Covariate Adjusted Precision Matrix Estimation with an Application in Genetical Genomics

BY T. TONY CAI

Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA
tcai@wharton.upenn.edu

AND HONGZHE LI

Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA
hongzhe@upenn.edu

AND WEIDONG LIU

Department of Mathematics and Institute of Natural Sciences, Shanghai Jiao Tong University, Shanghai 200240, China
liuweidong99@gmail.com

AND JICHUN XIE

Department of Statistics, Fox School of Business, Temple University, Philadelphia, Pennsylvania 19122, USA
jichunxie1@gmail.com

SUMMARY

Motivated by the analysis of eQTL data, we introduce a sparse high dimensional multivariate regression model for studying the conditional independent relationships among a set of genes adjusting for possible genetic effects, as well as the genetic architecture that influences the gene expressions. The precision matrix in the model specifies a covariate-adjusted Gaussian graph, which presents the conditional dependency structure of gene expression after the confounding genetic effects on gene expressions are taken into account. We present a covariate-adjusted precision matrix estimation (CAPME) method using constrained $\ell_1$ minimization, which can be easily implemented by linear programming. Asymptotic convergence rates and sign consistency are established for the estimators of the regression coefficients and the precision matrix, allowing both the number of genes and the number of the genetic variants to diverge. Numerical performance of the estimators is investigated using both simulated and real data sets. Simulation results show that CAPME results in significant improvements in both precision matrix estimation and in graphical structure selection when compared to the standard Gaussian graphical model assuming constant means. CAPME is also applied to analyze a yeast eQTL data for the identification of the gene network among a set of genes in the Mitogen-activated protein kinase (MAPK) pathway.

Some key words: eQTL data, high-dimensionality, Gaussian graphical model, multivariate regression.
1. Introduction

Gaussian graphical models have been applied to infer the relationship between genes at the transcriptional level (Schafer & Strimmer, 2005; Segal et al., 2005; Li & Gui, 2006; Peng et al., 2009), where the precision matrix (inverse covariance matrix) for multivariate normal data has an interpretation of conditional dependency and induces a graphical structure among the variables. Compared with marginal dependency, conditional dependency can capture the “direct” link between the variables. Since the variation at the expression level for a given gene can usually be explained by a small subset of other genes, the precision matrix for gene expression data is expected to be sparse. For example, Figure 1 shows the Mitogen-activated protein kinase (MAPK) signalling pathway from the KEGG database, where the links between the genes are sparse.

Estimation of high-dimensional Gaussian graphical models has been an active area of research in recent years. Meinshausen & Buhlmann (2006) proposed neighborhood selection by identifying edges for each node in the graph using \( \ell_1 \) penalized regression to select the variables that are associated with a particular node. This approach reduces the graphical model estimation problem to \( p \) separate high dimensional variable selection problems which have been well studied. Methods for variable selection in high dimensional regressions and their properties have been discussed in Candès & Tao (2007); Cai et al. (2010); James et al. (2009); Tibshirani (1996); Wainwright (2009); Zou (2006) with Lasso or Dantzig selector, in Fan et al. (2009); Zhang (2010) with non-concave SCAD or MCP penalties, and in Yuan & Lin (2006); Zou & Hastie (2005) with group LASSO or elastic net for regression problem with certain structures. Estimation of the precision matrix and the graphical structure can also be obtained through a penalized maximum likelihood approach, including Friedman et al. (2008); Rothman et al. (2008); Yuan & Lin (2007), Friedman et al. (2008) proposed a fast block coordinate descent algorithm, called GLASSO, to solve the penalized likelihood maximization problem. Cai et al. (In press) proposed a constrained \( \ell_1 \) minimization estimator (CLIME) for precision matrix estimation and obtained the results on convergence rates and sign consistency.

Although a direct application of the Gaussian graphical model to gene expression data alone provides some insights into gene regulation at the expression level, it does not fully utilize the available information. Gene expression data alone is unable to fully capture the gene activities. Estimation accuracy can be improved by incorporating other information such as genetic variants information from typical genetical genomic studies. It has now been well-established that gene expression levels are inheritable and can be viewed as quantitative traits in both model organisms and in human. The genetic loci that control the gene expressions are therefore referred as expression quantitative trait loci (eQTL) (Jansen & Nap, 2001). Gene expression data, together with the genetical variants information, have provided important insights into gene expression regulations in both model organisms and in human (Brem & Kruglyak, 2005; Cheung & Spielman, 2002; Schadt et al., 2003). Since some genetic variants have effects on expression levels of multiple genes and therefore may serve as confounders while detecting the association between the genes. Ignoring the effects of genetic variants on the gene expression levels can lead to both false positives and false negatives for associations in the gene network graph. In order to adjust for the effects of genetic variants on gene expressions, Yin & Li (2011) proposed a sparse conditional Gaussian graphical model (cGGM) for analysis of genetical genomics data. The model simultaneously identifies the genetic variants associated with gene expressions and constructs sparse Gaussian graphical model by maximizing a \( \ell_1 \) penalized joint likelihood of the regression coefficients and the precision matrix. However, the non-concavity of the objective function creates difficulties in both theoretical analysis and computational implementation.
In this paper, we introduce a sparse multivariate high dimensional regression model for studying the conditional independent relationships among a set of genes adjusting for possible genetic effects. The model allows the mean of gene expression levels to depend on other variables such as genetic markers and therefore allows different subjects to have different means. The model can be regarded as a covariate-adjusted Gaussian graphical model. We consider the estimation of both coefficient matrix and precision matrix and present a covariate adjusted precision matrix estimation (CAPME) method. CAPME is a two-stage method, where in each stage, we use constrained $\ell_1$ minimization approach to estimate coefficient matrix or precision matrix. The two matrices are estimated separately. Algorithm-wise, CAPME can be easily implemented by linear programming. An R package of CAPME has been developed and is available on the CRAN (http://cran.r-project.org/). We provide the rates of convergence and the estimation bounds for the estimates of both the regression coefficient matrix and the precision matrix, allowing both the number of the genetic variants and the number of genes to diverge as the sample size $n \to \infty$. In addition, a simple threshold procedure on the estimated precision matrix is proposed to recover the covariate-adjusted Gaussian graphical structure and show that the graph structure selection method is consistent. The method is applied to a yeast eQTL data to demonstrate the application of constructing gene regulatory network.

The rest of the paper is organized as follows. Section 2 presents the covariate-adjusted Gaussian graphical model and introduces the CAPME procedure. The asymptotic properties of CAPME are investigated in Section 3 and Section 4. Our method is further justified in Section 5 where numerical studies are carried out to compare the performances of CAPME and other competitive methods such as CLIME, EGGM and GLASSO. We use the yeast eQTL data to demonstrate the application of our method. The paper is concluded with a brief discussion in Section 7. The proofs are relegated to the Appendix.

2. Covariate Adjusted Precision Matrix Estimation

We first introduce the notation of matrix norms used in the rest of the paper. For a vector $a = (a_1, \ldots, a_p)^T \in \mathbb{R}$, define $|a|_1 = \sum_{i=1}^{p} |a_i|$ and $|a|_2 = \sqrt{\sum_{i=1}^{p} a_i^2}$. For a matrix $A = (a_{ij}) \in \mathbb{R}^{p \times q}$, we define the entrywise $\ell_1$ norm $|A|_1 = \sum_{i=1}^{p} \sum_{j=1}^{q} |a_{ij}|$ and the entrywise $\ell_\infty$ norm $|A|_\infty = \max_{i,j} |a_{ij}|$. The matrix $L_1$ norm is defined by $\|A\|_{L_1} = \max_{1 \leq i \leq p, 1 \leq j \leq q} \sum_{j=1}^{q} |a_{ij}|$, the matrix $L_\infty$ norm by $\|A\|_{L_\infty} = \max_{1 \leq i \leq p} \sum_{j=1}^{q} |a_{ij}|$, the spectral norm $\|A\|_2 = \max_{|x|_2=1} |Ax|_2$ and the Frobenius norm by $\|A\|_F = \sqrt{\sum_{i,j} a_{ij}^2}$. The notation $A \succ 0$ means that $A$ is positive definite.

2.1. Covariate Adjusted Gaussian Graphical Model

We consider the following model

$$y = \Gamma_0 x + z,$$  \hspace{1cm} (1)

where $y = (y_1, \ldots, y_p)^T$ is a random vector denoting the expression levels for $p$ genes, $x = (x_1, \ldots, x_q)^T$ is a random vector describe the coding for $q$ genetic markers, $\Gamma_0$ is a $p \times q$ unknown coefficient matrix, $z$ is a $p \times 1$ normal random vector with mean zero, covariance matrix $\Sigma_0 = (\sigma_{0_{ij}})$ and precision matrix $\Omega_0 = (\omega_{0_{ij}}) = \Sigma_0^{-1}$. Since the segregating population in the genetical genomics experiments can be viewed as random due to the random recombination process, $z$ can be treated as a random variable. We further assume $x$ and $z$ are independent. Assume that we have $n$ independent identically distributed observations $(x_k, y_k)$ ($k = 1, \ldots, n$) from model (1).
Model (1) is similar to the seemingly unrelated regression (SUR) model of Zeller (1962), which aimed to improve the estimation efficiency of the effects of genetic variants on gene expressions $\Gamma_0$ by considering the residual correlations of the gene expressions of many genes. However, the model is viewed differently here with a focus on improving the estimation accuracy of the conditional dependency structure of $y$ by adjusting for the covariates $x$. Different from the classical SUR model, we consider the case where $p$ and $q$ can be larger than $n$.

In eQTL studies, each row of $\Gamma_0$ is assumed to be sparse since one gene is expected to have only a few genetic regulators. The precision matrix $\Omega_0$ is also expected to be sparse, since typical genetic networks have limited links. $\Omega_0$ has an interpretation of conditional dependency and can be used to construct a conditional dependency graph. To be specific, let $G = (V, E)$ be a graph representing conditional independence relations between components of $y$. The vertex set $V$ has $p$ components $y_1, \ldots, y_p$ and the edge set $E$ consists of pairs $(i, j)$, where $(i, j) \in E$ if there is an edge between $y_i$ and $y_j$. The edge between $y_i$ and $y_j$ is excluded from $E$ if and only if $z_i$ and $z_j$ are independent given all other $z_k$'s $(k \neq i, j)$. Since $z$ follows a multivariate normal distribution, the conditional independence of $z_i$ and $z_j$ leads to $\omega_{ij} = 0$. We are interested in detecting the non-zero entries of $\Omega_0$ so that we can construct a conditional independency graph for $y$ after the effects of the covariates $x$ on $y$ are adjusted. Such a graphical model is called the covariate-adjusted Gaussian graphical model.

In this paper, we consider the setting where both $p$ and $q$ can be much larger than $n$. The assumption fits the real applications to the analysis of genetical genomics data where there are usually thousands of genes and genetic markers, but relatively small sample sizes.

2.2. Estimation of $\Gamma_0$

When $q = 1$, many novel methods have been developed for estimation of $\Gamma_0$, including the methods based on the $\ell_1$ minimization such as the LASSO in Tibshirani (1996) and the Dantzig selector in Candès & Tao (2007). We propose to develop a method for estimating $\Gamma_0$ using constrained $\ell_1$ minimization that can be treated as a multivariate extension of the Dantzig selector. Let $\bar{\gamma} = n^{-1}\sum_{k=1}^{n}y_k$, $\bar{x} = n^{-1}\sum_{k=1}^{n}x_k$, $\bar{z} = n^{-1}\sum_{k=1}^{n}z_k$. Then it follows that

$$y_k - \bar{\gamma} = \Gamma_0(x_k - \bar{x}) + z_k - \bar{z}. \quad (2)$$

Set $S_{xy} = n^{-1}\sum_{k=1}^{n}(y_k - \bar{\gamma})(x_k - \bar{x})^T$ and $S_{xx} = n^{-1}\sum_{k=1}^{n}(x_k - \bar{x})(x_k - \bar{x})^T$. We propose to estimate $\Gamma_0$ by the solution to the following the optimization problem:

$$\hat{\Gamma} \in \arg\min_{\Gamma \in \mathbb{R}^{p \times q}} \left\{ \|\Gamma\|_1 : \left\| S_{xy} - \Gamma S_{xx} \right\|_\infty \leq \lambda_n \right\}, \quad (3)$$

where $\lambda_n$ is a tuning parameter that will be specified later. Note that (3) is equivalent to the following $p$ optimization problems: for $1 \leq i \leq p$,

$$\min_{\gamma_i} \left\{ \|\gamma_i\|_1, \text{ subject to } \left\| S_{xy,i} - \gamma_i^T \Gamma S_{xx} \right\|_\infty \leq \lambda_n \right\}, \quad (4)$$

where $\Gamma = (\gamma_1, \ldots, \gamma_p)^T$ and $S_{xy} = (S_{xy,1}, \ldots, S_{xy,p})^T$. This is exactly the Dantzig selector formulation in the usual regression analysis for the $i$th gene and its solution can therefore be obtained by solving the corresponding linear programming problem (Candès & Romberg, 2005) or by some alternative methods (Becker et al., 2010; Lu, 2009; Lu et al., 2011). This simple observation is useful for implementation and technical analysis. In this paper and the R package CAPME we developed, CAPME is implemented by a linear programming optimization using the primal dual and interior point algorithm.
2.3. Estimation of $\Omega_0$

After plugging the estimator $\hat{\Gamma}$ given in (3) into the equations (2), $\Omega_0$ can be estimated by the method of constrained $\ell_1$-minimization proposed in Cai et al. (In press). Let

$$S_{yy} = \frac{1}{n} \sum_{k=1}^{n} (y_k - \hat{\Gamma} x_k)(y_k - \hat{\Gamma} x_k)^T.$$  

We estimate $\Omega_0$ by the solution to the following optimization problem:

$$\hat{\Omega}_1 \in \arg\min_{\Omega \in \mathbb{R}^{p \times p}} \left\{ |\Omega|_1 : |I_p \times p - S_{yy}\Omega|_\infty \leq \tau_n \right\},$$  

(5)

where $\tau_n$ is a tuning parameter. Let $\hat{\Omega}_1 = (\hat{\omega}_{ij}^1)$ be any solution to (5). This constrained $\ell_1$-minimization approach is the same as the CLIME proposed in Cai et al. (In press), except that $S_{yy}$ depend on the estimated regression coefficient matrix $\hat{\Gamma}$. Note that we do not impose the symmetry condition on $\hat{\Omega}_1$ and as a result the solution is not symmetric in general. The final CAPME estimator of $\Omega_0$, denoted by $\hat{\Omega} = (\hat{\omega}_{ij})$, is obtained by symmetrizing the estimator as follows.

$$\hat{\Omega} = (\hat{\omega}_{ij}), \quad \text{where} \quad \hat{\omega}_{ij} = \hat{\omega}_{ji} = \hat{\omega}_{ij}^1 I\{|\hat{\omega}_{ij}^1| \leq |\hat{\omega}_{ji}^1|\} + \hat{\omega}_{ji}^1 I\{|\hat{\omega}_{ij}^1| > |\hat{\omega}_{ji}^1|\}. \quad (6)$$

As in (4), the problem (6) can be decomposed into $p$ optimization problems. For $1 \leq i \leq p$, let $\hat{\omega}_i$ be the solution of the following convex optimization problem

$$\min |\omega_i|_1 \text{ subject to } |e_i - S_{yy}\omega_i|_\infty \leq \tau_n,$$

(7)

where $\omega_i$ is a vector in $\mathbb{R}^p$, $e_i$ is a standard unit vector in $\mathbb{R}^p$ with 1 in the $i$-th coordinate and 0 in all other coordinates. This can also be solved using the primal dual and interior point algorithm.

3. Rates of Convergence of the Estimators

We now turn to the theoretical properties of the estimators $\hat{\Gamma}$ and $\hat{\Omega}$. Write $x = (x_1, \ldots, x_q)^T$, $z = (z_1, \ldots, z_p)^T$ and $u = z^T \Omega_0 = (u_1, \ldots, u_p)$. The following conditions are needed for establishing the rates of convergence.

A1. Let $\log(p \vee q) = o(n)$. Suppose that there exists some $\eta > 0$ and $K > 0$ such that

$$E\{\exp(\eta x_i^2)\} \leq K, \quad E\{\exp(\eta z_j^2/\sigma_{jj}^2)\} \leq K, \quad E\{\exp(\eta u_i^2/\omega_{jj}^0)\} \leq K$$

for all $1 \leq i \leq q$ and $1 \leq j \leq p$. Furthermore, $\max_{1 \leq j \leq p} \sigma_{jj}^2 \leq K$.

A2. The regression coefficient matrix $\Gamma_0$ belongs to the following class with $0 \leq \delta_1 < 1$:

$$\mathcal{V}_{\delta_1} := \mathcal{V}_{\delta_1}(s_1(q)) = \left\{ \Gamma \in \mathbb{R}^{p \times q} : \max_{1 \leq i \leq p} \sum_{j=1}^{q} |\gamma_{ij}|^{\delta_1} \leq s_1(q) \right\}.$$

A3. The precision matrix $\Omega_0 = (\omega_{ij}^0)_{p \times q}$ belongs to the following class with $0 \leq \delta_2 < 1$:

$$\mathcal{U}_{\delta_2} := \mathcal{U}_{\delta_2}(s_2(p)) = \left\{ \Omega > 0 : \|\Omega\|_{L_{\infty}} \leq M_p, \quad \max_{1 \leq i \leq p} \sum_{j=1}^{p} |\omega_{ij}|^{\delta_2} \leq s_2(p) \right\}.$$

$$\lambda_{\max}(\Omega)/\lambda_{\min}(\Omega) \leq C_0.$$

A4. There exists some $N_q > 0$ such that the matrix $l_\infty$ norm of $\Sigma_{x}^{-1}$ satisfies

$$\|\Sigma_{x}^{-1}\|_{L_{\infty}} \leq N_q.$$
where $\Sigma_x = \text{cov}(x)$.

Condition A1 is a sub-gaussian condition on $x$, $z$ and $z^T \Omega_0$. Note that the variance of $u_j$ is $\omega_{jj}^0$. The dimensions $p$ and $q$ are of the order $\exp(o(n))$. Conditions A2 and A3 assume the uniformity class of matrices for the regression coefficient matrix and the precision matrix. Similar parameter spaces have also been used in Bickel & Levina (2008) and Cai et al. (In press). Note that $V_0$ and $U_0$ are classes of matrices with the sparsity measurements of $s_1(q)$ and $s_2(p)$, respectively. A2 and A3 also bound the matrix $L_\infty$ norm of $\Gamma_0$ and $\Omega_0$. Finally, Condition A4 bounds the matrix $L_\infty$ norm of the inverse covariance matrix of $x$.

### 3.1. Convergence rates of $\hat{\Gamma} - \Gamma_0$

The estimation error $\hat{\Gamma} - \Gamma_0$ can be measured by three types of matrix norms: the matrix $L_\infty$ norm, the Frobenius norm and the entry-wise $\ell_\infty$ norm. The matrix $L_\infty$ norm measures the accuracy of the estimation of $\Gamma_0$. The Frobenius norm is also a reasonable measure on the accuracy of the estimation of $\Gamma_0$ and can be viewed as the sum of squared errors for estimating individual rows. The entry-wise $\ell_\infty$ norm can be used to recover the support of $\Gamma_0$ by a further thresholding step. We have the following rate of convergence for the estimator $\hat{\Gamma}$.

**Theorem 1.** Suppose (A1), (A2) and (A4) hold. Let

$$\Gamma_0 \in V_{\delta_1}, \quad \lambda_n = C_1 \{(\log(pq))/n\}^{1/2},$$

where $C_1 > 0$ is a sufficiently large constant. If

$$s_1(q) = o \left\{ N_q^{\delta_1} \left( \frac{n}{\log(pq)} \right)^{(1-\delta_1)/2} \right\}, \quad (8)$$

then with probability greater than $1 - O((pq)^{-1})$, we have

$$\|\hat{\Gamma} - \Gamma_0\|_{L_\infty} \leq CN_q^{1-\delta_1} s_1(q) \left\{ \frac{\log(pq)}{n} \right\}^{(1-\delta_1)/2} \quad (9)$$

and

$$\frac{1}{p} \|\hat{\Gamma} - \Gamma_0\|_F^2 \leq CN_q^{2-\delta_1} s_1(q) \left\{ \frac{\log(pq)}{n} \right\}^{1-\delta_1/2} \quad (10)$$

for some constant $C > 0$.

Theorem 1 shows that the regression coefficient matrix $\Gamma_0$ can be estimated consistently under the Frobenius norm if the sparsity $s_1(q)$ of $\Gamma_0$ is of order $o \left\{ N_q^{\delta_1/2} \left( \frac{n}{\log(pq)} \right)^{1-\delta_1/2} \right\}$. The requirement on the dimension $p$ and $q$ is mild as they only appear in the logarithmic term. To see this, if $s_1(q) = O(n^{r_1})$ for some $r_1 < 1 - \delta_1/2$ and $N_q$ is bounded, then $p$ and $q$ can be as large as $\exp(n^{2r_2})$ for some $r_2 < 1 - \delta_1/2 - r_1$.

**Theorem 2.** Under the conditions of Theorem 1, with probability greater than $1 - O((pq)^{-1})$, we have

$$|\hat{\Gamma} - \Gamma_0|_{\infty} \leq C_0 N_q \left\{ \frac{\log(pq)}{n} \right\}^{1/2} \quad (11)$$

for some constant $C_0 > 0$.
The rate under the element-wise $\ell_\infty$ norm is critical to the support recovery. Define $\hat{\Gamma}_{\text{thr}} = (\hat{\gamma}_{ij})$ with

$$\hat{\gamma}_{ij} = \tilde{\gamma}_{ij} I \left[ |\tilde{\gamma}_{ij}| \geq C_0 N_q \left\{ \frac{\log(pq)}{n} \right\}^{1/2} \right],$$

where $(\tilde{\gamma}_{ij}) := \hat{\Gamma}$. Let $S(\Gamma_0) = \{(i,j) : \gamma_{ij}^0 \neq 0\}$ and $\gamma_{\min} = \min_{(i,j) \in S(\Gamma_0)} |\gamma_{ij}|$.

**Theorem 3.** Suppose the conditions in Theorem 1 hold and

$$\gamma_{\min} \geq 2C_0 N_q \left\{ \frac{\log(pq)}{n} \right\}^{1/2},$$

then with probability greater than $1 - O\{(pq)^{-1}\}$, we have $S(\hat{\Gamma}_{\text{thr}}) = S(\Gamma_0)$.

The lower bound condition (12) requires that the magnitude of the non-zero entries in $\Gamma_0$ cannot be too small in order to achieve the support recovery.

### 3.2. Convergence rates of $\hat{\Omega} - \Omega_0$

In this section, we consider the rate of $\hat{\Omega} - \Omega_0$ under the spectral norm and the element-wise $\ell_\infty$ norm. The rate under the spectral norm is important because it can lead to the consistency of the estimation of eigenvalues and eigenvectors. The rate under the spectral norm is essentially needed in developing theoretical properties for various statistical inference problems when the estimator of the precision matrix is used.

**Theorem 4.** Suppose (A1) – (A4) and (8) hold. Let $\Gamma_0 \in \mathcal{V}_{\delta_1}$, $\Omega_0 \in \mathcal{U}_{\delta_2}$ and

$$s_1(q) \leq C(1 + M_p)^{-1} N_q^{2+\delta_1} \left\{ \frac{n}{\log(pq)} \right\}^{(1-\delta_1)/2}.$$ \hspace{1cm} (13)

Let $\tau_n = C_2\{(\log(pq))/n\}^{1/2}$, where $C_2 > 0$ is a sufficiently large constant. Then with probability greater than $1 - O\{(pq)^{-1}\}$, we have

$$\|\hat{\Omega} - \Omega_0\|_2 \leq CM_p^{1-\delta_2} s_2(p) \left\{ \frac{\log(pq)}{n} \right\}^{(1-\delta_2)/2}$$

for some constant $C > 0$.

The condition (13) on the sparsity $s_1(q)$ of $\Gamma_0$ ensures that $\Gamma_0$ can be well estimated with certain rate so that $y - \Gamma_0 x$ can be used to replace the oracle one $y - \Gamma_0 x$. The convergence rate in (14) is optimal. In fact, even if $\Gamma_0 = 0$ or is known in advance, the minimax optimal rate of estimation of $\Omega_0$ is still $O\left(M_p^{1-\delta_2} s_2(p)(\log p/n)^{(1-\delta_2)/2}\right)$ (Cai et al., 2011). If $q = O(p)$, then the rate in (14) is the same as the oracle optimal rate and thus is also optimal.

The next theorem shows the convergence rate under the element-wise $\ell_\infty$ norm, which is useful for the support recovery for the precision matrix $\Omega$.

**Theorem 5.** Under the conditions of Theorem 3, we have with probability greater than $1 - O\{(pq)^{-1}\}$,

$$|\hat{\Omega} - \Omega_0|_\infty \leq CM_p \left\{ \frac{\log(pq)}{n} \right\}^{1/2},$$

where $C > 0$ is a constant.
As shown in Cai et al. (2011), the minimax optimal rate under element-wise l∞ norm of estimation of precision matrix is $O\left(M_p (\log p/n)^{1/2}\right)$ when $\Gamma_0 = 0$ or is known. Hence CAPME can achieve the same optimal rate as the case that $\Gamma_0$ is known.

4. Graphical Model Selection Consistency

The support recovery of the precision matrix $\Omega_0$ is closely related to the graphical model selection. When $\Gamma_0 = 0$, the problem is reduced to the Gaussian graphical model selection and has been the subject of research in many recent papers. See, for example, Meinshausen & Bühlmann (2006, Cai et al. (In press). Suppose that $\Omega_0$ belongs to $\mathcal{U}_0$. We are interested in estimating the support of $\Omega_0$, $S(\Omega_0) = \{(i, j) : \omega_{ij}^0 \neq 0\}$ when $\Gamma_0 \neq 0$. However, the estimator $\hat{\Omega}$ from the CAPME may not be sparse. Using the rate under the element-wise norm given in Theorem 5, we can further threshold the entries in $\hat{\Omega}$ and obtain a sparse estimator of the precision matrix. Let

$$\hat{\Omega}_r = (\hat{\omega}_{ij}^r), \quad \text{where} \quad \hat{\omega}_{ij}^r = \hat{\omega}_{ij} I\{|\hat{\omega}_{ij}| \geq \tau_n^r\},$$

where $\tau_n^r$ is a tuning parameter which will be specified later. Define $\theta_{\min} = \min_{(i, j) \in S(\Omega_0)} |\omega_{ij}^0|$, and $\Psi = \{\text{sign}(\omega_{ij}^0) : 1 \leq i, j \leq p\}$ and $\hat{\Psi} = \{\text{sign}(\hat{\omega}_{ij}^r) : 1 \leq i, j \leq p\}$ be the vector of the signs of the elements of the true and the estimated precision matrix, where $\text{sign}(t)$ is defined as

$$\text{sign}(t) = \begin{cases} 1 & \text{if } t > 0 \\ 0 & \text{if } t = 0 \\ -1 & \text{if } t < 0. \end{cases}$$

We have the following theorem on sign consistency of the estimator $\hat{\Psi}$, i.e., the estimator not only recovers the sparsity pattern of $\Gamma_0$, but also recovers the signs of the nonzero elements.

**Theorem 6.** Let $\tau_n^r = 4M_p \tau_n$. Suppose that $\theta_{\min} > 2\tau_n^r$. Then under the conditions of Theorem 4, as $n$ and $p$ tending to infinity, we have with probability tending to one, $\hat{\Psi} = \Psi$.

Theorem 6 shows that the support of $\Omega_0$ can be recovered exactly if the minimum of the nonzero entries in $\Omega_0$ has a lower bound which is not too small. The lower bound condition is necessary in order to recover the support exactly. In fact, suppose that $\Gamma_0 = 0$ or is known in advance. If $\theta_{\min} \leq c\tau_n^r$ for a sufficiently small constant $c > 0$, then for any estimator of $\Omega_0$, it is not possible to recover the support exactly uniformly over the class of $s_2(p)$ sparse precision matrices; see Cai et al. (2011).

5. Simulation Results

We conduct simulation studies to evaluate the numerical performance of the proposed procedure and to compare it with several other procedures for precision matrix estimation and support recovery. We consider four different settings of $p, q$ and $n$ as presented in Table 1.

For each model, we first generate a $p \times q$ sparse coefficient matrix $\Gamma$ and a $p \times p$ sparse precision matrix $\Omega$ randomly. For Model 1-3, the average nonzero proportion of each row of $\Gamma$ is set to be $15/q$. Therefore, for each $Y_j$, there are in average 15 covariates that are associated with its mean. For all the models, the average nonzero proportion of each row of $\Omega$ is set to be $5/q$, leading to relatively sparse precision matrix and the corresponding graph. On average, there are 5 neighbors for each gene node. As more genetic variables are included in the model, the coefficient and the precision matrix will become sparser. In order to ensure the numerical stability, the
diagonal level of the matrix is adjusted so that the condition number, defined as the ratio of the maximal and minimal singular values, is equal to the dimension of the matrix \( p \).

Model 1 has a relatively small \((p + q)/n\) ratio. This model is considered to mimic the application in elucidating small-scale gene regulatory pathways or constructing of networks in other areas such as in social sciences. In addition, we set the coefficient matrix \( \Gamma \) is non-zero everywhere, representing the 'extreme' case, in which the comparison between several methods can provide some insights into different procedures. Model 2 has a moderate \((p + q)/n\) ratio, while Model 3 has a large \((p + q)/n\) ratio, simulating the situation in most genomic applications. Model 4, on the other hand, simulates the other 'extreme' case, where we assume the coefficient matrix \( \Gamma \) is zero. For numerical experiments, we first generate random matrix according to the settings of Models 1-4. For each replication of the simulations, a \( q \times n \) covariate matrix \( X \) is generated with each entry following a standard normal distribution and the random error \( z \) following multivariate normal distribution with zero mean and the covariance matrix \( \Sigma = \Omega^{-1} \). Following the model \( y = \Gamma x + z \), the outcome matrix \( y \) is then generated.

The performance of CAPME ia compared with those of estimates, including CLIME, cGGM and GLASSO, where CLIME and GLASSO assume a constant mean and do not adjust for the covariate effects. In contrast, CAPME and cGGM both allow the means to depend on covariates. For each method, a 5-fold cross validation is used to choose the tuning parameters based on maximizing the cross-validated log-likelihood function

\[
\log \det(\hat{\Omega}) - \text{tr}(S_{yy}\hat{\Omega}).
\]

The methods are evaluated and compared using several different criteria. The estimation errors are evaluated in terms of operator norm, Frobenius norm and \( l_1 \) norm and the variable selection property is evaluated by Hamming distance (DIST), specificity (SPE), sensitivity (SEN) and Matthews correlation coefficient (MCC), which are defined as the following:

\[
\begin{align*}
\text{DIST}(\Omega_0, \hat{\Omega}) &= FN + FP, \quad \text{SPE} = \frac{TN}{TN + FP}, \quad \text{SEN} = \frac{TP}{TP + FN}, \\
\text{MCC} &= \frac{TP \times TN - FP \times FN}{\{ (TP + FP)(TP + FN)(TN + FP)(TN + FN) \}^{1/2}}.
\end{align*}
\]

Here, TP, TN, FP, FN are the numbers of true positives, true negatives, false positives and false negatives. Here, "positives" stands for non-zero entries in the estimators. The performances over 50 replications are reported in Table 2 and Table 3.

In general, when the regression coefficient matrix for the means is dense, as expected, both CAPME and cGGM that adjust for covariates perform significantly better. For example, in Model 1, the coefficient matrix is non-zero everywhere, CLIME and GLASSO fail to produce a sparse network, and the estimation errors are large when compared to CAPME and cGGM. CAPME and cGGM perform almost equally well, while cGGM outperforms a little bit in estimation and CAPME works better in variable selection.
In Model 2, when the coefficient matrix is moderately sparse and \((p + q)/n\) ratio is moderate, CLIME and GLASSO perform worse than CAPME. cGGM has a reasonable performance in variable selection, but its estimation error is relatively large. When the \((p + q)/n\) ratio becomes larger as in Model 3, cGGM and GLASSO perform poorly. CAPME outperforms other methods in both estimation error and variable selection.

In Model 4, the coefficient matrix \(\Gamma\) is set to be zero. The extra step CAPME takes to estimate \(\Gamma\) does not affect its numerical performance. It has comparable performance to the methods GLASSO and CLIME assuming a constant mean.

Overall, CAPME outperforms the other competing methods, especially in the relatively high dimensional setting. To further gain insight into how covariates can affect the estimation of the precision matrix, Table 4 shows the extra false links that are identified by GLASSO or CLIME when compared to CAPME or cGGM and how many of these links that can be due to shared genetic variants or confounders. CLIME and GLASSO identified many false links due to the fact that shared confounders are not adjusted. CAPME detects a large proportion of the confounders and therefore successfully avoids many false discoveries. In contrast, cGGM identifies a relatively smaller set of confounders.

Table 2. Simulation results: estimation errors of four different methods for the precision matrix as measured by different matrix norms based on 100 replications. Numbers in parentheses are the simulation standard errors.

<table>
<thead>
<tr>
<th>Model 1</th>
<th>Method</th>
<th>Operator norm</th>
<th>Frobenius norm</th>
<th>Matrix (L_1) norm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>CAPME</td>
<td>5.97(0.59)</td>
<td>15.00(0.73)</td>
<td>14.39(1.48)</td>
</tr>
<tr>
<td></td>
<td>CLIME</td>
<td>104(14.64)</td>
<td>64.42(2.51)</td>
<td>63.00(16.04)</td>
</tr>
<tr>
<td></td>
<td>cGGM</td>
<td>4.84(0.80)</td>
<td>13.86(0.32)</td>
<td>10.60(1.05)</td>
</tr>
<tr>
<td></td>
<td>GLASSO</td>
<td>7.68(0.05)</td>
<td>19.27(0.21)</td>
<td>16.45(0.39)</td>
</tr>
<tr>
<td>Model 2</td>
<td>CAPME</td>
<td>7.93(0.10)</td>
<td>25.13(0.09)</td>
<td>16.3(0.16)</td>
</tr>
<tr>
<td></td>
<td>CLIME</td>
<td>8.01(0.15)</td>
<td>25.73(0.09)</td>
<td>16.45(0.29)</td>
</tr>
<tr>
<td></td>
<td>cGGM</td>
<td>10.82(3.67)</td>
<td>23.86(0.55)</td>
<td>17.78(2.24)</td>
</tr>
<tr>
<td></td>
<td>GLASSO</td>
<td>8.56(0.07)</td>
<td>27.65(0.32)</td>
<td>17.87(0.28)</td>
</tr>
<tr>
<td>Model 3</td>
<td>CAPME</td>
<td>0.47(0.00)</td>
<td>12.02(0.01)</td>
<td>1.06(0.01)</td>
</tr>
<tr>
<td></td>
<td>CLIME</td>
<td>2.13(0.23)</td>
<td>19.29(0.50)</td>
<td>3.41(0.36)</td>
</tr>
<tr>
<td></td>
<td>cGGM</td>
<td>9.54(0.01)</td>
<td>50.42(0.09)</td>
<td>20.33(0.08)</td>
</tr>
<tr>
<td></td>
<td>GLASSO</td>
<td>9.54(0.01)</td>
<td>50.39(0.07)</td>
<td>20.32(0.08)</td>
</tr>
<tr>
<td>Model 4</td>
<td>CAPME</td>
<td>15.4(0.22)</td>
<td>20.43(0.28)</td>
<td>12.12(0.81)</td>
</tr>
<tr>
<td></td>
<td>CLIME</td>
<td>7.69(0.12)</td>
<td>24.55(0.19)</td>
<td>14.36(0.50)</td>
</tr>
<tr>
<td></td>
<td>cGGM</td>
<td>6.13(2.43)</td>
<td>21.38(0.28)</td>
<td>13.11(0.67)</td>
</tr>
<tr>
<td></td>
<td>GLASSO</td>
<td>4.68(0.36)</td>
<td>21.88(0.27)</td>
<td>12.53(0.91)</td>
</tr>
</tbody>
</table>

6. Real Data Analysis

We illustrate CAPME using the yeast eQTL data set generated by [Brem & Kruglyak (2005)]. The data set contains 112 yeast segregants grown from a cross involving BY4716 and wild wild isolate RM11-1a. RNA was isolated and cDNA was hybridized to microarrays with 6,216 yeast genes assayed on each array. Each of the 112 segregants were individually genotyped at 2,956 marker positions. Since many of the markers are in high linkage disequilibrium, we chose 179 markers so that their pairwise correlations will not exceed 0.6.

In order to demonstrate the performance of CAPME and other methods, we focused our analysis on 54 genes that belong to the yeast MAPK signaling pathway provided by the KEGG database (Kanehisa et al., 2010). Figure 1 shows the MAPK signaling pathway provided by the KEGG...
Simulation results: variable selection performances as measured by distance, sensitivity, specificity and the Matthew’s correlation, for four different procedures, based on 100 replications. Numbers in parentheses are the simulation standard errors.

<table>
<thead>
<tr>
<th>Model</th>
<th>Method</th>
<th>DIST</th>
<th>SPE</th>
<th>SEN</th>
<th>MCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>CAPME</td>
<td>212.52(25.87)</td>
<td>0.99(0.02)</td>
<td>0.27(0.12)</td>
<td>0.41(0.09)</td>
</tr>
<tr>
<td></td>
<td>CLIME</td>
<td>616.00(2.91)</td>
<td>(0)</td>
<td>1.00(0.01)</td>
<td>NA(NA)</td>
</tr>
<tr>
<td></td>
<td>cGGM</td>
<td>206.02(14.11)</td>
<td>0.90(0.03)</td>
<td>0.43(0.07)</td>
<td>0.38(0.05)</td>
</tr>
<tr>
<td></td>
<td>GLASSO</td>
<td>561.70(15.24)</td>
<td>0.15(0.02)</td>
<td>0.85(0.03)</td>
<td>0.00(0.05)</td>
</tr>
<tr>
<td>Model 2</td>
<td>CAPME</td>
<td>616.04(15.97)</td>
<td>0.96(0.01)</td>
<td>0.20(0.02)</td>
<td>0.23(0.02)</td>
</tr>
<tr>
<td></td>
<td>CLIME</td>
<td>852.44(105.51)</td>
<td>0.88(0.04)</td>
<td>0.20(0.04)</td>
<td>0.09(0.03)</td>
</tr>
<tr>
<td></td>
<td>cGGM</td>
<td>678.68(120.12)</td>
<td>0.92(0.05)</td>
<td>0.20(0.06)</td>
<td>0.15(0.05)</td>
</tr>
<tr>
<td></td>
<td>GLASSO</td>
<td>2125.98(77.58)</td>
<td>0.35(0.03)</td>
<td>0.64(0.05)</td>
<td>0.00(0.02)</td>
</tr>
<tr>
<td>Model 3</td>
<td>CAPME</td>
<td>2334.48(164.55)</td>
<td>0.99(0.00)</td>
<td>0.11(0.01)</td>
<td>0.19(0.02)</td>
</tr>
<tr>
<td></td>
<td>CLIME</td>
<td>2831.48(163.86)</td>
<td>0.98(0.00)</td>
<td>0.13(0.01)</td>
<td>0.15(0.01)</td>
</tr>
<tr>
<td></td>
<td>cGGM</td>
<td>9273.68(703.07)</td>
<td>0.80(0.02)</td>
<td>0.22(0.03)</td>
<td>0.01(0.01)</td>
</tr>
<tr>
<td></td>
<td>GLASSO</td>
<td>8878.04(528.67)</td>
<td>0.81(0.01)</td>
<td>0.21(0.02)</td>
<td>0.01(0.01)</td>
</tr>
<tr>
<td>Model 4</td>
<td>CAPME</td>
<td>472.88(22.25)</td>
<td>0.92(0.01)</td>
<td>0.59(0.03)</td>
<td>0.52(0.02)</td>
</tr>
<tr>
<td></td>
<td>CLIME</td>
<td>461.68(27.33)</td>
<td>0.95(0.02)</td>
<td>0.47(0.06)</td>
<td>0.49(0.03)</td>
</tr>
<tr>
<td></td>
<td>cGGM</td>
<td>631.94(46.66)</td>
<td>0.88(0.02)</td>
<td>0.52(0.04)</td>
<td>0.37(0.02)</td>
</tr>
<tr>
<td></td>
<td>GLASSO</td>
<td>872.82(82.49)</td>
<td>0.77(0.04)</td>
<td>0.65(0.05)</td>
<td>0.34(0.02)</td>
</tr>
</tbody>
</table>

**Table 4. Simulation results:** effects of genetic variants on Gaussian graph recovery

<table>
<thead>
<tr>
<th>Model</th>
<th>Methods</th>
<th>FP</th>
<th>Covariate(true)</th>
<th>Covariate(estimated)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>GLASSO/CAPME</td>
<td>258.31 (8.02)</td>
<td>258.31 (8.02)</td>
<td>258.31 (8.02)</td>
</tr>
<tr>
<td></td>
<td>GLASSO/cGGM</td>
<td>166.93 (126.63)</td>
<td>166.93 (126.63)</td>
<td>166.93 (126.63)</td>
</tr>
<tr>
<td></td>
<td>CLIME/CAPME</td>
<td>303.78 (9.22)</td>
<td>303.78 (5.22)</td>
<td>303.78 (5.22)</td>
</tr>
<tr>
<td></td>
<td>CLIME/cGGM</td>
<td>196.86 (149.15)</td>
<td>196.86 (149.15)</td>
<td>196.86 (149.15)</td>
</tr>
<tr>
<td>Model 2</td>
<td>GLASSO/CAPME</td>
<td>910.91 (49.08)</td>
<td>910.91 (49.08)</td>
<td>910.91 (49.08)</td>
</tr>
<tr>
<td></td>
<td>GLASSO/cGGM</td>
<td>946.77 (145.39)</td>
<td>946.77 (145.39)</td>
<td>923.01 (163.46)</td>
</tr>
<tr>
<td></td>
<td>CLIME/CAPME</td>
<td>166.04 (58.85)</td>
<td>166.04 (58.85)</td>
<td>166.04 (58.85)</td>
</tr>
<tr>
<td></td>
<td>CLIME/cGGM</td>
<td>178.10 (67.52)</td>
<td>178.10 (67.52)</td>
<td>173.36 (67.22)</td>
</tr>
<tr>
<td>Model 3</td>
<td>GLASSO/CAPME</td>
<td>3597.16 (283.49)</td>
<td>3423.10 (261.98)</td>
<td>2454.5 (355.33)</td>
</tr>
<tr>
<td></td>
<td>GLASSO/cGGM</td>
<td>516.65 (124.48)</td>
<td>478.39 (116.65)</td>
<td>0(0)</td>
</tr>
<tr>
<td></td>
<td>CLIME/CAPME</td>
<td>349.52 (77.90)</td>
<td>311.80 (70.28)</td>
<td>237.24 (82.62)</td>
</tr>
<tr>
<td></td>
<td>CLIME/cGGM</td>
<td>326.94 (58.70)</td>
<td>282.22 (51.39)</td>
<td>0(0)</td>
</tr>
</tbody>
</table>

FP, the number of wrong links that are identified using the Gaussian graphical model (GLASSO or CLIME) but not by the covariance adjusted Gaussian graphical model (cGGM or CAPME). Covariate (true), among the wrong links under the FP the number of pairs of genes that share at least one genetic variant based on the true model; Covariate (estimated), among the wrong links under the FP the number of pairs of genes that share at least one genetic variant based on the estimated covariance adjusted Gaussian graphical model.

Table 3. Simulation results: variable selection performances as measured by distance, sensitivity, specificity and the Matthew’s correlation, for four different procedures, based on 100 replications. Numbers in parentheses are the simulation standard errors.

website, showing how yeast cells respond to different cellular signals. Note that this pathway plot is not a graph since some gene nodes such as Ste20, Ste11 and Ste7 appear in multiple locations. The corresponding undirected graph as given in the Bioconductor database is shown in Figure 6, where the nodes with the same shape are involved in the same perturbation process and appeared to be clustered in Figure 6. In this undirected graph, the round nodes are for cell responses to the pheromone – mating process, the square ones are for cell responses to hypotonic shock – cell wall remodeling, the octagon ones are for cell responses to high osmolarity – osmolyte synthesis.
and the triangle nodes are for cell responses to starvation – filamentation. The diamond nodes are those involved in multiple perturbation processes.

![The yeast MAPK signaling pathway from the KEGG database, illustrating the signaling paths in responses to different signals. Note that some genes appear in multiple paths.](http://www.genome.jp/kegg/pathway/sce/sce04011.html)

We applied CAPME to this set of 54 genes and 179 markers and used 5-fold cross validation to choose the tuning parameters. The cross validation chose $\lambda = 0.15$ and $\tau = 0.19$, which lead to CAPME to select 138 non-zero entries for the precision matrix and 42 links among the 54 genes. In addition, CAPME identified 181 non-zero entries for the coefficient matrix, indicating many gene expression levels are affected by genetic variants. The number of genetic markers that are associated with a given gene ranges from 0 to 29, with a mean value of 3.35. There are 54 pairs of genes sharing at least one common genetic variant. If we do not adjust for the genetic effects on gene expressions, shared genetic variants can lead to false links between the gene pairs. Figure 6 shows the graph constructed by CAPME based on the precision matrix estimator. It is very interesting to see that the nodes of the same shape tend to linked together based on the conditional Gaussian graph identified by CAPME, indicating that the data and our model can indeed recover some important links related to the MAPK signaling pathway. For example, the kinase Sho1 is linked to its downstream gene MCM1 and the upper part of the pheromone mating process is captured well in the estimated graph.

As a comparison, we applied CLIME and GLASSO to the gene expressions of 54 genes without adjusting for the genetic marker effect on gene expressions. We used the 5-fold cross-validation to choose the tuning parameters for both methods, resulting $\lambda_{\text{CLIME}} = 0.25$ and $\lambda_{\text{GLASSO}} = 0.007$. CLIME identified 200 non-zero entries for the precision matrix and therefore 73 links of 54 genes, while GLASSO identified 1186 non-zero entries and therefore 566 links. The corresponding graphs are presented in Figure 6 and Figure 6, where the genes involved in different perturbation processes are mixed together. The results obtained from CLIME and GLASSO are less systematic when compared to the results of CAPME. There are too many false links so that the biologically
meaningful links cannot be distinguished. One potential reason that CLIME and GLASSO failed to cluster the genes well is that they do not adjust for genetic variants associated with the expression levels.

We noticed that some links presented in the original pathway graph Figure 6 are missed by the graph derived from the estimated precision matrix. One reason is that many of the links of the MAPK pathway can only be observed at the protein levels but not at the gene expression levels. Another reason is that the cellular systems might not be perturbed enough to provide enough information to infer these links. Figure 6 shows the histogram of the correlations of genes that are linked based on the MAPK signaling pathway, where some of the linked gene pairs have very small marginal correlations. CAPME and CLIME are not able to recover the links. On the other hand, the linked genes identified by these two methods tend to have higher marginal correlations...
(Figure 6 and Figure 6). In contrast, some linked genes identified by GLASSO have small marginal correlations in their expression levels (Figure 6).

![Graphs of marginal correlations for pairs of linked genes](image)

**Fig. 3.** Analysis of yeast eQTL data set: histograms of marginal correlations for pairs of linked genes on the MAPK signaling pathway (the upper left panel), and the linked genes identified by CAPME (the upper right panel), CLIME (the lower left panel) and GLASSO (the lower right panel).

### 7. Conclusion and Discussion

We have introduced a constrained $\ell_1$ minimization method for the sparse high dimensional multivariate regression model to estimate the gene network based on the genetical genomics data in order to account for genetic effects on gene expressions. By including the genetic information in the model, the variations and the dependency of gene expression levels are separated into two parts: the correlation due to common genetic factors and the coherence for gene transcriptional activities. In the multivariate regression model, the coefficient matrix is used to model the first part and the precision matrix is used to model the second. The numerical experiments demonstrated that the proposed method CAPME can capture the network more accurately than its competitors such as CLIME and GLASSO. We compared CAPME with several methods in the analysis of the yeast eQTL data and observed that the conventional methods for the Gaussian graphical model resulted in too many false links and fail to provide any biological insights of
the data. In contrast, CAPME provides a sparser and more interpretable graph. Due to the limited capacity of gene expression data, it is not expected that the gene regulatory network can be accurately recovered by gene expression data alone. However, as demonstrated in this paper, through the joint modeling of genetic markers and gene expressions, the proposed method CAPME can indeed provide important information about the gene networks.

In this paper, we have focused on estimating the sparse precision matrix for the gene expression data accounting for the genetics effects. The coefficient matrix is estimated without utilizing the dependency information of the gene expressions, which may lead to some loss of efficiency. In other applications, the focus could be on the inference of the regression coefficients of multiple regressions. It is interesting to study whether efficiency can be gained in estimating the regression coefficients when the estimated sparse precision matrix is utilized. In the low dimensional setting, such problem has been considered by SUR model (Zeller, 1962).

8. Appendix - Proofs of the Theorems

The first lemma is an exponential inequality on the partial sums of independent random variables (Cai & Liu, In press).

Lemma 1. Let \( \xi_1, \ldots, \xi_n \) be independent random variables with mean zero. Suppose that there exists some \( t > 0 \) and \( \tilde{B}_n \) such that

\[
\sum_{k=1}^{n} E\{\xi_k^2 e^{t|\xi_k|}\} \leq \tilde{B}_n^2.
\]

Then uniformly for \( 0 < x \leq \tilde{B}_n \),

\[
\Pr\left( \sum_{k=1}^{n} \xi_k \geq C_t \tilde{B}_n x \right) \leq \exp(-x^2),
\]

where \( C_t = t + t^{-1} \).

Proof of Theorem 1 and Theorem 2. Without loss of generality, we assume that \( E(x) = 0 \). Recall that \( E(z) = 0 \). We show that with probability greater than \( 1 - O\{pq^{-1}\} \),

\[
\left| S_{xy} - \Gamma_0 S_{xx} \right|_\infty \leq \lambda_n.
\]

(17)

To prove (17), it suffices to show that

\[
\left| \frac{1}{n} \sum_{k=1}^{n} (z_k - \bar{z})(x_k - \bar{x})^T \right|_\infty \leq \lambda_n.
\]

(18)

Taking \( \xi_k = z_k, x_k \) in Lemma 1 and note that \( \max_{i,j} E \exp(t|z_k x_k|) \leq K \) for all \( |t| \leq \min(\eta, \eta/K) \), we have

\[
\max_{i,j} \Pr\left( n^{-1} \sum_{k=1}^{n} z_k x_k \geq \lambda_n/2 \right) \leq 2(pq)^{-2}.
\]

(19)

By Lemma 1, we have

\[
\max_{j} \Pr\left( |\bar{x}| \geq C\{\log(pq)/n\}^{1/2} \right) \leq 2(pq)^{-2}
\]
and
\[
\max_i \Pr\left(|\tilde{z}_i| \geq C\left\{\log(pq)/n\right\}^{1/2}\right) \leq 2(pq)^{-2}
\]
for some constant \(C > 0\). This implies (18). Let \(\hat{\Gamma} = (\hat{\gamma}_{ij}) = (\hat{\gamma}_1, \ldots, \hat{\gamma}_p)^T\) be the solution of (3). Then by (17), we have
\[
|(|\hat{\Gamma} - \Gamma_0|S_{\text{max}}|) \leq 2\lambda_n.
\]
Moreover, by the equivalence between (3) and (4), we have \(\sum_{j=1}^q |\hat{\gamma}_{ij}| \leq \sum_{j=1}^q |\gamma_{ij}|\) for all \(1 \leq i \leq p\). Set \(\|\Gamma_0\|_{L_\infty} = \max_{1 \leq i \leq p} \sum_{j=1}^q |\gamma_{ij}|\). We have \(\|\hat{\Gamma}\|_{L_\infty} \leq \|\Gamma_0\|_{L_\infty}\). Also by Lemma I we have
\[
\Pr\left(|\Sigma_{\text{max}} - S_{\text{max}}| \geq C\left\{\log(pq)/n\right\}^{1/2}\right) \leq 2(pq)^{-1}
\]
for some constant \(C > 0\). Then, with probability greater than \(1 - O\{pq\}^{-1}\), we have
\[
|(|\hat{\Gamma} - \Gamma_0|\Sigma_{\text{max}}|) \leq |(\hat{\Gamma} - \Gamma_0)S_{\text{max}}|_{L_\infty} + |(\hat{\Gamma} - \Gamma_0)(\Sigma_{\text{max}} - S_{\text{max}})|_{L_\infty} \leq 2\lambda_n + C\|\hat{\Gamma} - \Gamma_0\|_{L_\infty}\{\log(pq)/n\}^{1/2}.
\]
It follows that
\[
|\hat{\Gamma} - \Gamma_0|_{L_\infty} \leq 2\|\Sigma_{\text{max}} - S_{\text{max}}\|_{L_1}\|\Sigma_{\text{max}}^{-1}\|_{L_1} \leq 2\|\Sigma_{\text{max}}^{-1}\|_{L_1}\lambda_n + C\|\hat{\Gamma} - \Gamma_0\|_{L_\infty}\{\log(pq)/n\}^{1/2}.
\]
Let \(t_n = |\hat{\Gamma} - \Gamma_0|_{L_\infty}\). Define
\[
h_j = (h_{j1}, \ldots, h_{jq})^T = \hat{\gamma}_j - \gamma_j^0;
\]
\[
h_j^1 = (\gamma_{ij}| |\hat{\gamma}_{ij}| \geq 2t_n : 1 \leq i \leq q)^T - \gamma_j^0;
\]
\[
h_j^2 = h_j - h_j^1.
\]
Then
\[
|h_j^2|_1 - |h_j^1|_1 + |\gamma_j^0|_1 \leq |h_j^2|_1 + |h_j^1|_1 + |\gamma_j^0|_1 = |h_j|_1 + |\gamma_j^0|_1 \leq |\gamma_j^0|_1.
\]
So we have \(|h_j|_1 \leq 2|h_j^1|_1\). It suffices to estimate \(|h_j^1|_1\). We have
\[
|h_j^1|_1 = \sum_{i=1}^q |\gamma_{ij}|I\{|\hat{\gamma}_{ij}| \geq 2t_n\} - |\gamma_j^0|_1
\]
\[
= \sum_{i=1}^q |\gamma_{ij} - \gamma_j^0|I\{|\hat{\gamma}_{ij}| \geq 2t_n\} + \sum_{i=1}^q |\gamma_j^0|I\{|\gamma_{ij}| < 2t_n\}
\]
\[
\leq \sum_{i=1}^q t_nI\{|\gamma_{ij}| \geq t_n\} + \sum_{i=1}^q |\gamma_j^0|I\{|\gamma_{ij}| < 3t_n\}
\]
\[
\leq t_n(3t_n)^{1-\delta_1} \sum_{i=1}^q \left(\sum_{j=1}^q |\gamma_{ij}|^{\delta_1} \right) + (3t_n)^{1-\delta_1} \sum_{i=1}^q |\gamma_j^0|^{\delta_1}.
\]
Therefore,
\[
|\hat{\Gamma} - \Gamma_0|_{L_\infty} \leq C s_1(q)N_q^{1-\delta_1}\lambda_n^{1-\delta_1} + C\|\hat{\Gamma} - \Gamma_0\|_{L_\infty}^{1-\delta_1} s_1(q)N_q^{1-\delta_1}\lambda_n^{1-\delta_1}.
\]
If \( \| \hat{\Gamma} - \Gamma_0 \|_{L_\infty} \leq 1 \), then we have \( \| \hat{\Gamma} - \Gamma_0 \|_{L_\infty} \leq C s_1(q) N q^{1-\delta_1} \lambda_n^{1-\delta_1} \). If \( \| \hat{\Gamma} - \Gamma_0 \|_{L_\infty} > 1 \), then by (8), we have for \( n \) large,

\[
\| \hat{\Gamma} - \Gamma_0 \|_{L_\infty} \leq C s_1(q) N q^{1-\delta_1} \lambda_n^{1-\delta_1} + \frac{1}{2} \| \hat{\Gamma} - \Gamma_0 \|_{L_\infty}.
\]

Thus (9) holds with probability greater than \( 1 - O((pq)^{-1}) \). By (9) and (8), we have \( \| \hat{\Gamma} - \Gamma_0 \|_{L_\infty} \leq 1 \) with probability greater than \( 1 - O\{ (pq)^{-1} \} \). This, together with (20), implies (11). Finally, (10) follows from (9), (11) and the inequality \( p^{-1} \| \hat{\Gamma} - \Gamma_0 \|_F^2 \leq \| \hat{\Gamma} - \Gamma_0 \|_{L_\infty} \). \( \square \)

Proof of Theorem 4 and Theorem 5

Recall that \( E(z) = 0 \). Set

\[
\hat{\Sigma}_z = \frac{1}{n} \sum_{k=1}^{n} z_k z_k^T.
\]

We suppose that

\[
|(S_{yy} - \hat{\Sigma}_z) \Omega_0|_\infty \leq \tau_n
\]

and

\[
|(\hat{\Sigma}_z - \Sigma_0) \Omega_0|_\infty \leq \tau_n.
\]

Then we have

\[
|I_{p \times p} - S_{yy} \Omega_0|_\infty = |(S_{yy} - \Sigma_0) \Omega_0|_\infty \leq 2\tau_n.
\]

It follows that

\[
|\Omega_1 - \Omega_0|_\infty \leq |(I_{p \times p} - \Omega_0 S_{yy}) \hat{\Omega}_1|_\infty + |\Omega_0 (I_{p \times p} - S_{xx} \hat{\Omega}_1)|_\infty \leq 2\|\Omega_0\|_{L_1} \tau_n.
\]

This proves Theorem 5. Following the arguments as the proof of Theorem 1, we can get Theorem 4.

It remains to prove (21) and (22). Write \( \Delta_n = \hat{\Gamma} - \Gamma_0 \). Then we have

\[
S_{xx} = \frac{1}{n} \sum_{k=1}^{n} (z_k - \Delta_n x_k) (z_k - \Delta_n x_k)^T.
\]

We now prove that with probability greater than \( 1 - O\{ (pq)^{-1} \} \),

\[
\left| \frac{1}{n} \sum_{k=1}^{n} z_k x_k^T \Delta_n^T \right|_\infty \leq CM_p^{-1} \left\{ \frac{\log(pq)}{n} \right\}^{1/2}
\]

and

\[
\left| \frac{1}{n} \sum_{k=1}^{n} \Delta_n x_k x_k^T \Delta_n^T \right|_\infty \leq CM_p^{-1} \left\{ \frac{\log(pq)}{n} \right\}^{1/2}.
\]

First, recall that

\[
\max_{i,j} \text{pr} \left( n^{-1} \sum_{k=1}^{n} z_{ki} x_{kj} \geq \lambda_n / 2 \right) \leq C (pq)^{-2}.
\]
Write $\Delta_n = (\delta_{ij})$, $x_k = (x_{kj})$ and $z_k = (z_{kj})$. To prove (23), we only need to show that with probability greater than $1 - O((pq)^{-1})$,

$$\max_{i,j} \left| \frac{1}{n} \sum_{k=1}^{n} \left( z_{ki} x_{kj} \delta_{ij} + \cdots + z_{ki} x_{kq} \delta_{iq} \right) \right| \leq C \left( \frac{\log(pq)}{n} \right)^{1/2}.$$ 

By (9), (13) and (25),

$$\max_{i,j} \left| \frac{1}{n} \sum_{k=1}^{n} \sum_{q=1}^{q} z_{ki} x_{kj} \delta_{ij} \right| \leq \| \hat{\Gamma} - \Gamma_0 \| \max_{i,j} \left| \frac{1}{n} \sum_{k=1}^{n} z_{ki} x_{kj} \right| \leq CM_p^{-1} \max_{i,j} \left| \frac{1}{n} \sum_{k=1}^{n} z_{ki} x_{kj} \right| \leq CM_p^{-1} \left( \log(pq)/n \right)^{1/2}. \quad (26)$$

Thus (23) holds. It remains to show (24), which is equivalent to show that with probability greater than $1 - O((pq)^{-1})$,

$$\max_{i,j} \left| \frac{1}{n} \sum_{k=1}^{n} \sum_{q=1}^{q} \delta_{ij} x_{kj} \right| \leq C M_p^{-1} \left( \log(pq)/n \right)^{1/2}. \quad (27)$$

By Lemma [1] we can get

$$\max_{j} \text{pr} \left( \frac{1}{n} \sum_{k=1}^{n} x_{kj}^2 \geq C \right) = O((pq)^{-2})$$

for some constant $C > 0$. By (11), (9) and (13),

$$\max_{i} \frac{1}{n} \sum_{k=1}^{n} \left( \sum_{j=1}^{q} \delta_{ij} x_{kj} \right)^2 \leq \max_{i} \sum_{j=1}^{q} \delta_{ij}^2 \frac{1}{n} \sum_{k=1}^{n} x_{kj}^2 \leq CM_p^{-1} \left( \log(pq)/n \right)^{1/2}$$

with probability greater than $1 - O((pq)^{-1})$. This implies (27).

We next prove (22). Write

$$(\hat{\Sigma}_2 - \Sigma_0) \Omega_0 = \frac{1}{n} \sum_{k=1}^{n} (z_k z_k^T \Omega_0 - E z_k z_k^T \Omega_0).$$

Note that $\text{var}(z_{ki}) = \sigma_{ki}^0$ and $\text{var}(z_k^T \Omega_0) = \omega_{ij}^0$. By A2, $\max_i \sigma_{ki}^0 \max_j \omega_{ij}^0 \leq C_0$. By Lemma 1, we have

$$\max_{i,j} \text{pr} \left( \left| \frac{1}{n} \sum_{k=1}^{n} (z_{ki} z_k^T \Omega_0 - E z_{ki} z_k^T \Omega_0) \right| \geq C \left( \frac{\log(pq)}{n} \right)^{1/2} \right) \leq C(pq)^{-2}.$$

for some bounded constant $C$ depending only on $C_0$, $\eta$ and $K$. This yields (22). \hfill \Box

\textbf{REFERENCES}


[Received month year. Revised month year]