SEQUENTIAL APPROACHES IN GRAPHICAL MODELS AND MULTIVARIATE RESPONSE REGRESSION MODELS

JIANG YIWEI

(B.Sc., WUHAN UNIVERSITY)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF STATISTICS AND APPLIED PROBABILITY

NATIONAL UNIVERSITY OF SINGAPORE

2015

DECLARATION

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Jiang Yiwei

8th May 2015

Thesis Supervisor

Chen Zehua Professor; Department of Statistics and Applied Probability, National University of Singapore, Singapore, 117546, Singapore.

ACKNOWLEDGMENTS

First and foremost, I would like to show my deepest gratitude to my supervisor, Professor Chen Zehua, for his patience, continuous support and valuable advice. He is truly a great mentor, who conscientiously led me into the field of statistical research. His knowledge, clarity of instruction and encouragement greatly motivated this research.

I gratefully acknowledge the National University of Singapore for awarding me the scholarship and the Department of Statistics and Applied Probability for providing me the opportunity to pursue my graduate study. I am also thankful to the other faculty members and the department secretarial staffs. Special appreciation goes to Mr. Zhang Rong and Ms. Chow Peck Ha, Yvonne for their IT support.

I would like to express my sincere gratitude to Dr. Luo Shan, Dr. He Yawei and Dr. Yin Teng, who have devoted their time and attention to facilitating my research. Also thanks to all my classmates and friends in the department for the company and encouragement. You have made my Ph.D life pleasant and memorable. I must not forget my dear overseas fellows. Thanks for your friendship, whenever and wherever.

Last but not least, I owe a lot to my parents, who instilled in me the

inspiration and perseverance to chase my dream. Thanks for your unconditional love, considerable understanding and constant support in my life. This thesis is also in memory of my dear grandfather.

Contents

SUMMARY

This thesis aims at expanding the idea of the sequential LASSO approach in linear regression models to the areas of graphical models and multivariate response regression models under the high-dimension-low-sample-size circumstances.

First, a sequential scaled pairwise selection (SSPS) method is developed for the edge detection in sparse high-dimensional nonparanormal graphical models. The extended Bayesian information criterion (EBIC) is adopted as the stopping rule for this sequential procedure. Its selection consistency is established under appropriate conditions. Extensive simulation studies are carried out to compare the edge detection accuracy among SSPS and other competitors. The results demonstrate that the SSPS method has an edge over the others. In addition, the computational efficiency makes it more appealing. Its applications on the precision matrix estimation are also explored. Specifically, the SSPS method can not only be used to directly identify the nonzero entries of the precision matrix, but also serve as a screening tool for the other existing methods. Follow-up simulations and a real example are employed to assess the proposed methods' estimation accuracy in comparison to the others.

Another aspect considered in this thesis is two sequential approaches to the joint estimation of the coefficient matrix and the precision matrix in multivariate response regression models. The ideas of sequential methods in linear regression models and Gaussian graphical models are combined and exploited to derive these two sequential methods in the conditional regression formulation (SCR). One relies on the alternate updating framework; the other depends on the simultaneous estimation scheme. Considerable simulation examples are used to show the SCR methods' overall advantages in terms of model selection and prediction. The implementations of these methods on the real data analysis are also examined.

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Context of Research

The graphical models play an important role in studying the network among a set of variables, where the variables are represented by vertices in a graph and their conditional dependent relationships are demonstrated by the edges connecting the relevant vertices (Lauritzen, 1996). Particularly assuming the multivariate normal distribution of these variables, the consequent model is termed as Gaussian graphical model. The pairwise conditional relationships are then completely encoded in the inverse of the covariance matrix, a.k.a. the precision matrix. Hence, by setting some precision matrix's entries to zero, Dempster (1972) introduced the concept of covariance selection to simplify these variables' network structure. The process of recovering such network is also called edge detection, which concentrates on discriminating the zero and nonzero entries of the precision matrix. Besides, precision matrix estimation is crucial for principal com-

ponent analysis, linear discriminant analysis and longitudinal studies, et cetera.

In practice, the precision matrix is estimated with i.i.d. samples of these variables. Since the dimensionality of the data surges exponentially along with the expeditious information technology evolution, the situation of the dimensionality exceeding the sample size challenges the traditional precision matrix estimation methods. They could become quite unstable and even infeasible, such as inverting the sample covariance matrix. With the multivariate Gaussian assumption, considerable research activities have been devoted to solving the sparse precision matrix in either the neighborhood selection approach or the penalized likelihood approach. Both of them employ the regularized estimation techniques, which add different levels of penalties to the parameters related to the precision matrix. Such techniques were first developed to perform variable selection and coefficients estimation simultaneously in the linear regression models. Benefiting from the trade-off between the estimation bias and variability, they can produce continuously shrunk estimates and more importantly stable estimation procedures. With penalties that are singular at the origin, some parameters can be estimated as exactly zero, achieving the parsimonious estimation.

Recent studies have extended the regularized estimation framework to solve the problems in the multivariate response regression models. These models have increasing applications, e.g., predicting the returns of multiple stocks simultaneously with a common set of econometric predictors, or recovering the genetic expressions' network with their means adjusted by the effects of some genetic variants. Other practices can be seen in disciplines

such as chemometrics and psychometrics.

One branch of the relevant work focuses exclusively on the sparse coefficient matrix estimation by imposing various forms of penalties on it. For instance, penalizing the coefficient matrix elementwise produces overall sparse estimates; penalizing the coefficients of the same covariate as a group can fit all the responses with only a subset of the total covariates. Another branch of the methods intend to incorporate the response variables' correlations into the coefficient matrix estimation and model prediction. Therefore, the response variables' precision matrix estimation can be involved. The joint estimation of the sparse coefficient matrix and precision matrix are usually formulated in two ways: the penalized likelihood and the penalized conditional regressions.

In the thesis, we propose some sequential methods for the edge detection and variable selection involved in the graphical models and the multivariate response regression models. In the rest of this chapter, we summarize the evolution of the regularized precision matrix estimation in the graphical models and the development of the relevant parametric estimation in the multivariate response regression models, which are followed by the objective of this study and the thesis outline.

1.2 Literature Review

This section first reviews a few representative studies in the literature of sparse precision matrix estimation. The regularized coefficients estimation techniques of the linear regression models are also evaluated, since they

are fundamental for the regularized precision matrix estimation methods. Then the methods for the estimation of the coefficient matrix as well as the precision matrix in the multivariate response regression models are discussed and compared.

1.2.1 Sparse Precision Matrix Estimation

Suppose $\boldsymbol{y}_j$ consists of n i.i.d. samples of random variable Y_j, for $j = 1, \ldots, p$. Let Σ be the covariance matrix of $Y_1, \ldots, Y_p$ and Ω be its inverse, i.e. the precision matrix. The estimation of the precision matrix is as elementary as that of the covariance matrix in statistical analysis. In the context of large p and small n, the sample covariance matrix $\hat{\Sigma}_n$ is probably singular and impossible to be inverted for the precision matrix estimation.

The current trend of the research on the precision matrix estimation can be mainly classified into two categories. In the first category, although the sample covariance behaves unsatisfactorily in high-dimensional settings, the regularized estimation based on its modified Cholesky decomposition and the inverse is still feasible (Wu and Pourahmadi, 2003; Huang et al., 2006; Bickel and Levina, 2008; Cai et al., 2010). Moreover, Cai et al. (2011) introduced the constrained ℓ_1-minimization for inverse matrix estimation (CLIME) method to estimate Ω by minimizing $\|\Omega\|_1$ subject to $\|\hat{\Sigma}_n\Omega - \boldsymbol{I}_p\|_\infty \leq \lambda$, where λ is a regularization parameter, $\boldsymbol{I}_p$ denotes the $p \times p$ identity matrix, $\|\cdot\|_1$ and $\|\cdot\|_\infty$ are elementwise matrix ℓ_1 norm and ℓ_∞ norm respectively. In its implementation, each column of Ω's estimate is obtained by solving a linear program without any outer iteration. Such algorithm is efficient and easy to parallel. After symmetrizing the estimator that formed

by combining the column solutions, they showed that the resulting CLIME estimator has a large probability to be positive definite.

The second category of methods resort to the Gaussian graphical models for the precision matrix estimation. Let $Y_1, \ldots, Y_p$ be the vertices of a graphical model $G(V, E)$, where V is the vertex set and E is the edge set. There is an edge between vertices j and k if and only if Y_j and Y_k are conditionally dependent given the remaining variables. Suppose that the random variables have a joint multivariate normal distribution with mean $\mathbf{0}$. This model is referred to as a Gaussian graphical model. The edge set E can be totally determined by the precision matrix Ω, since Y_j and Y_k are conditionally dependent if and only if $\omega_{jk} \neq 0$, where ω_{jk} is the (j,k)th entry of Ω. There are two major methodologies for the estimation of precision matrix under the Gaussian graphical models: the neighborhood selection approach and the penalized likelihood approach.

The study of Meinshausen and Bühlmann (2006) paved the way for the neighborhood selection approach. It is based on the relation between Ω and the coefficients of p linear regression models, where each component of $\boldsymbol{Y}$ is regressed on the remaining $p-1$ components. That is,

$$\boldsymbol{y}_j = \sum_{k \neq j} \beta_{jk} \boldsymbol{y}_k + \boldsymbol{\epsilon}_j, \quad \boldsymbol{\epsilon}_j \sim N_n(0, \sigma^{jj} \boldsymbol{I}_n), \quad \text{for } j = 1, \ldots, p. \tag{1.1}$$

A nonzero off-diagonal entry of Ω corresponds to a nonzero regression coefficient in (1.1). The identification and estimation of the nonzero ω_{jk}'s are boiled down to variable selection and estimation in these linear regression models. Various regularized estimation methods for linear regression

models have been used in this framework. The approach of LASSO was adopted by Meinshausen and Bühlmann (2006) separately for each of the p regression models in (1.1). The Dantzig selector was similarly applied in Yuan (2010). The LASSO neighborhood selection procedure was considered in Zhou et al. (2011) with further thresholding to screen off the small coefficients. Then the precision matrix is estimated by maximum likelihood estimation subject to the support of Ω given by the estimated neighborhood sets. In the above application of LASSO and Dantzig selector, a single penalty parameter is used universally and hence imposes the same regularization level on all the p models in (1.1). Sun and Zhang (2013) applied the scaled LASSO, which avoids the selection of penalty parameter, to each of the regressions. Implicitly, it has different penalty levels for different models. Since the aforementioned methods handle the p regression models independently, they did not leverage on the intrinsic symmetry of the precision matrix. Hence, the immediate neighborhood estimators are potentially contradictory and require to be refined by certain symmetrization rules. Regarding β_{jk} and β_{kj} as a group, a symmetric version of the neighborhood selection in Meinshausen and Bühlmann (2006) was derived by Friedman et al. (2010) with the paired grouped LASSO penalty on the joint least squares of the p regressions. Besides, integrating these p regression models together in a penalized joint weighted square error loss, Peng et al. (2009) proposed a method called Sparse PArtial Correlation Estimation (SPACE). This weighted square error loss has a similar effect to the scaled LASSO. Due to the heterogeneity of the error variances σ^{jj}'s, the imposition of different regularization levels has an edge over that of a single one, which is demonstrated in Section 7 of Sun and Zhang (2013).

Luo and Chen (2014a) studied two sequential approaches by employing the SLasso method. The first approach follows the separate neighborhood selection framework and solves each regression model by SLasso, which is referred to as SR-SLasso. In the second approach, SLasso is applied to an unweighted joint regression model formed by combining the p regression models together. Hence, it is dubbed as JR-SLasso. However, JR-SLasso overlooks the heterogeneity among these models, due to the equal weights.

The penalized likelihood approach maximizes the profile likelihood function of Ω with direct regularization on its entries. It can be equivalently expressed as the following minimization problem:

$$\arg\min_{\Omega \succ 0} \Big\{ \operatorname{tr}(\hat{\Sigma}_n \Omega) - \log\det(\Omega) + \sum_{j,k=1}^{p} p_{\lambda_{jk}}(|\omega_{jk}|) \Big\}, \tag{1.2}$$

where $\Omega \succ 0$ denotes that Ω is positive definite, and $p_{\lambda_{jk}}$ is a penalty function with tuning parameter λ_{jk}. For simplicity, let $\lambda_{jk} = \lambda$ for any $j, k \in \{1, \ldots, p\}$. Different from the neighborhood selection approach, this formulation can better incorporate Ω's symmetry and positive definite constraint into the estimation process. This approach was first considered by using the ℓ_1 penalty (Yuan and Lin, 2007; Banerjee et al., 2008), hence the problem in (1.2) is convex. Unfortunately, the interior-point algorithm in Yuan and Lin (2007), which is adapted from the algorithm for the "max-dat" problem in Vandenberghe et al. (1998), is generally not efficient to handle the high-dimensional data. Inspired by the block coordinate descent algorithm for solving Σ (rather than Ω) in Banerjee et al. (2008), Friedman et al. (2008) developed the graphical LASSO (GLasso) algorithm to effectively tackle the high-dimensional computation. In addition to the ℓ_1

penalty, other forms of regularization have also been considered. The adaptive Lasso was considered in Zhou et al. (2009) to improve the consistency of recovering the underlying graphs. The SCAD penalty was exploited by Fan et al. (2009) to reduce the estimation bias. To take advantage of the GLasso algorithm, the local linear approximation (Zou and Li, 2008) was suggested to transform the SCAD penalized problem into some LASSO penalized problems. The regularization with a general penalty was studied in Lam and Fan (2009). Instead of the above elementwise penalty, Friedman et al. (2010) took advantage of the grouped LASSO and considered the column-wise regularization $\lambda \sum_j \|\Omega_{-j,j}\|_2$, where $\Omega_{-j,j}$ is the jth column of Ω without ω_{jj} and $\|\cdot\|_2$ is the vector ℓ_2 norm. This penalty groups all the edges connected to a given vertex. Such consideration corresponds to the graph that is sparse in its vertices but not in its edges.

Nevertheless, the Gaussian graphical models have a major limitation due to its assumption of normality, since, in many practical problems, the networked random variables are rarely normal. The nonparanormal graphical models, proposed by Liu et al. (2009), greatly weaken this normality assumption and are more flexible in practice. They suppose that the Gaussian variables $Y_1, \ldots, Y_p$ are latent, and that the observable variables are $X_1, \ldots, X_p$. There are univariate monotone functions $\{f_j\}_{j=1}^p$ such that $f_j(X_j) = Y_j$, for $j = 1, \ldots, p$. That is, $f(\boldsymbol{X}) = (f_1(X_1), \ldots, f_p(X_p))^\tau \sim N_p(\mathbf{0}, \Sigma)$, where v^τ denotes the transposition of a vector v. If f_j's are differentiable, it has been shown that, the conditional dependent relationships among X_j's are preserved in $f_j(X_j)$'s and still encoded in $\Omega = \Sigma^{-1}$. After estimating the functions f_j's and transforming the data accordingly, the precision matrix Ω and its sparsity pattern can be estimated by the meth-

ods for Gaussian graphical models. The nonparanormal SKEPTIC method in Liu et al. (2012) further strengthens the estimation robustness with nonparametric rank-based statistics. Specifically, it adopts the Spearman's ρ and Kendall's τ to estimate the correlation matrix of $f_1(X_1), \ldots, f_p(X_p)$ without explicitly estimating the transformation functions f_j's.

1.2.2 Regularized Variable Selection and Estimation in Linear Regression Models

The aforementioned regularized precision matrix estimation methods evolve from their counterparts in the linear regression models. Now we discuss them in details. Consider the following linear regression model:

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1.3}$$

where ε_i's are i.i.d. $N(0, \sigma^2)$, and the number of covariates p is far more than the sample size n. It is reasonable to assume that only a small set of β_j's are nonzero. Traditional stepwise variable selection methods ignore the stochastic errors inherited from the discrete selection processes and suffer from high variabilities. They are also computationally infeasible in the presence of substantial covariates. The regularized estimation approach minimizes the following penalized least squares:

$$\sum_{j=1}^{p} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \sum_{j=1}^{p} p_\lambda(|\beta_j|), \tag{1.4}$$

where p_λ is a penalty function with $\lambda > 0$ to be its regularization parameter. If $\lambda = 0$, it reduces to the OLS. As λ increases, the coefficients are

continuously shrunk towards 0, which contributes to the stability of the estimation. If the penalty is singular at the origin, part of the coefficients can be constricted to exactly 0, leading to the parsimony and interpretability of the fitted model. Such penalized framework can be easily extended to the generalized linear models by adding the penalties to the likelihood functions.

Frank and Friedman (1993) outlined the ℓ_q penalty, where $p_\lambda(|\beta_j|) = \lambda\|\beta_j\|_q$, as a generalization of the ridge regression where $q = 2$. The pioneer work in this area was presented by Tibshirani (1996) with the well-known least absolute shrinkage and selection operator (LASSO), which is equivalent to the ℓ_1 penalty. It is parallel to the basis pursuit for the wavelet regressions (Chen et al., 1998). This ℓ_1 penalty, on the other hand, also induces significant bias to the estimation. To enhance the estimation accuracy, the smoothly clipped absolute deviation (SCAD) penalty (Fan and Li, 2001) was designed to be bounded by a constant, thus some large coefficients are not penalized. Fan and Li (2001) also concluded that a good regularized estimation procedure should satisfy the so-called oracle property. It covers the procedure's two aspects: variable selection is asymptotically consistent and the nonzero coefficients are estimated as if the true model were known in advance. The oracle property of SCAD has been established when p is finite or divergent, provided a properly chosen regularization parameter λ (Fan and Li, 2001; Fan et al., 2004). However, LASSO's oracle property is more involved. Its coefficients estimation asymptotics can be shown under proper conditions; whereas the variable selection consistency generally does not hold (Knight and Fu, 2000). In the effort to explore the LASSO procedure's oracle property, Zou (2006) assigned different weights to the penalized co-

efficients, i.e., $p_\lambda(|\beta_j|) = \lambda w_j|\beta_j|$ and developed the adaptive LASSO. On the condition of a proper λ and data-driven, cleverly selected weights w_j's, adaptive LASSO is an oracle procedure for fixed p. Its asymptotic properties in the situation of p diverging with n were examined by Huang et al. (2008). Finally, it is found that the so-called Irrepresentable Condition on the design matrix is almost necessary and sufficient for the LASSO procedure to be variable selection consistent (Meinshausen and Bühlmann, 2006; Zhao and Yu, 2006). Unfortunately, it is a very strong condition and can be easily violated if the covariates are high-dimensional and correlated. Candes and Tao (2007) proposed the Dantzig selector, which has its ℓ_2-norm loss bounded with large probability. Through their simulation study and the comparison analysis in Bickel et al. (2009), similar performance can be expected between LASSO and Dantzig selector. The minimax concave penalty (MCP) in the form of $p_\lambda(|\beta_j|) = \lambda \int_0^{|\beta_j|}(1 - x/(\gamma\lambda))_+ dx$ was considered by Zhang (2010), which approaches to the ℓ_1 penalty as $\gamma \to \infty$ and converges to a constant as $\gamma \to 0^+$. Similar to the design of SCAD penalty, large coefficients are not subject to shrinkage. Assuming a weaker condition, i.e., the sparse Riesz condition, on the design matrix, the MCP procedure can select the correct model with a high probability for a universal λ, even when p is much larger than n. Sun and Zhang (2012) proposed an adaptive way for the LASSO-type procedure to determine the penalty parameter λ by the estimated noise level σ^2. In their iterative algorithm, σ^2 is estimated by the current mean residual sum of squares and then used to scale λ; the coefficients are estimated with the updated λ. Moreover, Yuan and Lin (2006) introduced the concept of grouped regularization, where the model selection and estimation are based on groups of features.

It is worth noting that many foregoing methods' selection consistency depends on the choice of λ, which is commonly suggested to be tuned by the cross validation. However, this technique is computationally costly and usually designed to choose the penalty parameter that minimizes the prediction error. It is acknowledged that there is usually a contradiction between optimal prediction and variable selection consistency (Leng et al., 2006; Meinshausen and Bühlmann, 2006). In other words, the procedures that select λ in the criterion of prediction accuracy tend to include many noise features and thus are generally inconsistent in variable selection.

Recently, Luo and Chen (2014b) introduced a sequential variable selection procedure by solving the LASSO penalized least squares, which is dubbed as sequential LASSO (SLasso). Essentially, it is equivalent to the orthogonal matching pursuit (OMP) in Pati et al. (1993), as long as only one variable is selected at each step. Suppose the model is standardized. Compare SLasso and the LASSO solution path implied by the least angle regression (LARS) in Efron et al. (2004). Conceptually, both of them select the covariate that has the largest absolute correlation with the current residual at each step. However, the residual implemented by LARS is actually calculated with the shrunk estimated coefficients, thus it is larger than the true residual, which is used by SLasso. Such overestimation makes LARS susceptible to the covariates that have highly spurious correlations with the already selected ones. Furthermore, consider the forward stepwise selection (FSS), which selects the covariate such that the subsequent residual is minimized in ℓ_2 norm at each step. In other words, this procedure identifies variables based on a scaled version of the absolute correlations in SLasso. The scaler is the corresponding covariate's ℓ_2-norm projection

into the space spanned by the unselected features and thereby less than 1. It turns out that FSS inclines to selected a group of correlated covariates. Moreover, an appropriate stopping rule is also crucial for a sequential method. SLasso adopts the extended Bayesian information criterion (EBIC; Chen and Chen, 2008) to stop its selection procedure. The EBIC not only considers the newly identified variable's contribution to the reduction of the residual, but also penalizes it for the consequent augmentation to the size of the selected model. The latter aspect is often ignored by many sequential methods' stopping rules, which are likely to include many irrelevant features, such as those for the OMP in Cai and Wang (2011).

1.2.3 Sparse Estimation in Multivariate Response Regression Models

The regularized coefficient estimation of the univariate response regression models has been extensively studied in various literature. Their natural extensions are the multivariate response regression models. Let $Y \in \mathbf{R}^{n\times p}$ be the response matrix and $X \in \mathbf{R}^{n\times q}$ be the covariate matrix. Each row of Y is a sample of the p-dimensional response $\boldsymbol{Y} = (Y_1, \ldots, Y_p)^\tau$ and the corresponding row of X is a sample of the covariates $X_1, \ldots, X_q$. Define the multivariate response regression model as

$$Y = XB + E, \tag{1.5}$$

where $B = (\beta_{ij}) \in \mathbf{R}^{q\times p}$ is the coefficient matrix, E is the random error matrix and its row vectors are i.i.d. $N_p(\mathbf{0}, \Sigma)$. Naively, B can be fitted by decomposing (1.5) into a series of marginal linear regression models with

univariate response, thus the coefficients for different responses are estimated independently. This approach works well if $Y_1, \ldots, Y_p$ are truly independent. Nevertheless, in practice, these responses, rather than being independent, are probably correlated, especially under the high-dimensional circumstances. These correlations may arise from different responses sharing some common covariates or the correlated random errors. Therefore, ignoring the responses' dependent structure, the naive approach is inefficient and tends to be inaccurate, especially when this dependency is significant.

One way to jointly predict the multivariate response in (1.5) is through the reduced-rank regression (RRR), which optimizes the log-likelihood function or the least squares subject to a rank constraint on the coefficient matrix, i.e., $\text{rank}(B) = r \leq \min(p, q)$ (See Anderson, 1951; Izenman, 1975; Reinsel and Velu, 1998). Under this constraint, the coefficient matrix is the product of a factor matrix and its load matrix with the given rank. Such estimator typically lacks interpretability. Similar to the RRR, the method factor estimation and selection (FES), proposed by Yuan et al. (2007), imposes the penalty on the sum of the coefficient matrix's singular values, a.k.a. the Ky Fan norm penalty. Different from the discrete property of the previous RRR, FES yields reduced-rank estimates along a continuous regularization path. However, this method focuses on the dimension reduction and factor selection rather than the covariate selection. Chen and Huang (2012) enhanced the interpretability of RRR by the sparse reduced-rank regression (SRRR) method, where the grouped LASSO penalty is applied to the row vectors of the factor matrix. This can select the important factors and induce sparsity to B simultaneously.

Furthermore, when the number of covariates is large, sometimes, it is desirable to identify a small group of them to predict all the responses, resulting in more interpretable fitted model. Aiming at this goal, Turlach et al. (2005) penalized the least squares with the sum of the largest absolute coefficients in B's row vectors, i.e., $\sum_{i=1}^{q} \max(|\beta_{i1}|, \ldots, |\beta_{ip}|)$. It can be considered as an extension of the LASSO method to the multivariate response case. Due to the large bias induced by the ℓ_∞-norm related penalty, it is only suggested to perform covariate selection. On the other hand, this grouped penalty shrinks some covariates' coefficients for all responses to 0 and hence efficiently avoids the over-fitting problem that may be caused when separate variable selection is performed. It is noticeable that each of the selected predictors is used in the prediction of all the responses. In other words, there is no sparsity within the selected covariates' coefficients. Such sparsity was particularly considered by Peng et al. (2010) in the method called remMap, which employs the combination of two penalties. One is the grouped LASSO on the row vectors of B, which performs similarly to the above ℓ_∞-norm related penalty so as to identify the master predictors; the other is ℓ_1 penalty on each entry of B, which further induces sparsity over the master predictors' coefficients.

Effective as these methods are, they closely rely on the assumption that all the marginal univariate regressions reside in the same low-dimensional space, which may be too strong in practice. This approach also implicitly assumes that the responses' dependency attributes to the common predictors. Although Turlach et al. (2005) suggested to incorporate the responses' correlations via multiplying Y and X by $W^{1/2}$ in the data processing, where $W \succ 0$ is a weight matrix. Unfortunately, they did not specify the way to

obtain this weight matrix W, nor how it relates to the covariance matrix Σ. Therefore, none of these methods directly incorporates the dependent structure among the responses into the model fitting.

Breiman and Friedman (1997) first exploited the responses' correlations to improve the prediction accuracy. Suppose $\hat{B}^{OLS}$ is the OLS estimate of B. In their proposed Curds and Whey (CW) method, the final prediction has the form $X\hat{B}^{OLS}M$, where $M \in \mathbf{R}^{p\times p}$ can shrink the OLS prediction $X\hat{B}^{OLS}$ to attain an optimal linear combination. Obviously, CW is not suitable in the presence of high-dimensional challenge. Recent research incorporates the responses' correlations with the penalized likelihood framework and performs the simultaneous regularized estimation for the coefficient matrix B and the precision matrix $\Omega = \Sigma^{-1}$. It can be shown as the following optimization problem:

$$\arg\min_{B,\Omega} \Big\{ \operatorname{tr}[\frac{1}{n}(Y - XB)^{\tau}(Y - XB)\Omega] - \log\det(\Omega) + \lambda_1 p_1(B) + \lambda_2 p_2(\Omega) \Big\}. \tag{1.6}$$

This formulation was first considered by Rothman et al. (2010) in their method called multivariate regression with covariance estimation (MRCE). MRCE imposes the ℓ_1 penalty on both B and Ω, leading to their simultaneous sparse estimation. Although the optimization problem (1.6) for solving either B or Ω is convex with the other being fixed, it is not convex for solving them simultaneously. Hence, an alternate optimization scheme is adopted in the numerical implementation. If B is fixed, the problem is reduced to the precision matrix estimation in the Gaussian graphical models. Analogously, for fixed Ω, the coefficient matrix can be solved with a cyclical-coordinate descent version of the algorithm that is used to solve

the single LASSO problem in Friedman et al. (2007). They also mentioned that the estimation of the precision matrix is to improve the prediction accuracy, which can be dubbed as the supervised covariance estimation.

Instead of taking it as a multivariate response regression problem, Yin and Li (2011) formulated it as a Gaussian graphical model (GGM) but with heterogeneous means, which abandons the constant mean vector assumption and allows the vertices to rely on some other covariates. To differentiate it from the standard GGM, it is called the conditional Gaussian graphical model (cGGM). This cGGM and the multivariate response regression model are essentially equivalent, even though they model the same problem from different perspectives. Their initial motivation was from a practical problem: identifying genes' conditional dependent structure at the expression level (the responses) with adjustment for the effect of gene variants' shared regulations (the covariates). They dealt with with the same LASSO penalized problem as in Rothman et al. (2010). However, in practice, they aimed at improving the estimation accuracy of the precision matrix through the study of the coefficient matrix, which is opposite to the objective of MRCE.

Within the same penalized likelihood framework, Lee and Liu (2012) considered the adaptive LASSO penalty and provided three variants for the estimation of coefficient matrix and precision matrix. The first two are plug-in methods. Specifically, the first method, plug-in joint weighted LASSO (PWL), applies the GLasso to get the estimate of Ω, then plugs it in to the adaptive LASSO penalized likelihood to solve B; the second method, plug-in weighted graphical LASSO (PWGL), focuses more on esti-

mating Ω, provided that the each response's coefficients are first estimated separately with adaptive LASSO. These two methods are similar to the ideas implied by the first two algorithms in MRCE without iteration. The third method, doubly penalized maximum likelihood (DML), estimates B and Ω simultaneously by optimizing the adaptive Lasso penalized problem in (1.6).

All these above methods using the alternative updating scheme cannot ensure the global optimum of the final estimates, especially when the parameters are high-dimensional. Their computational cost is another concern. Other than the penalized likelihood formulation, Wang (2013) tackled the problem from the perspective of conditional regressions, where each response is fitted against all the covariates and the other responses, similar to the neighborhood selection framework in the Gaussian graphical models. For each individual conditional regression, the adaptive LASSO procedure is applied for the related parameters' estimation. Since each conditional regression's optimization problem is convex, this method is able to achieve the global optimal estimates of the coefficient matrix and the precision matrix. Moreover, its computation involves no outer iteration and can be easily parallelized. Additionally, its identification of the sparsity pattern in Ω can be regarded as the adaptive LASSO regularized neighborhood selection. Such separate estimation ignores the symmetry of Ω, thus it needs a symmetrization rule to attain the final estimate.

1.3 Research Aims

Besides the SLasso in Luo and Chen (2014b) for the common linear models, another similar sequential procedure was successfully developed for the interactive linear models by He and Chen (2014). It has been found that they not only excel at the feature selection in the large-p-small-n scenario, but also appear computationally attractive. Initial attempt on the sequential edge detection in the Gaussian graphical models has also been realized in Luo and Chen (2014a).

The research presented in this thesis further explores the sequential approaches to the relevant problems in the graphical models and the multivariate response regression models. The specific objectives of the thesis include:

- Development of a sequential edge detection method in the nonparanormal graphical models;
- Applications of this edge detection technique to the precision matrix estimation;
- Extension of the sequential framework to the joint estimation of the coefficient matrix and the precision matrix in the multivariate response regression models.

For the graphical models, we assume that they are high-dimensional but sparse. Particularly, there are a substantial number of vertices but only a few edges in the graph. This is equivalent to assuming the elementwise

sparsity in the corresponding precision matrix. For the multivariate regression models, we assume that both the coefficient matrix and the precision matrix are sparse. Specifically, we suppose that each marginal regression's coefficient vector is sparse; however, we do not require that the true model should only rely on a small subset of the covariates. Consequently, we consider that the responses' correlations originate from the correlated random errors.

1.4 Outline of Thesis

In Chapter 2, a sequential scaled pairwise selection (SSPS) method is derived for the edge detection in the nonparanormal graphical models. Its selection consistency is shown under proper conditions. Based on the SSPS method, two approaches to the precision matrix estimation in sparse Gaussian graphical models are discussed in Chapter 3. The first approach estimates the precision matrix by optimizing over the support recovered by the SSPS edge detection. The second approach adapts some existing methods by performing an SSPS screening in the beginning. Chapter 4 is on the joint covariance matrix and precision matrix estimation in multivariate response regression models. Two sequential conditional regression (SCR) methods are considered. Their selection consistency is provided. Chapter 5 concludes this research and discusses about the future work.

Chapter 2

A Sequential Scaled Pairwise Selection Approach to Edge Detection in Nonparanormal Graphical Models

In this chapter, we tackle the problem of edge detection in high-dimensional nonparanormal graphical models by the proposed sequential scaled pairwise selection (SSPS) method. This chapter is organized as follows. We first introduce the Gaussian graphical models and the nonparanormal graphical models in Section 1. Then our SSPS method is presented in Section 2. Its selection consistency is established in Section 3. In Section 4, we report the results of the simulation studies, where SSPS is compared with other existing methods. Further discussion is provided in Section 5.

2.1 Gaussian Graphical Model and Nonparanormal Graphical Model

An undirected graph $G(V, E)$, consisting of a vertex set V and an edge set E, is usually exploited to model the network structure of a finite number of random variables $Y_1, \ldots, Y_p$. The vertex set V is composed of the random variables. The edge set E describes their pairwise conditionally dependent relationships. An edge exists between vertices j and k if and only if Y_j and Y_k are conditionally dependent given the remaining variables. The assumption of the random variables' multivariate normality is widely adopted in the analysis of such relationships, resulting in the popular Gaussian graphical model. In the rest of this section, we introduce the Gaussian graphical model and its extension, the nonparanormal graphical model.

2.1.1 Gaussian Graphical Model

In a Gaussian graphical model, the p random variables are assumed to have a joint multivariate normal distribution, i.e., $\boldsymbol{Y} = (Y_1, \ldots, Y_p)^\tau \sim N_p(\boldsymbol{\mu}, \Sigma)$. Denote the precision matrix by $\Omega = \Sigma^{-1}$ with its (j, k)th entry to be ω_{jk}. Without loss of generality, let $\boldsymbol{\mu} = \mathbf{0}$. It is well known that $\omega_{jk} = 0$ if and only if Y_j and Y_k are conditionally independent given the remaining random variables. Therefore, the precision matrix Ω completely captures the pairwise conditional independence among $Y_1, \ldots, Y_p$, or equivalently the underlying graphical model's edge set E. It is of particular interest to estimate Ω based on n i.i.d. samples from the distribution of $\boldsymbol{Y}$, where n is far less than the number of unknown parameters $p(p+1)/2$. An important

part of this problem is to identify the nonzero entries in Ω, which is also termed as edge detection and can be regarded as a sub-problem of the precision matrix estimation. Specifically, the edge set E can be represented by an adjacent matrix $\Theta = (\theta_{jk})_{p\times p}$, where $\theta_{jk} = \theta_{kj} = 1$ if and only if there is an edge between vertices j and k, $\theta_{jk} = \theta_{kj} = 0$ otherwise. By convention, let $\theta_{jj} = 0$, for $j = 1, \ldots, p$. Due to the relation between Ω and E, $\theta_{jk} = 1$ if and only if $\omega_{jk} \neq 0$, for any $j \neq k$.

In the context of the sparse high-dimensional graph, a prevalent approach to precision matrix estimation along with edge detection is via optimizing the following penalized likelihood problem:

$$\arg\min_{\Omega \succ 0} \Big\{ \operatorname{tr}(\hat{\Sigma}_n \Omega) - \log\det(\Omega) + p_\lambda(\Omega) \Big\}, \tag{2.1}$$

where $\hat{\Sigma}_n$ is the sample covariance matrix and $p_\lambda(\cdot)$ is a penalty function on each entry of Ω with λ to be the regularization parameter. The penalty imposes elementwise shrinkage on the estimate of Ω to achieve estimation parsimony. Penalties such as LASSO, adaptive LASSO and SCAD have been considered.

An alternative formulation to tackle the sparse precision matrix estimation is through the neighborhood selection. In what follows, let $\boldsymbol{Y}_{-j} = (Y_1, \ldots, Y_{j-1}, Y_{j+1}, \ldots, Y_p)^\tau$ and $\Sigma_{-i,-j}$ be the submatrix of Σ that deletes its ith row and jth column. Analogously, $\Sigma_{-i,j}$ and $\Sigma_{i,-j}$ refer to the jth column of Σ with its ith element omitted and the ith row of Σ with its jth element omitted. The conditional distribution of Y_j given $\boldsymbol{Y}_{-j}$ can be

expressed as

$$Y_j|\boldsymbol{Y}_{-j} \sim N\Big(\Sigma_{j,-j}\Sigma_{-j,-j}^{-1}\boldsymbol{Y}_{-j}, \quad \Sigma_{jj} - \Sigma_{j,-j}\Sigma_{-j,-j}^{-1}\Sigma_{-j,j}\Big). \tag{2.2}$$

This can be transformed to the linear regression model that predicts Y_j with $Y_1, \ldots, Y_{j-1}, Y_{j+1}, \ldots, Y_p$. That is,

$$Y_j = \boldsymbol{Y}_{-j}^{\tau}\boldsymbol{\beta}_j + \epsilon_j, \qquad \epsilon_j \sim N(0, \sigma^{jj}), \quad j = 1, \ldots, p, \tag{2.3}$$

where $\boldsymbol{\beta}_j = \Sigma_{-j,-j}^{-1}\Sigma_{-j,j}$ is the regression coefficient vector, ϵ_j is the random error that is independent of $\boldsymbol{Y}_{-j}$ and has variance $\sigma^{jj} = \Sigma_{jj} - \Sigma_{j,-j}\Sigma_{-j,-j}^{-1}\Sigma_{-j,j}$. Denote $\boldsymbol{\beta}_j$'s coordinate that corresponds to the covariate Y_k $(k \neq j)$ by β_{jk}. Moreover, by inverting Σ blockwise, the precision matrix Ω is equal to

$$\begin{pmatrix} \Sigma_{jj} & \Sigma_{j,-j} \\ \Sigma_{-j,j} & \Sigma_{-j,-j} \end{pmatrix}^{-1} = \begin{pmatrix} \omega_{jj} & -\omega_{jj}\Sigma_{j,-j}\Sigma_{-j,-j}^{-1} \\ -\omega_{jj}\Sigma_{-j,-j}^{-1}\Sigma_{-j,j} & \omega_{jj}\Sigma_{-j,-j}^{-1}\Sigma_{-j,j}\Sigma_{j,-j}\Sigma_{-j,-j}^{-1} \end{pmatrix}, \tag{2.4}$$

where $\omega_{jj} = (\Sigma_{jj} - \Sigma_{j,-j}\Sigma_{-j,-j}^{-1}\Sigma_{-j,j})^{-1}$. The above equation implies that $\Omega_{-j,j} = -\omega_{jj}\Sigma_{-j,-j}^{-1}\Sigma_{-j,j}$, for $j = 1, \ldots, p$. Comparing (2.3) and (2.4), immediately, we have

$$\boldsymbol{\beta}_j = -\frac{\Omega_{-j,j}}{\omega_{jj}} \quad \text{and} \quad \sigma^{jj} = \frac{1}{\omega_{jj}}. \tag{2.5}$$

Consequently, the estimation of Ω can be transformed to the estimation of the regression coefficients β_{jk}'s and the random errors' variances σ^{jj}'s. Particularly, $\omega_{jk} = 0$ suggests that $\beta_{jk} = 0$. In conclusion, the neighborhood selection approach accomplishes the task by solving the regularized linear

regression models in (2.3) with several i.i.d. samples of $\boldsymbol{Y}$.

2.1.2 Nonparanormal Graphical Model

A nonparanormal graphical model is an extension of the Gaussian graphical model, but relieves the normality restriction. It is characterized as follows: there are monotone univariate functions $\{f_j\}_{j=1}^p$, such that $f(\boldsymbol{X}) = (f_1(X_1), \ldots, f_p(X_p))^\tau$ follows a multivariate normal distribution $N_p(\boldsymbol{\mu}, \Sigma)$. The distribution of $\boldsymbol{X} = (X_1, \ldots, X_p)^\tau$ is referred to as a nonparanormal distribution, denoted by $\boldsymbol{X} \sim \mathrm{NPN}_p(\boldsymbol{\mu}, \Sigma)$. Hence, $f(\boldsymbol{X})$, rather than $\boldsymbol{X}$, indicates a Gaussian graphical model. To make the model identifiable, the following constraints are imposed: $\mu_j = E[f_j(X_j)] = E[X_j]$ and $\sigma_j^2 = \mathrm{Var}[f_j(X_j)] = \mathrm{Var}[X_j]$, for $j = 1, \ldots, p$. In other words, f_j's preserve the means and variances.

We further look into the transformation functions f_j's. Let $F_j(x)$ be the marginal distribution function of X_j. Since f_j is monotone (without loss of generality, assume that it is increasing), we have

$$F_j(x_j) = P(X_j \leq x_j) = P(f_j(X_j) \leq f_j(x_j)) = \Phi\left(\frac{f_j(x_j) - \mu_j}{\sigma_j}\right), \quad (2.6)$$

where Φ is the c.d.f. of the standard normal distribution. Hence, the marginal transformation function f_j is determined by

$$f_j(x_j) = \sigma_j \Phi^{-1}(F_j(x_j)) + \mu_j. \quad (2.7)$$

For the sake of convenience, denote $f_j(X_j)$ by Y_j and let $\boldsymbol{Y} = (Y_1, \ldots, Y_p)^\tau$.

Hence, under the nonparanormal model, $\boldsymbol{Y} \sim N_p(\boldsymbol{\mu}, \Sigma)$. Without loss of generality, assume that $\boldsymbol{\mu} = \mathbf{0}$. Still let $\Omega = \Sigma^{-1}$ be the precision matrix of $\boldsymbol{Y}$. Interestingly, the undirected graphs of $\boldsymbol{X}$ and $\boldsymbol{Y}$ are identical and both are encoded in Ω (Liu et al., 2009). If the transformation functions f_j's are known, the problem of edge detection for the nonparanormal graphical model is then reduced to the same problem for the Gaussian graphical model. This can be described by the following two-step procedure:

(1) Estimate the functions $f_1, \ldots, f_p$ and transform the nonparanormal data to the Gaussian data;

(2) Estimate the graph with the transformed data by methods for Gaussian graphical models.

2.2 The Sequential Scaled Pairwise Selection Method

In this section, we detect edges in the nonparanormal graphical models with a sequential method, which belongs to the neighborhood selection approach. Before presenting our sequential scaled paired selection (SSPS) method, we first introduce the variable selection technique SLasso in Luo and Chen (2014b) and comment on the two sequential edge detection methods: SR-SLasso and JR-SLasso in Luo and Chen (2014a).

2.2.1 Preliminary: SLasso, SR-SLasso and JR-SLasso

SLasso Consider the variable selection in the linear regression model: $\boldsymbol{y} = X\boldsymbol{\xi} + \boldsymbol{e}$, where $\boldsymbol{y}$ and columns of X are normalized to have mean 0

and standard deviation 1. The SLasso approach selects the variables by sequentially minimizing the partially ℓ_1 penalized squared error losses, where the coefficients of the already selected variables are not penalized. This sequential procedure is terminated when the selected model's EBIC (Chen and Chen, 2008) attains a minimum. For the purpose of variable selection, it suffices to identify the active set of the minimization problem at each step, and the minimization does not need to be carried out completely. Here, the active set is the index set of the features with nonzero coefficients in the given optimization problem. Let S be the column index set of X, s be any of its subset and s^c be the complementary set of s. The submatrix of X consisting of columns with indices in s is denoted by $X(s)$. Its projection matrix is denoted by $H(s)$, i.e., $H(s) = X(s)[X^\tau(s)X(s)]^{-1}X^\tau(s)$. Suppose s_* is the index set of the already selected variables at a certain step. Then the active set at this step is given by the covariate that has the largest absolute correlation with the current residual. This gives rise to the following simple computation algorithm for the SLasso procedure.

The SLasso Algorithm

Initial step: Set $s_* = \emptyset$ and $\tilde{\boldsymbol{y}} = \boldsymbol{y}$.

Iterative step: Compute $\boldsymbol{x}_j^\tau \tilde{\boldsymbol{y}}$ for $j \in s_*^c$, and identify j^* such that $|\boldsymbol{x}_{j^*}^\tau \tilde{\boldsymbol{y}}| = \max_{j \in s_*^c} |\boldsymbol{x}_j^\tau \tilde{\boldsymbol{y}}|$. Let $s_{\text{tmp}} = s_* \cup \{j^*\}$. If $\text{EBIC}(s_{\text{tmp}}) < \text{EBIC}(s_*)$, let $s_* = s_{\text{tmp}}$, $\tilde{\boldsymbol{y}} = [\boldsymbol{I}_n - H(s_*)]\boldsymbol{y}$, and continue; otherwise, stop.

Output: s_* is the selected variables' index set and the nonzero coefficients are estimated by OLS, i.e., $\boldsymbol{\xi}(s_*) = [X^\tau(s_*)X(s_*)]^{-1}X^\tau(s_*)\boldsymbol{y}$.

In the above algorithm,

$$\text{EBIC}(s) = n \ln(\frac{\|[\boldsymbol{I}_n - H(s)]\boldsymbol{y}\|_2^2}{n}) + |s| \ln n + 2\gamma \ln \binom{|S|}{|s|},$$

where γ is a regularization parameter and $|s|$ denotes the cardinality of a set s. For more details of SLasso, see Luo and Chen (2014b).

SR-SLasso and JR-SLasso Both SR-SLasso and JR-SLasso aim at edge detection, or equivalently estimating the adjacent matrix Θ, in the Gaussian graphical models via solving the regression models in (2.3) with n i.i.d. samples of $\boldsymbol{Y}$ by the foregoing SLasso method. Their main difference is: SR-SLasso solves the p models separately; whereas JR-SLasso handles them in a joint fashion.

The SR-SLasso algorithm is briefly described here. For the first regression, apply SLasso directly to obtain Y_1's neighborhood index set, referred to as s_{1*}; for Y_j's regression ($j = 2, \ldots, p$), adjust the SLasso Algorithm's initial step with $s_* = \{i : j \in s_{i*}, i < j\}$ and $\tilde{\boldsymbol{y}} = [\boldsymbol{I}_n - H(s_*)]\boldsymbol{y}$, and obtain s_{j*}. Combine the p neighborhood sets $s_{1*}, \ldots, s_{p*}$ with the "OR" rule, which claims an edge between vertices j and k if either $k \in s_{j*}$ or $j \in s_{k*}$. Although the idea is feasible, this algorithm is asymmetric, which is caused by the modification to the initial step of the SLasso Algorithm. Consequently, solving the regressions in different orders probably leads to inconsistent results. It would be more appropriate if the original algorithm is directly implemented on all the regressions. Without using the information in the previously detected neighborhood sets, it appears less computationally efficient. However, such symmetrization correction makes it more

flexible for the parallel computation, which cannot be applied to the original SR-SLasso algorithm. Additionally, either the "OR" rule or the "AND" rule can be adopted to get the final estimate.

Conceptually, JR-SLasso employs SLasso to solve the single linear regression that results from merging the p regression models in (2.3), which can be written as:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{pmatrix} = \begin{pmatrix} \boldsymbol{Y}_{-1}^{\tau} & 0 & \cdots & 0 \\ 0 & \boldsymbol{Y}_{-2}^{\tau} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{Y}_{-p}^{\tau} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_p \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_p \end{pmatrix}. \tag{2.8}$$

Compared to SR-SLasso, this joint formulation considers all the relevant absolute correlations simultaneously and thus is more favorable. Nevertheless, dealing directly with the simply combined regression model (2.8) implies that a generic penalty is imposed on $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p$ uniformly. It is acknowledged that the penalty level should be proportional to the regression variance. In this case, the variances are $\frac{1}{\omega_{jj}}$'s, and it is impractical to assume that they are homogeneous. Therefore, JR-SLasso implicitly ignores such heterogeneity among ω_{jj}'s. The necessity of handling the manifold regression variances becomes our motivation to propose the following SSPS method.

2.2.2 SSPS Method

We now introduce the SSPS method under the nonparanormal graphical models, and follow the two-step procedure shown in Section 2.1.2 and the

notations therein.

Gaussian Transformation Let $(X_{1i}, \ldots, X_{pi})^\tau$ be a sample from the nonparanormal distribution $\text{NPN}_p(\mathbf{0}, \Sigma)$, where $i = 1, \ldots, n$. For $1 \leq j \leq p$, denote the empirical likelihood estimator of F_j by $\hat{F}_j$, which is based on n samples of X_j, i.e. $X_{j1}, \ldots, X_{jn}$, subject to $EX_j = 0$ and $\text{Var}(X_j) = 1$. Since $\Phi^{-1}(t)(t \in [0, 1])$ is extremely sensitive when t is close to 0 or 1, the following Winsorized estimator of F_j is suggested by Liu et al. (2012):

$$\tilde{F}_j(t) = \delta_n \cdot \mathbf{I}(\hat{F}_j(t) < \delta_n) + \hat{F}_j(t) \cdot \mathbf{I}(\delta_n \leq \hat{F}_j(t) \leq 1-\delta_n) + (1-\delta_n) \cdot \mathbf{I}(\hat{F}_j(t) > 1-\delta_n), \tag{2.9}$$

where $\delta_n = 1/(n+1)$. The transformation function f_j is then estimated by

$$\hat{f}_j(t) = \Phi^{-1}(\tilde{F}_j(t)), \qquad t \in (-\infty, +\infty). \tag{2.10}$$

Liu et al. (2012) also showed that, for any $0 < \gamma < 1$,

$$\sup_{f_j^{-1}(-c_n) \leq t \leq f_j^{-1}(c_n)} |\hat{f}_j(t) - f_j(t)| \to 0, \text{ in probability}, \tag{2.11}$$

where $c_n = \left[\frac{7}{4}(1-\gamma)\ln n\right]^{1/2}$. This result implies that the normal-score estimate $(\hat{f}_1(X_1), \ldots, \hat{f}_p(X_p))^\tau$ converges, in distribution, to $(f_1(X_1), \ldots, f_p(X_p))^\tau$. Asymptotically, $(\hat{f}_1(X_1), \ldots, \hat{f}_p(X_p))^\tau$ has the same distribution as $(f_1(X_1), \ldots, f_p(X_p))^\tau$. By an abuse of notation, in what follows, we denote $\hat{f}_j(X_j)$ by Y_j as well. Without loss of generality, we can still assume $\boldsymbol{Y} \sim N_p(\mathbf{0}, \Sigma)$.

Edge Detection Let $Y = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_p)$, where $\boldsymbol{y}_j = (Y_{j1}, \ldots, Y_{jn})^\tau = (\hat{f}_j(X_{j1}), \ldots, \hat{f}_j(X_{jn}))^\tau$ is the transformed sample of X_j. Let Z_j denote the matrix obtained by omitting the jth column of Y. According to the regression models in (2.3), we have

$$\boldsymbol{y}_j = Z_j \boldsymbol{\beta}_j + \boldsymbol{\epsilon}_j, \ \ j = 1, \ldots, p, \tag{2.12}$$

where $\boldsymbol{\beta}_j$ is a $(p-1)$-vector of regression coefficients and $\boldsymbol{\epsilon}_j$ is a n-vector of i.i.d. normal variables with mean zero and variance σ^{jj}. Because σ^{jj} varies with j, the p regressions in (2.12) are heterogeneous. The homogeneousness can be achieved by scaling the jth regression with $1/\sqrt{\sigma^{jj}}$. Since σ^{jj}'s are unknown, we estimate them by the iterative algorithm in the scaled Lasso (Sun and Zhang, 2013), which goes as follows. Let $\lambda_0 = A\sqrt{2(\log p/a)/n}$ for fixed $A > 1$ and $a > 0$. Let $\hat{\boldsymbol{\beta}}_j(\lambda)$ be the LASSO solution path of the jth regression in (2.12). Given the solution path, $\hat{\sigma}^{jj}$ can be obtained as follows. At the initial step, set $\lambda = \lambda_0$, then repeat the following procedure until convergence:

$$\hat{\boldsymbol{\beta}}_j \leftarrow \hat{\boldsymbol{\beta}}_j(\lambda), \qquad \hat{\sigma}^{jj} \leftarrow \frac{1}{n}\|\boldsymbol{y}_j - Z_j\hat{\boldsymbol{\beta}}_j\|_2^2, \qquad \lambda \leftarrow \sqrt{\hat{\sigma}^{jj}}\lambda_0.$$

Then $\boldsymbol{y}_j$ is scaled by $1/\sqrt{\hat{\sigma}^{jj}}$ and let $\tilde{\boldsymbol{y}}_j = \boldsymbol{y}_j/\sqrt{\hat{\sigma}^{jj}}$, for $j = 1, \ldots, p$. Now define

$$\boldsymbol{y} = \begin{pmatrix} \tilde{\boldsymbol{y}}_1 \\ \tilde{\boldsymbol{y}}_2 \\ \vdots \\ \tilde{\boldsymbol{y}}_p \end{pmatrix}, \ X = \begin{pmatrix} Z_1 & 0 & \cdots & 0 \\ 0 & Z_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Z_p \end{pmatrix}, \ \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1/\sqrt{\sigma^{11}} \\ \boldsymbol{\beta}_2/\sqrt{\sigma^{22}} \\ \vdots \\ \boldsymbol{\beta}_p/\sqrt{\sigma^{pp}} \end{pmatrix}. \tag{2.13}$$

We then have a homogeneous regression model: $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}$ is a np-vector of i.i.d. normal variables with mean zero and variance 1. Write the matrix X in the form of column blocks as

$$X = \begin{pmatrix} B_1 & \cdots & B_j & \cdots & B_p \end{pmatrix}, \text{ where } B_j = \begin{pmatrix} 0 \\ Z_j \\ 0 \end{pmatrix}.$$

We index the columns of X by double subscripts. The column containing $\boldsymbol{y}_k$ in the block B_j is denoted by $\boldsymbol{x}_{jk}$. Let s be a subset of the column indices of X. We denote by $X(s)$ the submatrix consisting of the columns of X with indices in s. By an abuse of notation, we also use s to denote the regression model with design matrix $X(s)$. Let $\boldsymbol{y}_s$ be the residual vector of $\boldsymbol{y}$ while projected into the space spanned by the columns of $X(s)$.

The sequential pairwise selection procedure is as follows. Let s be the index set of the columns in X which have already been selected (at the beginning, s is taken as the empty set $\emptyset$). To select the next pair, we project $\boldsymbol{y}_s$ into the space spanned by $(\boldsymbol{x}_{jk}, \boldsymbol{x}_{kj})$ for all $(jk) \in s^c$, and compute the ℓ_2 norm of each projected vector. The pair of indices with the largest norm are then selected. The model s is updated by including the newly selected pair. The EBIC is used as the stopping rule: if the current model's EBIC is smaller that of the previous model, the sequential procedure continues; otherwise, it stops. More details are described below.

In what follows, let $s_j = \{k : (jk) \in s\}$ and $\boldsymbol{y}_{js} = [\boldsymbol{I}_n - H(s_j)]\boldsymbol{y}_j$, where $H(s_j) = Z_j(s_j)(Z_j^\tau(s_j)Z_j(s_j))^{-1}Z_j^\tau(s_j)$, for $j = 1, \ldots, p$. Note that the residual $\boldsymbol{y}_s$ can be decomposed as $\boldsymbol{y}_s = (\tilde{\boldsymbol{y}}_{1s}^\tau, \cdots, \tilde{\boldsymbol{y}}_{js}^\tau, \cdots, \tilde{\boldsymbol{y}}_{ps}^\tau)^\tau$, where

$\tilde{\boldsymbol{y}}_{js} = \boldsymbol{y}_{js}/\sqrt{\hat{\sigma}^{jj}}$. Besides,

$$\begin{pmatrix} \boldsymbol{x}_{jk} & \boldsymbol{x}_{kj} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \boldsymbol{y}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{y}_j \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

Hence, the ℓ_2 norm of the projection of $\boldsymbol{y}_s$ into the space spanned by $(\boldsymbol{x}_{jk}, \boldsymbol{x}_{kj})$ can be expressed as

$$r_{jk} = \sqrt{\frac{\left(\boldsymbol{y}_k^\tau \boldsymbol{y}_{js}\right)^2}{\hat{\sigma}^{jj}\boldsymbol{y}_k^\tau \boldsymbol{y}_k} + \frac{\left(\boldsymbol{y}_j^\tau \boldsymbol{y}_{ks}\right)^2}{\hat{\sigma}^{kk}\boldsymbol{y}_j^\tau \boldsymbol{y}_j}}.$$

We now give the computation algorithm for the SSPS method.

The SSPS Algorithm

Initial step: Compute the estimated scale parameters $\hat{\sigma}^{jj}$ for $j = 1, \cdots, p$.

Set $s = \emptyset$ and $s_j = \emptyset$, $\boldsymbol{y}_{js} = \boldsymbol{y}_j$, for $j = 1, \ldots, p$.

Selecting step: For $j < k, (jk) \in s^c$, compute r_{jk}^2 and identify

$$s_{\text{TEMP}} = \left\{(jk), (kj) : r_{jk}^2 = \max_{(\tilde{j}\tilde{k}) \in s^c} r_{\tilde{j}\tilde{k}}^2\right\}.$$

Let $s_{\text{NEW}} = s \cup s_{\text{TEMP}}$. Compare $\text{EBIC}_\gamma(s_{\text{NEW}})$ with $\text{EBIC}_\gamma(s)$. If $\text{EBIC}_\gamma(s_{\text{NEW}}) < \text{EBIC}_\gamma(s)$, go to the updating step; otherwise, go to Output.

Updating step: Update s to $s = s_{\text{NEW}}$. For $j = 1, \ldots, p$, update s_j to

$s_j \cup \{k : (jk) \in s_{\text{TEMP}}\}$, update $\boldsymbol{y}_{js}$ to $\boldsymbol{y}_{js} = [\boldsymbol{I}_n - H(s_j)]\boldsymbol{y}_j$, and go to Selecting step.

Output: Estimate the adjacent matrix as $\hat{\Theta} = (\hat{\theta}_{jk})_{p\times p}$, where $\hat{\theta}_{jk} = 1$ if $(jk) \in s$.

The EBIC_γ in the above algorithm is given by

$$\text{EBIC}_\gamma(s) = n\sum_{j=1}^{p} \ln\left(\frac{\|\boldsymbol{y}_{js}\|_2^2}{n\hat{\sigma}^{jj}}\right) + |s|\ln n + 2\gamma \ln\binom{p(p-1)/2}{|s|/2}, \tag{2.14}$$

where γ is taken as $\gamma > 1 - \frac{\ln n}{2\ln p}$. Note that $|s|/2$ is the number of pairs of $(\boldsymbol{x}_{jk}, \boldsymbol{x}_{kj})$'s in model s and each pair associates with two parameters in Θ. The form of the above EBIC is in compliance with its original definition given in Chen and Chen (2008): $-2\ln$(maximum likelihood)$+\ln n\times$ the number of parameters in the model $+2\gamma\ln$(cardinality of the class that the model belongs to). Since we are interested in the high-dimensional sparse scenario where $p(p-1)/2$ is much larger than $|s|/2$, $\binom{p(p-1)/2}{|s|/2}$ can be approximated by $p^{|s|}$. Consequently, the EBIC in (2.14) is approximated as below, up to a constant,

$$\text{EBIC}_\gamma(s) \approx n\sum_{j=1}^{p} \ln\left(\frac{\|\boldsymbol{y}_{js}\|_2^2}{\hat{\sigma}^{jj}}\right) + |s|\ln n + 2\gamma|s|\ln p. \tag{2.15}$$

2.3 Selection Consistency of SSPS

In this section, we establish the selection consistency of the SSPS method. By the argument given in the previous section, without loss of generality, we assume that $\boldsymbol{Y} = (Y_1, \ldots, Y_p) = (\hat{f}_1(X_1), \ldots, \hat{f}_p(X_p))^\tau$ follows the mul-

tivariate normal distribution $N_p(\mathbf{0}, \Sigma)$. First, we introduce some notations. Denote the edge set by $E = \{(jk) : j \neq k\}$, thus $(jk) \in E$ if and only if there is an edge between vertices j and k. Let $s_{0j} = \{k : (jk) \in E\}$ and $p_{0j} = |s_{0j}|$, for $j = 1, \ldots, p$. Let s be a subset of E and $s_j = \{k : (jk) \in s\}$. Obviously, $s_j \subset s_{0j}$. Let $\Sigma_{s_j s_k}$ denote the matrix obtained from Σ by deleting its rows whose indices are not in s_j and the columns whose indices are not in s_k. If s_j consists of a single index i, the notation $\Sigma_{i s_k}$ is used for simplicity. Let $S_j = \{1, \ldots, p\} \setminus \{j\}$ and, for $j \neq k$, define

$$\tau_{jk} = \frac{(\Sigma_{kS_j} - \Sigma_{ks_j}\Sigma^{-1}_{s_j s_j}\Sigma_{s_j S_j})\boldsymbol{\beta}_j}{\sqrt{\sigma^{jj}\sigma_k^2}},$$

where $\boldsymbol{\beta}_j$ is the regression coefficients vector in (2.12), $\sigma^{jj} = \sigma_j^2 - \Sigma_{jS_j}\Sigma^{-1}_{S_j S_j}\Sigma_{S_j j}$, and σ_k^2 is the variance of Y_k. Let $v_{jk} = (\tau_{jk}, \tau_{kj})^\tau$.

We assume the following conditions:

C1 $p = \exp(\alpha n^\kappa)$, where $\alpha > 0$ and $\kappa \in (0, 1/3)$. $\max_j p_{0j} = O(n^{1/6-\delta})$, for some $\delta \in (0, 1/6 - \kappa/2)$.

C2 For any $s \subset E$ but $s \neq E$,

$$\max_{(jk)\in E\setminus s} \|v_{jk}\|_2^2 > \max_{(jk)\notin E} \|v_{jk}\|_2^2.$$

C3 $\lim_{n\to\infty} \min_{1\le j\le p} \left\{ \frac{\sqrt{n}}{\ln p} \lambda_{min}(\Sigma_{s_{0j}s_{0j}}) \min_{k\in s_{0j}} |\frac{\beta_{jk}}{\sqrt{\sigma^{jj}}}| \right\} = +\infty$, where $\lambda_{min}(A)$ denotes the smallest eigenvalue of any symmetric matrix A.

Theorem 2.1. Assume conditions C1 - C3. Further suppose that the correlations between the components in $\boldsymbol{Y}$ are bounded by a constant less than 1, the variances of the components of $\boldsymbol{Y}$ are also bounded and that

the γ in (2.15) is taken to be larger than $1 - \ln n/(2 \ln p)$. Then the SSPS method is selection consistent, i.e.,

$$P(\hat{\Theta}_n = \Theta) \to 1, \text{ as } n \to +\infty,$$

where $\hat{\Theta}_n$ is the SSPS estimate of the adjacent matrix Θ.

Theorem 2.1 can be explained as follows. Let $s_{*1}, s_{*2}, \ldots, s_{*m^*}$ be the series of index sets obtained at the selecting steps of the SSPS procedure until it stops, where m^* is the final step. The theorem implies that (i) for $m \leq m^*$, $P(s_{*m} \subset E) \to 1$ uniformly in m; (ii) for $m < m^*$, $P(\text{EBIC}(s_{*m}) > \text{EBIC}(s_{*m+1})) \to 1$ uniformly in m, and (iii) $P(s_{*m^*} = E, \min_{s:E\subset s} \text{EBIC}(s) > \text{EBIC}(s_{*m^*})) \to 1$. In other words, it indicates that, asymptotically with probability 1, only true edges can be selected at each step of the SSPS procedure, and that the procedure stops only when all the true edges have been selected. In order to establish Theorem 2.1, we first provide two lemmas as follows.

Lemma 2.1. Assume that (a) the correlation between Y_j and Y_k is bounded by a constant less than 1, (b) the variances of the components in $\boldsymbol{Y}$, i.e. the σ_j^2's, are bounded, and (c) p and p_{0j} are as given in C1. Then we have

(1) $P(\max_{j,k} |\frac{1}{n}\boldsymbol{y}_j^\tau \boldsymbol{y}_k - \Sigma_{jk}| > n^{-1/3}\sigma_{\max}) \to 0$, where $\sigma_{\max} = \max_{j,k} \sqrt{\text{Var}(Y_j Y_k)}$;

(2) Let $\Sigma_{jk|s} = \Sigma_{jk} - \Sigma_{js}\Sigma_{ss}^{-1}\Sigma_{sk}$ and $\hat{\Sigma}_{jk|s} = \frac{1}{n}\boldsymbol{y}_j^\tau [I - H(s)]\boldsymbol{y}_k$, where $s \subset \{1, \ldots, p\}$,

$$\max_{j,k,s:|s|\leq \max_i p_{0i}} |\hat{\Sigma}_{jk|s} - \Sigma_{jk|s}| = o_p(1)$$

Lemma 2.1 follows from the facts that, under the normality of $(Y_1, \ldots, Y_p)$, (1) $\sigma_{\max} < \infty$ and (2) $\max_{1\le j,k\le p} \boldsymbol{E}\exp(tY_jY_k)$ and $\max_{1\le j,k\le p} \boldsymbol{E}\exp(tY_j\epsilon_k)$ are finite for t in a neighborhood of 0. The proof of this lemma is analogous to that of Lemma 3.1 in Luo and Chen (2014b) and hence is omitted.

Define $\hat{\tau}_{jk} = \frac{\boldsymbol{y}_k^\tau[I-H(s_j)]Z_j\boldsymbol{\beta}_j}{n\sqrt{\sigma^{jj}\sigma_k^2}}$ and let $\hat{v}_{jk} = (\hat{\tau}_{jk}, \hat{\tau}_{kj})^\tau$. It follows from Lemma 2.1 that the conditions $C2$ and $C3$ in Theorem 2.1 can be replaced by $C2^{'}$ and $C3^{'}$ given below.

C$2^{'}$ For any $s \subset E$ but $s \neq E$ and some $0 < q < 1$,

$$q \max_{(jk)\in E\setminus s} \|\hat{v}_{jk}\|_2^2 > \max_{(jk)\notin E} \|\hat{v}_{jk}\|_2^2.$$

C$3^{'}$ $\lim_{n\to\infty} \min_{1\le j\le p} \left(\frac{\sqrt{n}}{\ln p}\lambda_{min}(\frac{1}{n}Z_j(s_{0j})^\tau Z_j(s_{0j})) \min_{k\in s_{0j}} |\frac{\beta_{jk}}{\sqrt{\sigma^{jj}}}|\right) = +\infty.$

In fact, we have

Lemma 2.2. Under the conditions of $C2$, $C3$, and those in Lemma 2.1, $C2^{'}$ and $C3^{'}$ hold almost surely.

Proof. From the definitions,

$$\tau_{jk} = \frac{1}{\sqrt{\sigma^{jj}}\sigma_k} \sum_{l\in s_j} \Sigma_{kl|s_j}\beta_{jl}, \qquad \hat{\tau}_{jk} = \frac{1}{\sqrt{\sigma^{jj}}\sigma_k} \sum_{l\in s_j} \hat{\Sigma}_{kl|s_j}\beta_{jl}.$$

It follows from (2) of Lemma 2.1 that, for any $1 \le k \neq j \le p$ and $s_j \subset s_{0j}$,

$$\hat{\tau}_{jk} = \tau_{jk}(1 + o_p(1)), \quad \|\hat{v}_{jk}\|_2^2 = \|v_{jk}\|_2^2(1 + o_p(1)).$$

Therefore, $C2^{'}$ holds almost surely when $C2$ is true.

For any vector $\boldsymbol{w}_j \in R^{p_{0j}}$ and $\|\boldsymbol{w}_j\|_2 = 1$,

$$\begin{aligned} & \max_j \left|\boldsymbol{w}_j^\tau \big(\frac{1}{n} Z_j(s_{0j})^\tau Z_j(s_{0j}) - \Sigma_{s_{0j}s_{0j}}\big)\boldsymbol{w}_j\right| \\ \leq & \max_j \Big(\sum_{1\leq k.l\leq p} |w_{jk}w_{jl}| \max_{1\leq k.l\leq p} |\frac{1}{n}\boldsymbol{y}_k\tau\boldsymbol{y}_l - \Sigma_{kl}|\Big) \\ \leq & (\max_j p_{0j})^2 \max_{1\leq k.l\leq p} |\frac{1}{n}\boldsymbol{y}_k\tau\boldsymbol{y}_l - \Sigma_{kl}|. \end{aligned}$$

Therefore, by (1) of Lemma 2.1 and $C1$, uniformly for $1 \leq j \leq p$, we have

$$\lambda_{min}(\frac{1}{n} Z_j(s_{0j})^\tau Z_j(s_{0j})) = \lambda_{min}(\Sigma_{s_{0j}s_{0j}})(1 + o_p(1))$$

That is, $C3$ implies $C3^{'}$ with probability 1. □

In the following proof of Theorem 2.1, conditions $C2$ and $C3$ are replaced by $C2^{'}$ and $C3^{'}$.

Proof of Theorem 2.1. Let $s_{*1} \subset s_{*2} \subset \cdots \subset s_{*m} \subset \cdots$ be the sequence of the edge sets selected in the SSPS procedure without any stopping rule. The results of Theorem 2.1 will be shown if we can establish the following two facts: (i) there is a positive integer m^* such that $P(s_{*m^*} = E) \to 1$ as n goes to ∞, and (ii) the sequence $\text{EBIC}_\gamma(s_{*m})$ is decreasing when $m < m^*$ and increasing when $m^* \leq m < kp_0$ for any fixed $k > 1$. Given that fact (i) is true, fact (ii) can be established similarly to Theorem 3.3 of Luo and Chen (2014b). Hence we only establish fact (i) in what follows.

Let $s_{*0} = \emptyset$. It suffices to show that, given $s_{*m} \subset E$, $P(s_{*m+1} \subset E) \to 1$ uniformly for all $m < m^*$. Denote $s_{j*m} = \{k : (jk) \in s_{*m}\}$ and $\hat{\tau}_{jk}^m = \frac{\boldsymbol{y}_k^\tau[I - H(s_{j*m})]Z_j\boldsymbol{\beta}_j}{n\sqrt{\sigma^{jj}\sigma_k^2}}$. Let $\tilde{\tau}_{jk}^m = \frac{1}{n\sigma_k}\boldsymbol{y}_k^\tau[I - H(s_{j*m})]\tilde{\boldsymbol{y}}_j$ and $\tilde{v}_{jk}^m = (\tilde{\tau}_{jk}^m, \tilde{\tau}_{kj}^m)^\tau$. We

need to show that the inequality

$$\max_{(jk)\in E\setminus s_{*m}} \|\tilde{v}_{jk}^m\|_2^2 > \max_{(jk)\notin E} \|\tilde{v}_{jk}^m\|_2^2 \tag{2.16}$$

holds with probability tending to 1 uniformly for $m < m^*$. The above inequality implies that $s_{*m+1} \subset E$. Note that

$$\begin{aligned} |\tilde{v}_{jk}^m\|_2^2 &= \|\hat{v}_{jk}^m\|_2^2 + (\frac{1}{n\sigma_k}\boldsymbol{y}_k^\tau[I - H(s_{j*m})]\frac{\boldsymbol{\epsilon}_j}{\sqrt{\sigma^{jj}}})^2 + (\frac{1}{n\sigma_j}\boldsymbol{y}_j^\tau[I - H(s_{k*m})]\frac{\boldsymbol{\epsilon}_k}{\sqrt{\sigma^{kk}}})^2 \\ &\quad +2\frac{1}{n\sigma_k}\boldsymbol{y}_k^\tau[I - H(s_{j*m})]\frac{\boldsymbol{\epsilon}_j}{\sqrt{\sigma^{jj}}}\hat{\tau}_{jk} + 2\frac{1}{n\sigma_k}\boldsymbol{y}_j^\tau[I - H(s_{k*m})]\frac{\boldsymbol{\epsilon}_k}{\sqrt{\sigma^{kk}}}\hat{\tau}_{kj}. \end{aligned}$$

Inequality (2.16) holds if we have

(a) $\frac{1}{n\sigma_k}\boldsymbol{y}_k^\tau[I - H(s_{j*m})]\frac{\boldsymbol{\epsilon}_j}{\sqrt{\sigma^{jj}}} = O_p(n^{-1/2}\ln p)$ uniformly for all $(jk) \in s_{*m}^c$ and all $m < m^*$;

(b) $\max_{(jk)\in E\setminus s_{*m}} \|\hat{v}_{jk}^m\|_2^2 > C_n(\ln p)^2/n$ where $C_n \to \infty$;

(c) $\max_{(jk)\in E\setminus s_{*m}} \|\hat{v}_{jk}^m\|_2^2 - \max_{(jk)\notin E} \|\hat{v}_{jk}^m\|_2^2$ has the same order as $\max_{(jk)\in E\setminus s_{*m}} \|\hat{v}_{jk}^m\|_2^2$.

These facts are established in the following.

Proof of (a): Let $\tau^2 = \sigma_k^2 - \Sigma_{ks_{j*m}}\Sigma_{s_{j*m}s_{j*m}}^{-1}\Sigma_{s_{j*m}k}$ and $\hat{\tau}^2 = \boldsymbol{y}_k^\tau[I - H(s_{j*m})]\boldsymbol{y}_k/n$. We have

$$\begin{aligned} & P(\frac{1}{n\sigma_k}|\boldsymbol{y}_k^\tau[I - H(s_{j*m})]\frac{\boldsymbol{\epsilon}_j}{\sqrt{\sigma^{jj}}}| > n^{-1/2}\ln p) \\ = \; & P(\frac{1}{\sqrt{n}}|\boldsymbol{y}_k^\tau[I - H(s_{j*m})]\frac{\boldsymbol{\epsilon}_j}{\sqrt{\sigma^{jj}}}| > \sigma_k \ln p) \\ \leq \; & P(\frac{1}{\sqrt{n}}|\boldsymbol{y}_k^\tau[I - H(s_{j*m})]\frac{\boldsymbol{\epsilon}_j}{\sqrt{\sigma^{jj}}}| > \tau \ln p) \\ = \; & P(\frac{\hat{\tau}}{\tau}|Z| > \ln p), \end{aligned}$$

where $Z = \frac{1}{\sqrt{n}\hat{\tau}} \boldsymbol{y}_k^\tau [I - H(s_{j*m})] \frac{\boldsymbol{\epsilon}_j}{\sqrt{\sigma^{jj}}}$. Conditioning on Y_{-j}, Z is standard normal. By Slutsky's theorem, $\frac{\hat{\tau}}{\tau} Z \to Z$ in distribution. Therefore,

$$\begin{aligned} & P\Big(\frac{1}{n\sigma_k} |\boldsymbol{y}_k^\tau [I - H(s_{j*m})] \frac{\boldsymbol{\epsilon}_j}{\sqrt{\sigma^{jj}}}| > n^{-1/2} \ln p\Big) \\ \leq \;\; & P\Big(\frac{\hat{\tau}}{\tau} |Z| > \ln p\Big) \quad \to \quad P(|Z| > \ln p) \leq \frac{2}{\ln p} \exp\{-\frac{(\ln p)^2}{2}\}. \end{aligned}$$

Since the bound does not depend on Y_{-j}, it holds unconditionally as well. By the Bonferroni inequality,

$$P\Big(\max_{m<m^*} \max_{(jk)\in s_{*m}^c} \frac{1}{n\sigma_k} |\boldsymbol{y}_k^\tau [I - H(s_{j*m})] \frac{\boldsymbol{\epsilon}_j}{\sqrt{\sigma^{jj}}}| > n^{-1/2} \ln p\Big) \leq \frac{2}{\ln p} \exp\{-\frac{(\ln p)^2}{2} + 3 \ln p\} \to 0.$$

Hence, (a) is verified.

Proof of (b) and (c): By an abuse of notaion, in the following, we denote $\boldsymbol{\beta}_j / \sqrt{\sigma^{jj}}$ still by $\boldsymbol{\beta}_j$ and denote $s_{0j} \setminus s_{j*m}$ by s_{j*m}^-. First, we show that, for any j,

$$\max_{k \in s_{j*m}^-} |\hat{\tau}_{jk}^m| \geq \lambda_{\min}\Big(\frac{1}{n}(Z_j(s_{0j})^\tau Z_j(s_{0j})\Big) \min_{k \in s_{0j}} |\beta_{jk}| / \sigma_k. \tag{2.17}$$

Let $\Delta(s_{j*m}) = \boldsymbol{\mu}_j^\tau [\boldsymbol{I}_n - H(s_{j*m})] \boldsymbol{\mu}_j$ where $\boldsymbol{\mu}_j = Z_j \boldsymbol{\beta}_j$. We have

$$\Delta(s_{j*m}) = \sum_{k \in s_{j*m}^-} \beta_{jk} \boldsymbol{y}_k^\tau [\boldsymbol{I}_n - H(s_{j*m})] \boldsymbol{\mu}_j \leq n \|\boldsymbol{\beta}_j(s_{j*m}^-)\|_1 \max_{k \in s_{j*m}^-} \sigma_k |\hat{\tau}_{jk}^m|, \tag{2.18}$$

and

$$\begin{aligned}\Delta(s_{j*m}) =& \boldsymbol{\beta}_j^\tau(s_{j*m}^-)Z_j^\tau(s_{j*m}^-)[\boldsymbol{I}_n - H(s_{j*m})]Z_j(s_{j*m}^-)\boldsymbol{\beta}_j(s_{j*m}^-) \\ \geq& \lambda_{\min}(Z_j^\tau(s_{j*m}^-)[\boldsymbol{I}_n - H(s_{j*m})]Z_j(s_{j*m}^-))\|\boldsymbol{\beta}_j(s_{j*m}^-)\|_2^2 \\ \geq& \lambda_{\min}(Z_j^\tau(s_{0j})Z_j(s_{0j}))\|\boldsymbol{\beta}_j(s_{j*m}^-)\|_2^2.\end{aligned} \tag{2.19}$$

The second inequality in (2.19) holds since $s_{j*m} \cup s_{j*m}^- = s_{0j}$ and $(Z_j^\tau(s_{j*m}^-)[\boldsymbol{I}_n - H(s_{j*m})]Z_j(s_{j*m}^-))^{-1}$ is a major sub-matrix of $(Z_j^\tau(s_{0j})Z_j(s_{0j}))^{-1}$. Thus (2.18) and (2.19) imply that

$$\begin{aligned}\max_{k\in s_{j*m}^-}|\hat{\tau}_{jk}^m| &\geq \sigma_k^{-1}\lambda_{\min}(\frac{1}{n}Z_j^\tau(s_{0j})Z_j(s_{0j}))\frac{\|\boldsymbol{\beta}_j(s_{j*m}^-)\|_2^2}{\|\boldsymbol{\beta}_j(s_{j*m}^-)\|_1} \\ &\geq \lambda_{\min}(\frac{1}{n}Z_j^\tau(s_{0j})Z_j(s_{0j}))\min_{k\in s_{0j}}|\beta_{jk}|/\sigma_k.\end{aligned}$$

This verifies (2.17). Since $\max_{(jk)\in E\setminus s_{*m}}\|\hat{v}_{jk}^m\|_2^2 \geq \max_{(jk)\in E\setminus s_{*m}}\|\hat{v}_{jk}^m\|_2^2 \geq \max_{k\in s_{j*m}^-}\|\hat{v}_{jk}^m\|_2^2$, for all j, (b) follows from condition $C3'$.

By condition $C2'$, we have

$$\max_{(jk)\in E\setminus s_{*m}}\|\hat{v}_{jk}^m\|_2^2 \geq \max_{(jk)\in E\setminus s_{*m}}\|\hat{v}_{jk}^m\|_2^2 - \max_{(jk)\notin E}\|\hat{v}_{jk}^m\|_2^2 > (1-q)\max_{(jk)\in E\setminus s_{*m}}\|\hat{v}_{jk}^m\|_2^2.$$

Hence, (c) is verified.

□

2.4 Simulation Study

In this section, we report the simulation study that compares SSPS with the following methods: Glasso [Friedman et al. (2008)], Clime [Cai

et al. (2011)], graphical scad (G-scad) [Fan et al. (2009)], Space [Peng et al. (2009)], SR-SLasso and JR-SLasso [Luo and Chen (2014a)]. The two-step procedure in Liu et al. (2009) is followed in the implementation of all the methods. That is, the transformation of the nonparanormal data is made first, then these methods are applied to the transformed data. Corresponding R packages, glasso [Friedman et al. (2014)], clime [Cai et al. (2012)], and space [Peng et al. (2010)], are employed in the computation accordingly. Besides, the Glasso algorithm combined with the local linear approximation (LLA) [Zou and Li (2008)] is used in the G-scad procedure, where set $a = 3.7$ [Fan and Li (2001)]. For the comparison to be fair, the EBIC is used in the tuning parameter selection of all the non-sequential methods, among which Glasso, Clime, G-scad use the cross validation in the original work, while Space adopts a BIC-type criterion.

2.4.1 Simulation Settings and Measures

The nine Gaussian graphs in Luo and Chen (2014a) are considered. They are AR(1), AR(2), Circle, Cluster, Erdös-Rényi (ER), Tridiagonal (Tridiag), Hub, Random Proximity (RP) and Barabási-Albert (BA), where the first six are non-random graphs and the other three are random graphs. For completeness, we briefly describe them as follows.

AR(1): $\omega_{ii} = 1$, $\omega_{i,i+1} = \omega_{i+1,i} = 0.45$, for $i, i+1 \in \{1, \ldots, p\}$.

AR(2): $\omega_{ii} = 1$, $\omega_{i,i+1} = \omega_{i+1,i} = 0.3$, $\omega_{i,i+2} = \omega_{i+2,i} = 0.4$, for $i, i+1, i+2 \in \{1, \ldots, p\}$.

Circle: $\omega_{1p} = \omega_{p1} = 0.1$ and $\omega_{ii} = 1$, $\omega_{i,i+1} = \omega_{i+1,i} = 0.45$, for $i, i+1 \in$

$\{1, \ldots, p\}$.

Cluster: Ω is a block diagonal matrix, where each main diagonal block is a 5×5 matrix with its (i, j)th entry to be $\mathbf{I}(i = j) + 0.5\mathbf{I}(i \neq j)$, $i, j = 1, \ldots, 5$.

ER: Let $B = (b_{ij})_{p \times p}$ be a symmetric matrix with b_{ij}'s i.i.d. Bernoulli(0.05). Then $\Omega = 0.5B + \delta \boldsymbol{I}_p$, where δ is chosen such that Ω is positive definite and has conditional number p.

Tridiag: Let $U_1, \ldots, U_{p-1}$ be i.i.d. Uniform(0.5,1). For $i < j$, the (i, j)th entry of Σ is $\sigma_{ij} = \tilde{\sigma}_{ij}\mathbf{I}(\tilde{\sigma}_{ij} \geq 10^{-7})$, where $\tilde{\sigma}_{ij} = \exp(-0.5 \sum_{k=i}^{j-1} U_k)$. The diagonal elements of Σ are all 1. Then $\Omega = \Sigma^{-1}$ with its entries truncated at 10^{-7}.

Hub: Θ is a block diagonal matrix with 5 identical diagonal blocks Θ_0, where only the entries in the first row and column of Θ_0 are 1.

RP: Let $U_1, \ldots, U_p$ be i.i.d. uniform random points over the unit square $[0, 1]^2$. Define the neighborhood set of U_i as n_i, such that $\|U_i - U_j\|_2$ is among the five smallest of $\{\|U_i - U_k\|_2 : k \neq i\}$ if $j \in n_i$. Hence, $\theta_{ij} = 1$ if and only if either $i \in n_j$ or $j \in n_i$.

BA: The graph is generated using the Barabási-Albert algorithm. The initial graph has two connected vertices and a new vertex is connected to only one vertex in the existing graph with the probability proportional to the current degree of that vertex.

The precision matrix of a non-random graph is generated directly. However, in the random graphs, the adjacent matrices are generated first accordingly,

and the precision matrices are obtained based on the adjacent matrices by the following procedure. For $i < j$, the lower triangle entries ω_{ij}'s are i.i.d. Uniform$[-1, -0.5] \cup [0.5, 1]$; then in the upper triangle, $\omega_{ji} = \omega_{ij}$. The ith diagonal entry is $\omega_{ii} = \delta \sum_{j \neq i} |\omega_{ji}|$, where δ is set to be 1.2, 1 and 1.01 for the graphs Hub, RP, and BA respectively. The nine graphs are shown in Figure 2.3 with 50 vertices.

The Gaussian CDF transformation, symmetric power transformation and linear transformation considered in Liu et al. (2012) are used for generating the nonparanormal data from the Gaussian data. The two transformation functions are given as follows.

Gaussian CDF Transformation: Assume g_0 to be the c.d.f. of the univariate Gaussian distribution with mean μ_0 and standard deviation σ_0, i.e. $g_0(t) = \Phi(\frac{t-\mu_{g0}}{\sigma_{g0}})$. The inverse transformation g_j is given by

$$g_j(y_j) = \frac{g_0(y_j) - \int g_0(t)\phi((t-\mu_j)/\sigma_j)dt}{\sqrt{\int [g_0(x) - \int g_0(t)\phi((t-\mu_j)/\sigma_j)dt]^2 \phi((x-\mu_j)/\sigma_j)dx}},$$

where $\mu_j = EY_j$, $\sigma_j^2 = \mathrm{Var}(Y_j)$, and $\phi(\cdot)$ is the p.d.f. of the standard normal distribution.

Symmetric Power Transformation: Let $g_0(t) = \mathrm{sign}(t)|t|^\alpha$ for a given $\alpha > 0$. Then,

$$g_j(y_j) = \frac{g_0(y_j - \mu_j)}{\sqrt{\int g_0^2(t-\mu_j)\phi((t-\mu_j)/\sigma_j)dt}}.$$

Linear Transformation: $g_j(y_j) = a + by_j$.

The nonparanormal data are simulated as follows. First, a multivariate normal vector $\boldsymbol{Y} = (Y_1, \ldots, Y_p)^\tau$ with its covariance matrix determined by a graph setting is generated. Then, the Y_j's are transformed to the X_j's by the inverse transformation $X_j = g_j(Y_j)$. To comply with Liu et al. (2012)'s simulation design, we also set $\mu_{g_0} = 0.05$ and $\sigma_{g_0} = 0.4$ for the Gaussian CDF transformation and $\alpha = 3$ for the power transformation. In the linear transformation, set $a = 0$ and $b = 1$, which amounts to an identity transformation. By using this identity inverse transformation, the $\boldsymbol{Y}$-data are in fact unchanged. For illustration, we apply these three inverse transformations on the standard normal random variable $N(0, 1)$, respectively. Figure 2.1 shows the density functions of the three inverse transformed variables. Figure 2.2 visualizes the relation between the $\boldsymbol{Y}$-data (Gaussian) and the $\boldsymbol{X}$-data (nonparanormal) under different transformations in 1-dimension. As can be seen, the two nontrivial transformations lead to the nonparanormal data quite different the Gaussian ones. Their direct impacts on the edge detection will be reflected on the estimation of f.

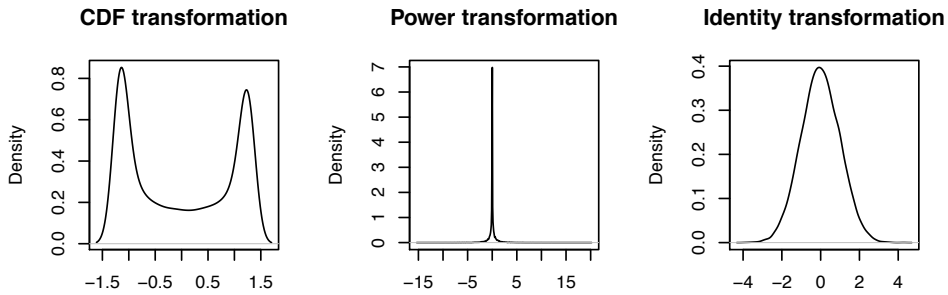

Figure 2.1: Density plots of the random variables transformed from $N(0, 1)$ by the three inverse transformations

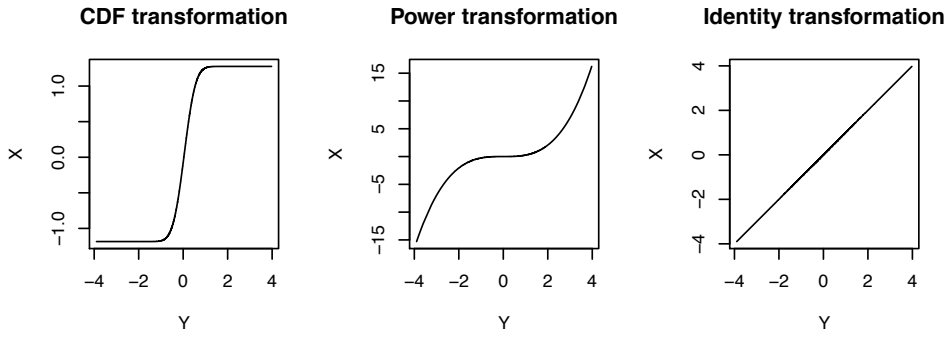

Figure 2.2: Three nonparanormal inverse transformations on the standard normal distributed data (Y) in 1-dimension

The comparison is concentrated on the accuracy of edge detection. The performance of a method is measured by its positive detection rate (PDR) and the false detection rate (FDR) defined below:

$$\text{PDR} = \frac{|\{(i,j) : \theta_{ij} = \hat{\theta}_{ij} = 1, i \neq j\}|}{|\{(i,j) : \theta_{ij} = 1, i \neq j\}|},$$

$$\text{FDR} = \frac{|\{(i,j) : \theta_{ij} = 0, \hat{\theta}_{ij} = 1, i \neq j\}|}{|\{(i,j) : \hat{\theta}_{ij} = 1, i \neq j\}|},$$

where θ_{ij}'s and $\hat{\theta}_{ij}$'s are the entries of the true adjacent matrix Θ and the estimated adjacent matrix $\hat{\Theta}$ respectively. In other words, the PDR is the proportion of the true edges which have been detected and the FDR is the proportion of the falsely detected edges among all the detected ones. Additionally, the number of estimated edges is also examined, which is $|\hat{\Theta}|/2$. For simplicity, in the following results, we denote it by $|\hat{E}|$.

We considered two scenarios: $p < n$ and $p > n$. For $p < n$, the nonparanormal data are generated as described with dimension $p = 50$ and sample

size $n = 100, 200$. In the scenario of $p > n$, we set $p = 200$ and $n = 100$. The corresponding nonparanormal data are generated as follow. The first 50 variables are simulated the same as in the $p < n$ case; with the same inverse transformation, the rest 150 variables are derived from i.i.d. $N(0, 1)$, also independent from the previous 50 variables. We report the average PDR, FDR and $|\hat{E}|$ of the various methods under different graph settings over 100 simulations in Table 2.2 to Table 2.4, where the corresponding standard deviations are in the parentheses.

2.4.2 Results

Before the comparison, we first evaluate some properties of the nine Gaussian graphs in the simulation settings. Define

$$R_j^2 = \frac{\boldsymbol{\beta}_j^T \Sigma_{-j,-j} \boldsymbol{\beta}_j}{\boldsymbol{\beta}_j^T \Sigma_{-j,-j} \boldsymbol{\beta}_j + \sigma^{jj}} = \frac{\Sigma_{j,-j} \Sigma_{-j,-j}^{-1} \Sigma_{-j,j}}{\Sigma_{j,-j} \Sigma_{-j,-j}^{-1} \Sigma_{-j,j} + \omega_{jj}^{-1}} = 1 - \frac{1}{\omega_{jj} \sigma_j^2},$$

which is the R^2 of the jth regression in (2.12). Denote the median of $R_1^2, \ldots, R_p^2$ by MED(R^2). Let MED(DG) be the median of all the vertices' degrees, where the degree of a vertex is the number of edges connected to it. $|E|$ means the total number of edges in the graph. The values of these three measures for each graph is given in the Table 2.1 with $p = 50$. As can be seen, the graphs AR(1), AR(2), Circle, Tridiag and BA have comparatively higher MED(R^2) values; on the other hand, those of Hub and RP are only around 0.2. Low MED(R^2) indicates that the regressions in (2.12) are dominated by the random errors. Hence, it might be more difficult to solve the linear regressions for the underlying edge sets in these two graphs with limited samples. Inferred from the values of MED(DG) and $|E|$, Hub

is sparsest; on the contrary, RP has the densest structure. The estimation under such graph may be compromised by its complexity. Additionally, it is noticeable that, in the graph Cluster, there is no sparsity within each component cluster, where all vertices are connected with each other. The graph Hub also consists of several small clusters. There is also no sparsity for the hub vertex in each cluster, since it is connected to all the other vertices in the same cluster.

Table 2.1: A summary of the simulation graphs' MED(R^2), MED(DG) and $|E|$ when $p = 50$

Graph	MED(R^2)	MED(DG)	$\|E\|$
AR(1)	0.564	2	49
AR(2)	0.505	4	97
Circle	0.564	2	50
Cluster	0.400	4	100
ER	0.353	2	60
Tridiag	0.647	2	73
Hub	0.207	1	45
RP	0.193	6	153
BA	0.506	1	49

The findings of the simulation study are discussed in the following. First, it is interesting to notice that the estimated transformation $\hat{f}$ has a quite diverse impact on the analysis of the nonparanormal data with different methods. The difference in the results between the inverse transformed data and the $\boldsymbol{Y}$-data reflects the impact of $\hat{f}$. Comparing the results of the same method across the three inverse transformations, those of the neighborhood selection methods, i.e. SSPS, SR-SLasso, JR-SLasso, and Space, remain almost unchanged. In contrast, the differences in the results obtained by Glasso, Clime and G-scad between the non-identity and the identity inverse transformations are quite noticeable, especially, between the symmetric power transformation and the identity transforma-

tion. For instance, under the Hub graph setting in the scenario of $p = 50$ and $n = 100$, with the identity transformation, the PDRs of Glasso, Clime and G-scad are all more than 0.8 and their FDRs range from 0.086 to 0.09; however, their PDRs drop to around 0.4 and FDRs increase to about the same level with the symmetric power transformation. The above findings indicate that the four considered neighborhood selection methods seem to be insensitive to the variation between the exact f and its estimate $\hat{f}$ caused by the different inverse transformations; on the contrary, Glasso, Clime and G-scad could be greatly affected by this variation. This difference may be attributed to the fact that, at the second step of the nonparanormal procedure, Glasso, Clime and G-scad use the Gaussian profile likelihood function of the precision matrix while the transformed data might not be so close to normality; whereas, these neighborhood selection methods only rely on the data's correlation relationships, thus they are more robust when the estimated $\hat{f}$ transformed data are deviated from the exact f transformed ones. Therefore, in the application of the two-step procedure with Glasso, Clime and G-scad, a highly accurate estimate of the transformation f is probably required.

We now consider the comparison of the various methods in terms of their PDR, FDR and the number of detected edges. In what follows, we omit the discussion related to the graph RP, since none of the methods' performance is satisfactory. This may caused by its less predictable and more complicated structure as mentioned before.

In the scenario of $p = 50$ and $n = 100$, apparently, SSPS outperforms Glasso, Clime and G-scad with higher PDRs and lower FDRs in all the sim-

ulated graphs. Moreover, among the three sequential methods, SR-SLasso has higher PDRs than SSPS in some graphs, such as Hub, but it also produces obviously higher FDRs. Meanwhile, JR-SLasso generates slightly smaller FDRs than SSPS's, but the corresponding PDRs are also declined. Considering the trade-off between PDR and FDR, SSPS is in general better than SR-SLasso and JR-SLasso. Furthermore, the performance of the Space is close to that of SSPS in AR(1), Circle, Tridiag, Hub and BA. This observation can be explained by the similarity of these two methods: by scaling (SSPS) or weighting (Space), effectively, different regularization levels are imposed in the corresponding regression models. However, SSPS has an advantage over Space in AR(2), Cluster and ER. Particularly, Space has difficulty in recovering the Cluster graph. As n increases to 200, the comparison is parallel to that of $n = 100$, whereas every method's estimation improves.

Now turn to the scenario of $p > n$. In this case, the advantage of the neighborhood selection methods over the other three methods in terms of FDR becomes more obvious. Essentially, Glasso, Clime and G-scad fail the task of edge selection. The proportion of average edges selected by them across the three inverse transformations and nine graphs, ranges from a minimum 17% to a maximum 46%, see the table below:

A summary of the proportion of selected edges by Glasso, Clime and G-scad

Min.	1st Quartile	Median	Mean	3rd Quartile	Max.
0.168	0.433	0.440	0.436	0.447	0.462

Their FDRs, no matter in which graph, are approximately 1, implying that they mistakenly discover many nonexistent edges. Hence, their high PDRs

are meaningless due to the extremely high FDRs. Besides, the FDRs of the SR-SLasso also inflated substantially to around 0.5 to 0.6. In marked contrast to the performance of the above four methods, the other two sequential methods and Space can still control the FDRs at reasonably low levels. Unfortunately, such low FDRs sometimes accompany with reduced PDRs. Among the sequential methods, again, SPSS is the best taking into account both PDR and FDR, except for the Hub graph, where JR-SLasso yields the best estimation. In the Cluster graph, as the dimension increases, none of the methods can effectively detect the edges, where the sequential methods and Space have very low PDRs while the others have extremely high FDRs. Except for this graph, when comparing SSPS and Space, it is indecisive which one offers the better results, although SSPS is slightly better in ER while Space has an advantage in Hub and BA.

In summary, the simulation studies in both scenarios demonstrate that, from the perspective of edge detection, SSPS is the winner among all the methods considered here and the sequential approaches have an edge over the non-sequential ones. It is worth noting that all the current methods have difficulties in edge detection of certain graphs, which calls for further research to be done in this field.

2.5 Conclusion

In this chapter, we consider the edge detection in the sparse high-dimensional nonparanormal graphical models. In fact, the proposed SSPS method is essentially derived under the Gaussian graphical models. Mo-

tivated by its predecessor JR-SLasso, SSPS resolves the problem of the heterogeneous variances when integrating different regression models into one. Moreover, inspired by the idea of grouped LASSO, it considers the pair of coefficients β_{jk} and β_{kj} jointly in the sequential procedure in respond to the intrinsic symmetry of the precision matrix. Specifically, at each step, the current residual vector is projected into the space spanned by each pair of the design matrix' columns, and the selected model is then augmented by the pair with the largest ℓ_2-norm projection. This sequential procedure's selection consistency has been established accordingly.

Through our extensive simulation studies, SSPS shows its advantage over the other competitors. Moreover, its estimation is quite stable across different inverse transformations. Hence, SSPS can analyze the nonparanormal data almost the same as the Gaussian data. Free from tuning parameter selection, SSPS and the other two sequential methods are more computationally efficient. Besides, it is very convenient to adapt SSPS for further computation reduction when dealing with ultra-high-dimensional data. Under such circumstances, we suggest a screening step appended to the updating step of the iterative algorithm. In the screening step, index pairs with current ℓ_2-norm projections smaller than a threshold are identified, such that the relevant calculation is skipped thereafter. For instance, if r_{jk}^2 is less than the threshold value, then if the iteration continues, r_{jk}^2 will not be calculated henceforward. This is equivalent to estimate β_{jk} and β_{kj} as 0 and remove them from the current model. However, it was found that SSPS is less effective if the underlying graph has a relatively low overall sparsity level even when $p < n$, e.g. the RP graph. Its ability to detect the true edges will also be compromised if the graph is not sparse within

some clusters, e.g. the Cluster graph, especially when $p > n$. This weakness may result from the violation of the sparsity condition, which is required by SSPS to be selection consistent. Similar problems can also be observed in the results of other methods.

Rather than edge detection, most existing methods focus on the precision matrix estimation. This topic will be covered in the next chapter, where the SSPS method is applied to estimate the precision matrix.

Table 2.2: Average (SD) of PDR, FDR and $|\hat{E}|$ of the various methods in scenario $p < n$ ($p = 50, n = 100$)

		Transformation								
		Gaussian CDF			Power			Identity		
Graph	Method	PDR	FDR	$\lvert\hat{E}\rvert$	PDR	FDR	$\lvert\hat{E}\rvert$	PDR	FDR	$\lvert\hat{E}\rvert$
AR(1)	Glasso	0.961(0.028)	0.122(0.050)	54(4)	0.888(0.050)	0.349(0.066)	68(8)	0.973(0.023)	0.054(0.036)	50(2)
	Clime	0.981(0.021)	0.065(0.042)	52(3)	0.878(0.050)	0.308(0.081)	63(9)	0.996(0.009)	0.033(0.035)	51(2)
	G-scad	0.964(0.027)	0.109(0.048)	53(4)	0.903(0.049)	0.335(0.069)	67(9)	0.981(0.020)	0.052(0.043)	51(3)
	Space	0.992(0.013)	0.018(0.017)	49(1)	0.992(0.013)	0.018(0.017)	50(1)	0.992(0.013)	0.018(0.017)	50(1)
	SR-SLasso	0.972(0.012)	0.218(0.054)	61(4)	0.973(0.012)	0.222(0.056)	62(4)	0.973(0.012)	0.222(0.056)	62(4)
	JR-SLasso	0.988(0.018)	0.027(0.025)	50(1)	0.988(0.017)	0.028(0.026)	50(1)	0.988(0.017)	0.028(0.026)	50(1)
	SSPS	0.994(0.011)	0.032(0.024)	50(1)	0.995(0.010)	0.031(0.022)	50(1)	0.995(0.010)	0.031(0.022)	50(1)
AR(2)	Glasso	0.456(0.030)	0.171(0.059)	54(6)	0.375(0.061)	0.432(0.064)	65(15)	0.480(0.024)	0.091(0.042)	51(4)
	Clime	0.484(0.034)	0.164(0.056)	56(6)	0.392(0.054)	0.422(0.070)	67(14)	0.524(0.047)	0.116(0.061)	58(9)
	G-scad	0.459(0.030)	0.172(0.056)	54(6)	0.387(0.063)	0.429(0.068)	67(16)	0.484(0.024)	0.095(0.041)	52(4)
	Space	0.443(0.021)	0.030(0.024)	44(2)	0.444(0.021)	0.030(0.024)	44(2)	0.444(0.021)	0.030(0.024)	44(2)
	SR-SLasso	0.767(0.050)	0.191(0.044)	92(7)	0.772(0.048)	0.191(0.041)	93(6)	0.772(0.048)	0.191(0.041)	93(6)
	JR-SLasso	0.547(0.052)	0.031(0.025)	55(6)	0.547(0.051)	0.031(0.025)	55(5)	0.547(0.051)	0.031(0.025)	55(5)
	SSPS	0.596(0.054)	0.059(0.031)	61(6)	0.600(0.058)	0.059(0.031)	62(6)	0.600(0.058)	0.059(0.031)	62(6)
Circle	Glasso	0.943(0.027)	0.111(0.051)	53(4)	0.876(0.049)	0.376(0.057)	71(8)	0.959(0.022)	0.058(0.038)	51(3)
	Clime	0.964(0.018)	0.063(0.037)	52(2)	0.861(0.051)	0.334(0.068)	65(8)	0.977(0.009)	0.036(0.037)	51(2)
	G-scad	0.948(0.026)	0.100(0.051)	53(4)	0.882(0.050)	0.359(0.054)	69(8)	0.965(0.019)	0.055(0.044)	51(3)
	Space	0.971(0.012)	0.017(0.019)	49(1)	0.972(0.012)	0.018(0.018)	50(1)	0.972(0.012)	0.018(0.018)	50(1)
	SR-SLasso	0.954(0.011)	0.220(0.054)	62(4)	0.954(0.011)	0.219(0.053)	62(4)	0.954(0.011)	0.219(0.053)	62(4)
	JR-SLasso	0.968(0.019)	0.029(0.025)	50(1)	0.969(0.018)	0.030(0.026)	50(1)	0.969(0.018)	0.030(0.026)	50(1)
	SSPS	0.976(0.009)	0.034(0.027)	51(1)	0.976(0.009)	0.032(0.026)	50(1)	0.976(0.009)	0.032(0.026)	50(1)
Cluster	Glasso	0.139(0.050)	0.287(0.108)	20(8)	0.011(0.011)	0.481(0.421)	2(2)	0.297(0.078)	0.264(0.070)	41(13)
	Clime	0.147(0.065)	0.279(0.111)	21(11)	0.011(0.012)	0.498(0.416)	2(2)	0.605(0.227)	0.359(0.077)	99(43)
	G-scad	0.139(0.050)	0.287(0.108)	20(8)	0.009(0.012)	0.240(0.377)	1(2)	0.296(0.082)	0.260(0.079)	41(13)
	Space	0.043(0.021)	0.081(0.156)	5(2)	0.044(0.021)	0.080(0.155)	5(2)	0.044(0.021)	0.080(0.155)	5(2)
	SR-SLasso	0.841(0.078)	0.184(0.038)	103(9)	0.842(0.081)	0.185(0.039)	104(10)	0.842(0.081)	0.185(0.039)	104(10)
	JR-SLasso	0.512(0.163)	0.058(0.038)	54(16)	0.511(0.165)	0.059(0.038)	54(16)	0.511(0.165)	0.059(0.038)	54(16)
	SSPS	0.581(0.176)	0.060(0.039)	61(18)	0.580(0.182)	0.061(0.039)	61(18)	0.580(0.182)	0.061(0.039)	61(18)
ER	Glasso	0.597(0.131)	0.354(0.074)	56(12)	0.438(0.135)	0.468(0.064)	50(17)	0.621(0.128)	0.305(0.076)	54(11)
	Clime	0.691(0.109)	0.230(0.061)	54(10)	0.474(0.116)	0.437(0.074)	51(14)	0.782(0.101)	0.139(0.049)	55(9)
	G-scad	0.642(0.152)	0.333(0.072)	58(14)	0.479(0.142)	0.468(0.064)	55(18)	0.767(0.174)	0.283(0.078)	65(18)
	Space	0.664(0.093)	0.123(0.056)	45(6)	0.664(0.096)	0.120(0.061)	45(6)	0.664(0.096)	0.120(0.061)	45(6)
	SR-SLasso	0.782(0.084)	0.445(0.046)	85(9)	0.785(0.084)	0.443(0.045)	85(9)	0.785(0.084)	0.443(0.045)	85(9)
	JR-SLasso	0.689(0.105)	0.052(0.042)	44(6)	0.693(0.104)	0.053(0.042)	44(6)	0.693(0.104)	0.053(0.042)	44(6)
	SSPS	0.741(0.121)	0.068(0.047)	48(7)	0.738(0.122)	0.058(0.041)	47(7)	0.738(0.122)	0.058(0.041)	47(7)

Continued on next page

Table 2.2: Average (SD) of PDR, FDR and $|\hat{\mathbf{E}}|$ of the various methods in scenario $p < n$ $(p = 50, n = 100)$

		Transformation								
		Gaussian CDF			Power			Identity		
Graph	Method	PDR	FDR	$\lvert\hat{E}\rvert$	PDR	FDR	$\lvert\hat{E}\rvert$	PDR	FDR	$\lvert\hat{E}\rvert$
Tridiag	Glasso	0.638(0.043)	0.152(0.055)	55(5)	0.614(0.053)	0.354(0.075)	70(11)	0.649(0.045)	0.110(0.044)	53(3)
	Clime	0.656(0.045)	0.065(0.048)	51(3)	0.607(0.052)	0.294(0.076)	64(9)	0.666(0.047)	0.030(0.031)	50(2)
	G-scad	0.646(0.047)	0.144(0.060)	55(5)	0.627(0.050)	0.339(0.067)	70(8)	0.668(0.047)	0.111(0.047)	55(3)
	Space	0.668(0.045)	0.029(0.022)	50(1)	0.668(0.045)	0.029(0.022)	50(1)	0.668(0.045)	0.029(0.022)	50(1)
	SR-SLasso	0.655(0.045)	0.220(0.059)	62(4)	0.655(0.045)	0.220(0.059)	62(4)	0.655(0.045)	0.220(0.059)	62(4)
	JR-SLasso	0.660(0.046)	0.031(0.032)	50(2)	0.660(0.046)	0.031(0.032)	50(2)	0.660(0.046)	0.031(0.032)	50(2)
	SSPS	0.667(0.045)	0.022(0.022)	50(1)	0.667(0.045)	0.022(0.022)	50(1)	0.667(0.045)	0.022(0.022)	50(1)
Hub	Glasso	0.718(0.084)	0.181(0.064)	40(6)	0.435(0.112)	0.450(0.089)	37(11)	0.803(0.069)	0.087(0.050)	40(4)
	Clime	0.717(0.077)	0.188(0.072)	40(6)	0.419(0.095)	0.441(0.093)	35(10)	0.848(0.058)	0.090(0.051)	42(4)
	G-scad	0.724(0.079)	0.179(0.057)	40(6)	0.437(0.115)	0.444(0.087)	36(12)	0.819(0.068)	0.086(0.043)	40(4)
	Space	0.757(0.068)	0.040(0.030)	36(3)	0.758(0.068)	0.040(0.030)	36(3)	0.758(0.068)	0.040(0.030)	36(3)
	SR-SLasso	0.923(0.032)	0.255(0.049)	56(4)	0.923(0.032)	0.255(0.049)	56(4)	0.923(0.032)	0.255(0.049)	56(4)
	JR-SLasso	0.828(0.070)	0.044(0.035)	39(3)	0.830(0.070)	0.044(0.035)	39(3)	0.830(0.070)	0.044(0.035)	39(3)
	SSPS	0.801(0.071)	0.086(0.051)	40(4)	0.796(0.072)	0.085(0.049)	39(4)	0.796(0.072)	0.085(0.049)	39(4)
RP	Glasso	0.121(0.042)	0.158(0.091)	22(7)	0.027(0.028)	0.357(0.310)	6(7)	0.160(0.042)	0.090(0.067)	27(7)
	Clime	0.122(0.037)	0.175(0.093)	23(7)	0.028(0.026)	0.356(0.286)	7(6)	0.168(0.040)	0.113(0.070)	29(7)
	G-scad	0.121(0.042)	0.160(0.089)	22(7)	0.027(0.029)	0.271(0.280)	6(7)	0.161(0.041)	0.094(0.073)	27(7)
	Space	0.078(0.031)	0.040(0.054)	12(5)	0.078(0.031)	0.040(0.054)	12(5)	0.078(0.031)	0.040(0.054)	12(5)
	SR-SLasso	0.200(0.036)	0.482(0.053)	59(8)	0.200(0.036)	0.482(0.053)	59(8)	0.200(0.036)	0.482(0.053)	59(8)
	JR-SLasso	0.141(0.037)	0.092(0.062)	24(6)	0.141(0.037)	0.092(0.062)	24(6)	0.141(0.037)	0.092(0.062)	24(6)
	SSPS	0.164(0.031)	0.168(0.066)	30(5)	0.164(0.031)	0.168(0.066)	30(5)	0.164(0.031)	0.168(0.066)	30(5)
BA	Glasso	0.899(0.045)	0.274(0.076)	61(8)	0.822(0.064)	0.501(0.091)	84(19)	0.920(0.043)	0.211(0.081)	58(7)
	Clime	0.870(0.050)	0.227(0.084)	56(7)	0.791(0.062)	0.438(0.088)	71(12)	0.912(0.045)	0.159(0.076)	54(6)
	G-scad	0.907(0.044)	0.259(0.073)	61(7)	0.822(0.065)	0.490(0.093)	82(17)	0.939(0.033)	0.175(0.081)	56(6)
	Space	0.959(0.024)	0.184(0.058)	58(5)	0.956(0.025)	0.171(0.049)	57(4)	0.955(0.027)	0.172(0.058)	57(5)
	SR-SLasso	0.828(0.048)	0.466(0.058)	77(5)	0.828(0.054)	0.459(0.063)	76(6)	0.843(0.052)	0.445(0.062)	75(5)
	JR-SLasso	0.829(0.034)	0.064(0.040)	43(2)	0.830(0.039)	0.056(0.040)	43(2)	0.830(0.038)	0.052(0.039)	43(2)
	SSPS	0.891(0.056)	0.112(0.056)	49(3)	0.883(0.059)	0.113(0.055)	49(3)	0.878(0.059)	0.120(0.053)	49(3)

Table 2.3: Average (SD) of PDR, FDR and $|\hat{E}|$ of the various methods in scenario $p < n$ ($p = 50, n = 200$)

		Transformation								
		Gaussian CDF			Power			Identity		
Graph	Method	PDR	FDR	$\lvert\hat{E}\rvert$	PDR	FDR	$\lvert\hat{E}\rvert$	PDR	FDR	$\lvert\hat{E}\rvert$
AR(1)	Glasso	0.994(0.010)	0.077(0.045)	53(3)	0.984(0.018)	0.266(0.071)	66(7)	0.999(0.005)	0.021(0.023)	50(1)
	Clime	0.999(0.003)	0.024(0.027)	50(1)	0.976(0.020)	0.203(0.072)	61(6)	1.000(0.000)	0.018(0.025)	50(1)
	G-scad	0.996(0.008)	0.070(0.043)	53(3)	0.985(0.017)	0.248(0.071)	65(7)	1.000(0.000)	0.016(0.022)	50(1)
	Space	1.000(0.000)	0.017(0.020)	50(1)	1.000(0.000)	0.015(0.019)	50(1)	1.000(0.000)	0.015(0.019)	50(1)
	SR-SLasso	0.980(0.000)	0.188(0.047)	59(3)	0.980(0.000)	0.191(0.046)	60(3)	0.980(0.000)	0.191(0.046)	60(3)
	JR-SLasso	1.000(0.000)	0.015(0.020)	50(1)	1.000(0.000)	0.015(0.021)	50(1)	1.000(0.000)	0.015(0.021)	50(1)
	SSPS	1.000(0.000)	0.029(0.022)	50(1)	1.000(0.000)	0.032(0.025)	51(1)	1.000(0.000)	0.032(0.025)	51(1)
AR(2)	Glasso	0.518(0.025)	0.157(0.077)	60(8)	0.519(0.036)	0.375(0.073)	82(13)	0.525(0.032)	0.148(0.091)	61(11)
	Clime	0.623(0.077)	0.224(0.091)	80(18)	0.539(0.051)	0.354(0.073)	83(16)	0.871(0.052)	0.270(0.058)	117(14)
	G-scad	0.520(0.024)	0.163(0.081)	61(9)	0.521(0.038)	0.373(0.078)	82(14)	0.542(0.063)	0.173(0.116)	66(20)
	Space	0.521(0.018)	0.036(0.026)	52(2)	0.521(0.017)	0.037(0.026)	53(2)	0.521(0.017)	0.037(0.026)	53(2)
	SR-SLasso	0.968(0.020)	0.127(0.035)	108(4)	0.970(0.019)	0.127(0.037)	108(4)	0.970(0.019)	0.127(0.037)	108(4)
	JR-SLasso	0.878(0.079)	0.023(0.017)	87(8)	0.879(0.080)	0.023(0.016)	87(8)	0.879(0.080)	0.023(0.016)	87(8)
	SSPS	0.942(0.040)	0.037(0.020)	95(4)	0.944(0.040)	0.037(0.020)	95(3)	0.944(0.040)	0.037(0.020)	95(3)
Circle	Glasso	0.975(0.010)	0.083(0.047)	53(3)	0.962(0.021)	0.265(0.072)	66(7)	0.978(0.006)	0.023(0.024)	50(1)
	Clime	0.980(0.003)	0.021(0.026)	50(1)	0.961(0.020)	0.220(0.070)	62(6)	0.981(0.004)	0.016(0.023)	50(1)
	G-scad	0.976(0.009)	0.070(0.043)	53(3)	0.964(0.020)	0.254(0.068)	65(6)	0.979(0.004)	0.016(0.018)	50(1)
	Space	0.980(0.003)	0.016(0.017)	50(1)	0.980(0.003)	0.013(0.016)	50(1)	0.980(0.003)	0.013(0.016)	50(1)
	SR-SLasso	0.963(0.007)	0.194(0.052)	60(4)	0.963(0.008)	0.191(0.049)	60(4)	0.963(0.008)	0.191(0.049)	60(4)
	JR-SLasso	0.984(0.008)	0.015(0.017)	50(1)	0.984(0.008)	0.016(0.015)	50(1)	0.984(0.008)	0.016(0.015)	50(1)
	SSPS	0.982(0.005)	0.034(0.025)	51(1)	0.981(0.005)	0.033(0.025)	51(1)	0.981(0.005)	0.033(0.025)	51(1)
Cluster	Glasso	0.730(0.088)	0.517(0.078)	157(41)	0.218(0.165)	0.446(0.168)	46(42)	0.909(0.039)	0.428(0.062)	161(23)
	Clime	0.910(0.048)	0.468(0.049)	173(20)	0.681(0.191)	0.618(0.068)	189(68)	0.988(0.014)	0.208(0.049)	125(8)
	G-scad	0.738(0.090)	0.519(0.078)	160(42)	0.236(0.177)	0.453(0.155)	51(48)	0.921(0.039)	0.430(0.059)	164(21)
	Space	0.237(0.034)	0.024(0.031)	24(3)	0.240(0.032)	0.024(0.029)	25(3)	0.240(0.032)	0.024(0.029)	25(3)
	SR-SLasso	0.983(0.006)	0.145(0.028)	115(4)	0.983(0.006)	0.145(0.028)	115(4)	0.983(0.006)	0.145(0.028)	115(4)
	JR-SLasso	0.999(0.010)	0.014(0.011)	101(2)	0.999(0.010)	0.014(0.011)	101(2)	0.999(0.010)	0.014(0.011)	101(2)
	SSPS	1.000(0.001)	0.023(0.014)	102(2)	1.000(0.001)	0.022(0.014)	102(1)	1.000(0.001)	0.022(0.014)	102(1)
ER	Glasso	0.822(0.083)	0.398(0.071)	83(14)	0.742(0.095)	0.489(0.068)	89(16)	0.852(0.077)	0.374(0.076)	83(14)
	Clime	0.896(0.060)	0.197(0.062)	68(11)	0.750(0.101)	0.409(0.069)	78(16)	0.954(0.041)	0.123(0.057)	66(10)
	G-scad	0.845(0.077)	0.370(0.073)	82(14)	0.764(0.095)	0.474(0.066)	89(17)	0.920(0.080)	0.360(0.093)	89(19)
	Space	0.909(0.051)	0.107(0.049)	61(7)	0.909(0.051)	0.094(0.045)	61(7)	0.909(0.051)	0.094(0.045)	61(7)
	SR-SLasso	0.952(0.034)	0.395(0.041)	95(11)	0.953(0.035)	0.392(0.041)	95(11)	0.953(0.035)	0.392(0.041)	95(11)
	JR-SLasso	0.946(0.052)	0.037(0.025)	59(7)	0.945(0.055)	0.035(0.025)	59(8)	0.945(0.055)	0.035(0.025)	59(8)
	SSPS	0.966(0.040)	0.032(0.030)	60(8)	0.962(0.045)	0.028(0.029)	60(8)	0.962(0.045)	0.028(0.029)	60(8)

Continued on next page

Table 2.3: Average (SD) of PDR, FDR and $|\hat{\mathbf{E}}|$ of the various methods in scenario $p < n$ $(p = 50, n = 200)$

		Transformation								
		Gaussian CDF			Power			Identity		
Graph	Method	PDR	FDR	$\lvert\hat{E}\rvert$	PDR	FDR	$\lvert\hat{E}\rvert$	PDR	FDR	$\lvert\hat{E}\rvert$
Tridiag	Glasso	0.668(0.047)	0.120(0.049)	55(3)	0.659(0.045)	0.291(0.074)	69(8)	0.669(0.046)	0.073(0.046)	53(3)
	Clime	0.672(0.046)	0.027(0.039)	50(2)	0.658(0.049)	0.213(0.084)	62(7)	0.672(0.046)	0.011(0.028)	50(2)
	G-scad	0.669(0.047)	0.097(0.047)	54(3)	0.661(0.045)	0.267(0.073)	66(7)	0.671(0.046)	0.055(0.040)	52(2)
	Space	0.672(0.046)	0.032(0.026)	51(1)	0.672(0.046)	0.032(0.026)	51(1)	0.672(0.046)	0.032(0.026)	51(1)
	SR-SLasso	0.660(0.045)	0.172(0.051)	58(4)	0.660(0.045)	0.172(0.051)	58(4)	0.660(0.045)	0.172(0.051)	58(4)
	JR-SLasso	0.672(0.046)	0.015(0.021)	50(1)	0.672(0.046)	0.015(0.021)	50(1)	0.672(0.046)	0.015(0.021)	50(1)
	SSPS	0.672(0.046)	0.019(0.020)	50(1)	0.672(0.046)	0.019(0.020)	50(1)	0.672(0.046)	0.019(0.020)	50(1)
Hub	Glasso	0.946(0.036)	0.095(0.050)	47(4)	0.792(0.071)	0.390(0.073)	60(10)	0.968(0.023)	0.048(0.033)	46(2)
	Clime	0.958(0.031)	0.072(0.053)	47(3)	0.759(0.064)	0.392(0.093)	58(11)	0.990(0.015)	0.038(0.036)	46(2)
	G-scad	0.954(0.033)	0.090(0.049)	47(3)	0.800(0.072)	0.393(0.076)	60(11)	0.977(0.021)	0.045(0.032)	46(2)
	Space	0.993(0.011)	0.017(0.020)	46(1)	0.993(0.011)	0.018(0.019)	46(1)	0.993(0.011)	0.018(0.019)	46(1)
	SR-SLasso	0.978(0.010)	0.223(0.049)	57(4)	0.978(0.010)	0.223(0.050)	57(4)	0.978(0.010)	0.223(0.050)	57(4)
	JR-SLasso	0.996(0.009)	0.043(0.036)	47(2)	0.996(0.009)	0.042(0.035)	47(2)	0.996(0.009)	0.042(0.035)	47(2)
	SSPS	0.990(0.015)	0.039(0.037)	46(2)	0.990(0.015)	0.039(0.037)	46(2)	0.990(0.015)	0.039(0.037)	46(2)
RP	Glasso	0.348(0.063)	0.075(0.039)	58(11)	0.168(0.065)	0.317(0.087)	38(15)	0.358(0.068)	0.046(0.036)	58(12)
	Clime	0.356(0.065)	0.090(0.045)	60(12)	0.165(0.067)	0.324(0.091)	38(17)	0.373(0.063)	0.060(0.038)	61(11)
	G-scad	0.349(0.063)	0.075(0.039)	58(11)	0.169(0.068)	0.317(0.087)	39(16)	0.357(0.071)	0.046(0.035)	57(12)
	Space	0.235(0.048)	0.024(0.025)	37(7)	0.235(0.048)	0.024(0.025)	37(7)	0.235(0.048)	0.024(0.025)	37(7)
	SR-SLasso	0.402(0.043)	0.369(0.040)	98(9)	0.402(0.043)	0.369(0.040)	98(9)	0.402(0.043)	0.369(0.040)	98(9)
	JR-SLasso	0.302(0.041)	0.051(0.036)	49(6)	0.302(0.041)	0.051(0.036)	49(6)	0.302(0.041)	0.051(0.036)	49(6)
	SSPS	0.352(0.041)	0.081(0.040)	59(6)	0.352(0.041)	0.081(0.040)	59(6)	0.352(0.041)	0.081(0.040)	59(6)
BA	Glasso	0.964(0.028)	0.252(0.088)	64(8)	0.929(0.041)	0.449(0.091)	85(16)	0.965(0.028)	0.194(0.084)	59(7)
	Clime	0.945(0.031)	0.214(0.110)	60(9)	0.904(0.046)	0.397(0.095)	75(12)	0.961(0.028)	0.156(0.084)	56(6)
	G-scad	0.970(0.026)	0.232(0.082)	63(7)	0.936(0.038)	0.431(0.088)	83(14)	0.973(0.027)	0.176(0.087)	59(7)
	Space	0.980(0.019)	0.240(0.051)	63(4)	0.975(0.022)	0.213(0.056)	61(4)	0.977(0.020)	0.209(0.054)	61(4)
	SR-SLasso	0.928(0.036)	0.380(0.070)	74(6)	0.931(0.038)	0.368(0.086)	73(7)	0.929(0.041)	0.370(0.082)	73(7)
	JR-SLasso	0.898(0.029)	0.048(0.035)	46(2)	0.896(0.032)	0.044(0.029)	46(2)	0.897(0.030)	0.041(0.027)	46(2)
	SSPS	0.984(0.021)	0.060(0.036)	51(2)	0.986(0.019)	0.067(0.038)	52(2)	0.983(0.020)	0.070(0.038)	52(2)

Table 2.4: Average (SD) of PDR, FDR and $|\hat{E}|$ of the various methods in scenario $p > n$ ($p = 200, n = 100$)

		Transformation								
		Gaussian CDF			Power			Identity		
Graph	Method	PDR	FDR	$\lvert\hat{E}\rvert$	PDR	FDR	$\lvert\hat{E}\rvert$	PDR	FDR	$\lvert\hat{E}\rvert$
AR(1)	Glasso	1.000(0.000)	0.994(0.000)	8550(196)	0.995(0.010)	0.994(0.000)	8713(299)	1.000(0.000)	0.994(0.000)	8550(196)
	Clime	0.995(0.010)	0.995(0.000)	9035(424)	0.971(0.025)	0.995(0.000)	8724(92)	1.000(0.000)	0.995(0.000)	8954(324)
	G-scad	1.000(0.000)	0.994(0.000)	8629(237)	0.987(0.016)	0.994(0.000)	8659(644)	0.999(0.004)	0.994(0.000)	8561(309)
	Space	0.978(0.021)	0.026(0.022)	49(2)	0.980(0.020)	0.024(0.020)	49(2)	0.980(0.020)	0.024(0.020)	49(2)
	SR-SLasso	0.990(0.013)	0.531(0.050)	105(11)	0.990(0.013)	0.530(0.050)	105(11)	0.990(0.013)	0.530(0.050)	105(11)
	JR-SLasso	0.854(0.058)	0.111(0.062)	47(6)	0.853(0.058)	0.111(0.064)	47(6)	0.853(0.058)	0.111(0.064)	47(6)
	SSPS	0.988(0.017)	0.060(0.044)	52(3)	0.989(0.015)	0.058(0.040)	52(2)	0.989(0.015)	0.058(0.040)	52(2)
AR(2)	Glasso	0.959(0.020)	0.989(0.000)	8770(185)	0.820(0.036)	0.991(0.001)	8763(414)	0.959(0.020)	0.989(0.000)	8770(185)
	Clime	0.832(0.035)	0.991(0.001)	9084(456)	0.758(0.041)	0.992(0.000)	8801(219)	0.937(0.024)	0.990(0.001)	9038(418)
	G-scad	0.843(0.025)	0.991(0.000)	8734(223)	0.786(0.036)	0.991(0.001)	8715(681)	0.949(0.025)	0.989(0.000)	8611(245)
	Space	0.405(0.033)	0.042(0.030)	41(4)	0.408(0.033)	0.041(0.031)	41(4)	0.408(0.033)	0.041(0.031)	41(4)
	SR-SLasso	0.689(0.046)	0.463(0.046)	125(10)	0.690(0.045)	0.461(0.046)	125(10)	0.690(0.045)	0.461(0.046)	125(10)
	JR-SLasso	0.427(0.033)	0.145(0.068)	49(6)	0.428(0.032)	0.145(0.068)	49(6)	0.428(0.032)	0.145(0.068)	49(6)
	SSPS	0.504(0.037)	0.125(0.062)	56(6)	0.503(0.037)	0.120(0.060)	56(6)	0.503(0.037)	0.120(0.060)	56(6)
Circle	Glasso	0.993(0.010)	0.994(0.000)	8568(207)	0.985(0.015)	0.994(0.000)	8710(316)	0.993(0.010)	0.994(0.000)	8568(207)
	Clime	0.986(0.015)	0.994(0.000)	8943(468)	0.965(0.023)	0.994(0.000)	8712(162)	0.991(0.010)	0.994(0.000)	8898(388)
	G-scad	0.990(0.012)	0.994(0.000)	8652(197)	0.971(0.020)	0.994(0.000)	8603(653)	0.992(0.010)	0.994(0.000)	8466(187)
	Space	0.961(0.021)	0.030(0.022)	50(2)	0.963(0.020)	0.028(0.022)	50(2)	0.963(0.020)	0.028(0.022)	50(2)
	SR-SLasso	0.967(0.017)	0.534(0.046)	105(10)	0.967(0.017)	0.536(0.046)	105(10)	0.967(0.017)	0.536(0.046)	105(10)
	JR-SLasso	0.837(0.052)	0.121(0.062)	48(5)	0.844(0.049)	0.127(0.062)	49(5)	0.844(0.049)	0.127(0.062)	49(5)
	SSPS	0.969(0.019)	0.074(0.046)	52(3)	0.969(0.016)	0.072(0.049)	52(3)	0.969(0.016)	0.072(0.049)	52(3)
Cluster	Glasso	0.982(0.012)	0.989(0.000)	8959(325)	0.780(0.039)	0.991(0.001)	8796(502)	0.982(0.012)	0.989(0.000)	8959(325)
	Clime	0.821(0.042)	0.991(0.001)	8948(468)	0.699(0.047)	0.992(0.001)	8885(257)	0.972(0.016)	0.989(0.001)	9049(482)
	G-scad	0.867(0.033)	0.990(0.001)	8884(259)	0.741(0.051)	0.991(0.001)	8706(599)	0.996(0.007)	0.988(0.001)	8678(371)
	Space	0.008(0.006)	0.368(0.438)	1(1)	0.009(0.006)	0.357(0.428)	1(1)	0.009(0.006)	0.357(0.428)	1(1)
	SR-SLasso	0.405(0.128)	0.595(0.090)	99(15)	0.407(0.129)	0.595(0.090)	99(15)	0.407(0.129)	0.595(0.090)	99(15)
	JR-SLasso	0.108(0.047)	0.585(0.105)	26(9)	0.108(0.043)	0.586(0.103)	26(8)	0.108(0.043)	0.586(0.103)	26(8)
	SSPS	0.113(0.056)	0.560(0.134)	26(10)	0.115(0.057)	0.561(0.135)	27(10)	0.115(0.057)	0.561(0.135)	27(10)
ER	Glasso	0.997(0.007)	0.993(0.001)	8775(188)	0.924(0.040)	0.994(0.001)	8783(363)	0.997(0.007)	0.993(0.001)	8775(188)
	Clime	0.935(0.037)	0.994(0.001)	9158(224)	0.841(0.056)	0.994(0.001)	8661(200)	0.983(0.017)	0.993(0.001)	9003(212)
	G-scad	0.957(0.029)	0.993(0.001)	8715(260)	0.877(0.062)	0.994(0.001)	8620(602)	0.993(0.008)	0.993(0.001)	8562(312)
	Space	0.573(0.108)	0.118(0.064)	39(6)	0.572(0.108)	0.114(0.064)	39(6)	0.572(0.108)	0.114(0.064)	39(6)
	SR-SLasso	0.703(0.100)	0.645(0.043)	119(13)	0.706(0.102)	0.643(0.043)	119(13)	0.706(0.102)	0.643(0.043)	119(13)
	JR-SLasso	0.577(0.099)	0.156(0.084)	41(7)	0.581(0.102)	0.155(0.083)	41(7)	0.581(0.102)	0.155(0.083)	41(7)
	SSPS	0.595(0.134)	0.079(0.054)	39(8)	0.592(0.129)	0.071(0.051)	38(8)	0.592(0.129)	0.071(0.051)	38(8)

Continued on next page

Table 2.4: Average (SD) of PDR, FDR and $|\hat{\mathbf{E}}|$ of the various methods in scenario $p > n$ ($p = 200, n = 100$)

		Transformation								
		Gaussian CDF			Power			Identity		
Graph	Method	PDR	FDR	$\lvert\hat{E}\rvert$	PDR	FDR	$\lvert\hat{E}\rvert$	PDR	FDR	$\lvert\hat{E}\rvert$
Tridiag	Glasso	0.761(0.043)	0.993(0.000)	8366(226)	0.781(0.047)	0.993(0.000)	8566(256)	0.761(0.043)	0.993(0.000)	8366(226)
	Clime	0.783(0.046)	0.994(0.000)	8949(301)	0.770(0.045)	0.993(0.000)	8568(144)	0.765(0.046)	0.994(0.000)	8796(232)
	G-scad	0.776(0.044)	0.993(0.000)	8459(181)	0.798(0.043)	0.994(0.000)	9185(406)	0.753(0.054)	0.993(0.000)	8374(445)
	Space	0.664(0.048)	0.035(0.023)	50(1)	0.664(0.048)	0.035(0.023)	50(1)	0.664(0.048)	0.035(0.023)	50(1)
	SR-SLasso	0.663(0.047)	0.543(0.050)	107(11)	0.663(0.047)	0.543(0.050)	107(11)	0.663(0.047)	0.543(0.050)	107(11)
	JR-SLasso	0.526(0.047)	0.103(0.067)	43(5)	0.526(0.046)	0.103(0.067)	43(5)	0.526(0.046)	0.103(0.067)	43(5)
	SSPS	0.664(0.047)	0.036(0.033)	50(2)	0.664(0.047)	0.036(0.033)	50(2)	0.664(0.047)	0.036(0.033)	50(2)
Hub	Glasso	0.997(0.008)	0.995(0.000)	8833(319)	0.926(0.038)	0.995(0.000)	8750(488)	0.997(0.008)	0.995(0.000)	8833(319)
	Clime	0.936(0.036)	0.995(0.000)	9081(478)	0.814(0.057)	0.996(0.000)	8856(165)	0.975(0.025)	0.995(0.000)	9125(435)
	G-scad	0.964(0.029)	0.995(0.000)	8704(327)	0.856(0.052)	0.996(0.000)	8846(645)	0.987(0.017)	0.995(0.000)	8752(242)
	Space	0.653(0.083)	0.014(0.023)	30(4)	0.654(0.084)	0.014(0.023)	30(4)	0.654(0.084)	0.014(0.023)	30(4)
	SR-SLasso	0.926(0.046)	0.557(0.043)	95(9)	0.925(0.046)	0.557(0.042)	95(9)	0.925(0.046)	0.557(0.042)	95(9)
	JR-SLasso	0.688(0.081)	0.059(0.060)	33(5)	0.689(0.083)	0.059(0.059)	33(5)	0.689(0.083)	0.059(0.059)	33(5)
	SSPS	0.683(0.071)	0.143(0.085)	36(6)	0.682(0.074)	0.141(0.084)	36(6)	0.682(0.074)	0.141(0.084)	36(6)
RP	Glasso	0.806(0.034)	0.986(0.001)	8983(309)	0.652(0.040)	0.989(0.001)	8883(452)	0.806(0.034)	0.986(0.001)	8983(309)
	Clime	0.686(0.041)	0.988(0.001)	8924(475)	0.569(0.039)	0.990(0.001)	8867(282)	0.713(0.041)	0.988(0.001)	9022(488)
	G-scad	0.723(0.030)	0.988(0.000)	8947(52)	0.587(0.042)	0.990(0.001)	8699(569)	0.744(0.038)	0.987(0.001)	8857(218)
	Space	0.027(0.024)	0.099(0.250)	4(4)	0.027(0.024)	0.099(0.250)	4(4)	0.027(0.024)	0.099(0.250)	4(4)
	SR-SLasso	0.123(0.032)	0.772(0.048)	83(11)	0.123(0.032)	0.772(0.048)	83(11)	0.123(0.032)	0.772(0.048)	83(11)
	JR-SLasso	0.076(0.025)	0.395(0.168)	20(6)	0.076(0.025)	0.395(0.168)	20(6)	0.076(0.025)	0.395(0.168)	20(6)
	SSPS	0.091(0.028)	0.509(0.113)	28(6)	0.091(0.028)	0.509(0.113)	28(6)	0.091(0.028)	0.509(0.113)	28(6)
BA	Glasso	0.998(0.008)	0.994(0.000)	8229(437)	0.936(0.040)	0.994(0.000)	8300(261)	0.998(0.008)	0.994(0.000)	8229(437)
	Clime	0.945(0.038)	0.995(0.000)	8747(577)	0.918(0.041)	0.995(0.000)	8680(322)	0.986(0.019)	0.994(0.000)	8808(486)
	G-scad	0.934(0.042)	0.994(0.000)	8227(274)	0.807(0.094)	0.694(0.273)	3348(3793)	0.985(0.030)	0.953(0.181)	7606(1743)
	Space	0.956(0.025)	0.185(0.074)	58(6)	0.967(0.024)	0.182(0.073)	58(6)	0.963(0.024)	0.162(0.069)	57(5)
	SR-SLasso	0.831(0.054)	0.637(0.041)	113(10)	0.828(0.056)	0.635(0.042)	112(10)	0.823(0.052)	0.635(0.045)	112(10)
	JR-SLasso	0.770(0.037)	0.112(0.077)	43(5)	0.761(0.034)	0.107(0.077)	42(5)	0.776(0.036)	0.102(0.072)	43(4)
	SSPS	0.828(0.075)	0.103(0.046)	45(3)	0.834(0.065)	0.104(0.048)	46(3)	0.813(0.075)	0.108(0.048)	45(4)

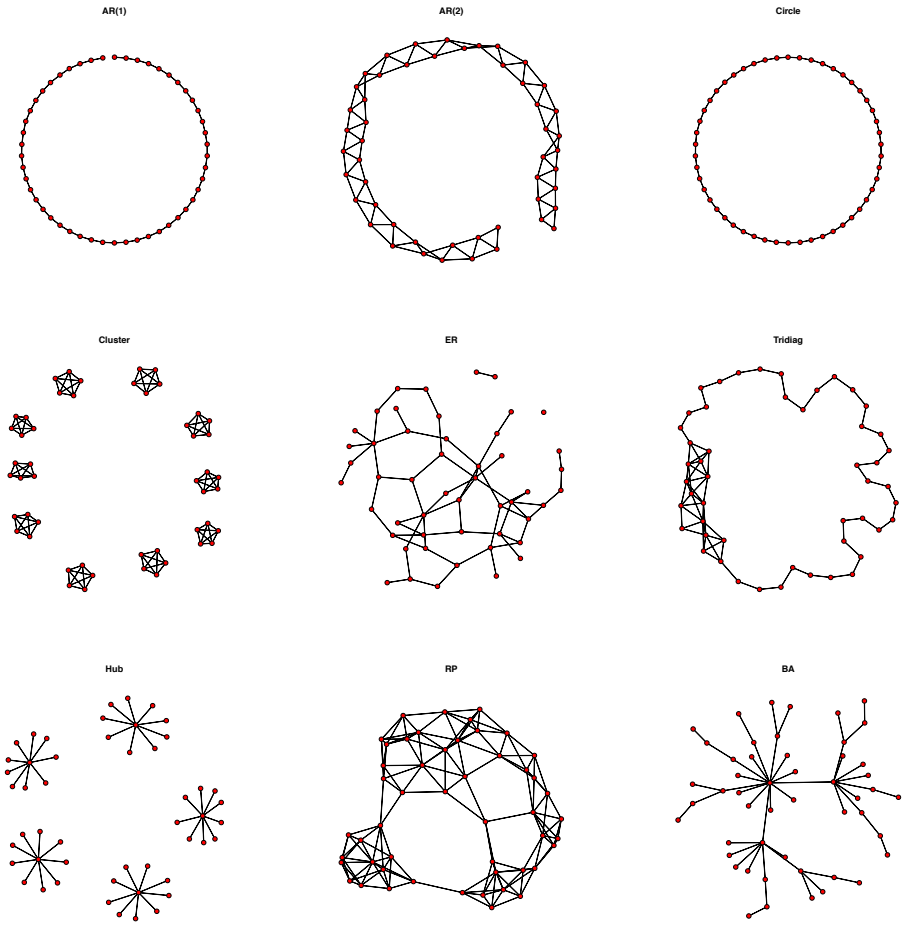

Figure 2.3: Nine simulation graphs with fifty vertices

Chapter 3

SSPS Based Precision Matrix Estimation

After the edge detection, we discuss the subsequent precision matrix estimation problem in the sparse Gaussian graphical models. Hence, the relevant notations follow those in Chapter 2. The organization of this chapter is as follows. The precision matrix estimation by the constrained optimization with SSPS edge detection is elaborated in Section 1. The idea of precision matrix estimation with SSPS screening is introduced in Section 2. In Section 3, we continue the simulation study in Chapter 2, but focus on the precision matrix estimation accuracy. In Section 4, a real example is analyzed with the proposed methods. We conclude this chapter in Section 5.

3.1 Constrained Optimization with SSPS Edge Detection

Through the SSPS edge detection, we obtain an estimate of the adjacent matrix Θ, which implies the zero and nonzero patterns in the corresponding precision matrix Ω's off-diagonal entries, as well as those in the coefficient vectors $\boldsymbol{\beta}_j$'s. Consequently, in the precision matrix estimation, we can filter out Ω's off-diagonal entries whose counterparts in the adjacent matrix are estimated as zero, and concentrate on estimating the remaining ones. With appropriate assumption of sparsity, the number of nonzero entries in Ω should be only a small proportion of the total number of the parameters, which is $p(p+1)/2$. Hence, even under the large-p-small-n circumstances, some ordinary optimization methods, such as OLS and MLE, may be sufficient to solve the precision matrix based on the sub-model, which only includes the detected nonzero entries. In fact, similar refitting procedures can be found in Cai et al. (2011) and Zhou et al. (2011), where the final estimate is improved by further optimization over the estimated support without any regularization. In summary, we estimate Ω with a two-step procedure:

(1) Estimate Θ with SSPS;

(2) Fit Ω subject to the reduced model indicated by the estimated Θ.

In what follows, denote the SSPS estimated adjacent matrix by $\hat{\Theta}$ and the detected jth vertex's neighborhood index set by s_j, for $j = 1, \ldots, p$.

3.1.1 Constrained Least Squares

We first consider to fit the coefficients in the linear regressions (2.12) by constraining those with indices not in the detected neighborhood sets to be 0. This is equivalent to the following reduced linear models:

$$\boldsymbol{y}_j = Z_j(s_j)\boldsymbol{\beta}_j(s_j) + \boldsymbol{\epsilon}_j,\ \ j = 1, \ldots, p, \tag{3.1}$$

where $\boldsymbol{\epsilon}_j$ consists of n i.i.d. normal random errors with mean zero and variance σ^{jj}. Hence, by OLS, $\boldsymbol{\beta}_j$'s coordinates with indices in s_j are estimated as

$$\hat{\boldsymbol{\beta}}_j(s_j) = [Z_j^\tau(s_j)Z_j(s_j)]^{-1}Z_j^\tau(s_j)\boldsymbol{y}_j;$$

otherwise, $\hat{\beta}_{jk} = 0$ for $k \in s_j^c$. Besides, the random error's variance σ^{jj} is estimated by the mean residual sum of squares (RSS), i.e.,

$$\hat{\sigma}^{jj} = \frac{1}{n}\big\|[\boldsymbol{I}_n - H(s_j)]\boldsymbol{y}_j\big\|_2^2.$$

Recall that $\sigma^{jj} = 1/\omega_{jj}$ and $\beta_{jk} = -\omega_{jk}/\omega_{jj}$ for $k \neq j$. Accordingly, the estimates of the Ω's diagonals are given by $\tilde{\omega}_{jj} = 1/\hat{\sigma}^{jj} = n/\|[\boldsymbol{I}_n - H(s_j)]\boldsymbol{y}_j\|_2^2$, and its off-diagonals are estimated as $\tilde{\omega}_{jk} = -\hat{\beta}_{jk}\tilde{\omega}_{jj}$, where $k \neq j$. To ensure the estimate of Ω being a symmetric matrix, we let the final estimate of Ω to be $\hat{\Omega}$ with its (j,k)th entry $\hat{\omega}_{jk}$ given by the following symmetrization:

$$\hat{\omega}_{jk} = \frac{\tilde{\omega}_{jk} + \tilde{\omega}_{kj}}{2}.$$

3.1.2 Constrained MLE

Another way to estimate Ω based on the reduced model is the constrained MLE. Denote the sample covariance as $\hat{\Sigma}_n$. Since Θ's diagonals are always 0 according to its definition, so are those of $\hat{\Theta}$. On the other hand, the diagonals of Ω should be positive given that Σ is positive definite. Let $\tilde{\Theta} = \hat{\Theta} + \boldsymbol{I}_p$, which indicates the nonzero patterns of Ω. Define the sets of zero and one entries' indices in $\tilde{\Theta}$ respectively as

$$\Gamma_0(\hat{\Theta}) = \{(j,k) : \tilde{\theta}_{jk} = 0\}, \qquad \Gamma_1(\hat{\Theta}) = \{(j,k) : \tilde{\theta}_{jk} = 1\}.$$

Consequently, the entries in Ω with indices in $\Gamma_0(\hat{\Theta})$ are forced to be 0, while the nonzero entries ω_{jk}'s, where $(j,k) \in \Gamma_1(\hat{\Theta})$, can be estimated by maximizing the log-likelihood function. This optimization problem can be expressed as:

$$\underset{\substack{\Omega_{\Gamma_0}=0 \\ \Omega \succ 0}}{\arg\min} \Big\{ \operatorname{tr}(\hat{\Sigma}_n \Omega) - \log\det(\Omega) \Big\}, \tag{3.2}$$

where $\Omega_{\Gamma_0} = \{\omega_{jk} : (j,k) \in \Gamma_0(\hat{\Theta})\}$.

3.2 Precision Matrix Estimation with SSPS Screening

Apparently, the above methods' performance closely depends on the accuracy of the estimated adjacent matrix. The numerical study in Section 2.4 demonstrates SSPS's superiority in edge detection, especially in terms of FDR. Nevertheless, implied by the values of PDR, it may fail to detect

part of the nonzero entries under some circumstances. With a noticeable number of the true nonzero entries in Θ estimated as 0 in the first step, the corresponding entries in Ω are accordingly taken as 0, which definitely compromises the estimation accuracy of the above constrained optimization methods. On the contrary, the non-sequential methods, namely Glasso, G-scad and Clime, generally have large PDRs, but suffer severely from the extremely high FDRs, which are almost 1 in the scenario of $p > n$. In other words, despite the ability to discover the majority of true nonzero components in Ω, they mistakenly estimate considerable actually null components as nonzero. Such estimates are also problematic. Furthermore, even though their penalty parameters are selected by EBIC rather than cross validation, we noticed that the computational cost of G-scad and Clime is still substantial when $p > n$.

In an effort to take advantage of both SSPS and these non-sequential methods, we propose another two-stage precision matrix estimation method, where SSPS is employed as a screening tool for the estimation of other prevalent methods. Through this process, we hope to ameliorate the overfitting problem in the existing methods and the issue of underestimation in SSPS simultaneously. The details are shown in the following.

3.2.1 SSPS Screening

The screening stage aims at selecting a subset of the precision matrix's entries for the evaluation of the next stage. Subsequently, the unselected entries are estimated as 0 and removed from the model of the next stage. This dimension reduction objective can be achieved efficiently by the SSPS

framework.

In SSPS's original procedure, a pair of nonzero entries are detected at each step, which continues until the selected model's EBIC starts to increase. Let $\hat{\Theta}_{*m}$ be the SSPS estimated adjacent matrix at the mth step and m^* be the stopping step suggested by EBIC, thus the corresponding adjacent matrix estimate is denoted by $\hat{\Theta}_{*m^*}$. Moreover, without any stopping rule, the selection procedure is carried on until all the off-diagonal entries are included. The whole sequence of the SSPS estimated adjacent matrices can be expressed as $\hat{\Theta}_{*1}, \ldots, \hat{\Theta}_{*d}$, where $d = p(p-1)/2$.

The priority of the screening is to detect the nonzero Ω entries as comprehensive as possible. Therefore, more steps than that indicated by the EBIC stopping rule may be necessary to avoid omitting the true nonzero ones, particularly when the graph is less sparse or the dimension is extremely high. We suggest to take the EBIC stopping step m^* as the reference and end the selection procedure at the $[\alpha m^*]$th step, where $\alpha \geq 1$. This scaler α shouldn't be too large; otherwise, the effect of screening is less significant. Without loss of generality, assume that αm^* is an integer. Hence, the relevant adjacent matrix is $\hat{\Theta}_{*\alpha m^*}$.

Similar to the constrained MLE method in Section 3.1.2, define $\tilde{\Theta}_{\alpha m^*} = \hat{\Theta}_{\alpha m^*} + \boldsymbol{I}_p$; let $\Gamma_0(\tilde{\Theta}_{\alpha m^*})$ be the index set of the null entries in $\tilde{\Theta}_{\alpha m^*}$ and $\Gamma_1(\tilde{\Theta}_{\alpha m^*})$ be that of the nonzero ones. Moreover, let $\Omega_{\Gamma_0^\alpha} = \{\omega_{jk} : (j,k) \in \Gamma_0(\tilde{\Theta}_{\alpha m^*})\}$ and define $\Omega_{\Gamma_1^\alpha}$ likewise. Consequently, the components in $\Omega_{\Gamma_0^\alpha}$ are estimated as 0 and exempt from future examination. Since $\Omega_{\Gamma_1^\alpha}$ inevitably contains some actually null entries, further regularization is required.

3.2.2 Estimation in the Reduced Model

At this stage, we estimate $\Omega_{\Gamma_1^\alpha}$ by some prevalent methods. We mainly consider two of them, the penalized likelihood approach and Clime. It is equivalent to apply the these methods subject to the entries in $\Omega_{\Gamma_0^\alpha}$ being estimated as 0. Instead of the K-fold cross validation, we tune the regularization parameter with the EBIC exclusively derived for the reduced model, so as to be more computationally efficient.

Penalized Likelihood Approach In this approach, the precision matrix is estimated by solving the following optimization problem:

$$\underset{\substack{\Omega_{\Gamma_0^\alpha}=0 \\ \Omega \succ 0}}{\arg\min} \Big\{ \operatorname{tr}(\hat{\Sigma}_n \Omega) - \log\det(\Omega) + \sum_{j,k=1}^{p} p_{\lambda_{jk}}(|\omega_{jk}|) \Big\}, \tag{3.3}$$

where $p_\lambda(\cdot)$ denotes a penalty function with regularization parameter λ. With p_λ to be the ℓ_1 penalty, the optimization problem in (3.3) can be delicately tackled by the Glasso algorithm (Friedman et al., 2008), which allows to customize part of the entries directly to be 0. However, due to the bias caused by the ℓ_1 penalty, the regularized optimization with the SCAD penalty (Fan and Li, 2001) is expected to be more accurate and favorable. In what follows, we briefly show the SSPS screening based G-scad method with local linear approximation (LLA, Zou and Li, 2008), in order to exploit the Glasso algorithm.

For simplicity, let $\lambda_{jk} = \lambda$ in (3.3). The first order Taylor expansion of

a general penalty function p_λ at a fixed point ω_0 is

$$p_\lambda(|\omega|) \approx p_\lambda(|\omega_0|) + p'_\lambda(|\omega_0|)(|\omega| - |\omega_0|).$$

For p_λ to be the SCAD penalty, its first derivative is

$$p'_{\lambda,a}(x) = \lambda I(|x| \leq \lambda) + \frac{(a\lambda - |x|)_+}{(a-1)\lambda} I(|x| > \lambda),$$

where $a > 2$. It is another turning parameter and recommended to be set as 3.7 (Fan and Li, 2001). Therefore, in the iterative estimation procedure of the LLA, given the ith-step estimate $\hat{\Omega}^{(i)} = (\hat{\omega}_{jk}^{(i)})$, $\hat{\Omega}^{(i+1)}$ can be taken to minimize the following penalized likelihood problem, up to a constant,

$$\arg\min_{\substack{\Omega_{\Gamma_0^c}=0 \\ \Omega \succ 0}} \left\{ \mathrm{tr}(\hat{\Sigma}_n \Omega) - \log\det(\Omega) + \sum_{j,k=1}^{p} p'_{\lambda,a}(|\hat{\omega}_{jk}^{(i)}|)|\omega_{jk}| \right\}. \tag{3.4}$$

In fact, (3.4) is a weighted ℓ_1 regularized likelihood optimization problem with further constraints on certain entries, thus it can be immediately solved by the Glasso algorithm. It has been shown that, given the sparsity of Ω, one-step algorithm of the above iterative LLA performs asymptotically the same as the fully iterated version. Hence, we initialize $\Omega^{(0)}$ with the Glasso optimizer of (3.3), and solve (3.4) for the G-scad estimate with $i = 0$.

Clime Approach Alternatively, we can add the constraints about the null entries to the ℓ_1 penalized optimization problem studied by Cai et al. (2011). For any matrix $A = (a_{ij})$, define the matrix elementwise ℓ_1 norm by $\|A\|_1 = \sum_{ij} |a_{ij}|$ and ℓ_∞ norm by $\|A\|_\infty = \max_{i,j} |a_{ij}|$. Then the afore-

mentioned optimization problem can be written as

$$\min \|\Omega\|_1, \quad \text{subject to} \quad \|\hat{\Sigma}_n\Omega - \boldsymbol{I}_p\|_\infty \leq \lambda \quad \text{and} \quad \Omega_{\Gamma_0^\alpha} = 0, \tag{3.5}$$

where λ is still the tuning parameter. With proper modifications in the original Clime algorithm to include the new constraint, the solution of (3.5) can be described as below.

Let $\boldsymbol{e}_j$ be the unit vector in $\mathbf{R}^p$, where only its jth coordinate is 1 and the others are 0, and $s_{j0}^\alpha = \{i : (i,j) \in \Gamma_0^\alpha\}$ for $1 \leq j \leq p$. Hence, (3.5) is equivalent to a series of optimization problems, which solve Ω column by column. That is, for $j = 1, \ldots, p$,

$$\min \|\boldsymbol{\phi}_j\|_1, \quad \text{subject to} \quad \|\hat{\Sigma}_n\boldsymbol{\phi}_j - \boldsymbol{e}_j\|_\infty \leq \lambda \quad \text{and} \quad \boldsymbol{\phi}_j(s_{j0}^\alpha) = 0, \tag{3.6}$$

where $\boldsymbol{\phi}_j = (\phi_{ij}) \in \mathbf{R}^p$. For numerical implementation, the regularization problem (3.6) can be further relaxed to the linear program below:

$$\min \sum_{i=1}^{p} u_{ij} \quad \text{subject to} \begin{cases} -\phi_{ij} \leq u_{ij} & \text{for } 1 \leq i \leq p, \\ \phi_{ij} \leq u_{ij} & \text{for } 1 \leq i \leq p, \\ -\hat{\Sigma}_i^\tau\boldsymbol{\phi}_j + I(i=j) \leq \lambda & \text{for } 1 \leq i \leq p, \\ \hat{\Sigma}_i^\tau\boldsymbol{\phi}_j + I(i=j) \leq \lambda & \text{for } 1 \leq i \leq p, \\ u_{ij} = 0 & \text{for } i \in s_{j0}^\alpha. \end{cases} \tag{3.7}$$

where $\hat{\Sigma}_i \in \mathbf{R}^p$ is the ith row vector in $\hat{\Sigma}_n$. Since (3.7) is a typical convex linear program, it can be solve by various linear optimization methods, such as the primal dual optimization suggested by Cai et al. (2011).

Denote the solution of (3.7) by $\hat{\boldsymbol{\phi}}_j$, for $j = 1, \ldots, p$. Let $\hat{\Phi} = (\hat{\boldsymbol{\phi}}_1, \ldots, \hat{\boldsymbol{\phi}}_p)$, which is the optimizer of (3.5). However, its asymmetry must be rectified to produce a symmetric estimate. The final estimate $\hat{\Omega} = (\hat{\omega}_{ij})_{p\times p}$, is given by the following symmetrization:

$$\hat{\omega}_{ij} = \hat{\omega}_{ji} = \hat{\phi}_{ij} I(|\hat{\phi}_{ij}| \leq |\hat{\phi}_{ji}|) + \hat{\phi}_{ji} I(|\hat{\phi}_{ij}| > |\hat{\phi}_{ji}|), \quad 1 \leq i, j \leq p. \tag{3.8}$$

Model Selection with EBIC Other than the K-fold cross validation, we resort to the EBIC as the model selection criterion. For a precision matrix estimate $\hat{\Omega}$, denote its indicating edge set by $\hat{E}$. Referring to the EBIC function for the Gaussian graphical models in Foygel and Drton (2010), we incorporate the effect of the screening stage and arrive at the following adapted version:

$$\text{EBIC}_\gamma(\hat{\Omega}) = \text{tr}(\hat{\Sigma}_n \hat{\Omega}) - \log\det(\hat{\Omega}) + |\hat{E}| \log n + 4\gamma |\hat{E}| \log p^*, \tag{3.9}$$

where γ is the model selection penalty parameter and p^* is the model complexity parameter. Here γ is set to be larger than $1 - \log n/(4 \log p^*)$, which is different from that in the SSPS's EBIC function (2.14). Moreover, p^* is positively related to the number of unknown parameters at the second stage, i.e. $p^* \propto |\Gamma_1^\alpha|$. Without the screening process, $p^* = p$, hence the EBIC in Foygel and Drton (2010) is recovered. Since some of the parameters are excluded from the second-stage estimation, then $p^* \leq p$. For $1 \leq j \leq p$, let $s_{j_1}^\alpha = \{i : (i, j) \in \Gamma_1^\alpha\}$, which consists of the indices of the jth vertex's potential neighbors. We set p^* as the cardinality of the largest potential

neighborhood index set, i.e.,

$$p^* = \max_{1 \le j \le p} |s^{\alpha}_{j_1}|. \tag{3.10}$$

Eventually, for a series of precision matrix estimators produced by a sequence of tuning parameters, we choose the one corresponding to the smallest EBIC in (3.9).

3.3 Simulation Study

In this section, we compare the methods of constrained optimization with SSPS edge detection and the methods based on SSPS screening to the other competitors, namely Glasso [Friedman et al. (2008)], G-scad [Fan et al. (2009)], Clime [Cai et al. (2011)] and Space [Peng et al. (2009)]. These methods' tuning parameters are selected as described in the original work. Specifically, Glasso, G-scad and Clime use cross validation to select the parameter that optimize an object function. Space employs a BIC-type criterion. For simplicity, we denote the SSPS edge detection based constrained LS and MLE methods by SSPS_1 and SSPS_2, and the G-scad and Clime with SSPS screening by G-scad_s and Clime_s.

The Gaussian graphical models' simulation setups are the same as the those described in Section 2.4.1. Due to the similarity between AR(1) and Circle, we only include the graph AR(1) here. Besides, we don't consider the Random Proximity graph, since the involved methods' performance in edge detection is unsatisfactory.

For any matrix $A = (a_{ij}) \in \mathbf{R}^{p\times q}$, denote its spectral norm by $\|A\|_S = \sup_{\|\boldsymbol{x}\|_2\leq 1} \|A\boldsymbol{x}\|_2$, the matrix ℓ_1 norm by $\|A\|_{\ell_1} = \max_{1\leq j\leq q} \sum_{i=1}^{p} |a_{ij}|$, and the Frobenius norm by $\|A\|_F = \sqrt{\sum_{i,j} a_{ij}^2}$. Following the work of Cai et al. (2011), we evaluate the methods' performance based on the aforementioned three matrix norms of $\Omega - \hat{\Omega}$, which is the different matrix between the true precision matrix and its estimate. In addition, the corresponding PDR, FDR and $|\hat{E}|$, defined in Section 2.4.1, are also calculated.

We still consider the two scenarios, where $p < n$ and $p > n$ respectively. Likewise, set the same values for p and n as in Section 2.4.1. For the SSPS screening, the parameter α controls the scale of possible nonzero candidates. We briefly examine the SSPS's edge detection PDR-FDR plots of the simulation examples in Figures 3.1 to Figure 3.3, with $\alpha = 1, 2, 3$ marked therein. For AR(1) and BA, $\alpha = 1$ is sufficient to include almost all the true edges, even when $p > n$. However, for Cluster in the scenario of $p > n$, the PDR just passes 0.5 when $\alpha = 3$. It is also noticeable that, for Tridiag, the PDR is barely enhanced by adding more steps after the one suggested by the EBIC ($\alpha = 1$). For simplicity, we naively take $\alpha = 2$ for the examples with $p < n$ and $\alpha = 3$ for the examples with $p > n$. We replicate the estimation 100 times and report each method's mean and standard deviation (in the parentheses) of the above measures in Table 3.1 to Table 3.3.

Overall, SSPS_1 and SSPS_2 have the best edge detection performance. Compared to the results in Section 2.4.2, without the EBIC model selection rule, the FDRs of Glasso, G-scad and Clime surge substantially even when $p < n$. Moderate augmentations in Space's FDRs can also be ob-

served. Their problem of high FDR deteriorates when p exceeds n. Such observations are apparently in contrast to those from SSPS_1 and SSPS_2. Moreover, each of the two methods with SSPS screening ameliorates the original method's high FDR issue to some extent, especially Clime_s. For illustration, based on Table 3.1 and Table 3.3, the comparison of all the methods in terms of FDR is summarized in the tables below, where we omit the results for the scenario of $p = 50$ and $n = 200$.

A summary of the FDR when $p = 50$ and $n = 100$

	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.
Glasso	0.752	0.799	0.861	0.838	0.878	0.901
G-scad	0.623	0.671	0.764	0.759	0.855	0.873
Clime	0.562	0.751	0.822	0.803	0.889	0.960
Space	0.104	0.241	0.318	0.290	0.360	0.405
SSPS	**0.023**	**0.041**	**0.054**	**0.055**	**0.067**	**0.095**
G-scad_s	0.139	0.369	0.473	0.420	0.519	0.551
Clime_s	0.095	0.118	0.141	0.157	0.198	0.230

A summary of the FDR when $p = 200$ and $n = 100$

	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.
Glasso	0.896	0.921	0.949	0.937	0.956	0.961
G-scad	0.805	0.856	0.909	0.888	0.924	0.942
Clime	0.805	0.839	0.849	0.880	0.924	0.977
Space	0.279	0.320	0.330	0.383	0.440	0.555
SSPS	**0.031**	**0.064**	**0.086**	**0.147**	**0.124**	**0.537**
G-scad_s	0.464	0.690	0.741	0.692	0.748	0.762
Clime_s	0.232	0.298	0.323	0.450	0.542	0.911

From the perspective of matrix losses, SSPS_2 generally produces estimates with the smallest loss in matrix ℓ_1 norm and Frobenius norm, regardless of the different dimensions. As $p = 50$ and n increases to 200, all the methods' estimation errors are narrowed. SSPS_2 surpasses G-scad_s in spectral norm and uniformly outperforms the others. Moreover, G-scad_s clearly

enhances the estimation of G-Scad, especially in terms of spectral norm. Comparing SSPS_1 and SSPS_2, we conclude that the constrained MLE is generally more accurate than the constrained LS. In addition, despite the significant improvement in edge detection, Clime_s surprisingly yields the estimates with largest matrix losses. We summarize each method's average ranks of the matrix losses over the seven graphs in three norms when $n = 100$, $p = 50$ and 200, respectively, in the following table.

Average ranks of matrix losses over seven graphs

	$p = 50$			$p = 200$		
	$\|\cdot\|_S$	$\|\cdot\|_{\ell_1}$	$\|\cdot\|_F$	$\|\cdot\|_S$	$\|\cdot\|_{\ell_1}$	$\|\cdot\|_F$
Glasso	6.43	6.93	6.43	6.29	7.71	7.00
G-scad	4.14	4.71	4.29	5.29	5.57	5.14
Clime	5.43	6.00	5.71	6.14	5.00	6.00
Space	3.71	3.14	4.29	3.86	3.00	3.57
SSPS_1	4.00	3.29	3.00	3.00	3.29	1.93
SSPS_2	2.43	**1.43**	**1.71**	2.14	**2.00**	**1.36**
G-scad_s	**2.00**	2.86	2.71	**1.86**	3.14	3.29
Clime_s	7.86	7.64	7.86	7.43	6.29	7.71

3.4 Application to the Breast Cancer Data

We now apply the constrained MLE with SSPS edge detection (SSPS) and the two screening based methods (G-scad_s and Clime_s) to the breast cancer data, which are available at http://bioinformatics.mdanderson.org/pubdata.html. The data consist of 22,823 gene expression levels of 133 subjects and were analyzed by Hess et al. (2006). Among the 133 subjects, 34 of them achieved pathologic Complete Response (pCR), while the remaining 99 subjects did not achieve pCR and had residual disease (RD). Compared to those with RD, patients with pCR after the neoadjuvant

chemotherapy have a better chance to survive in the long term (Kuerer et al., 1999). Hence, it is of particular interest to predict whether a patient may achieve pCR. This goal can be realized by the linear discriminant analysis (LDA), which involves the precision matrix of the gene expressions.

To compare with the other methods, we follow the analysis scheme in Cai et al. (2011) and the references therein. Its steps are briefly described as follows. 5 pCR subjects and 16 RD subjects are randomly selected to form the testing set, while the rest 112 subjects belong to the training set. A two-sample t-test is performed between the pCR group and RD group for all genes in the training dataset and the 113 most significant ones are used for the analysis. Such design aims at evaluating the methods' performance in the scenario of $p > n$. Afterwards, both the training data and the testing data are gene-wise divided by the standard deviations estimated from the training data. The resulting training data are used to obtain the precision matrix's estimate $\hat{\Omega}$. In LDA, the standardized data from two groups are assumed to follow $N_p(\boldsymbol{\mu}_k, \Sigma)$, where $p = 113$, $k = 1$ for the pCR group and $k = 2$ for the RD group. They have the same covariance Σ, but different mean vectors. For the training data, let $\hat{\pi}_k = n_k/n$ be the ratio of the observations in group k and $\hat{\boldsymbol{\mu}}_k = 1/n_k \sum_{i \in \text{group k}} \boldsymbol{x}_i$ be the within-group mean vector. An observation $\boldsymbol{x}$'s linear discriminant scores are defined as

$$\delta_k(\boldsymbol{x}) = \boldsymbol{x}^\tau \hat{\Omega} \hat{\boldsymbol{\mu}}_k - \frac{1}{2} \hat{\boldsymbol{\mu}}_k^\tau \hat{\Omega} \hat{\boldsymbol{\mu}}_k + \log \hat{\pi}_k, \quad k = 1, 2. \tag{3.11}$$

Hence, the classification rule is $\arg\max_k \delta_k(\boldsymbol{x})$, for $k = 1, 2$. The testing data are then used to assess the accuracy of the classification based on specificity, sensitivity and Mathews Correlation Coefficient (MCC). They

are defined as

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \qquad \text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}},$$

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}},$$

where TP, TN, FP and FN are the numbers of the true positives, true negatives, false positives and false negatives, respectively. MCC is widely used to evaluate the binary classifiers. The classification with the higher MCC value is usually better. The whole process is replicated 100 times. The averages and standard errors (in parentheses) of these criteria and the number of nonzero entries in $\hat{\Omega}$ are reported in Table 3.4. For comparison, we copy the results of Glasso, Adaptive Lasso, and G-Scad from Fan et al. (2009) and those of Clime from Cai et al. (2011), which are also shown in Table 3.4.

As can been seen, the proposed methods are superior in terms of specificity, but underperform the others in term of sensitivity. Overall, their MCC values follow that of Clime and are larger than the rest. The screening procedure improves the specificity values of G-scad and Clime, as well as the MCC of G-scad. The resulting estimated matrices are usually sparser than the counterparts without screening. Both SSPS and Clime$_s$ averagely produce relatively sparser estimates than the others, which are more interpretable in practice. Furthermore, we use the whole dataset to recover the network among the 113 significant genes identified by the foregoing two-sample t-test. Figure 3.4 displays the networks among the 60 most significant genes, which are recovered by SSPS, G-scad$_s$ and Clime$_s$, respectively.

Compared to those in Fan et al. (2009), similar cluster structure can be observed in all the three networks.

3.5 Conclusion

This chapter covers two approaches to the precision matrix estimation under the sparse Gaussian assumption. The first approach applies the SSPS edge detection result immediately in the relevant constrained optimization problems. The second approach modifies some existing methods by performing a screening before the estimation, where SSPS is an ad hoc dimension reduction technique. Extensive simulation examples and a real data example are employed to examine the proposed methods and compare them with the other existing ones.

Our simulation studies showed that the first approach with constrained MLE is in general more precise and less computationally intensive than the others, thus superior in precision matrix estimation. It is worth noting that, practically, the estimate of first approach with constrained LS might not be positive definite, which is an important property of Ω. Moreover, the SSPS screening not only significantly reduces the computational cost, but also improves the edge detection accuracy and the model interpretability of G-scad and Clime. In the simulations and real data analysis, we roughly set the value of α. In practice, it could be selected more delicately. The screening can be easily adapted to incorporate the prior knowledge about the real data. Due to the flexibility of SSPS framework, if we have some ex ante information, such as the sparsity of the graph or the number of edges,

we can demand the selection procedure to stop at any step, other than the suggested one. Furthermore, the G-scad with screening works quite well in the view of estimation accuracy. Although G-scad$_s$ does not have significant improvement in edge detection, its precision matrix estimation accuracy is obviously enhanced. Contrary to expectations, Clime$_s$ succeeds to reduce the FDR of Clime and enhance the overall edge detection, but fails to downscale the estimation error. Further work is needed to fix this problem.

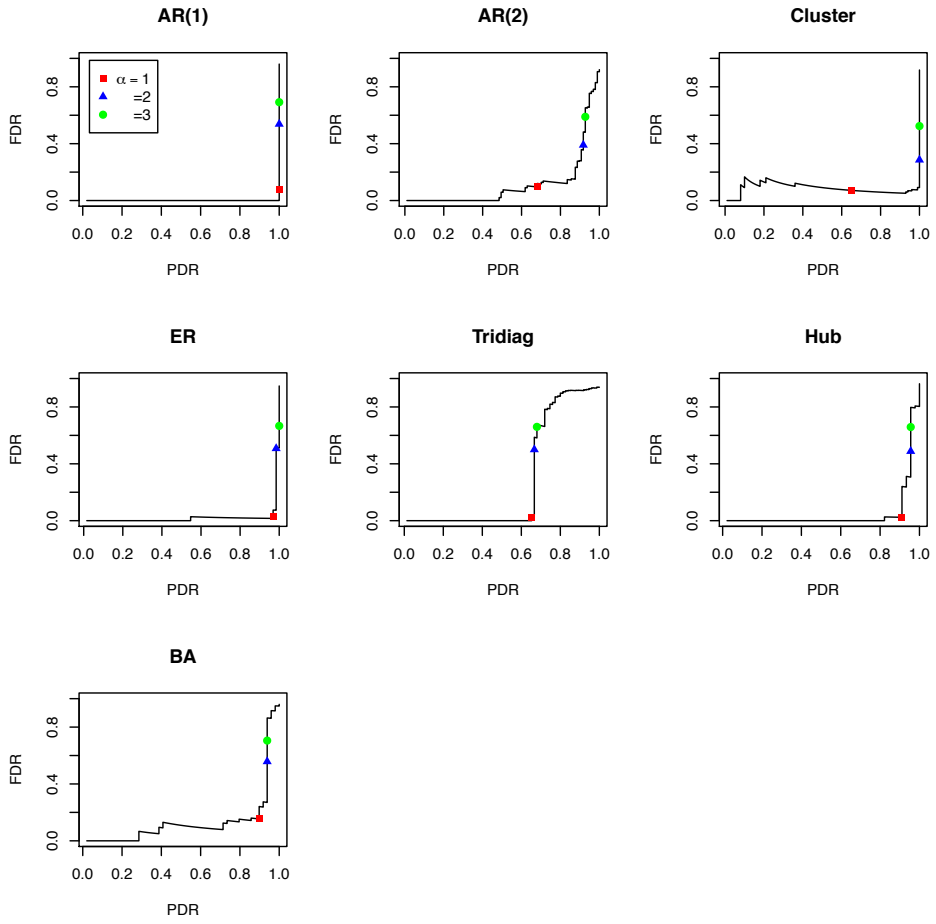

Figure 3.1: SSPS edge detection PDR-FDR path when $p = 50$, $n = 100$

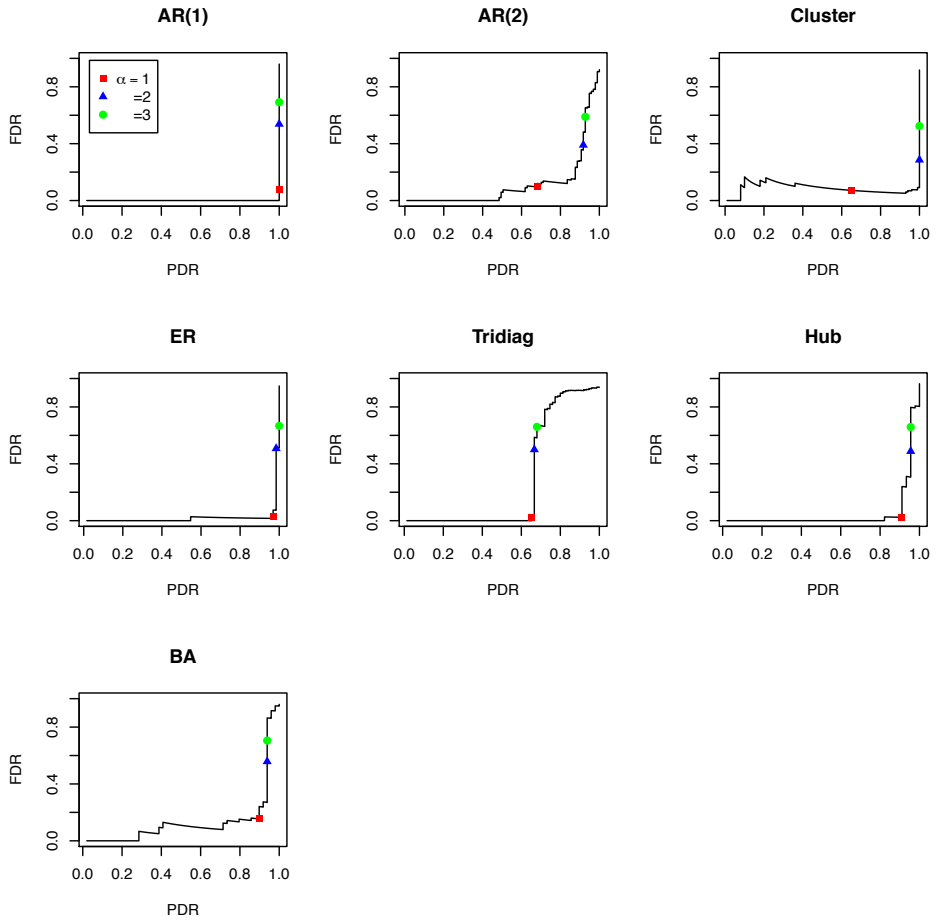

Figure 3.2: SSPS edge detection PDR-FDR path when $p = 50$, $n = 200$

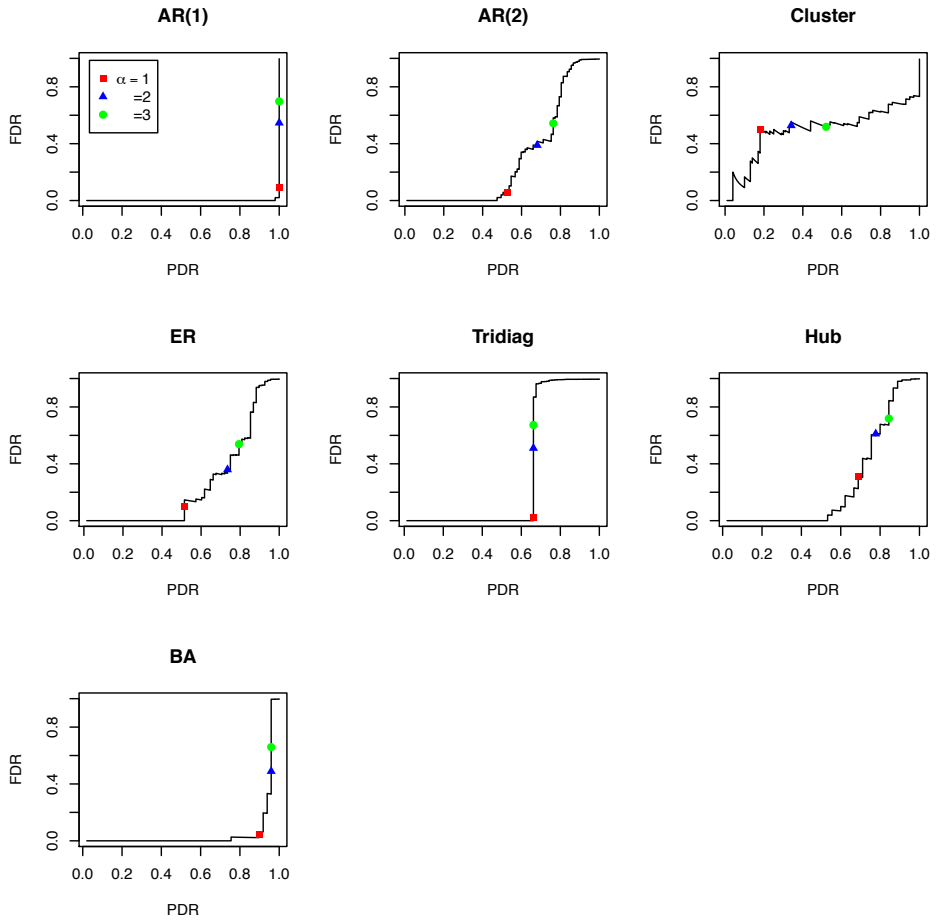

Figure 3.3: SSPS edge detection PDR-FDR path when $p = 200$, $n = 100$

Table 3.1: Average (SD) of matrix loss in three norms, PDR, FDR and number of detected edges when $p = 50$, $n = 100$

Graphs	Method	$\|\cdot\|_S$	$\|\cdot\|_{\ell_1}$	$\|\cdot\|_F$	PDR	FDR	$\|\hat{E}\|$
AR(1)	Glasso	0.841 (0.052)	1.283 (0.074)	2.907 (0.218)	1.000(0.000)	0.861(0.014)	356(40)
	G-scad	0.691 (0.084)	1.126 (0.086)	2.109 (0.267)	1.000(0.000)	0.838(0.016)	305(28)
	Clime	0.780 (0.056)	1.161 (0.083)	2.531 (0.192)	1.000(0.000)	0.822(0.045)	285(38)
	Space	0.601 (0.063)	0.806 (0.085)	1.940 (0.111)	1.000(0.000)	0.258(0.048)	66(4)
	$SSPS_1$	0.697 (0.188)	0.918 (0.271)	1.594 (0.216)	0.999(0.006)	0.036(0.028)	51(2)
	$SSPS_2$	0.626 (0.176)	0.790 (0.229)	1.460 (0.196)	0.999(0.006)	0.036(0.028)	51(2)
	$G\text{-scad}_s$	0.541 (0.065)	0.936 (0.130)	1.652 (0.101)	0.999(0.003)	0.515(0.013)	101(3)
	$Clime_s$	1.133 (0.036)	1.283 (0.049)	4.393 (0.145)	0.999(0.005)	0.121(0.044)	56(3)
AR(2)	Glasso	1.291 (0.071)	1.888 (0.088)	3.845 (0.228)	0.935(0.024)	0.792(0.017)	440(43)
	G-scad	1.168 (0.108)	1.779 (0.101)	3.474 (0.293)	0.906(0.040)	0.764(0.021)	376(42)
	Clime	1.148 (0.127)	1.703 (0.102)	3.361 (0.398)	0.927(0.059)	0.691(0.087)	313(83)
	Space	1.204 (0.054)	1.510 (0.068)	3.678 (0.120)	0.673(0.043)	0.334(0.051)	99(10)
	$SSPS_1$	0.950 (0.096)	1.409 (0.184)	3.021 (0.184)	0.661(0.057)	0.060(0.026)	68(6)
	$SSPS_2$	0.931 (0.081)	1.327 (0.156)	2.953 (0.193)	0.661(0.057)	0.060(0.026)	68(6)
	$G\text{-scad}_s$	0.878 (0.098)	1.481 (0.188)	2.476 (0.193)	0.893(0.050)	0.364(0.053)	137(15)
	$Clime_s$	1.690 (0.030)	1.893 (0.051)	5.331 (0.108)	0.613(0.034)	0.141(0.046)	69(5)
Cluster	Glasso	2.222 (0.067)	2.686 (0.071)	6.850 (0.226)	0.880(0.058)	0.752(0.031)	362(64)
	G-scad	2.283 (0.071)	2.624 (0.061)	7.051 (0.237)	0.752(0.073)	0.676(0.045)	239(56)
	Clime	2.115 (0.059)	2.507 (0.078)	6.188 (0.134)	0.944(0.022)	0.562(0.022)	216(11)
	Space	2.433 (0.027)	2.537 (0.032)	7.497 (0.069)	0.083(0.050)	0.104(0.111)	9(6)
	$SSPS_1$	2.148 (0.338)	2.439 (0.279)	4.485 (1.018)	0.710(0.168)	0.042(0.025)	74(17)
	$SSPS_2$	2.157 (0.327)	2.424 (0.295)	4.584 (1.025)	0.710(0.168)	0.042(0.025)	74(17)
	$G\text{-scad}_s$	1.911 (0.676)	2.289 (0.514)	5.213 (2.622)	0.557(0.461)	0.139(0.134)	71(62)
	$Clime_s$	2.582 (0.019)	2.662 (0.021)	8.021 (0.051)	0.058(0.031)	0.095(0.126)	6(3)
ER	Glasso	1.521 (0.248)	2.801 (0.445)	4.215 (0.550)	0.990(0.017)	0.806(0.050)	332(85)
	G-scad	1.321 (0.244)	2.245 (0.408)	3.597 (0.525)	0.969(0.039)	0.666(0.067)	184(37)
	Clime	1.710 (0.206)	3.285 (0.400)	4.959 (0.513)	0.996(0.009)	0.891(0.031)	585(90)
	Space	1.345 (0.273)	2.162 (0.479)	3.721 (0.510)	0.933(0.058)	0.386(0.063)	94(15)
	$SSPS_1$	1.443 (0.330)	2.319 (0.588)	3.774 (0.643)	0.816(0.088)	0.089(0.043)	55(6)
	$SSPS_2$	1.423 (0.346)	2.222 (0.589)	3.674 (0.675)	0.816(0.088)	0.089(0.043)	55(6)
	$G\text{-scad}_s$	1.307 (0.257)	2.108 (0.454)	3.616 (0.541)	0.929(0.053)	0.473(0.038)	105(12)
	$Clime_s$	2.325 (0.276)	3.557 (0.555)	7.150 (0.644)	0.814(0.128)	0.194(0.091)	61(13)
Tridiag	Glasso	2.791 (0.275)	4.145 (0.299)	8.425 (0.783)	0.778(0.046)	0.864(0.015)	418(42)
	G-scad	1.767 (0.401)	2.375 (0.507)	4.326 (0.483)	0.708(0.042)	0.623(0.043)	137(14)
	Clime	3.076 (0.254)	4.278 (0.546)	9.628 (0.737)	0.748(0.057)	0.811(0.045)	316(130)
	Space	1.896 (0.217)	2.458 (0.270)	5.633 (0.362)	0.681(0.040)	0.223(0.042)	63(3)
	$SSPS_1$	2.049 (0.541)	2.665 (0.775)	4.532 (0.657)	0.678(0.041)	0.031(0.026)	50(1)
	$SSPS_2$	1.798 (0.461)	2.255 (0.618)	4.046 (0.581)	0.678(0.041)	0.031(0.026)	50(1)
	$G\text{-scad}_s$	1.953 (0.617)	2.666 (0.930)	4.627 (0.904)	0.687(0.047)	0.374(0.072)	81(11)
	$Clime_s$	4.098 (0.245)	4.797 (0.294)	14.471(0.680)	0.679(0.047)	0.115(0.064)	56(6)
Hub	Glasso	6.696 (0.360)	11.862(0.649)	13.690(0.532)	0.977(0.026)	0.891(0.009)	407(36)
	G-scad	5.194 (0.632)	9.464 (1.164)	10.378(1.198)	0.951(0.039)	0.872(0.014)	338(37)
	Clime	4.897 (0.401)	9.921 (0.711)	9.801 (0.669)	0.812(0.090)	0.887(0.017)	329(49)
	Space	2.815 (0.607)	5.905 (0.957)	5.003 (0.696)	0.980(0.023)	0.318(0.076)	65(8)
	$SSPS_1$	3.142 (1.075)	5.604 (1.498)	5.076 (0.954)	0.855(0.058)	0.130(0.048)	44(3)
	$SSPS_2$	2.305 (0.528)	4.336 (0.886)	4.134 (0.606)	0.855(0.058)	0.130(0.048)	44(3)
	$G\text{-scad}_s$	2.666 (0.740)	5.031 (1.184)	4.758 (0.788)	0.930(0.039)	0.523(0.037)	88(7)
	$Clime_s$	6.197 (0.438)	11.822(0.839)	12.712(0.780)	0.557(0.219)	0.230(0.108)	34(15)
BA	Glasso	8.641 (2.550)	16.156(5.121)	11.436(1.850)	0.994(0.010)	0.901(0.008)	494(36)
	G-scad	7.618 (2.832)	14.256(5.563)	9.604 (2.710)	0.993(0.012)	0.873(0.015)	388(45)
	Clime	10.676(2.690)	19.893(5.463)	15.649(1.654)	1.000(0.000)	0.960(0.000)	1216(3)
	Space	2.677 (0.781)	5.642 (1.666)	4.039 (0.631)	0.986(0.015)	0.405(0.043)	82(6)
	$SSPS_1$	3.115 (1.151)	7.271 (3.010)	4.655 (0.989)	0.841(0.050)	0.211(0.066)	52(3)
	$SSPS_2$	3.760 (1.530)	7.390 (3.280)	4.945 (1.336)	0.841(0.050)	0.211(0.066)	52(3)
	$G\text{-scad}_s$	4.971 (1.970)	9.186 (3.878)	5.976 (1.721)	0.902(0.039)	0.551(0.031)	99(5)
	$Clime_s$	10.926(2.628)	20.085(5.401)	16.170(1.528)	0.801(0.081)	0.202(0.070)	49(5)

Table 3.2: Average (SD) of matrix loss in three norms, PDR, FDR and number of detected edges when $p = 50$, $n = 200$

Graphs	Method	$\|\cdot\|_S$	$\|\cdot\|_{\ell_1}$	$\|\cdot\|_F$	PDR	FDR	$\|\hat{E}\|$
AR(1)	Glasso	0.669 (0.041)	1.035 (0.067)	2.225 (0.107)	1.000(0.000)	0.872(0.007)	385(20)
	G-scad	0.544 (0.063)	0.809 (0.084)	1.606 (0.159)	1.000(0.000)	0.795(0.016)	241(22)
	Clime	0.547 (0.059)	0.800 (0.070)	1.626 (0.162)	1.000(0.000)	0.766(0.015)	210(13)
	Space	0.447 (0.050)	0.589 (0.070)	1.371 (0.084)	1.000(0.000)	0.234(0.051)	64(4)
	SSPS_1	0.452 (0.104)	0.595 (0.140)	1.064 (0.134)	1.000(0.000)	0.036(0.030)	51(2)
	SSPS_2	0.413 (0.101)	0.527 (0.123)	0.971 (0.122)	1.000(0.000)	0.036(0.030)	51(2)
	G-scad_s	0.504 (0.093)	0.819 (0.164)	1.279 (0.111)	1.000(0.000)	0.514(0.012)	101(2)
	Clime_s	0.823 (0.049)	0.969 (0.058)	2.996 (0.176)	1.000(0.000)	0.177(0.047)	60(3)
AR(2)	Glasso	1.065 (0.044)	1.573 (0.072)	3.079 (0.100)	0.998(0.005)	0.798(0.008)	480(19)
	G-scad	0.766 (0.079)	1.334 (0.097)	2.185 (0.186)	0.998(0.006)	0.784(0.015)	449(28)
	Clime	0.858 (0.054)	1.254 (0.091)	2.393 (0.111)	0.995(0.007)	0.636(0.015)	266(11)
	Space	0.984 (0.043)	1.241 (0.063)	3.002 (0.087)	0.903(0.030)	0.299(0.032)	125(7)
	SSPS_1	0.663 (0.111)	0.974 (0.183)	1.572 (0.228)	0.955(0.033)	0.037(0.020)	96(3)
	SSPS_2	0.636 (0.105)	0.937 (0.175)	1.490 (0.240)	0.955(0.033)	0.037(0.020)	96(3)
	G-scad_s	0.653 (0.075)	1.244 (0.143)	1.747 (0.128)	0.997(0.006)	0.497(0.018)	193(7)
	Clime_s	1.308 (0.146)	1.556 (0.117)	4.000 (0.502)	0.886(0.067)	0.191(0.107)	109(25)
Cluster	Glasso	1.741 (0.040)	2.351 (0.084)	5.248 (0.107)	1.000(0.000)	0.823(0.007)	564(21)
	G-scad	1.338 (0.109)	2.019 (0.132)	3.642 (0.303)	1.000(0.000)	0.818(0.007)	551(20)
	Clime	1.383 (0.084)	1.768 (0.113)	3.685 (0.162)	1.000(0.000)	0.638(0.016)	277(12)
	Space	1.781 (0.058)	1.970 (0.067)	5.259 (0.100)	0.998(0.005)	0.314(0.034)	146(7)
	SSPS_1	0.712 (0.184)	0.997 (0.257)	1.450 (0.172)	1.000(0.000)	0.029(0.018)	103(2)
	SSPS_2	0.705 (0.193)	0.966 (0.245)	1.426 (0.171)	1.000(0.000)	0.029(0.018)	103(2)
	G-scad_s	0.893 (0.157)	1.548 (0.215)	2.011 (0.178)	1.000(0.000)	0.513(0.008)	206(4)
	Clime_s	2.346 (0.289)	2.476 (0.202)	7.166 (1.103)	0.406(0.422)	0.110(0.151)	56(65)
ER	Glasso	1.330 (0.225)	2.324 (0.406)	3.597 (0.502)	0.998(0.007)	0.780(0.057)	296(76)
	G-scad	0.872 (0.225)	1.467 (0.316)	2.338 (0.425)	0.996(0.008)	0.654(0.106)	192(52)
	Clime	1.444 (0.190)	2.648 (0.452)	4.113 (0.503)	1.000(0.002)	0.872(0.037)	500(91)
	Space	1.002 (0.213)	1.605 (0.369)	2.691 (0.385)	0.992(0.013)	0.376(0.062)	99(15)
	SSPS_1	0.821 (0.212)	1.225 (0.372)	2.043 (0.421)	0.979(0.027)	0.050(0.032)	64(8)
	SSPS_2	0.772 (0.217)	1.141 (0.385)	1.929 (0.434)	0.979(0.027)	0.050(0.032)	64(8)
	G-scad_s	0.833 (0.222)	1.291 (0.325)	2.245 (0.490)	0.996(0.009)	0.403(0.105)	101(18)
	Clime_s	1.863 (0.186)	2.887 (0.443)	5.636 (0.548)	0.989(0.015)	0.255(0.099)	80(12)
Tridiag	Glasso	2.403 (0.309)	3.466 (0.288)	7.337 (1.000)	0.771(0.061)	0.856(0.021)	397(80)
	G-scad	1.125 (0.238)	1.449 (0.270)	2.733 (0.273)	0.695(0.048)	0.466(0.056)	95(11)
	Clime	2.142 (0.226)	3.407 (0.429)	6.363 (0.699)	0.777(0.065)	0.846(0.071)	421(128)
	Space	1.363 (0.164)	1.780 (0.205)	3.972 (0.257)	0.682(0.050)	0.210(0.042)	62(3)
	SSPS_1	1.254 (0.284)	1.622 (0.377)	2.925 (0.308)	0.681(0.050)	0.020(0.022)	50(1)
	SSPS_2	1.131 (0.256)	1.402 (0.298)	2.640 (0.278)	0.681(0.050)	0.020(0.022)	50(1)
	G-scad_s	1.126 (0.190)	1.439 (0.262)	2.930 (0.366)	0.681(0.052)	0.235(0.062)	65(6)
	Clime_s	3.155 (0.262)	3.758 (0.337)	10.955(0.823)	0.676(0.048)	0.147(0.059)	58(6)
Hub	Glasso	5.825 (0.313)	10.229(0.559)	11.863(0.442)	1.000(0.000)	0.899(0.005)	448(23)
	G-scad	2.372 (1.002)	4.389 (1.688)	4.413 (1.905)	0.997(0.009)	0.875(0.015)	364(40)
	Clime	3.056 (0.434)	6.438 (0.740)	5.852 (0.728)	0.995(0.009)	0.918(0.003)	544(15)
	Space	1.863 (0.434)	4.016 (0.776)	3.377 (0.464)	1.000(0.000)	0.273(0.074)	63(7)
	SSPS_1	1.853 (0.697)	3.118 (1.005)	2.826 (0.697)	0.992(0.012)	0.079(0.039)	49(2)
	SSPS_2	1.329 (0.437)	2.308 (0.616)	2.247 (0.451)	0.992(0.012)	0.079(0.039)	49(2)
	G-scad_s	1.576 (0.439)	2.959 (0.655)	2.815 (0.458)	0.998(0.007)	0.535(0.019)	97(4)
	Clime_s	4.507 (0.396)	9.016 (0.661)	8.983 (0.691)	0.922(0.044)	0.278(0.059)	58(5)
BA	Glasso	8.680 (2.750)	16.172(5.459)	11.378(2.073)	0.999(0.004)	0.887(0.008)	435(30)
	G-scad	6.254 (2.855)	11.679(5.585)	7.166 (2.668)	0.998(0.006)	0.867(0.011)	369(30)
	Clime	10.819(2.872)	20.185(5.794)	15.707(1.853)	1.000(0.000)	0.960(0.000)	1213(4)
	Space	1.977 (0.833)	4.152 (1.664)	2.858 (0.668)	0.995(0.011)	0.386(0.052)	80(7)
	SSPS_1	1.892 (0.860)	3.997 (2.125)	2.716 (0.775)	0.951(0.037)	0.157(0.069)	56(3)
	SSPS_2	2.021 (1.220)	3.775 (2.431)	2.616 (1.139)	0.951(0.037)	0.157(0.069)	56(3)
	G-scad_s	3.688 (1.626)	6.561 (3.128)	4.383 (1.531)	0.977(0.022)	0.529(0.025)	102(5)
	Clime_s	10.408(3.042)	19.041(6.197)	15.857(1.824)	0.945(0.053)	0.392(0.193)	86(33)

Table 3.3: Average (SD) of matrix loss in three norms, PDR, FDR and number of detected edges when $p = 200$, $n = 100$

Graphs	Method	$\|\cdot\|_S$	$\|\cdot\|_{\ell_1}$	$\|\cdot\|_F$	PDR	FDR	$\|\hat{E}\|$
AR(1)	Glasso	0.966 (0.040)	1.591 (0.100)	4.379 (0.178)	1.000(0.000)	0.949(0.011)	997(192)
	G-scad	0.960 (0.047)	1.425 (0.084)	4.138 (0.175)	1.000(0.000)	0.909(0.016)	554(99)
	Clime	0.959 (0.046)	1.161 (0.063)	4.244 (0.121)	1.000(0.003)	0.830(0.010)	288(17)
	Space	0.771 (0.055)	0.954 (0.064)	3.311 (0.123)	0.999(0.004)	0.318(0.090)	73(10)
	$SSPS_1$	0.794 (0.140)	1.001 (0.241)	2.620 (0.216)	0.992(0.014)	0.057(0.046)	52(3)
	$SSPS_2$	0.760 (0.140)	0.909 (0.196)	2.571 (0.196)	0.992(0.014)	0.057(0.046)	52(3)
	$G\text{-scad}_s$	0.849 (0.092)	1.409 (0.226)	4.100 (0.202)	0.998(0.006)	0.742(0.017)	191(12)
	$Clime_s$	1.347 (0.027)	1.491 (0.041)	6.739 (0.121)	0.974(0.026)	0.307(0.175)	74(23)
AR(2)	Glasso	1.512 (0.040)	2.177 (0.112)	5.302 (0.156)	0.818(0.039)	0.911(0.017)	936(227)
	G-scad	1.521 (0.048)	2.045 (0.104)	5.134 (0.157)	0.758(0.038)	0.857(0.028)	536(125)
	Clime	1.539 (0.035)	1.807 (0.056)	5.292 (0.088)	0.661(0.035)	0.805(0.014)	330(14)
	Space	1.424 (0.050)	1.684 (0.071)	4.735 (0.111)	0.528(0.028)	0.330(0.084)	78(12)
	$SSPS_1$	1.111 (0.124)	1.586 (0.223)	4.111 (0.170)	0.516(0.038)	0.112(0.052)	57(6)
	$SSPS_2$	1.106 (0.105)	1.518 (0.166)	4.111 (0.173)	0.516(0.038)	0.112(0.052)	57(6)
	$G\text{-scad}_s$	1.079 (0.067)	1.716 (0.203)	4.968 (0.209)	0.681(0.053)	0.661(0.035)	197(20)
	$Clime_s$	1.939 (0.027)	2.087 (0.034)	7.415 (0.085)	0.372(0.032)	0.232(0.163)	50(14)
Cluster	Glasso	2.368 (0.027)	2.884 (0.115)	7.798 (0.108)	0.704(0.062)	0.896(0.019)	710(184)
	G-scad	2.407 (0.023)	2.722 (0.071)	7.771 (0.079)	0.571(0.049)	0.805(0.035)	305(70)
	Clime	2.482 (0.025)	2.623 (0.026)	7.979 (0.049)	0.508(0.041)	0.848(0.015)	335(18)
	Space	2.433 (0.023)	2.539 (0.025)	7.773 (0.051)	0.045(0.027)	0.555(0.175)	10(6)
	$SSPS_1$	2.419 (0.027)	2.742 (0.116)	7.688 (0.179)	0.119(0.054)	0.537(0.128)	26(9)
	$SSPS_2$	2.414 (0.027)	2.731 (0.110)	7.717 (0.167)	0.119(0.054)	0.537(0.128)	26(9)
	$G\text{-scad}_s$	2.340 (0.095)	2.723 (0.153)	6.953 (0.448)	0.413(0.164)	0.464(0.102)	78(32)
	$Clime_s$	2.626 (0.139)	2.701 (0.151)	9.090 (0.176)	0.011(0.011)	0.911(0.182)	28(18)
ER	Glasso	2.084 (0.263)	3.718 (0.607)	6.585 (0.606)	0.914(0.069)	0.931(0.026)	944(373)
	G-scad	1.892 (0.287)	3.333 (0.570)	5.665 (0.596)	0.857(0.100)	0.854(0.052)	413(169)
	Clime	2.117 (0.306)	3.457 (0.729)	6.417 (0.608)	0.863(0.086)	0.849(0.037)	398(220)
	Space	1.740 (0.266)	2.719 (0.539)	5.065 (0.531)	0.794(0.107)	0.338(0.080)	75(15)
	$SSPS_1$	1.572 (0.283)	2.500 (0.621)	4.853 (0.676)	0.610(0.136)	0.070(0.053)	40(8)
	$SSPS_2$	1.540 (0.281)	2.451 (0.628)	4.801 (0.690)	0.610(0.136)	0.070(0.053)	40(8)
	$G\text{-scad}_s$	1.489 (0.276)	2.355 (0.507)	5.220 (0.456)	0.826(0.091)	0.720(0.033)	181(23)
	$Clime_s$	2.682 (0.243)	4.243 (0.638)	8.929 (0.488)	0.342(0.141)	0.323(0.200)	35(21)
Tridiag	Glasso	4.516 (0.260)	5.591 (0.330)	15.459(0.686)	0.685(0.046)	0.961(0.007)	1317(240)
	G-scad	3.098 (0.327)	3.883 (0.385)	10.287(0.838)	0.679(0.045)	0.925(0.015)	687(140)
	Clime	3.163 (0.259)	3.878 (0.325)	10.701(0.526)	0.678(0.044)	0.877(0.005)	400(15)
	Space	2.497 (0.224)	3.066 (0.308)	8.090 (0.424)	0.673(0.045)	0.321(0.077)	73(9)
	$SSPS_1$	2.070 (0.518)	2.709 (0.816)	4.975 (0.612)	0.667(0.045)	0.031(0.031)	50(2)
	$SSPS_2$	1.813 (0.400)	2.261 (0.532)	4.580 (0.562)	0.667(0.045)	0.031(0.031)	50(2)
	$G\text{-scad}_s$	1.756 (0.252)	2.444 (0.378)	6.886 (0.367)	0.674(0.045)	0.741(0.017)	190(12)
	$Clime_s$	4.229 (0.358)	4.985 (0.409)	15.611(1.299)	0.676(0.044)	0.290(0.232)	78(28)
Hub	Glasso	7.535 (0.342)	13.431(0.640)	15.908(0.569)	0.858(0.078)	0.958(0.009)	959(226)
	G-scad	6.600 (0.334)	12.315(0.674)	13.849(0.535)	0.726(0.119)	0.923(0.015)	443(119)
	Clime	4.566 (0.597)	10.146(0.872)	9.678 (1.073)	0.164(0.094)	0.972(0.015)	274(134)
	Space	3.941 (0.497)	8.119 (0.905)	7.598 (0.658)	0.916(0.052)	0.279(0.084)	58(8)
	$SSPS_1$	3.320 (1.015)	6.512 (1.589)	6.086 (0.905)	0.706(0.076)	0.136(0.077)	37(6)
	$SSPS_2$	2.761 (0.679)	5.597 (1.226)	5.509 (0.637)	0.706(0.076)	0.136(0.077)	37(6)
	$G\text{-scad}_s$	2.427 (0.563)	4.776 (1.034)	7.224 (0.482)	0.844(0.060)	0.754(0.027)	156(17)
	$Clime_s$	6.168 (0.612)	12.182(0.895)	12.912(1.448)	0.095(0.074)	0.670(0.269)	18(13)
BA	Glasso	10.074(3.025)	18.730(6.083)	13.961(1.917)	0.996(0.008)	0.953(0.011)	1087(247)
	G-scad	9.219 (2.999)	17.227(5.998)	12.478(2.141)	0.995(0.010)	0.942(0.014)	890(202)
	Clime	9.841 (3.046)	19.073(6.457)	13.556(2.060)	0.838(0.119)	0.977(0.010)	1983(659)
	Space	4.306 (1.035)	8.867 (2.124)	6.878 (0.829)	0.973(0.022)	0.542(0.064)	106(15)
	$SSPS_1$	3.169 (1.525)	7.302 (3.890)	5.037 (1.292)	0.842(0.081)	0.086(0.042)	45(4)
	$SSPS_2$	3.287 (1.607)	6.599 (3.472)	4.882 (1.443)	0.842(0.081)	0.086(0.042)	45(4)
	$G\text{-scad}_s$	5.233 (2.308)	9.546 (4.514)	7.262 (1.830)	0.841(0.060)	0.762(0.029)	175(18)
	$Clime_s$	9.959 (2.890)	18.401(5.879)	15.932(1.697)	0.793(0.067)	0.415(0.175)	74(28)

Table 3.4: Comparison of average (SE) pCR classification results

Methods	Specificity	Sensitivity	MCC	$\lvert\hat{\Omega}\rvert$
Glasso	0.768(0.009)	0.630(0.021)	0.366(0.018)	3923(2)
Adaptive Lasso	0.787(0.009)	0.622(0.022)	0.381(0.018)	1233(1)
G-scad	0.794(0.005)	0.634(0.022)	0.402(0.020)	674(1)
Clime	0.749(0.005)	0.806(0.017)	0.506(0.020)	492(7)
SSPS	0.821(0.009)	0.660(0.020)	0.461(0.020)	436(3)
G-scad$_s$	0.839(0.008)	0.598(0.021)	0.430(0.021)	606(4)
Clime$_s$	0.846(0.010)	0.582(0.022)	0.432(0.021)	436(5)

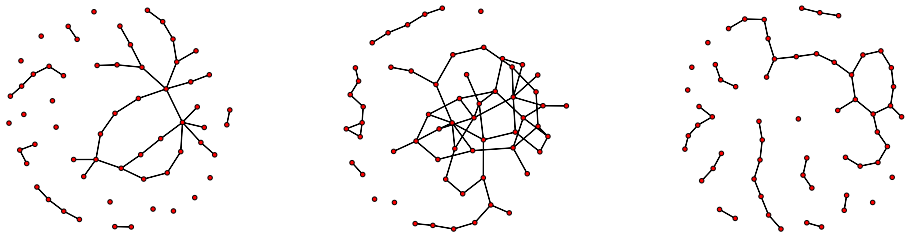

Figure 3.4: Gene networks recovered by three methods: SSPS (left panel), G-scad$_s$ (middle panel), and Clime$_s$ (right panel)

Chapter 4

Joint Estimation of the Coefficient Matrix and Precision Matrix in Multivariate Response Regression Models

In this chapter, we fit the sparse multivariate response regression models with two sequential conditional regression (SCR) methods. The outline is listed as follows. Section 1 introduces the multivariate response regression models in two formulations to jointly estimate the coefficient matrix and precision matrix. In Section 2, we elaborate on the two SCR methods. The selection consistency of each method is discussed in Section 3. In Section 4, we report the results of the simulation study in three scenarios. In Section 5, the proposed methods are applied to a real data. We conclude this study in Section 6.

4.1 Multivariate Response Regression Model

Assume $\boldsymbol{X}$ is a q-vector covariate variable and $\boldsymbol{Y}$ is a p-vector response variable. There are n samples of them: $\{\boldsymbol{x}_i, \boldsymbol{y}_i\}_{i=1}^n$, where $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{iq})^\tau \in \mathbf{R}^q$ and $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{ip})^\tau \in \mathbf{R}^p$. Let $X = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^\tau \in \mathbf{R}^{n\times q}$ be the covariate matrix and $Y = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)^\tau \in \mathbf{R}^{n\times p}$ be the response matrix. Denote the jth column of Y by $\boldsymbol{Y}_j$ and the kth column of X by $\boldsymbol{X}_k$, for $j = 1, \ldots, p$ and $k = 1, \ldots, q$. Without loss of generality, we suppose that every column vector in Y and X is centered to have mean 0. We have the following multivariate response regression model:

$$Y = XB + E, \tag{4.1}$$

where $B = (\beta_{ij})_{q\times p}$ is the covariate coefficient matrix, and $E = (e_{ij})_{n\times p}$ is the error matrix with its row vectors i.i.d. $N_p(\mathbf{0}, \Sigma)$. Hence, Σ is the response variables' covariance matrix. Fitting the multivariate response regression model (4.1) involves solving the coefficient matrix B and the covariance matrix Σ, or equivalently its inverse, the precision matrix, denoted by $\Omega = (\omega_{ij})_{p\times p}$. If the data are of high dimension, it is reasonable to assume the sparsity in B and Ω accordingly.

Explicitly, in the model (4.1), the responses' marginal distributions are

$$\boldsymbol{Y}_j | X \sim N_n(X\boldsymbol{\beta}_j, \sigma_j^2 \boldsymbol{I}_n), \quad j = 1, \ldots, p. \tag{4.2}$$

where $\boldsymbol{\beta}_j = (\beta_{1j}, \ldots, \beta_{qj})^\tau$ and σ_j^2 is the jth diagonal entry of Σ. This can

be equivalently expressed as following marginal regressions:

$$\boldsymbol{Y}_j = X\boldsymbol{\beta}_j + \boldsymbol{e}_j, \quad j = 1, \ldots, p, \tag{4.3}$$

where $\boldsymbol{e}_j = (e_{1j}, \ldots, e_{nj})^\tau \sim N_n(\mathbf{0}, \sigma_j^2 \boldsymbol{I}_n)$. Therefore, a naive approach to solve B is to fit the p marginal regressions in (4.3) separately, which completely neglects the mutual influence among them caused by the correlated errors $\boldsymbol{e}_j$'s. This approach will be feasible if all the $\boldsymbol{e}_j$'s are actually independent. However, this is seldom the case in practice. Ignoring the effect of such correlations when only limited number of samples are available, the coefficient matrix B's estimation efficiency and accuracy will definitely be compromised.

Alternatively, the multivariate response regression model (4.1) can be interpreted as a p-dimension Gaussian graphical model with precision matrix Ω. Different from the standard one, this Gaussian graphical model has a non-constant mean vector $(\boldsymbol{X}^\tau\boldsymbol{\beta}_1, \ldots, \boldsymbol{X}^\tau\boldsymbol{\beta}_p)^\tau$. If all the $\boldsymbol{\beta}_j$'s are $\mathbf{0}$, then it is reduced to the standard Gaussian graphical model that has been studied in various literature of this area, and the precision matrix Ω can be solved accordingly with the techniques therein (e.g., Friedman et al., 2008; Fan et al., 2009; Cai et al., 2011; Peng et al., 2009, etc.). Despite its sparsity, it is inappropriate to assume that B is a null matrix. Subsequently, in the multivariate response regression model (4.1), estimating Ω without adjusting the non-constant means is likely to suffer from false discovery of conditional dependent correlations among the response variables.

In conclusion, the standalone estimation of B and Ω is ineffective, even problematic, under the circumstances of high-dimensional parameters and

low sample size; instead, methods that take their interaction into account can be more efficient and thus favorable.

4.1.1 Penalized Likelihood Formulation

One plausible way to formulate the coefficient matrix B and the precision matrix Ω together is the maximum likelihood estimation. The log-likelihood function of (B, Ω), up to a constant, can be written as

$$\log\det(\Omega) - \mathrm{tr}[\frac{1}{n}(Y - XB)^\tau(Y - XB)\Omega].$$

It turns out that the MLE of B is $(X^\tau X)^{-1}X^\tau Y$, which is the simple combination of the p marginal regressions' OLS estimates. It is independent of Ω, thus does not leverage on the information of the responses' correlations. Fortunately, the joint estimation of B and Ω can be realized in the penalized likelihood formulation by solving the following optimization problem.

$$\arg\min_{B,\Omega} \Big\{ \mathrm{tr}[\frac{1}{n}(Y - XB)^\tau(Y - XB)\Omega] - \log\det(\Omega) + \lambda_1 p_1(B) + \lambda_2 p_2(\Omega) \Big\}, \tag{4.4}$$

where $p_1()$ and $p_2()$ are two penalty functions with regularization parameters λ_1 and λ_2 respectively. Heretofore, the LASSO and adaptive LASSO penalties have been considered in the studies of Rothman et al. (2010), Yin and Li (2011) and Lee and Liu (2012).

However, the optimization problem (4.4) is not convex. Instead, it is convex for either B or Ω with the other fixed. Solving (4.4) with B fixed

at B^* yields the optimization over Ω as shown below:

$$\hat{\Omega}(B^*) = \arg\min_{\Omega} \Big\{ \operatorname{tr}[\hat{\Sigma}_n(B^*)\Omega] - \log\det(\Omega) + \lambda_2 p_2(\Omega) \Big\}, \tag{4.5}$$

where $\hat{\Sigma}_n(B^*) = \frac{1}{n}(Y - XB^*)^\tau(Y - XB^*)$. Likewise, solving (4.4) for B with Ω fixed at Ω^*, we have

$$\hat{B}(\Omega^*) = \arg\min_{B} \Big\{ \operatorname{tr}[\frac{1}{n}(Y - XB)^\tau(Y - XB)\Omega^*] + \lambda_1 p_1(B) \Big\}. \tag{4.6}$$

Therefore, an alternate updating scheme is employed to break the non-convex optimization problem (4.4) into a serial of convex optimization problems, where the estimates of B and Ω are updated alternatively by solving (4.6) and (4.5). This iterative procedure continues until the estimates converge. Nevertheless, the main drawback of this scheme is that the converged estimates are probably only the local optimizers of (4.4) rather than the global ones, because there are probably many stationary points in the high-dimensional space.

4.1.2 Conditional Regression Formulation

An alternative way to achieve the joint estimation of B and Ω is via the conditional regression formulation (Wang, 2013). Besides the marginal distributions in (4.2), the multivariate response regression model (4.1) also contains the information about these p response variables' conditional dependent structure, which is encoded in the precision matrix Ω. Denote the matrix that omitting the jth column of Y by Y_{-j}. For $j = 1, \ldots, p$, given

X and Y_{-j}, the conditional distribution of $\boldsymbol{Y}_j$ is given as follows:

$$\boldsymbol{Y}_j|(X, Y_{-j}) \sim N_n\Big(X\boldsymbol{\beta}_j + (Y_{-j} - XB_{-j})\boldsymbol{\theta}_j, \frac{1}{\omega_{jj}}\boldsymbol{I}_n\Big), \tag{4.7}$$

where B_{-j} represents the covariate coefficient matrix that deletes the jth column of B and $\boldsymbol{\theta}_j = (\theta_{1j}, \ldots, \theta_{j-1,j}, \theta_{j+1,j}, \ldots, \theta_{pj})^\tau \in \mathbf{R}^{p-1}$ is the response coefficient vector. Let $\Theta = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p)$ be the corresponding response coefficient matrix. This conditional distribution (4.7) can also be transformed into the conditional regression model for $\boldsymbol{Y}_j$. That is, for $j = 1, \ldots, p$,

$$\boldsymbol{Y}_j = X\boldsymbol{\beta}_j + (Y_{-j} - XB_{-j})\boldsymbol{\theta}_j + \boldsymbol{\epsilon}_j, \tag{4.8}$$

where $\boldsymbol{\epsilon}_j \in \mathbf{R}^n$ consists of n i.i.d. normal random variables with mean 0 and variance $\frac{1}{\omega_{jj}}$. Implied by the conditional regression (4.8), $\boldsymbol{Y}_j$ consists of three components: the part that can be explained by the covariates $X\boldsymbol{\beta}_j$, the part that attributes to the other centered responses $(Y_{-j} - XB_{-j})\boldsymbol{\theta}_j$ and the random error $\boldsymbol{\epsilon}_j$. Analogous to the standard Gaussian graphical model, it can be shown that $\theta_{kj} = -\frac{\omega_{kj}}{\omega_{jj}}$ for all $k \neq j$. Compared to the marginal regression of $\boldsymbol{Y}_j$ in (4.3), its conditional regression decomposes the random error $\boldsymbol{e}_j$ into two parts: $(Y_{-j} - XB_{-j})\boldsymbol{\theta}_j$ and $\boldsymbol{\epsilon}_j$, reducing the variation of the regression random error. Specifically, if the p responses are mutually independent or their pairwise correlations are negligible, suggesting that $\boldsymbol{\theta}_j$'s are $\mathbf{0}$, then $\boldsymbol{Y}_j$'s conditional regression can be simplified as its marginal regression. As these correlations increase, the part $(Y_{-j} - XB_{-j})\boldsymbol{\theta}_j$ accounts for more variation of $\boldsymbol{e}_j$.

Subsequently, to solve the original multivariate response regression model (4.1) is equivalent to fit these p conditional regressions in (4.8). Such con-

ditional regression formulation translates the joint estimation of B and Ω into that of the regression coefficients $\boldsymbol{\beta}_j$'s and $\boldsymbol{\theta}_j$'s by incorporating the responses' partial correlation into the response coefficient matrix Θ. Furthermore, considering the sparse property of B and Ω, penalties on $\boldsymbol{\beta}_j$ and $\boldsymbol{\theta}_j$ (e.g. adaptive Lasso in Wang, 2013) are appended to the least squares optimization to achieve the estimation parsimony, which can be expressed as

$$\underset{\boldsymbol{\beta}_j,\boldsymbol{\theta}_j}{\arg\min} \left\{ \|\boldsymbol{Y}_j - X\boldsymbol{\beta}_j - (Y_{-j} - XB_{-j})\boldsymbol{\theta}_j\|_2^2 + \lambda_1 p_1(\boldsymbol{\beta}_j) + \lambda_2 p_2(\boldsymbol{\theta}_j) \right\}. \quad (4.9)$$

In contrast to (4.4), the above optimization problem is convex. However, the main challenge arises when one tries to solve $\boldsymbol{\beta}_j$ and $\boldsymbol{\theta}_j$ directly due to their interdependence in the term $XB_{-j}\boldsymbol{\theta}_j$.

4.2 Sequential Methods in Conditional Regression Formulation

We solve the multivariate response regression problem with the foregoing conditional regression formulation, where sequential methods are applied to solve the p conditional regressions in (4.8) for the covariate coefficients $\boldsymbol{\beta}_j$'s, the response coefficients $\boldsymbol{\theta}_j$'s and the random error variances $\frac{1}{\omega_{jj}}$'s. To handle the complicity caused by $XB_{-j}\boldsymbol{\theta}_j$, we first consider to employ an alternate updating scheme, similar to that in the penalized likelihood formulation, to estimate the covariate coefficients and response coefficients separately. Moreover, a simultaneous approach is considered to estimate $\boldsymbol{\beta}_j$ and $\boldsymbol{\theta}_j$ synchronously without any outer iteration for each j,

provided a proper initialization of B.

4.2.1 Alternate Updating Approach

Analogous to the penalized likelihood formulation, we restate the conditional regression formulation in the alternate updating framework. When B is fixed at $B^* = (\boldsymbol{\beta}_1^*, \ldots, \boldsymbol{\beta}_p^*)$, subtract XB^* from Y and denote the consequent centered response matrix by $Y^c(B^*) = (Y_1^c(\boldsymbol{\beta}_1^*), \ldots, Y_p^c(\boldsymbol{\beta}_p^*))$. Each row vector follows a multivariate normal distribution with the same covariance matrix but zero mean vector, i.e. $N_p(\mathbf{0}, \Sigma)$. Therefore, the conditional regression in (4.8) can be rearranged as

$$\boldsymbol{Y}_j^c(\boldsymbol{\beta}_j^*) = Y_{-j}^c(B_{-j}^*)\boldsymbol{\theta}_j + \boldsymbol{\epsilon}_j, \qquad \boldsymbol{\epsilon}_j \sim N_n(0, \frac{1}{\omega_{jj}}\boldsymbol{I}_n). \tag{4.10}$$

This implies a standard Gaussian graphical model with the same precision matrix Ω, where n samples of the jth vertex are contained in $\boldsymbol{Y}_j^c(\boldsymbol{\beta}_j^*)$. Consequently, existing methods can be exploited to estimate the response coefficients $\boldsymbol{\theta}_j$'s in (4.10) and ultimately the precision matrix Ω. Assuming the true graph is high-dimensional and sparse, we use the SSPS method proposed in Section 2.2.2 for the edge detection and estimate Ω with the constrained MLE as described in Section 3.1.2.

Furthermore, suppose the response variables' partial correlations are encoded in $\Theta^* = (\boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_p^*)$. For the jth response, given the other responses' covariate coefficient vectors $\{\boldsymbol{\beta}_k^*\}_{k \neq j}$, we are able to detach their influence from $\boldsymbol{Y}_j$. Denote the resulting shifted jth response as $\boldsymbol{Y}_j^s(B_{-j}^*, \boldsymbol{\theta}_j^*) = \boldsymbol{Y}_j - (Y_{-j} - XB_{-j}^*)\boldsymbol{\theta}_j^*$. Then the conditional regression (4.8) becomes a

simple univariate response regression

$$\boldsymbol{Y}_j^s(B_{-j}^*, \boldsymbol{\theta}_j^*) = X\boldsymbol{\beta}_j + \boldsymbol{\epsilon}_j, \qquad \boldsymbol{\epsilon}_j \sim N_n(\mathbf{0}, \frac{1}{\omega_{jj}}\boldsymbol{I}_n). \tag{4.11}$$

As a result, $\boldsymbol{\beta}_j$ can be solved by the prevalent coefficients estimation methods for the linear regression models. In particular, we apply the SLasso method (Luo and Chen, 2014b, see Section 2.2.1) to fit the sparse $\boldsymbol{\beta}_j$ in (4.11). Note that solving $\boldsymbol{\beta}_j$ requires not only the $\boldsymbol{\theta}_j^*$ but also the B_{-j}^*. In an iteration procedure, all the $\boldsymbol{\beta}_j$'s can be estimated in sequence on condition of a consistent initialization of B.

To sum up, we employ the alternate updating approach to solve the sparse B and Θ with the aforementioned sequential methods. Usually, the methods with iterative algorithms choose to stop the iteration when the estimation converges. However, the sequential methods are designed exclusively for feature selection, thereby we terminate the iteration when the relevant feature selection is stable. Let A be the matrix that is of our interest, i.e. $A = B$, or Θ, where its (i, j)th entry is a_{ij}. Denote its estimates in two consecutive steps by $\hat{A}^{(k-1)}$ and $\hat{A}^{(k)}$, $k \geq 1$. Define

$$r_1(\hat{A}) = \frac{|\{(i,j) : \hat{a}_{ij}^{(k-1)} \neq 0 \text{ and } \hat{a}_{ij}^{(k)} \neq 0\}|}{|\{(i,j) : \hat{a}_{ij}^{(k-1)} \neq 0\}|}$$

and

$$r_2(\hat{A}) = \frac{|\{(i,j) : \hat{a}_{ij}^{(k-1)} \neq 0 \text{ and } \hat{a}_{ij}^{(k)} \neq 0\}|}{|\{(i,j) : \hat{a}_{ij}^{(k)} \neq 0\}|}.$$

These two ratios together depict the similarity between $\hat{A}^{(k-1)}$'s and $\hat{A}^{(k)}$'s feature selection. If both $r_1(\hat{A})$ and $r_2(\hat{A})$ are larger than a threshold ratio r (e.g. 0.95, 0.99), then we regard the current estimate $\hat{A}^{(k)}$ as feature

selection stable.

In what follows, we elaborate this alternate updating approach by considering the multivariate response regression model in three high-dimensional scenarios, which are determined by the comparative relationships between the sample size (n) and the numbers of responses and covariates (p and q). The three scenarios are: (1) p is small, but q is large; (2) p is large, but q is small; (3) both p and q are large.

Scenario 1 In this scenario, we consider that the model in (4.1) has a huge number of covariates but a few responses. In the presence of substantial covariates, it is believed that each response variable only relies on a small number of covariates. Therefore, the SLasso method can be employed to achieve the sparsity of $\boldsymbol{\beta}_j$'s estimation. In contrast, due to Θ's low dimensionality, sparsity may be irrelevant or not important for its estimation, and it can be easily estimated from the inverse of the responses' sample covariance, according to the relation between Θ and Ω. Rather as the ultimate objective, the estimation of Θ is more like a vehicle that incorporates the responses' correlations and thus improves the estimation accuracy of B. B and Θ are estimated alternatively until the feature selection in B's estimate is stable. More details are shown in Algorithm 1.

Algorithm 1

Initialization. $\hat{B}^{(0)}$.

Iteration. In the kth ($k \geq 1$) step, update the estimates of Θ and B alternatively.

- Update Θ: Compute the centered response $\boldsymbol{Y}^c(\hat{B}^{(k-1)})$ and the inverse of its sample covariance $\hat{\Sigma}_n(\hat{B}^{(k-1)})$, which is denoted by $\hat{\Omega}^{(k)}$ with its (i,j)th entry to be $\hat{\omega}_{ij}^{(k)}$. For $j = 1, \ldots, p$, obtain $\hat{\boldsymbol{\theta}}_j^{(k)}$ with components in $\{\hat{\theta}_{ij}^{(k)} = -\frac{\hat{\omega}_{ij}^{(k)}}{\hat{\omega}_{jj}^{(k)}} : i \neq j\}$. Let $\hat{\Theta}^{(k)} = (\hat{\boldsymbol{\theta}}_1^{(k)}, \ldots, \hat{\boldsymbol{\theta}}_p^{(k)})$.
- Update B: With $\hat{B}^{(k-1)}$ and the updated $\hat{\Theta}^{(k)}$, solve the $\boldsymbol{\beta}_j$ in (4.11) by SLasso for $j = 1, \ldots, p$ in turn. Specifically, compute $\boldsymbol{Y}_1^s(\hat{B}^{(k-1)}, \hat{\boldsymbol{\theta}}_1^{(k)})$ and solve $\boldsymbol{\beta}_1$ with SLasso. Denote the solution by $\hat{\boldsymbol{\beta}}_1^{(k)}$ and update the 1st column of $\hat{B}^{(k-1)}$ with $\hat{\boldsymbol{\beta}}_1^{(k)}$. Analogously, for $j = 2, \ldots, p$, applying SLasso with $\hat{\boldsymbol{\theta}}_j^{(k)}$ and the partially updated $\hat{B}^{(k-1)}$, we can obtain $\hat{\boldsymbol{\beta}}_2^{(k)}, \ldots, \hat{\boldsymbol{\beta}}_p^{(k)}$ successively; meanwhile, $\hat{B}^{(k-1)}$ is updated column by column and eventually evolves into $\hat{B}^{(k)}$.
- Stop: Compute $r_1(\hat{B})$ and $r_2(\hat{B})$. If $r_1(\hat{B}) > r_b$ and $r_2(\hat{B}) > r_b$, stop and estimate B as $\hat{B} = \hat{B}^{(k)}$; otherwise, continue.

In the above algorithm, $\hat{B}^{(0)}$ can be any consistent estimate of B. We set $\hat{B}^{(0)}$ to be composed of the SLasso coefficients estimates of the marginal regressions in (4.3). Such initialization results from the naive assumption that all the response variables are independent. Additionally, r_b is a specified threshold ratio that determines B's estimation stability.

Scenario 2 Opposite to Scenario 1, now the response $\boldsymbol{Y}$ is high-dimensional while the covariate $\boldsymbol{X}$ is of low dimension. In this case, it is reasonable to suppose the sparsity in Ω. Even if the number of covariates is substantial, some dimension reduction methods, such as PCA, can be exploited to pro-

duce a low-dimensional explanatory data X. Hence, there is no need to assume sparsity in B, it is sufficient to estimate the $\boldsymbol{\beta}_j$ in (4.11) by the ordinary least squares (OLS). However, as the dimension of Y increases, the sample covariance tends to be singular and cannot serve as a suitable approach to estimate Ω. Obviously, the edge detection and the estimation of Ω are of our concern. We can take advantage of B's estimate to center the response matrix Y. Therefore, the $\boldsymbol{\theta}_j$'s in (4.10) can be analyzed by SSPS as in the standard Gaussian graphical models. In conclusion, we fit the conditional regressions in (4.8) by integrating the methods OLS and SSPS into the alternate updating scheme and terminate the iteration if the edge detection is stable. The corresponding algorithm is described as follows.

Algorithm 2

Initialization. $\hat{B}^{(0)}$ and $\hat{\Theta}^{(0)}$.

Iteration. In the kth ($k \geq 1$) step, update the estimates of B and Θ alternatively.

- Update B: For $j = 1, \ldots, p$, compute the shifted the response $\boldsymbol{Y}_j^s(\hat{B}^{(k-1)}, \hat{\boldsymbol{\theta}}_j^{(k-1)})$ and solve the $\boldsymbol{\beta}_j$ in (4.11) with OLS. Hence, $\hat{\boldsymbol{\beta}}_j^{(k)} = (X^\tau X)^{-1} X^\tau \boldsymbol{Y}_j^s(\hat{B}^{(k-1)}, \hat{\boldsymbol{\theta}}_j^{(k-1)})$. $\hat{B}^{(k)}$ is obtained by updating the corresponding columns in $\hat{B}^{(k-1)}$ with the $\hat{\boldsymbol{\beta}}_j^{(k)}$'s in sequence, analogous to that in Algorithm 1.
- Update Θ: With B's current estimate $\hat{B}^{(k)}$, compute the centered response $Y^c(\hat{B}^{(k)})$. Apply the SSPS method to the Gaussian graphical model implied by (4.10), detect the nonzero elements

in $\boldsymbol{\theta}_j$'s and estimate them with OLS. Denote by $\hat{\boldsymbol{\theta}}_j^{(k)}$ the current estimate of $\boldsymbol{\theta}_j$ for $j = 1, \ldots, p$, thereby $\hat{\Theta}^{(k)} = (\hat{\boldsymbol{\theta}}_1^{(k)}, \ldots, \hat{\boldsymbol{\theta}}_p^{(k)})$.

- Stop: Compute $r_1(\hat{\Theta})$ and $r_2(\hat{\Theta})$. If $r_1(\hat{\Theta}) > r_\theta$ and $r_2(\hat{\Theta}) > r_\theta$, stop and estimate Θ and B as $\hat{\Theta} = \hat{\Theta}^{(k)}$ and $\hat{B} = \hat{B}^{(k)}$; otherwise, continue.

In Algorithm 2, $\hat{B}^{(0)} = 0 \cdot \boldsymbol{I}_{q\times p}$ and $\hat{\Theta}^{(0)}$ is the SSPS estimate of (4.10) with all the $\boldsymbol{\beta}_j$'s to be $\mathbf{0}$. Likewise, r_θ is a predetermined threshold to evaluate the selection stability in $\hat{\Theta}^{(k)}$'s. Nevertheless, the above algorithm does not offer an explicit estimation of the precision matrix Ω. We use the constrained MLE method as shown in Section 3.1.2 to derive Ω's estimate. The process is briefly described here. Denote the sample covariance of the estimated centered response $Y^c(\hat{B})$ by $\hat{\Sigma}_n(\hat{B})$. Let $\Gamma_0 = \{(i,j) : \hat{\theta}_{ij} = 0\}$, which is the index set of the estimated zero entries in Θ as well as Ω from Algorithm 2. We estimate Ω from the following optimization problem:

$$\hat{\Omega} = \underset{\substack{\Omega_{\Gamma_0}=0 \\ \Omega\succ 0}}{\arg\min} \Big\{ \operatorname{tr}[\hat{\Sigma}_n(\hat{B})\Omega] - \log\det(\Omega) \Big\}, \tag{4.12}$$

where $\Omega_{\Gamma_0} = \{\omega_{ij} : (i,j) \in \Gamma_0\}$.

Scenario 3 In the last scenario, there are numerous responses and covariates, which is an extension of both Scenario 1 and Scenario 2. Consequently, the covariate selection and covariance selection are indispensable. This is the common situation considered by the relevant research on this topic. Accordingly, SLasso and SSPS are employed to impose sparsity on the estimates of the coefficient matrix B and the precision matrix Ω respectively.

We adapt these two sequential methods into the alternate updating framework to achieve the joint sparse estimation of the multivariate response regression model in (4.1). For the stopping rule, it is essential to check the stability in the estimates of Θ and B synchronously. The corresponding algorithm is summarized as follows.

Algorithm 3

Initialization. $\hat{B}^{(0)}$ and $\hat{\Theta}^{(0)}$.

Iteration. In the kth ($k \geq 1$) step, update the estimates of B and Θ alternatively.

- Update Θ: Compute the centered response $Y^c(\hat{B}^{(k-1)})$. Same as in Algorithm 2, employ SSPS to identify the nonzero θ_{ij}'s in (4.10) and estimate them with OLS. Denote the current estimate of Θ by $\hat{\Theta}^{(k)}$ with its jth column to be $\hat{\boldsymbol{\theta}}_j^{(k)}$.
- Update B: Same as in Algorithm 1, apply SLasso to solve the $\boldsymbol{\beta}_j$ in (4.11) and update the jth column in $\hat{B}^{(k-1)}$ accordingly, for $j = 1, \ldots, p$. Denote the current estimate by $\hat{B}^{(k)}$.
- Stop: Compute $r_1(\hat{B})$, $r_2(\hat{B})$, $r_1(\hat{\Theta})$ and $r_2(\hat{\Theta})$. If $r_1(\hat{B})$ and $r_2(\hat{B}) > r_b$, $r_1(\hat{\Theta})$ and $r_2(\hat{\Theta}) > r_\theta$ are satisfied simultaneously, stop and estimate B and Θ as $\hat{B} = \hat{B}^{(k)}$ and $\hat{\Theta} = \hat{\Theta}^{(k)}$; otherwise, continue.

With the initial assumption of independent responses, $\hat{\Theta}^{(0)} = 0 \cdot \boldsymbol{I}_{(p-1)\times p}$ and $\hat{B}^{(0)}$ is the same as that in Algorithm 1, consisting of the SLasso estimates of the marginal regressions. Furthermore, the precision matrix Ω is estimated by the constrained MLE described in Scenario 2.

4.2.2 Simultaneous Estimation Approach

Suppose that the multivariate response regression model (4.1) is in the context of Scenario 3, i.e., both responses and covariates are in high dimension. Following the conditional regression formulation, we explore the simultaneous sequential estimation of the $\boldsymbol{\beta}_j$ and $\boldsymbol{\theta}_j$ in (4.8) without the alternate updating scheme, for $j = 1, \ldots, p$.

Inspired by the study in Wang (2013), we propose to solve the penalized least squares optimization problem in (4.9) with the SLasso framework, where we let $p_1(\boldsymbol{\beta}_j) = \|\boldsymbol{\beta}_j\|_1$ and $p_2(\boldsymbol{\theta}_j) = \|\boldsymbol{\theta}_j\|_1$. However, as mentioned before, the interaction of $\boldsymbol{\beta}_i$ and $\boldsymbol{\theta}_j$ $(i \neq j)$ in the term $X_{-j}B_{-j}\boldsymbol{\theta}_j$ complicates the implementation. To simplify the optimization problem (4.9), we first substitute the parameter b_{ij} $(i \neq j)$ in $X_{-j}B_{-j}\boldsymbol{\theta}_j$ with any of its consistent estimate and thereby remove the interdependence between the coefficients. Subsequently, we have the following adjustment. Denote a consistent initial estimate of B by $\hat{B}^0$, let $Z = Y - X\hat{B}^0$. Then the conditional regression in (4.8) can be rewritten as

$$\boldsymbol{Y}_j = X\boldsymbol{\beta}_j + Z_{-j}\boldsymbol{\theta}_j + \boldsymbol{\epsilon}_j, \qquad \boldsymbol{\epsilon}_j \sim N_n(\mathbf{0}, \frac{1}{\omega_{jj}}\boldsymbol{I}_n). \tag{4.13}$$

Its ℓ_1 penalized least square optimization is

$$\underset{\boldsymbol{\beta}_j, \boldsymbol{\theta}_j}{\arg\min} \Big\{ \|\boldsymbol{Y}_j - X\boldsymbol{\beta}_j - Z_{-j}\boldsymbol{\theta}_j\|_2^2 + \|\boldsymbol{\beta}_j\|_1 + \|\boldsymbol{\theta}_j\|_1 \Big\}. \tag{4.14}$$

This regression in (4.13) reminds us the interactive linear models with main-effect features and interaction features. However, here we do not dis-

criminate between them; instead, we simply recognize that there are two groups of features, included in X and Z_{-j} respectively. Their coefficients $\boldsymbol{\beta}_j$ and $\boldsymbol{\theta}_j$ in (4.14) are solved by a sequential procedure, which is essentially a modified version of the SLasso method.

Without loss of generality, we assume that Y, X and Z are normalized such that each column has mean 0 and standard deviation 1. Denote by $\boldsymbol{Z}_{ji}$ the column in Z_{-j} that corresponds to the coefficient θ_{ij}. Let $S_{\beta_j} = \{1, \ldots, q\}$ and $S_{\theta_j} = \{1, \ldots, p\} \setminus \{j\}$, thus S_{β_j} and S_{θ_j} are the index sets for features in X and Z_{-j}. Let $S_j = S_{\beta_j} \cup S_{\theta_j}$ denote the index set for all features. Suppose $s_{\beta_j} \subset S_{\beta_j}$, $s_{\theta_j} \subset S_{\theta_j}$ and $s_j \subset S_j$. Denote by $X(s)$ the submatrix of X consisting of columns with indices in the set s. Parallel to the EBIC for the common linear regression models, the EBIC for the regression in (4.13) is derived as below.

$$\begin{aligned}\mathrm{EBIC}_{\gamma_\beta,\gamma_\theta}(s_j) =& n \ln\left(\frac{\|[I - H(s_j)]\boldsymbol{Y}_j\|_2^2}{n}\right) + \ln(n)(|s_{\beta_j}| + |s_{\theta_j}|) \\ &+ 2\gamma_\beta \ln\binom{q}{|s_{\beta_j}|} + 2\gamma_\theta \ln\binom{p-1}{|s_{\theta_j}|}, \end{aligned} \tag{4.15}$$

where $H(s_j)$ is the projection matrix of the matrix $M(s_j) = \Big(X(s_{\beta_j})Z(s_{\theta_j})\Big)$, i.e., $H(s_j) = M(s_j)[M^\tau(s_j)M(s_j)]^{-1}M^\tau(s_j)$. Let $\gamma_\beta > 1 - \ln n/(2\ln q)$ and $\gamma_\theta > 1 - \ln n/(2\ln(p-1))$. In every step of the sequential procedure, we select one candidate feature from each group and compute the resulting model's EBIC. The one with smaller EBIC will be eventually selected. This procedure goes on until the selected model's EBIC starts to increase. Therefore, the algorithm for the simultaneous estimation of the conditional regressions in (4.8) goes as below.

Algorithm 4

Step 1. Initialize $\hat{B}^0$ and obtain $Z = Y - X\hat{B}^0$.

Step 2. For each $j \in \{1, \ldots, p\}$, solve the $\boldsymbol{\beta}_j$ and $\boldsymbol{\theta}_j$ in (4.14) by the following sequential procedure.

Initialization. Let $s_{\beta_j} = s_{\theta_j} = \emptyset$, $s_j = s_{\beta_j} \cup s_{\theta_j}$ and $\tilde{\boldsymbol{Y}}_j = \boldsymbol{Y}_j$.

Iteration.

- Compute $|\boldsymbol{X}_i^\tau \tilde{\boldsymbol{Y}}_j|$ for $i \in s_{\beta_j}^c$ and identify i^* such that

$$|\boldsymbol{X}_{i^*}^\tau \tilde{\boldsymbol{Y}}_j| = \max_{i \in s_{\beta_j}^c} |\boldsymbol{X}_i^\tau \tilde{\boldsymbol{Y}}_j|.$$

  Let $s_{\beta_j}^* = s_{\beta_j} \cup \{i^*\}$ and $s_{+\beta_j}^* = s_{\beta_j}^* \cup s_{\theta_j}$.

- Compute $|\boldsymbol{Z}_{jk}^\tau \tilde{\boldsymbol{Y}}_j|$ for $k \in s_{\theta_j}^c$ and identify k^* such that

$$|\boldsymbol{Z}_{jk^*}^\tau \tilde{\boldsymbol{Y}}_j| = \max_{k \in s_{\theta_j}^c} |\boldsymbol{Z}_{jk}^\tau \tilde{\boldsymbol{Y}}_j|.$$

  Let $s_{\theta_j}^* = s_{\theta_j} \cup \{k^*\}$ and $s_{+\theta_j}^* = s_{\beta_j} \cup s_{\theta_j}^*$.

- If $\text{EBIC}_{\gamma_\beta,\gamma_\theta}(s_{+\beta_j}^*) < \text{EBIC}_{\gamma_\beta,\gamma_\theta}(s_{+\theta_j}^*)$, let $s_j^* = s_{+\beta_j}^*$; otherwise, $s_j^* = s_{+\theta_j}^*$.

- If $\text{EBIC}_{\gamma_\beta,\gamma_\theta}(s_j^*) < \text{EBIC}_{\gamma_\beta,\gamma_\theta}(s_j)$, let $s_j = s_j^*$ and $\tilde{\boldsymbol{Y}}_j = [\boldsymbol{I} - H(s_j)]\boldsymbol{Y}_j$; otherwise, go to the Output.

Output. s_j is the index set of selected features, where $s_j = s_{\beta_j} \cup s_{\theta_j}$. Estimate the nonzero components in $\boldsymbol{\beta}_j$ and $\boldsymbol{\theta}_j$ by OLS. That is, $(\hat{\boldsymbol{\beta}}_j^\tau(s_{\beta_j}), \hat{\boldsymbol{\theta}}_j^\tau(s_{\theta_j}))^\tau = [M^\tau(s_j)M(s_j)]^{-1}M^\tau(s_j)\boldsymbol{Y}_j$. Estimate the random error's variance $\frac{1}{\omega_{jj}}$ by the selected model's mean residual sum of squares (RSS), thus $\hat{\omega}_{jj} = n/\|[\boldsymbol{I} - H(s_j)]\boldsymbol{Y}_j\|_2^2$.

In the above algorithm, we initialize $\hat{B}^0$ with the SLasso solutions of the marginal regressions in (4.3), same as those in Algorithm 1 and Algorithm 3. Similar to the alternate updating approach, solving the parameters of the p conditional regressions alone does not directly lead to the estimate of Ω. As described in Section 3.1.1, a simple way to form Ω's estimate from $\hat{\Theta}$ and $\hat{\omega}_{jj}$'s is based on the relation between Ω and Θ. For $i \neq j$, let $\tilde{\omega}_{ij} = -\hat{\theta}_{ij}\hat{\omega}_{jj}$. Since Ω is symmetric, the final estimation is $\hat{\Omega}$ with its (i,j)th entry to be $\hat{\omega}_{ij} = (\tilde{\omega}_{ij} + \tilde{\omega}_{ji})/2$. Alternatively, Ω can be estimated by the constrained MLE adopted by the alternate updating approach.

4.3 Selection Consistency

In this section, we consider the selection consistency of the two SCR methods, which is expected to closely associate with that of the SLasso and SSPS. In fact, the consistency of the simultaneous estimation approach is implied by that of the SLasso with some adjustments for the two groups of features. In what follows, we discuss the selection consistency of the alternate updating approach.

Suppose X is a deterministic design matrix and its columns are standardized. Let $s_{\beta_{0j}} = \{i : \beta_{ij} \neq 0\}$, $q_{0j} = |s_{\beta_{0j}}|$, $s_{\theta_{0j}} = \{i : \theta_{ij} \neq 0\}$, $p_{0j} = |s_{\theta_{0j}}|$, for $1 \leq j \leq p$. Let $S_{B_0} = \{(i,j) : \beta_{ij} \neq 0\}$, s_B be its subset and $s_{\beta_j} = \{i : (i,j) \in s_B\}$. Similarly define S_{Θ_0}, s_Θ and s_{θ_j}. Obviously, $s_{\beta_j} \subset s_{\beta_{0j}}$ and $s_{\theta_j} \subset s_{\theta_{0j}}$. Follow the notations of S_{β_j} and S_{θ_j} defined in

Section 4.2.2. Define

$$\phi_{jk} = \frac{1}{n}\boldsymbol{X}_k^\tau[\boldsymbol{I}_n - H(s_{\beta_j})]X\boldsymbol{\beta}_j,$$

and for $k \neq j$,

$$\tau_{jk} = \frac{(\Sigma_{kS_{\theta_j}} - \Sigma_{ks_{\theta_j}}\Sigma^{-1}_{s_{\theta_j}s_{\theta_j}}\Sigma_{s_{\theta_j}S_{\theta_j}})\boldsymbol{\theta}_j}{\sqrt{\sigma^{jj}\sigma_k^2}},$$

where $\boldsymbol{\beta}_j$ and $\boldsymbol{\theta}_j$ are the two coefficient vectors in (4.8), $H(s_{\beta_j})$ is the projection matrix of $X(s_{\beta_j})$, $\sigma^{jj} = \sigma_j^2 - \Sigma_{jS_{\theta_j}}\Sigma^{-1}_{S_{\theta_j}S_{\theta_j}}\Sigma_{S_{\theta_j}j}$ and σ_k^2 is the kth diagonal entry of Σ. Let $v_{jk} = (\tau_{jk}, \tau_{kj})^\tau$. We assume the following conditions:

C1 $\ln q = O(n^{\kappa_1})$, where $\kappa_1 \in (0, 1/2)$. $\max_j q_{0j} = O(n^{\delta_1})$, for some $\delta_1 \in (0, 1)$.

C2 $\ln p = O(n^{\kappa_2})$, where $\kappa_2 \in (0, 1/3)$. $\max_j p_{0j} = O(n^{\delta_2})$, for some $\delta_2 \in (\kappa_2/2, 1/6)$.

C3 For any $s_{\beta_j} \subset s_{\beta_{0j}}$ but $s_{\beta_j} \neq s_{\beta_{0j}}$,

$$\max_{k \in s_{\beta_{0j}} \setminus s_{\beta_j}} |\phi_{jk}| > \max_{k \notin s_{\beta_{0j}}} |\phi_{jk}|$$

holds uniformly for $1 \leq j \leq p$.

C4 For any $s_\Theta \subset S_{\Theta_0}$ but $s_\Theta \neq S_{\Theta_0}$,

$$\max_{(j,k) \in S_{\Theta_0} \setminus s_\Theta} \|v_{jk}\|_2^2 > \max_{(j,k) \notin S_{\Theta_0}} \|v_{jk}\|_2^2.$$

C5 $\lim_{n\to\infty} \min_{1\leq j\leq p} \left\{ \frac{\sqrt{n}}{\ln q} \lambda_{min}[\frac{1}{n}X^\tau(s_{\beta_{0j}})X(s_{\beta_{0j}})] \min_{k \in s_{\beta_{0j}}} |\beta_{kj}| \right\} = +\infty.$

C6 $\lim_{n\to\infty} \min_{1\le j\le p} \left\{ \frac{\sqrt{n}}{\ln p} \lambda_{min}(\Sigma_{s_{\beta_{0j}} s_{\beta_{0j}}}) \min_{k\in s_{\beta_{0j}}} |\frac{\theta_{kj}}{\sqrt{\sigma^{jj}}}| \right\} = +\infty.$

Let $\hat{B}$ and $\hat{\Theta}$ be the estimated covariate coefficient matrix and precision matrix obtained by the SCR method with the alternate updating scheme. Define $\hat{S}_B = \{(i,j) : \hat{\beta}_{ij} \neq 0\}$, which is the index set of the nonzero entries in $\hat{B}$. Similarly define $\hat{S}_\Theta = \{(i,j) : \hat{\theta}_{ij} \neq 0\}$. The selection consistency of the alternate updating SCR method is stated in the following theorem.

Theorem 4.1. Assume conditions C1 - C6. Moreover, suppose that the correlations between the components of $\boldsymbol{Y}$ are bounded by a constant less than 1, the variances of the components of $\boldsymbol{Y}$ are bounded and that the regularization parameter γ in the EBIC functions of SLasso and SSPS is taken to be larger than $1 - \ln n/(2\ln q)$ and $1 - \ln n/(2\ln p)$ respectively. Then the alternate updating SCR method is selection consistent, i.e.,

$$P(\hat{S}_B = S_{B_0}, \hat{S}_\Theta = S_{\Theta_0}) \to 1, \text{ as } n \to +\infty.$$

Conditions C1, C3 and C5 ensure the SLasso selection consistency in the estimation of B. Conditions C2, C4 and C6 guarantee the edge detection consistency of the SSPS in the estimation of Θ. Provided that $\hat{B}$ is consistent, the consistency of $\hat{\Theta}$ is implied by that of SSPS. Conversely, if $\hat{\Theta}$ is consistent, then the consistency of $\hat{B}$ can be shown by the fact that each SLasso estimate $\hat{\boldsymbol{\beta}}_j$ is consistent for $1 \le j \le p$. Therefore, Theorem 4.1 can be established according to the selection consistency proofs of the SLasso and SSPS. For more details, refer to Luo and Chen (2014b) and Section 2.3. Here we only consider the case of deterministic design matrix. For the case of random design matrix, the consistency property can be shown similarly,

with some modifications in C1, C3 and C5 based on the Theorem 3.2 in Luo and Chen (2014b).

4.4 Simulation Study

In the simulation study, we compare the methods for the multivariate response regression models in the three foregoing scenarios separately. In Scenario 1, the sequential conditional regression method in Algorithm 1 (SCR_1) is compared with SLasso [Luo and Chen (2014b)], which follows the naive approach to estimate the covariate coefficient matrix only based on the marginal regressions in (4.3). Similarly, in Scenario 2, we compare our method in Algorithm 2 (SCR_2) with SSPS [Section 2.2.2] on the precision matrix estimation, where SSPS only relies on the response data without adjusting the covariates' effect on the mean. In Scenario 3, the alternate updating method in Algorithm 3 (SCR_3) and the simultaneous estimation method in Algorithm 4 (SCR_4) are compared to MRCE [Rothman et al. (2010)] and aMCR [Wang (2013)]. The precision matrix estimation of the simultaneous estimation approach is obtained by the constrained MLE, same as that of the alternate updating approach. Moreover, for each of the three methods with alternate updating scheme, its counterpart without iteration is also considered as a computational approximation, denoted as SCR_i^*, for $i = 1, 2, 3$. Since the settings vary among these three scenarios, their results are shown respectively.

4.4.1 Scenario 1

In this scenario, the estimates from SLasso, SCR_1 and SCR_1^* are compared on their accuracy of covariate selection and prediction. The simulation settings are adapted from those in Luo and Chen (2014b) and incorporate different correlation designs of the response variables. The details go as below. The covariates in $\boldsymbol{X}$ are multivariate normal with mean $\mathbf{0}$ and covariance Σ_X in three forms:

- I: $\Sigma_X = \boldsymbol{I}_q$, i.e., the covariates are i.i.d. $N(0, 1)$;
- PD: $\Sigma_X = (0.5^{|i-j|})_{q\times q}$, i.e., the covariates have power decay correlations;
- Eq: Σ_X's off-diagonal entries are 0.5 and its diagonals are 1, i.e., the covariates are equally correlated.

For each response, we randomly select q_0 covariates to be its true features, where $q_0 = [4n^{0.16}]$. If Σ_X is of type PD, then the q_0 true features are consecutive. Let u follow Bernoulli(0.4), z be a normal random variable with mean 0 and satisfying that $P(|z| \geq 0.1) = 0.25$. The true features' coefficients are independently generated from the random variable $(-1)^u(4n^{-0.15} + |z|)$, such that each of their absolute values has approximate order $O(n^{-0.15})$; the rest coefficients are 0. Moreover, the random error matrix E consists of n i.i.d. samples from $N_p(\mathbf{0}, \Sigma)$, where Σ is determined by the following process. First, its jth diagonal entry is given by

$$\sigma_j^2 = \frac{1-h}{h}\boldsymbol{\beta}_j^\tau \Sigma_X \boldsymbol{\beta}_j, j = 1, \ldots, p,$$

where h is analogous to the R^2 of the marginal regressions in (4.3). Let $D = \text{diag}(\sigma_1, \ldots, \sigma_p)$. Then, four kinds of correlation matrix $\text{Cor}(E) = (\rho_{ij})_{p\times p}$ are considered, which are

- I: $\text{Cor}(E) = \boldsymbol{I}_p$, i.e., the random errors are independent;
- Band: $\rho_{i,i+1} = \rho_{i+1,i} = 0.5$, i.e., only two neighboring random errors are correlated;
- PD: $\rho_{ij} = 0.8^{|i-j|}$, i.e., the random errors have power decay correlations;
- Eq: $\rho_{ij} = 0.5$, $i \neq j$, i.e., the random errors are equally correlated.

Therefore, $\Sigma = D\,\text{Cor}(E)D$.

We evaluate the two methods in terms of PDR, FDR of the estimated coefficient matrix and prediction mean square error (PMSE). Their definitions are given as follows.

$$\text{PDR} = \frac{|\{(i,j) : b_{ij} \neq 0 \text{ and } \hat{\beta}_{ij} \neq 0\}|}{|\{(i,j) : b_{ij} \neq 0\}|}, \quad \text{FDR} = \frac{|\{(i,j) : b_{ij} = 0 \text{ and } \hat{\beta}_{ij} \neq 0\}|}{|\{(i,j) : \hat{\beta}_{ij} \neq 0\}|},$$

$$\text{PMSE} = \frac{1}{n}\|Y - X\hat{B}\|_F^2,$$

where $\hat{B} = (\hat{\beta}_{ij})_{q\times p}$ is an estimate of B.

We fix the sample size n at 100 and consider the settings with $q = [5e^{n^{0.3}}]$ and 500, $p = 10$ and 20, $h = 0.75$ and 0.8. The above procedure is replicated 200 times. Due to the calculation of PMSE, it is necessary to generate two i.i.d. samples of Y in each replication, where one sample is employed

to estimate B and the other is used to test the prediction accuracy. The results are shown in Table 4.1 to Table 4.4.

We first examine the trend of each method's estimation when one of p, q and h changes. When h increases from 0.75 to 0.8, the estimation of the same method apparently improves in the terms of PDR, FDR and PMSE. On the contrary, even though the number of true features remains the same, when q rises from $[5e^{n^{0.3}}]$ (i.e. 267) to 500, all the methods' performance deteriorates with decreasing PDRs and inflated PMSEs. On the other hand, their FDRs are not significantly influenced by the growth of irrelevant features. Additionally, each method's estimation is almost insusceptible to the augmentation of p. There is no obvious change in PDR and FDR, and the increase in PMSE is basically proportional to that in p.

Now we compare among these methods. Generally, SLasso performs slightly better than the other two SCR_1 methods when $\text{Cor}(E) = \boldsymbol{I}_p$, which suggests that all the p marginal regressions in (4.3) are actually independent. SLasso assumes this independence by default; whereas the two SCR_1 methods use the sample covariance to explore this relation, which inevitably induces some estimation errors. When $\text{Cor}(E)$ is not the identity matrix, SLasso still performs as if the responses were independent. In the meanwhile, the two SCR_1 methods' estimates apparently benefit from the incorporation of the responses' correlations. Notwithstanding their moderately higher FDRs in most cases, this disadvantage can be completely compensated by the larger improvements in terms of PDR and PMSE. As Σ_X and $\text{Cor}(E)$ become more complicated, their superiority in PDR and PMSE appears more conspicuous. Furthermore, between SCR_1 and SCR_1^*,

the former somewhat outperforms the latter, but their results are basically parallel. It suggests that SCR_1 achieves the estimation stability very fast and that SCR_1^* could be an efficient computational approximation of SCR_1.

4.4.2 Scenario 2

We now compare SCR_2 to SSPS on the edge detection and precision matrix estimation. The simulation examples synthesize the standard Gaussian graphical models and the design for the non-constant means with several covariates. First, the precision matrix Ω is generated from the eight graphs described in Section 2.4.1, which are AR(1), AR(2), Circle, Cluster, ER, Tridiag, Hub and BA. Here we exclude the graph RP due to the observed difficulty in its edge detection. Parallel to the settings in Section 2.4.1, two specific dimensions are considered, i.e. $p = 50$ and 200. When $p = 50$, Ω exactly follows the graph's definition. When $p = 200$, two precision matrix designs are considered. In the first design, Ω is a block diagonal matrix with four identical component matrices, where each of them is the precision matrix when $p = 50$. Consequently, the graph consists of four exactly the same clusters. In the second design, Ω is also a block diagonal matrix, consisting of two main diagonal blocks of size 50×50 and 150×150. The former is the same as the precision matrix when $p = 50$; the latter is the identity matrix. Such design is equivalent to add 150 independent noises to the 50 conditionally dependent variables. Let $\Sigma = \Omega^{-1}$. Then the row vectors in the error matrix E are i.i.d. $N_p(\mathbf{0}, \Sigma)$. Moreover, $\boldsymbol{X}$ is also multivariate normal with mean $\mathbf{0}$ and all the components mutually independent, i.e., $\Sigma_X = \boldsymbol{I}_q$. The coefficients in B are first generated in the same way as

the nonezero coefficients in Scenario 1, denoted by $\{\boldsymbol{\beta}_j^0\}_{j=1}^p$. Then each $\boldsymbol{\beta}_j^0$ is scaled to $\boldsymbol{\beta}_j$ such that $\frac{\boldsymbol{\beta}_j^\tau \Sigma_X \boldsymbol{\beta}_j}{\boldsymbol{\beta}_j^\tau \Sigma_X \boldsymbol{\beta}_j + \sigma_j^2}$ is equal to a fixed ratio h. That is, for $1 \leq j \leq p$, $\boldsymbol{\beta}_j = \alpha_j \boldsymbol{\beta}_j^0$, where α_j is a scaler given by the following equation:

$$\alpha_j = \sqrt{\frac{h\sigma_j^2}{(1-h)\boldsymbol{\beta}_j^{0\tau} \Sigma_X \boldsymbol{\beta}_j^0}}.$$

Let $\hat{\Omega}$ be an estimate of Ω. Its PDR, FDR, number of detected edges and its difference with Ω in spectral norm ($\|\cdot\|_S$), matrix ℓ_1 norm (maximum column sum, $\|\cdot\|_{\ell_1}$) and Frobenius norm ($\|\cdot\|_F$), are examined. In the absence of ambiguity, we refer to the number of detected edges implied by $\hat{\Omega}$ as $|\hat{\Omega}|$ in the simulation results, which actually should be calculated by $|\hat{\Omega}|/2 - p$. Fixing $n = 100$, we consider the settings formed by the different combinations among the eight graphs, $q = 10, 20$, $h = 0.2, 0.5$ and $p = 50, 200$. We replicate the estimation of each setting 100 times. The average values of the above criteria are shown in Table 4.5 to Table 4.7 with their standard deviations in the parentheses.

The increase of q from 10 to 20 adversely affects SCR_2's estimation accuracy, leading to lower PDRs, higher FDRs and larger errors. This phenomenon is more conspicuous in graphs like ER and Hub. As q increases, the coefficients' scales become much smaller, which makes it more difficult for SCR_2 to estimate the responses' means from the covariates. Nevertheless, this change of q has very limited influence on SSPS and sometimes even results in somewhat better estimation. This interesting observation may attribute to SSPS's insensitivity to such change, since it never considers the covariates' influence on the means. As h increases from 0.2 to 0.5,

SSPS's performance declines dramatically. Its PDRs considerably decrease, the FDRs increase and even exceed the PDRs, and the three norms of the matrix losses also inflate. In contrast, SCR_2's performance remains quite stable despite the alteration of h.

Now we compare the two methods in the same settings. As can be seen, compared to SSPS, SCR_2 significantly enhances the edge detection's PDR and FDR. Even so, its edge detection is still unsatisfactory in the graph Cluster when $p = 200$. Generally, SCR_2 produces Ω estimate with smaller matrix losses. However, we also notice some breaches, where SCR_2's estimate has larger matrix losses in spectral norm and matrix ℓ_1 norm. In the block precision matrix design for $p = 200$, most of the SCR_2's FDRs stay at an obviously low level, e.g. in AR(1), AR(2), Circle and Tridiag; whereas, in the noise precision matrix design, SCR_2's FDRs are apparently larger than those in the previous design, although it still compares favorably with SSPS.

Although we took SCR_2's non-iterative approximation SCR_2^* into account in the simulation, we observed that their results are almost the same. In fact, SCR_2 usually stops at the first step, i.e., without any iteration. Hence, the results of SCR_2^* are omitted.

4.4.3 Scenario 3

In this scenario, both the coefficient matrix B and the precision matrix Ω are high-dimensional. We refer to the two methods MRCE and aMCR for comparison. The R package MRCE [Rothman (2013)] is directly used

for the implementation of MRCE. The data are simulated in a procedure similar to that in Scenario 2. For the error matrix E, only five graphs, namely AR(1), ER, Tridiag, Hub and BA, are considered in the generation of Ω. Note that $s_{\beta_{0j}}$ is the index set of the jth response's relevant covariates, where $s_{\beta_{0j}} \subset \{1, \ldots, q\}$, for $1 \leq j \leq p$. Let $(q, q_0) = ([5e^{n^{0.3}}], [4n^{0.16}])$. The design matrix X and $s_{\beta_{0j}}$'s are generated according to the following descriptions.

- Type I: $X_1, \ldots, X_q$ are i.i.d. $N(0, 1)$. Each $s_{\beta_{0j}}$ is a random sample of size q_0 from $\{1, \ldots, q\}$.

- Type II: $(X_1, \ldots, X_q)^\tau$ follows $N_q(\mathbf{0}, \Sigma_X)$, where Σ_X's (i, j)th entry is $0.5^{|i-j|}$ such that $\boldsymbol{X}$ has a power decay correlation structure. $s_{\beta_{0j}}$ consists of q_0 consecutive random numbers from $\{1, \ldots, q\}$.

- Type III: $X_1, , \ldots, X_{[\frac{q}{2}]}$ are i.i.d $N(0, 1)$ and the rest are power decay correlated as in Type II. $s_{\beta_{0j}}$ consists of $[\frac{q_0}{2}]$ random numbers from $\{1, \ldots, [\frac{q}{2}]\}$ and $q_0 - [\frac{q_0}{2}]$ consecutive random numbers from $\{[\frac{q}{2}] + 1, \ldots, q\}$.

- Type IV: $X_1, , \ldots, X_{q_{0\min}}$ have the above power decay correlations, where $q_{0\min} = [\frac{q_0}{2}] + 1$; the rest X_j's are generated as

$$X_j = \varepsilon_j + \frac{\sum_{k=1}^{q_{0\min}} X_k}{q_{0\min}},$$

  where ε_j's are i.i.d. $N(0, 0.08)$. $s_{\beta_{0j}}$ consists of $1, \ldots, q_{0\min}$ and a random set of size φ_j from $\{q_{0\min} + 1, \ldots, q\}$, where φ_j is a random number from $\{q_{0\min}, \ldots, q_0\}$.

The coefficient matrix B is first generated as described in Scenario 1 and then scaled as shown in Scenario 2. The evaluation measures in Scenario 1 and Scenario 2 are used. Specifically, PDR, FDR, $|\hat{\Omega}|$ and three norms of the matrix losses are examined for $\hat{\Omega}$; PDR, FDR and PMSE are calculated for the assessment of $\hat{B}$. Two i.i.d. samples of Y are generated synchronously, one of which is used in the estimation of B and Ω and the other is for the computation of PMSE. Since there is no explicit estimate of Ω given by Wang (2013), the matrix losses of aMCR are not calculated. When $p = 50$, we consider the settings that are the combinations of the five graphs, $n = 100$ and 200, $h = 0.8$ and 0.6. Likewise, when $p = 200$, Ω in each of the five graphs is obtained by either the block design or the noise design; but only $n = 100$ and $h = 0.8$ are considered. We repeat the estimation 100 times for each setting. The results are shown in Table 4.8 to Table 4.12.

Table 4.8 shows the results when $p = 50$, $n = 100$ and $h = 0.8$. Even though the same method's estimation quality changes across different graphs, the relative performance among these methods is determined by the types of X. When the X is of type I and III, the SCR methods uniformly outperform aMCR and MRCE. Although SCR_4 has sightly higher $\hat{\Omega}$-PDRs than SCR_3, the latter excels in terms of the other criteria. For the settings with X of type II and IV, aMCR provides $\hat{B}$ and $\hat{\Omega}$ with moderately higher PDRs and smaller PMSEs. On the other hand, aMCR's advantage in PDRs is largely compromised by its even higher augmentations in the corresponding FDRs. Additionally, it is worth noting that smaller PMSE does not necessarily indicate better estimation. Therefore, SCR_3 is the better choice when model selection counts. With n growing to 200 and other

parameters unchanged, the results are displayed in Table 4.9. The SCR methods outperform the other two competitors, especially the SCR_3. Due to the increase of the sample size, the SCR methods overcome the deficiency in $\hat{B}$-PDR, $\hat{\Omega}$-PDR and PMSE as shown when $n = 100$, and remain the advantage in terms of two FDRs. Moreover, Table 4.10 shows the changes when h decreases to 0.6. The SCR methods' performance degenerates with incremental feature selection errors in both $\hat{B}$ and $\hat{\Omega}$.

The results when $p = 200$ are reported in Table 4.11 and Table 4.12. Compared to the results when $p = 50$, there is no essential alteration in aMCR and the two SCR methods' comparative performance in the estimation of B. However, aMCR's $\hat{\Omega}$-FDRs soar to at least 0.811. Therefore, SCR_3's advantage in edge detection becomes more obvious. Comparing across these two precision matrix designs, we found that SCR_3's $\hat{\Omega}$-FDRs in the second case are uniformly larger than those in the first case, although they remain at relatively low levels. This suggests that the noise responses can significantly increase SCR_3's risk of including false edges.

Compared to SCR_4, SCR_3 constantly yields much lower $\hat{\Omega}$-FDRs but slightly smaller $\hat{\Omega}$-PDRs. Therefore, SCR_3 has better estimation of Ω, which can also be inferred from its generally smaller matrix losses. Furthermore, SCR_4's B estimates are comparable to, or almost the same as, those of SCR_3's in the settings with the first three types of X. However, the discrepancies between these two methods' PDR, FDR and PMSE of $\hat{B}$ clearly inflate in the case of type IV. Such differences are largely narrowed when n increases from 100 to 200.

Generally, SCR_3^* can serve as a proper computational approximation of SCR_3 at the cost of some estimation accuracy. The discrepancies between the two methods' PDRs and FDRs are averagely less than 0.05 and the variations in the matrix losses and PMSEs are usually not substantial. Particularly, this approximation works best when X is of type I, which suggests that the alternate updating approach achieves feature selection stability very fast if the covariates are mutually independent. However, for the other three types of X, SCR_3^*'s approximation deteriorates as the correlations among covariates become complex. For instance, significant differences between the estimates of SCR_3 and SCR_3^* can be observed in Table 4.8 under the BA graph and type II X. Fortunately, as $n = 200$, such approximation errors are largely reduced, sometimes even negligible.

In our simulation study, we noticed that MRCE has considerable computational cost due to the two-dimensional grid search of the optimal pair of tuning parameters and the iterative algorithm. It seems hard to balance the feature selection between B and Ω, thereby it is less competitive in terms of all the considered criteria.

4.5 Application to the Glioblastoma Multiforme Cancer Data

In this section, we apply our SCR methods, the alternate updating approach and the simultaneous estimation approach, on a real example: the Glioblastoma multiforme (GBM) cancer data. GBM is the most familiar type of malignant brain cancer among adults, without efficient therapy.

The data were analyzed by the Cancer Genome Atlas(TCGA) Research Network (McLendon et al., 2008; Verhaak et al., 2010) and are available on the TCGA Data Portal (https://tcga-data.nci.nih.gov/tcga/tcgaHome2.jsp). The dataset consists of 11861 gene expression levels and 534 microRNA values from 202 subjects. We utilize the sparse multivariate response regression model to predict the microRNAs with the gene expressions and explore the network among the microRNAs. Due to the small sample size, it is impractical to study the whole dataset. We follow the analysis process in Wang (2013) and the references therein. Its steps are briefly described here. First, 6 samples with missing microRNA values are excluded from the analysis. Afterwards, a prescreening procedure is conducted to sort the genes and microRNAs based on their median absolute deviation (MAD) values and then extract a subset from the complete dataset, which is composed of the gene expressions with the 500 largest MADs and the microRNA values with the top 20 MADs. The resulting dataset is randomly divided into a training dataset with 120 subjects and a testing dataset with the remaining 76 subjects. The training dataset is used to estimate the 500 genes' coefficient matrix B, along with the precision matrix Ω of the 20 microRNAs. The testing dataset is used in the calculation of PSE, where

$$\text{PSE} = \frac{1}{\text{testing sample size}} \sum_{i \in \text{testing set}} \frac{\|\boldsymbol{y}_i - \hat{B}^\tau \boldsymbol{x}_i\|_2^2}{\text{dimension of } \boldsymbol{y}_i}.$$

Furthermore, the number of involved genes in the fitted model is also examined. Due to the dimension of the data, we do not consider the method SCR_2. The other SCR methods' averages and standard errors of the fore-

going two measures over 50 replications are reported in Table 4.13. In addition, before Wang (2013), this dataset was studied by Lee and Liu (2012) in a slightly different procedure and over 10 replications. For comparison, their results are presented in Table 4.14, where those of CW, PWL and DML are copied from the latter and those of aMCR originate from the former. CW denotes the Curds and Whey method in Breiman and Friedman (1997). PWL and DWL are proposed by Lee and Liu (2012), which are similar to MRCR but impose the adaptive Lasso penalty on both B and Ω.

As can be seen, the PSEs produced by SCR_1, SCR_3, SCR_4 are close to that of aMCR, which is the smallest among the previous methods' results; SCR_1^*, SCR_3^* have even better performance. From the perspective of the included genes' number, the SCR methods fit the model with around 44 genes, significantly less than the others. Overall, the SCR methods use less gene expressions to achieve better or comparable prediction accuracy.

Figure 4.1 visualizes the networks of the 20 microRNAs detected by SCR_3, SCR_3^* and SCR_4. The number of the edges in each estimated network is summaries in the table below. Apparently, the alternate update

Number of edges detected by SCR_3, SCR_3^* and SCR_4

SCR_3	SCR_3^*	SCR_4	common edges
22	21	35	18

approach (SCR_3 and SCR_3^*) yields sparser graphical structure than the simultaneous approach (SCR_4). This observation is consistent with the results shown in the simulation study. In addition, networks obtained by SCR_3 and SCR_3^* share 19 edges and are very close. It is worth noting that

two microRNAs "hsa.miR.136" and "hsa.miR.377" are included in the previous analyses; however, they are excluded in our study due to their small MAD values and are replaced by another two microRNAs "hsa.miR.9" and "hsa.miR.127". Except the edges connected to the discrepant miRNAs, most of the rest edges captured by the SCR methods are also detected by DML and aMCR, e.g. the subnetwork among "hsa.miR.630", "hsa.miR.630" and "hsa.miR.801". Moreover, graphs in Figure 4.1 further screen off some weak conditional correlations and are much sparser.

4.6 Conclusion

In this chapter, we employed sequential methods and conditional regression formulation to fit the sparse multivariate response regression models in two approaches. The first approach embeds SLasso for the estimation of B and SSPS for the estimation of Θ into the alternate updating framework. This approach is elaborated in three scenarios to get a more comprehensive analysis. Provided an initial estimate of B, the second approach solves the conditional regressions separately with a sequential method similar to SLasso, where $\boldsymbol{\beta}_j$ and $\boldsymbol{\theta}_j$ are estimated simultaneously.

The simulation study of Scenario 1 shows the benefit to incorporate the responses' correlations into the coefficient matrix estimation. Likewise, the results of Scenario 2 prove the necessity to adjust for the non-constant means in a general Gaussian graphical model, when the information about the related covariates is available. In Scenario 3, compared to the other methods for the joint estimation of B and Ω in multivariate response regres-

sion models, the proposed SCR methods are computationally more efficient, since there is no tuning parameter selection involved. More importantly, they inherit the sequential methods' variable selection accuracy and normally outperform the others, especially the alternate updating approach. This advantage is also supported by real data analysis.

A common drawback of methods utilizing the alternate updating scheme to solve the nonconvex penalized likelihood optimization is that they are likely to be trapped in the local optimizers. However, this problem is irrelevant for our alternate updating method, because no optimization is involved in the sequential feature selection procedures. To further relieve the computation burden, we also considered the implementation of this approach without iteration. Such approximation generally works.

It was recognized that the SCR methods could underestimate the number of relevant features, although the selected features are usually the true ones. It may not be the best choice if only the minimization of the prediction errors matters. Furthermore, we realized that the simultaneous estimation approach ignores the intrinsic symmetry among those $\boldsymbol{\theta}_j$'s and thus has relatively higher edge detection FDR. Instead of exploring the edges over the whole graph, it identifies the neighborhood of each vertex independently and combines all the resulting neighborhood sets with a simple "OR" rule to estimate the edge set. In the view of edge detection, it is analogous to SR-SLasso in Luo and Chen (2014a) and tends to include many false edges.

Table 4.1: Average (SD) of $\hat{B}$ related PDR, FDR and PMSE when $n = 100$, $p = 10$ and $h = 0.75$

			$q = [5e^{n^{0.3}}]$			$q = 500$		
Σ_X	Cor(E)	Mtd.	PDR	FDR	PMSE	PDR	FDR	PMSE
I	I	SLasso	0.886(0.069)	0.073(0.029)	154.886(17.000)	0.774(0.102)	0.081(0.036)	183.320(25.303)
		SCR_1	0.894(0.069)	0.089(0.034)	155.670(17.468)	0.786(0.097)	0.109(0.043)	182.682(24.580)
		SCR_1^*	0.894(0.069)	0.089(0.034)	155.671(17.429)	0.786(0.100)	0.105(0.041)	182.388(25.096)
I	Band	SLasso	0.883(0.073)	0.071(0.032)	155.884(18.755)	0.759(0.096)	0.079(0.035)	185.925(24.169)
		SCR_1	0.982(0.035)	0.122(0.043)	129.849(10.775)	0.940(0.070)	0.131(0.047)	140.886(17.649)
		SCR_1^*	0.980(0.036)	0.118(0.040)	130.386(10.900)	0.929(0.074)	0.122(0.042)	143.782(18.166)
I	PD	SLasso	0.883(0.090)	0.073(0.035)	155.165(24.897)	0.757(0.113)	0.084(0.042)	187.293(30.901)
		SCR_1	0.994(0.022)	0.107(0.046)	121.704(11.724)	0.986(0.034)	0.109(0.044)	124.959(13.535)
		SCR_1^*	0.994(0.022)	0.107(0.045)	121.887(11.650)	0.983(0.038)	0.106(0.042)	126.167(14.051)
I	Eq	SLasso	0.889(0.081)	0.066(0.031)	153.492(20.067)	0.769(0.113)	0.083(0.043)	181.391(29.214)
		SCR_1	0.982(0.036)	0.080(0.035)	128.581(11.427)	0.935(0.068)	0.092(0.033)	138.449(17.824)
		SCR_1^*	0.982(0.036)	0.079(0.034)	128.575(11.438)	0.933(0.068)	0.090(0.031)	138.600(18.005)
PD	I	SLasso	0.436(0.062)	0.108(0.052)	209.997(22.806)	0.384(0.052)	0.109(0.055)	218.163(23.533)
		SCR_1	0.439(0.064)	0.140(0.056)	212.310(23.575)	0.386(0.053)	0.145(0.065)	220.063(24.281)
		SCR_1^*	0.438(0.063)	0.135(0.054)	211.833(23.590)	0.387(0.053)	0.143(0.064)	219.631(24.134)
PD	Band	SLasso	0.449(0.060)	0.103(0.051)	209.638(23.105)	0.385(0.046)	0.115(0.059)	222.780(24.662)
		SCR_1	0.547(0.090)	0.153(0.053)	194.764(23.320)	0.446(0.060)	0.156(0.060)	211.438(23.356)
		SCR_1^*	0.532(0.083)	0.139(0.048)	196.007(22.720)	0.440(0.058)	0.145(0.056)	211.441(23.679)
PD	PD	SLasso	0.434(0.062)	0.107(0.056)	208.852(23.355)	0.379(0.049)	0.111(0.052)	223.441(24.341)
		SCR_1	0.736(0.105)	0.140(0.046)	159.655(24.199)	0.593(0.102)	0.152(0.049)	183.042(23.261)
		SCR_1^*	0.696(0.101)	0.131(0.042)	165.178(23.127)	0.568(0.097)	0.136(0.049)	186.440(24.021)
PD	Eq	SLasso	0.442(0.061)	0.106(0.050)	210.356(27.566)	0.391(0.055)	0.111(0.051)	220.643(23.407)
		SCR_1	0.561(0.078)	0.137(0.050)	188.951(25.702)	0.481(0.064)	0.156(0.061)	201.927(22.408)
		SCR_1^*	0.553(0.077)	0.130(0.046)	189.550(26.133)	0.479(0.065)	0.142(0.053)	201.684(22.294)
Eq	I	SLasso	0.412(0.102)	0.175(0.080)	234.489(43.688)	0.301(0.091)	0.240(0.104)	253.769(48.849)
		SCR_1	0.421(0.111)	0.191(0.084)	235.037(45.373)	0.300(0.089)	0.262(0.102)	255.594(48.135)
		SCR_1^*	0.421(0.107)	0.188(0.084)	234.365(45.037)	0.303(0.087)	0.254(0.098)	254.366(47.562)
Eq	Band	SLasso	0.405(0.110)	0.174(0.083)	235.004(43.466)	0.285(0.086)	0.250(0.106)	255.051(39.274)
		SCR_1	0.563(0.150)	0.185(0.066)	208.543(45.604)	0.383(0.131)	0.248(0.102)	237.397(42.551)
		SCR_1^*	0.542(0.139)	0.174(0.069)	210.866(43.098)	0.361(0.116)	0.246(0.101)	240.199(41.488)
Eq	PD	SLasso	0.410(0.110)	0.172(0.080)	233.485(48.536)	0.296(0.085)	0.230(0.103)	250.660(43.912)
		SCR_1	0.780(0.117)	0.138(0.056)	171.326(42.055)	0.634(0.146)	0.175(0.069)	190.273(42.776)
		SCR_1^*	0.734(0.126)	0.142(0.058)	178.580(43.184)	0.576(0.139)	0.171(0.074)	198.498(42.856)
Eq	Eq	SLasso	0.404(0.107)	0.177(0.079)	238.293(45.335)	0.298(0.091)	0.237(0.102)	250.465(44.639)
		SCR_1	0.611(0.122)	0.152(0.059)	203.059(47.871)	0.441(0.121)	0.212(0.086)	222.615(45.470)
		SCR_1^*	0.597(0.122)	0.149(0.066)	204.694(47.758)	0.432(0.119)	0.206(0.087)	223.122(44.602)

Table 4.2: Average (SD) of $\hat{B}$ related PDR, FDR and PMSE when $n = 100$, $p = 20$ and $h = 0.75$

			$q = [5e^{n^{0.3}}]$			$q = 500$		
Σ_X	Cor(E)	Mtd.	PDR	FDR	PMSE	PDR	FDR	PMSE
I	I	SLasso	0.885(0.049)	0.069(0.022)	309.575(24.244)	0.764(0.070)	0.080(0.024)	368.612(34.797)
		SCR_1	0.893(0.046)	0.097(0.027)	313.472(23.116)	0.772(0.070)	0.120(0.036)	374.810(35.081)
		SCR_1^*	0.893(0.046)	0.096(0.027)	313.466(23.091)	0.772(0.070)	0.118(0.034)	373.943(34.696)
I	Band	SLasso	0.890(0.051)	0.069(0.021)	307.686(26.751)	0.764(0.074)	0.081(0.026)	367.721(37.129)
		SCR_1	0.972(0.031)	0.133(0.028)	266.055(17.950)	0.891(0.071)	0.148(0.035)	306.512(35.535)
		SCR_1^*	0.971(0.031)	0.129(0.025)	266.738(17.961)	0.886(0.070)	0.135(0.030)	308.613(34.847)
I	PD	SLasso	0.884(0.058)	0.067(0.027)	309.306(32.326)	0.763(0.078)	0.082(0.026)	368.461(43.336)
		SCR_1	0.991(0.017)	0.117(0.032)	246.855(18.427)	0.962(0.037)	0.131(0.032)	262.081(24.448)
		SCR_1^*	0.990(0.018)	0.115(0.030)	247.592(18.914)	0.957(0.041)	0.125(0.029)	264.937(25.666)
I	Eq	SLasso	0.881(0.063)	0.069(0.025)	312.816(34.535)	0.772(0.086)	0.081(0.027)	364.496(44.038)
		SCR_1	0.969(0.034)	0.101(0.026)	267.037(21.771)	0.912(0.051)	0.114(0.030)	292.469(28.695)
		SCR_1^*	0.968(0.036)	0.100(0.025)	267.316(22.441)	0.910(0.051)	0.110(0.027)	292.778(28.827)
PD	I	SLasso	0.440(0.044)	0.105(0.036)	423.462(38.550)	0.381(0.037)	0.117(0.039)	443.905(33.148)
		SCR_1	0.446(0.042)	0.157(0.044)	429.725(39.721)	0.386(0.036)	0.179(0.048)	449.714(32.129)
		SCR_1^*	0.446(0.043)	0.153(0.041)	428.362(39.363)	0.386(0.036)	0.173(0.043)	448.117(31.486)
PD	Band	SLasso	0.439(0.039)	0.104(0.034)	427.515(36.524)	0.383(0.033)	0.113(0.039)	441.257(31.991)
		SCR_1	0.504(0.049)	0.168(0.042)	409.028(36.158)	0.427(0.043)	0.187(0.050)	426.498(29.998)
		SCR_1^*	0.498(0.048)	0.158(0.036)	408.649(36.137)	0.425(0.041)	0.174(0.046)	425.375(30.101)
PD	PD	SLasso	0.440(0.040)	0.104(0.036)	426.221(37.890)	0.381(0.038)	0.118(0.040)	441.525(34.009)
		SCR_1	0.645(0.071)	0.160(0.037)	355.283(38.242)	0.530(0.063)	0.188(0.044)	383.324(31.420)
		SCR_1^*	0.628(0.066)	0.144(0.036)	359.589(37.193)	0.515(0.058)	0.165(0.043)	386.735(31.028)
PD	Eq	SLasso	0.432(0.047)	0.098(0.039)	423.825(40.530)	0.382(0.041)	0.114(0.043)	442.354(34.612)
		SCR_1	0.547(0.058)	0.156(0.039)	383.661(36.923)	0.465(0.047)	0.177(0.047)	408.700(32.132)
		SCR_1^*	0.541(0.056)	0.143(0.037)	383.382(36.331)	0.462(0.045)	0.161(0.044)	407.987(32.533)
Eq	I	SLasso	0.404(0.066)	0.170(0.054)	469.488(61.884)	0.287(0.057)	0.237(0.073)	503.862(61.442)
		SCR_1	0.408(0.066)	0.204(0.055)	475.051(64.159)	0.285(0.060)	0.279(0.074)	509.644(62.457)
		SCR_1^*	0.409(0.066)	0.200(0.054)	472.633(63.386)	0.286(0.059)	0.274(0.073)	507.453(61.536)
Eq	Band	SLasso	0.417(0.081)	0.160(0.056)	459.006(65.671)	0.293(0.059)	0.231(0.071)	503.296(63.705)
		SCR_1	0.527(0.105)	0.197(0.058)	424.436(70.964)	0.354(0.077)	0.271(0.068)	480.921(61.252)
		SCR_1^*	0.515(0.103)	0.185(0.053)	426.615(69.555)	0.348(0.075)	0.256(0.067)	481.475(61.653)
Eq	PD	SLasso	0.404(0.079)	0.166(0.058)	466.825(69.883)	0.292(0.064)	0.238(0.078)	501.890(64.210)
		SCR_1	0.702(0.101)	0.170(0.046)	365.308(68.196)	0.527(0.101)	0.222(0.064)	413.918(62.829)
		SCR_1^*	0.667(0.099)	0.160(0.045)	374.818(66.045)	0.489(0.101)	0.209(0.060)	425.265(62.316)
Eq	Eq	SLasso	0.415(0.079)	0.166(0.050)	463.455(65.530)	0.298(0.067)	0.228(0.071)	498.444(69.334)
		SCR_1	0.579(0.088)	0.172(0.050)	406.119(66.782)	0.434(0.075)	0.223(0.057)	444.783(67.935)
		SCR_1^*	0.568(0.085)	0.164(0.045)	408.426(64.568)	0.426(0.076)	0.213(0.058)	446.233(68.138)

Table 4.3: Average (SD) of $\hat{B}$ related PDR, FDR and PMSE when $n = 100$, $p = 10$ and $h = 0.8$

			$q = [5e^{n^{0.3}}]$			$q = 500$		
Σ_X	Cor(E)	Mtd.	PDR	FDR	PMSE	PDR	FDR	PMSE
I	I	SLasso	0.962(0.048)	0.062(0.029)	105.198(11.679)	0.903(0.075)	0.067(0.031)	120.428(18.796)
		SCR_1	0.970(0.041)	0.077(0.028)	105.140(10.118)	0.910(0.071)	0.092(0.041)	120.709(17.199)
		SCR_1^*	0.970(0.041)	0.077(0.029)	105.124(10.106)	0.912(0.070)	0.090(0.040)	120.293(17.153)
I	Band	SLasso	0.971(0.039)	0.056(0.027)	103.184(10.824)	0.892(0.076)	0.068(0.031)	121.486(18.287)
		SCR_1	0.996(0.018)	0.121(0.042)	95.081 (7.138)	0.965(0.052)	0.119(0.045)	102.103(12.714)
		SCR_1^*	0.996(0.017)	0.120(0.039)	95.065 (7.079)	0.962(0.055)	0.113(0.039)	102.717(13.183)
I	PD	SLasso	0.965(0.044)	0.057(0.031)	104.223(13.328)	0.901(0.078)	0.069(0.034)	119.052(20.048)
		SCR_1	0.999(0.006)	0.093(0.038)	90.645 (8.580)	0.992(0.027)	0.110(0.044)	92.311 (9.799)
		SCR_1^*	0.999(0.006)	0.093(0.038)	90.653 (8.589)	0.991(0.028)	0.107(0.042)	92.714 (10.215)
I	Eq	SLasso	0.957(0.054)	0.064(0.032)	107.127(13.788)	0.888(0.084)	0.068(0.034)	122.519(20.795)
		SCR_1	0.992(0.024)	0.078(0.034)	95.810 (8.668)	0.969(0.049)	0.086(0.034)	100.732(13.011)
		SCR_1^*	0.992(0.024)	0.079(0.034)	95.903 (8.740)	0.968(0.049)	0.086(0.034)	100.920(13.109)
PD	I	SLasso	0.518(0.085)	0.101(0.048)	163.178(19.150)	0.447(0.063)	0.110(0.049)	174.418(19.505)
		SCR_1	0.525(0.080)	0.128(0.050)	163.742(18.971)	0.449(0.059)	0.145(0.054)	175.995(19.689)
		SCR_1^*	0.526(0.080)	0.124(0.050)	163.307(18.812)	0.450(0.060)	0.139(0.051)	175.252(19.555)
PD	Band	SLasso	0.520(0.071)	0.095(0.042)	165.151(17.935)	0.452(0.061)	0.103(0.047)	177.653(19.146)
		SCR_1	0.626(0.100)	0.146(0.048)	148.930(19.042)	0.514(0.080)	0.155(0.059)	167.683(18.944)
		SCR_1^*	0.608(0.091)	0.136(0.045)	151.154(18.517)	0.509(0.082)	0.139(0.053)	167.823(18.954)
PD	PD	SLasso	0.519(0.072)	0.099(0.051)	165.782(21.468)	0.448(0.062)	0.102(0.052)	177.966(21.251)
		SCR_1	0.812(0.093)	0.127(0.044)	121.127(19.909)	0.703(0.112)	0.144(0.053)	135.176(23.154)
		SCR_1^*	0.782(0.097)	0.121(0.043)	125.027(20.083)	0.665(0.104)	0.127(0.048)	140.902(22.637)
PD	Eq	SLasso	0.525(0.073)	0.100(0.046)	164.442(21.419)	0.447(0.068)	0.109(0.049)	178.913(19.451)
		SCR_1	0.670(0.079)	0.123(0.046)	141.705(20.325)	0.556(0.078)	0.147(0.056)	158.860(19.368)
		SCR_1^*	0.664(0.084)	0.116(0.044)	142.236(21.059)	0.552(0.079)	0.137(0.052)	158.840(19.259)
Eq	I	SLasso	0.545(0.109)	0.136(0.061)	173.987(34.855)	0.389(0.107)	0.198(0.085)	203.022(37.185)
		SCR_1	0.549(0.113)	0.158(0.064)	175.433(34.924)	0.396(0.108)	0.222(0.086)	203.166(37.034)
		SCR_1^*	0.553(0.112)	0.154(0.063)	174.290(34.481)	0.399(0.112)	0.217(0.086)	201.916(37.104)
Eq	Band	SLasso	0.538(0.110)	0.135(0.058)	177.883(33.385)	0.402(0.107)	0.197(0.086)	199.281(36.979)
		SCR_1	0.697(0.137)	0.159(0.055)	152.361(33.528)	0.508(0.152)	0.208(0.087)	181.217(38.196)
		SCR_1^*	0.676(0.137)	0.149(0.052)	155.462(33.750)	0.487(0.133)	0.199(0.086)	183.755(37.312)
Eq	PD	SLasso	0.546(0.118)	0.135(0.065)	174.281(41.824)	0.408(0.109)	0.190(0.091)	193.208(36.695)
		SCR_1	0.864(0.104)	0.123(0.053)	123.276(35.208)	0.730(0.142)	0.152(0.067)	139.578(34.499)
		SCR_1^*	0.847(0.111)	0.121(0.050)	126.113(36.557)	0.698(0.149)	0.149(0.060)	143.953(35.486)
Eq	Eq	SLasso	0.537(0.116)	0.135(0.060)	175.933(37.885)	0.392(0.115)	0.196(0.094)	199.210(38.852)
		SCR_1	0.714(0.112)	0.126(0.051)	146.438(34.545)	0.567(0.140)	0.176(0.078)	169.537(39.425)
		SCR_1^*	0.714(0.116)	0.122(0.048)	146.419(35.821)	0.556(0.142)	0.173(0.084)	170.863(40.212)

Table 4.4: Average (SD) of $\hat{B}$ related PDR, FDR and PMSE when $n = 100$, $p = 20$ and $h = 0.8$

			$q = [5e^{n^{0.3}}]$			$q = 500$		
Σ_X	Cor(E)	Mtd.	PDR	FDR	PMSE	PDR	FDR	PMSE
I	I	SLasso	0.964(0.034)	0.060(0.020)	209.980(14.550)	0.903(0.057)	0.072(0.023)	239.719(28.252)
		SCR_1	0.966(0.035)	0.087(0.026)	214.418(15.156)	0.906(0.054)	0.105(0.031)	244.549(27.234)
		SCR_1^*	0.966(0.035)	0.087(0.026)	214.400(15.143)	0.907(0.054)	0.104(0.029)	244.215(26.961)
I	Band	SLasso	0.961(0.037)	0.063(0.020)	211.087(17.982)	0.899(0.055)	0.068(0.019)	240.251(26.058)
		SCR_1	0.991(0.018)	0.132(0.029)	193.458(12.698)	0.965(0.039)	0.135(0.031)	205.853(19.293)
		SCR_1^*	0.991(0.018)	0.131(0.028)	193.453(12.588)	0.964(0.039)	0.130(0.028)	206.521(19.358)
I	PD	SLasso	0.961(0.036)	0.058(0.023)	210.737(21.913)	0.907(0.055)	0.071(0.026)	235.770(28.354)
		SCR_1	0.996(0.013)	0.109(0.030)	184.410(14.241)	0.986(0.023)	0.126(0.037)	188.312(14.847)
		SCR_1^*	0.996(0.013)	0.108(0.029)	184.432(14.223)	0.985(0.023)	0.123(0.035)	188.853(15.232)
I	Eq	SLasso	0.964(0.034)	0.060(0.023)	210.383(21.098)	0.901(0.056)	0.070(0.025)	240.642(30.747)
		SCR_1	0.991(0.019)	0.091(0.027)	193.515(16.309)	0.966(0.037)	0.106(0.029)	205.908(20.488)
		SCR_1^*	0.991(0.018)	0.090(0.027)	193.339(16.142)	0.966(0.036)	0.104(0.028)	205.704(20.320)
PD	I	SLasso	0.523(0.050)	0.093(0.033)	329.544(27.458)	0.443(0.041)	0.107(0.034)	352.977(26.589)
		SCR_1	0.520(0.048)	0.139(0.041)	336.737(28.249)	0.439(0.038)	0.166(0.047)	361.151(27.513)
		SCR_1^*	0.521(0.048)	0.134(0.039)	335.536(27.846)	0.440(0.038)	0.158(0.040)	359.453(27.042)
PD	Band	SLasso	0.531(0.054)	0.095(0.029)	325.265(25.656)	0.446(0.046)	0.109(0.036)	351.058(27.050)
		SCR_1	0.602(0.065)	0.156(0.038)	305.912(27.299)	0.491(0.055)	0.172(0.047)	338.100(26.828)
		SCR_1^*	0.595(0.062)	0.146(0.035)	306.642(27.088)	0.488(0.053)	0.162(0.043)	337.574(26.786)
PD	PD	SLasso	0.521(0.054)	0.092(0.034)	327.865(27.546)	0.443(0.046)	0.107(0.039)	353.673(29.082)
		SCR_1	0.746(0.073)	0.151(0.040)	259.167(28.317)	0.616(0.074)	0.180(0.041)	295.458(31.419)
		SCR_1^*	0.723(0.071)	0.136(0.034)	265.363(28.335)	0.597(0.071)	0.154(0.037)	300.186(30.903)
PD	Eq	SLasso	0.513(0.052)	0.097(0.034)	331.109(32.957)	0.445(0.048)	0.107(0.035)	351.722(28.381)
		SCR_1	0.634(0.060)	0.144(0.034)	293.686(32.360)	0.538(0.062)	0.173(0.044)	319.856(29.213)
		SCR_1^*	0.629(0.061)	0.135(0.032)	294.108(32.749)	0.536(0.059)	0.154(0.037)	318.907(29.132)
Eq	I	SLasso	0.537(0.084)	0.134(0.041)	351.521(53.232)	0.397(0.077)	0.187(0.054)	394.637(48.872)
		SCR_1	0.541(0.085)	0.167(0.047)	356.163(53.235)	0.395(0.074)	0.230(0.062)	399.437(48.470)
		SCR_1^*	0.541(0.086)	0.163(0.045)	355.230(53.612)	0.396(0.073)	0.222(0.059)	397.976(48.275)
Eq	Band	SLasso	0.537(0.088)	0.133(0.043)	349.302(55.237)	0.402(0.074)	0.189(0.059)	399.907(54.872)
		SCR_1	0.658(0.099)	0.173(0.045)	313.623(56.387)	0.468(0.090)	0.236(0.058)	378.349(54.308)
		SCR_1^*	0.645(0.102)	0.161(0.042)	315.888(57.445)	0.464(0.088)	0.219(0.057)	377.860(55.047)
Eq	PD	SLasso	0.524(0.088)	0.140(0.044)	357.925(53.362)	0.414(0.087)	0.179(0.059)	385.445(54.112)
		SCR_1	0.809(0.080)	0.149(0.037)	265.598(44.791)	0.657(0.097)	0.184(0.053)	303.084(51.038)
		SCR_1^*	0.782(0.085)	0.144(0.037)	273.923(46.591)	0.629(0.103)	0.171(0.049)	311.717(53.461)
Eq	Eq	SLasso	0.534(0.088)	0.135(0.044)	350.591(53.344)	0.406(0.081)	0.178(0.060)	385.334(55.356)
		SCR_1	0.686(0.086)	0.148(0.042)	300.268(49.717)	0.547(0.085)	0.194(0.050)	337.082(54.267)
		SCR_1^*	0.680(0.086)	0.141(0.040)	301.524(49.343)	0.534(0.082)	0.181(0.047)	339.159(53.261)

Table 4.5: Average (SD) of $\hat{\Omega}$ related PDR, FDR, $|\hat{\Omega}|$ and matrix loss in three norms when $p = 50$, $n = 100$

			$h = 0.2$						$h = 0.5$					
Graph	q	Mtd.	PDR	FDR	$\lvert\hat{\Omega}\rvert$	$\lVert\cdot\rVert_S$	$\lVert\cdot\rVert_{\ell_1}$	$\lVert\cdot\rVert_F$	PDR	FDR	$\lvert\hat{\Omega}\rvert$	$\lVert\cdot\rVert_S$	$\lVert\cdot\rVert_{\ell_1}$	$\lVert\cdot\rVert_F$
AR(1)	10	SSPS	0.823	0.385	66	1.269	1.757	3.854	0.549	0.760	113	1.620	2.310	5.404
			(0.056)	(0.067)	(7)	(0.090)	(0.164)	(0.204)	(0.088)	(0.033)	(15)	(0.048)	(0.157)	(0.213)
		SCR_2	0.996	0.060	52	1.037	1.392	2.371	0.998	0.062	52	1.047	1.428	2.408
			(0.008)	(0.032)	(2)	(0.208)	(0.317)	(0.286)	(0.006)	(0.035)	(2)	(0.186)	(0.326)	(0.278)
	20	SSPS	0.850	0.292	59	1.235	1.635	3.890	0.462	0.716	80	1.647	2.103	6.000
			(0.046)	(0.064)	(5)	(0.086)	(0.144)	(0.169)	(0.064)	(0.041)	(11)	(0.033)	(0.118)	(0.121)
		SCR_2	0.992	0.104	54	1.536	2.105	3.738	0.996	0.098	54	1.526	2.054	3.770
			(0.013)	(0.044)	(3)	(0.278)	(0.496)	(0.408)	(0.009)	(0.047)	(3)	(0.312)	(0.486)	(0.421)
AR(2)	10	SSPS	0.435	0.340	64	1.632	2.239	4.695	0.271	0.734	99	2.017	2.742	5.919
			(0.048)	(0.067)	(7)	(0.087)	(0.168)	(0.187)	(0.048)	(0.036)	(13)	(0.057)	(0.145)	(0.159)
		SCR_2	0.656	0.076	69	1.209	1.840	3.412	0.664	0.084	70	1.224	1.891	3.451
			(0.058)	(0.033)	(6)	(0.219)	(0.350)	(0.235)	(0.062)	(0.033)	(7)	(0.242)	(0.381)	(0.250)
	20	SSPS	0.434	0.270	58	1.629	2.152	4.740	0.203	0.717	70	2.082	2.595	6.402
			(0.042)	(0.064)	(5)	(0.071)	(0.145)	(0.163)	(0.040)	(0.048)	(10)	(0.045)	(0.119)	(0.132)
		SCR_2	0.659	0.124	73	1.805	2.790	4.542	0.660	0.124	73	1.744	2.643	4.500
			(0.060)	(0.045)	(7)	(0.318)	(0.552)	(0.401)	(0.055)	(0.043)	(6)	(0.273)	(0.476)	(0.362)
Circle	10	SSPS	0.803	0.392	67	1.254	1.720	3.864	0.531	0.759	110	1.619	2.280	5.437
			(0.050)	(0.062)	(6)	(0.092)	(0.156)	(0.192)	(0.076)	(0.027)	(13)	(0.045)	(0.134)	(0.177)
		SCR_2	0.977	0.058	52	1.010	1.356	2.360	0.979	0.063	52	0.995	1.339	2.372
			(0.010)	(0.032)	(2)	(0.185)	(0.319)	(0.283)	(0.009)	(0.033)	(2)	(0.182)	(0.305)	(0.262)
	20	SSPS	0.832	0.282	58	1.225	1.603	3.926	0.454	0.716	80	1.647	2.108	6.001
			(0.050)	(0.068)	(5)	(0.075)	(0.156)	(0.187)	(0.067)	(0.035)	(9)	(0.034)	(0.113)	(0.131)
		SCR_2	0.975	0.093	54	1.514	2.036	3.733	0.975	0.100	54	1.497	2.052	3.749
			(0.012)	(0.040)	(2)	(0.274)	(0.428)	(0.401)	(0.013)	(0.041)	(3)	(0.297)	(0.484)	(0.411)
Cluster	10	SSPS	0.209	0.399	35	2.518	2.861	7.263	0.134	0.809	69	2.699	3.206	7.941
			(0.062)	(0.107)	(8)	(0.036)	(0.112)	(0.218)	(0.048)	(0.058)	(8)	(0.036)	(0.087)	(0.140)
		SCR_2	0.750	0.065	80	2.088	2.580	4.724	0.734	0.059	78	2.155	2.613	4.792
			(0.153)	(0.033)	(16)	(0.347)	(0.400)	(0.767)	(0.155)	(0.036)	(16)	(0.308)	(0.408)	(0.762)
	20	SSPS	0.170	0.368	27	2.526	2.797	7.419	0.083	0.832	49	2.728	3.119	8.232
			(0.064)	(0.108)	(8)	(0.033)	(0.103)	(0.226)	(0.033)	(0.059)	(9)	(0.022)	(0.096)	(0.085)
		SCR_2	0.747	0.094	83	2.477	3.266	5.756	0.724	0.100	80	2.442	3.239	5.754
			(0.159)	(0.040)	(18)	(0.509)	(0.773)	(0.562)	(0.147)	(0.042)	(16)	(0.468)	(0.732)	(0.557)

Continued on next page

Table 4.5: Average (SD) of $\hat{\Omega}$ related PDR, FDR, $|\hat{\Omega}|$ and matrix loss in three norms when $p = 50$, $n = 100$

			$h = 0.2$						$h = 0.5$					
Graph	q	Mtd.	PDR	FDR	$\lvert\hat{\Omega}\rvert$	$\lVert\cdot\rVert_S$	$\lVert\cdot\rVert_{\ell_1}$	$\lVert\cdot\rVert_F$	PDR	FDR	$\lvert\hat{\Omega}\rvert$	$\lVert\cdot\rVert_S$	$\lVert\cdot\rVert_{\ell_1}$	$\lVert\cdot\rVert_F$
ER	10	SSPS	0.489	0.468	56	2.333	3.845	6.068	0.302	0.813	95	2.796	4.963	7.815
			(0.084)	(0.067)	(8)	(0.269)	(0.594)	(0.525)	(0.074)	(0.039)	(12)	(0.258)	(0.686)	(0.527)
		SCR_2	0.813	0.121	56	1.842	2.835	4.748	0.801	0.130	54	1.955	3.057	4.914
			(0.083)	(0.047)	(6)	(0.343)	(0.632)	(0.724)	(0.094)	(0.054)	(6)	(0.416)	(0.805)	(0.818)
	20	SSPS	0.502	0.394	50	2.371	3.881	6.068	0.251	0.794	75	2.960	4.831	8.456
			(0.097)	(0.071)	(6)	(0.252)	(0.664)	(0.558)	(0.063)	(0.043)	(10)	(0.243)	(0.615)	(0.561)
		SCR_2	0.784	0.182	58	2.704	4.188	6.973	0.789	0.179	59	2.756	4.210	6.938
			(0.092)	(0.057)	(6)	(0.548)	(1.056)	(1.052)	(0.090)	(0.062)	(7)	(0.505)	(0.897)	(0.821)
Tridiag	10	SSPS	0.576	0.391	70	4.105	5.491	11.765	0.454	0.741	129	5.180	7.534	16.069
			(0.049)	(0.060)	(7)	(0.449)	(0.659)	(0.776)	(0.061)	(0.030)	(15)	(0.353)	(0.675)	(0.838)
		SCR_2	0.667	0.048	51	3.063	3.955	6.777	0.667	0.050	52	2.970	3.909	6.768
			(0.046)	(0.033)	(2)	(0.789)	(1.070)	(1.006)	(0.044)	(0.030)	(2)	(0.603)	(0.938)	(0.817)
	20	SSPS	0.596	0.269	60	4.046	5.152	12.248	0.396	0.688	93	5.397	7.096	18.247
			(0.047)	(0.058)	(4)	(0.362)	(0.486)	(0.725)	(0.055)	(0.045)	(11)	(0.296)	(0.505)	(0.737)
		SCR_2	0.670	0.084	53	4.401	5.830	10.689	0.673	0.082	53	4.461	5.985	10.686
			(0.048)	(0.038)	(2)	(0.767)	(1.088)	(1.094)	(0.046)	(0.037)	(2)	(0.984)	(1.481)	(1.300)
Hub	10	SSPS	0.533	0.464	45	5.141	9.474	9.324	0.343	0.813	83	6.634	12.808	12.801
			(0.063)	(0.056)	(5)	(0.590)	(1.102)	(0.922)	(0.062)	(0.031)	(8)	(0.549)	(1.031)	(0.994)
		SCR_2	0.843	0.165	46	4.810	8.211	7.400	0.852	0.157	46	4.704	8.109	7.365
			(0.054)	(0.058)	(3)	(1.495)	(2.209)	(1.629)	(0.045)	(0.054)	(3)	(1.517)	(2.240)	(1.624)
	20	SSPS	0.542	0.386	40	5.279	9.663	9.923	0.298	0.782	62	7.188	13.347	14.525
			(0.065)	(0.069)	(4)	(0.520)	(1.048)	(0.839)	(0.061)	(0.044)	(8)	(0.360)	(0.731)	(0.718)
		SCR_2	0.803	0.257	49	6.467	11.236	10.483	0.807	0.245	48	6.884	11.682	10.770
			(0.054)	(0.060)	(4)	(1.874)	(2.734)	(2.168)	(0.057)	(0.063)	(4)	(2.440)	(3.679)	(2.677)
BA	10	SSPS	0.566	0.624	74	8.187	15.645	12.111	0.442	0.838	134	9.570	19.519	13.870
			(0.059)	(0.051)	(9)	(2.383)	(4.898)	(1.595)	(0.076)	(0.020)	(19)	(2.709)	(5.601)	(1.801)
		SCR_2	0.833	0.234	54	3.434	8.300	5.455	0.829	0.228	53	4.052	9.744	5.939
			(0.059)	(0.072)	(3)	(1.060)	(3.301)	(0.916)	(0.061)	(0.073)	(3)	(1.825)	(4.367)	(1.531)
	20	SSPS	0.584	0.555	65	9.167	17.366	13.117	0.388	0.813	103	10.358	19.820	15.521
			(0.061)	(0.065)	(7)	(2.353)	(4.949)	(1.485)	(0.065)	(0.032)	(16)	(3.118)	(6.426)	(1.798)
		SCR_2	0.818	0.255	54	5.484	12.317	8.251	0.812	0.266	54	5.376	12.147	8.110
			(0.058)	(0.075)	(4)	(2.765)	(5.866)	(2.346)	(0.062)	(0.069)	(3)	(2.243)	(5.306)	(1.761)

Table 4.6: Average (SD) of $\hat{\Omega}$ related PDR, FDR, $|\hat{\Omega}|$ and matrix loss in three norms for block precision matrix design when $p = 200$, $n = 100$

			$h = 0.2$						$h = 0.5$					
Graph	q	Mtd.	PDR	FDR	$\lvert\hat{\Omega}\rvert$	$\lVert\cdot\rVert_S$	$\lVert\cdot\rVert_{\ell_1}$	$\lVert\cdot\rVert_F$	PDR	FDR	$\lvert\hat{\Omega}\rvert$	$\lVert\cdot\rVert_S$	$\lVert\cdot\rVert_{\ell_1}$	$\lVert\cdot\rVert_F$
AR(1)	10	SSPS	0.734	0.310	210	1.387	1.871	8.078	0.409	0.744	314	1.652	2.222	11.265
			(0.038)	(0.048)	(18)	(0.080)	(0.125)	(0.281)	(0.053)	(0.019)	(36)	(0.029)	(0.079)	(0.293)
		SCR_2	0.993	0.031	201	1.186	1.585	4.646	0.994	0.032	201	1.203	1.635	4.706
			(0.007)	(0.014)	(3)	(0.162)	(0.243)	(0.291)	(0.006)	(0.014)	(3)	(0.182)	(0.307)	(0.247)
	20	SSPS	0.746	0.191	181	1.348	1.737	8.270	0.318	0.733	234	1.675	2.113	12.334
			(0.038)	(0.043)	(14)	(0.083)	(0.123)	(0.262)	(0.041)	(0.027)	(25)	(0.020)	(0.074)	(0.183)
		SCR_2	0.989	0.061	207	1.689	2.257	7.305	0.988	0.062	207	1.695	2.318	7.291
			(0.009)	(0.021)	(5)	(0.249)	(0.414)	(0.398)	(0.009)	(0.022)	(5)	(0.236)	(0.398)	(0.371)
AR(2)	10	SSPS	0.314	0.299	175	1.762	2.362	10.096	0.148	0.784	265	2.068	2.667	12.353
			(0.029)	(0.045)	(17)	(0.059)	(0.115)	(0.263)	(0.026)	(0.026)	(28)	(0.025)	(0.085)	(0.193)
		SCR_2	0.513	0.052	210	1.259	1.972	7.006	0.512	0.052	210	1.260	1.992	7.003
			(0.021)	(0.017)	(10)	(0.187)	(0.306)	(0.140)	(0.021)	(0.018)	(11)	(0.148)	(0.302)	(0.153)
	20	SSPS	0.305	0.205	149	1.775	2.300	10.309	0.106	0.788	194	2.112	2.580	13.117
			(0.024)	(0.041)	(14)	(0.052)	(0.087)	(0.207)	(0.019)	(0.029)	(28)	(0.019)	(0.080)	(0.149)
		SCR_2	0.511	0.093	219	1.726	2.749	8.292	0.504	0.092	215	1.716	2.712	8.234
			(0.021)	(0.027)	(13)	(0.285)	(0.425)	(0.286)	(0.020)	(0.026)	(11)	(0.239)	(0.444)	(0.253)
Circle	10	SSPS	0.716	0.299	206	1.374	1.854	8.108	0.403	0.745	316	1.656	2.231	11.241
			(0.040)	(0.048)	(19)	(0.088)	(0.128)	(0.309)	(0.047)	(0.022)	(32)	(0.027)	(0.085)	(0.259)
		SCR_2	0.974	0.036	202	1.186	1.577	4.655	0.973	0.034	201	1.190	1.589	4.682
			(0.006)	(0.017)	(4)	(0.190)	(0.279)	(0.291)	(0.006)	(0.015)	(3)	(0.183)	(0.293)	(0.315)
	20	SSPS	0.747	0.195	186	1.323	1.739	8.182	0.316	0.734	238	1.678	2.099	12.323
			(0.042)	(0.040)	(14)	(0.073)	(0.121)	(0.297)	(0.041)	(0.024)	(29)	(0.020)	(0.060)	(0.189)
		SCR_2	0.969	0.057	206	1.693	2.255	7.277	0.969	0.061	207	1.678	2.244	7.255
			(0.009)	(0.021)	(5)	(0.237)	(0.360)	(0.402)	(0.008)	(0.023)	(5)	(0.233)	(0.373)	(0.378)
Cluster	10	SSPS	0.057	0.684	73	2.581	2.961	15.427	0.040	0.926	216	2.738	3.232	16.181
			(0.016)	(0.061)	(13)	(0.013)	(0.082)	(0.116)	(0.011)	(0.019)	(16)	(0.013)	(0.068)	(0.085)
		SCR_2	0.146	0.267	79	2.408	2.834	13.915	0.139	0.276	77	2.396	2.829	13.951
			(0.035)	(0.057)	(17)	(0.108)	(0.199)	(0.238)	(0.037)	(0.071)	(17)	(0.023)	(0.165)	(0.258)
	20	SSPS	0.045	0.639	50	2.580	2.893	15.554	0.026	0.937	165	2.753	3.167	16.572
			(0.011)	(0.074)	(9)	(0.013)	(0.065)	(0.076)	(0.009)	(0.020)	(17)	(0.010)	(0.069)	(0.071)
		SCR_2	0.158	0.369	100	2.387	3.265	13.781	0.156	0.391	102	2.421	3.327	13.828
			(0.035)	(0.057)	(18)	(0.227)	(0.375)	(0.217)	(0.028)	(0.053)	(15)	(0.350)	(0.422)	(0.193)

Continued on next page

Table 4.6: Average (SD) of $\hat{\Omega}$ related PDR, FDR, $|\hat{\Omega}|$ and matrix loss in three norms for block precision matrix design when $p = 200$, $n = 100$

			$h = 0.2$						$h = 0.5$					
Graph	q	Mtd.	PDR	FDR	$\lvert\hat{\Omega}\rvert$	$\lVert\cdot\rVert_S$	$\lVert\cdot\rVert_{\ell_1}$	$\lVert\cdot\rVert_F$	PDR	FDR	$\lvert\hat{\Omega}\rvert$	$\lVert\cdot\rVert_S$	$\lVert\cdot\rVert_{\ell_1}$	$\lVert\cdot\rVert_F$
ER	10	SSPS	0.351	0.440	154	2.588	4.357	13.014	0.188	0.826	263	2.957	5.153	16.361
			(0.075)	(0.061)	(24)	(0.271)	(0.638)	(1.113)	(0.044)	(0.026)	(34)	(0.247)	(0.686)	(1.106)
		SCR_2	0.641	0.099	175	2.091	3.405	10.086	0.655	0.096	178	2.079	3.380	10.000
			(0.104)	(0.035)	(21)	(0.325)	(0.640)	(1.281)	(0.110)	(0.033)	(22)	(0.346)	(0.724)	(1.245)
	20	SSPS	0.335	0.359	131	2.682	4.484	13.366	0.138	0.821	191	3.078	5.104	17.482
			(0.069)	(0.066)	(18)	(0.264)	(0.664)	(1.136)	(0.033)	(0.039)	(33)	(0.234)	(0.641)	(1.096)
		SCR_2	0.612	0.163	183	2.932	4.742	13.818	0.631	0.162	184	2.851	4.423	13.489
			(0.097)	(0.047)	(22)	(0.369)	(0.831)	(1.307)	(0.100)	(0.043)	(21)	(0.366)	(0.716)	(1.340)
Tridiag	10	SSPS	0.544	0.309	228	4.507	5.994	24.527	0.351	0.714	363	5.464	7.560	33.434
			(0.046)	(0.041)	(15)	(0.339)	(0.560)	(1.215)	(0.037)	(0.019)	(33)	(0.281)	(0.480)	(1.157)
		SCR_2	0.674	0.021	199	3.519	4.550	13.046	0.662	0.021	199	3.447	4.442	13.117
			(0.053)	(0.010)	(3)	(0.571)	(0.788)	(0.886)	(0.039)	(0.011)	(2)	(0.588)	(0.763)	(0.933)
	20	SSPS	0.563	0.194	203	4.486	5.676	25.326	0.306	0.684	284	5.693	7.411	37.236
			(0.039)	(0.038)	(11)	(0.334)	(0.507)	(1.165)	(0.030)	(0.028)	(26)	(0.265)	(0.395)	(1.245)
		SCR_2	0.668	0.040	202	4.763	6.139	20.127	0.666	0.038	202	5.071	6.484	20.282
			(0.041)	(0.015)	(3)	(0.597)	(0.877)	(1.176)	(0.038)	(0.019)	(4)	(0.763)	(1.050)	(1.227)
Hub	10	SSPS	0.298	0.511	111	6.520	12.113	23.350	0.164	0.877	240	7.614	14.545	29.275
			(0.035)	(0.059)	(13)	(0.429)	(0.804)	(1.116)	(0.026)	(0.017)	(20)	(0.370)	(0.752)	(1.047)
		SCR_2	0.562	0.197	126	4.500	8.910	13.325	0.560	0.188	124	4.536	9.030	13.477
			(0.044)	(0.036)	(11)	(1.141)	(1.445)	(0.820)	(0.048)	(0.036)	(11)	(0.919)	(1.227)	(0.915)
	20	SSPS	0.291	0.426	92	6.602	12.236	24.125	0.125	0.877	183	7.879	14.528	31.486
			(0.029)	(0.053)	(11)	(0.371)	(0.726)	(0.926)	(0.024)	(0.022)	(19)	(0.337)	(0.628)	(1.003)
		SCR_2	0.518	0.288	131	5.799	11.076	16.344	0.521	0.288	132	6.157	11.582	16.732
			(0.039)	(0.044)	(11)	(1.491)	(1.991)	(1.218)	(0.042)	(0.046)	(11)	(1.631)	(2.264)	(1.538)
BA	10	SSPS	0.496	0.552	218	9.375	17.737	26.215	0.333	0.826	377	10.328	19.969	29.552
			(0.041)	(0.048)	(19)	(2.477)	(5.127)	(3.111)	(0.039)	(0.018)	(49)	(2.618)	(5.504)	(3.045)
		SCR_2	0.660	0.247	172	6.664	13.538	15.218	0.643	0.264	172	6.962	14.246	15.689
			(0.058)	(0.069)	(7)	(2.205)	(4.799)	(3.194)	(0.054)	(0.069)	(8)	(2.423)	(5.215)	(3.484)
	20	SSPS	0.512	0.471	190	9.733	18.126	27.532	0.287	0.812	300	10.883	20.586	31.853
			(0.046)	(0.059)	(13)	(2.636)	(5.470)	(3.174)	(0.040)	(0.025)	(34)	(3.371)	(6.946)	(4.053)
		SCR_2	0.626	0.282	172	6.701	14.359	16.470	0.620	0.283	170	6.972	15.026	16.697
			(0.052)	(0.072)	(9)	(2.360)	(5.224)	(3.075)	(0.066)	(0.082)	(8)	(2.998)	(6.531)	(3.926)

Table 4.7: Average (SD) of $\hat{\Omega}$ related PDR, FDR, $|\hat{\Omega}|$ and matrix loss in three norms for noise precision matrix design when $p = 200$, $n = 100$

			$h = 0.2$						$h = 0.5$					
Graph	q	Mtd.	PDR	FDR	$\lvert\hat{\Omega}\rvert$	$\lVert\cdot\rVert_S$	$\lVert\cdot\rVert_{\ell_1}$	$\lVert\cdot\rVert_F$	PDR	FDR	$\lvert\hat{\Omega}\rvert$	$\lVert\cdot\rVert_S$	$\lVert\cdot\rVert_{\ell_1}$	$\lVert\cdot\rVert_F$
AR(1)	10	SSPS	0.778	0.583	93	1.244	1.778	5.358	0.453	0.897	216	1.573	2.294	8.094
			(0.056)	(0.064)	(14)	(0.096)	(0.190)	(0.225)	(0.087)	(0.017)	(30)	(0.052)	(0.130)	(0.130)
		SCR_2	0.991	0.322	72	1.385	1.795	5.015	0.991	0.315	71	1.342	1.757	5.016
			(0.015)	(0.054)	(6)	(0.219)	(0.349)	(0.366)	(0.014)	(0.050)	(5)	(0.185)	(0.295)	(0.364)
	20	SSPS	0.782	0.466	73	1.212	1.636	5.162	0.368	0.882	154	1.625	2.167	8.311
			(0.060)	(0.061)	(9)	(0.091)	(0.158)	(0.183)	(0.069)	(0.020)	(25)	(0.038)	(0.097)	(0.130)
		SCR_2	0.986	0.451	89	1.996	2.672	8.090	0.984	0.456	89	1.921	2.566	8.063
			(0.017)	(0.046)	(8)	(0.303)	(0.463)	(0.521)	(0.017)	(0.050)	(8)	(0.300)	(0.429)	(0.515)
AR(2)	10	SSPS	0.334	0.592	81	1.675	2.297	6.139	0.177	0.918	210	1.974	2.759	8.495
			(0.039)	(0.073)	(15)	(0.082)	(0.160)	(0.182)	(0.041)	(0.017)	(27)	(0.047)	(0.109)	(0.110)
		SCR_2	0.518	0.330	76	1.357	1.951	5.662	0.520	0.340	77	1.365	1.987	5.700
			(0.034)	(0.064)	(8)	(0.193)	(0.287)	(0.362)	(0.030)	(0.055)	(7)	(0.218)	(0.360)	(0.325)
	20	SSPS	0.331	0.492	65	1.680	2.216	6.006	0.132	0.914	150	2.040	2.658	8.615
			(0.040)	(0.075)	(12)	(0.068)	(0.140)	(0.164)	(0.038)	(0.021)	(26)	(0.035)	(0.097)	(0.108)
		SCR_2	0.511	0.464	93	1.975	2.752	8.318	0.511	0.466	94	1.956	2.821	8.324
			(0.031)	(0.053)	(9)	(0.324)	(0.512)	(0.541)	(0.038)	(0.054)	(11)	(0.275)	(0.444)	(0.551)
Circle	10	SSPS	0.740	0.583	91	1.256	1.757	5.386	0.453	0.896	219	1.570	2.265	8.096
			(0.056)	(0.063)	(13)	(0.097)	(0.179)	(0.232)	(0.067)	(0.015)	(26)	(0.047)	(0.110)	(0.114)
		SCR_2	0.972	0.317	72	1.386	1.805	5.015	0.973	0.324	72	1.358	1.759	5.029
			(0.015)	(0.056)	(6)	(0.240)	(0.382)	(0.377)	(0.012)	(0.050)	(5)	(0.189)	(0.302)	(0.404)
	20	SSPS	0.760	0.471	73	1.237	1.670	5.215	0.355	0.886	158	1.623	2.181	8.314
			(0.059)	(0.070)	(10)	(0.089)	(0.186)	(0.187)	(0.060)	(0.023)	(25)	(0.038)	(0.113)	(0.111)
		SCR_2	0.965	0.444	87	1.962	2.600	8.011	0.967	0.451	89	1.960	2.659	8.107
			(0.017)	(0.050)	(7)	(0.290)	(0.471)	(0.506)	(0.016)	(0.047)	(8)	(0.265)	(0.462)	(0.496)
Cluster	10	SSPS	0.075	0.868	57	2.528	2.979	8.494	0.049	0.976	209	2.633	3.353	10.073
			(0.033)	(0.050)	(12)	(0.025)	(0.117)	(0.129)	(0.023)	(0.011)	(20)	(0.023)	(0.105)	(0.075)
		SCR_2	0.190	0.510	39	2.332	2.784	7.871	0.177	0.516	37	2.339	2.791	7.923
			(0.075)	(0.095)	(12)	(0.059)	(0.174)	(0.213)	(0.066)	(0.114)	(11)	(0.106)	(0.199)	(0.250)
	20	SSPS	0.056	0.853	38	2.538	2.915	8.412	0.036	0.978	162	2.675	3.270	10.048
			(0.028)	(0.064)	(9)	(0.022)	(0.112)	(0.101)	(0.019)	(0.012)	(20)	(0.017)	(0.082)	(0.068)
		SCR_2	0.213	0.648	60	2.293	3.229	9.481	0.194	0.668	58	2.298	3.264	9.544
			(0.076)	(0.080)	(13)	(0.261)	(0.424)	(0.403)	(0.077)	(0.086)	(15)	(0.236)	(0.460)	(0.408)

Continued on next page

Table 4.7: Average (SD) of $\hat{\Omega}$ related PDR, FDR, $|\hat{\Omega}|$ and matrix loss in three norms for noise precision matrix design when $p = 200$, $n = 100$

			$h = 0.2$						$h = 0.5$					
Graph	q	Mtd.	PDR	FDR	$\lvert\hat{\Omega}\rvert$	$\lVert\cdot\rVert_S$	$\lVert\cdot\rVert_{\ell_1}$	$\lVert\cdot\rVert_F$	PDR	FDR	$\lvert\hat{\Omega}\rvert$	$\lVert\cdot\rVert_S$	$\lVert\cdot\rVert_{\ell_1}$	$\lVert\cdot\rVert_F$
ER	10	SSPS	0.354	0.743	84	2.445	3.993	7.583	0.200	0.946	225	2.786	4.875	10.195
			(0.089)	(0.057)	(13)	(0.281)	(0.683)	(0.549)	(0.065)	(0.014)	(28)	(0.250)	(0.645)	(0.430)
		SCR_2	0.611	0.425	64	1.799	2.929	6.866	0.608	0.424	65	1.761	2.851	6.868
			(0.116)	(0.063)	(7)	(0.311)	(0.581)	(0.572)	(0.122)	(0.069)	(9)	(0.253)	(0.552)	(0.558)
	20	SSPS	0.348	0.663	64	2.462	3.930	7.474	0.160	0.941	166	2.881	4.768	10.300
			(0.087)	(0.073)	(10)	(0.276)	(0.667)	(0.528)	(0.047)	(0.018)	(25)	(0.242)	(0.604)	(0.493)
		SCR_2	0.579	0.556	81	2.516	3.946	9.984	0.599	0.555	81	2.464	3.794	9.824
			(0.099)	(0.054)	(9)	(0.381)	(0.681)	(0.689)	(0.108)	(0.059)	(10)	(0.395)	(0.739)	(0.601)
Tridiag	10	SSPS	0.519	0.618	100	4.212	5.303	13.226	0.363	0.896	255	5.054	6.825	17.685
			(0.053)	(0.048)	(11)	(0.439)	(0.572)	(0.703)	(0.056)	(0.014)	(20)	(0.357)	(0.495)	(0.718)
		SCR_2	0.647	0.313	69	2.823	3.599	8.119	0.643	0.300	67	2.863	3.610	8.001
			(0.052)	(0.050)	(5)	(0.629)	(0.829)	(0.788)	(0.044)	(0.048)	(5)	(0.614)	(0.792)	(0.716)
	20	SSPS	0.542	0.495	78	4.167	5.189	13.354	0.312	0.884	197	5.346	6.870	19.332
			(0.048)	(0.049)	(8)	(0.385)	(0.516)	(0.761)	(0.053)	(0.020)	(21)	(0.307)	(0.507)	(0.633)
		SCR_2	0.640	0.434	82	3.932	4.997	12.257	0.636	0.435	82	4.095	5.171	12.428
			(0.049)	(0.047)	(7)	(0.770)	(1.058)	(0.916)	(0.049)	(0.043)	(6)	(0.804)	(1.028)	(0.975)
Hub	10	SSPS	0.327	0.804	76	5.914	10.986	12.050	0.186	0.962	222	7.110	13.739	15.742
			(0.056)	(0.035)	(10)	(0.484)	(0.911)	(0.752)	(0.046)	(0.009)	(17)	(0.448)	(0.853)	(0.673)
		SCR_2	0.589	0.506	54	3.573	7.379	8.010	0.577	0.523	55	3.440	7.416	7.983
			(0.066)	(0.067)	(8)	(1.072)	(1.484)	(0.829)	(0.069)	(0.062)	(6)	(0.755)	(1.128)	(0.667)
	20	SSPS	0.328	0.716	53	6.025	11.088	12.161	0.139	0.963	171	7.489	13.962	16.601
			(0.053)	(0.050)	(9)	(0.511)	(0.999)	(0.655)	(0.046)	(0.012)	(19)	(0.415)	(0.743)	(0.571)
		SCR_2	0.554	0.665	75	4.792	9.565	11.242	0.530	0.670	73	4.590	9.322	11.031
			(0.069)	(0.045)	(9)	(1.371)	(2.032)	(1.027)	(0.063)	(0.045)	(8)	(1.257)	(1.802)	(0.863)
BA	10	SSPS	0.503	0.704	85	9.096	17.237	13.546	0.359	0.922	227	10.240	20.038	15.986
			(0.061)	(0.051)	(13)	(2.756)	(5.701)	(1.826)	(0.064)	(0.015)	(30)	(2.754)	(5.764)	(1.700)
		SCR_2	0.665	0.471	62	5.547	11.977	8.620	0.664	0.471	62	5.602	12.209	8.614
			(0.068)	(0.080)	(8)	(2.442)	(5.376)	(1.880)	(0.063)	(0.066)	(7)	(2.405)	(5.300)	(1.775)
	20	SSPS	0.528	0.624	70	9.584	17.869	14.025	0.314	0.900	159	10.245	19.400	16.663
			(0.067)	(0.075)	(10)	(2.709)	(5.581)	(1.676)	(0.054)	(0.023)	(27)	(2.612)	(5.409)	(1.499)
		SCR_2	0.652	0.552	73	5.549	12.918	10.403	0.650	0.551	72	5.108	11.759	10.190
			(0.075)	(0.087)	(12)	(2.385)	(5.544)	(1.463)	(0.070)	(0.076)	(11)	(1.881)	(4.394)	(1.146)

Table 4.8: Average (SD) of $\hat{\Omega}$ related PDR, FDR, $|\hat{\Omega}|$ and matrix loss in three norms and $\hat{B}$ related PDR, FDR and PMSE when $p = 50$, $n = 100$ and $h = 0.8$

			$\hat{\Omega}$						$\hat{B}$		
Graphs	X	Mtd.	PDR	FDR	$\lvert\hat{\Omega}\rvert$	$\lVert\cdot\rVert_S$	$\lVert\cdot\rVert_{\ell_1}$	$\lVert\cdot\rVert_F$	PDR	FDR	PMSE
AR(1)	I	MRCE	0.570	0.894	396	3.142	3.496	8.164	0.315	0.337	463.671
			(0.237)	(0.065)	(271)	(1.879)	(1.988)	(1.094)	(0.338)	(0.321)	(108.399)
		aMCR	0.887	0.695	145	-	-	-	0.998	0.663	155.381
			(0.044)	(0.043)	(21)	-	-	-	(0.003)	(0.033)	(5.914)
		SCR_3	0.973	0.070	51	1.502	1.917	3.424	0.993	0.077	127.336
			(0.027)	(0.037)	(2)	(0.366)	(0.435)	(0.449)	(0.011)	(0.015)	(4.492)
		SCR_3^*	0.967	0.076	51	1.321	1.713	3.053	0.990	0.077	128.802
			(0.030)	(0.038)	(2)	(0.239)	(0.315)	(0.367)	(0.012)	(0.015)	(4.600)
		SCR_4	0.981	0.369	77	1.560	2.181	3.522	0.983	0.078	130.770
			(0.021)	(0.055)	(7)	(0.349)	(0.474)	(0.434)	(0.018)	(0.014)	(6.327)
	II	MRCE	0.427	0.869	192	3.588	3.937	8.346	0.218	0.306	453.568
			(0.143)	(0.036)	(156)	(2.132)	(2.335)	(1.074)	(0.099)	(0.189)	(50.010)
		aMCR	0.807	0.718	143	-	-	-	0.887	0.677	170.376
			(0.066)	(0.041)	(21)	-	-	-	(0.021)	(0.025)	(6.426)
		SCR_3	0.621	0.347	47	1.531	1.989	4.949	0.621	0.125	194.451
			(0.076)	(0.079)	(5)	(0.145)	(0.208)	(0.332)	(0.048)	(0.022)	(12.072)
		SCR_3^*	0.557	0.352	42	1.537	1.891	5.228	0.584	0.116	201.877
			(0.073)	(0.077)	(5)	(0.075)	(0.128)	(0.270)	(0.038)	(0.021)	(10.003)
		SCR_4	0.707	0.569	81	1.506	1.985	5.069	0.581	0.122	203.931
			(0.072)	(0.050)	(9)	(0.057)	(0.123)	(0.281)	(0.041)	(0.021)	(11.127)
	III	MRCE	0.947	0.958	1109	14.683	17.495	18.430	0.115	0.373	518.764
			(0.034)	(0.002)	(22)	(18.356)	(20.421)	(16.348)	(0.025)	(0.263)	(15.738)
		aMCR	0.850	0.714	147	-	-	-	0.960	0.673	162.134
			(0.057)	(0.031)	(19)	-	-	-	(0.010)	(0.028)	(5.490)
		SCR_3	0.901	0.119	50	1.515	1.968	3.653	0.940	0.088	137.825
			(0.055)	(0.051)	(3)	(0.518)	(0.639)	(0.494)	(0.025)	(0.017)	(6.358)
		SCR_3^*	0.851	0.153	49	1.290	1.713	3.460	0.903	0.089	146.621
			(0.061)	(0.059)	(3)	(0.225)	(0.272)	(0.381)	(0.028)	(0.017)	(7.480)
		SCR_4	0.937	0.424	80	1.655	2.423	3.894	0.925	0.098	141.777
			(0.039)	(0.047)	(6)	(0.460)	(0.717)	(0.504)	(0.027)	(0.018)	(7.126)
	IV	MRCE	0.425	0.940	345	1.756	2.027	6.925	0.008	0.197	576.900
			(0.089)	(0.010)	(29)	(0.027)	(0.033)	(0.131)	(0.007)	(0.328)	(47.607)
		aMCR	0.981	0.676	154	-	-	-	0.619	0.625	141.103
			(0.020)	(0.062)	(31)	-	-	-	(0.042)	(0.046)	(6.237)
		SCR_3	0.944	0.142	54	1.250	1.717	2.888	0.583	0.267	152.838
			(0.037)	(0.060)	(3)	(0.305)	(0.395)	(0.377)	(0.056)	(0.048)	(14.707)
		SCR_3^*	0.911	0.180	55	1.157	1.691	2.792	0.570	0.266	149.923
			(0.044)	(0.081)	(5)	(0.158)	(0.280)	(0.299)	(0.053)	(0.045)	(10.737)
		SCR_4	0.903	0.617	119	1.332	2.352	3.686	0.399	0.359	194.697
			(0.050)	(0.083)	(18)	(0.155)	(0.417)	(0.419)	(0.078)	(0.060)	(27.699)
ER	I	MRCE	0.183	0.801	69	3.982	5.088	11.847	0.071	0.267	286.144
			(0.114)	(0.068)	(84)	(1.060)	(0.881)	(0.942)	(0.071)	(0.315)	(36.417)
		aMCR	0.616	0.694	123	-	-	-	0.998	0.667	81.345
			(0.092)	(0.043)	(21)	-	-	-	(0.003)	(0.027)	(10.440)
		SCR_3	0.691	0.165	50	2.675	3.979	6.546	0.978	0.080	68.800
			(0.103)	(0.059)	(7)	(0.700)	(1.075)	(0.971)	(0.016)	(0.015)	(8.354)
		SCR_3^*	0.671	0.168	49	2.315	3.512	6.018	0.976	0.077	69.561
			(0.100)	(0.060)	(7)	(0.457)	(0.797)	(0.789)	(0.016)	(0.015)	(8.843)
		SCR_4	0.777	0.438	84	2.685	4.290	6.808	0.961	0.079	71.484
			(0.089)	(0.053)	(10)	(0.529)	(0.842)	(0.862)	(0.025)	(0.016)	(9.197)
	II	MRCE	0.154	0.788	46	3.611	5.084	11.346	0.115	0.112	264.267
			(0.060)	(0.061)	(19)	(0.959)	(0.786)	(0.786)	(0.066)	(0.127)	(36.413)
		aMCR	0.524	0.721	116	-	-	-	0.879	0.667	86.329
			(0.078)	(0.040)	(20)	-	-	-	(0.022)	(0.023)	(10.364)
		SCR_3	0.320	0.490	38	2.653	4.266	7.667	0.569	0.122	100.020
			(0.088)	(0.097)	(6)	(0.501)	(0.851)	(0.744)	(0.038)	(0.024)	(12.177)
		SCR_3^*	0.271	0.514	34	2.695	4.312	7.791	0.548	0.112	103.292
			(0.077)	(0.103)	(6)	(0.269)	(0.630)	(0.668)	(0.036)	(0.020)	(12.642)
		SCR_4	0.393	0.652	69	2.638	4.346	7.678	0.536	0.121	105.079
			(0.087)	(0.058)	(10)	(0.255)	(0.641)	(0.630)	(0.035)	(0.020)	(13.539)

Continued on next page

Table 4.8: Average (SD) of $\hat{\Omega}$ related PDR, FDR, $|\hat{\Omega}|$ and matrix loss in three norms and $\hat{B}$ related PDR, FDR and PMSE when $p = 50$, $n = 100$ and $h = 0.8$

			$\hat{\Omega}$						$\hat{B}$		
Graphs	X	Mtd.	PDR	FDR	$\lvert\hat{\Omega}\rvert$	$\lVert\cdot\rVert_S$	$\lVert\cdot\rVert_{\ell_1}$	$\lVert\cdot\rVert_F$	PDR	FDR	PMSE
	III	MRCE	0.161	0.781	61	4.952	5.794	12.125	0.123	0.396	276.151
			(0.127)	(0.074)	(142)	(1.667)	(1.258)	(1.205)	(0.050)	(0.241)	(39.929)
		aMCR	0.575	0.709	122	-	-	-	0.962	0.673	84.158
			(0.073)	(0.046)	(20)	-	-	-	(0.010)	(0.026)	(10.420)
		SCR_3	0.575	0.243	46	2.399	3.804	6.452	0.876	0.091	76.394
			(0.108)	(0.072)	(6)	(0.532)	(0.882)	(0.840)	(0.028)	(0.017)	(8.741)
		SCR_3^*	0.507	0.282	43	2.192	3.603	6.229	0.854	0.089	79.776
			(0.097)	(0.076)	(6)	(0.331)	(0.712)	(0.693)	(0.027)	(0.016)	(9.852)
		SCR_4	0.669	0.497	81	2.715	4.519	6.811	0.858	0.097	78.457
			(0.095)	(0.056)	(10)	(0.716)	(1.296)	(0.886)	(0.033)	(0.019)	(8.979)
	IV	MRCE	0.108	0.706	42	4.231	5.606	9.205	0.553	0.625	152.464
			(0.061)	(0.222)	(43)	(1.922)	(1.411)	(1.599)	(0.132)	(0.061)	(74.110)
		aMCR	0.804	0.653	147	-	-	-	0.614	0.631	72.736
			(0.080)	(0.055)	(25)	-	-	-	(0.038)	(0.038)	(9.446)
		SCR_3	0.608	0.267	52	2.235	3.684	5.939	0.553	0.260	77.853
			(0.103)	(0.079)	(7)	(0.547)	(0.911)	(0.842)	(0.056)	(0.049)	(11.467)
		SCR_3^*	0.569	0.282	49	2.085	3.576	5.642	0.551	0.253	76.609
			(0.096)	(0.081)	(7)	(0.348)	(0.676)	(0.663)	(0.054)	(0.048)	(11.108)
		SCR_4	0.617	0.654	114	2.350	4.970	6.504	0.360	0.349	102.691
			(0.094)	(0.084)	(18)	(0.306)	(1.249)	(0.669)	(0.078)	(0.054)	(17.474)
Tridiag	I	MRCE	0.049	0.803	20	6.243	7.232	23.330	0.113	0.175	239.109
			(0.048)	(0.094)	(35)	(0.259)	(0.354)	(0.691)	(0.112)	(0.269)	(18.320)
		aMCR	0.626	0.668	139	-	-	-	0.998	0.627	66.626
			(0.046)	(0.035)	(17)	-	-	-	(0.003)	(0.038)	(2.567)
		SCR_3	0.654	0.056	50	4.282	5.404	9.620	0.991	0.082	57.116
			(0.049)	(0.035)	(2)	(0.980)	(1.247)	(1.531)	(0.011)	(0.015)	(2.367)
		SCR_3^*	0.647	0.066	50	3.771	4.882	8.454	0.988	0.085	58.003
			(0.049)	(0.040)	(2)	(0.746)	(1.038)	(1.306)	(0.014)	(0.015)	(2.677)
		SCR_4	0.665	0.363	77	4.081	5.514	9.201	0.981	0.081	58.728
			(0.048)	(0.056)	(6)	(0.809)	(1.036)	(1.308)	(0.021)	(0.015)	(3.333)
	II	MRCE	0.030	0.813	13	6.167	7.173	22.953	0.208	0.108	202.526
			(0.022)	(0.162)	(6)	(0.307)	(0.417)	(0.725)	(0.081)	(0.121)	(18.534)
		aMCR	0.597	0.701	147	-	-	-	0.886	0.653	74.018
			(0.053)	(0.036)	(21)	-	-	-	(0.021)	(0.028)	(3.057)
		SCR_3	0.484	0.328	52	4.985	6.514	14.232	0.687	0.119	81.126
			(0.065)	(0.081)	(5)	(0.481)	(0.710)	(1.190)	(0.059)	(0.018)	(5.876)
		SCR_3^*	0.428	0.361	49	5.178	6.462	15.745	0.614	0.117	87.634
			(0.061)	(0.075)	(5)	(0.360)	(0.525)	(1.056)	(0.040)	(0.021)	(5.215)
		SCR_4	0.516	0.598	93	5.056	6.643	15.359	0.616	0.126	88.334
			(0.057)	(0.042)	(9)	(0.382)	(0.613)	(1.022)	(0.044)	(0.021)	(5.324)
	III	MRCE	0.374	0.697	247	10.498	12.705	25.806	0.529	0.455	158.372
			(0.380)	(0.328)	(299)	(5.912)	(7.122)	(5.182)	(0.406)	(0.384)	(70.428)
		aMCR	0.614	0.680	140	-	-	-	0.961	0.649	70.220
			(0.058)	(0.035)	(15)	-	-	-	(0.013)	(0.028)	(3.340)
		SCR_3	0.640	0.087	51	4.419	5.600	9.839	0.967	0.086	58.906
			(0.054)	(0.046)	(2)	(1.107)	(1.329)	(1.374)	(0.019)	(0.017)	(2.830)
		SCR_3^*	0.601	0.146	51	4.012	5.272	9.438	0.928	0.096	63.365
			(0.058)	(0.059)	(3)	(0.678)	(0.825)	(1.185)	(0.025)	(0.017)	(3.313)
		SCR_4	0.659	0.406	81	4.575	6.629	10.587	0.954	0.098	60.349
			(0.054)	(0.058)	(7)	(1.301)	(1.987)	(1.462)	(0.022)	(0.018)	(3.102)
	IV	MRCE	0.015	0.453	15	5.533	6.560	19.324	0.629	0.610	84.470
			(0.033)	(0.469)	(32)	(0.344)	(0.494)	(0.997)	(0.104)	(0.050)	(37.701)
		aMCR	0.695	0.649	149	-	-	-	0.627	0.613	62.991
			(0.051)	(0.074)	(34)	-	-	-	(0.037)	(0.046)	(3.297)
		SCR_3	0.650	0.112	53	3.667	4.830	7.934	0.625	0.253	67.471
			(0.050)	(0.063)	(3)	(0.984)	(1.298)	(1.150)	(0.058)	(0.046)	(7.004)
		SCR_3^*	0.631	0.171	55	3.545	5.010	7.965	0.606	0.254	66.637
			(0.050)	(0.072)	(4)	(0.781)	(1.272)	(1.217)	(0.056)	(0.049)	(5.006)
		SCR_4	0.639	0.608	122	4.350	6.882	11.105	0.431	0.339	87.151
			(0.056)	(0.090)	(21)	(0.626)	(1.326)	(1.770)	(0.088)	(0.055)	(14.028)

Continued on next page

Table 4.8: Average (SD) of $\hat{\Omega}$ related PDR, FDR, $|\hat{\Omega}|$ and matrix loss in three norms and $\hat{B}$ related PDR, FDR and PMSE when $p = 50$, $n = 100$ and $h = 0.8$

			$\hat{\Omega}$						$\hat{B}$		
Graphs	X	Mtd.	PDR	FDR	$\lvert\hat{\Omega}\rvert$	$\lVert\cdot\rVert_S$	$\lVert\cdot\rVert_{\ell_1}$	$\lVert\cdot\rVert_F$	PDR	FDR	PMSE
Hub	I	MRCE	0.014	0.995	74	8.724	15.182	18.338	0.028	0.272	334.997
			(0.047)	(0.008)	(74)	(0.429)	(0.698)	(1.228)	(0.020)	(0.357)	(13.534)
		aMCR	0.540	0.766	107	-	-	-	0.998	0.684	101.587
			(0.090)	(0.050)	(20)	-	-	-	(0.002)	(0.022)	(4.591)
		SCR_3	0.709	0.218	41	4.392	8.198	7.487	0.973	0.074	83.545
			(0.075)	(0.059)	(4)	(1.365)	(2.117)	(1.302)	(0.020)	(0.015)	(5.250)
		SCR_3^*	0.695	0.217	40	3.893	7.520	6.923	0.970	0.072	84.008
			(0.070)	(0.058)	(3)	(1.328)	(2.120)	(1.234)	(0.020)	(0.014)	(5.311)
		SCR_4	0.833	0.469	71	4.884	8.851	8.115	0.956	0.076	86.560
			(0.067)	(0.067)	(6)	(1.515)	(2.450)	(1.483)	(0.026)	(0.016)	(6.441)
	II	MRCE	0.005	0.997	63	8.621	15.058	18.280	0.038	0.164	331.542
			(0.012)	(0.007)	(19)	(0.387)	(0.630)	(0.728)	(0.016)	(0.177)	(11.942)
		aMCR	0.456	0.792	101	-	-	-	0.879	0.683	106.990
			(0.094)	(0.045)	(19)	-	-	-	(0.025)	(0.023)	(4.039)
		SCR_3	0.299	0.553	30	7.035	12.790	12.709	0.536	0.121	128.292
			(0.076)	(0.082)	(6)	(0.819)	(1.263)	(1.606)	(0.034)	(0.021)	(6.852)
		SCR_3^*	0.259	0.557	26	7.244	13.061	13.564	0.531	0.110	127.999
			(0.066)	(0.089)	(5)	(0.729)	(1.176)	(1.482)	(0.032)	(0.022)	(6.567)
		SCR_4	0.398	0.685	57	7.131	12.975	13.193	0.514	0.118	131.250
			(0.093)	(0.061)	(9)	(0.692)	(1.186)	(1.401)	(0.034)	(0.023)	(6.970)
	III	MRCE	0.335	0.984	468	9.913	16.262	16.907	0.083	0.712	328.580
			(0.387)	(0.017)	(482)	(3.080)	(3.173)	(2.530)	(0.023)	(0.189)	(12.035)
		aMCR	0.513	0.783	109	-	-	-	0.960	0.684	103.778
			(0.099)	(0.040)	(23)	-	-	-	(0.011)	(0.025)	(4.273)
		SCR_3	0.551	0.310	36	5.046	9.596	8.318	0.848	0.091	96.857
			(0.081)	(0.095)	(4)	(1.687)	(2.606)	(1.574)	(0.025)	(0.018)	(4.857)
		SCR_3^*	0.505	0.318	33	5.399	10.065	8.668	0.836	0.084	97.846
			(0.092)	(0.097)	(5)	(1.420)	(2.281)	(1.568)	(0.028)	(0.017)	(4.961)
		SCR_4	0.679	0.537	66	6.239	11.569	10.059	0.836	0.094	98.496
			(0.092)	(0.067)	(7)	(1.810)	(2.836)	(1.939)	(0.028)	(0.016)	(5.234)
	IV	MRCE	0.002	1.000	239	6.879	12.985	13.750	0.106	0.663	324.971
			(0.012)	(0.002)	(24)	(0.652)	(0.904)	(1.449)	(0.038)	(0.138)	(27.849)
		aMCR	0.742	0.744	133	-	-	-	0.605	0.642	88.239
			(0.069)	(0.041)	(21)	-	-	-	(0.038)	(0.039)	(4.426)
		SCR_3	0.637	0.322	42	3.803	7.659	6.913	0.528	0.265	92.621
			(0.081)	(0.081)	(5)	(1.075)	(1.904)	(1.108)	(0.049)	(0.042)	(6.769)
		SCR_3^*	0.596	0.330	40	4.191	8.213	7.216	0.528	0.260	91.562
			(0.071)	(0.079)	(4)	(1.145)	(1.928)	(1.224)	(0.050)	(0.043)	(6.166)
		SCR_4	0.662	0.687	98	4.977	9.684	8.752	0.349	0.350	121.847
			(0.087)	(0.075)	(13)	(1.201)	(2.095)	(1.632)	(0.075)	(0.050)	(17.219)
BA	I	MRCE	0.675	0.952	694	10.029	18.967	16.195	0.298	0.616	800.088
			(0.088)	(0.007)	(74)	(3.225)	(6.232)	(2.672)	(0.128)	(0.186)	(72.364)
		aMCR	0.697	0.830	202	-	-	-	0.998	0.648	245.365
			(0.060)	(0.019)	(17)	-	-	-	(0.003)	(0.044)	(17.651)
		SCR_3	0.761	0.278	52	4.485	10.072	6.982	0.998	0.086	205.652
			(0.068)	(0.088)	(3)	(1.774)	(4.187)	(1.466)	(0.006)	(0.017)	(14.062)
		SCR_3^*	0.730	0.314	53	5.250	11.064	7.206	0.995	0.093	209.772
			(0.076)	(0.092)	(3)	(1.988)	(4.406)	(1.749)	(0.008)	(0.019)	(14.525)
		SCR_4	0.819	0.582	97	4.685	9.954	6.729	0.983	0.085	214.496
			(0.052)	(0.051)	(9)	(2.173)	(4.301)	(1.866)	(0.017)	(0.014)	(18.008)
	II	MRCE	0.738	0.941	653	10.858	20.464	16.723	0.297	0.357	711.443
			(0.086)	(0.017)	(153)	(2.790)	(5.629)	(1.871)	(0.240)	(0.266)	(165.324)
		aMCR	0.651	0.846	208	-	-	-	0.887	0.667	270.467
			(0.077)	(0.023)	(21)	-	-	-	(0.027)	(0.035)	(19.685)
		SCR_3	0.646	0.430	56	6.654	13.621	9.653	0.780	0.120	268.890
			(0.093)	(0.097)	(5)	(2.564)	(5.074)	(2.264)	(0.055)	(0.021)	(22.052)
		SCR_3^*	0.479	0.587	57	9.334	17.833	13.425	0.656	0.123	308.791
			(0.082)	(0.079)	(4)	(2.501)	(5.105)	(1.851)	(0.044)	(0.020)	(22.665)
		SCR_4	0.647	0.720	114	9.258	17.729	13.094	0.663	0.130	308.565
			(0.087)	(0.043)	(10)	(2.656)	(5.287)	(2.046)	(0.054)	(0.019)	(24.597)

Continued on next page

Table 4.8: Average (SD) of $\hat{\Omega}$ related PDR, FDR, $|\hat{\Omega}|$ and matrix loss in three norms and $\hat{B}$ related PDR, FDR and PMSE when $p = 50$, $n = 100$ and $h = 0.8$

			$\hat{\Omega}$						$\hat{B}$		
Graphs	X	Mtd.	PDR	FDR	$\lvert\hat{\Omega}\rvert$	$\lVert\cdot\rVert_S$	$\lVert\cdot\rVert_{\ell_1}$	$\lVert\cdot\rVert_F$	PDR	FDR	PMSE
	III	MRCE	0.879	0.959	1092	22.066	30.758	28.650	0.107	0.683	889.077
			(0.085)	(0.011)	(161)	(21.659)	(22.684)	(23.228)	(0.168)	(0.236)	(113.131)
		aMCR	0.687	0.836	206	-	-	-	0.963	0.663	257.452
			(0.065)	(0.020)	(19)	-	-	-	(0.012)	(0.039)	(18.337)
		SCR_3	0.778	0.275	53	4.045	8.925	6.616	0.980	0.089	212.634
			(0.062)	(0.074)	(3)	(1.710)	(3.897)	(1.413)	(0.013)	(0.017)	(14.000)
		SCR_3^*	0.686	0.369	54	6.094	11.978	8.448	0.946	0.099	228.755
			(0.080)	(0.091)	(4)	(2.292)	(4.573)	(2.004)	(0.023)	(0.019)	(16.985)
		SCR_4	0.811	0.587	97	5.304	11.634	8.080	0.964	0.105	220.333
			(0.054)	(0.051)	(9)	(2.367)	(4.380)	(2.064)	(0.017)	(0.016)	(15.508)
	IV	MRCE	0.581	0.938	455	10.222	18.999	15.929	0.002	0.000	966.933
			(0.124)	(0.007)	(77)	(2.367)	(4.944)	(1.313)	(0.002)	(0.000)	(82.846)
		aMCR	0.796	0.778	181	-	-	-	0.641	0.625	232.100
			(0.065)	(0.046)	(29)	-	-	-	(0.043)	(0.049)	(17.666)
		SCR_3	0.745	0.311	53	4.969	10.120	7.158	0.661	0.263	245.231
			(0.073)	(0.088)	(4)	(1.989)	(4.350)	(1.725)	(0.066)	(0.053)	(28.170)
		SCR_3^*	0.683	0.402	57	6.271	12.097	8.747	0.626	0.266	243.749
			(0.087)	(0.107)	(5)	(2.289)	(4.818)	(2.069)	(0.065)	(0.049)	(21.246)
		SCR_4	0.720	0.732	135	7.332	14.393	10.779	0.444	0.340	330.514
			(0.075)	(0.058)	(18)	(2.257)	(4.830)	(1.814)	(0.084)	(0.055)	(51.697)

Table 4.9: Average (SD) of $\hat{\Omega}$ related PDR, FDR, $|\hat{\Omega}|$ and matrix loss in three norms and $\hat{B}$ related PDR, FDR and PMSE when $p = 50$, $n = 200$ and $h = 0.8$

			$\hat{\Omega}$						$\hat{B}$		
Graphs	X	Mtd.	PDR	FDR	$\lvert\hat{\Omega}\rvert$	$\lVert\cdot\rVert_S$	$\lVert\cdot\rVert_{\ell_1}$	$\lVert\cdot\rVert_F$	PDR	FDR	PMSE
AR(1)	I	MRCE	0.440	0.709	107	4.855	5.262	9.143	0.543	0.551	423.955
			(0.185)	(0.090)	(130)	(1.732)	(1.748)	(1.064)	(0.071)	(0.080)	(25.514)
		aMCR	0.999	0.497	99	-	-	-	1.000	0.396	123.770
			(0.004)	(0.057)	(11)	-	-	-	(0.000)	(0.055)	(3.124)
		SCR_3	1.000	0.036	51	0.650	0.814	1.540	1.000	0.045	118.240
			(0.000)	(0.027)	(1)	(0.140)	(0.177)	(0.188)	(0.000)	(0.009)	(2.514)
		SCR_3^*	1.000	0.036	51	0.640	0.800	1.509	1.000	0.045	118.298
			(0.000)	(0.028)	(1)	(0.143)	(0.178)	(0.183)	(0.000)	(0.009)	(2.502)
		SCR_4	1.000	0.230	64	0.710	0.946	1.668	1.000	0.043	118.178
			(0.000)	(0.053)	(4)	(0.151)	(0.192)	(0.206)	(0.000)	(0.009)	(2.519)
	II	MRCE	0.679	0.664	135	4.235	4.581	8.131	0.611	0.453	297.873
			(0.316)	(0.191)	(87)	(2.055)	(2.197)	(1.066)	(0.223)	(0.261)	(88.416)
		aMCR	0.995	0.546	109	-	-	-	0.975	0.517	131.844
			(0.010)	(0.048)	(12)	-	-	-	(0.010)	(0.044)	(3.534)
		SCR_3	0.997	0.050	51	0.766	0.993	1.701	0.987	0.048	120.379
			(0.008)	(0.029)	(2)	(0.200)	(0.284)	(0.227)	(0.012)	(0.011)	(3.295)
		SCR_3^*	0.994	0.059	52	0.770	1.033	1.667	0.976	0.054	122.663
			(0.012)	(0.035)	(2)	(0.203)	(0.291)	(0.242)	(0.016)	(0.011)	(4.013)
		SCR_4	0.998	0.283	69	0.832	1.169	1.839	0.977	0.049	122.373
			(0.006)	(0.053)	(5)	(0.186)	(0.269)	(0.236)	(0.014)	(0.011)	(3.583)
	III	MRCE	0.989	0.909	638	17.551	21.020	26.860	0.759	0.696	254.612
			(0.016)	(0.037)	(286)	(10.409)	(12.077)	(14.334)	(0.372)	(0.355)	(152.260)
		aMCR	0.998	0.529	105	-	-	-	0.992	0.464	127.793
			(0.007)	(0.045)	(10)	-	-	-	(0.004)	(0.054)	(3.192)
		SCR_3	1.000	0.040	51	0.683	0.852	1.587	0.999	0.044	118.917
			(0.000)	(0.025)	(1)	(0.155)	(0.200)	(0.212)	(0.002)	(0.011)	(2.776)
		SCR_3^*	1.000	0.039	51	0.640	0.795	1.498	0.999	0.048	119.378
			(0.002)	(0.026)	(1)	(0.134)	(0.175)	(0.206)	(0.003)	(0.011)	(2.703)
		SCR_4	1.000	0.234	64	0.759	1.059	1.735	0.999	0.047	119.070
			(0.000)	(0.052)	(4)	(0.197)	(0.272)	(0.245)	(0.003)	(0.011)	(2.812)
	IV	MRCE	0.390	0.945	350	1.767	2.021	6.980	0.008	0.127	569.789
			(0.086)	(0.010)	(25)	(0.027)	(0.031)	(0.109)	(0.007)	(0.274)	(31.026)
		aMCR	1.000	0.509	104	-	-	-	0.641	0.639	129.950
			(0.000)	(0.092)	(23)	-	-	-	(0.049)	(0.038)	(3.210)
		SCR_3	1.000	0.050	52	0.689	0.855	1.552	0.717	0.244	125.752
			(0.003)	(0.030)	(2)	(0.176)	(0.211)	(0.193)	(0.061)	(0.056)	(3.490)
		SCR_3^*	1.000	0.049	52	0.573	0.724	1.360	0.711	0.239	126.559
			(0.003)	(0.029)	(2)	(0.118)	(0.153)	(0.145)	(0.062)	(0.056)	(3.514)
		SCR_4	0.999	0.234	65	0.633	0.892	1.519	0.685	0.260	127.512
			(0.004)	(0.074)	(8)	(0.155)	(0.292)	(0.248)	(0.061)	(0.056)	(4.925)
ER	I	MRCE	0.165	0.663	39	4.785	5.565	12.276	0.130	0.483	277.828
			(0.116)	(0.123)	(52)	(1.581)	(1.192)	(1.192)	(0.110)	(0.335)	(42.420)
		aMCR	0.927	0.529	121	-	-	-	1.000	0.478	65.120
			(0.045)	(0.050)	(16)	-	-	-	(0.000)	(0.046)	(9.009)
		SCR_3	0.975	0.065	64	1.177	1.639	2.775	1.000	0.039	61.326
			(0.026)	(0.031)	(8)	(0.300)	(0.441)	(0.507)	(0.000)	(0.008)	(8.599)
		SCR_3^*	0.975	0.064	64	1.148	1.601	2.728	1.000	0.040	61.355
			(0.027)	(0.031)	(8)	(0.275)	(0.411)	(0.495)	(0.000)	(0.008)	(8.592)
		SCR_4	0.985	0.218	77	1.281	1.891	3.068	1.000	0.040	61.327
			(0.019)	(0.048)	(9)	(0.299)	(0.510)	(0.504)	(0.000)	(0.009)	(8.610)
	II	MRCE	0.117	0.604	19	4.289	5.370	11.174	0.368	0.245	219.201
			(0.071)	(0.104)	(16)	(1.661)	(1.247)	(1.097)	(0.103)	(0.197)	(36.882)
		aMCR	0.887	0.565	127	-	-	-	0.976	0.549	67.616
			(0.054)	(0.037)	(18)	-	-	-	(0.009)	(0.043)	(7.328)
		SCR_3	0.940	0.088	64	1.351	1.995	3.313	0.938	0.050	63.801
			(0.044)	(0.043)	(8)	(0.300)	(0.571)	(0.528)	(0.025)	(0.011)	(6.728)
		SCR_3^*	0.922	0.099	63	1.432	2.175	3.419	0.924	0.054	65.141
			(0.051)	(0.049)	(7)	(0.317)	(0.633)	(0.597)	(0.026)	(0.011)	(6.839)
		SCR_4	0.949	0.294	83	1.420	2.264	3.584	0.927	0.051	64.760
			(0.036)	(0.063)	(9)	(0.267)	(0.529)	(0.530)	(0.025)	(0.011)	(7.037)

Continued on next page

Table 4.9: Average (SD) of $\hat{\Omega}$ related PDR, FDR, $|\hat{\Omega}|$ and matrix loss in three norms and $\hat{B}$ related PDR, FDR and PMSE when $p = 50$, $n = 200$ and $h = 0.8$

			$\hat{\Omega}$						$\hat{B}$		
Graphs	X	Mtd.	PDR	FDR	$\lvert\hat{\Omega}\rvert$	$\lVert\cdot\rVert_S$	$\lVert\cdot\rVert_{\ell_1}$	$\lVert\cdot\rVert_F$	PDR	FDR	PMSE
	III	MRCE	0.585	0.792	437	6.199	7.267	12.346	0.525	0.548	205.306
			(0.409)	(0.169)	(358)	(2.250)	(2.396)	(1.606)	(0.242)	(0.292)	(45.832)
		aMCR	0.902	0.546	125	-	-	-	0.993	0.527	66.092
			(0.051)	(0.040)	(15)	-	-	-	(0.005)	(0.045)	(7.133)
		SCR_3	0.963	0.068	65	1.194	1.742	2.900	0.991	0.043	61.080
			(0.040)	(0.033)	(8)	(0.337)	(0.597)	(0.595)	(0.007)	(0.009)	(6.670)
		SCR_3^*	0.960	0.066	64	1.129	1.633	2.812	0.990	0.044	61.295
			(0.040)	(0.033)	(7)	(0.249)	(0.469)	(0.546)	(0.008)	(0.009)	(6.646)
		SCR_4	0.972	0.227	79	1.297	2.055	3.237	0.990	0.045	61.189
			(0.029)	(0.043)	(9)	(0.296)	(0.582)	(0.570)	(0.007)	(0.010)	(6.629)
	IV	MRCE	0.121	0.687	45	2.971	4.598	8.206	0.529	0.638	159.928
			(0.079)	(0.234)	(50)	(0.212)	(0.598)	(0.564)	(0.134)	(0.065)	(89.441)
		aMCR	0.964	0.508	121	-	-	-	0.629	0.640	68.798
			(0.034)	(0.081)	(24)	-	-	-	(0.041)	(0.038)	(9.816)
		SCR_3	0.953	0.088	63	1.181	1.742	2.814	0.679	0.233	66.771
			(0.049)	(0.041)	(7)	(0.314)	(0.549)	(0.595)	(0.050)	(0.051)	(9.505)
		SCR_3^*	0.941	0.086	62	1.089	1.639	2.704	0.668	0.231	67.291
			(0.051)	(0.041)	(7)	(0.269)	(0.466)	(0.587)	(0.049)	(0.048)	(9.628)
		SCR_4	0.960	0.256	80	1.204	2.056	2.987	0.646	0.254	68.446
			(0.036)	(0.097)	(16)	(0.344)	(1.009)	(0.691)	(0.057)	(0.047)	(10.267)
Tridiag	I	MRCE	0.006	0.705	2	6.254	7.584	22.163	0.667	0.658	155.660
			(0.011)	(0.387)	(2)	(0.479)	(0.864)	(0.789)	(0.055)	(0.061)	(11.967)
		aMCR	0.680	0.507	101	-	-	-	1.000	0.324	54.404
			(0.041)	(0.049)	(10)	-	-	-	(0.000)	(0.049)	(1.423)
		SCR_3	0.672	0.030	51	1.891	2.349	4.333	1.000	0.049	52.847
			(0.041)	(0.025)	(1)	(0.467)	(0.551)	(0.600)	(0.000)	(0.011)	(1.368)
		SCR_3^*	0.672	0.030	51	1.826	2.277	4.183	1.000	0.052	52.934
			(0.041)	(0.025)	(1)	(0.388)	(0.481)	(0.546)	(0.000)	(0.012)	(1.374)
		SCR_4	0.675	0.221	63	1.971	2.578	4.469	1.000	0.048	52.860
			(0.041)	(0.052)	(4)	(0.535)	(0.717)	(0.636)	(0.000)	(0.011)	(1.358)
	II	MRCE	0.042	0.377	5	6.113	7.246	21.888	0.573	0.315	135.803
			(0.137)	(0.453)	(16)	(0.434)	(0.745)	(0.792)	(0.152)	(0.302)	(24.631)
		aMCR	0.686	0.557	113	-	-	-	0.974	0.461	57.732
			(0.052)	(0.044)	(12)	-	-	-	(0.010)	(0.039)	(1.613)
		SCR_3	0.677	0.035	51	2.095	2.659	4.570	0.995	0.048	53.173
			(0.049)	(0.025)	(1)	(0.602)	(0.845)	(0.628)	(0.007)	(0.012)	(1.436)
		SCR_3^*	0.676	0.042	51	2.098	2.740	4.315	0.989	0.061	54.010
			(0.049)	(0.032)	(2)	(0.743)	(1.046)	(0.794)	(0.011)	(0.013)	(1.603)
		SCR_4	0.681	0.283	69	2.481	3.469	4.976	0.984	0.052	54.244
			(0.048)	(0.064)	(6)	(0.646)	(0.967)	(0.857)	(0.013)	(0.012)	(1.762)
	III	MRCE	0.730	0.776	253	15.694	18.262	24.376	0.943	0.872	75.425
			(0.056)	(0.057)	(74)	(2.298)	(2.735)	(2.470)	(0.034)	(0.021)	(9.295)
		aMCR	0.684	0.542	109	-	-	-	0.993	0.404	55.929
			(0.051)	(0.044)	(10)	-	-	-	(0.004)	(0.046)	(1.349)
		SCR_3	0.677	0.028	51	1.915	2.328	4.415	1.000	0.045	52.906
			(0.049)	(0.022)	(1)	(0.397)	(0.489)	(0.558)	(0.001)	(0.009)	(1.281)
		SCR_3^*	0.677	0.030	51	1.768	2.162	4.077	1.000	0.054	53.191
			(0.049)	(0.022)	(1)	(0.380)	(0.483)	(0.550)	(0.001)	(0.011)	(1.300)
		SCR_4	0.677	0.226	64	2.154	2.969	4.928	1.000	0.051	52.999
			(0.049)	(0.052)	(4)	(0.446)	(0.666)	(0.575)	(0.001)	(0.010)	(1.290)
	IV	MRCE	0.029	0.515	29	5.671	6.758	19.916	0.597	0.603	97.701
			(0.050)	(0.470)	(48)	(0.299)	(0.423)	(0.986)	(0.138)	(0.065)	(48.683)
		aMCR	0.682	0.463	96	-	-	-	0.647	0.626	57.374
			(0.043)	(0.094)	(21)	-	-	-	(0.044)	(0.040)	(1.664)
		SCR_3	0.672	0.033	51	1.966	2.434	4.361	0.753	0.240	55.441
			(0.042)	(0.027)	(1)	(0.412)	(0.521)	(0.554)	(0.054)	(0.049)	(1.598)
		SCR_3^*	0.671	0.038	51	1.737	2.193	3.866	0.738	0.246	56.162
			(0.043)	(0.025)	(1)	(0.450)	(0.621)	(0.494)	(0.054)	(0.046)	(1.657)
		SCR_4	0.676	0.236	65	1.863	2.552	4.321	0.709	0.263	56.391
			(0.043)	(0.065)	(6)	(0.501)	(0.798)	(0.680)	(0.052)	(0.048)	(1.733)

Continued on next page

Table 4.9: Average (SD) of $\hat{\Omega}$ related PDR, FDR, $|\hat{\Omega}|$ and matrix loss in three norms and $\hat{B}$ related PDR, FDR and PMSE when $p = 50$, $n = 200$ and $h = 0.8$

			$\hat{\Omega}$						$\hat{B}$		
Graphs	X	Mtd.	PDR	FDR	$\lvert\hat{\Omega}\rvert$	$\lVert\cdot\rVert_S$	$\lVert\cdot\rVert_{\ell_1}$	$\lVert\cdot\rVert_F$	PDR	FDR	PMSE
Hub	I	MRCE	0.009	0.994	41	8.681	15.088	18.052	0.039	0.445	332.114
			(0.036)	(0.018)	(60)	(0.443)	(0.684)	(1.144)	(0.016)	(0.388)	(11.137)
		aMCR	0.896	0.640	113	-	-	-	1.000	0.580	82.116
			(0.048)	(0.042)	(13)	-	-	-	(0.000)	(0.039)	(3.269)
		SCR_3	0.988	0.107	50	2.324	3.863	3.701	1.000	0.039	71.851
			(0.017)	(0.046)	(2)	(0.650)	(0.927)	(0.757)	(0.000)	(0.009)	(2.495)
		SCR_3^*	0.988	0.106	50	2.236	3.731	3.576	1.000	0.038	71.833
			(0.017)	(0.047)	(2)	(0.602)	(0.856)	(0.689)	(0.000)	(0.009)	(2.488)
		SCR_4	0.996	0.243	60	2.478	4.224	3.978	1.000	0.039	71.868
			(0.009)	(0.055)	(4)	(0.663)	(1.046)	(0.775)	(0.000)	(0.009)	(2.535)
	II	MRCE	0.010	0.998	34	8.342	14.695	17.554	0.081	0.256	324.390
			(0.038)	(0.008)	(72)	(0.423)	(0.659)	(0.848)	(0.032)	(0.185)	(13.014)
		aMCR	0.852	0.671	118	-	-	-	0.974	0.609	85.369
			(0.049)	(0.043)	(15)	-	-	-	(0.010)	(0.039)	(3.382)
		SCR_3	0.950	0.135	50	2.168	3.874	3.754	0.920	0.051	80.056
			(0.032)	(0.045)	(2)	(0.854)	(1.336)	(0.759)	(0.022)	(0.011)	(3.524)
		SCR_3^*	0.936	0.140	49	2.430	4.326	3.985	0.908	0.051	81.315
			(0.043)	(0.055)	(2)	(1.377)	(2.226)	(1.168)	(0.027)	(0.012)	(3.858)
		SCR_4	0.969	0.317	64	2.631	4.684	4.170	0.913	0.052	80.770
			(0.033)	(0.060)	(5)	(1.312)	(2.102)	(1.214)	(0.026)	(0.012)	(3.866)
	III	MRCE	0.402	0.982	468	8.137	14.462	16.531	0.079	0.529	330.040
			(0.448)	(0.020)	(503)	(0.507)	(0.754)	(1.185)	(0.013)	(0.254)	(11.178)
		aMCR	0.868	0.659	116	-	-	-	0.993	0.598	84.099
			(0.057)	(0.041)	(13)	-	-	-	(0.004)	(0.042)	(3.833)
		SCR_3	0.979	0.115	50	2.244	3.809	3.621	0.990	0.042	73.738
			(0.022)	(0.048)	(3)	(0.688)	(0.968)	(0.667)	(0.007)	(0.010)	(2.882)
		SCR_3^*	0.978	0.115	50	2.110	3.605	3.460	0.988	0.043	74.006
			(0.022)	(0.047)	(3)	(0.666)	(0.947)	(0.650)	(0.008)	(0.010)	(2.988)
		SCR_4	0.990	0.260	61	3.225	5.739	4.899	0.989	0.044	73.878
			(0.013)	(0.060)	(5)	(1.137)	(1.700)	(1.142)	(0.007)	(0.012)	(2.815)
	IV	MRCE	0.000	1.000	210	7.051	13.304	14.625	0.170	0.639	311.660
			(0.000)	(0.000)	(41)	(0.357)	(0.557)	(0.527)	(0.068)	(0.092)	(35.937)
		aMCR	0.922	0.607	108	-	-	-	0.627	0.656	80.408
			(0.049)	(0.058)	(15)	-	-	-	(0.035)	(0.034)	(2.952)
		SCR_3	0.967	0.132	50	2.144	3.679	3.462	0.663	0.228	77.599
			(0.026)	(0.046)	(2)	(0.594)	(0.895)	(0.637)	(0.046)	(0.044)	(2.863)
		SCR_3^*	0.959	0.132	50	2.042	3.525	3.326	0.657	0.227	77.763
			(0.030)	(0.048)	(2)	(0.638)	(0.982)	(0.645)	(0.049)	(0.044)	(2.808)
		SCR_4	0.983	0.276	63	2.114	3.765	3.430	0.640	0.243	78.913
			(0.019)	(0.105)	(13)	(0.789)	(1.293)	(0.782)	(0.056)	(0.045)	(4.609)
BA	I	MRCE	0.889	0.940	724	10.520	19.740	16.276	0.220	0.528	840.413
			(0.054)	(0.005)	(37)	(2.433)	(4.963)	(1.369)	(0.047)	(0.089)	(60.367)
		aMCR	0.873	0.773	190	-	-	-	1.000	0.331	201.050
			(0.051)	(0.028)	(18)	-	-	-	(0.000)	(0.064)	(11.635)
		SCR_3	0.939	0.163	55	2.127	4.224	3.208	1.000	0.048	194.849
			(0.043)	(0.067)	(3)	(1.248)	(2.605)	(1.096)	(0.000)	(0.014)	(11.715)
		SCR_3^*	0.933	0.173	56	2.252	4.454	3.203	1.000	0.060	195.799
			(0.043)	(0.070)	(3)	(1.201)	(2.554)	(1.066)	(0.000)	(0.015)	(11.805)
		SCR_4	0.951	0.355	73	1.820	3.975	2.815	1.000	0.051	195.730
			(0.026)	(0.056)	(5)	(0.611)	(1.549)	(0.580)	(0.000)	(0.012)	(11.851)
	II	MRCE	0.926	0.909	505	9.579	17.690	15.323	0.596	0.371	514.348
			(0.052)	(0.010)	(45)	(2.530)	(5.121)	(1.513)	(0.071)	(0.058)	(65.625)
		aMCR	0.861	0.790	203	-	-	-	0.974	0.476	214.592
			(0.053)	(0.025)	(17)	-	-	-	(0.009)	(0.047)	(14.819)
		SCR_3	0.936	0.175	56	2.201	4.276	3.305	0.995	0.048	197.489
			(0.041)	(0.065)	(3)	(1.248)	(2.620)	(1.067)	(0.007)	(0.012)	(13.615)
		SCR_3^*	0.909	0.223	58	3.311	6.192	4.181	0.989	0.064	200.786
			(0.057)	(0.086)	(4)	(2.173)	(4.289)	(2.019)	(0.011)	(0.015)	(13.958)
		SCR_4	0.948	0.387	77	2.498	4.870	3.648	0.985	0.056	201.981
			(0.031)	(0.067)	(8)	(1.499)	(2.830)	(1.314)	(0.013)	(0.013)	(14.475)

Continued on next page

Table 4.9: Average (SD) of $\hat{\Omega}$ related PDR, FDR, $|\hat{\Omega}|$ and matrix loss in three norms and $\hat{B}$ related PDR, FDR and PMSE when $p = 50$, $n = 200$ and $h = 0.8$

			$\hat{\Omega}$						$\hat{B}$		
Graphs	X	Mtd.	PDR	FDR	$\lvert\hat{\Omega}\rvert$	$\lVert\cdot\rVert_S$	$\lVert\cdot\rVert_{\ell_1}$	$\lVert\cdot\rVert_F$	PDR	FDR	PMSE
	III	MRCE	0.878	0.920	593	38.127	44.991	57.665	0.578	0.894	588.154
			(0.062)	(0.021)	(242)	(19.526)	(18.926)	(21.559)	(0.231)	(0.137)	(153.392)
		aMCR	0.874	0.782	198	-	-	-	0.993	0.433	209.271
			(0.065)	(0.027)	(15)	-	-	-	(0.004)	(0.061)	(14.865)
		SCR_3	0.939	0.171	56	2.121	4.167	3.254	1.000	0.049	197.153
			(0.053)	(0.071)	(3)	(1.462)	(3.412)	(1.249)	(0.001)	(0.010)	(14.110)
		SCR_3^*	0.928	0.187	56	2.449	4.680	3.422	1.000	0.061	198.536
			(0.055)	(0.077)	(3)	(1.685)	(3.660)	(1.469)	(0.001)	(0.014)	(14.409)
		SCR_4	0.946	0.359	73	3.050	6.533	4.508	1.000	0.065	198.522
			(0.033)	(0.060)	(6)	(1.338)	(3.015)	(1.184)	(0.002)	(0.014)	(14.273)
	IV	MRCE	0.526	0.939	425	10.764	20.035	16.418	0.001	0.000	955.081
			(0.105)	(0.008)	(60)	(2.369)	(4.923)	(1.256)	(0.002)	(0.000)	(74.004)
		aMCR	0.871	0.713	153	-	-	-	0.664	0.642	211.506
			(0.048)	(0.057)	(27)	-	-	-	(0.046)	(0.042)	(12.670)
		SCR_3	0.929	0.175	55	2.349	4.707	3.429	0.782	0.252	203.123
			(0.038)	(0.060)	(3)	(1.071)	(2.280)	(0.932)	(0.052)	(0.051)	(12.230)
		SCR_3^*	0.897	0.217	56	3.579	6.660	4.547	0.763	0.250	205.825
			(0.059)	(0.085)	(3)	(1.942)	(3.706)	(1.831)	(0.059)	(0.050)	(12.099)
		SCR_4	0.934	0.368	73	3.090	5.869	4.426	0.733	0.262	208.162
			(0.032)	(0.072)	(10)	(1.263)	(2.460)	(1.141)	(0.056)	(0.053)	(12.389)

Table 4.10: Average (SD) of $\hat{\Omega}$ related PDR, FDR, $|\hat{\Omega}|$ and matrix loss in three norms and $\hat{B}$ related PDR, FDR and PMSE when $p = 50$, $n = 100$ and $h = 0.6$

			$\hat{\Omega}$						$\hat{B}$		
Graphs	X	Mtd.	PDR	FDR	$\lvert\hat{\Omega}\rvert$	$\lVert\cdot\rVert_S$	$\lVert\cdot\rVert_{\ell_1}$	$\lVert\cdot\rVert_F$	PDR	FDR	PMSE
AR(1)	I	MRCE	0.964	0.949	919	25.530	30.186	27.000	0.378	0.642	238.886
			(0.028)	(0.002)	(31)	(9.670)	(10.066)	(9.588)	(0.037)	(0.051)	(6.966)
		aMCR	0.890	0.674	136	-	-	-	0.897	0.680	160.482
			(0.044)	(0.039)	(18)	-	-	-	(0.024)	(0.025)	(5.006)
		SCR_3	0.755	0.223	48	1.582	2.087	4.326	0.722	0.119	157.379
			(0.069)	(0.066)	(4)	(0.375)	(0.446)	(0.407)	(0.053)	(0.022)	(8.140)
		SCR_3^*	0.658	0.262	44	1.425	1.814	4.496	0.604	0.114	173.038
			(0.067)	(0.062)	(4)	(0.074)	(0.131)	(0.284)	(0.049)	(0.020)	(8.111)
		SCR_4	0.815	0.528	85	1.569	2.316	4.491	0.640	0.136	169.286
			(0.057)	(0.050)	(7)	(0.350)	(0.644)	(0.372)	(0.052)	(0.024)	(8.663)
	II	MRCE	0.931	0.895	717	16.325	19.432	18.607	0.490	0.647	194.063
			(0.181)	(0.169)	(193)	(10.628)	(11.760)	(9.837)	(0.100)	(0.144)	(18.775)
		aMCR	0.892	0.672	135	-	-	-	0.641	0.651	161.714
			(0.046)	(0.041)	(18)	-	-	-	(0.030)	(0.033)	(5.376)
		SCR_3	0.731	0.251	48	1.343	1.719	4.272	0.377	0.152	169.364
			(0.064)	(0.064)	(4)	(0.078)	(0.120)	(0.253)	(0.026)	(0.031)	(5.297)
		SCR_3^*	0.704	0.245	46	1.372	1.717	4.518	0.359	0.143	171.346
			(0.066)	(0.067)	(4)	(0.071)	(0.115)	(0.230)	(0.022)	(0.031)	(4.936)
		SCR_4	0.791	0.457	72	1.341	1.798	4.362	0.373	0.162	170.498
			(0.054)	(0.053)	(7)	(0.076)	(0.160)	(0.207)	(0.023)	(0.028)	(5.253)
	III	MRCE	0.970	0.944	853	26.755	31.359	28.517	0.486	0.671	212.734
			(0.025)	(0.003)	(51)	(7.598)	(7.832)	(7.948)	(0.054)	(0.046)	(8.818)
		aMCR	0.893	0.671	135	-	-	-	0.782	0.667	160.274
			(0.046)	(0.037)	(16)	-	-	-	(0.026)	(0.027)	(5.499)
		SCR_3	0.744	0.223	47	1.342	1.764	3.923	0.568	0.127	161.573
			(0.063)	(0.064)	(4)	(0.139)	(0.179)	(0.298)	(0.040)	(0.024)	(7.018)
		SCR_3^*	0.691	0.227	44	1.378	1.728	4.340	0.506	0.120	168.974
			(0.064)	(0.066)	(4)	(0.079)	(0.100)	(0.287)	(0.039)	(0.024)	(7.028)
		SCR_4	0.816	0.447	73	1.340	1.873	4.084	0.546	0.136	164.575
			(0.052)	(0.049)	(6)	(0.125)	(0.299)	(0.264)	(0.041)	(0.027)	(7.081)
	IV	MRCE	0.109	0.029	11	2.350	2.525	6.245	0.628	0.582	132.445
			(0.106)	(0.106)	(55)	(2.021)	(2.132)	(1.248)	(0.042)	(0.049)	(6.181)
		aMCR	0.991	0.612	129	-	-	-	0.512	0.623	133.468
			(0.015)	(0.070)	(25)	-	-	-	(0.039)	(0.042)	(4.696)
		SCR_3	0.930	0.211	58	1.089	1.631	2.634	0.353	0.250	156.855
			(0.037)	(0.074)	(5)	(0.197)	(0.330)	(0.328)	(0.035)	(0.051)	(8.216)
		SCR_3^*	0.899	0.281	62	1.165	1.791	2.860	0.350	0.270	151.538
			(0.043)	(0.073)	(5)	(0.154)	(0.282)	(0.315)	(0.037)	(0.046)	(6.583)
		SCR_4	0.913	0.595	112	1.241	2.154	3.273	0.258	0.357	166.819
			(0.039)	(0.043)	(12)	(0.147)	(0.331)	(0.270)	(0.034)	(0.058)	(8.527)
ER	I	MRCE	0.210	0.511	132	5.493	6.324	11.287	0.232	0.501	135.659
			(0.291)	(0.258)	(300)	(1.654)	(1.517)	(1.302)	(0.083)	(0.144)	(22.603)
		aMCR	0.592	0.679	115	-	-	-	0.885	0.674	82.120
			(0.099)	(0.041)	(20)	-	-	-	(0.025)	(0.026)	(11.411)
		SCR_3	0.401	0.379	39	2.543	4.157	7.124	0.585	0.121	85.348
			(0.113)	(0.095)	(6)	(0.485)	(0.942)	(0.805)	(0.054)	(0.023)	(10.917)
		SCR_3^*	0.324	0.438	35	2.550	4.179	7.208	0.529	0.113	90.999
			(0.094)	(0.097)	(6)	(0.364)	(0.780)	(0.675)	(0.042)	(0.023)	(12.374)
		SCR_4	0.477	0.604	75	2.608	4.707	7.347	0.531	0.129	90.105
			(0.100)	(0.063)	(10)	(0.640)	(1.186)	(0.773)	(0.048)	(0.023)	(12.557)
	II	MRCE	0.265	0.554	162	5.870	6.851	11.053	0.339	0.475	118.137
			(0.331)	(0.260)	(297)	(3.196)	(3.326)	(2.609)	(0.093)	(0.167)	(18.832)
		aMCR	0.592	0.673	114	-	-	-	0.626	0.655	83.176
			(0.093)	(0.042)	(18)	-	-	-	(0.029)	(0.030)	(10.538)
		SCR_3	0.356	0.431	39	2.455	3.977	6.912	0.337	0.155	88.195
			(0.096)	(0.082)	(7)	(0.247)	(0.619)	(0.630)	(0.021)	(0.028)	(11.476)
		SCR_3^*	0.336	0.431	36	2.542	4.041	7.064	0.327	0.142	88.978
			(0.090)	(0.080)	(6)	(0.250)	(0.607)	(0.615)	(0.020)	(0.030)	(11.709)
		SCR_4	0.449	0.560	63	2.464	4.064	6.943	0.332	0.152	88.629
			(0.099)	(0.063)	(9)	(0.256)	(0.662)	(0.616)	(0.021)	(0.029)	(11.785)

Continued on next page

Table 4.10: Average (SD) of $\hat{\Omega}$ related PDR, FDR, $|\hat{\Omega}|$ and matrix loss in three norms and $\hat{B}$ related PDR, FDR and PMSE when $p = 50$, $n = 100$ and $h = 0.6$

			$\hat{\Omega}$						$\hat{B}$		
Graphs	X	Mtd.	PDR	FDR	$\lvert\hat{\Omega}\rvert$	$\lVert\cdot\rVert_S$	$\lVert\cdot\rVert_{\ell_1}$	$\lVert\cdot\rVert_F$	PDR	FDR	PMSE
	III	MRCE	0.138	0.441	47	5.640	6.230	11.186	0.279	0.475	133.881
			(0.168)	(0.201)	(171)	(1.614)	(1.261)	(1.325)	(0.077)	(0.142)	(19.879)
		aMCR	0.609	0.670	113	-	-	-	0.774	0.667	85.494
			(0.092)	(0.042)	(20)	-	-	-	(0.024)	(0.026)	(10.747)
		SCR_3	0.394	0.362	37	2.299	3.680	6.534	0.489	0.131	88.297
			(0.097)	(0.091)	(7)	(0.366)	(0.767)	(0.678)	(0.036)	(0.026)	(10.045)
		SCR_3^*	0.341	0.395	34	2.430	3.801	6.837	0.451	0.127	91.814
			(0.086)	(0.092)	(6)	(0.256)	(0.601)	(0.621)	(0.030)	(0.025)	(10.852)
		SCR_4	0.500	0.540	66	2.305	3.873	6.610	0.468	0.137	89.847
			(0.086)	(0.059)	(9)	(0.321)	(0.688)	(0.655)	(0.035)	(0.026)	(10.293)
	IV	MRCE	0.067	0.226	6	3.716	4.993	8.139	0.616	0.575	72.675
			(0.056)	(0.212)	(5)	(1.776)	(1.398)	(1.345)	(0.035)	(0.060)	(10.482)
		aMCR	0.858	0.599	133	-	-	-	0.498	0.631	69.027
			(0.071)	(0.050)	(21)	-	-	-	(0.038)	(0.042)	(8.595)
		SCR_3	0.598	0.357	57	2.018	3.547	5.658	0.305	0.280	81.051
			(0.110)	(0.083)	(6)	(0.332)	(0.670)	(0.782)	(0.040)	(0.047)	(11.051)
		SCR_3^*	0.559	0.385	55	2.036	3.562	5.634	0.308	0.284	78.782
			(0.105)	(0.074)	(7)	(0.293)	(0.633)	(0.667)	(0.039)	(0.049)	(9.907)
		SCR_4	0.626	0.640	107	2.151	4.145	6.037	0.206	0.386	87.355
			(0.096)	(0.052)	(12)	(0.252)	(0.677)	(0.623)	(0.031)	(0.049)	(11.875)
Tridiag	I	MRCE	0.860	0.915	825	5.613	8.022	19.617	0.199	0.185	116.659
			(0.096)	(0.093)	(86)	(0.252)	(0.400)	(0.711)	(0.029)	(0.081)	(3.163)
		aMCR	0.629	0.647	131	-	-	-	0.888	0.653	69.421
			(0.059)	(0.036)	(14)	-	-	-	(0.026)	(0.029)	(2.439)
		SCR_3	0.581	0.158	50	4.795	6.136	11.720	0.840	0.108	63.603
			(0.060)	(0.069)	(3)	(1.690)	(1.840)	(1.647)	(0.056)	(0.021)	(3.608)
		SCR_3^*	0.492	0.263	49	4.621	5.926	12.897	0.667	0.117	73.779
			(0.054)	(0.070)	(4)	(0.462)	(0.570)	(1.160)	(0.052)	(0.022)	(3.080)
		SCR_4	0.607	0.512	91	4.957	7.275	12.560	0.749	0.134	69.608
			(0.061)	(0.044)	(7)	(1.302)	(2.037)	(1.236)	(0.059)	(0.021)	(3.754)
	II	MRCE	0.815	0.888	640	5.199	7.724	16.657	0.444	0.426	89.117
			(0.125)	(0.128)	(99)	(0.592)	(0.667)	(1.063)	(0.046)	(0.066)	(3.957)
		aMCR	0.633	0.656	136	-	-	-	0.635	0.629	71.326
			(0.049)	(0.035)	(13)	-	-	-	(0.027)	(0.033)	(2.649)
		SCR_3	0.523	0.268	52	4.499	5.702	12.947	0.414	0.155	73.934
			(0.049)	(0.062)	(4)	(0.456)	(0.578)	(0.983)	(0.029)	(0.028)	(2.846)
		SCR_3^*	0.504	0.266	50	4.629	5.751	14.081	0.383	0.147	75.535
			(0.049)	(0.064)	(3)	(0.406)	(0.539)	(0.963)	(0.024)	(0.029)	(2.866)
		SCR_4	0.560	0.480	79	4.569	5.946	13.536	0.404	0.161	74.880
			(0.049)	(0.051)	(7)	(0.406)	(0.614)	(0.854)	(0.025)	(0.027)	(2.633)
	III	MRCE	0.737	0.791	668	5.923	8.171	19.321	0.312	0.278	106.134
			(0.301)	(0.321)	(272)	(1.673)	(1.609)	(1.549)	(0.043)	(0.107)	(3.888)
		aMCR	0.634	0.652	133	-	-	-	0.778	0.642	69.821
			(0.055)	(0.036)	(15)	-	-	-	(0.025)	(0.032)	(2.728)
		SCR_3	0.558	0.185	50	4.325	5.529	10.873	0.671	0.127	67.338
			(0.065)	(0.063)	(3)	(1.205)	(1.326)	(1.333)	(0.046)	(0.022)	(3.233)
		SCR_3^*	0.510	0.238	48	4.521	5.686	12.867	0.560	0.124	72.886
			(0.062)	(0.064)	(4)	(0.468)	(0.625)	(0.981)	(0.038)	(0.020)	(2.859)
		SCR_4	0.593	0.460	80	4.608	6.456	11.939	0.634	0.136	69.411
			(0.056)	(0.047)	(6)	(1.127)	(1.900)	(1.049)	(0.042)	(0.022)	(2.917)
	IV	MRCE	0.064	0.044	9	5.324	6.251	18.292	0.633	0.563	59.423
			(0.180)	(0.142)	(28)	(0.602)	(0.645)	(2.142)	(0.035)	(0.070)	(2.195)
		aMCR	0.689	0.570	119	-	-	-	0.516	0.604	58.997
			(0.047)	(0.065)	(20)	-	-	-	(0.039)	(0.047)	(2.150)
		SCR_3	0.638	0.193	58	3.414	4.773	7.532	0.374	0.243	69.860
			(0.047)	(0.072)	(4)	(0.698)	(1.078)	(0.962)	(0.039)	(0.051)	(3.954)
		SCR_3^*	0.610	0.306	64	3.752	5.529	8.670	0.356	0.276	67.831
			(0.049)	(0.084)	(7)	(0.541)	(0.919)	(1.001)	(0.043)	(0.053)	(3.107)
		SCR_4	0.639	0.591	114	3.827	6.407	9.514	0.281	0.346	74.207
			(0.055)	(0.044)	(11)	(0.572)	(1.106)	(0.845)	(0.038)	(0.056)	(4.199)

Continued on next page

Table 4.10: Average (SD) of $\hat{\Omega}$ related PDR, FDR, $|\hat{\Omega}|$ and matrix loss in three norms and $\hat{B}$ related PDR, FDR and PMSE when $p = 50$, $n = 100$ and $h = 0.6$

			$\hat{\Omega}$						$\hat{B}$		
Graphs	X	Mtd.	PDR	FDR	$\lvert\hat{\Omega}\rvert$	$\lVert\cdot\rVert_S$	$\lVert\cdot\rVert_{\ell_1}$	$\lVert\cdot\rVert_F$	PDR	FDR	PMSE
Hub	I	MRCE	0.541	0.882	587	8.515	14.300	14.756	0.082	0.525	166.294
			(0.399)	(0.279)	(423)	(2.086)	(2.130)	(2.815)	(0.055)	(0.349)	(5.650)
		aMCR	0.537	0.739	96	-	-	-	0.884	0.687	102.628
			(0.082)	(0.054)	(19)	-	-	-	(0.023)	(0.022)	(4.281)
		SCR_3	0.343	0.454	28	6.600	12.112	11.285	0.505	0.125	112.542
			(0.084)	(0.093)	(5)	(0.808)	(1.342)	(1.450)	(0.044)	(0.025)	(5.117)
		SCR_3^*	0.310	0.461	26	6.790	12.402	12.154	0.487	0.116	113.595
			(0.076)	(0.087)	(5)	(0.644)	(1.106)	(1.306)	(0.039)	(0.023)	(4.840)
		SCR_4	0.490	0.631	60	6.699	12.223	11.571	0.467	0.134	116.228
			(0.093)	(0.060)	(8)	(1.797)	(2.661)	(1.973)	(0.045)	(0.020)	(5.668)
	II	MRCE	0.138	0.793	149	7.744	14.124	15.511	0.098	0.309	159.852
			(0.318)	(0.399)	(334)	(0.592)	(0.822)	(1.879)	(0.093)	(0.233)	(10.558)
		aMCR	0.544	0.748	100	-	-	-	0.624	0.666	101.627
			(0.090)	(0.048)	(19)	-	-	-	(0.025)	(0.026)	(3.425)
		SCR_3	0.359	0.435	29	6.387	11.701	11.868	0.315	0.153	107.052
			(0.074)	(0.098)	(5)	(0.620)	(1.129)	(1.185)	(0.022)	(0.035)	(4.002)
		SCR_3^*	0.337	0.437	27	6.510	11.897	12.408	0.312	0.141	107.049
			(0.069)	(0.098)	(5)	(0.562)	(1.050)	(1.096)	(0.021)	(0.032)	(4.065)
		SCR_4	0.490	0.553	50	6.417	11.704	11.834	0.311	0.153	107.335
			(0.078)	(0.069)	(6)	(0.600)	(1.071)	(1.139)	(0.021)	(0.027)	(4.190)
	III	MRCE	0.142	0.613	164	7.645	13.947	14.883	0.099	0.422	163.228
			(0.312)	(0.482)	(349)	(1.339)	(1.584)	(2.404)	(0.072)	(0.316)	(7.387)
		aMCR	0.538	0.756	101	-	-	-	0.773	0.679	102.344
			(0.086)	(0.049)	(18)	-	-	-	(0.027)	(0.025)	(4.363)
		SCR_3	0.362	0.437	29	6.243	11.524	10.976	0.438	0.133	108.212
			(0.078)	(0.104)	(5)	(0.730)	(1.277)	(1.191)	(0.032)	(0.027)	(4.410)
		SCR_3^*	0.331	0.443	27	6.505	11.895	11.969	0.423	0.126	108.879
			(0.065)	(0.097)	(5)	(0.659)	(1.162)	(1.015)	(0.032)	(0.027)	(4.485)
		SCR_4	0.494	0.572	52	6.260	11.480	10.960	0.428	0.137	109.006
			(0.081)	(0.063)	(7)	(0.934)	(1.633)	(1.300)	(0.030)	(0.025)	(4.668)
	IV	MRCE	0.000	0.180	1	6.543	12.610	12.422	0.592	0.545	86.941
			(0.000)	(0.386)	(5)	(0.554)	(0.810)	(1.249)	(0.063)	(0.096)	(10.746)
		aMCR	0.777	0.718	126	-	-	-	0.488	0.647	83.017
			(0.069)	(0.041)	(19)	-	-	-	(0.038)	(0.043)	(4.176)
		SCR_3	0.601	0.446	49	4.274	8.378	7.424	0.285	0.302	94.520
			(0.078)	(0.084)	(6)	(1.058)	(1.787)	(1.062)	(0.039)	(0.047)	(5.274)
		SCR_3^*	0.575	0.438	46	4.631	8.827	7.794	0.295	0.299	92.234
			(0.081)	(0.082)	(6)	(0.968)	(1.809)	(1.122)	(0.042)	(0.048)	(4.909)
		SCR_4	0.655	0.674	91	4.701	9.648	8.254	0.193	0.397	102.221
			(0.075)	(0.050)	(10)	(0.786)	(1.552)	(0.961)	(0.030)	(0.054)	(6.408)
BA	I	MRCE	0.892	0.951	902	17.350	23.609	22.992	0.302	0.779	418.579
			(0.062)	(0.003)	(50)	(15.919)	(16.677)	(15.017)	(0.046)	(0.032)	(36.060)
		aMCR	0.702	0.822	194	-	-	-	0.898	0.673	254.757
			(0.074)	(0.023)	(16)	-	-	-	(0.026)	(0.036)	(23.016)
		SCR_3	0.736	0.310	52	4.973	11.020	7.611	0.911	0.099	217.633
			(0.076)	(0.085)	(3)	(2.780)	(6.344)	(2.120)	(0.044)	(0.020)	(18.663)
		SCR_3^*	0.557	0.513	56	8.701	16.491	11.921	0.728	0.121	260.018
			(0.076)	(0.082)	(5)	(3.520)	(7.267)	(2.630)	(0.054)	(0.021)	(23.023)
		SCR_4	0.764	0.641	105	6.403	13.952	9.248	0.819	0.129	241.014
			(0.076)	(0.058)	(10)	(3.777)	(6.902)	(3.112)	(0.049)	(0.019)	(19.522)
	II	MRCE	0.912	0.946	826	11.965	17.899	17.255	0.452	0.742	355.540
			(0.048)	(0.004)	(50)	(13.332)	(13.883)	(12.154)	(0.045)	(0.033)	(29.582)
		aMCR	0.700	0.816	188	-	-	-	0.643	0.638	268.695
			(0.077)	(0.026)	(17)	-	-	-	(0.037)	(0.044)	(19.174)
		SCR_3	0.623	0.448	56	7.606	14.338	11.038	0.495	0.153	267.664
			(0.088)	(0.091)	(3)	(2.856)	(5.992)	(2.156)	(0.047)	(0.024)	(19.325)
		SCR_3^*	0.537	0.534	57	9.182	17.257	13.391	0.408	0.150	281.936
			(0.072)	(0.071)	(4)	(2.943)	(6.167)	(1.757)	(0.027)	(0.027)	(20.255)
		SCR_4	0.689	0.655	99	8.208	15.830	11.972	0.459	0.158	274.466
			(0.076)	(0.054)	(9)	(3.120)	(6.395)	(2.134)	(0.037)	(0.021)	(18.966)

Continued on next page

Table 4.10: Average (SD) of $\hat{\Omega}$ related PDR, FDR, $|\hat{\Omega}|$ and matrix loss in three norms and $\hat{B}$ related PDR, FDR and PMSE when $p = 50$, $n = 100$ and $h = 0.6$

			$\hat{\Omega}$						$\hat{B}$		
Graphs	X	Mtd.	PDR	FDR	$\lvert\hat{\Omega}\rvert$	$\lVert\cdot\rVert_S$	$\lVert\cdot\rVert_{\ell_1}$	$\lVert\cdot\rVert_F$	PDR	FDR	PMSE
	III	MRCE	0.895	0.948	854	15.253	21.310	21.185	0.386	0.769	386.047
			(0.060)	(0.004)	(53)	(15.205)	(15.936)	(13.781)	(0.053)	(0.027)	(28.735)
		aMCR	0.687	0.824	192	-	-	-	0.787	0.658	256.345
			(0.078)	(0.026)	(18)	-	-	-	(0.030)	(0.041)	(16.419)
		SCR_3	0.714	0.337	53	6.000	12.224	8.260	0.789	0.116	227.924
			(0.076)	(0.085)	(3)	(2.501)	(5.410)	(2.049)	(0.039)	(0.022)	(13.624)
		SCR_3^*	0.541	0.521	56	9.422	17.707	12.882	0.605	0.124	260.072
			(0.071)	(0.077)	(4)	(3.276)	(6.524)	(2.390)	(0.046)	(0.023)	(16.234)
		SCR_4	0.791	0.580	93	6.160	12.868	8.853	0.738	0.126	237.520
			(0.063)	(0.056)	(8)	(2.933)	(5.413)	(2.447)	(0.043)	(0.021)	(14.602)
	IV	MRCE	0.723	0.873	327	9.690	18.345	14.057	0.379	0.565	328.311
			(0.170)	(0.039)	(177)	(3.975)	(7.716)	(3.601)	(0.100)	(0.145)	(60.396)
		aMCR	0.809	0.743	158	-	-	-	0.537	0.619	219.475
			(0.069)	(0.047)	(23)	-	-	-	(0.043)	(0.054)	(16.473)
		SCR_3	0.726	0.382	58	5.757	11.572	7.735	0.422	0.245	254.866
			(0.075)	(0.089)	(5)	(2.598)	(5.719)	(2.164)	(0.058)	(0.058)	(21.140)
		SCR_3^*	0.652	0.517	67	7.155	13.870	9.824	0.380	0.281	248.932
			(0.071)	(0.070)	(7)	(2.873)	(6.120)	(2.178)	(0.049)	(0.054)	(18.870)
		SCR_4	0.717	0.730	131	6.423	13.667	9.083	0.290	0.353	275.807
			(0.071)	(0.033)	(10)	(3.003)	(6.234)	(2.310)	(0.044)	(0.061)	(22.198)

Table 4.11: Average (SD) of $\hat{\Omega}$ related PDR, FDR, $|\hat{\Omega}|$ and matrix loss in three norms and $\hat{B}$ related PDR, FDR and PMSE for block precision matrix design when $p = 200$, $n = 100$ and $h = 0.8$

			$\hat{\Omega}$						$\hat{B}$		
Graphs	X	Mtd.	PDR	FDR	$\lvert\hat{\Omega}\rvert$	$\lVert\cdot\rVert_S$	$\lVert\cdot\rVert_{\ell_1}$	$\lVert\cdot\rVert_F$	PDR	FDR	PMSE
AR(1)	I	MRCE	0.002	0.976	16	4.390	4.790	15.775	0.817	0.507	1232.233
			(0.004)	(0.042)	(6)	(2.102)	(2.121)	(1.294)	(0.029)	(0.051)	(66.793)
		aMCR	0.837	0.849	1090	-	-	-	0.998	0.682	634.426
			(0.028)	(0.011)	(85)	-	-	-	(0.001)	(0.013)	(12.870)
		SCR_3	0.915	0.044	188	1.845	2.313	7.096	0.982	0.081	527.082
			(0.027)	(0.016)	(6)	(0.373)	(0.483)	(0.446)	(0.009)	(0.008)	(12.961)
		SCR_3^*	0.901	0.046	185	1.577	2.041	6.468	0.979	0.080	533.867
			(0.028)	(0.017)	(6)	(0.275)	(0.331)	(0.412)	(0.010)	(0.007)	(14.411)
		SCR_4	0.945	0.474	353	2.824	4.550	9.117	0.947	0.097	578.348
			(0.022)	(0.034)	(19)	(0.786)	(1.465)	(0.723)	(0.020)	(0.008)	(29.264)
	II	MRCE	0.005	0.974	66	2.470	2.617	15.286	0.231	0.181	1812.271
			(0.005)	(0.099)	(18)	(1.649)	(1.770)	(0.542)	(0.081)	(0.109)	(135.878)
		aMCR	0.763	0.870	1156	-	-	-	0.882	0.684	692.143
			(0.035)	(0.011)	(109)	-	-	-	(0.014)	(0.014)	(14.996)
		SCR_3	0.376	0.375	118	1.648	2.108	11.030	0.566	0.127	829.382
			(0.056)	(0.044)	(17)	(0.219)	(0.275)	(0.335)	(0.025)	(0.011)	(23.192)
		SCR_3^*	0.317	0.380	101	1.625	1.979	11.464	0.547	0.116	842.243
			(0.054)	(0.053)	(18)	(0.028)	(0.096)	(0.309)	(0.022)	(0.011)	(20.813)
		SCR_4	0.588	0.689	371	1.698	2.635	10.775	0.509	0.144	896.721
			(0.037)	(0.017)	(16)	(0.326)	(0.816)	(0.237)	(0.020)	(0.012)	(23.927)
	III	MRCE	0.003	0.380	7	3.737	4.073	14.886	0.711	0.533	1158.763
			(0.006)	(0.451)	(13)	(2.034)	(2.113)	(1.040)	(0.298)	(0.241)	(457.948)
		aMCR	0.801	0.857	1107	-	-	-	0.961	0.686	655.552
			(0.031)	(0.012)	(89)	-	-	-	(0.006)	(0.012)	(12.717)
		SCR_3	0.768	0.092	166	1.839	2.347	7.961	0.907	0.094	580.320
			(0.047)	(0.028)	(9)	(0.402)	(0.509)	(0.443)	(0.019)	(0.009)	(18.830)
		SCR_3^*	0.695	0.119	155	1.533	1.970	7.880	0.873	0.093	612.762
			(0.049)	(0.030)	(10)	(0.220)	(0.322)	(0.441)	(0.019)	(0.008)	(18.959)
		SCR_4	0.871	0.511	350	2.514	4.166	8.838	0.888	0.105	600.471
			(0.027)	(0.029)	(18)	(0.788)	(1.527)	(0.572)	(0.018)	(0.009)	(18.321)
	IV	MRCE	0.169	0.983	1925	1.785	2.016	14.609	0.031	0.332	2280.273
			(0.035)	(0.003)	(154)	(0.011)	(0.024)	(0.125)	(0.024)	(0.170)	(148.065)
		aMCR	0.980	0.862	1418	-	-	-	0.629	0.636	568.345
			(0.010)	(0.020)	(196)	-	-	-	(0.021)	(0.023)	(23.119)
		SCR_3	0.848	0.097	184	1.498	2.007	6.215	0.571	0.264	623.223
			(0.046)	(0.037)	(9)	(0.319)	(0.343)	(0.433)	(0.033)	(0.028)	(45.138)
		SCR_3^*	0.794	0.132	179	1.334	1.943	6.337	0.559	0.263	613.254
			(0.057)	(0.048)	(12)	(0.132)	(0.194)	(0.545)	(0.029)	(0.026)	(32.427)
		SCR_4	0.770	0.700	504	1.661	3.508	8.369	0.280	0.417	960.764
			(0.037)	(0.020)	(27)	(0.318)	(0.817)	(0.429)	(0.037)	(0.040)	(66.423)
ER	I	MRCE	0.005	0.802	7	4.885	5.684	21.802	0.617	0.601	843.912
			(0.008)	(0.241)	(7)	(1.591)	(1.271)	(1.801)	(0.157)	(0.152)	(165.635)
		aMCR	0.520	0.841	804	-	-	-	0.998	0.672	323.042
			(0.074)	(0.016)	(87)	-	-	-	(0.001)	(0.015)	(37.333)
		SCR_3	0.530	0.148	152	3.182	4.729	13.108	0.973	0.078	274.619
			(0.101)	(0.044)	(19)	(0.692)	(1.024)	(1.441)	(0.009)	(0.008)	(30.738)
		SCR_3^*	0.510	0.148	146	2.690	4.222	12.287	0.971	0.075	277.059
			(0.096)	(0.040)	(18)	(0.485)	(0.767)	(1.324)	(0.010)	(0.008)	(31.903)
		SCR_4	0.645	0.540	344	4.859	8.662	15.851	0.926	0.090	306.452
			(0.092)	(0.040)	(29)	(1.829)	(3.780)	(1.819)	(0.022)	(0.009)	(36.272)
	II	MRCE	0.007	0.822	12	4.088	5.237	21.764	0.252	0.255	991.202
			(0.008)	(0.220)	(10)	(1.478)	(1.222)	(1.387)	(0.060)	(0.097)	(174.344)
		aMCR	0.467	0.865	840	-	-	-	0.880	0.671	357.354
			(0.068)	(0.015)	(101)	-	-	-	(0.011)	(0.012)	(43.573)
		SCR_3	0.191	0.576	107	2.850	4.687	15.720	0.549	0.121	428.638
			(0.049)	(0.065)	(13)	(0.267)	(0.573)	(1.151)	(0.020)	(0.013)	(52.909)
		SCR_3^*	0.161	0.590	94	2.877	4.701	16.011	0.539	0.112	435.270
			(0.042)	(0.063)	(13)	(0.221)	(0.582)	(1.078)	(0.018)	(0.012)	(53.418)
		SCR_4	0.286	0.782	315	2.886	5.310	16.072	0.485	0.135	470.017
			(0.060)	(0.031)	(27)	(0.293)	(0.998)	(1.106)	(0.017)	(0.014)	(60.244)

Continued on next page

Table 4.11: Average (SD) of $\hat{\Omega}$ related PDR, FDR, $|\hat{\Omega}|$ and matrix loss in three norms and $\hat{B}$ related PDR, FDR and PMSE for block precision matrix design when $p = 200$, $n = 100$ and $h = 0.8$

			$\hat{\Omega}$						$\hat{B}$		
Graphs	X	Mtd.	PDR	FDR	$\lvert\hat{\Omega}\rvert$	$\lVert\cdot\rVert_S$	$\lVert\cdot\rVert_{\ell_1}$	$\lVert\cdot\rVert_F$	PDR	FDR	PMSE
	III	MRCE	0.005	0.665	5	4.312	5.353	22.897	0.221	0.253	1014.938
			(0.006)	(0.347)	(4)	(1.398)	(1.051)	(1.964)	(0.178)	(0.228)	(145.867)
		aMCR	0.484	0.849	796	-	-	-	0.959	0.675	334.296
			(0.075)	(0.016)	(106)	-	-	-	(0.006)	(0.015)	(47.009)
		SCR_3	0.381	0.266	126	2.916	4.586	13.725	0.852	0.092	310.761
			(0.093)	(0.070)	(16)	(0.626)	(0.961)	(1.619)	(0.019)	(0.008)	(38.950)
		SCR_3^*	0.338	0.290	116	2.635	4.431	13.458	0.836	0.087	320.785
			(0.083)	(0.072)	(16)	(0.408)	(0.802)	(1.438)	(0.018)	(0.009)	(41.950)
		SCR_4	0.511	0.600	313	4.027	7.664	15.155	0.825	0.098	323.972
			(0.096)	(0.042)	(28)	(1.579)	(3.209)	(1.968)	(0.021)	(0.009)	(39.933)
	IV	MRCE	0.269	0.966	1956	3.201	5.469	19.859	0.006	0.041	1173.614
			(0.065)	(0.008)	(346)	(0.515)	(0.641)	(1.486)	(0.036)	(0.117)	(161.634)
		aMCR	0.775	0.829	1138	-	-	-	0.609	0.637	293.599
			(0.061)	(0.025)	(146)	-	-	-	(0.021)	(0.020)	(33.634)
		SCR_3	0.448	0.252	147	2.573	4.198	12.025	0.534	0.267	320.155
			(0.079)	(0.072)	(16)	(0.571)	(0.740)	(1.124)	(0.034)	(0.026)	(38.383)
		SCR_3^*	0.411	0.268	138	2.403	4.074	11.845	0.531	0.262	314.227
			(0.074)	(0.071)	(16)	(0.319)	(0.628)	(0.981)	(0.029)	(0.025)	(36.286)
		SCR_4	0.424	0.762	440	2.692	6.173	13.886	0.243	0.427	503.839
			(0.061)	(0.029)	(33)	(0.312)	(1.247)	(0.974)	(0.030)	(0.035)	(66.881)
Tridiag	I	MRCE	0.000	0.005	0	6.229	7.434	43.279	0.795	0.562	535.729
			(0.000)	(0.050)	(0)	(0.388)	(0.721)	(1.321)	(0.026)	(0.058)	(23.367)
		aMCR	0.599	0.811	928	-	-	-	0.998	0.639	269.958
			(0.043)	(0.014)	(71)	-	-	-	(0.001)	(0.018)	(6.536)
		SCR_3	0.638	0.039	193	5.353	6.736	19.188	0.989	0.084	230.333
			(0.045)	(0.015)	(4)	(1.068)	(1.424)	(1.370)	(0.006)	(0.007)	(5.252)
		SCR_3^*	0.629	0.047	192	4.671	6.134	17.254	0.985	0.087	234.255
			(0.046)	(0.018)	(5)	(0.641)	(0.973)	(1.254)	(0.007)	(0.007)	(5.539)
		SCR_4	0.642	0.478	360	7.963	12.391	25.071	0.951	0.098	254.626
			(0.043)	(0.037)	(22)	(2.298)	(3.994)	(2.226)	(0.018)	(0.008)	(11.861)
	II	MRCE	0.000	0.108	0	6.242	7.256	45.017	0.299	0.138	747.138
			(0.000)	(0.309)	(1)	(0.281)	(0.378)	(1.860)	(0.171)	(0.224)	(120.766)
		aMCR	0.558	0.848	1082	-	-	-	0.884	0.658	299.988
			(0.038)	(0.011)	(87)	-	-	-	(0.013)	(0.014)	(6.587)
		SCR_3	0.374	0.364	172	5.571	7.205	30.869	0.631	0.129	348.104
			(0.045)	(0.040)	(12)	(0.722)	(0.794)	(1.367)	(0.032)	(0.012)	(13.665)
		SCR_3^*	0.324	0.393	156	5.582	6.995	33.287	0.582	0.119	364.330
			(0.036)	(0.039)	(14)	(0.301)	(0.393)	(1.260)	(0.022)	(0.012)	(10.135)
		SCR_4	0.447	0.694	428	5.545	8.235	32.118	0.545	0.147	388.364
			(0.036)	(0.019)	(22)	(0.646)	(1.849)	(1.142)	(0.021)	(0.012)	(11.785)
	III	MRCE	0.001	0.212	1	6.223	7.670	41.811	0.814	0.656	446.587
			(0.002)	(0.376)	(3)	(0.461)	(0.939)	(1.433)	(0.018)	(0.047)	(17.189)
		aMCR	0.577	0.825	963	-	-	-	0.961	0.656	282.527
			(0.043)	(0.015)	(78)	-	-	-	(0.006)	(0.014)	(6.555)
		SCR_3	0.594	0.074	186	5.380	6.830	20.901	0.952	0.090	242.068
			(0.048)	(0.024)	(6)	(1.013)	(1.127)	(1.245)	(0.012)	(0.008)	(6.792)
		SCR_3^*	0.546	0.123	181	4.745	6.252	20.660	0.910	0.097	260.994
			(0.046)	(0.029)	(7)	(0.533)	(0.642)	(1.186)	(0.014)	(0.009)	(7.194)
		SCR_4	0.621	0.498	360	7.048	11.135	24.254	0.928	0.106	252.869
			(0.042)	(0.030)	(18)	(2.012)	(3.245)	(1.555)	(0.015)	(0.009)	(7.514)
	IV	MRCE	0.135	0.629	1162	5.920	7.171	41.545	0.293	0.565	737.285
			(0.113)	(0.464)	(928)	(0.355)	(0.504)	(3.269)	(0.306)	(0.137)	(369.746)
		aMCR	0.675	0.827	1155	-	-	-	0.627	0.626	248.413
			(0.043)	(0.022)	(152)	-	-	-	(0.020)	(0.024)	(6.451)
		SCR_3	0.611	0.088	195	4.839	6.362	16.905	0.611	0.262	269.183
			(0.046)	(0.035)	(6)	(1.215)	(1.322)	(1.358)	(0.034)	(0.027)	(18.658)
		SCR_3^*	0.582	0.138	197	4.288	6.090	17.176	0.586	0.267	267.354
			(0.045)	(0.054)	(11)	(0.624)	(0.801)	(1.700)	(0.032)	(0.025)	(13.755)
		SCR_4	0.539	0.703	530	5.061	9.536	24.354	0.310	0.406	420.449
			(0.046)	(0.023)	(34)	(0.479)	(1.633)	(1.179)	(0.041)	(0.035)	(30.144)

Continued on next page

Table 4.11: Average (SD) of $\hat{\Omega}$ related PDR, FDR, $|\hat{\Omega}|$ and matrix loss in three norms and $\hat{B}$ related PDR, FDR and PMSE for block precision matrix design when $p = 200$, $n = 100$ and $h = 0.8$

			$\hat{\Omega}$						$\hat{B}$		
Graphs	X	Mtd.	PDR	FDR	$\lvert\hat{\Omega}\rvert$	$\lVert\cdot\rVert_S$	$\lVert\cdot\rVert_{\ell_1}$	$\lVert\cdot\rVert_F$	PDR	FDR	PMSE
Hub	I	MRCE	0.000	1.000	31	8.716	15.156	35.046	0.071	0.478	1330.418
			(0.000)	(0.000)	(43)	(0.278)	(0.496)	(1.631)	(0.014)	(0.252)	(40.136)
		aMCR	0.424	0.894	728	-	-	-	0.998	0.690	406.984
			(0.046)	(0.015)	(86)	-	-	-	(0.001)	(0.011)	(13.511)
		SCR_3	0.527	0.193	118	5.701	10.955	15.397	0.969	0.075	335.630
			(0.044)	(0.041)	(10)	(1.468)	(2.162)	(1.384)	(0.012)	(0.006)	(12.736)
		SCR_3^*	0.511	0.190	114	5.384	10.517	14.871	0.968	0.072	336.128
			(0.043)	(0.041)	(9)	(1.589)	(2.282)	(1.341)	(0.012)	(0.006)	(13.334)
		SCR_4	0.723	0.565	300	10.751	19.393	22.187	0.920	0.087	376.488
			(0.053)	(0.043)	(18)	(7.503)	(12.288)	(6.309)	(0.026)	(0.008)	(22.803)
	II	MRCE	0.000	1.000	15	8.543	14.931	32.625	0.122	0.588	1268.409
			(0.000)	(0.000)	(6)	(0.374)	(0.631)	(1.223)	(0.014)	(0.050)	(52.890)
		aMCR	0.358	0.912	740	-	-	-	0.878	0.687	434.989
			(0.043)	(0.013)	(91)	-	-	-	(0.012)	(0.011)	(15.648)
		SCR_3	0.138	0.651	71	7.920	14.390	28.275	0.526	0.117	520.578
			(0.031)	(0.057)	(12)	(0.421)	(0.673)	(1.365)	(0.018)	(0.011)	(20.173)
		SCR_3^*	0.126	0.655	66	7.982	14.428	29.028	0.524	0.110	518.882
			(0.027)	(0.059)	(11)	(0.427)	(0.679)	(1.252)	(0.017)	(0.011)	(19.460)
		SCR_4	0.259	0.817	254	7.870	14.691	27.962	0.469	0.132	561.591
			(0.038)	(0.023)	(17)	(0.480)	(1.104)	(1.183)	(0.018)	(0.013)	(22.037)
	III	MRCE	0.000	0.770	2	8.598	15.034	33.257	0.095	0.625	1302.651
			(0.000)	(0.423)	(2)	(0.320)	(0.554)	(1.429)	(0.008)	(0.078)	(42.925)
		aMCR	0.393	0.902	733	-	-	-	0.959	0.687	411.872
			(0.048)	(0.015)	(92)	-	-	-	(0.006)	(0.012)	(14.271)
		SCR_3	0.334	0.325	89	7.126	13.119	19.982	0.825	0.090	395.238
			(0.042)	(0.057)	(10)	(0.839)	(1.241)	(1.616)	(0.017)	(0.008)	(16.540)
		SCR_3^*	0.308	0.330	83	7.223	13.226	21.015	0.819	0.085	396.690
			(0.041)	(0.055)	(10)	(0.763)	(1.157)	(1.579)	(0.017)	(0.008)	(16.633)
		SCR_4	0.532	0.640	267	8.172	15.205	21.759	0.801	0.098	409.218
			(0.052)	(0.035)	(19)	(3.057)	(4.943)	(2.616)	(0.018)	(0.008)	(16.824)
	IV	MRCE	0.007	0.999	1716	7.749	14.039	29.032	0.091	0.633	1344.845
			(0.015)	(0.001)	(244)	(0.649)	(0.876)	(2.478)	(0.038)	(0.111)	(103.589)
		aMCR	0.690	0.888	1127	-	-	-	0.605	0.643	353.961
			(0.041)	(0.014)	(126)	-	-	-	(0.020)	(0.022)	(14.687)
		SCR_3	0.438	0.306	114	5.595	10.857	16.191	0.520	0.274	375.037
			(0.043)	(0.062)	(11)	(0.948)	(1.610)	(1.310)	(0.029)	(0.023)	(19.911)
		SCR_3^*	0.402	0.311	106	6.030	11.604	17.416	0.518	0.268	371.718
			(0.045)	(0.068)	(12)	(0.895)	(1.517)	(1.514)	(0.026)	(0.021)	(17.997)
		SCR_4	0.442	0.796	391	6.701	13.745	21.539	0.230	0.443	598.584
			(0.045)	(0.023)	(26)	(0.569)	(1.600)	(1.273)	(0.033)	(0.039)	(36.984)
BA	I	MRCE	0.094	0.966	539	11.271	21.040	33.155	0.182	0.500	3454.254
			(0.030)	(0.007)	(135)	(2.882)	(6.090)	(2.717)	(0.047)	(0.181)	(260.490)
		aMCR	0.668	0.904	1376	-	-	-	0.999	0.665	981.557
			(0.052)	(0.011)	(96)	-	-	-	(0.001)	(0.020)	(58.986)
		SCR_3	0.678	0.238	175	6.451	13.753	15.769	0.990	0.091	834.578
			(0.055)	(0.066)	(8)	(2.501)	(5.407)	(3.229)	(0.006)	(0.010)	(51.068)
		SCR_3^*	0.659	0.251	173	7.344	14.589	16.497	0.984	0.097	856.464
			(0.058)	(0.067)	(8)	(2.738)	(5.673)	(3.655)	(0.008)	(0.010)	(54.269)
		SCR_4	0.758	0.622	396	8.618	17.611	18.691	0.950	0.106	928.247
			(0.051)	(0.042)	(24)	(3.201)	(6.823)	(3.253)	(0.019)	(0.011)	(71.045)
	II	MRCE	0.073	0.958	344	10.973	20.360	32.009	0.368	0.569	2797.879
			(0.030)	(0.011)	(115)	(2.793)	(5.795)	(2.801)	(0.058)	(0.056)	(310.072)
		aMCR	0.639	0.919	1553	-	-	-	0.889	0.676	1099.499
			(0.044)	(0.008)	(108)	-	-	-	(0.013)	(0.018)	(62.743)
		SCR_3	0.493	0.429	169	10.023	19.067	23.437	0.711	0.129	1190.068
			(0.062)	(0.058)	(13)	(2.880)	(5.870)	(3.364)	(0.033)	(0.011)	(64.488)
		SCR_3^*	0.329	0.536	139	10.763	20.064	29.214	0.594	0.125	1341.996
			(0.054)	(0.053)	(21)	(2.808)	(5.781)	(3.189)	(0.027)	(0.011)	(82.683)
		SCR_4	0.545	0.804	545	10.321	20.061	27.017	0.581	0.152	1394.789
			(0.049)	(0.019)	(24)	(2.788)	(5.718)	(3.400)	(0.025)	(0.013)	(74.802)

Continued on next page

Table 4.11: Average (SD) of $\hat{\Omega}$ related PDR, FDR, $|\hat{\Omega}|$ and matrix loss in three norms and $\hat{B}$ related PDR, FDR and PMSE for block precision matrix design when $p = 200$, $n = 100$ and $h = 0.8$

			$\hat{\Omega}$						$\hat{B}$		
Graphs	X	Mtd.	PDR	FDR	$\lvert\hat{\Omega}\rvert$	$\lVert\cdot\rVert_S$	$\lVert\cdot\rVert_{\ell_1}$	$\lVert\cdot\rVert_F$	PDR	FDR	PMSE
	III	MRCE	0.085	0.964	466	10.835	20.193	32.168	0.337	0.549	3092.597
			(0.035)	(0.009)	(147)	(2.922)	(6.130)	(2.834)	(0.061)	(0.060)	(291.341)
		aMCR	0.655	0.908	1395	-	-	-	0.963	0.676	1033.160
			(0.053)	(0.010)	(96)	-	-	-	(0.006)	(0.015)	(51.867)
		SCR_3	0.678	0.235	174	6.583	13.551	15.815	0.969	0.093	867.681
			(0.058)	(0.060)	(7)	(2.509)	(5.502)	(2.941)	(0.012)	(0.009)	(43.730)
		SCR_3^*	0.587	0.309	167	8.905	16.923	20.215	0.919	0.104	954.123
			(0.054)	(0.058)	(10)	(2.889)	(5.876)	(3.428)	(0.015)	(0.009)	(55.987)
		SCR_4	0.756	0.622	393	8.210	16.347	18.315	0.941	0.108	909.956
			(0.053)	(0.038)	(20)	(2.955)	(6.113)	(2.830)	(0.014)	(0.009)	(42.446)
	IV	MRCE	0.209	0.978	1825	11.458	21.108	34.826	0.009	0.159	3782.157
			(0.034)	(0.003)	(59)	(2.777)	(5.724)	(3.240)	(0.004)	(0.168)	(316.921)
		aMCR	0.772	0.892	1421	-	-	-	0.639	0.639	923.662
			(0.049)	(0.015)	(153)	-	-	-	(0.022)	(0.026)	(60.607)
		SCR_3	0.650	0.264	173	7.720	15.151	17.169	0.638	0.275	989.863
			(0.058)	(0.064)	(9)	(2.385)	(4.990)	(3.132)	(0.039)	(0.029)	(85.741)
		SCR_3^*	0.615	0.323	179	8.888	16.838	19.850	0.606	0.278	980.495
			(0.055)	(0.068)	(14)	(2.552)	(5.194)	(3.339)	(0.031)	(0.026)	(71.845)
		SCR_4	0.587	0.788	544	9.584	18.764	23.302	0.329	0.401	1582.389
			(0.046)	(0.020)	(31)	(2.564)	(5.195)	(3.127)	(0.041)	(0.035)	(142.161)

Table 4.12: Average (SD) of $\hat{\Omega}$ related PDR, FDR, $|\hat{\Omega}|$ and matrix loss in three norms and $\hat{B}$ related PDR, FDR and PMSE for noise precision matrix design when $p = 200$, $n = 100$ and $h = 0.8$

			$\hat{\Omega}$						$\hat{B}$		
Graphs	X	Mtd.	PDR	FDR	$\lvert\hat{\Omega}\rvert$	$\lVert\cdot\rVert_S$	$\lVert\cdot\rVert_{\ell_1}$	$\lVert\cdot\rVert_F$	PDR	FDR	PMSE
AR(1)	I	MRCE	0.008	0.938	7	5.382	5.382	13.512	0.600	0.585	961.219
			(0.011)	(0.109)	(3)	(1.742)	(1.741)	(2.046)	(0.063)	(0.087)	(43.056)
		aMCR	0.884	0.919	548	-	-	-	0.998	0.687	386.622
			(0.054)	(0.012)	(82)	-	-	-	(0.001)	(0.012)	(6.780)
		SCR_3	0.927	0.294	65	2.002	2.471	6.629	0.969	0.068	322.759
			(0.043)	(0.055)	(5)	(0.461)	(0.634)	(0.480)	(0.010)	(0.006)	(7.435)
		SCR_3^*	0.915	0.295	64	1.808	2.252	6.276	0.968	0.068	324.130
			(0.045)	(0.055)	(5)	(0.353)	(0.457)	(0.405)	(0.010)	(0.006)	(7.796)
		SCR_4	0.941	0.750	186	3.273	5.269	9.017	0.921	0.080	360.233
			(0.037)	(0.028)	(18)	(1.145)	(2.218)	(0.977)	(0.024)	(0.009)	(18.444)
	II	MRCE	0.012	0.957	13	2.343	2.383	11.800	0.181	0.089	1146.964
			(0.016)	(0.065)	(4)	(1.627)	(1.613)	(0.633)	(0.017)	(0.054)	(21.653)
		aMCR	0.830	0.932	604	-	-	-	0.878	0.683	414.470
			(0.053)	(0.009)	(77)	-	-	-	(0.011)	(0.011)	(7.469)
		SCR_3	0.421	0.681	65	1.604	2.098	7.504	0.529	0.115	499.556
			(0.069)	(0.052)	(10)	(0.171)	(0.298)	(0.225)	(0.017)	(0.011)	(12.368)
		SCR_3^*	0.384	0.681	60	1.602	2.027	7.501	0.527	0.110	500.483
			(0.066)	(0.056)	(10)	(0.172)	(0.252)	(0.211)	(0.016)	(0.011)	(11.912)
		SCR_4	0.564	0.876	223	1.741	2.797	8.093	0.474	0.129	539.094
			(0.067)	(0.015)	(15)	(0.503)	(1.002)	(0.282)	(0.016)	(0.011)	(13.398)
	III	MRCE	0.010	0.853	5	2.272	2.298	12.443	0.106	0.089	1236.430
			(0.027)	(0.277)	(10)	(1.355)	(1.346)	(0.643)	(0.170)	(0.200)	(124.379)
		aMCR	0.847	0.925	561	-	-	-	0.959	0.686	396.394
			(0.055)	(0.010)	(65)	-	-	-	(0.006)	(0.011)	(6.882)
		SCR_3	0.787	0.387	63	1.808	2.318	6.385	0.831	0.084	372.341
			(0.077)	(0.062)	(6)	(0.451)	(0.568)	(0.466)	(0.017)	(0.009)	(10.399)
		SCR_3^*	0.738	0.395	60	1.596	2.023	6.107	0.823	0.083	379.314
			(0.079)	(0.062)	(6)	(0.354)	(0.411)	(0.397)	(0.017)	(0.008)	(10.368)
		SCR_4	0.869	0.773	189	2.544	4.166	8.085	0.802	0.094	389.788
			(0.060)	(0.024)	(16)	(0.723)	(1.359)	(0.566)	(0.017)	(0.009)	(10.865)
	IV	MRCE	0.558	0.987	2157	1.643	2.452	9.309	0.003	0.162	1338.924
			(0.106)	(0.002)	(252)	(0.033)	(0.084)	(0.546)	(0.003)	(0.289)	(87.292)
		aMCR	0.983	0.929	706	-	-	-	0.601	0.645	341.323
			(0.020)	(0.014)	(128)	-	-	-	(0.020)	(0.021)	(10.673)
		SCR_3	0.877	0.423	76	1.655	2.188	5.606	0.521	0.265	362.142
			(0.059)	(0.077)	(10)	(0.377)	(0.434)	(0.429)	(0.031)	(0.023)	(21.579)
		SCR_3^*	0.842	0.429	73	1.467	2.020	5.328	0.519	0.262	359.220
			(0.064)	(0.081)	(10)	(0.352)	(0.395)	(0.353)	(0.029)	(0.023)	(17.514)
		SCR_4	0.774	0.893	356	1.699	4.219	7.290	0.234	0.430	575.807
			(0.060)	(0.012)	(25)	(0.320)	(1.234)	(0.337)	(0.031)	(0.036)	(40.985)
ER	I	MRCE	0.006	0.676	2	4.285	5.291	14.970	0.249	0.271	922.902
			(0.012)	(0.420)	(3)	(1.460)	(1.070)	(1.023)	(0.329)	(0.266)	(164.903)
		aMCR	0.552	0.936	529	-	-	-	0.998	0.685	312.121
			(0.105)	(0.013)	(82)	-	-	-	(0.001)	(0.012)	(11.192)
		SCR_3	0.549	0.426	57	2.548	3.738	8.700	0.967	0.069	262.224
			(0.121)	(0.074)	(8)	(0.519)	(0.736)	(0.793)	(0.010)	(0.007)	(10.244)
		SCR_3^*	0.534	0.423	56	2.307	3.498	8.244	0.967	0.067	262.080
			(0.116)	(0.074)	(8)	(0.442)	(0.673)	(0.713)	(0.010)	(0.007)	(10.290)
		SCR_4	0.661	0.781	182	3.881	6.624	11.329	0.917	0.080	294.623
			(0.105)	(0.031)	(20)	(1.392)	(2.686)	(1.261)	(0.025)	(0.008)	(16.330)
	II	MRCE	0.007	0.579	2	5.426	6.171	14.828	0.242	0.199	833.759
			(0.012)	(0.455)	(2)	(1.813)	(1.404)	(0.943)	(0.033)	(0.076)	(49.310)
		aMCR	0.479	0.949	589	-	-	-	0.880	0.680	329.864
			(0.105)	(0.011)	(82)	-	-	-	(0.012)	(0.012)	(11.580)
		SCR_3	0.187	0.814	62	2.722	4.446	9.628	0.527	0.112	398.265
			(0.063)	(0.053)	(10)	(0.282)	(0.647)	(0.547)	(0.015)	(0.012)	(15.268)
		SCR_3^*	0.167	0.818	57	2.803	4.504	9.614	0.527	0.108	397.671
			(0.058)	(0.054)	(10)	(0.346)	(0.627)	(0.558)	(0.016)	(0.012)	(15.463)
		SCR_4	0.280	0.918	210	2.762	4.811	10.168	0.472	0.125	429.791
			(0.073)	(0.019)	(16)	(0.410)	(0.758)	(0.549)	(0.015)	(0.013)	(17.141)

Continued on next page

Table 4.12: Average (SD) of $\hat{\Omega}$ related PDR, FDR, $|\hat{\Omega}|$ and matrix loss in three norms and $\hat{B}$ related PDR, FDR and PMSE for noise precision matrix design when $p = 200$, $n = 100$ and $h = 0.8$

			$\hat{\Omega}$						$\hat{B}$		
Graphs	X	Mtd.	PDR	FDR	$\lvert\hat{\Omega}\rvert$	$\lVert\cdot\rVert_S$	$\lVert\cdot\rVert_{\ell_1}$	$\lVert\cdot\rVert_F$	PDR	FDR	PMSE
	III	MRCE	0.005	0.255	1	5.313	5.957	15.629	0.087	0.268	986.534
			(0.010)	(0.396)	(1)	(1.687)	(1.296)	(0.944)	(0.017)	(0.153)	(39.754)
		aMCR	0.512	0.944	559	-	-	-	0.959	0.685	317.742
			(0.095)	(0.011)	(84)	-	-	-	(0.006)	(0.012)	(10.777)
		SCR_3	0.403	0.548	54	2.320	3.693	8.464	0.820	0.084	304.392
			(0.092)	(0.078)	(7)	(0.408)	(0.615)	(0.641)	(0.013)	(0.008)	(10.082)
		SCR_3^*	0.369	0.554	50	2.234	3.658	8.154	0.817	0.081	306.222
			(0.090)	(0.087)	(7)	(0.345)	(0.610)	(0.578)	(0.013)	(0.007)	(10.850)
		SCR_4	0.521	0.821	177	3.142	5.666	10.150	0.792	0.091	318.299
			(0.095)	(0.030)	(17)	(0.917)	(1.714)	(0.705)	(0.015)	(0.008)	(10.993)
	IV	MRCE	0.280	0.992	2114	3.071	5.284	11.823	0.006	0.119	1086.407
			(0.083)	(0.003)	(323)	(0.574)	(0.741)	(0.864)	(0.019)	(0.247)	(79.509)
		aMCR	0.799	0.933	731	-	-	-	0.600	0.641	274.199
			(0.085)	(0.013)	(116)	-	-	-	(0.022)	(0.023)	(13.690)
		SCR_3	0.475	0.573	67	2.159	3.513	7.709	0.518	0.259	290.417
			(0.108)	(0.085)	(12)	(0.367)	(0.723)	(0.632)	(0.029)	(0.027)	(18.342)
		SCR_3^*	0.446	0.574	63	2.054	3.441	7.402	0.518	0.256	288.261
			(0.099)	(0.081)	(11)	(0.293)	(0.659)	(0.564)	(0.028)	(0.026)	(16.961)
		SCR_4	0.441	0.921	334	2.403	5.307	9.120	0.229	0.427	466.535
			(0.091)	(0.015)	(25)	(0.329)	(1.093)	(0.511)	(0.028)	(0.037)	(33.710)
Tridiag	I	MRCE	0.000	0.030	0	6.396	7.249	23.816	0.797	0.565	535.034
			(0.000)	(0.171)	(0)	(0.863)	(0.576)	(1.217)	(0.021)	(0.057)	(22.071)
		aMCR	0.596	0.925	590	-	-	-	0.998	0.679	297.297
			(0.053)	(0.010)	(72)	-	-	-	(0.001)	(0.013)	(4.789)
		SCR_3	0.629	0.285	65	4.341	5.488	11.366	0.970	0.071	249.519
			(0.045)	(0.054)	(5)	(1.059)	(1.211)	(1.157)	(0.011)	(0.008)	(6.721)
		SCR_3^*	0.619	0.289	64	3.853	5.079	10.388	0.969	0.071	250.474
			(0.046)	(0.050)	(5)	(0.778)	(1.052)	(0.941)	(0.011)	(0.007)	(6.675)
		SCR_4	0.633	0.747	185	6.387	10.189	15.467	0.925	0.083	278.343
			(0.048)	(0.027)	(16)	(2.280)	(3.942)	(2.279)	(0.025)	(0.008)	(15.469)
	II	MRCE	0.000	0.100	0	6.447	7.251	24.528	0.245	0.108	777.184
			(0.000)	(0.302)	(0)	(0.743)	(0.492)	(0.804)	(0.019)	(0.058)	(18.914)
		aMCR	0.556	0.936	644	-	-	-	0.878	0.678	317.061
			(0.059)	(0.009)	(78)	-	-	-	(0.013)	(0.011)	(4.874)
		SCR_3	0.375	0.655	79	5.075	6.451	16.389	0.545	0.116	379.231
			(0.068)	(0.046)	(10)	(0.389)	(0.477)	(1.033)	(0.021)	(0.010)	(9.384)
		SCR_3^*	0.333	0.667	73	5.217	6.470	17.397	0.536	0.110	381.145
			(0.062)	(0.048)	(10)	(0.351)	(0.437)	(0.989)	(0.018)	(0.011)	(8.903)
		SCR_4	0.447	0.863	238	5.087	7.049	17.333	0.481	0.130	412.153
			(0.055)	(0.015)	(16)	(0.474)	(1.117)	(0.879)	(0.019)	(0.012)	(10.477)
	III	MRCE	0.000	0.020	0	6.262	7.196	25.131	0.110	0.064	923.802
			(0.000)	(0.141)	(0)	(0.341)	(0.417)	(0.790)	(0.165)	(0.161)	(114.191)
		aMCR	0.576	0.929	598	-	-	-	0.959	0.679	304.284
			(0.063)	(0.010)	(82)	-	-	-	(0.007)	(0.012)	(5.830)
		SCR_3	0.581	0.364	67	4.374	5.631	11.847	0.838	0.086	289.357
			(0.052)	(0.061)	(6)	(1.086)	(1.231)	(1.126)	(0.014)	(0.009)	(6.633)
		SCR_3^*	0.537	0.394	65	4.131	5.419	11.560	0.829	0.086	293.130
			(0.049)	(0.060)	(6)	(0.599)	(0.767)	(1.034)	(0.016)	(0.008)	(6.978)
		SCR_4	0.618	0.760	188	6.001	9.541	14.540	0.812	0.095	301.748
			(0.056)	(0.022)	(15)	(3.170)	(5.706)	(2.521)	(0.015)	(0.008)	(6.977)
	IV	MRCE	0.166	0.754	1490	5.785	7.041	21.825	0.205	0.575	865.088
			(0.113)	(0.426)	(895)	(0.340)	(0.509)	(1.541)	(0.278)	(0.189)	(397.917)
		aMCR	0.673	0.930	728	-	-	-	0.603	0.639	262.742
			(0.044)	(0.012)	(125)	-	-	-	(0.019)	(0.018)	(6.703)
		SCR_3	0.602	0.413	76	3.461	4.761	9.536	0.534	0.260	277.903
			(0.047)	(0.070)	(9)	(0.739)	(1.005)	(0.829)	(0.030)	(0.025)	(11.831)
		SCR_3^*	0.581	0.423	75	3.578	5.179	9.575	0.531	0.258	275.970
			(0.048)	(0.072)	(10)	(0.610)	(0.957)	(0.910)	(0.028)	(0.024)	(9.693)
		SCR_4	0.535	0.890	358	4.299	7.196	13.495	0.238	0.429	446.196
			(0.049)	(0.012)	(27)	(0.555)	(1.286)	(0.874)	(0.033)	(0.039)	(31.724)

Continued on next page

Table 4.12: Average (SD) of $\hat{\Omega}$ related PDR, FDR, $|\hat{\Omega}|$ and matrix loss in three norms and $\hat{B}$ related PDR, FDR and PMSE for noise precision matrix design when $p = 200$, $n = 100$ and $h = 0.8$

			$\hat{\Omega}$						$\hat{B}$		
Graphs	X	Mtd.	PDR	FDR	$\lvert\hat{\Omega}\rvert$	$\lVert\cdot\rVert_S$	$\lVert\cdot\rVert_{\ell_1}$	$\lVert\cdot\rVert_F$	PDR	FDR	PMSE
Hub	I	MRCE	0.000	0.810	5	8.230	14.417	18.695	0.024	0.698	1085.975
			(0.000)	(0.394)	(11)	(0.463)	(0.725)	(0.845)	(0.004)	(0.061)	(17.426)
		aMCR	0.461	0.961	536	-	-	-	0.998	0.689	330.918
			(0.085)	(0.008)	(84)	-	-	-	(0.002)	(0.010)	(6.272)
		SCR_3	0.529	0.496	48	4.226	8.700	9.735	0.965	0.069	276.185
			(0.080)	(0.069)	(7)	(1.286)	(1.972)	(0.998)	(0.012)	(0.006)	(7.824)
		SCR_3^*	0.516	0.492	46	3.984	8.294	9.298	0.964	0.066	276.031
			(0.080)	(0.070)	(7)	(1.316)	(1.986)	(0.952)	(0.012)	(0.006)	(7.736)
		SCR_4	0.718	0.812	172	7.279	14.096	13.592	0.910	0.080	312.853
			(0.103)	(0.030)	(14)	(3.428)	(5.690)	(2.806)	(0.023)	(0.007)	(15.098)
	II	MRCE	0.000	0.720	6	8.217	14.483	18.445	0.158	0.228	952.142
			(0.000)	(0.451)	(19)	(0.568)	(0.826)	(1.710)	(0.087)	(0.104)	(81.101)
		aMCR	0.395	0.967	550	-	-	-	0.879	0.683	349.499
			(0.085)	(0.009)	(84)	-	-	-	(0.012)	(0.012)	(6.278)
		SCR_3	0.151	0.871	53	7.417	13.742	15.030	0.521	0.112	421.058
			(0.054)	(0.044)	(10)	(0.527)	(0.829)	(1.051)	(0.018)	(0.011)	(9.183)
		SCR_3^*	0.140	0.871	50	7.453	13.725	15.280	0.522	0.108	419.085
			(0.047)	(0.043)	(8)	(0.517)	(0.817)	(1.027)	(0.018)	(0.011)	(9.234)
		SCR_4	0.269	0.938	195	7.338	13.956	15.216	0.467	0.126	453.611
			(0.063)	(0.014)	(18)	(0.769)	(1.401)	(1.015)	(0.015)	(0.012)	(10.808)
	III	MRCE	0.000	0.180	2	8.067	14.225	18.150	0.072	0.452	1046.472
			(0.000)	(0.386)	(13)	(0.510)	(0.785)	(1.165)	(0.020)	(0.130)	(21.319)
		aMCR	0.430	0.963	538	-	-	-	0.960	0.688	337.036
			(0.089)	(0.008)	(81)	-	-	-	(0.005)	(0.011)	(6.301)
		SCR_3	0.360	0.627	44	5.471	10.694	10.755	0.813	0.084	326.088
			(0.092)	(0.082)	(8)	(1.519)	(2.204)	(1.448)	(0.015)	(0.008)	(8.156)
		SCR_3^*	0.330	0.637	41	5.875	11.176	11.125	0.811	0.082	326.215
			(0.083)	(0.083)	(7)	(1.319)	(2.001)	(1.488)	(0.014)	(0.008)	(7.818)
		SCR_4	0.536	0.853	164	6.402	12.390	12.831	0.784	0.092	340.961
			(0.094)	(0.026)	(16)	(1.871)	(3.174)	(1.694)	(0.015)	(0.009)	(8.559)
	IV	MRCE	0.054	0.969	1957	7.453	13.658	15.747	0.063	0.576	1096.793
			(0.067)	(0.171)	(486)	(0.916)	(1.282)	(2.072)	(0.125)	(0.255)	(260.048)
		aMCR	0.717	0.952	692	-	-	-	0.595	0.651	289.348
			(0.069)	(0.010)	(124)	-	-	-	(0.023)	(0.020)	(9.000)
		SCR_3	0.438	0.657	59	4.528	9.200	9.523	0.508	0.260	305.404
			(0.082)	(0.080)	(11)	(1.055)	(1.747)	(1.100)	(0.030)	(0.022)	(13.719)
		SCR_3^*	0.410	0.655	55	4.959	9.810	9.803	0.510	0.256	302.722
			(0.072)	(0.083)	(11)	(1.082)	(1.789)	(1.152)	(0.029)	(0.021)	(12.149)
		SCR_4	0.442	0.938	322	5.923	12.007	12.426	0.217	0.437	485.836
			(0.081)	(0.012)	(24)	(0.791)	(1.663)	(1.013)	(0.034)	(0.040)	(33.909)
BA	I	MRCE	0.113	0.926	74	11.376	20.842	20.728	0.588	0.579	1340.454
			(0.055)	(0.030)	(20)	(2.510)	(5.208)	(1.820)	(0.019)	(0.064)	(53.656)
		aMCR	0.708	0.937	554	-	-	-	0.998	0.683	474.999
			(0.067)	(0.010)	(64)	-	-	-	(0.001)	(0.013)	(20.765)
		SCR_3	0.696	0.430	60	4.931	10.807	9.300	0.969	0.072	401.614
			(0.073)	(0.083)	(6)	(1.852)	(4.257)	(1.422)	(0.008)	(0.008)	(16.799)
		SCR_3^*	0.681	0.433	59	5.466	11.293	9.321	0.968	0.072	406.645
			(0.076)	(0.078)	(6)	(2.099)	(4.457)	(1.515)	(0.008)	(0.008)	(18.046)
		SCR_4	0.746	0.811	195	6.097	13.239	11.922	0.923	0.084	447.722
			(0.076)	(0.028)	(16)	(2.478)	(4.715)	(1.740)	(0.023)	(0.008)	(28.566)
	II	MRCE	0.123	0.931	87	10.114	18.245	19.233	0.170	0.094	1538.776
			(0.048)	(0.024)	(17)	(2.235)	(4.673)	(1.125)	(0.017)	(0.059)	(52.065)
		aMCR	0.677	0.944	604	-	-	-	0.880	0.678	520.683
			(0.075)	(0.009)	(79)	-	-	-	(0.013)	(0.011)	(21.511)
		SCR_3	0.535	0.684	84	7.045	13.425	12.271	0.568	0.116	595.683
			(0.095)	(0.058)	(10)	(2.268)	(4.558)	(1.826)	(0.020)	(0.012)	(25.206)
		SCR_3^*	0.416	0.728	76	8.380	15.572	14.331	0.546	0.110	622.745
			(0.082)	(0.054)	(10)	(1.967)	(4.096)	(1.372)	(0.019)	(0.010)	(25.918)
		SCR_4	0.551	0.898	266	8.193	16.013	14.371	0.495	0.131	667.707
			(0.088)	(0.017)	(18)	(2.133)	(4.457)	(1.477)	(0.020)	(0.011)	(31.128)

Continued on next page

Table 4.12: Average (SD) of $\hat{\Omega}$ related PDR, FDR, $|\hat{\Omega}|$ and matrix loss in three norms and $\hat{B}$ related PDR, FDR and PMSE for noise precision matrix design when $p = 200$, $n = 100$ and $h = 0.8$

			$\hat{\Omega}$						$\hat{B}$		
Graphs	X	Mtd.	PDR	FDR	$\lvert\hat{\Omega}\rvert$	$\lVert\cdot\rVert_S$	$\lVert\cdot\rVert_{\ell_1}$	$\lVert\cdot\rVert_F$	PDR	FDR	PMSE
	III	MRCE	0.136	0.924	87	10.483	18.930	19.724	0.108	0.118	1631.707
			(0.059)	(0.026)	(27)	(2.499)	(5.338)	(1.033)	(0.172)	(0.221)	(125.101)
		aMCR	0.690	0.940	577	-	-	-	0.960	0.684	500.015
			(0.068)	(0.010)	(72)	-	-	-	(0.005)	(0.012)	(18.654)
		SCR_3	0.699	0.457	64	4.801	10.141	8.936	0.848	0.085	451.447
			(0.062)	(0.060)	(5)	(1.684)	(3.843)	(1.115)	(0.014)	(0.008)	(14.852)
		SCR_3^*	0.623	0.504	62	6.535	12.515	10.328	0.838	0.086	468.992
			(0.076)	(0.068)	(6)	(2.340)	(4.804)	(1.739)	(0.016)	(0.008)	(17.410)
		SCR_4	0.744	0.813	196	6.195	12.992	11.471	0.821	0.095	474.286
			(0.061)	(0.024)	(15)	(2.634)	(5.148)	(1.842)	(0.017)	(0.009)	(17.872)
	IV	MRCE	0.558	0.985	1803	10.889	20.299	18.040	0.000	0.009	1733.840
			(0.169)	(0.003)	(469)	(2.642)	(5.419)	(1.535)	(0.001)	(0.086)	(114.002)
		aMCR	0.788	0.940	667	-	-	-	0.604	0.643	431.305
			(0.065)	(0.013)	(127)	-	-	-	(0.020)	(0.023)	(20.429)
		SCR_3	0.681	0.536	73	5.663	11.367	9.277	0.534	0.265	465.611
			(0.076)	(0.081)	(9)	(2.191)	(4.670)	(1.616)	(0.030)	(0.026)	(39.213)
		SCR_3^*	0.643	0.551	72	6.768	12.892	10.275	0.528	0.263	459.273
			(0.075)	(0.081)	(10)	(2.444)	(5.039)	(1.898)	(0.029)	(0.025)	(29.518)
		SCR_4	0.575	0.922	363	7.902	16.357	12.856	0.243	0.428	735.612
			(0.080)	(0.013)	(26)	(2.425)	(5.146)	(1.681)	(0.032)	(0.036)	(55.233)

Table 4.13: Average PSE, number of included genes of the SCR methods with the standard errors in parentheses

	SCR_1	SCR_1^*	SCR_3	SCR_3^*	SCR_4
PSE	1.195	1.182	1.193	1.187	1.193
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Num. Genes	46	45	43	44	44
	(0.687)	(0.640)	(0.624)	(0.648)	(0.634)

Table 4.14: Average PSE, number of included genes of CW, PWL DML and aMCR with the standard errors in parentheses

	CW	PWL	DML	aMCR
PSE	1.298	1.248	1.229	1.190
	(0.038)	(0.032)	(0.032)	(0.012)
Num. Genes	500	17	78	65
	(0.000)	(13.565)	(32.151)	(1.750)

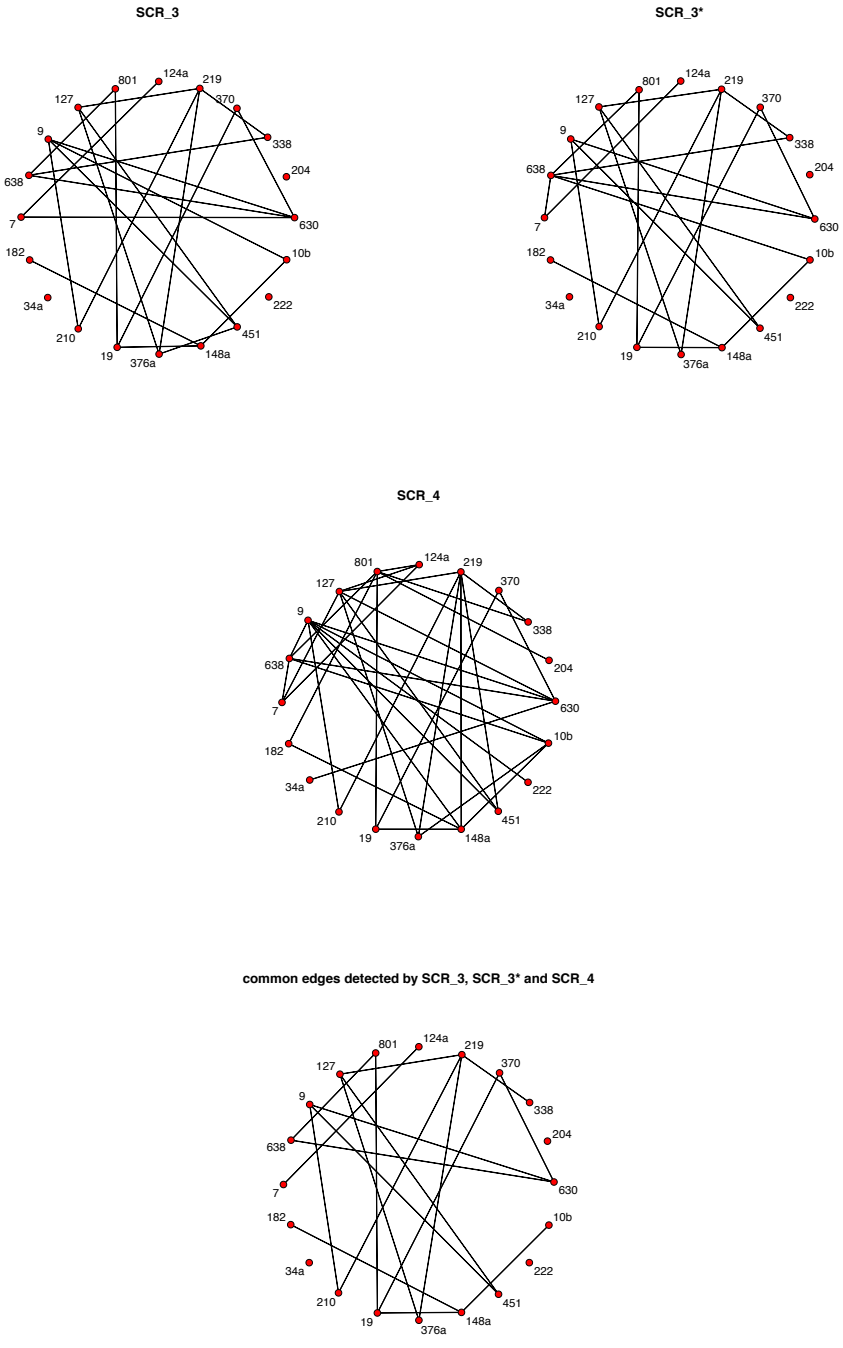

Figure 4.1: Graphical networks of the twenty selected microRNAs detected by the SCR methods

Chapter 5

Conclusion and Future Research

This study proposed to solve the relevant problems in graphical models and multivariate response regression models with several sequential methods. First, the SSPS edge detection method was introduced in the context of nonparanormal graphical models. Essentially, it is used under Gaussian graphical models. Enabled by the nonparanormal formulation, it has wider applications. Our simulation results showed that its performance under different nonparanormal inverse transformations is quite stable and parallel to that under the corresponding Gaussian graphical models. In general, it is a competitive edge detection tool. However, the unsatisfactory performance of all the considered methods in the Cluster graph and RP graph drew our attention. For the former, there is no sparsity within each clusters, but the overall network may be still sparse. The network in the latter graph is denser and more complicated. Further research can be devoted to the edge detection of such graphs. Moreover, it is also of our interest to expand the sequential approach to the matrix-variate graphical models, where the observations are structured data or matrix-variate data. Relevant work can

be found in Leng and Tang (2012) and the references therein.

After the edge detection, the precision matrix estimation with SSPS was discussed. The constrained MLE approach generally outperforms the others regardless of the dimensionality. Note that its preciseness closely depends on that of the SSPS edge detection. More than a dimension reduction tool, the SSPS screening provides an insight on integrating the prior knowledge into the estimation procedure and can contribute to more flexibility in practice. This screening process has been applied to some existing methods, where the regularization parameter is tuned by a special version of EBIC. Improvements in edge detection and precision matrix estimation can be observed in the simulations and the real data analysis. Even so, the choice of the screening parameter α worths additional examination.

In the study of multivariate response regression models, two sequential conditional regression (SCR) methods were proposed for the joint estimation of the covariance matrix and the precision matrix. They cleverly utilize the sequential methods in univariate response regression models and Gaussian graphical models. Accordingly, their selection consistency has been established on that of the component sequential methods. The numerical results demonstrated that the method with the alternate updating scheme is quite competitive in terms of model selection and prediction accuracy. However, the other with the simultaneous estimation scheme, to some extent, is limited by the separate treatment of the conditional regressions. It is desirable to improve the second approach by certain joint formulation of these conditional regressions in future.

Bibliography

Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *The Annals of Mathematical Statistics 22*(3), 327–351.

Banerjee, O., L. El Ghaoui, and A. d'Aspremont (2008). Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. *The Journal of Machine Learning Research 9*, 485–516.

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

Bickel, P. J., Y. Ritov, and A. B. Tsybakov (2009). Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics 37*(4), 1705–1732.

Breiman, L. and J. H. Friedman (1997). Predicting multivariate responses in multiple linear regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 59*(1), 3–54.

Cai, T., W. Liu, and X. Luo (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association 106*(494), 594–607.

Cai, T. T., W. Liu, and X. Luo (2012). *clime: Constrained L1-minimization for Inverse (covariance) Matrix Estimation*. R package version 0.4.1.

Cai, T. T. and L. Wang (2011). Orthogonal matching pursuit for sparse signal recovery with noise. *IEEE Transactions on Information Theory 57*(7), 4680–4688.

Cai, T. T., C.-H. Zhang, and H. H. Zhou (2010). Optimal rates of convergence for covariance matrix estimation. *The Annals of Statistics 38*(4), 2118–2144.

Candes, E. and T. Tao (2007). The dantzig selector: statistical estimation when p is much larger than n. *The Annals of Statistics 35*(6), 2313–2351.

Chen, J. and Z. Chen (2008). Extended bayesian information criteria for model selection with large model spaces. *Biometrika 95*(3), 759–771.

Chen, L. and J. Z. Huang (2012). Sparse reduced-rank regression for simultaneous dimension reduction and variable selection. *Journal of the American Statistical Association 107*(500), 1533–1545.

Chen, S. S., D. L. Donoho, and M. A. Saunders (1998). Atomic decomposition by basis pursuit. *Journal on Scientific Computing 20*(1), 33–61.

Dempster, A. P. (1972). Covariance selection. *Biometrics 28*(1), 157–175.

Efron, B., T. Hastie, I. Johnstone, R. Tibshirani, et al. (2004). Least angle regression. *The Annals of Statistics 32*(2), 407–499.

Fan, J., Y. Feng, and Y. Wu (2009). Network exploration via the adaptive lasso and scad penalties. *The Annals of Applied Statistics 3*(2), 521–541.

Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association 96*(456), 1348–1360.

Fan, J., H. Peng, et al. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *The Annals of Statistics 32*(3), 928–961.

Foygel, R. and M. Drton (2010). Extended bayesian information criteria for gaussian graphical models. In *Advances in Neural Information Processing Systems*, pp. 604–612.

Frank, L. E. and J. H. Friedman (1993). A statistical view of some chemometrics regression tools. *Technometrics 35*(2), 109–135.

Friedman, J., T. Hastie, H. Höfling, R. Tibshirani, et al. (2007). Pathwise coordinate optimization. *The Annals of Applied Statistics 1*(2), 302–332.

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

Friedman, J., T. Hastie, and R. Tibshirani (2010). Applications of the lasso and grouped lasso to the estimation of sparse graphical models. Technical report, Stanford University.

Friedman, J., T. Hastie, and R. Tibshirani (2014). *glasso: Graphical lasso-estimation of Gaussian graphical models*. R package version 1.8.

He, Y. and Z. Chen (2014). The ebic and a sequential procedure for feature selection in interactive linear models with high-dimensional data. *Annals of the Institute of Statistical Mathematics*, 1–26.

Hess, K. R., K. Anderson, W. F. Symmans, V. Valero, N. Ibrahim, J. A. Mejia, D. Booser, R. L. Theriault, A. U. Buzdar, P. J. Dempsey, et al. (2006). Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. *Journal of Clinical Oncology 24*(26), 4236–4244.

Huang, J., S. Ma, and C.-H. Zhang (2008). Adaptive lasso for sparse high-dimensional regression models. *Statistica Sinica 18*(4), 1603–1618.

Huang, J. Z., N. Liu, M. Pourahmadi, and L. Liu (2006). Covariance matrix selection and estimation via penalised normal likelihood. *Biometrika 93*(1), 85–98.

Izenman, A. J. (1975). Reduced-rank regression for the multivariate linear model. *Journal of Multivariate Analysis 5*(2), 248–264.

Knight, K. and W. Fu (2000). Asymptotics for lasso-type estimators. *The Annals of Statistics 28*(5), 1356–1378.

Kuerer, H. M., L. A. Newman, T. L. Smith, F. C. Ames, K. K. Hunt, K. Dhingra, R. L. Theriault, G. Singh, S. M. Binkley, N. Sneige, et al. (1999). Clinical course of breast cancer patients with complete pathologic primary tumor and axillary lymph node response to doxorubicin-based neoadjuvant chemotherapy. *Journal of Clinical Oncology 17*(2), 460–460.

Lam, C. and J. Fan (2009). Sparsistency and rates of convergence in large covariance matrix estimation. *The Annals of Statistics 37*(6B), 4254–4278.

Lauritzen, S. L. (1996). *Graphical models*. Oxford University Press.

Lee, W. and Y. Liu (2012). Simultaneous multiple response regression and inverse covariance matrix estimation via penalized gaussian maximum likelihood. *Journal of Multivariate Analysis 111*, 241–255.

Leng, C., Y. Lin, and G. Wahba (2006). A note on the lasso and related procedures in model selection. *Statistica Sinica 16*(4), 1273–1284.

Leng, C. and C. Y. Tang (2012). Sparse matrix graphical models. *Journal of the American Statistical Association 107*(499), 1187–1200.

Liu, H., F. Han, M. Yuan, J. Lafferty, and L. Wasserman (2012). High-dimensional semiparametric gaussian copula graphical models. *The Annals of Statistics 40*(4), 2293–2326.

Liu, H., J. Lafferty, and L. Wasserman (2009). The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *The Journal of Machine Learning Research 10*, 2295–2328.

Luo, S. and Z. Chen (2014a). Edge detection in sparse gaussian graphical models. *Computational Statistics & Data Analysis 70*, 138–152.

Luo, S. and Z. Chen (2014b). Sequential lasso cum ebic for feature selection with ultra-high dimensional feature space. *Journal of the American Statistical Association 109*(507), 1229–1240.

McLendon, R., A. Friedman, D. Bigner, E. G. Van Meir, D. J. Brat, G. M. Mastrogianakis, J. J. Olson, T. Mikkelsen, N. Lehman, K. Aldape, et al. (2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature 455*(7216), 1061–1068.

Meinshausen, N. and P. Bühlmann (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics 34*(3), 1436–1462.

Pati, Y. C., R. Rezaiifar, and P. Krishnaprasad (1993). Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*, pp. 40–44. IEEE.

Peng, J., P. Wang, N. Zhou, and J. Zhu (2009). Partial correlation estimation by joint sparse regression models. *Journal of the American Statistical Association 104*(486), 735–746.

Peng, J., P. Wang, N. Zhou, and J. Zhu. (2010). *space: Sparse PArtial Correlation Estimation*. R package version 0.1-1.

Peng, J., J. Zhu, A. Bergamaschi, W. Han, D.-Y. Noh, J. R. Pollack, and P. Wang (2010). Regularized multivariate regression for identifying master predictors with application to integrative genomics study of breast cancer. *The Annals of Applied Statistics 4*(1), 53–77.

Reinsel, G. C. and R. P. Velu (1998). *Multivariate reduced-rank regression*. Springer.

Rothman, A. J. (2013). *MRCE: Multivariate regression with covariance estimation*. R package version 2.0.

Rothman, A. J., E. Levina, and J. Zhu (2010). Sparse multivariate regression with covariance estimation. *Journal of Computational and Graphical Statistics 19*(4), 947–962.

Sun, T. and C.-H. Zhang (2012). Scaled sparse linear regression. *Biometrika 99*(4), 879–898.

Sun, T. and C.-H. Zhang (2013). Sparse matrix inversion with scaled lasso. *The Journal of Machine Learning Research 14*(1), 3385–3418.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological) 58*(1), 267–288.

Turlach, B. A., W. N. Venables, and S. J. Wright (2005). Simultaneous variable selection. *Technometrics 47*(3), 349–363.

Vandenberghe, L., S. Boyd, and S.-P. Wu (1998). Determinant maximization with linear matrix inequality constraints. *Journal on Matrix Analysis and Applications 19*(2), 499–533.

Verhaak, R. G., K. A. Hoadley, E. Purdom, V. Wang, Y. Qi, M. D. Wilkerson, C. R. Miller, L. Ding, T. Golub, J. P. Mesirov, et al. (2010). Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in pdgfra, idh1, egfr, and nf1. *Cancer cell 17*(1), 98–110.

Wang, J. (2013). Joint estimation of sparse multivariate regression and conditional graphical models. *arXiv preprint arXiv:1306.4410*.

Wu, W. B. and M. Pourahmadi (2003). Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika 90*(4), 831–844.

Yin, J. and H. Li (2011). A sparse conditional gaussian graphical model for analysis of genetical genomics data. *The Annals of Applied Statistics 5*(4), 2630–2650.

Yuan, M. (2010). High dimensional inverse covariance matrix estimation via linear programming. *The Journal of Machine Learning Research 99*, 2261–2286.

Yuan, M., A. Ekici, Z. Lu, and R. Monteiro (2007). Dimension reduction and coefficient estimation in multivariate linear regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69*(3), 329–346.

Yuan, M. and Y. Lin (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68*(1), 49–67.

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

Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics 38*(2), 894–942.

Zhao, P. and B. Yu (2006). On model selection consistency of lasso. *The Journal of Machine Learning Research 7*, 2541–2563.

Zhou, S., P. Rütimann, M. Xu, and P. Bühlmann (2011). High-dimensional covariance estimation based on gaussian graphical models. *The Journal of Machine Learning Research 12*, 2975–3026.

Zhou, S., S. van de Geer, and P. Bühlmann (2009). Adaptive lasso for high dimensional regression and gaussian graphical modeling. *arXiv preprint arXiv:0903.2515*.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association 101*(476), 1418–1429.

Zou, H. and R. Li (2008). One-step sparse estimates in nonconcave penalized likelihood models. *The Annals of Statistics 36*(4), 1509–1533.