QBS
Vol. 36, No. 1, pp. 33-38 (2017)
https://doi.org/10.22283/qbs.2017.36.1.33
pISSN 2288-1344 eISSN 2508-7185 **Original Article**

Quantitative Bio-Science
Official Journal of the Institute of Natural Science at Keimyung University
Available Online at https://qbs.kmu.ac.kr:442/

Joint Modeling for Mean Vector and Covariance Estimation with l_1-Penalty

Jae-Hwan Jhong[1], JungJun Lee[1], SungHwan Kim[2], Ja-Yong Koo[1,*]

[1]Department of Statistics, Korea University, Seoul 02841, Korea
[2]Department of Statistics, Keimyung University, Daegu 42601, Korea
(Received March 27, 2017; Revised May 2, 2017; Accepted May 14, 2017)

ABSTRACT

In this study, we develop a novel updating-based method for penalized estimators for the mean vector and the covariance matrix. With a linear combination of predictors, the coefficients can be estimated by maximizing a penalized log likelihood function, and using coordinate descent algorithm is used to handle the l_1-penalized function. In order to estimate the inverse covariance matrix estimation, we adopt a modified Cholesky decomposition so that to guarantee the positive definiteness of the estimators. In the genomic data analysis setting, we show that the proposed method can be efficiently used to detect the conditional independence among a group of genes, while adjusting for shared genetic effects. Simulation experiments benchmark the performance of the proposed method against another existing method.

Key words : Cholesky decomposition, Coordinate descent algorithm, Lasso, Penalized likelihood, Variable selection

1. Introduction

In recent years, the estimation of the sparse inverse covariance matrix using penalized methods, has been proposed by a number of authors. For example, Yin and Li [1] suggested the sparse conditional graphical Gaussian model (cGGM) with lasso and adaptive lasso penalty functions applied to both, the mean and the precision matrix. Furthermore, Li et al. [2] proposed a two-stage estimation structure: (1) estimating a non-sparse conditional covariance of genes by reproducing kernel Hilbert spaces of genes and makers, and then (2) using the lasso and adaptive lasso penalty to obtain sparse estimates of a precision matrix under the cGGM. Similarly, Cai et al. [3] researched a precision matrix estimation method under the consideration of covariates' effect on the matrix using the constrained l_1 optimization without making multivariate normality assumption on the error. Chun et al. [4] studied a method of inferring multiple gene networks using the cGGM in a single model framework. Zhang and Kim [5] consider sparse cGGM, in order to learn gene network structure under SNPs perturbations, while improving in terms of computing direct perturbations of gene-expression traits of SNPs.

In this study, we consider a penalized joint mean and a constant covariance model (PJM) with lasso penalties for both the estimations, the mean and the precision matrix. A modified Cholesky decomposition, inspired by Pourahmadi [6], is used to reparametrize the precision matrix, ensuring the positive definiteness of the estimates. This approach also converts the problem of precision matrix estimation into that of linear regressions [7], such that when estimating elements of the Cholesky factor, the precision matrix is ultimately equivalent to estimating the regression coefficients. Thus, variable selection techniques such as ridge and lasso [8], can be used to shrink the elements of the Cholesky factor, and identify any existing structural zeros [9]. We achieve sparsity of both, the

* Correspondence should be addressed to Dr. Ja-Yong Koo, Department of Statistics, Korea University, Seoul 02841, Korea. Tel: +82-2-3290-2240, Fax: +82-2-3290-2240, E-mail: jykoo@korea.ac.kr

mean and the precision matrix, by imposing lasso penalty on both, the mean regression coefficients and the entries of the Cholesky factor. While studies have been previously conducted on sparse covariance estimation based on Cholesky decomposition using various penalized methods [7] for example, do not consider the mean and the covariance estimation simultaneously. We also derive a formula for the maximum candidate of tuning parameter, in that all regression coefficients are shrunk to zero. For the implementation of our method, we use a coordinate descent algorithm [10], which is applied to solve the mean and the covariance optimization problem.

In numerical studies, simulation scenarios experimentally show the performance of PJM in estimation and identification of the sparse structure of the precision matrix with a finite sample. In this study, we compare our approach with the sparse conditional Gaussian graphical model (SCGGM) of Zhang and Kim [5].

The paper is structured as follows. In Section 2, we describe our PJM with the modified Cholesky decomposition and penalized maximum likelihood estimator. An updated algorithm, using a coordinate descent method and its implementation for our estimator, is presented in Section 3. Section 4 deals with simulation studies to demonstrate the performance of variable selection and the estimation of the proposed model. Concluding remarks and discussions are presented in Section 5.

2. Model

2.1 Joint model for mean vector and covariance matrix

Suppose (x, y) is a pair of a $p\times 1$ vector x and an $m\times 1$ random vector y. Assume that when the predictor value x is given, the response variable y follows the multivariate normal distribution

$$y|x\sim\mathcal{N}_m\left(\mathsf{B}^\top x,\ \Sigma\right), \tag{1}$$

where $\mathsf{B}=[\beta_1\dots\beta_m]$ is a $p\times m$ coefficient matrix and Σ is an $m\times m$ covariance matrix. We consider a modified Cholesky decomposition of the precision matrix $\Omega=\Sigma^{-1}$ to be

$$\Omega(\phi,\tau)=\mathsf{C}(\phi)^\top\mathsf{D}(\tau)\mathsf{C}(\phi),$$

where C is an upper triangular matrix with diagonal entries 1, the above-diagonal elements are negative with $\phi=(\phi_{12},\dots,\phi_{1m},\phi_{23},\dots,\phi_{2m},\dots,\phi_{m-1,m})$, and D is a diagonal matrix with diagonal entries $\tau=(\tau_1,\dots,\tau_m)$ with positive value τ_j's for $j=1,\dots,m$ [6]. Since Ω is a symmetric positive definite, it is desirable to ensure the symmetry and positive definiteness of our estimator. The parametrization with a modified Cholesky decomposition assists us to guarantee this property [6].

Given a random sample $(x^1,y^1),\dots,(x^n,y^n)$ from the model (1), our goal is to estimate the mean vector and the precision matrix by solving an optimization problem of a penalized log-likelihood function presented in Section 2.2. Considering the cGGM [1], our main goal is to identify the zero and non-zero entries of the precision matrix.

2.2 Maximum penalized log-likelihood estimator

Suppose $X=(x_1,\dots,x_p)$ is the $n\times p$ matrix and $Y=(y_1,\dots,y_m)$ is the $n\times m$ observations. We write the vectorization of the coefficient matrix B and the observation matrix Y as $\beta=\mathrm{vec}(\mathsf{B})$ and $\mathsf{Y}=\mathrm{vec}(Y)$, respectively. We also adopt the Kronecker product '$\otimes$' for notational convenience so that

$$\mathsf{X}=I_m\otimes X \text{ and } \mathsf{W}(\phi,\tau)=\Omega(\phi,\tau)\otimes I_n,$$

where I_m and I_n are $m\times m$ and $n\times n$ identity matrices, respectively.

Let $x^i=(x_1^i,\dots,x_p^i)$ and $y^i=(y_1^i,\dots,y_m^i)$. Under the likelihood framework, we can define the log-likelihood function as

$$\begin{aligned}\ell(\theta)&=\frac{n}{2}\{\log|\Omega(\phi,\tau)|-(\mathsf{Y}-\mathsf{X}\beta)^\top\mathsf{W}(\phi,\tau)(\mathsf{Y}-\mathsf{X}\beta)\} \quad (2)\\ &=\frac{n}{2}\{\log|\Omega(\phi,\tau)|-\mathrm{tr}(\Omega(\phi,\tau)\mathsf{V}(\beta))\}\end{aligned}$$

up to the constants independent of $\theta=(\beta,\phi,\tau)$, where $|A|$ denotes the determinant of the matrix A, "tr" denotes the trace operation and

$$\mathsf{V}(\beta)=\frac{1}{n}\sum_{i=1}^{n}\left(y^i-\mathsf{B}^\top x^i\right)\left(y^i-\mathsf{B}^\top x^i\right)^\top.$$

Denote by V_{jj}, $\mathsf{V}_{j,21}$ and $\mathsf{V}_{j,22}$ the (1,1), (2,1) and (2,2) components of the lower jth principal submatrix is $\mathsf{V}(\beta)_j$ of $\mathsf{V}(\beta)$, respectively. Let $\phi_j=(\phi_{j,j+1},\dots,\phi_{jm})$ for $j=1,\dots,m-1$. Note that the log-likelihood function (2) can be represented by

$$\begin{aligned}\ell(\theta)&=\frac{n}{2}\left\{-m\log 2\pi+\sum_{j=1}^{m}\log\tau_j\right.\\ &\qquad\left.+\mathrm{tr}\left(\sum_{j=1}^{m}\tau_j\,\mathsf{C}(\phi_j)\mathsf{C}(\phi_j)^\top\mathsf{V}(\beta)\right)\right\}\\ &=\frac{n}{2}\left\{-m\log 2\pi+\sum_{j=1}^{m}\log\tau_j\right.\end{aligned}$$

$$-\sum_{j=1}^{m}\tau_j\left(\mathsf{V}_{jj}-2\phi_j^{\top}\mathsf{V}_{j,21}+\phi_j^{\top}\mathsf{V}_{j,22}\phi_j\right)\Big\}$$
$$=\sum_{j=1}^{m}\ell_j(\theta_j),$$

where $\theta_j=(\beta,\phi_j,\tau_j)$ and

$$\ell_j(\theta_j)=\frac{n}{2}\Big\{-m\log 2\pi+\sum_{j=1}^{m}\log\tau_j$$
$$-\sum_{j=1}^{m}\tau_j\left(\mathsf{V}_{jj}-2\phi_j^{\top}\mathsf{V}_{j,21}+\phi_j^{\top}\mathsf{V}_{j,22}\phi_j\right)\Big\}$$

The estimate of θ is the solution to the following optimization problem for the penalized likelihood function:

$$\max\left\{\ell^{\lambda}(\theta)\equiv\ell(\theta)-\lambda_1|\beta|-\lambda_2|\phi|\right\},$$

where $\lambda_1>0$ and $\lambda_2>0$ are the two tuning parameters that control the sparsity of the estimators. We consider the l_1 as the lasso regularization for the penalty of both the mean vector and the precision matrix. The first constraint encourages sparsity in the coefficients for the mean vector, and the second encourages sparsity in the precision matrix. Finally, we denote the PJM estimator $\hat{\theta}^{\lambda}=\underset{\theta}{\operatorname{argmin}}-\ell^{\lambda}(\theta)$.

3. Implementation

Consider a coordinate descent algorithm for minimizing $-\ell^{\lambda}(\theta)$. To obtain $\hat{\theta}^{\lambda}$, we separately compute $\hat{\beta}^{\lambda}$, $\hat{\phi}^{\lambda}$, and $\hat{\tau}^{\lambda}$ in the proposed algorithm. Note that we do not penalize τ, so that $\hat{\tau}^{\lambda}$ is easily computed by the maximum likelihood estimator. Since we adopt Cholesky decomposition, minimizing each coefficient for jth row $\phi_j=(\phi_{j,j+1},\ldots,\phi_{j,m})$ can be represented by the regression problem of the response Y_j and the predictors $Y_{j+1},\ldots,Y_m$. Thus, we can consider the algorithm to update β and $\phi_1,\ldots,\phi_{m-1}$ as a l_1-penalized regression problem between some specific responses and predictors.

3.1 A coordinate descent algorithm for PJM

Consider a quadratic function q defined as

$$q(z)=\frac{b}{2}(z-c)^2 \text{ for } z\in\mathbb{R},$$

where $b>0$ and $c\in\mathbb{R}$. Let q^{λ} be a univariate penalized quadratic function given as

$$q^{\lambda}(z)=q(z)+\lambda|z| \text{ for } \lambda>0$$

and denote $z^{\lambda}=\underset{z\in\mathbb{R}}{\operatorname{argmin}}\,q^{\lambda}(z)$. Note that Then the minimizer z^{λ} of q^{λ} is then given by

$$z^{\lambda}=\mathrm{ST}\left(\mathrm{c},\frac{\lambda}{b}\right),\tag{3}$$

where the soft-thresholding operator is defined as

$$\mathrm{ST}(\mathrm{y},\lambda)=\begin{cases}\mathrm{y}-\lambda & \text{if } \mathrm{y}>\lambda\\ \mathrm{y}+\lambda & \text{if } \mathrm{y}<-\lambda\\ 0 & \text{if } |\mathrm{y}|\le\lambda\end{cases}\text{ for } \mathrm{y}\in\mathbb{R}.$$

Expanding this result, for a quadratic function Q defined on $\mathbb{R}^p$, we minimize Q^{λ} which is defined as

$$Q^{\lambda}(x)=Q(x)+\lambda\sum_{j=1}^{p}|x_j| \text{ for } x\in\mathbb{R}^p.$$

Let $x^{\lambda}=\underset{x\in\mathbb{R}^p}{\operatorname{argmin}}\,Q^{\lambda}(x)$. Note that the Hessian matrix of $\nabla^2Q(x)$ at any $x\in\mathbb{R}^p$ is a constant matrix, denoted by I.

For a current vector $\tilde{x}=(\tilde{x}_1,\ldots,\tilde{x}_p)$, denote

$$q_j(z)=Q(\tilde{x}_1,\ldots,\tilde{x}_{j-1},z,\tilde{x}_{j+1},\ldots,\tilde{x}_p) \text{ for } z\in\mathbb{R}$$

and

$$q_j^{\lambda}(z)=q_j(z)+\lambda|z|.$$

By (3), the minimizer z_j^{λ} of q_j^{λ} is given by

$$z_j^{\lambda}=\mathrm{ST}\left(z_j^0,\frac{\lambda}{\mathsf{I}_{jj}}\right),$$

where z_j^0 is the solution to $q_j'(z)=0$ and I_{jj} denote the j-th diagonal element of I. It can be observed that

$$q_j'(z)=\left[\nabla Q(\tilde{x}_1,\ldots,\tilde{x}_{j-1},z,\tilde{x}_{j+1},\ldots,\tilde{x}_p)\right]_j \text{ for } j=1,\ldots,p.$$

The coordinate descent algorithm is summarized as follows:

(1) Initialize x as $\tilde{x}=(\tilde{x}_1,\ldots,\tilde{x}_p)$.

(2) Iterating for $j=1,\ldots,p$, we update $\tilde{x}_j$ as follows:

a. solve for x_j^0 satisfying

$$\left[\nabla Q(\tilde{x}_1,\ldots,\tilde{x}_{j-1},x_j^0,\tilde{x}_{j+1},\ldots,\tilde{x}_p)\right]_j=0.$$

b. update

$$\tilde{x}_j\leftarrow\mathrm{ST}\left(\mathrm{x}_j^0,\frac{\lambda}{\mathsf{I}_{jj}}\right).$$

(3) Until convergence of Q^{λ}.

3.1.1 Compute $\hat{\beta}^{\lambda}$

The solution to the optimization problem with respect to β given ϕ and τ has the form of the l_1-penalized weighted regression problem

$$\min\left\{\frac{1}{2}(\mathsf{Y}-\mathsf{X}\beta)^{\top}\mathsf{W}(\phi,\tau)(\mathsf{Y}-\mathsf{X}\beta)+\lambda_1|\beta|\right\}.$$

According to the coordinate descent algorithm presented in Section 1, we update $\hat{\beta}^{\lambda}$ by using the score function $\mathsf{X}^{\mathsf{T}}\mathsf{W}(\phi,\tau)\mathsf{Y}-\mathsf{X}^{\mathsf{T}}\mathsf{W}(\phi,\tau)\mathsf{X}\beta$ and the Hessian matrix $\mathsf{X}^{\mathsf{T}}\mathsf{W}(\phi,\tau)\mathsf{X}$.

3.1.2 Compute $\hat{\phi}^{\lambda}$ and $\hat{\lambda}^{\lambda}$

As we use a modified Cholesky decomposition to reparameterize the precision matrix Ω, we transform the problem into $m-1$ linear regressions by each row. We run total $m-1$ l_1-penalized regressions to estimate $\phi_1,\ldots,\phi_{m-1}$. The jth Lasso regression becomes

$$\min\left\{\frac{1}{2}\left\|y_j-y_{j+1:M}\phi_j\right\|^2+\lambda_2\left|\phi_j\right|\right\} \text{ for } j=1,\ldots,m-1,$$

where $y_{j+1:m}=(y_{j+1},\ldots,y_m)$ is the $n\times(m-j)$ observation matrix for the $(j+1)$th to mth variables. When β and τ are fixed, we can update ϕ_j coordinate-wise using $\mathsf{V}(\beta)$. Finally, the maximum likelihood estimator $\hat{\tau}_j^{\lambda}$, which does not actually depend on λ, is obtained by solving $\nabla\ell_j(\theta_j)$ with respect to τ_j by

$$\frac{1}{\hat{\tau}_j^{\lambda}}=\mathsf{V}(\beta)_{jj}-\mathsf{V}_{j,12}\mathsf{V}_{j,22}^{-1}\mathsf{V}_{j,21}.$$

3.2 Complexity parameter selection

We compute the PJM estimators using the following steps. First, we designate λ_{1K_1} and λ_{2K_2} which are sufficiently large positive numbers. Second, we choose the numbers K_1 and K_2 which are the numbers of λ_1^j's and λ_2^j's, respectively. We compute the PJM estimator for the grid for the decreasing sequences $\{\lambda_{1K_1},\ldots,\lambda_{11}=\lambda_{1K_1}\times10^{-4}\}$ and $\{\lambda_{2K_2},\ldots,\lambda_{21}=\lambda_{2K_2}\times10^{-4}\}$, which are equally spaced on the log-scale. Third, the Bayesian information criterion (BIC) is adopted to choose the optimal tuning parameters $\hat{\lambda}_1$ and $\hat{\lambda}_2$ which minimizes

$$\mathrm{BIC}(\hat{\theta}^{\lambda})=-\mathrm{n}\left\{\log\left|\Omega(\hat{\phi}^{\lambda},\hat{\tau}^{\lambda})\right|-\mathrm{tr}\left(\Omega(\hat{\phi}^{\lambda},\hat{\tau}^{\lambda})\mathsf{V}(\hat{\beta}^{\lambda})\right)\right\}+\log\mathrm{n}\left(\mathrm{d}_{\beta}+\mathrm{d}_{\phi}+\mathrm{m}\right),$$

where d_{β} and d_{ϕ} are the numbers of nonzero elements of β and ϕ, respectively. In the simulation study, we also try the validation approach to select tuning parameters. We generate additional n validation data in each run with the same model, and choose the tuning parameters that maximize the likelihood of $\hat{\theta}^{\lambda}$, given the validation data. These two methods yield similar results; therefore, we only report the results of the method using the validation approach.

4. Numerical Studies

In this section, we investigate the empirical properties of the proposed estimator. We compare our approach with the SCGGM of Zhang and Kim [5]. We use a training set of $n=100$ samples to train the different methods. An independent validation set of size n is used to select the prediction optimal tuning parameters λ_1 and λ_2. We use grids (on the log-scale) for both, λ_1 and λ_2, where the grid for λ_1 is of size 30 and the grid for λ_2 is typically of size 50.

Table 1. Experimental scenarios

Model	m	p	$\mathbb{P}(w_{ij}\neq0)$	$\mathbb{P}(\beta_j^l\neq0)$
1	10	10	$2/m$	$3.5/p$
2	20	10	$2/m$	$3.5/p$
3	30	20	$2/m$	$3.5/p$

4.1 Experiment scenarios

The experiment scenarios used in this study are motivated by Zhang and Kim [5] and Lee et al. [11]. We generate simulation data sets such that nonzero entries of the precision matrix are randomly assigned with probability c_1/p with a positive constant c_1. To make the (i, j)th entry of the precision matrix to be denoted by w_{ij}, the corresponding element is observed from uniform distribution over $[-1,0.5]\cup[0.5,1]$. For each row, off-diagonal components are divided by the sum of their absolute values multiplied by 1.5. Further, we obtain Ω by symmetrizing. To create B, we generate a $p\times m$ indicator matrix which has 1 as its entry, with probability c_2/p for a positive constant c_2. If the (l, j)th element of this matrix has the value 1, β_j^l is generated from $Unif([d_m,1]\cup[-1,-d_m])$, where d_m is the smallest absolute value of Ω.

Producing B and Ω, we generate $X=(X_1,\ldots,X_p)$ using $X_l\sim Unif(-1,1)$ for $l=1,\ldots,p$. Finally, we generate y from the multivariate normal distribution given X, $Y|X\sim\mathcal{N}_m(X^{\mathsf{T}}\mathsf{B},\Omega^{-1})$. We also generate a data set of n i.i.d random vectors (X,Y). Table 1 specifies the three simulation scenarios. A total of 50 simulation runs are used for each of setting.

4.2 Performance measures

In order to measure the estimation performance of our proposed method, we use the Steins loss and Frobenius norm, whose respective definitions are

$$\delta_{Stein}(\Omega,\widehat{\Omega})=\left\langle\Omega,\widehat{\Omega}^{-1}\right\rangle-\log\left|\Omega\widehat{\Omega}^{-1}\right|-m \text{ and}$$
$$\|\Delta\|_F=\left\|\Omega-\widehat{\Omega}\right\|_F,$$

where $\widehat{\Omega}$ is an estimate of the true precision matrix Ω. In each simulation run, we compute these quantities of the two meth-

Table 2. Comparisons of the performance of PJM and SCGGM with each model (standard errors are presented in parentheses.)

Model	Method	δ_{Stein}	$\|\Delta\|_F$	SPE	SEN	Youden
1	PJM	0.467 (0.015)	1.023 (0.030)	0.26 (0.022)	0.99 (0.004)	0.25 (0.021)
	SCGGM	0.539 (0.017)	1.122 (0.029)	0.44 (0.029)	0.93 (0.010)	0.37 (0.027)
2	PJM	1.163 (0.02)	1.683 (0.033)	0.49 (0.031)	0.7 (0.024)	0.2 (0.014)
	SCGGM	1.622 (0.043)	2.095 (0.065)	0.55 (0.035)	0.7 (0.023)	0.25 (0.017)
3	PJM	2.362 (0.036)	2.544 (0.038)	0.61 (0.032)	0.62 (0.031)	0.23 (0.010)
	SCGGM	2.935 (0.100)	3.003 (0.103)	0.64 (0.029)	0.62 (0.022)	0.26 (0.012)

ods and summarize them over all the simulation runs.

To measure the performance of identifying nonzero entries in the precision matrix, we count the number of true positives (TP) and false positives (FP) in each simulation run. We check how efficiently our model recovers the true conditional dependent relationship among the genes, with specificity (SPE) and sensitivity (SEN) defined by

$$\text{SPE} = \frac{TN}{TN+FP} \text{ and } \text{SEN} = \frac{TP}{TP+FN},$$

where TN and FN are the numbers of true negatives and false negatives, respectively, with regard to off-diagonal elements of the precision matrix. In all the experiments, we treat a nonzero entry as a "positive." Combining SPE and SEN, the Youden's index ($=\text{SPE}+\text{SEN}-1$) is used to measure the overall selection performance.

4.3 Results

The simulation results are summarized in Table 2. First, we confirm that the PJM estimator has an advantage over the SCGGM in estimation accuracy, measured by the Stein loss and the Frobenius norm. On the other hand, the PJM is almost equal to, or slightly underperforms, the SCGGM, in terms of variable selection. We checked all the possible estimates for each λ given the appropriate λ sequences. In some experiments, rather than the optimal value of λ through the BIC or validation approach, the PJM estimator for other λ values yields better performance for variable selection. In these cases, it is observed that the values of δ_{Stein} and $\|\Delta\|_F$ tend to increase. This indicates that the performance depends on how the criterion for selecting the optimal complexity parameter λ is defined. Overall, the two methods perform fairly well in our simulation studies.

5. Concluding Remarks

In this study, we demonstrate that the penalizing method for estimating the mean vector and the covariance matrix. Contrary to the existing conditional graphical models, we parameterize the precision matrix using a modified Cholesky decomposition. We consider a coordinate-wise updating method to separately estimate the mean and covariance parts, the solutions of which, in theory, are equally obtained from solving l_1-penalized regression. The results of numerical studies in various scenarios demonstrate that the finite-sample performance of the proposed estimator is satisfactory. Although we focus on the l_1 lasso penalty term, it can be extended to other forms of regularization, such as SCAD (Smoothly Clipped Absolute Deviation [12]) and MCP (Minimax Concave Penalty [13]). These issues deserve further research in future studies.

Acknowledgements

The research of Ja-Yong Koo 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). The research of SungHwan Kim is supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP; Ministry of Science, ICT and Future Planning) (NRF-2017R1C1B5017528).

References

1. Yin J, Li H. A sparse conditional gaussian graphical model for analysis of genetical genomics data. Ann Appl Stat 2011;5: 2630-2650.
2. Li B, Chun H, Zhao H. Sparse estimation conditional graphical models with applications to gene network. J Am Stat Assoc 2012;107:152-167.
3. Cai T, Li H, Liu W, Xie J. Covariate-adjusted precision matrix estimation with an application in genetical genomics. Biometrika 2013;100:139-156.

4. Chun H, Chen M, Li B, Zhao H. Joint conditional gaussian graphical models with multiple sources of genomic data. Front Genet 2013;4:294.
5. Zhang L, Kim S. Learning gene networks under snp perturbations using eqtl datasets. PLOS Comput Biol 2014;10:e1003420.
6. Pourahmadi M. Joint mean-covariance models with applications to longitudinal data: unconstrained parameterization. Biometrika 1999;86:677-690.
7. Chang C, Tsay RS. Estimation of covariance matrix via the sparse cholesky factor with lasso. J Stat Plan Infer 2010;140: 3858-3873.
8. Tibshirani R. Regression shrinkage and selection via the lasso. J Roy Stat Soc B Met 1996;58:267-288.
9. Huang JZ, Liu N, Pourahmadi M, Liu L. Covariance matrix selection and estimation via penalized normal likelihood. Biometrika 2006;93:85-98.
10. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010;33:1-22.
11. Lee JJ, Kim SH, Jhong JH, Koo JY. Variable selection and joint estimation of mean and covariance models with an application to eQTL data. Submitted to J Appl Stat 2016.
12. Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. J Am Stat Assoc 2001;96:1348-1360.
13. Zhang CH. Nearly unbiased variable selection under minimax concave penalty. Ann Stat 2010;38:894-942.