is 0, it can be estimated as a value far from 0.

To overcome this, in this study, we proposed a new approach to estimate sparse low-rank correlation matrices. The proposed approach, combined with a heatmap, provides a visual interpretation of the relationships between the variables. There are two types of techniques available for the sparse methods of the correlation matrix and the covariance matrix [9,10]. The first involves adding a sparsity penalty to the objective functions [11–19]. The other type uses thresholding values to achieve a sparse structure. Bickel and Levina (2008) proposed the thresholding matrix estimator [20] and various related methods have been developed [21–24]. In addition, to estimate the sparse correlation matrix, Refs. [25–27] used generalized thresholding operator-based methods [25]. For the estimation of sparse low-rank matrices, methods based on penalty terms have also been proposed [28,29]. In addition to that, there is another approach to estimate the covariance matrix based on the modified Cholesky decomposition (MCD) [30]. This covariance matrix estimation certainly has beneficial properties when the number of variables is high. However, the estimate depends on the order of variables [31]. To tackle the problem, several methods have been proposed [32–35].

The proposed approach adopts an approach that uses hard thresholding based [20,24] therefore, it is easy to use and provides interpretable results. We describe why we aimed to estimate a low-rank sparse correlation matrix rather than a low-rank sparse covariance matrix. The covariance matrix depends on the original multivariate data scale. Since the threshold corresponding to all covariances cannot be treated with a single value, multiple thresholds must be selected by cross-validation. In other words, estimating a low-rank covariance matrix requires multiple threshold-based approaches, which takes an enormous amount of time. On the other hand, estimating a low-rank correlation matrix can be implemented with a single-threshold approach because the scale is uniform, and can be computed relatively quickly. Furthermore, when estimating the covariance matrix, the variance of each variable must also be estimated, which increases the number of parameters to be estimated compared to the estimation of the correlation matrix. Based on the above, this method focuses on the correlation matrix from the feasibility viewpoint and develops a method for estimating sparse low-rank correlation matrices. For more on this discussion, see Refs. [14,36].

The summary of the proposed method is as follows: for small sample sizes and noisy data, a sparse low-rank correlation matrix can be used to emphasize the correlation coefficients between related variables. In addition, the inclusion of sparse constraints eliminates the problem of misunderstanding and facilitates interpretation. Furthermore, the direct estimation of the correlation matrix instead of the covariance matrix reduces the computation time comparatively. We introduce a new cross-validation function to estimate sparse low-rank correlation matrices modifying those used in Refs. [20,24]. This cross-validation function is the mean difference between the low-rank correlation matrix and the original-scale correlation matrix. Compared to the cross-validation function of Refs. [20,24], the value of the proposed cross-validation function tends to be higher when the rank is set to lower. The proposed cross-validation function is expected to choose threshold values corresponding to a more sparse correlation matrix from the property. To calculate the values of the cross-validation function, the majorize-minimization algorithm (MM algorithm) [3,4] and the hard thresholding approach are used. The proposed method has two advantages; first, the estimated sparse low-rank correlation matrix allows an easy and visual interpretation of the correlation matrix using a heatmap and avoids a misinterpretation of the correlation matrix. The proposed approach can estimate the low-rank correlation matrix, and more sparse and specific relations between variables can be emphasized. If the true correlation coefficient is zero, the proposed method estimates the corresponding coefficient as zero. In addition, we focus only on positive correlation

coefficients, not negative correlation coefficients. With the focus on only positive relations, it becomes easy to interpret the features of the relations.

The rest of this paper is structured as follows. We explain the model and algorithm in Section 2. Section 3 evaluates the proposed approach and describes the numerical simulation. The results of applying the proposed method to real data are provided in Section 3. Finally, we present our conclusions in Section 4.

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, an estimation of a low-rank correlation matrix is described based on the MM algorithm [3,4]. Next, the hard thresholding operator and the proposed cross-validation function are described to achieve the sparse low-rank correlation structure.

2.1.1. Optimization Problem of Low-Rank Correlation Matrices

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

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

subject to

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

where $Y = (y_1, y_2, \cdots, y_p)^T$, $y_j = (y_{j1}, y_{j2}, \cdots, y_{jd})^T$, $y_{jo} \in \mathbb{R}$ $(j = 1,2,\cdots,p; \; o = 1,2,\cdots,d)$ is the coordinate matrix of variables on dimensions d, $\odot$ is the Hadamard product, $\| \cdot \|_F$ is the Frobenius norm, and $\| \cdot \|$ is L_2 norm. The objective function in Equation (1) was explained by Ref. [37]. From the constraint (2), $Y^T Y$ becomes the correlation matrix.

When Y is estimated, W is fixed. In this situation, Y is calculated based on the MM algorithm. The estimation is described in Section 2.1.2. We introduced a modified cross-validation function to determine W and chose W in Section 2.1.3.

2.1.2. Estimation of Low-Rank Correlation Matrices Based on MM Algorithm

The MM algorithm for estimating a low-rank correlation matrix proposed by Ref. [4] is explained (Algorithm 1). To estimate Y in the closed-form under constraint (2), the quadratic optimization problem for Y must be converted to a linear optimization one. We can derive the updated formula combined with the Lagrange multiplier in the closed-form using the linear function. Let $y^{(t)} \in \mathbb{R}^d$ be the parameter of the t step of the optimization problem algorithm, and $g(y|y^{(t)})$ be a real function $g : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}$. If $g(y|y^{(t)})$ satisfy the following conditions such that

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

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

$g(y|y^{(t)})$ is defined as the majorizing function of $f(y)$ at the point $y^{(t)}$, where $f : \mathbb{R}^d \mapsto \mathbb{R}$ is the original function. Simply put, to estimate the parameters in the MM algorithm, $g(y|y^{(t)})$, not $f(y)$, should be minimized. In several situations, $g(y|y^{(t)})$ is expected to easily minimize the value. For more details on the MM algorithm, see Ref. [38].

Before deriving the majorizing function, the objective function (1) can be re-described as follows:

$$
\begin{aligned}
f(Y\mid R, W) &= \left\| R - W \odot YY^T \right\|_F^2 \\
&= \sum_{i=1}^{p}\sum_{j\neq i}^{p}(r_{ij} - w_{ij}y_i^T y_j)^2 \\
&= \sum_{i=1}^{p}\sum_{j\neq i}^{p} r_{ij}^2 + \sum_{i=1}^{p} y_i^T \left(\sum_{j\neq i} w_{ij}y_j y_j^T \right) y_i - 2\sum_{i=1}^{p} y_i^T \left(\sum_{j\neq i} w_{ij}r_{ij}y_j \right) \\
&= \sum_{i=1}^{p}\sum_{j\neq i}^{p} r_{ij}^2 + \sum_{i=1}^{p} y_i^T B_i y_i - 2\sum_{i=1}^{p} y_i^T \left(\sum_{j\neq i} w_{ij}r_{ij}y_j \right),
\end{aligned}
\tag{5}
$$

where $B_i = \sum_{j\neq i} w_{ij}y_j y_j^T$. Here, the parameter estimation of Y is conducted by y_i. The corresponding part of Equation (5) and the majorizing function can be described as follows:

$$
f_i(y_i\mid \{y_j\}_{j\neq i}) = y_i^T B_i y_i - 2y_i^T \left(\sum_{j\neq i} w_{ij}r_{ij}y_j \right)
\tag{6}
$$

$$
\leq -y_i^{(t-1)T} B_i y_i^{(t-1)} + 2\lambda_i - 2y_i^T(\lambda_i I_d - B_i)y_i^{(t-1)} - 2y_i^T \left(\sum_{j\neq i} w_{ij}r_{ij}y_j \right)
\tag{7}
$$

$$
= g(y_i\mid y_i^{(t-1)}, \{y_j\}_{j\neq i}),
$$

where $g(y_i\mid y_i^{(t-1)}, \{y_j\}_{j\neq i})$ represents the majorizing function of Equation (5), I_d is $d \times d$ identity matrix, λ_i is the maximum eigenvalue of B_i, and $y_i^{(t-1)}$ is y_i of $(t-1)$ step in the algorithm. Here, the inequality of Equation (7) is satisfied because $B_i - \lambda_i I_d$ is semi-definitely negative. In fact, if $y_i = y_i^{(t-1)}$, Equations (6) and (7) become equal.

Using the Lagrange multiplier method and Equation (7), the updated formula of y_i is derived as follows:

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

Algorithm 1 Algorithm for estimating the low-rank correlation matrix

---

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

    Initialisation: Set $Y^{(0)}$ satisfying $\|y_j^{(0)}\| = 1$ for all j and set $t \leftarrow 1$

1: **while** $f(Y^{(t-1)}) - f(Y^{(t)}) \geq \varepsilon$ **do**
2:    **for** $i = 1$ to p
3:       Calculate $B_i^{(t)} \leftarrow \sum_{j\neq i} w_{ij}y_j^{(t)}y_j^{(t)T}$
4:       Calculate λ_i to be the largest eigenvalue of B_i
5:       Update y_i based on Equation (8).
6:    **End for**
7:    $t \leftarrow t + 1$
8: **End while**
9: **return** $Y^{(t)}$ with the constraint $\|y_j^{(t)}\| = 1$ for all j

---

2.1.3. Proposed Algorithm of Cross-Validation to Determine Hard Thresholds

We adopt hard thresholding to estimate the sparse low-rank correlation matrix in the proposed approach. To determine threshold values, we introduce a cross-validation func-

tion based on Ref. [20]. The purpose of this approach is quite simple; that is, to determine the threshold values related to sparse estimation by considering the corresponding rank.

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

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

where

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

For example, the proportional threshold operator is used in the domain of neural science [39]. Let $\boldsymbol{W}_{h(\alpha),\boldsymbol{R}} = (w_{ij}^{(h(\alpha))}) = (\mathbb{1}_{h(\alpha)}[r_{ij} \geq h(\alpha)]) \in \{0, 1\}^{p \times p}$ for correlation matrix $\boldsymbol{R}$. By using $\boldsymbol{W}_{h(\alpha),\boldsymbol{R}}$, Equation (9) can be described as follows:

$$T_{h(\alpha)}(\boldsymbol{R}) = \boldsymbol{W}_{h(\alpha),\boldsymbol{R}} \odot \boldsymbol{R}. \tag{11}$$

Here, Equation (10) is modified for the original function of Ref. [20]. Initially, $\mathbb{1}_{h(\alpha)}[|r_{ij}| \geq h(\alpha)]$ was used; however, we focused only on higher correlation coefficients and not on negative correlation coefficients. Using modifications, it becomes easy to interpret the results.

To estimate a sparse low-rank correlation matrix, we introduce the modified proportional threshold operator based on Equation (9) because the interpretation of the proportional threshold is quite simple. Given an α percentile, rank d, and the correlation matrix $\boldsymbol{R}$, the modified threshold operator is defined as follows;

$$T_{h(\alpha),d}(\boldsymbol{R}) = \boldsymbol{W}_{h(\alpha),\boldsymbol{R}} \odot \boldsymbol{Y}\boldsymbol{Y}^{T} \tag{12}$$

where $\boldsymbol{Y}$ is the estimated correlation matrix with rank d, such as minimizing $f(\boldsymbol{Y} \mid \boldsymbol{R}, \boldsymbol{W}_{h(\alpha),\boldsymbol{R}})$. Equation (12) is different from Equation (11) using a low-rank correlation matrix, although $\boldsymbol{W}_{h(\alpha),\boldsymbol{R}}$ is calculated from the original correlation matrix, not from a low-rank correlation matrix. For the choice of the threshold value $h(\alpha)$, cross-validation was introduced (e.g., Refs. [20,24]). The cross-validation procedure for estimating $h(\alpha)$ consists of three steps, as shown in Figure 1. First, the original multivariate data $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ is split into two groups randomly, such as $\boldsymbol{X}_{(1,k)} \in \mathbb{R}^{n_1 \times p}$ and $\boldsymbol{X}_{(2,k)} \in \mathbb{R}^{n_2 \times p}$, where $n_1 = n - \lfloor n/\log n \rfloor$, $n_2 = \lfloor n/\log n \rfloor$, and k represent the index of the number of iterations for cross-validation, and $\lfloor \cdot \rfloor$ represents floor function. For n_1 and n_2, Ref. [20] determines both n_1 and n_2 from the point of view of theory. Second, the 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, the correlation matrix with rank d, $\boldsymbol{Y}$ is estimated, such as minimizing $f(\boldsymbol{Y} \mid \boldsymbol{R}_{(1,k)}, \boldsymbol{W}_{h(\alpha),\boldsymbol{R}_{(1,k)}})$ with constraint (2). Fourth, for fixed $h(\alpha)$, the procedure from the first step to the third step is repeated K times and the proposed cross-validation function is calculated as follows.

$$\mathrm{CV}(h(\alpha), d) = \frac{1}{K} \sum_{k=1}^{K} \left\| T_{h(\alpha),d}(\boldsymbol{R}_{(1,k)}) - \boldsymbol{R}_{(2,k)} \right\|_{F}^{2} \tag{13}$$

where K is the number of iterations for the cross-validation. Among the candidates for threshold values, $h(\alpha)$ is selected as the value, such that the expression in Equation (13) 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 $\{h(\alpha)\}_\alpha$, $\boldsymbol{X}$ and $d (\leq p)$
Output: $\mathrm{CV}(h(\alpha)^\dagger, d)$
    Initialisation: Set $\boldsymbol{v} = (\mathrm{NA}, \mathrm{NA}, \cdots, \mathrm{NA})$ with the length of $|\{h(\alpha)\}_\alpha|$
 1: **For** $i = 1$ **to** $|\{h(\alpha)\}_\alpha|$
 2:    **For** $k = 1$ to K
 3:       Split $\boldsymbol{X}$ into $\boldsymbol{X}_{(1,k)}$ and $\boldsymbol{X}_{(2,k)}$
 4:       Calculate both $\boldsymbol{R}_{(1,k)}$ and $\boldsymbol{R}_{(2,k)}$
 5:       Calculate $\boldsymbol{W}_{h(\alpha),\boldsymbol{R}_{(1,k)}}$ from corresponding $h(\alpha)$ and $\boldsymbol{R}_{(1,k)}$.
 6:       Given $\boldsymbol{R}_{(1,k)}$ and $\boldsymbol{W}_{h(\alpha),\boldsymbol{R}_{(1,k)}}$, **Algorithm 1** is applied to estimate $\boldsymbol{Y}$,
 7:    **End For**
 8:    calculate $\mathrm{CV}(h(\alpha), d)$
 9:    $\boldsymbol{v}[i] \leftarrow \mathrm{CV}(h(\alpha), d)$
10: **End For**
11: $h(\alpha)^\dagger \leftarrow \mathrm{argmin}\{\boldsymbol{v}[1], \boldsymbol{v}[2], \cdots, \boldsymbol{v}[|\{h(\alpha)\}_\alpha|]\}$
12: **return** $h(\alpha)^\dagger$

Finally, $h(\alpha)^\dagger$, corresponding to the minimum value of Equation (13) among the candidate threshold values, is selected, and $\boldsymbol{W}_{h(\alpha)^\dagger,\boldsymbol{R}} \odot \boldsymbol{Y}\boldsymbol{Y}^T$ is estimated based on Equation (1).
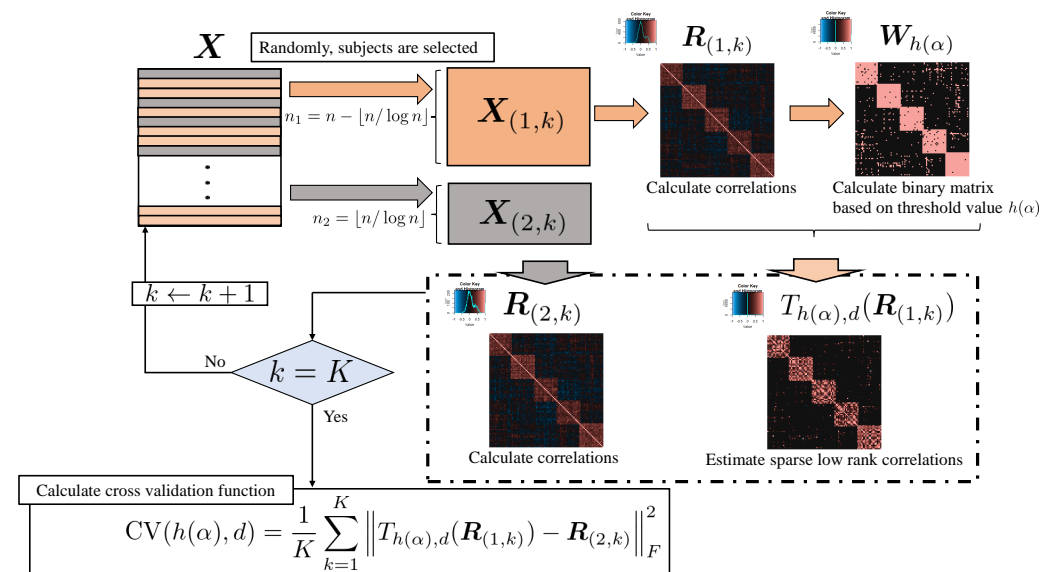

Figure 1. The framework of the proposed cross-validation.

2.2. Numerical Simulation and Real Example

This section presents a numerical simulation to evaluate the proposed approach. The numerical simulation was conducted based on Ref. [19], with some modifications. Practically, the size of the numerical data is matched to that of the real-data example in Section 2.2.2. In addition, we present a real example of applying the proposed method to a microarray gene expression data set from Ref. [40].

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 a true correlation matrix are generated. Second, sparse low-rank correlation matrices are estimated using two methods, including the proposed method. In addition, a sample correlation matrix and sparse correlation matrix based on the threshold also apply. 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 the correlation matrices, respectively. In addition, I_k and J_k are defined as follows:

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

where i_o and j_o $(o = 1, 2, \cdots, 100)$ indicate the number of rows and columns, respectively. Using this notation, 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

$$r_{ij}^{(1)} = \left(1 - \frac{|i-j|}{10}\right)_+ \quad (i, j = 1, 2, \cdots, p), \tag{14}$$

$$r_{ij}^{(2)} = 0.3^{|i-j|}, \quad (i, j = 1, 2, \cdots, p) \quad \text{and}, \tag{15}$$

$$r_{ij}^{(3)} = 0.6\mathbb{1}_{i=j} + 0.4\sum_{k=1}^{5}\mathbb{1}_{i\in I_k, j\in J_k} + 0.4\sum_{k=1}^{5}(\mathbb{1}_{i=i_k, j\in J_{k+1}} + \mathbb{1}_{i\in I_{k+1}, j=i_{k+1}}), \tag{16}$$

respectively, where i_k and j_k are the maximum number of indices of I_k and J_k, respectively, and $\mathbb{1}$ represents the indicator function. The models for Equations (14) and (15) are called sparse models, while the model for Equation (16) is called a non-sparse model by Ref. [19]. The models for Equations (14) and (15) are used in Refs. [12,13,20]; for these, see Figure 2. These artificial data are generated as $\boldsymbol{x}_i \sim N(\boldsymbol{0}_p, \boldsymbol{R}^{(\ell)})$ $(i = 1, 2, \cdots, n; \ \ell = 1, 2, 3)$, where $\boldsymbol{0}_p$ is a zero vector with a length of p. In this simulation, we set $p = 100$ and the number of cross-validations $K = 5$. In this simulation, there are several types of scenarios. For the number of scenarios in the estimation of sparse low-rank correlation matrices, there are 2 (setting 1) × 3 (setting 2) × 3 (setting 3) × 2 (setting 4: proposal and tandem) = 36 patterns. In addition to that, for the number of scenarios in the estimation of sparse correlation matrices without low-rank approximation, there are 2 (setting 1) × 3 (setting 3) × 3 (setting 4: sample correlation, Jiang (2013) with modification and graphical lasso [41]) = 18 patterns. Simply, there are 54 patterns in this numerical simulation. In each pattern, artificial data are generated 100 times and evaluated using several indices. Given $\boldsymbol{W}$ and rank d, the result of the estimation $\boldsymbol{Y}$ depends on the initial parameters $\boldsymbol{Y}^{(0)}$ in both the proposed approach and the tandem approach. Therefore, the low-rank matrix is estimated from 50 randomly generated initial values, and the solution with the smallest value of the objective function is adopted. For $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$, the candidates of α are set from 0.66 to 0.86 in steps of 0.02. However, for $\boldsymbol{R}^{(3)}$, the candidates of α are set from 0.66 to 0.82 in steps of 0.02. From the point of view of computation time, we set the range of threshold values. The threshold range is different in the case of $\boldsymbol{R}^{(3)}$ because above 0.84 it is too sparse to compute the solution of the low-rank matrix.

Next, the settings of the numerical simulation are presented. For the summary, see Table 1. Setting one 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, setting two was set. The variance between the estimated sparse low-rank correlation coefficients becomes larger when a smaller rank is set. Therefore, it becomes easy to interpret the results. Next, as explained in Equations (14)–(16), there are three levels in setting three.
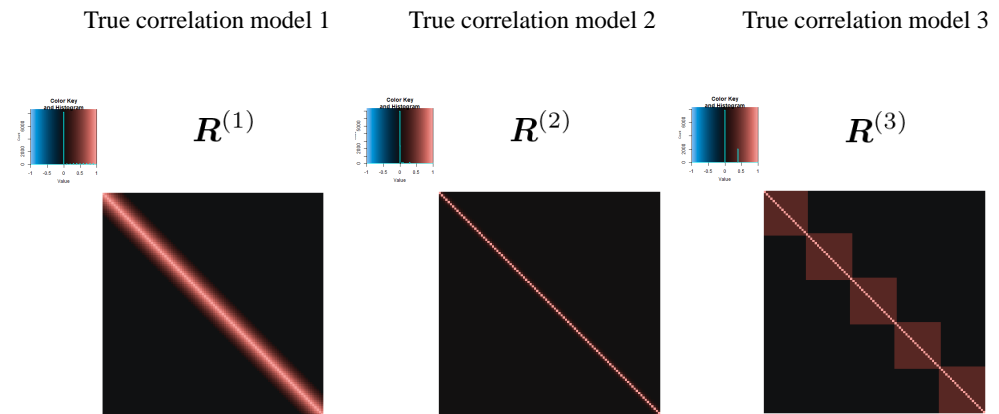
True correlation model 1   True correlation model 2   True correlation model 3

Figure 2. True correlation models.

Table 1. Settings of numerical simulation.

Setting Name	Levels	Description
Setting 1: The number of subjects	2	$n = 50,\ 75$
Setting 2: Rank	3	$d = 2,\ 3,\ 5$
Setting 3: True correlation model	3	Equations (14)–(16)
Setting 4: Methods	5	proposed approach, tandem approach, sample correlation, Jiang (2013) with modification, and graphical lasso

Finally, in setting four, we selected five methods: the proposed approach, tandem approach, sample correlation matrix calculation, and the sparse correlation matrix estimation based on threshold value [24] with modifications and graphical lasso [41]. The tandem approach was included in the comparison method to compare the performance of the proposed method, which considers low ranks when computing the cross-validation function, and the tandem approach, which combines existing methods. Jiang (2013) with modifications was included in the comparison method to compare the performance of the proposed method with the case where the low-rank approximation is not considered. Graphical lasso is included as a comparison method because it is often used to infer relationships between variables. However, since the method does not directly estimate the sparse correlation matrix, the correlation matrix is calculated by estimating the inverse of the sparse covariance matrix and then computing the inverse of the covariance matrix. Finally, the sample correlation matrix results are also presented for reference. Next, we explain the compared methods. The purpose of both the proposed and tandem approaches is to estimate a sparse low-rank correlation matrix. In the tandem procedure, there are two steps. The threshold value $h(\alpha)$ is determined based on the following cross-validation function in the first step.

$$\frac{1}{K}\sum_{k=1}^{K}\left\|\boldsymbol{W}_{h(\alpha),\boldsymbol{R}_{(1,k)}}\odot \boldsymbol{R}_{(1,k)} - \boldsymbol{R}_{(2,k)}\right\|^2 \tag{17}$$

Equation (17) is a modification of the cross-validation function in Ref. [24]. The modification part is $\boldsymbol{W}_{h(\alpha),\boldsymbol{R}} = (w_{ij}^{(h(\alpha))}) = (\mathbb{1}_{h(\alpha)}[r_{ij} \geq h(\alpha)]) \in \{0,1\}^{p\times p}$, although $\boldsymbol{W}_{h(\alpha),\boldsymbol{R}} = (w_{ij}^{(h(\alpha))}) = (\mathbb{1}_{h(\alpha)}[|r_{ij}| \geq h(\alpha)]) \in \{0,1\}^{p\times p}$ in Ref. [24]. In the second step, using $h(\alpha)$ in the first step, the low rank correlation matrix is estimated based on Ref. [4]. In short, given $h(\alpha)$ and $\boldsymbol{R}$, $\boldsymbol{Y}$ is estimated as follows:

$$\|\boldsymbol{R} - \boldsymbol{W}_{h(\alpha),\boldsymbol{R}}\odot \boldsymbol{Y}\boldsymbol{Y}^T\|_F^2 \to \text{Min}$$

with the constraint $\|y_j\| = 1(j = 1, 2, \cdots, p)$. To estimate the sparse correlation matrix without dimensional reduction, there are two methods in this simulation. First, Jiang (2013) with modifications is explained, and this approach is also two steps. The first step is the same procedure as the first step in the tandem approach, and the threshold value $h(\alpha)$ is determined. In the second step, by using $h(\alpha)$ in the first step, the sparse correlation matrix is calculated as $W_{h(\alpha),R} \odot R$. For the modification, Equation (10) is used as the threshold function (although $r_{ij} \cdot \mathbb{1}_{h(\alpha)}[|r_{ij}| \geq h(\alpha)]$ was used in Ref. [24]). The second approach is graphical lasso [41]. To select the tuning parameters, the `CVglasso` package is used [42] in `R`. The correlation matrix is then calculated from the estimated sparse inverse covariance matrix.

Likewise, as in the approach pursued in Ref. [19], we adopt four evaluation indices. To evaluate the fitting between the estimated sparse low-rank correlation matrix and the true correlation matrix, the average relative errors of the Frobenius norm (F-norm) and the spectral norm (S-norm) are adopted as follows:

$$\text{F-norm}(\hat{R}) = \frac{\|\hat{R} - R^{(\ell)}\|_F}{\|R^{(\ell)}\|_F} \qquad \text{and} \tag{18}$$

$$\text{S-norm}(\hat{R}) = \frac{\|\hat{R} - R^{(\ell)}\|_S}{\|R^{(\ell)}\|_S}, \tag{19}$$

where $\|\cdot\|_S$ indicates the spectral norm, $\hat{R}$ is an estimator of the sparse low-rank correlation matrix, and $R^{(\ell)}$ ($\ell = 1, 2, 3$) is the true correlation matrix corresponding to Equation (14), Equation (15), and Equation (16), respectively. In addition, to evaluate the results on sparseness, the true positive rate (TPR) and the false positive rate (FPR) are defined as follows:

$$\text{TPR} = \frac{|\{(i,j)| \ \hat{r}_{ij} \neq 0, \ r_{ij}^{(\ell)} \neq 0\}|}{|\{(i,j)| \ r_{ij}^{(\ell)} \neq 0\}|} \qquad \text{and} \tag{20}$$

$$\text{FPR} = \frac{|\{(i,j)| \ \hat{r}_{ij} \neq 0, \ r_{ij}^{(\ell)} = 0\}|}{|\{(i,j)| \ r_{ij}^{(\ell)} = 0\}|} \tag{21}$$

where $|\cdot|$ indicates the cardinality of a set, $\hat{R} = (\hat{r}_{ij})$, and $R^{(\ell)} = (r_{ij}^{(\ell)})$ ($\ell = 1, 2, 3$). In addition to that, to evaluate the interpretability, we adopt the following index.

$$\text{Sparsity} = \frac{|\{(i,j)| \ \hat{r}_{ij} = 0\}|}{p \times p - p}. \tag{22}$$

Equation (22) ranges from 0 to 1. If Equation (22) is close to 1, the estimated correlation matrix is very sparse, otherwise, that is not sparse.

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 Ref. [40]. This real application aims to evaluate the differences between two classes of genes in the results of estimating sparse low-rank correlation matrices. Concretely, in this real example, true correlation coefficients between classes are assumed to be zero, and the FPR of the proposed approach is compared with that of the tandem approach.

In Ref. [25], the same dataset was used to apply their method. Specifically, the data set provided by the R package "MADE4" [43] is used in this example. The data set includes the 64 training samples and the 306 genes. In addition, there are four types of small round blue cell tumors of childhood (SRBCT), such as neuroblastoma (NB), rhabdomyosarcoma(RMS), and 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 in Ref. [25],

these genes are classified into two classes: "informative" and "non-informative", where genes belonging to "informative" have information to discriminate four classes and those belonging to "non-informative" do not.

Next, to construct the "informative" class and the "non-informative" class, the F statistics are 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, $\bar{x}_{gj}$ is the mean of class g for gene j, $\bar{x}_j$ is the mean of gene j, and s_{gj} is the sample variance of 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 and bottom 60 genes are set as being in the "informative" and "non-informative" classes, respectively. The correlation matrix for 100 genes was then calculated and used as input data. See Figure 3.

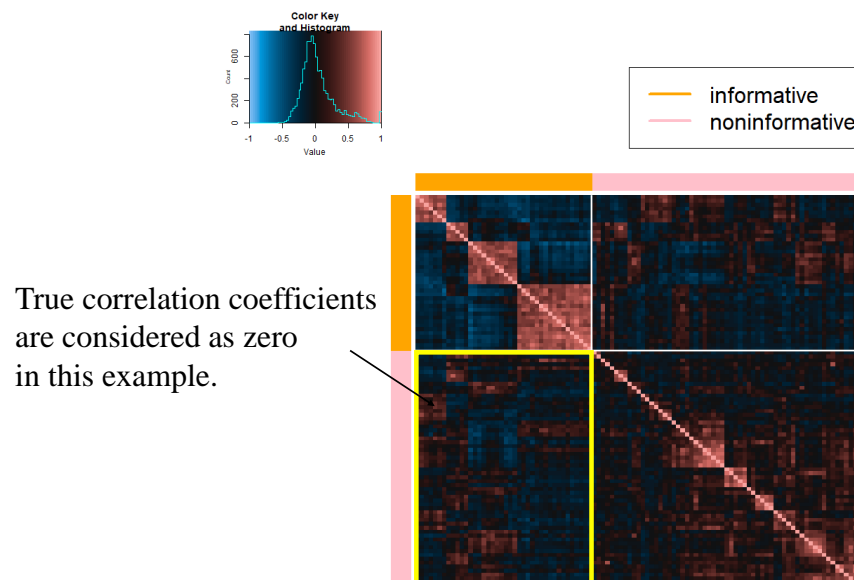True correlation coefficients are considered as zero in this example.

Figure 3. Sample correlation matrix among 100 selected genes.

To compare the results of the proposed approach with those of the tandem approach, the FPR is calculated. For the tandem approach, see Section 2.2.1. In this application, true correlations between genes belonging to the "informative" class and genes belonging to the "non-informative" class are considered to be 0. Therefore, the FPR denominator is $2 \times 40 \times 60 = 4800$. For TPR, it is difficult to determine the true structure because correlations within each class are not necessarily non-zero. In addition to that, to evaluate the interpretability of the estimated correlation matrix, we adopt the following index:

$$\text{within class Sparsity} = \frac{|\{\hat{r}_{ij}| \left((i, j \in \text{IS}) \vee (i, j \in \text{NS}) \right), \hat{r}_{ij} = 0 \}|}{|\{(i, j)| \ i, j \in \text{IS}\}| + |\{(i, j)| \ i, j \in \text{NS}\}|} \quad (23)$$

where IS and NS are a set of genes belonging to the "informative" and "non-informative" classes, respectively. If Equation (23) is larger, correlation within the same class is considered sparse. For rank, we set 2, 3, and 5. The candidates of α for determining the threshold value are set from 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 performed for the numerical

simulation, the sample correlation matrix, Jiang (2013) with modifications, and graphical lasso are also employed.

3. Results

This section presents 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. Tables 2–4 indicate the FPRs, TPRs, and Sparsity for applying $R^{(1)}$, $R^{(2)}$, and $R^{(3)}$, respectively. Each cell indicates the mean of these indices. Here, $R^{(2)}$ 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 was the lowest among those of all the methods in all the 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. In addition to that, the result of sparsity in the proposed approach was higher than that in the other methods. From the results, we found that the proposed method provides us with interpretable results compared to the tandem method. For the result of the graphical lasso, these correlation matrices are estimated as non-sparse, although these estimated inverse covariance matrices are estimated as sparse. Therefore, both TPRs and FPRs were estimated as higher, and the sparsity in the graphical lasso tends to be 0.00 without the case of $R^{(2)}$.

Table 2. Results of FPRs and TPRs for $R^{(1)}$. Each value indicates the mean.

$p \mid n$	Methods	d	FPR	TPR	Sparsity
100 | 50	Sample correlation matrix		—	—	—
	Jiang (2013) with modifications		0.07	**0.82**	0.79
	Graphical lasso		1.00	1.00	0.00
	Tandem	2	0.07	**0.82**	0.79
	Proposal		**0.02**	0.76	0.85
	Tandem	3	0.07	**0.82**	0.80
	Proposal		0.03	0.78	0.84
	Tandem	5	0.08	**0.82**	0.79
	Proposal		0.06	0.81	0.81
100 | 75	Sample correlation matrix		—	—	—
	Jiang (2013) with modifications		0.06	**0.85**	0.80
	Graphical lasso		1.00	1.00	0.00
	Tandem	2	0.06	**0.85**	0.80
	Proposal		**0.01**	0.79	0.85
	Tandem	3	0.07	**0.85**	0.79
	Proposal		0.02	0.81	0.84
	Tandem	5	0.07	**0.85**	0.79
	Proposal		0.05	0.84	0.81

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

$p \mid n$	Methods	d	FPR	TPR	Sparsity
100 | 50	Sample correlation matrix	—	—	—	—
	Jiang (2013) with modifications	—	**0.17**	0.84
	Graphical lasso	—	0.47	0.52
	Tandem	2	—	**0.17**	0.84
	Proposal		—	0.15	0.86
	Tandem	3	—	**0.17**	0.83
	Proposal		—	0.15	0.86
	Tandem	5	—	0.16	0.84
	Proposal		—	0.15	0.85
100 | 75	Sample correlation matrix	—	—	—	—
	Jiang (2013) with modifications	—	**0.17**	0.84
	Graphical lasso	—	0.97	0.02
	Tandem	2	—	**0.17**	0.84
	Proposal		—	0.15	0.86
	Tandem	3	—	**0.17**	0.83
	Proposal		—	0.15	0.86
	Tandem	5	—	**0.17**	0.83
	Proposal		—	0.15	0.85

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

$p \mid n$	Methods	d	FPR	TPR	Sparsity
100 | 50	Sample correlation matrix	—	—	—	—
	Jiang (2013) with modifications		0.08	0.93	0.74
	Graphical lasso		1.00	1.00	0.00
	Tandem	2	0.08	0.93	0.74
	Proposal		**0.02**	0.83	0.81
	Tandem	3	0.08	0.93	0.74
	Proposal		**0.02**	0.84	0.81
	Tandem	5	0.09	**0.94**	0.73
	Proposal		0.03	0.86	0.80
100 | 75	Sample correlation matrix		—	—	—
	Jiang (2013) with modifications		0.07	**0.97**	0.73
	Graphical lasso		1.00	1.00	0.00
	Tandem	2	0.07	**0.97**	0.73
	Proposal		**0.01**	0.86	0.81
	Tandem	3	0.07	**0.97**	0.74
	Proposal		**0.01**	0.87	0.81
	Tandem	5	0.07	**0.97**	0.74
	Proposal		0.02	0.90	0.80

For the relative error of the F-norm, Figure 4–6 indicate 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. Furthermore, the interquartile range of the proposed approach was smaller than that of the 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 F-norm in the case of $\boldsymbol{R}^{(1)}$ and $\boldsymbol{R}^{(3)}$, although that of the graphical lasso (glasso) is the best in the case of $\boldsymbol{R}^{(2)}$. However, it is a natural thing from the properties of low-rank approximation. As was done for the F-norm, those of S-norm for $\boldsymbol{R}^{(1)}$, $\boldsymbol{R}^{(2)}$, and $\boldsymbol{R}^{(3)}$ are shown in Figures 7–9, respectively. The tendency of the results for S-norm is quite similar to that for F-norm. From the results for F-norm, we observe

that the result of the proposed approach with rank 5 is quite close to that of Jiang (2013) with modifications.

For the estimated correlation matrices, Figures 10–12 correspond to the true correlation model one, the true correlation model two, and the true correlation model three with $n = 50$, respectively. In the same way, Figures 13–15 correspond to true correlation model one, true correlation model two, and true correlation model three with $n = 75$, respectively. From Figures 10, 12, 13 and 15, we found that the estimated correlation matrices of the proposed approach tend to estimate zero correctly compared to those of the tandem approach. Especially, the tendency can be confirmed when the rank is set as lower, visually. In addition, rank is set larger, and the estimated correlation matrices tend to be close to the results of Jiang (2013) with modifications.
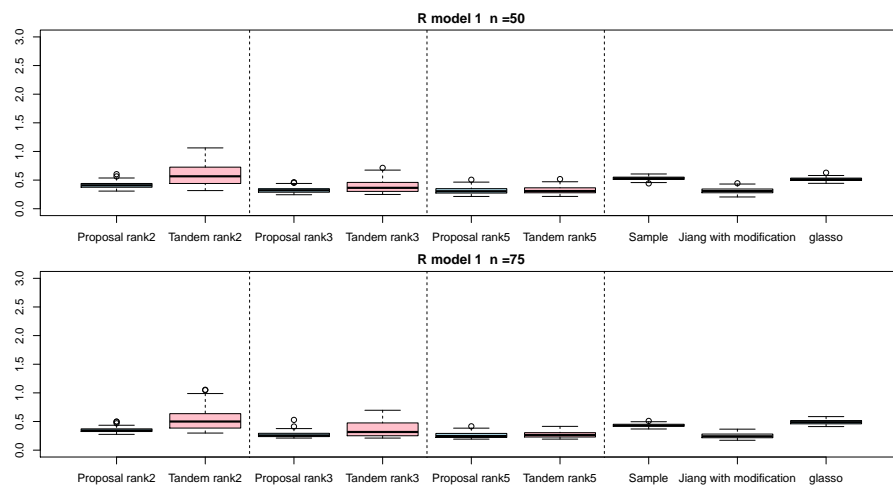
Figure 4. Relative errors of the F-norm for $\boldsymbol{R}^{(1)}$ with $n = 50$ and $n = 75$; the vertical axis indicates the results of relative errors of the F-norm.
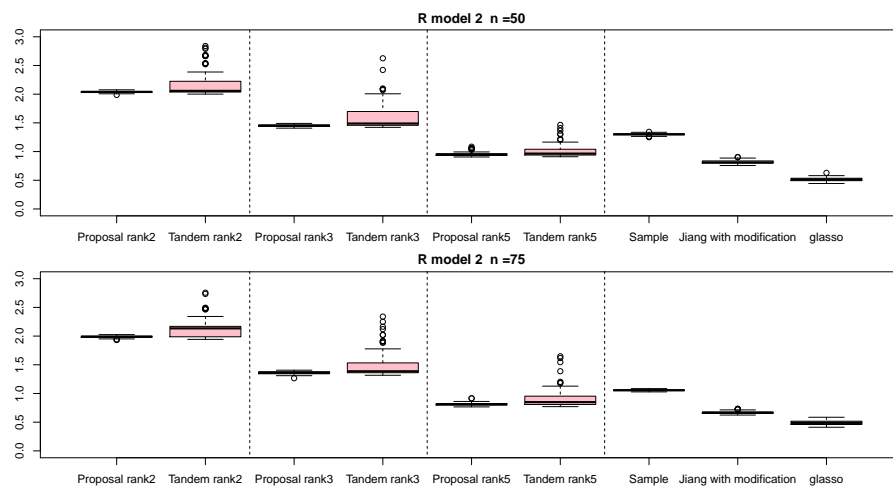
Figure 5. Relative errors of F-norm for $\boldsymbol{R}^{(2)}$ with $n = 50$ and $n = 75$; the vertical axis indicates the results of relative errors of F-norm.
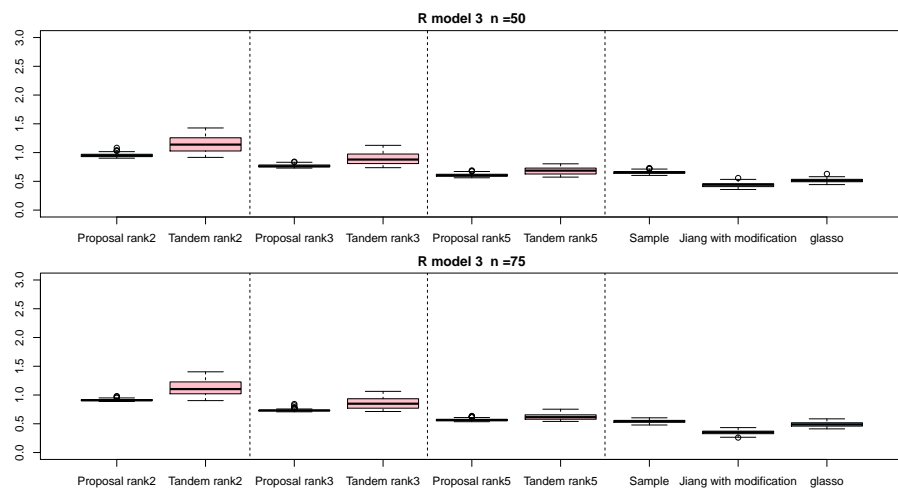
Figure 6. Relative errors of F-norm for $R^{(3)}$ with $n = 50$ and $n = 75$; the vertical axis indicates the results of relative errors of F-norm.
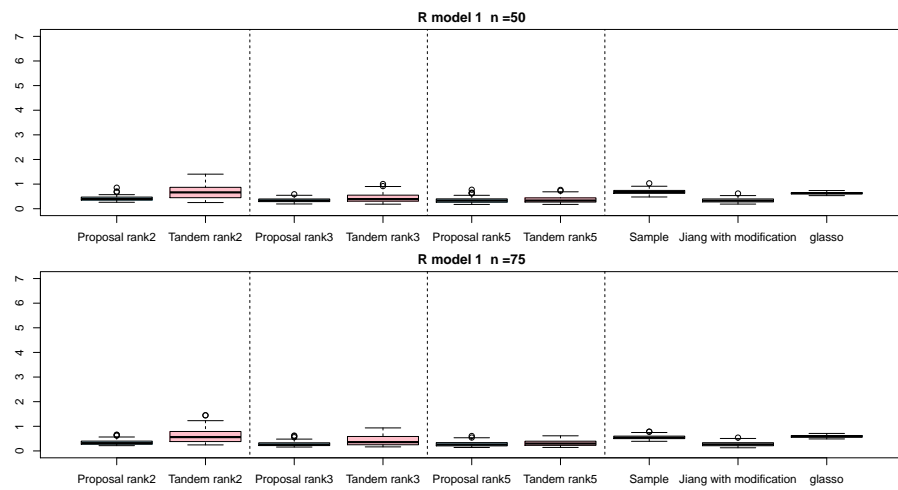

Figure 7. Relative errors of the S-norm for $R^{(1)}$ with $n = 50$ and $n = 75$; the vertical axis indicates the results of relative errors of the S-norm.
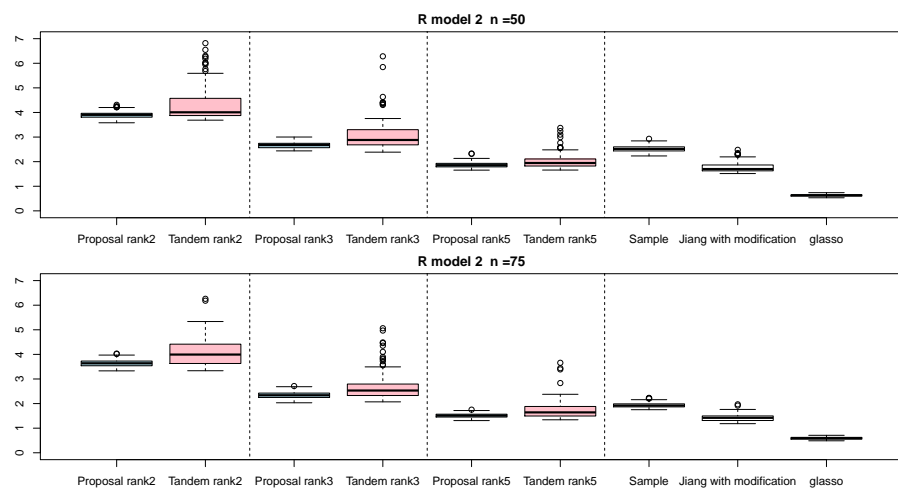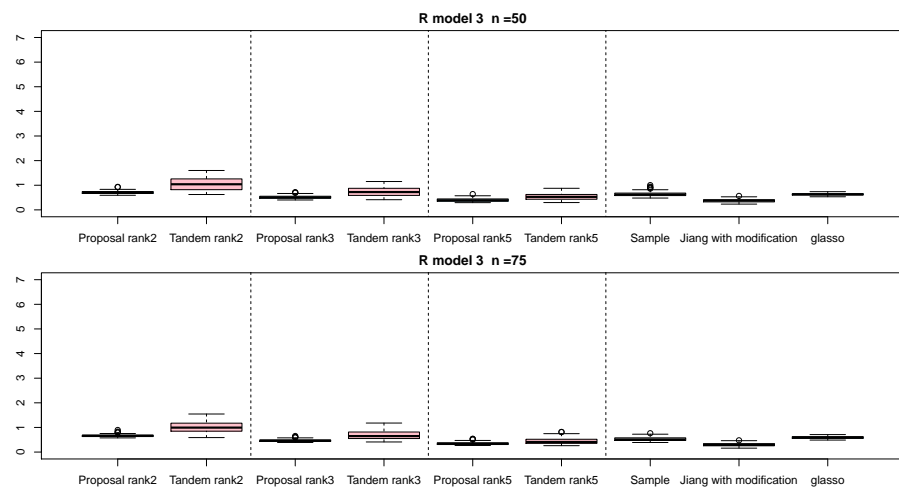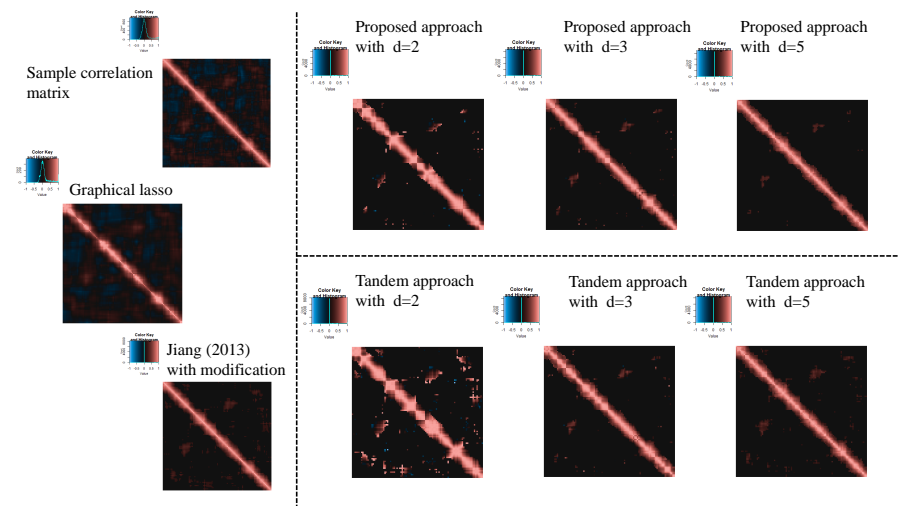

Figure 8. Relative errors of the S-norm for $R^{(2)}$ with $n = 50$ and $n = 75$; the vertical axis indicates the results of relative errors of the S-norm.

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

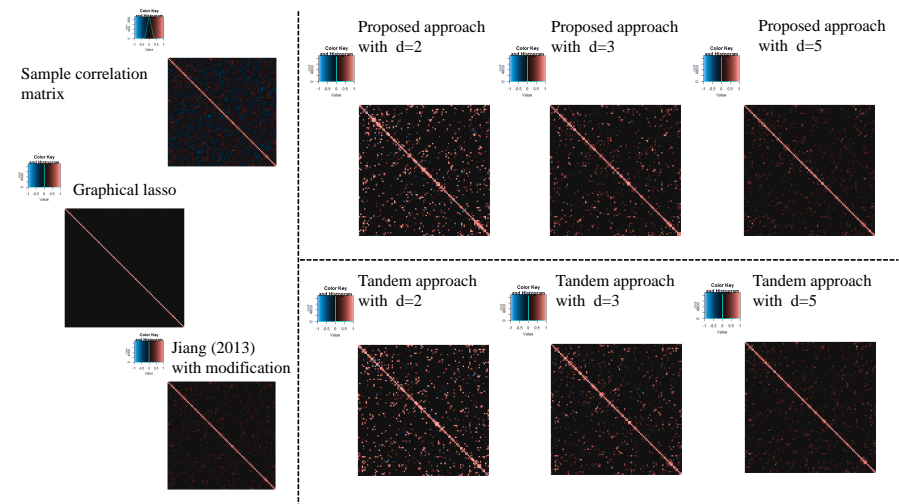**Figure 10.** Examples of estimated correlation matrices for true correlation model 1 ($n = 50$).

Figure 11. Examples of estimated correlation matrices for the true correlation model 2 ($n = 50$).
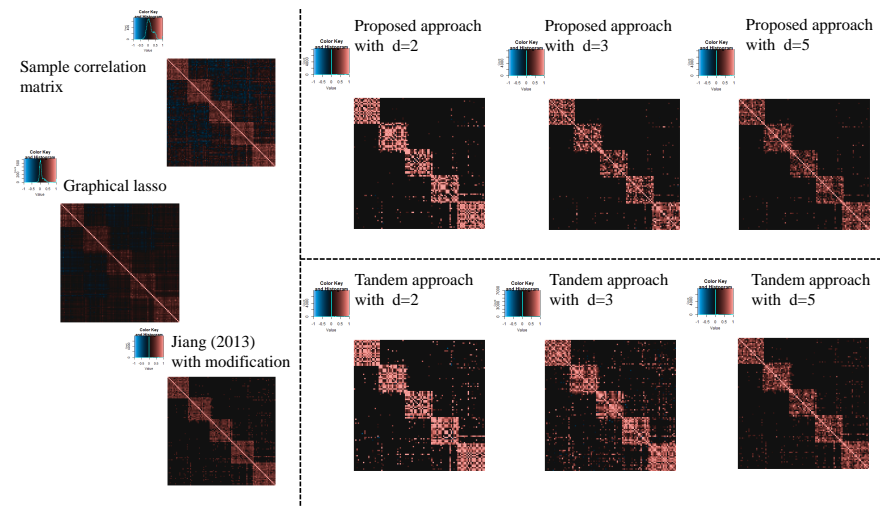
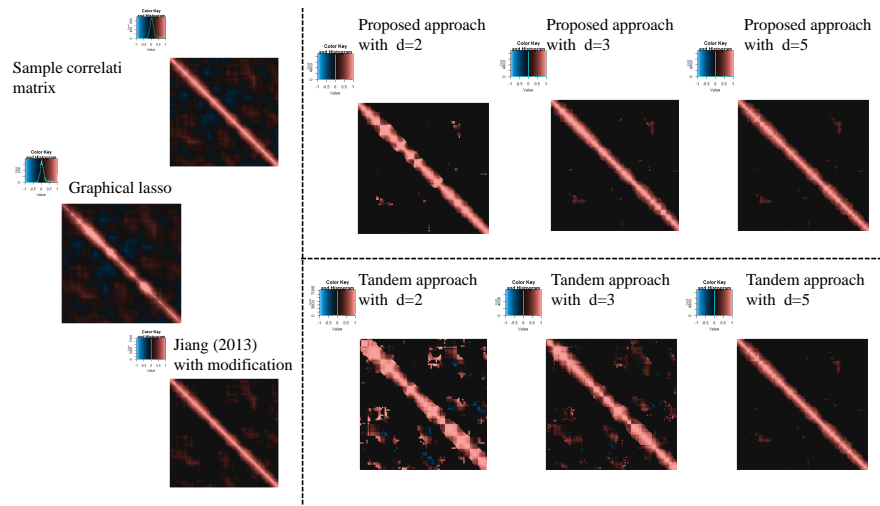**Figure 12.** Examples of estimated correlation matrices for the true correlation model 3 ($n = 50$).

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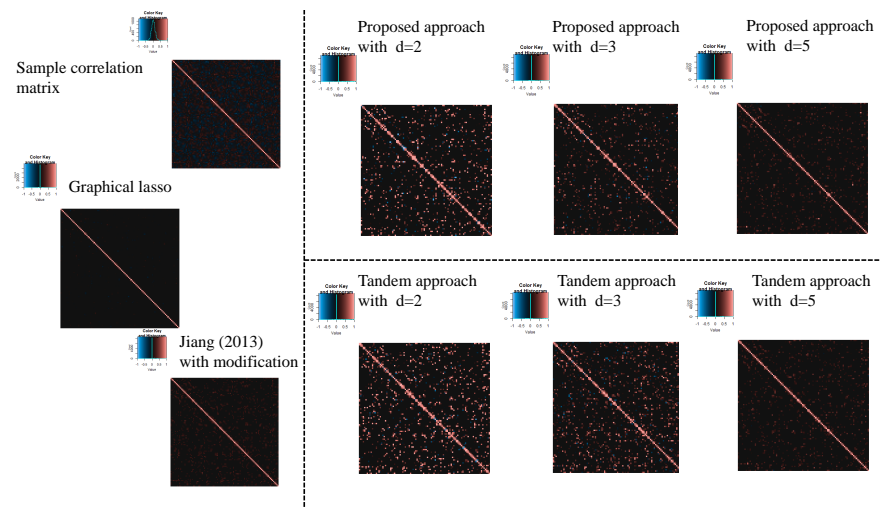

Figure 14. Examples of estimated correlation matrices for the true correlation model 2 ($n = 75$).
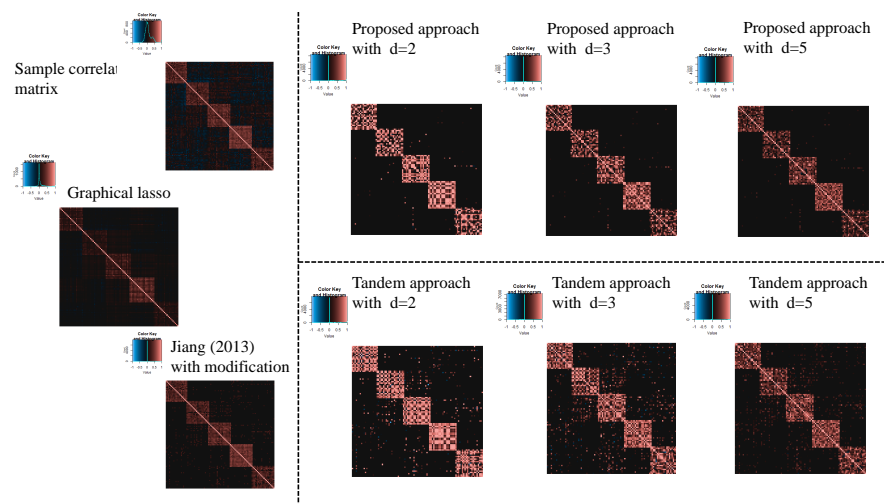
Figure 15. Examples of estimated correlation matrices for true correlation model 3 ($n = 75$).

3.2. Result of Application of Microarray Gene Expression Dataset

In this subsection, the results of the application of the microarray gene expression dataset are shown. For the estimated original correlation matrix, Jiang (2013) with modifications, the proposed approach tandem approach, and graphical lasso, see Figure 16. Therefore, the percentage points of $d = 2$, 3 and 5 in the proposed approach were estimated as $\alpha = 0.82$, 0.81, and 0.75, respectively, while the percentage points in the tandem approach and Jiang (2013) with modifications were both $\alpha = 0.65$. The estimated results of Jiang (2013) with modifications are presented in Figure 16. However, FPRs were higher than those of the proposed approach. Here, the FPR is unaffected 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 confirmed 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 the "informative" class tends to be similar to the results obtained in Ref. [25] using the heatmap.

Next, Table 5 shows the results of the FPR of the proposed approach, the tandem approach, Jiang (2013) with modifications, and the graphical lasso. The FPRs of the proposed method with $d = 2$, 3, and 5 were all lower than those of the tandem approach, Jiang (2013) with modifications, and graphical lasso. In addition, the tendency can be confirmed in Figure 16 visually. In fact, we can confirm that the proposed method was able to estimate the correlation coefficient between the classes as zero, compared to the tandem approach. The tendency was observed to be with respect to the results of the numerical simulations.

Table 5. Results of applying the microarray gene expression dataset.

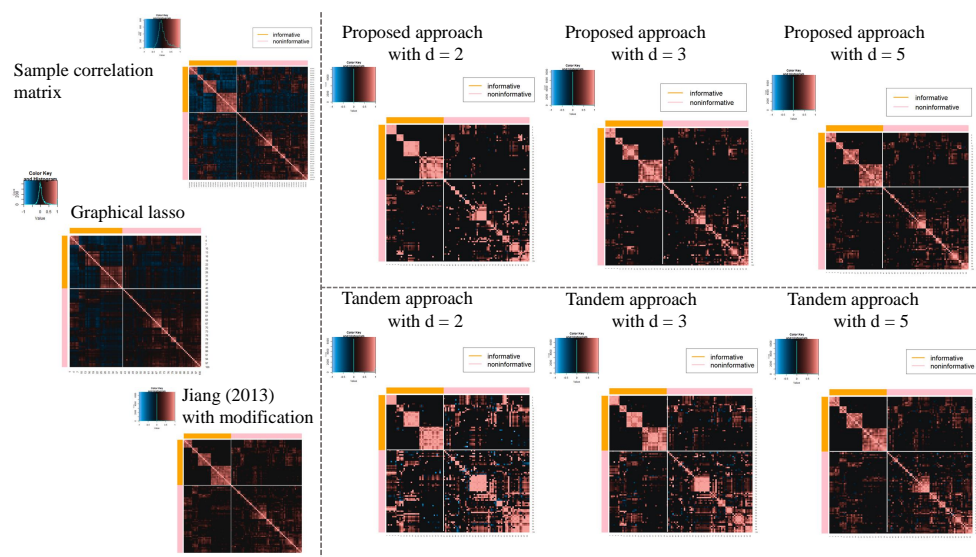| Method	d	FPR	Within Class Sparsity
Proposal	2	**0.128**	0.757
Proposal	3	0.139	0.748
Proposal	5	0.197	0.687
Tandem	–	0.285	0.578
Jiang (2013) with modifications	–	0.285	0.578
Graphical lasso	–	1.000	0.000

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. Conclusions

This study proposed a novel estimation method for sparse low-rank correlations based on the MM algorithm. The approach overcomes the problem of estimating low-rank correlation matrices. The low-rank approximation is an a potent tool, and the approach provides us with a straightforward interpretation of the features because the contrast of the estimated coefficients becomes larger. However, these estimates sometimes lead to misinterpretations. 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. To confirm the efficiency of the proposed method, we performed numerical simulations and experimented using a real example, which involved the use of a microarray gene expression dataset. In the case of the real example, the FPR of the proposed approach with $d = 2, 3$, and 5 were found to be 0.128, 0.139, and 0.197, respectively, although those of the tandem approach and Jiang (2013) with modifications were found to be 0.285 and 0.285, respectively. We were, therefore able to confirm that the FPR of the proposed approach was the best, irrespective of the rank. Similarly, from the numerical simulations, we confirmed that the FPR of the proposed approach was superior to that of the tandem approach and Jiang (2013) with modifications. In addition to that, we also found that these relative errors of the proposed approach were superior to those of the tandem approach through numerical simulations. The proposed approach is considered an approximate true correlation matrix compared to the tandem approach.

Next, we refer to each method compared. The sample correlation matrix was used as a reference and for comparison. For Jiang (2013), concerning modifications, the numerical simulation results confirm that the F-norm and the S-norm are better than the proposed approach and the tandem approach because the low-rank approximation is not performed. Furthermore, Jiang (2013) with modifications is better than the sample correlation matrix without sparse constraints. Graphical lasso is a method for the sparse estimation of the inverse of the covariance, and the computed correlation matrix is not necessarily a sparse. Therefore, FPR, sparsity, and within-class sparsity were not good compared to the other methods through numerical simulations and examples of real-data applications. In other words, when a correlation matrix is estimated, it would be better to estimate a direct correlation matrix or a covariance matrix. The TPR, FPR, and sparsity of the tandem approach are almost identical to Jiang (2013) with modifications. The tandem method has a higher FPR than the proposed method and worse results for sparsity and within-class sparsity, which are indicators of interpretability. Furthermore, the proposed method performed better for the F-norm and the S-norm because the tandem method does not take

threshold values into account. However, the numerical simulation results show that the TPR tends to be better than the proposed method.

Here, we show the recommendations of the proposed approach. First, the proposed approach can provide the interpretable sparse low-rank correlation matrix from the result of both the sparsity of the numerical simulation and the within-class sparsity of the real example. Second, in a real example, the sparse low-rank correlation matrix estimated by the proposed approach was close to previous studies [21]. However, additional constraint, such as a low-rank approximation, is added to the proposed method. Third, the proposed approach can reduce the FPR compared to the tandem approach and the graphical lasso in this study. In short, the proposed approach can avoid misunderstandings compared to those methods.

Although the proposed approach showed promising results in both experiments, several limitations need further investigation. First, when the rank is set low, the TPRs of the proposed method were low compared to those of the tandem approach. The relationship between the determination of the rank and the corresponding TPR, F-norm, S-norm, sparsity, and within-class sparsity was a trade-off. Therefore, a method for a more effective determination of the rank should be developed. As one of the solutions to the problem, we can consider the introduction of a nuclear-norm regularization [44]. Second, when the percentage point is set significantly higher, it may be challenging to obtain the sparse low-rank correlation matrix because it becomes difficult to calculate an updated formula of the low-rank correlation matrix. Third, the proposed approach was more sparse than the tandem approach through numerical simulations and real examples, but this was not shown theoretically. This point also needs to be studied theoretically, but this is a topic for future research. Finally, it should be noted that the proposed approach only focuses on the positive correlation matrix and does not consider the negative correlation matrix for the simplicity of the interpretation. However, the proposed method can be extended for considering both positive and negative correlation coefficients.

Author Contributions: K.T. constructed the proposed approach and designed the numerical simulation and real application; Y.F. and K.T. implemented the algorithm for artificial and real data, and wrote the manuscript.; S.H. modified the proposed method to solve the practical problems and reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The gene expression data set belongs to the package R 'MADE4' in: https://www.bioconductor.org/packages/release/bioc/html/made4.html (accessed on 17 April 2022).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MM	Majorize-minimization
DOAJ	Directory of open access journals
TLA	Three letter acronym
LD	Linear dichroism
FPR	False Positive Rate
TPR	True Positive Rate
F-norm	Frobenius norm
S-norm	Spectral norm
SRBCT	Small round blue cell tumours of childhood
NB	Neuroblastoma
RMS	Rhabdomyosarcoma
BL	Burkitt lymphoma, a subset of non-Hodgkin lymphoma
EWS	Ewing family of tumours

References

1. Wilkinson, L.; Friendly, M. The history of the cluster heat map. *Am. Stat.* **2009**, *36*, 179–184. [CrossRef]
2. Ten Berge, J. *Least Squares Optimization in Multivariate Analysis*; DSWO Press: Leiden, The Netherlands, 1993.
3. Pietersz, R.; Groenen, P. Rank reduction of correlation matrices by majorization. *Quant Financ.* **2004**, *4*, 649–662. [CrossRef]
4. Simon, D.; Abell, J. A majorization algorithm for constrained approximation. *Linear Algebra Its Appl.* **2010**, *432*, 1152–1164. [CrossRef]
5. Grubisic, I.; Pietersz, R. Efficient rank reduction of correlation matrices. *Linear Algebra Its Appl.* **2007**, *422*, 629–653. [CrossRef]
6. Duan, X.; Bai, J.; Li, J.; Peng, J. On the low rank solution of the q-weighted nearest correlation matrix problem. *Numer. Linear Algebra Appl.* **2016**, *23*, 340–355. [CrossRef]
7. Ding, C.; He, X. *k*-means clustering via principle component analysis. In Proceedings of the International Conference on Machine Learning (ICML), Banff, AB, Canada, 4–8 July 2004.
8. Eckart, C.; Young, G. The approximation of one matrix by another of lower rank. *Psychometrika* **1936**, *1*, 211–218. [CrossRef]
9. Engel, J.; Buydens, L.; Blanchet, L. An overview of large-dimensional covariance and precision matrix estimators with applications in chemometrics. *J. Chemom.* **2017**, *31*, e2880. [CrossRef]
10. Lam, C. High-dimensional covariance matrix estimation. *Wiley Interdiscip. Rev. Comput. Stat.* **2020**, *12*, e1485. [CrossRef]
11. Bien, J.; Tibshirani, R. Sparse estimation of a covariance matrix. *Biometrika* **2011**, *99*, 807–820. [CrossRef]
12. Rothman, A. Positive definite estimators of large covariance matrices. *Biometrika* **2012**, *99*, 733–740. [CrossRef]
13. Xue, L.; Ma, S.; Zou, H. Positive definite L_1 penalized estimation of large covariance matrices. *J. Am. Stat. Assoc.* **2012**, *107*, 1480–1491. [CrossRef]
14. Cai, T.; Liu, W.; Luo, X. A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *J. Am. Stat. Assoc.* **2011**, *106*, 594–607. [CrossRef]
15. D'aspremont, A.; Banerjee, O.; Ghaoui, L. First-order methods for sparse covariance selection. *SIAM J. Matrix Anal. Appl.* **2008**, *30*, 56–66. [CrossRef]
16. Friedman, J.; Hastie, T.; Tibshirani, R. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **2008**, *9*, 432–441. [CrossRef] [PubMed]
17. Rothman, A.; Bickel, P.; Levina, E.; Zhu, J. Sparse permutation invariant covariance estimation. *Electron. J. Stat.* **2008**, *2*, 495–515. [CrossRef]
18. Yuan, M.; Lin, Y. Model selection and estimation in Gaussian graphical model. *Biometrika* **2007**, *94*, 19–35. [CrossRef]
19. Cui, Y.; Leng, C.; Sun, D. Sparse estimation of high-dimensional correlation matrices. *Comput. Stat. Data Anal.* **2016**, *93*, 390–403. [CrossRef]
20. Bickel, P.; Levina, E. Covariance regularization by thresholding. *Ann. Stat.* **2008**, *36*, 2577–2604. [CrossRef]
21. Cai, T.; Liu, W. Adaptive thresholding for sparse covariance matrix estimation. *J. Am. Stat. Assoc.* **2011**, *106*, 594–607. [CrossRef]
22. Bickel, P.; Levina, E. Regularized estimation of large covariance matrices. *Ann. Stat.* **2008**, *36*, 199–227. [CrossRef]
23. El Karoui, N. Operator norm consistent estimation of large-dimensional sparse covariance matrices. *Ann. Stat.* **2008**, *36*, 2717–2756. [CrossRef]
24. Jiang, B. Covariance selection by thresholding the sample correlation matrix. *Stat. Probab. Lett.* **2013**, *83*, 2492–2498. [CrossRef]
25. Rothman, A.; Levina, E.; Zhu, J. Generalized thresholding of large covariance matrices. *J. Am. Stat. Assoc.* **2009**, *104*, 177–186. [CrossRef]
26. Lam, C.; Fan, J. Sparsistency and rates of convergence in large covariance matrices estimation. *Ann. Stat.* **2009**, *37*, 4254–4278. [CrossRef]
27. Liu, H.; Wang, L.; Zhao, T. Sparse covariance matrix estimation with eigenvalue constraints. *J. Comput. Graph. Stat.* **2014**, *37*, 439–459. [CrossRef] [PubMed]
28. Zhou, S.; Xiu, N.; Luo, Z.; Kong, L. Sparse and low-rank covariance matrix estimation. *J. Oper. Res. Soc. China* **2015**, *3*, 231–250. [CrossRef]
29. Savalle, P.; Richard, E.; Vayatis, N. Estimation of simultaneously sparse and low rank matrices. In Proceedings of the International Conference on Machine Learning (ICML), Edinburgh, UK, 26 June–1 July 2012.
30. Pourahmadi, M. Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika* **1999**, *86*, 677–690. [CrossRef]
31. Chang, C.; Tsay, R. Estimation of covariance matrix via the sparse Cholesky factor with lasso. *J. Stat. Plan. Inference* **2010**, *86*, 677–690. [CrossRef]
32. Kang, X.; Wang, M. Ensemble sparse estimation of covariance structure for exploring genetic desease data. *Comput. Stat. Data Anal.* **2021**, *159*, 107220. [CrossRef]
33. Li, C.; Yang, M.; Wang, M.; Kang, H.; Kang, X. A Cholesky-based sparse covariance estimation with an application to genes data. *J. Biopharm. Stat.* **2021**, *31*, 603–616. [CrossRef]
34. Yang, W.; Kang, X. An improved banded estimation for large covariance matrix. *Commun.-Stat.-Theory Methods* **2021**. [CrossRef]
35. Kang, X.; Deng, X. On variable ordination of Cholesky-based estimation for a sparse covariance matrix. *Can. J. Stat.* **2021**, *49*, 283–310. [CrossRef]
36. Niu, L.; Liu, X.; Zhao, J. Robust estimator of the correlation matrix with sparse Kronecker structure for a high-dimensional matrix-variate. *J. Multivar. Anal.* **2020**, *177*, 104598. [CrossRef]

37. Knol, D.; Ten Berge, J. Least-squares approximation of an improper correlation matrix by a proper one. *Psychometrika* **2012**, *54*, 53–61. [CrossRef]

38. Hunter, D.; Lange, K. A tutorial on MM algorithm. *Am. Stat.* **2004**, *58*, 30–37. [CrossRef]

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

40. Khan, J.; Wei, J.S.; Ringner, M.; Saal, L.H.; Ladanyi, M.; Westermann, F.; Berthold, F.; Schwab, M.; Antonescu, C.R.; Meltzer, P.S.; et al. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat. Med.* **2001**, *7*, 673–679. [CrossRef]

41. Friedman, J.; Hastie, T.; Tibshirani, R. Sparse inverse covariance estimation with the lasso. *arXiv* **2007**, arXiv:0708.3517.

42. Galloway, M. *CVglasso: Lasso Penalized Precision Matrix Estimation*; R package version 1.0.; 2018. Available online: https://cran.r-project.org/web/packages/CVglasso/index.html (accessed on 17 April 2022).

43. Culhane, A.; Thioulouse, J.; Perriere, G.; Higgins, D. MADE4: An R package for multivariate analysis of gene expression data. *Bioinformatics* **2005**, *21*, 2789–2790. [CrossRef]

44. Candes, E.; Tao, T. The power of convex relaxation: Near-optimal matrix completion. *IEEE Trans. Inf. Theory* **2010**, *56*, 2053–2080. [CrossRef]