

Research Article

Variable Selection and Joint Estimation of Mean and Covariance Models with an Application to eQTL Data

JungJun Lee ¹, SungHwan Kim ², Jae-Hwan Jhong,¹ and Ja-Yong Koo ¹

¹Department of Statistics, Korea University, Seoul 02841, Republic of Korea

²Department of Applied Statistics, Konkuk University, Seoul 05029, Republic of Korea

Correspondence should be addressed to SungHwan Kim; swiss747@gmail.com

Received 1 October 2017; Revised 25 December 2017; Accepted 18 April 2018; Published 25 June 2018

Academic Editor: Nadia A. Chuzhanova

Copyright © 2018 JungJun Lee et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In genomic data analysis, it is commonplace that underlying regulatory relationship over multiple genes is hardly ascertained due to unknown genetic complexity and epigenetic regulations. In this paper, we consider a joint mean and constant covariance model (JMCCM) that elucidates conditional dependent structures of genes with controlling for potential genotype perturbations. To this end, the modified Cholesky decomposition is utilized to parametrize entries of a precision matrix. The JMCCM maximizes the likelihood function to estimate parameters involved in the model. We also develop a variable selection algorithm that selects explanatory variables and Cholesky factors by exploiting the combination of the GCV and BIC as benchmarks, together with Rao and Wald statistics. Importantly, we notice that sparse estimation of a precision matrix (or equivalently gene network) is effectively achieved via the proposed variable selection scheme and contributes to exploring significant hub genes shown to be concordant to *a priori* biological evidence. In simulation studies, we confirm that our model selection efficiently identifies the true underlying networks. With an application to miRNA and SNPs data from yeast (a.k.a. eQTL data), we demonstrate that constructed gene networks reproduce validated biological and clinical knowledge with regard to various pathways including the cell cycle pathway.

1. Introduction

Generally, joint estimation of mean and covariance has been developed to address problems related to biomedical data. In longitudinal data analysis, for instance, identifying correct correlation structures within each subject is a major focus and many studies come up with variants of joint mean and covariance estimation to enhance statistical efficiency (see Ye and Pan [1] and references therein). With regard to graphical models, particularly, conditional Gaussian graphical models (cGGM) aiming to elucidate conditional dependence structures subject to the mean have increasingly received much attention [2–5] and conceptually can also be viewed as examples of joint mean and covariance estimation in essence. On a methodological side, joint estimation has been found to be practically applicable in the sense that applications of joint estimation methods cover bioinformatics, such as hormone and transcriptome [1, 4, 6–8], and stock price analysis [9]. A majority of existing methods involve variable selection of covariates for the mean vector and sparsity of covariance

mainly based on the penalization methods. However, in this paper, we purposely focus on variable selection and joint estimation with a pure implementation of the likelihood-based method, unlike common penalization techniques, to better analyze eQTL data in light of genetic feature selection.

Over the decades, genomics research has focused on comprehensive understanding of regulatory networks in the context of system biology. Commonly we are interested in a gene network, which pictures the interplay among genetic factors (e.g., gene regulation and activation). Particularly, it is important to investigate how a given genotype (genetic variants) at a particular quantitative trait locus (QTL) affects measured phenotypes and traits at that locus. For instance, gene expression quantitative loci (eQTL) make use of gene expressions as quantitative traits. eQTL analysis has been widely applied to figure out the effect of genetic perturbations associated with diseases as well as to construct regulatory networks describing how genes regulate expressions of other genes [10, 11]. More precisely, the location of single nucleotide polymorphisms (SNPs) may affect multiple gene expression

levels, and this accidentally causes misleading inference for dependency structure among genes [4]. Many popular methods [2–5, 12] have been introduced to identify gene networks, aiming at learning networks subject to perturbation effects by genetic variants on the basis of population gene expression and genotype data.

Yin and Li [4] proposed the sparse conditional graphical Gaussian model with ℓ_1 and the adaptive lasso penalty function. Li *et al.* [3] suggested the two-stage estimation framework: (1) estimating a nonsparse conditional covariance matrix of genes based on a conditional variance operator between the reproducing kernel Hilbert spaces of marker genes and then (2) using ℓ_1 and adaptive lasso penalty to obtain sparse estimates of a precision matrix under the cGGM. Cai *et al.* [12] studied the covariate-adjusted precision matrix estimation (CAPME) method using the constrained ℓ_1 optimization.

While most recent studies encourage sparsity estimation of a precision matrix based on the penalized likelihood, we instead rely on classical variable methods based on the standard likelihood. Strictly speaking, the penalized likelihood, according to its definition, cannot be viewed as the likelihood. It naturally poses a question whether the likelihood-based method performs better than the penalized likelihood approaches. To address this question, we consider a joint mean and constant covariance model (JMCCM) inspired by Pourahmadi [13] and propose methods for variable selection which effectively identify sparse conditional gene networks and covariates relevant to gene regulations. We employ the modified Cholesky decomposition to guarantee positive definiteness of an estimated precision matrix [13]. With this reparametrization of the precision matrix, the log-likelihood function corresponding to our model can be decomposed into an additive form of each response in terms of Cholesky parameters. This facilitates a coordinate descent type implementation we adopt for the precision matrix estimation. The combination of both generalized cross-validation (GCV) and Bayesian information criterion (BIC) performs variable selection as a benchmark. Rao and Wald statistics are also utilized to add and delete genetic markers for each gene expression and the Cholesky factors for the precision matrix. To the best of our knowledge, only a few works use Rao and Wald statistics for variable selection in the joint estimation problem, particularly with an application to eQTL analysis.

In simulation studies, extensive simulation scenarios experimentally confirm improved estimation efficiency and precise variable selection of the proposed method. For real data applications, we perform eQTL analysis via the JMCCM with gene expressions and SNPs of yeast data, in pursuit of detecting the effects of variant perturbations and underlying gene regulations. We find that the JMCCM effectively uncovers biological pathways that may potentially account for known biological processes. Taken together, the JMCCM is shown to be effective in identifying conditional dependence structures among variables compared to the existing graphical models with penalization methods such as the sparse cGGM [5] and CAPME [12].

The paper is outlined as follows. In Section 2, we describe our JMCCM with modified Cholesky decomposition and

maximum likelihood estimates. The variable selection algorithm using Rao and Wald statistics for SNPs and Cholesky factors is explained in Section 3. Section 4 deals with simulation studies to demonstrate performance of variable selection and estimation of the proposed model. The yeast cell cycle pathway genes with SNPs data analysis are presented in Section 5. Concluding remarks and discussion in Section 6 are followed by the Appendix, which provides mathematical details of our method.

2. Model And Estimators

2.1. JMCCM with the Modified Cholesky Decomposition. Contrary to previous methods [3, 4], the JMCCM primarily aims at simultaneous estimation over the mean and precision matrix of the Gaussian graphical model. Suppose that (x, y) is a pair of the $p \times 1$ vector x of genetic markers and the $m \times 1$ random vector y of expression levels. Let x denote the vector of SNPs and let y denote the gene expression traits following m -variate multivariate normal distribution with the mean $\mu(x; \beta)$ and covariance matrix Σ as follows:

$$y | x \sim N_m(\mu(x; \beta), \Sigma), \quad (1)$$

where the j th entry of $\mu(x; \beta)$ is $x^\top \beta_j$ for $j = 1, \dots, m$, and $\beta_j = (\beta_j^1, \dots, \beta_j^p)$ is the $p \times 1$ linear regression coefficient vector indicating effects of SNPs perturbations to gene expressions and $\beta = (\beta_1, \dots, \beta_m)$. Importantly, note that Σ does not depend on x . The coefficient β is assumed to be sparse, since each gene is known to have only a few genetic regulators according to Cai *et al.* [12]. The precision matrix represents a gene network (graph), as in an undirected GGM, by corresponding the nonzero (i, j) th element of the precision matrix with an edge between two vertices i and j [14]. This edge represents conditional dependence of genes i and j given all other gene expression levels. Thus, our goal is to identify nonzero entries of the precision matrix in order to construct a conditional dependence genetic network after the effects of SNPs perturbations are removed. The precision matrix Σ^{-1} is also expected to be sparse [4].

One of our primary interests is to estimate the precision matrix Σ^{-1} , which is symmetrically positive definite, so we need to ensure that the estimate of the precision matrix also satisfies symmetrical positive definiteness. To this end, we apply the modified Cholesky decomposition [13] to the precision matrix, denoted by K , as follows:

$$K(\phi, \tau) = C(\phi)^\top D(\tau) C(\phi), \quad (2)$$

where C is an upper triangular matrix with diagonal entries 1 and above-diagonal elements consisting of negative of $\phi = (\phi_{12}, \dots, \phi_{1m}, \phi_{23}, \dots, \phi_{2m}, \dots, \phi_{m-1,m})$ and D is a diagonal matrix containing $\tau = (\tau_1, \dots, \tau_m)$ with $\tau_j > 0$ for all $j = 1, \dots, m$ as diagonal entries. Here, the superscript " \top " denotes the transpose of a matrix. Positive definiteness of K is shown in Appendix A. Throughout this paper, a vector in the parenthesis is considered as a column vector. Let $\phi_j = (\phi_{j,j+1}, \dots, \phi_{jm})$ for $j = 1, \dots, m-1$. Then we can write $\phi = (\phi_1, \dots, \phi_{m-1})$. The parameter space for JMCCM is defined by

$\Theta = \{\theta = (\beta, \phi, \tau) : \beta \in \mathbb{R}^{mp}, \phi \in \mathbb{R}^{m(m-1)/2} \text{ and } \tau \in \mathbb{R}_+^m\}$, where \mathbb{R}_+^m represents the set of m -dimensional vectors of positive real numbers.

2.2. Maximum Likelihood Estimation. Suppose that we have the N independent observations (x^i, y^i) , $i = 1, \dots, N$, sampled from (1), where $y^i = (y_1^i, \dots, y_m^i)$ and $x^i = (x_1^i, \dots, x_p^i)$ represent the i th observation of x and y , respectively. With $\Sigma^{-1} = K = K(\phi, \tau)$, $\mu(x) = \mu(x; \beta)$, and $\mu^i = \mu(x^i)$, the log-likelihood function corresponding to (1) is given by

$$\ell(\theta) = \frac{N}{2} \left\{ -m \log(2\pi) + \sum_{j=1}^m \log \tau_j - \text{tr}(KV(\beta)) \right\}, \quad (3)$$

where $V(\beta) = (1/N) \sum_{i=1}^N (y^i - \mu^i)(y^i - \mu^i)^\top$ and "tr" denotes the trace of a matrix. A derivation of (3) can be found in Appendix B. The maximum likelihood estimator (MLE) of θ is defined by

$$\hat{\theta} = \underset{\theta \in \Theta}{\text{argmax}} \ell(\theta). \quad (4)$$

Some notations are needed to express MLE of θ in conjunction with variable selection. For $j = 1, \dots, m$, let G_j be an index set of active (or significant) coefficients of β_j and denote by g_j the number of elements in G_j . Let $G = \{G_1, \dots, G_m\}$ and $g = g_1 + \dots + g_m$. Obviously, $g \leq mp$ and if $g = mp$, then all variables are significant for all y_1, \dots, y_m . Define $\beta(G) = (\beta_1(G_1), \dots, \beta_m(G_m)) \in \mathbb{R}^g$ and $\beta_j(G_j) = [\beta_j^l]_{l \in G_j} \in \mathbb{R}^{g_j}$ for $j = 1, \dots, m$. Write $y_k = (y_k^1, \dots, y_k^N)$ for $k = 1, \dots, m$, $X = (1/N)(x_1, \dots, x_p)$, and $X(G_j) = [X_l]_{l \in G_j} \in \mathbb{R}^{N \times g_j}$, where X_l denotes l th column of X . Let $Y = (y_1^\top, \dots, y_m^\top) \in \mathbb{R}^{mN}$, $X = \oplus_{j=1}^m X(G_j) \in \mathbb{R}^{mN \times g}$, and $W = K \otimes I_N \in \mathbb{R}^{mN \times mN}$, where \oplus and \otimes represent direct sum and Kronecker product, respectively, and I_N is the $N \times N$ identity matrix. Then the MLE $\hat{\beta}(G)$ of $\beta(G)$ can be expressed as

$$\hat{\beta}(G) = (X^\top W X)^{-1} X^\top W Y. \quad (5)$$

Observe that estimating $\beta(G)$ involves K .

We can estimate K by obtaining MLE of ϕ and τ . For $j = 1, \dots, m-1$, let $\theta_j = (\beta, \phi_j, \tau_j)$ with $\theta_m = (\beta, \tau_m)$. Let $V_{jj}, V_{j,21}$, and $V_{j,22}$ represent the $(1, 1)$, $(2, 1)$, and $(2, 2)$ components of the lower j th principal submatrix $V_j(\beta)$ of $V(\beta)$. Then we can express

$$\ell(\theta) = \sum_{j=1}^m \ell_j(\theta_j), \quad (6)$$

where

$$\begin{aligned} \ell_j(\theta_j) = & \frac{N}{2} \left\{ -\log(2\pi) + \log \tau_j \right. \\ & \left. - \tau_j (V_{jj} - 2\phi_j^\top V_{j,21} + \phi_j^\top V_{j,22} \phi_j) \right\}. \end{aligned} \quad (7)$$

By (6), we can obtain $(\hat{\phi}, \hat{\tau})$ from $(\hat{\phi}_j, \hat{\tau}_j)$ for each j with $\hat{\phi}_j = \underset{\phi_j}{\text{argmax}} \ell_j(\theta_j)$ and $\hat{\tau}_j = \underset{\tau_j}{\text{argmax}} \ell_j(\theta_j)$. Thus, instead of finding a solution to the optimization problem with (3), we optimize each $\ell_j(\theta_j)$ with respect to ϕ_j and τ_j . Thus, the MLE $\hat{\phi}_j$ and $\hat{\tau}_j$ are expressed in terms of $V(\beta)$ in a way that

$$\begin{aligned} \hat{\phi}_j &= V_{j,22}^{-1} V_{j,21}, \\ \frac{1}{\hat{\tau}_j} &= V_{jj} - V_{j,12} V_{j,22}^{-1} V_{j,21} \end{aligned} \quad (8)$$

for $j = 1, \dots, m-1$

with $\hat{\beta}$ from (5) and $1/\hat{\tau}_m = V_{mm}$. We estimate Σ^{-1} by \hat{K} and it is defined by

$$\hat{K} = K(\hat{\phi}, \hat{\tau}). \quad (9)$$

Derivations of (5), (6), and (8) are presented in Appendix C.

3. Consecutive Variable Selection Algorithm

As mentioned above, β and K are commonly believed to be sparse in genomic data analysis. To address sparsity, the lasso-type penalty imposed on both regression coefficients and precision matrix has been popularly applied to diverse graphical models [4, 5]. Stepping aside the lasso-type approach, we develop a variable selection technique that mainly relies on the combination of classical variable selection methods. Generally, the numbers of SNPs and genes tend to be considerably huge so that computational costs normally become prohibitive. In order to address this problem, the proposed variable selection algorithm proceeds with largely two stages: (1) preliminary variable selection for mean and precision matrix and (2) secondary variable selection in the middle of the joint model estimation. It is important to note that the first stage leads possible variables to be limited in scope (i.e., working parameters in the model) in order to circumvent high computational complexity.

3.1. Preliminary Variable Selection. Preliminary variable selection is largely twofold: variable selection for the mean part and covariance part. The idea behind that is to add variables (or equivalently parameters) to the joint model with the maximum Rao statistic and to delete ones with the minimum Wald statistic. You may refer to Koo [15] for the basis selection method or Kooperberg *et al.* [16] that explain variable selection schemes based on Rao and Wald statistics.

3.1.1. Mean Part. When it comes to the mean part, we carry out selecting predictor variables (i.e., SNPs) for each response variable (i.e., gene expression), dealing with a univariate multiple regression problem. In the addition stage, we start off with a model including only an intercept term. The MLE is used as estimator for β and maximum Rao statistic (Rao [17]) is the criterion for adding a predictor together with GCV (Friedman [18]) as a stopping rule. In the deletion stage, Wald statistic is calculated to exclude predictor variables such that

```

(1) for  $j$  in  $1 : m$  do
(2)   Set the minimal linear regression model:  $Y_j = \beta_j^0$ .
(3)   Compute  $\hat{\beta}_j^0$  and GCV.
(4)   while GCV decreases do
(5)     Among predictors not in the current model, add a predictor  $X_b$  having the
     maximum Rao statistic.
(6)     Update  $\hat{\beta}_j(G_j)$  using (5) and compute GCV.
(7)   end while
(8)   while GCV decreases do
(9)     Among predictors in the current model, delete a predictor  $X_{b'}$  having the
     minimum Wald statistic.
(10)    Update  $\hat{\beta}_j(G_j)$  using (5) and compute GCV.
(11)  end while
(12)  The optimal model is chosen by the minimum GCV.
(13) end for

```

ALGORITHM 1: Variable selection in the mean vector estimation.

the updated model minimizes Wald statistics. The final model for the mean regression is chosen by the minimum GCV. Details are summarized in Algorithm 1. Once Algorithm 1 is done, the number of predictor variables included in the joint model no longer increases, while variable reduction can happen in Algorithm 3 (Section 3.2).

3.1.2. Covariance Part. The rationale behind variable selection in the precision matrix estimation is that each Y_j , $j = 1, \dots, m-1$, is regressed on Y_{j+1}, \dots, Y_m , and the regression coefficients are negative of off-diagonal entries of $C(\phi)$, the by-product of the modified Cholesky decomposition (2) [13, 19, 20]. Clearly, this is one of the major benefits of reparametrization via Cholesky decomposition [13] in pursuit of improved interpretation.

Subsequent to selection for predictor variables, we compute $V(\hat{\beta}(G))$. Given initial $\hat{\phi}_j^{(0)} = 0$, we start with computing $\hat{\tau}_j$ for $j = 1, \dots, m$ as in (8) using $V(\hat{\beta}(G))$ obtained from Algorithm 1. Then compute Rao statistic to add one variable out of Y_{j+1}, \dots, Y_m to the current model that builds on the response variable Y_j . We repeatedly update for Y_j , where $j = 1, \dots, m-1$. Next, we choose one variable, say ϕ_{lb} , for $l = 1, \dots, m-1$ and $b = l+1, \dots, m$, such that Rao statistic is maximized and thereby update $\hat{\phi}_l$ and $\hat{\tau}_l$, respectively. When calculating $\hat{\tau}$ and $\hat{\phi}$, the BIC is computed, and the addition process stops if the BIC no longer decreases. At the completion of addition, we build a model with all selected variables as a full model and subsequently begin deletion from a full model. Deletion process is similar to the addition process except that the minimum Wald statistic is used for deletion in place of the maximum Rao statistic. Successive deletion continues until the BIC stops decreasing. In the last stage, the final model is also selected by the BIC. Details are presented in Algorithm 2. Once Algorithm 2 is finished, the number of ϕ_{jk} included in the joint model no longer increases, while variable reduction could occur in the joint estimation (Algorithm 3, Section 3.2). Sparsity of \hat{K} is

achieved through this variable selection scheme along with Algorithm 3.

3.2. Secondary Variable Selection in the Middle of the Joint Model Estimation. With variables (i.e., parameters) selected by the preliminary stage above, we implement simultaneous parameter estimation for both mean and precision in the joint model and additional variable selection. Once \hat{K} is computed via Algorithm 2 with fixed $\hat{\beta}(G)$, we start joint estimation by updating $\hat{\beta}(G)$, which is formed with weight least square estimates as in (5) and weights coming from \hat{K} . Then $\hat{\phi}$ and $\hat{\tau}$, ultimately \hat{K} , are newly computed using updated $V(\hat{\beta}(G))$, and again this updated \hat{K} serves as a weight for updating $\hat{\beta}(G)$. Over the updates, the interplay between $V(\hat{\beta}(G))$ and \hat{K} continues until the log-likelihood function converges. Afterwards, deletion to current parameters begins by Wald statistic, excluding one parameter with the minimum Wald statistic. Finally, the BIC is used as a stopping rule for deletion and selection to finalize model estimation. Algorithm 3 contains details about this procedure. While the preliminary variable selection works for each mean and covariance part under the assumption of $\hat{\beta}(G)$ or \hat{K} are fixed, this joint estimation procedure is designed to improve estimation and selection accuracy by reflecting the changes of $\hat{\beta}(G)$ and \hat{K} .

4. Experimental Studies

In order to assess the performance of our proposed method, we carry out experimental simulations and compare the sparse conditional Gaussian graphical model (SCGGM), Zhang and Kim [5], covariate-adjusted precision matrix estimation (CAPME), Cai *et al.* [12] and joint model with lasso penalty (JML), and Jhong *et al.* [21], all of which are based on penalized likelihood approaches. The competing methods are run with their default setting regarding tuning parameters. To evaluate similarity, the estimated precision

- (1) Compute $V(\widehat{\beta}(G))$ using $\widehat{\beta}(G)$ obtained from Algorithm 1.
- (2) Compute $\widehat{\tau}_1, \dots, \widehat{\tau}_m$ by (8).
- (3) Combining initial $\phi_j^{(0)} = 0$ for all $j = 1, \dots, m - 1$ with $\widehat{\tau}$, compute \widehat{K} and BIC.
- (4) **while** BIC decreases **do**
- (5) **for** j in $1 : (m - 1)$ **do**
- (6) **if** no elements of $\phi_j = (\phi_{j,j+1}, \dots, \phi_{j,m})$ are included in the joint model **then**
- (7) compute Rao statistics for ϕ_{jk} by V_{kj}^2/V_{kk} , $k = j + 1, \dots, m$.
- (8) **else**
- (9) Set a linear regression model with response Y_j and predictors Y_b 's whose corresponding coefficients ϕ_{jb} are already included in the joint model. Here $b \in \{j + 1, \dots, m\}$. Compute Rao statistic for adding one predictor among Y_{j+1}, \dots, Y_m whose corresponding ϕ_{jk} 's are not in this linear model.
- (10) **end if**
- (11) **end for**
- (12) Among all ϕ_{jk} 's not in the joint model, add one with the maximum Rao value. Denote this by ϕ_{lb} for $l = 1, \dots, m - 1$ and $b = l + 1, \dots, m$.
- (13) Update $\widehat{\phi}_l, \widehat{\tau}_l$ using (8) as well as \widehat{K} and compute BIC.
- (14) **end while**
- (15) **while** BIC decreases **do**
- (16) Compute Wald statistics for all ϕ in the current model.
- (17) Delete one $\phi_{l'b'}$ with the minimum Wald.
- (18) Update $\widehat{\phi}_{l'}, \widehat{\tau}_{l'}$ using (8) as well as \widehat{K} and compute BIC.
- (19) **end while**
- (20) The optimal model is chosen by the minimum BIC.

ALGORITHM 2: Variable selection in the precision matrix estimation.

- (1) **repeat**
- (2) Update $\widehat{\beta}(G)$ by (5) and compute $V(\widehat{\beta}(G))$ with \widehat{K} from Algorithm 2.
- (3) Update $\widehat{\tau}, \widehat{\phi}$ and \widehat{K} using $V(\widehat{\beta}(G))$ obtained from the previous step.
- (4) **until** the log-likelihood function converges.
- (5) **while** BIC decreases **do**
- (6) Compute Wald statistic of deleting each β_j^p and ϕ_{jb} in the current model.
- (7) Delete one variable (i.e., parameter) with the minimum Wald statistic.
- (8) **repeat**
- (9) Update $\widehat{\beta}(G)$ by (5) and compute $V(\widehat{\beta}(G))$ with \widehat{K} from the previous step.
- (10) Update $\widehat{\tau}, \widehat{\phi}$ and \widehat{K} using $V(\widehat{\beta}(G))$ obtained from the previous step.
- (11) **until** the log-likelihood function converges.
- (12) **end while**
- (13) The optimal model is chosen by BIC.

ALGORITHM 3: Variable selection in the joint estimation.

matrix and true matrix are benchmarked by the Steins loss function:

$$\delta_{Stein}(\mathbf{K}, \widehat{\mathbf{K}}) = \text{tr}(\mathbf{K}\widehat{\mathbf{K}}^{-1}) - \log|\mathbf{K}\widehat{\mathbf{K}}^{-1}| - m, \quad (10)$$

where $\widehat{\mathbf{K}}$ is an estimate of the true precision matrix \mathbf{K} and $|\cdot|$ denotes the determinant of a matrix. The Frobenius norm of difference between \mathbf{K} and $\widehat{\mathbf{K}}$, denoted by $\|\Delta\|_F$, where $\Delta = \mathbf{K} - \widehat{\mathbf{K}}$, is also considered. In addition to the Steins loss function, to measure how efficiently our model recovers

the true conditional dependent relationship among genes, specificity (SPE) and sensitivity (SEN) are used, as defined by

$$\begin{aligned} \text{SPE} &= \frac{\text{TN}}{\text{TN} + \text{FP}}, \\ \text{SEN} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \end{aligned} \quad (11)$$

where TN, TP, FN, and FP are the numbers of true negatives, true positives, false negatives, and false positives with regard to off-diagonal elements of a precision matrix. Here, we treat a nonzero entry of a precision matrix as "positive." To combine sensitivity and specificity, Youden's index (=SPE + SEN - 1) is

TABLE 1: Six scenarios for small-scale experimental study.

Model	m	p	$\mathbb{P}(k_{ij} \neq 0)$	$\mathbb{P}(\beta_j^l \neq 0)$
1	10	10	$2/m$	$3.5/p$
2	20	10	$2/m$	$3.5/p$
3	40	10	$2/m$	$3.5/p$
4	20	20	$2/m$	$4/p$
5	30	30	$2/m$	$4/p$
6	40	40	$2/m$	$4/p$

TABLE 2: Three scenarios for large-scale experimental study.

Model	m	p	$\mathbb{P}(k_{ij} \neq 0)$	$\mathbb{P}(\beta_j^l \neq 0)$
7	100	100	$2/m$	$3/p$
8	200	200	$2.5/m$	$15/p$
9	400	200	$1.5/m$	$20/p$

used. The smaller values of δ_{Stein} and $\|\Delta\|_F$ are better, whereas the larger values of SPE, SEN, and Youden are better.

Inspired by Yin and Li [4], we generate simulation data sets in the form of eQTL data sets such that nonzero entries of a precision matrix, commonly called a link (or edge), are randomly assigned with probability c_1/m , where m is the number of genes and c_1 is some positive constant. For a link generated at the (i, j) th entry of the precision matrix, denoted by k_{ij} , the corresponding element is sampled from the uniform distribution over $[-1, 0.5] \cup [0.5, 1]$. For each row, off-diagonal elements are divided by the sum of their absolute values multiplied by 1.5. And we obtain the true precision matrix K by symmetrizing and setting diagonals as 1. To create the $p \times m$ regression coefficients matrix β , we first generate a $p \times m$ indicator matrix that has 1 as entry with probability c_2/p for some positive constant c_2 . If the (l, j) th element of this indicator matrix is 1, β_j^l is randomly generated from $Unif([d_m, 1] \cup [-1, -d_m])$, where d_m is the smallest absolute value of K generated.

Producing β and K , we generate SNPs, $X = (X_1, \dots, X_p)$ with $X_l \sim \text{Bernoulli}(1, 0.5)$ for $l = 1, \dots, p$. Finally, we simulated gene expressions by generating y from the multivariate normal distribution given x , $Y | X \sim N(X\beta, K^{-1})$. We generate a data set of N i.i.d. random vectors (X, Y) , and simulations are repeated 50 times. In Table 1, we outline the six simulation scenarios of small-scale setup. Table 2 contains the three simulation scenarios of large-scale setup.

The small-scale simulation results in Table 3 suggest that the JMCCM produces better estimates than all other methods across all six models. Due to computational issues, we drop some results of JML from Table 3. By comparing the results of model 3 with 6 and 2 with 4, when $N = 1000$ and the number of genes is fixed, we can see that JMCCM and CAPME show less changes in δ_{Stein} and $\|\Delta\|_F$ than they appear in the SCGGM as the number of SNPs increases. This indicates that JMCCM and CAPME are less subject to the increment of the number of SNPs than SCGGM, when modest size of genes is involved. Estimation performance of CAPME seems to be affected by the number of genes more easily than JMCCM because Stein loss and Frobenius norm of CAPME for models

1, 2, and 3 with fixed p increase more rapidly than those of JMCCM do. For identifying structures of the precision matrix, JMCCM surpasses SCGGM in discovering nonzero elements (higher sensitivity) as complexity of the model rises. This is possibly due to the fact that SCGGM tends to produce sparse estimates more than the true precision matrix. Higher sensitivity implies that our proposed model is less likely to miss the influential conditional dependency among genes. While CAPME and JML score near 1 in sensitivity, they mark poor number in specificity because 5-fold cross-validation for CAPME and validation approach for JML with default setting for tuning parameters selection tend to choose small ones, leading to dense estimates. JMCCM produces higher Youden's index across all simulation scenarios compared to all other methods, and the performance gap between JMCCM and others is increasingly widened as the models increase in sample size and complexity. The results for the large-scale simulations (models 7–9) are summarized in Table 4 and Table S3. Due to long computing time, the results of CAPME for models 8 and 9 are not reported. Overall, the results are consistent with the small-scale simulations. The gap between JMCCM and SCGGM in estimation performance widened as the complexity of the model increases (Table 4).

We also simulate SNPs to mimic the linkage disequilibrium (LD) which is known to be a common phenomenon in DNA sequence and to assess the performance of our approach. We randomly generate two groups of SNPs which lie on LD, each of which contains LD block including 10 correlated SNPs such that correlation is greater than 0.9. Together with these SNPs, 10 SNPs are generated independently from Bernoulli trials with probability of 0.5 for a total of 20 SNPs and 20 genes: $m = 20$ and $p = 20$. Table 5 presents the simulation results using these data of sample size $N = 500$ and $N = 1000$ with 50 repetitions. Compared to the results of model 4 in Table 3, which is based on 20 independent SNPs with $N = 500$, LD is found to have little impact on all methods, except for slight decreases in Stein loss and Frobenius norm of SCGGM.

5. eQTL Analysis of Yeast Data

In this section, we apply the proposed algorithm to genomic eQTL (i.e., expression quantitative trait loci) data in order to examine whether the proposed method effectively recovers true dependency of gene expressions, which builds on known molecular mechanisms. To this end, we collect a set of yeast data [22], including polymorphic genotypes and mRNA expressions. The yeast data have been widely applied to elucidate the biological interactions between nucleotide polymorphisms and their responding genes (e.g., perturbation effects) [23–26]. Thus, our primary goal is to identify conditional dependency among genes with an adjustment of SNPs perturbations to each gene expression level.

The data sets are collected for two yeast parent strains, BY4716 (BY) and RM11-1a (RM), and their 112 segregants. We obtain SNPs for 1,260 loci to the exclusion of the redundant SNPs observed in neighboring genetic regions and leave 3,684 expression genes after screening out genes of missing more than 5%. In order to validate whether or

TABLE 3: Comparisons of the performance of JMCCM with SCGGM, CAPME, and JML for models 1–6. Standard errors are presented in parenthesis.

Model	Method	δ_{Stein}	$\ \Delta\ _F$	SPE	SEN	Youden
$N = 500$						
1	JMCCM	0.139 (0.004)	0.519 (0.008)	0.864 (0.010)	0.723 (0.014)	0.587 (0.012)
	SCGGM	0.171 (0.007)	0.556 (0.010)	0.865 (0.030)	0.624 (0.027)	0.506 (0.014)
	CAPME	0.113 (0.003)	0.512 (0.008)	0.002 (0.001)	1.000 (0.000)	0.002 (0.001)
	JML	0.099 (0.003)	0.450 (0.005)	0.160 (0.026)	0.980 (0.005)	0.140 (0.024)
$N = 500$						
2	JMCCM	0.241 (0.005)	0.677 (0.009)	0.915 (0.004)	0.801 (0.007)	0.716 (0.006)
	SCGGM	0.349 (0.008)	0.821 (0.012)	0.661 (0.029)	0.882 (0.016)	0.542 (0.017)
	CAPME	0.430 (0.007)	1.028 (0.011)	0.005 (0.001)	1.000 (0.000)	0.005 (0.001)
	JML	0.232 (0.004)	0.670 (0.008)	0.210 (0.009)	0.980 (0.003)	0.200 (0.009)
$N = 500$						
3	JMCCM	0.705 (0.011)	1.145 (0.009)	0.947 (0.002)	0.653 (0.006)	0.600 (0.005)
	SCGGM	1.158 (0.034)	1.506 (0.032)	0.716 (0.036)	0.706 (0.029)	0.422 (0.010)
	CAPME	1.703 (0.012)	2.217 (0.015)	0.005 (0.000)	1.000 (0.000)	0.005 (0.000)
	JML	-	-	-	-	-
$N = 500$						
4	JMCCM	0.239 (0.006)	0.678 (0.011)	0.910 (0.004)	0.802 (0.007)	0.712 (0.007)
	SCGGM	0.402 (0.010)	0.869 (0.011)	0.825 (0.031)	0.772 (0.017)	0.597 (0.019)
	CAPME	0.447 (0.007)	1.081 (0.010)	0.004 (0.001)	1.000 (0.000)	0.004 (0.001)
	JML	0.284 (0.005)	0.793 (0.008)	0.110 (0.006)	1.000 (0.001)	0.100 (0.006)
$N = 500$						
5	JMCCM	0.463 (0.007)	0.953 (0.007)	0.941 (0.003)	0.731 (0.007)	0.672 (0.006)
	SCGGM	0.743 (0.011)	1.188 (0.010)	0.978 (0.002)	0.567 (0.007)	0.545 (0.007)
	CAPME	0.983 (0.008)	1.683 (0.012)	0.005 (0.001)	1.000 (0.000)	0.005 (0.001)
	JML	0.524 (0.524)	1.074 (0.009)	0.140 (0.006)	0.980 (0.002)	0.120 (0.006)
$N = 500$						
6	JMCCM	0.711 (0.010)	1.153 (0.010)	0.951 (0.002)	0.639 (0.006)	0.590 (0.005)
	SCGGM	1.126 (0.011)	1.438 (0.007)	0.988 (0.001)	0.440 (0.004)	0.428 (0.004)
	CAPME	1.781 (0.013)	2.406 (0.013)	0.006 (0.000)	0.999 (0.000)	0.006 (0.000)
	JML	-	-	-	-	-
$N = 1000$						
1	JMCCM	0.068 (0.002)	0.364 (0.006)	0.814 (0.007)	0.886 (0.009)	0.700 (0.009)
	SCGGM	0.139 (0.010)	0.483 (0.018)	0.888 (0.024)	0.668 (0.031)	0.556 (0.016)
	CAPME	0.055 (0.002)	0.346 (0.006)	0.003 (0.001)	1.000 (0.000)	0.003 (0.001)
	JML	0.046 (0.001)	0.301 (0.005)	0.150 (0.015)	1.000 (0.001)	0.150 (0.015)
$N = 1000$						
2	JMCCM	0.117 (0.003)	0.466 (0.006)	0.894 (0.003)	0.898 (0.005)	0.793 (0.004)
	SCGGM	0.168 (0.007)	0.550 (0.010)	0.652 (0.018)	0.960 (0.006)	0.612 (0.016)
	CAPME	0.215 (0.003)	0.696 (0.007)	0.003 (0.001)	1.000 (0.000)	0.003 (0.001)
	JML	0.119 (0.002)	0.459 (0.006)	0.320 (0.120)	0.990 (0.003)	0.310 (0.012)
$N = 1000$						
3	JMCCM	0.321 (0.006)	0.766 (0.007)	0.916 (0.002)	0.808 (0.004)	0.724 (0.003)
	SCGGM	0.645 (0.015)	1.107 (0.011)	0.674 (0.021)	0.847 (0.024)	0.521 (0.008)
	CAPME	0.824 (0.006)	1.405 (0.007)	0.005 (0.000)	1.000 (0.000)	0.005 (0.000)
	JML	-	-	-	-	-
$N = 1000$						
4	JMCCM	0.115 (0.003)	0.467 (0.006)	0.893 (0.003)	0.904 (0.006)	0.797 (0.005)
	SCGGM	0.286 (0.009)	0.745 (0.012)	0.786 (0.028)	0.822 (0.019)	0.609 (0.015)
	CAPME	0.215 (0.003)	0.703 (0.006)	0.003 (0.001)	1.000 (0.000)	0.003 (0.001)
	JML	0.123 (0.002)	0.486 (0.005)	0.160 (0.006)	1.000 (0.002)	0.150 (0.005)

TABLE 3: Continued.

Model	Method	δ_{Stein}	$\ \Delta\ _F$	SPE	SEN	Youden
$N = 1000$						
5	JMCCM	0.217 (0.004)	0.643 (0.006)	0.924 (0.002)	0.867 (0.005)	0.792 (0.004)
	SCGGM	0.671 (0.011)	1.141 (0.001)	0.995 (0.002)	0.542 (0.009)	0.536 (0.008)
	CAPME	0.473 (0.004)	1.065 (0.007)	0.004 (0.000)	1.000 (0.000)	0.004 (0.000)
	JML	-	-	-	-	-
$N = 1000$						
6	JMCCM	0.331 (0.005)	0.781 (0.006)	0.926 (0.002)	0.789 (0.005)	0.715 (0.004)
	SCGGM	1.024 (0.010)	1.380 (0.006)	0.998 (0.000)	0.414 (0.003)	0.412 (0.003)
	CAPME	0.844 (0.006)	1.466 (0.008)	0.005 (0.000)	0.999 (0.000)	0.004 (0.000)
	JML	-	-	-	-	-

TABLE 4: Comparisons of the performance of JMCCM with SCGGM and CAPME for models 7–9. Standard errors are presented in parenthesis.

Model	Method	δ_{Stein}	$\ \Delta\ _F$	SPE	SEN	Youden
$N = 500$						
7	JMCCM	2.234 (0.021)	2.071 (0.011)	0.969 (0.001)	0.638 (0.004)	0.608 (0.003)
	SCGGM	2.670 (0.017)	2.160 (0.007)	0.994 (0.000)	0.512 (0.002)	0.506 (0.002)
	CAPME	12.914 (0.038)	9.257 (0.028)	0.002 (0.000)	1.000 (0.000)	0.002 (0.000)
$N = 500$						
8	JMCCM	6.560 (0.073)	3.669 (0.028)	0.982 (0.001)	0.475 (0.004)	0.457 (0.003)
	SCGGM	10.687 (0.035)	4.318 (0.006)	0.925 (0.000)	0.430 (0.002)	0.355 (0.002)
	CAPME	-	-	-	-	-
$N = 500$						
9	JMCCM	16.703 (0.254)	5.998 (0.062)	0.986 (0.001)	0.466 (0.003)	0.452 (0.002)
	SCGGM	34.025 (0.722)	7.277 (0.048)	0.874 (0.005)	0.453 (0.003)	0.328 (0.003)
	CAPME	-	-	-	-	-

TABLE 5: Comparisons of the performance of JMCCM with SCGGM and CAPME for the experiment with highly correlated SNPs. 20 genes (m) and 20 SNPs (p) are involved. Standard errors are presented in parenthesis.

Method	δ_{Stein}	$\ \Delta\ _F$	SPE	SEN	Youden
$N = 500$					
JMCCM	0.239 (0.006)	0.678 (0.011)	0.910 (0.004)	0.802 (0.007)	0.712 (0.007)
SCGGM	0.402 (0.010)	0.869 (0.011)	0.825 (0.031)	0.772 (0.017)	0.597 (0.019)
CAPME	0.447 (0.007)	1.081 (0.010)	0.004 (0.001)	1.000 (0.000)	0.0004 (0.001)
$N = 1000$					
JMCCM	0.115 (0.003)	0.467 (0.006)	0.893 (0.003)	0.904 (0.006)	0.797 (0.005)
SCGGM	0.286 (0.009)	0.745 (0.012)	0.786 (0.028)	0.822 (0.019)	0.609 (0.015)
CAPME	0.215 (0.003)	0.703 (0.006)	0.003 (0.001)	1.000 (0.000)	0.003 (0.001)

not the estimated gene network is consistent with unknown biological knowledge, we take the true signaling pathway for comparison. Out of 3,684 yeast genes, we purposely focus on a set of 64 genes that are ascertained in the cell cycle yeast pathway available in the KEGG database [24]. Together with 1,260 SNPs as predictors, we construct a gene network model of 64 genes. JMCCM selects 222 nonzero elements of the precision matrix and leaves nonzero regression coefficients, for a total of 111 links among genes. A total of 489 regression coefficients of SNPs over gene expression levels are included in the final model. Figure 2 displays the conditional gene

network estimated by the proposed joint model. A pair of two genes is linked if an off-diagonal element of the estimated precision matrix is nonzero, and so 51 genes are shown to be linked and assemble in a module.

Given this estimated gene network, we notice that the gene network is found to be concordant with the true cell cycle pathway. For instance, according to the true cell cycle pathway in Figure 1, MCM3 is linked to ORC3, ORC5, MCM7, and MCM4. MCM5 is connected to TAH11 and ORC1. PHO4 is closely linked to CLN2 and SIC1, whose molecular function is related to MAP kinase orthologs (i.e., MAPK pathway

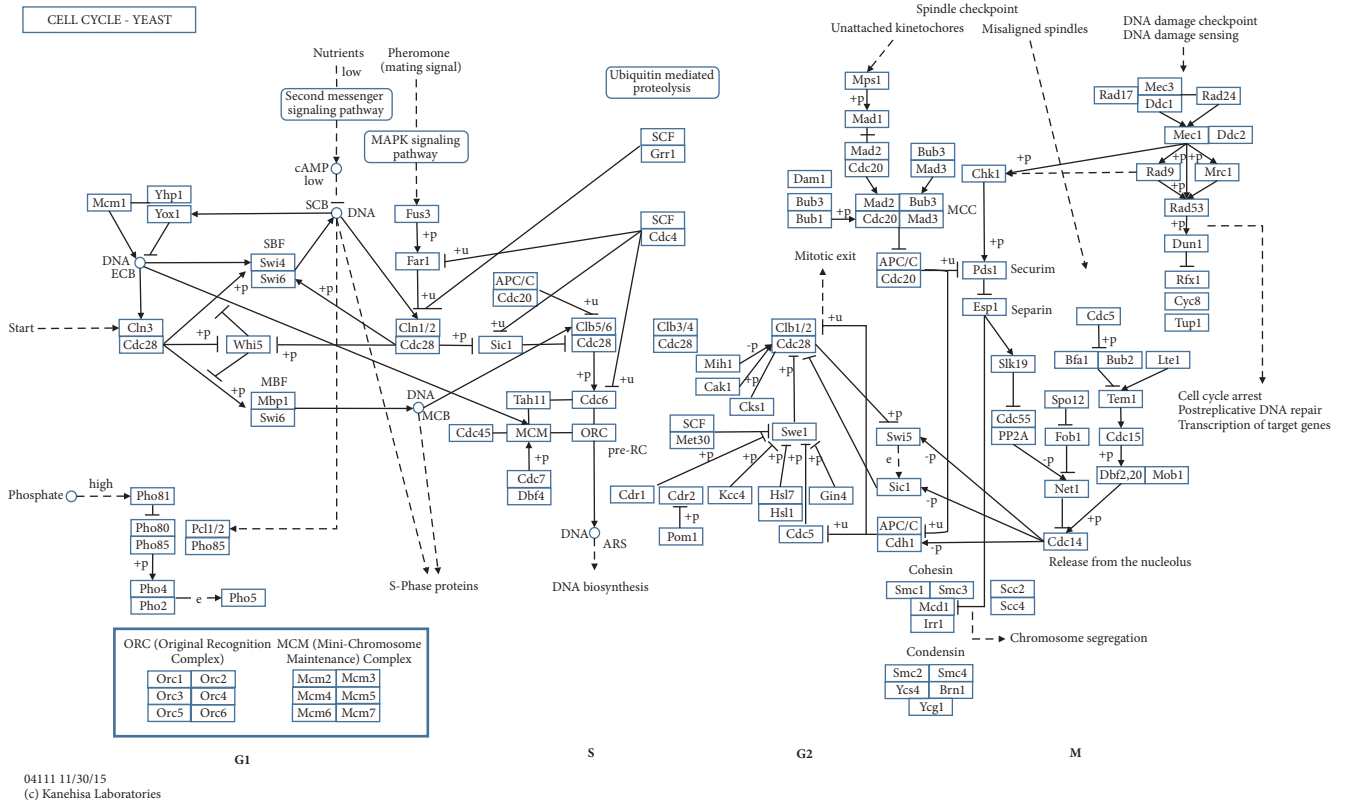


FIGURE 1: The yeast cell cycle pathway from the KEGG database. Source: http://www.kegg.jp/kegg-bin/highlight_pathway?scale=1.0&map=map041111&keyword=cell%20cycle.

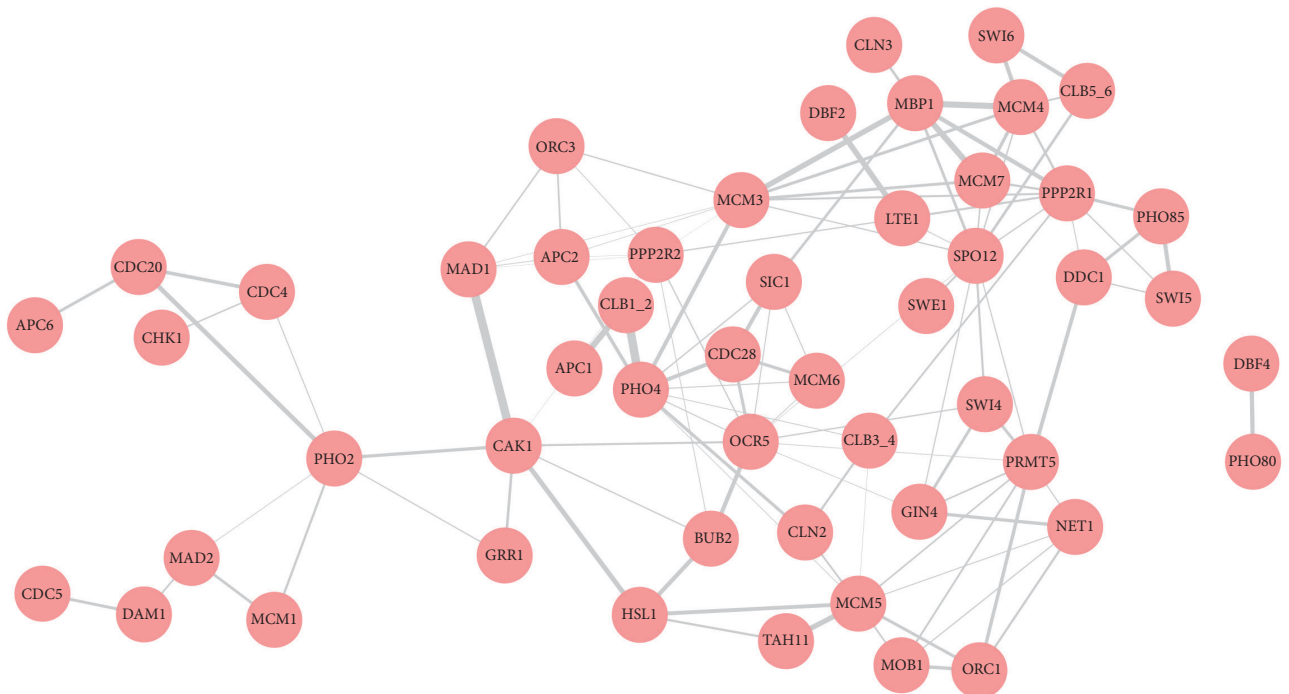


FIGURE 2: The gene network of the yeast data related to the cell cycle pathway via the JMCCM.

TABLE 6: Gene ontology (GO) enrichment analysis over the genes in detected module from the JMCCM.

Module	Module size	GO category	GO enrichment p value
Module 1	94	Carboxylic acid metabolic process	0.00101
		Carboxypeptidase activity	0.00904
		Catabolic process	0.00101
		Cellular catabolic process	0.00101
		Exopeptidase activity	0.00904
		Metalloexopeptidase activity	0.00904
		Metallopeptidase activity	0.00904
		Organic acid metabolic process	0.00101
		Peptidase activity	0.00904

genes). It is interesting to note that CLN3 and SWI4 are mutually connected together, linking to their downstream gene MBP1. Hence, this undirected graph contributes to recovering lots of links among the 64 genes of the pathway.

With regard to genetic variant effects, JMCCM is shown to effectively identify some of the well-known direct genetic perturbations. Gene expressions are regulated by some genetic variants, which, unless otherwise taken into account, may falsely capture the interplay of genes in a network. Our founding includes Clb-specific Cdk inhibitor (SIC1) as influencing the molecular interface between cyclin-Cdk complexes (e.g., binding to and blocking Cdk1/Clb activity, ultimately to maneuver the timing of DNA replication (see Table S1 in Supplementary Materials; [27])). In addition, our gene networks demonstrate that SIC1 is strongly perturbed by CLB3, while SCGGM did not detect any perturbation effects. More interestingly, many previous experiments validated the association between SIC1 and CLB3 to uncover the underlying mechanism of the cell cycle [28–30]. Clearly, this implies that JMCCM does a better job in accounting for SNPs perturbation as compared to SCGGM.

Moreover, the gene module detected from JMCCM is shown in Table 6. Pertaining to these gene modules, we hypothesize if gene modules, each containing hub genes and their neighboring genes, are enriched with common biological processes or not. To test this hypothesis, we conduct the gene ontology (GO) enrichment analysis [31] over the detected gene network modules (see Table S2 in Supplementary Materials) from both JMCCM and SCGGM using Fisher’s exact tests. Table 6 demonstrates that the proposed method outstandingly performs detecting modules biologically associated with many molecular processes, while none is detected by SCGGM. More importantly, the pathway of the organic acid metabolic process enriched in gene module 1 is also reported and validated on the basis of the network modules constructed with large-scale integration of yeast data in Zhu *et al.* [32].

Putting all things together, we conclude that the proposed method facilitates recovering the true SNPs perturbations in the midst of gene regulations and elucidating the underlying interplay of gene interactions. These fortes of the proposed algorithm are favored for reinforcing *a priori* biological knowledge and address a novel hypothesis related to clinical and translational potential.

6. Conclusion and Discussion

In this paper, we propose JMCCM to efficiently identify conditional dependent structures of gene expressions with adjustments to perturbation effects of SNPs. Contrary to the existing conditional graphical models, the precision matrix commonly used to reveal the true relationship among genes is parameterized via the modified Cholesky decomposition. The maximum likelihood estimates of the precision matrix were computed, while variable selection of SNPs and Cholesky factors are carried out separately and jointly by the GCV and BIC criterion. From experimental studies, it is clearly shown that JMCCM performs better than the existing penalization methods. Besides, JMCCM in the application to yeast cell cycle data successfully recovers many parts of the cell cycle pathway with adjustments of SNPs to each gene expressions level. Notably, the model entails the estimation of precision matrix, of which components are assumed to be constant. So, in the future, we may relax this somewhat strong assumption in the way that the model can parametrize over τ and ϕ in pursuit of more accurate estimation [13]. We leave this for next study.

Appendix

A. Positive Definiteness of K

The matrix representation of the modified Cholesky decomposition (2) is given by

$$C(\phi) = \begin{bmatrix} 1 & -\phi_{12} & \cdots & -\phi_{1m} \\ & 1 & \cdots & -\phi_{2m} \\ & & \ddots & \vdots \\ & & & 1 \end{bmatrix}, \quad (A.1)$$

$$D(\tau) = \begin{bmatrix} \tau_1 & & & \\ & \tau_2 & & \\ & & \ddots & \\ & & & \tau_m \end{bmatrix}.$$

For simplicity, write $C = C(\phi)$ and $D = D(\tau)$. For any nonzero vector $u \in \mathbb{R}^m$,

$$\begin{aligned} u^\top K u &= u^\top C^\top D C u = (C u)^\top D (C u) \\ &= \sum_{j=1}^m ((C u)_j)^2 \tau_j > 0. \end{aligned} \quad (\text{A.2})$$

Since $\tau_j > 0$ for all $j = 1, \dots, m$, $\sum_{j=1}^m ((C u)_j)^2 \tau_j = 0$ if and only if $C u = 0$, which cannot happen here because u is a nonzero vector. Thus K is a positive definite matrix.

B. A Derivation of the Log-Likelihood Function

The probability density function of (1) is given by

$$\begin{aligned} f(y | x; \theta) &= \frac{|K|^{1/2}}{(2\pi)^{m/2}} \exp \left\{ -\frac{1}{2} (y - \mu(x))^\top K (y - \mu(x)) \right\}, \end{aligned} \quad (\text{B.1})$$

where $\theta = (\beta, \phi, \tau)$. The log-likelihood function corresponding to JMCCM can be derived as follows:

$$\begin{aligned} \ell(\theta) &= \log \sum_{i=1}^N f(y^i | x^i; \theta) \\ &= \frac{-mN}{2} \log(2\pi) + \frac{N}{2} \log |K| \\ &\quad - \frac{1}{2} \sum_{i=1}^N (y^i - \mu^i)^\top K (y^i - \mu^i) \\ &= \frac{N}{2} \left\{ -m \log(2\pi) + \sum_{j=1}^m \log \tau_j - \text{tr}(KV(\beta)) \right\} \end{aligned} \quad (\text{B.2})$$

where the last equality comes from $|C| = |C^\top| = 1$, $|D| = \tau_1 \tau_2 \cdots \tau_m$, and

$$\log |K| = \log |C^\top| |D| |C| = \log |D| = \sum_{j=1}^m \log \tau_j. \quad (\text{B.3})$$

C. Derivations of MLE $\hat{\beta}$, $\hat{\phi}$, and $\hat{\tau}$

Using notations presented in Section 2.2, optimization problem for $\hat{\beta}(G)$ is expressed as

$$\max \left\{ \frac{1}{2} \left[\log |K| - (Y - X\beta(G))^\top W (Y - X\beta(G)) \right] \right\}. \quad (\text{C.1})$$

Denote the score function of $\beta(G)$ by $S(\beta(G))$. Observe that

$$\begin{aligned} S(\beta(G)) &= \frac{\partial}{\partial \beta(G)} \left\{ -\frac{1}{2} \left(Y^\top W Y \right. \right. \\ &\quad \left. \left. - 2\beta(G)^\top X^\top W Y + \beta(G)^\top X^\top W X \beta(G) \right. \right. \\ &\quad \left. \left. - \sum_{j=1}^m \log \tau_j \right) \right\} = X^\top W Y - X^\top W X \beta(G) \end{aligned} \quad (\text{C.2})$$

By solving the normal equation $S(\beta(G)) = 0$, we have (5).

To derive (6), let $C(\phi_j) = (0, \dots, 0, 1, \phi_{j,j+1}, \dots, \phi_{j,m})$, j th row of C . Then we can see that $C^\top D C = \sum_{j=1}^m \tau_j C(\phi_j) C(\phi_j)^\top$. The log-likelihood function (3) can be expressed as the sum of $\ell_j(\theta_j)$ as follows:

$$\begin{aligned} \ell(\theta) &= \frac{N}{2} \left\{ -\log(2\pi) + \sum_{j=1}^m \log \tau_j - \text{tr}(KV(\beta)) \right\} \\ &= \frac{N}{2} \left\{ -m \log(2\pi) + \sum_{j=1}^m \log \tau_j \right. \\ &\quad \left. - \text{tr} \left(\sum_{j=1}^m \tau_j C(\phi_j) C(\phi_j)^\top V(\beta) \right) \right\} \\ &= \frac{N}{2} \left\{ -m \log(2\pi) + \sum_{j=1}^m \log \tau_j \right. \\ &\quad \left. - \sum_{j=1}^m \tau_j C(\phi_j)^\top V(\beta) C(\phi_j) \right\} \\ &= \frac{N}{2} \left\{ -m \log(2\pi) + \sum_{j=1}^m \log \tau_j \right. \\ &\quad \left. - \sum_{j=1}^m \tau_j (V_{jj} - 2\phi_j^\top V_{j,21} + \phi_j^\top V_{j,22} \phi_j) \right\} \\ &= \sum_{j=1}^m \ell_j(\theta_j). \end{aligned} \quad (\text{C.3})$$

Finally, for any given $j = 1, \dots, m-1$, the gradient of log-likelihood function with respect to ϕ_j and τ_j is given by

$$\begin{aligned} \nabla \ell_j(\theta_j) &= \left[\begin{array}{c} N \tau_j (V_{j,21} - V_{j,22} \phi_j) \\ \frac{N}{2} \left\{ \frac{1}{\tau_j} - (V_{jj} - 2\phi_j^\top V_{j,21} + \phi_j^\top V_{j,22} \phi_j) \right\} \end{array} \right]. \end{aligned} \quad (\text{C.4})$$

Then (8) is obtained by solving $\nabla \ell_j(\theta_j) = 0$ with respect to ϕ_j and τ_j .

Data Availability

The source of the eQTL data analyzed in this paper is [22]. An R package JMCCM and the eQTL data set are available at author's webpage: <https://sites.google.com/site/sunghwan-shome/>.

Disclosure

A part of results from the earlier version of this research were presented in the poster session at 2016 Spring Seminar held by the Korean Statistical Society.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

Acknowledgments

This research is supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2015R1D1A1A01057747 and NRF-2017R1C1B5017528).

Supplementary Materials

Table S1 shows the estimated coefficient matrix from JMCCM applied to the genes of cell cycle pathway. The columns refer to genes and the rows to SNPs. Table S2 contains the gene lists of each module estimated from both SCGGM and JMCCM. Table S3 shows results of simulation with a large data set of $m = 1000$ (genes), $p = 4000$ (SNPs), and $N = 500$. (*Supplementary Materials*)

References

- [1] H. Ye and J. Pan, "Modelling of covariance structures in generalised estimating equations for longitudinal data," *Biometrika*, vol. 93, no. 4, pp. 927–941, 2006.
- [2] H. Chun, M. Chen, B. Li, and H. Zhao, "Joint conditional Gaussian graphical models with multiple sources of genomic data," *Frontiers in Genetics*, vol. 4, article 294, 2013.
- [3] B. Li, H. Chun, and H. Zhao, "Sparse estimation of conditional graphical models with application to gene networks," *Journal of the American Statistical Association*, vol. 107, no. 497, pp. 152–167, 2012.
- [4] J. Yin and H. Li, "A sparse conditional Gaussian graphical model for analysis of genetical genomics data," *The Annals of Applied Statistics*, vol. 5, no. 4, pp. 2630–2650, 2011.
- [5] L. Zhang and S. Kim, "Learning gene networks under SNP perturbations using eQTL datasets," *PLoS Computational Biology*, vol. 10, no. 2, Article ID e1003420, 2014.
- [6] X. He, Z.-Y. Zhu, and W.-K. Fung, "Estimation in a semiparametric model for longitudinal data with unspecified dependence structure," *Biometrika*, vol. 89, no. 3, pp. 579–590, 2002.
- [7] D. Xu, Z. Zhang, and L. Wu, "Joint Variable Selection of Mean-Covariance Model for Longitudinal Data," *Open Journal of Statistics*, vol. 03, no. 01, pp. 27–35, 2013.
- [8] X. Zheng, W. K. Fung, and Z. Zhu, "Variable selection in robust joint mean and covariance model for longitudinal data analysis," *Statistica Sinica*, vol. 24, no. 2, pp. 515–531, 2014.
- [9] K. A. Sohn and S. Kim, "Joint Estimation of Structured Sparsity and Output Structure in Multiple-Output Regression via Inverse-Covariance Regularization," in *Proceedings of the 15th International Conference on Artificial Intelligence and Statistics (AISTATS)*, vol. XX, pp. 1081–1089, 2012.
- [10] M. Morloy, C. M. Molony, T. M. Weber et al., "Genetic analysis of genome-wide variation in human gene expression," *Nature*, vol. 430, no. 7001, pp. 743–747, 2004.
- [11] B. E. Stranger, M. S. Forrest, A. G. Clark et al., "Genome-wide associations of gene expression variation in humans," *PLoS Genetics*, vol. 1, no. 6, p. e78, 2005.
- [12] T. T. Cai, H. Li, W. Liu, and J. Xie, "Covariate-adjusted precision matrix estimation with an application in genetical genomics," *Biometrika*, vol. 100, no. 1, pp. 139–156, 2013.
- [13] M. Pourahmadi, "Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation," *Biometrika*, vol. 86, no. 3, pp. 677–690, 1999.
- [14] S. r. Hojsgaard, D. Edwards, and S. Lauritzen, *Graphical models with R*, Springer, New York, NY, USA, 2012.
- [15] J.-Y. Koo, "Spline estimation of discontinuous regression functions," *Journal of Computational and Graphical Statistics*, vol. 6, no. 3, pp. 266–284, 1997.
- [16] C. Kooperberg, S. Bose, and C. J. Stone, "Polychotomous Regression," *Journal of the American Statistical Association*, vol. 92, no. 437, pp. 117–127, 1997.
- [17] C. R. Rao, *Linear Methods of Statistical Induction and their Applications*, Wiley, New York, NY, USA, 2nd edition, 1973.
- [18] J. H. Friedman, "Multivariate adaptive regression splines," *The Annals of Statistics*, vol. 19, no. 1, pp. 1–67, 1991.
- [19] M. Pourahmadi, *High-Dimensional Covariance Estimation*, Wiley Series in Probability and Statistics, John Wiley & Sons, Hoboken, NJ, USA, 2013.
- [20] N. Verzelen, "Adaptive estimation of covariance matrices via Cholesky decomposition," *Electronic Journal of Statistics*, vol. 4, pp. 1113–1150, 2010.
- [21] J. H. Jhong, J. Lee, S. Kim, and J. Y. Koo, "Joint Modeling for Mean Vector and Covariance Estimation with l_1 -Penalty," *Quantitative Bio-Science*, vol. 36, no. 1, pp. 33–38, 2017.
- [22] R. B. Brem and L. Kruglyak, "The landscape of genetic complexity across 5,700 gene expression traits in yeast," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 5, pp. 1572–1577, 2005.
- [23] S.-S. C. Huang and E. Fraenkel, "Integrating proteomic, transcriptional, and interactome data reveals hidden components of signaling and regulatory networks," *Science Signaling*, vol. 2, no. 81, p. ra40, 2009.
- [24] M. Kanehisa, S. Goto, M. Furumichi, M. Tanabe, and M. Hirakawa, "KEGG for representation and analysis of molecular networks involving diseases and drugs," *Nucleic Acids Research*, vol. 38, no. 1, pp. D355–D360, 2009.
- [25] F. Markowetz, D. Kostka, O. G. Troyanskaya, and R. Spang, "Nested effects models for high-dimensional phenotyping screens," *Bioinformatics*, vol. 23, no. 13, pp. i305–i312, 2007.
- [26] E. Yeger-Lotem, L. Riva, L. J. Su et al., "Bridging high-throughput genetic and transcriptional data reveals cellular responses to alpha-synuclein toxicity," *Nature Genetics*, vol. 41, no. 3, pp. 316–323, 2009.

- [27] M. Barberis, "Sic1 as a timer of Clb cyclin waves in the yeast cell cycle - Design principle of not just an inhibitor," *FEBS Journal*, vol. 279, no. 18, pp. 3386–3410, 2012.
- [28] V. Archambault, E. J. Chang, B. J. Drapkin, F. R. Cross, B. T. Chait, and M. P. Rout, "Targeted proteomic study of the cyclin-Cdk module," *Molecular Cell*, vol. 14, no. 6, pp. 699–711, 2004.
- [29] A. Breitskreutz, H. Choi, J. R. Sharom et al., "A global protein kinase and phosphatase interaction network in yeast," *Science*, vol. 328, no. 5981, pp. 1043–1046, 2010.
- [30] N. J. Krogan, G. Cagney, H. Yu et al., "Global landscape of protein complexes in the yeast *saccharomyces cerevisiae*," *Nature*, vol. 44, p. 0, 2006.
- [31] S. Carbon, A. Ireland, C. J. Mungall et al., "AmiGO: Online access to ontology and annotation data," *Bioinformatics*, vol. 25, no. 2, pp. 288–289, 2009.
- [32] J. Zhu, B. Zhang, E. N. Smith et al., "Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks," *Nature Genetics*, vol. 40, no. 7, pp. 854–861, 2008.