Statistics and Computing (2025) 35:94
https://doi.org/10.1007/s11222-025-10625-7

ORIGINAL PAPER

On sufficient dimension reduction methods based on a graphical model with non-concave penalty

Yujin Park[1] · **Kyongwon Kim**[2] · **Jae Keun Yoo**[1]

Received: 12 September 2024 / Accepted: 22 April 2025

Abstract
Ultra high-dimensional datasets, which refer to scenarios where the number of covariates grows at an exponential rate relative to the sample size, are frequently encountered in modern data analysis across fields such as genomics, finance, and social sciences. These datasets pose significant challenges due to the large number of variables relative to the number of observations, potentially resulting in issues such as multicollinearity, overfitting, and computational difficulties. Traditional sufficient dimension reduction (SDR) methods struggle with these challenges, making it necessary to develop new approaches. To address these limitations, we introduce a graphical model-based SDR method that incorporates a smoothly clipped absolute deviation (SCAD) penalty. This method effectively reduces dimensionality while managing sparsity in the dataset. Additionally, we extend directional regression for high-dimensional data by integrating them with graphical LASSO, which enhances the model's ability to estimate sparse precision matrices. This combined approach not only mitigates computational infeasibility in estimating covariance matrices but also helps avoid overfitting, making it particularly suitable for high-dimensional contexts. Through extensive simulation studies and real-world data analyses, we validate the robustness and effectiveness of our proposed methods. Moreover, we provide a theoretical framework that discusses the convergence rate of these methods, offering insights into their performance under various conditions. Finally, we outline potential avenues for future research, including exploring alternative penalty functions and expanding the applicability of these methods to other types of data structures.

Keywords sufficient dimension reduction · graphical model · precision matrix · SCAD · graphical LASSO

1 Introduction

SDR is one of the most powerful dimension reduction methods (Li 21; Cook and Weisberg 8; Li 24; Cook 7). The main objective of SDR is to identify a subspace so that we can project the original data into a low-dimensional subspace without loss of information. SDR has been widely applied in a variety of research areas, including image classification (Benouareth 2), graphical models (Li and Kim 25; Kim 18), causal inference (Ma et al. 33; Cheng et al. 5), and biomedical research (Yoo 47; Tomassi et al. 39).

Despite recent advances in SDR methods, theoretical developments remain limited for datasets in which the dimension (p) exceeds the number of observations (n). This is because of the necessity of computing inverse of the sample covariance matrix. However, estimating the precision matrix in an ultra high-dimensional dataset is by no means easy (Fan et al. 14) as the sample covariance matrix is not invertible.

To overcome this limitation, [23] approximated sparse SDR by merging shrinkage estimation with existing SDR methods such as sliced inverse regression (SIR; Li 21) and sliced average variance estimation (SAVE; Cook and Weisberg 8). Sparse SDR converts the SDR problem as an optimization framework. They added an L_1 penalty to achieve sparsity. They confirmed that this estimation strategy converges fast enough and reaches the global minimum, most of the time. [45] invented a sequential SDR algorithm to address large p and small n problems. This algorithm

✉ Kyongwon Kim
kimk@yonsei.ac.kr

✉ Jae Keun Yoo
peter.yoo@ewha.ac.kr

Yujin Park
yujin.alex.p@gmail.com

[1] Department of Statistics, Ewha Womans University, 52 Ewhayeodae-gil, Seodaemun-gu, Seoul 03760, Republic of Korea

[2] Department of Applied Statistics, Department of Statistics and Data Science, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea

Published online: 15 May 2025

partitions the data into blocks and reduces the data sequentially. [29] developed a diagonal thresholding screening SIR, which focuses on applying the method in high-dimensional cases. [41] introduced a convex optimization approach to fit sparse SIR in high-dimensional data. Using a group-Dantzig selector type formulation approach, [16] induced row-sparsity to the SIR dimension reduction vectors. [34] developed the graph informed SIR (GraphSIR) and graph informed SAVE (GraphSAVE). They employed the graphical LASSO (Glasso) (Friedman et al. 12) to address the challenges associated with SIR and SAVE in high-dimensional settings. For additional advances, see [35] and [40].

In this paper, we propose a graphical model based SDR method using a SCAD penalty (Fan and Li 13) to overcome the difficulties from traditional SDR methods for analyzing high dimension, low sample size dataset. SCAD combines variable selection and coefficient estimation simultaneously and this allows for more accurate and efficient modeling. The penalty functions employed are symmetric, nonconcave and having singularities at the origin. These penalty functions reduce bias in large coefficients and ensure continuous solutions, in contrast to methods such as LASSO, which may introduce bias. The SCAD method is widely applicable across different types of models, including parametric, nonparametric, and generalized linear models. In terms of sparsity and model selection, the SCAD penalty promotes sparse solutions similarly to Glasso, but it provides the additional advantage of being nonconvex. This nonconvex nature allows our proposed method to provide even sparser solutions than Glasso based methods (GraphSIR, GraphSAVE) by shrinking smaller coefficients to zero more effectively with a more accurate identification of non-zero entries in the precision matrix.

Our method combines the SCAD-penalized precision matrix estimation with SDR methods, including SIR, SAVE, and directional regression (DR; Li and Wang 27), to estimate the SDR subspaces. We further present the Glasso based DR, which is an extension of GraphSIR and GraphSAVE. The proposed method prevents users from the computational infeasibility of estimating the covariance matrix and overfitting problems.

We demonstrate that our method provides a robust estimation of the SDR subspace. In a theoretical perspective, we show that our method can consistently estimate the SDR subspace in high-dimensional settings. Moreover, simulation studies and real data analyses confirm that our method outperforms in statistical accuracy in high-dimensional settings.

The paper is organized as follows. In Section 2, we introduce methods of SDR and precision matrix estimation. In Section 3, we present our method for the data with large p small n and provide technical proof of our method . We report the simulation study results in Section 4 and give numerical results from real datasets in Section 5. Finally, Section 6 contains the conclusion remarks and the potential future studies. Computational time and memory usage for both simulation studies and real data analyses are provided in the Appendix.

2 Preliminaries

2.1 Sufficient dimension reduction

SDR in a regression focuses on identifying linear combinations $\nu^{\mathsf{T}}\mathbf{X}$, which can explain all the distributional relationship in $Y \in \mathbb{R}$ given $\mathbf{X} \in \mathbb{R}^p$. Here, ν stands for a $p \times q$ matrix, with $q \ll p$ and the relationship can be expressed as

$$Y \perp\!\!\!\perp \mathbf{X} | \nu^{\mathsf{T}}\mathbf{X},$$

where $\perp\!\!\!\perp$ denotes statistical independence. This also can be regarded as the original p-dimensional predictors $\mathbf{X}$ can be replaced by reduced q-dimensional predictors $\nu^{\mathsf{T}}\mathbf{X}$, without loss of information on the conditional distribution of Y given $\mathbf{X}$. Since the matrix ν itself is not identifiable, it is essential to focus on the space spanned by its columns. This space is referred to as the dimension reduction subspace (DRS), denoted by $\mathcal{S}(\nu)$. In this framework, the matrix ν is not uniquely defined. The intersection of all DRS is itself a DRS and is called the central subspace (CS), which is denoted by $\mathcal{S}_{Y|\mathbf{X}}$. The primary goal of SDR is to estimate $\mathcal{S}_{Y|\mathbf{X}}$. Let

$$\mathbf{Z} = \Sigma^{-1/2}(\mathbf{X} - \mu),$$

denotes the standardized version of $\mathbf{X}$, where μ and Σ are the mean and covariance of $\mathbf{X}$, respectively. Consequently, there exists a relationship between the CSs of $Y|\mathbf{X}$ and $Y|\mathbf{Z}$ as

$$\mathcal{S}_{Y|\mathbf{X}} = \Sigma^{-1/2}\mathcal{S}_{Y|\mathbf{Z}}.$$

This process significantly enhances the accuracy of CS estimation and ensures stable results. Many SDR methods, including SIR, SAVE, and DR, rely on matrix operations such as covariance estimation, eigenvalue decomposition, or conditional mean calculations. These operations are highly sensitive to the scale and variability of the predictors. If the variables in X are not standardized, the differences in their scales can lead to numerical instability, skewing the results and potentially introducing bias or inaccuracies in the estimation process. Standardizing X secures that all variables are on a comparable scale, allowing matrix computations to remain stable and reliable. Intuitively, standardization prevents dominant variables with large variances from disproportionately influencing the results, which could overshadow the contributions of other variables. This step is particularly crucial in

high-dimensional settings where variable scaling differences are amplified.

We denote the basis matrix for the CS of $\mathcal{S}_{Y|\mathbf{Z}}$ as $\nu_{\mathbf{Z}}$. For most SDR methods, either one or both of the following linear conditional mean and constant conditional variance conditions are required.

Assumption 1 (Linear conditional mean condition) $\mathbf{E}[\mathbf{Z}|\nu_{\mathbf{Z}}^{\mathsf{T}}\mathbf{Z}]$ is linear in $\nu_{\mathbf{Z}}^{\mathsf{T}}\mathbf{Z}$ resulting in $\mathbf{E}[\mathbf{Z}|\nu_{\mathbf{Z}}^{\mathsf{T}}\mathbf{Z}] = P_{\nu_{\mathbf{Z}}}\mathbf{Z}$ where $P_{\nu_{\mathbf{Z}}}$ is a projection onto a subspace of $\nu_{\mathbf{Z}}$.

This can be satisfied when $\mathbf{Z}$ having an elliptical distribution (Eaton 10).

Assumption 2 (Constant conditional variance condition) $\mathbf{var}(\mathbf{Z}|\nu_{\mathbf{Z}}^{\mathsf{T}}\mathbf{Z})$ is nonrandom, which is equivalent to $\mathbf{Z}$ having a normal distribution (Cook 6).

A number of studies have been conducted for estimating a CS. SIR was introduced by [21], utilizing the first conditional moment $E(\mathbf{Z}|Y)$. Under Assumption 1, the following relationship can be established

$$\mathcal{S}\{E(\mathbf{Z}|Y)\} \subseteq \mathcal{S}_{Y|\mathbf{Z}},$$

which implies that a CS can be recovered from $E(\mathbf{Z}|Y)$.

Although SIR has significantly contributed to advancements in dimension reduction, it fails to enable the recovery of $\mathcal{S}_{Y|\mathbf{Z}}$ when $E(\mathbf{Z}|Y) = 0$ for all Y, which means a regression surface is symmetric with respect to 0. To address this limitation of SIR, [8] proposed SAVE. This indicated $\mathrm{cov}(\mathbf{Z}|Y)$ changes from one slice to another when $E(\mathbf{Z}|Y) = 0$ for all Y and considered the second moment of $E(\mathbf{Z}|Y)$ to estimate $\mathcal{S}_{Y|\mathbf{Z}}$. Under Assumptions 1 and 2, we can derive $E[I_p - \mathbf{var}(\mathbf{Z}|Y)] = P_{\nu_{\mathbf{Z}}}E[I_p - \mathbf{var}(\mathbf{Z}|Y)]P_{\nu_{\mathbf{Z}}}$. It is possible to establish the following,

$$\begin{aligned}&\mathcal{S}\{I_p - \mathrm{cov}(\mathbf{Z}|Y)\}\\ &= \mathcal{S}\big[\{I_p - \mathrm{cov}(\mathbf{Z}|Y)\}^2\big] \subseteq \mathcal{S}_{Y|\mathbf{Z}}.\end{aligned}$$

While both SIR and SAVE use the conditional moment of $\mathbf{Z}$, DR (Li and Wang 27) is based on the conditional moment of $\mathbf{Z} - \tilde{\mathbf{Z}}$, where $\tilde{\mathbf{Z}}$ is an independent copy of $\mathbf{Z}$ (Li and Wang 27; Kim and Yoo 19). [28] introduced a contour regression by using the idea of empirical directions $\{\mathbf{Z}_i - \mathbf{Z}_j : 1 \le i \le j \le n\}$. Since empirical directions can capture positional information about $\mathbf{Z}$, it is helpful to uncover the empirical distribution. [27] improved the computational inefficiency of empirical directions by introducing DR which regresses $(\mathbf{Z} - \tilde{\mathbf{Z}})(\mathbf{Z} - \tilde{\mathbf{Z}})^{\mathsf{T}}$ onto the space of $(Y - \tilde{Y})$. Suppose Assumptions 1 and 2 hold and let $(\tilde{\mathbf{Z}}, \tilde{Y})$ be an independent copy of $(\mathbf{Z}, Y)$. Then, it is trivial that $\mathbf{E}\big[(\mathbf{Z} - \tilde{\mathbf{Z}})(\mathbf{Z} - \tilde{\mathbf{Z}})^{\mathsf{T}}|(Y, \tilde{Y})\big] - 2I_p = P_{\nu_{\mathbf{Z}}}\big[\mathbf{E}[(\mathbf{Z} - \tilde{\mathbf{Z}})(\mathbf{Z} - \tilde{\mathbf{Z}})^{\mathsf{T}}|(Y, \tilde{Y})] - 2I_p\big]P_{\nu_{\mathbf{Z}}}$. This can be expressed as

$$\mathcal{S}\big[\mathbf{E}[(\mathbf{Z} - \tilde{\mathbf{Z}})(\mathbf{Z} - \tilde{\mathbf{Z}})^{\mathsf{T}}|(Y, \tilde{Y})] - 2I_p\big] = \mathcal{S}_{Y|\mathbf{Z}}.$$

2.2 Sparse precision matrix

Several methods have been proposed for estimating sparse covariance and inverse of the covariance matrices in high-dimensional settings. These methods include penalized likelihood approaches (Fan and Li 13; Friedman et al. 12; Zhang 50), column-by-column estimation methods (Meinshausen and Bühlmann 30; Yuan 48; Sun and Zhang 38), and a method that focuses on symmetric precision matrix estimation (Zhao and Liu 52). To address the estimation problem of $\mathcal{S}_{Y|\mathbf{Z}}$ associated with a large covariance matrix in an high-dimensional dataset, we incorporate well-known penalized likelihood algorithms such as SCAD and Glasso to the SDR problem. These algorithms are based on the extensively studied theoretical properties of the underlying penalties, as demonstrated by [36], and [20]. Moreover, their computational efficiency enables them to be utilized in practical applications. The basic idea of the penalized likelihood method for estimating precision matrix is minimizing the penalized negative log-likelihood function, denoted by

$$l(\Theta) = \mathrm{trace}(S\Theta) - \log\det\Theta + \sum_{a\neq b} P_\lambda(|\theta_{ab}|),$$

where Θ is an inverse of a covariance matrix, called precision matrix, θ_{ab} is the (a, b)-element of Θ, and λ is the tuning parameter. $P_\lambda(\cdot)$ denotes the penalty function that increases sparsity in the precision matrix and $S = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^{\mathsf{T}}$. This formula assumes that $\mathbf{X}$ follows an independent and identically distributed multivariate normal distribution with mean μ and the positive definite covariance matrix Σ. This approach provides a balanced solution to the problem of high dimensions, striving for accuracy in estimation while moderating computational demands and promoting interpretability through sparsity. For more details, see [13].

The Glasso considers L_1 regularization, expressed by $P_\lambda(|\tau|) = \lambda|\tau|$. They proposed a method of determining conditional independence based on the sparsity of the precision matrix. That is, the variables a and b are conditionally independent, given the other variables if Σ_{ab}^{-1} is equal to zero. For more information on the Glasso, see [46], [12], and [31].

In this paper, we utilize the SCAD penalty to achieve better accuracy in estimating a precision matrix and eventually, SDR. The thresholding penalty of SCAD can effectively combine the thresholding and shrinkage, which is defined as

$$P_\lambda(|\tau|) = \begin{cases}\lambda|\tau|, & \text{if } |\tau| \le \lambda,\\ \frac{-(\tau^2 - 2a\lambda|\tau| + \lambda^2)}{2(a-1)}, & \text{if } \lambda < |\tau| \le a\lambda,\\ \frac{(a+1)\lambda^2}{2}, & \text{if } |\tau| > a\lambda,\end{cases} \tag{1}$$

for $\lambda > 0$ and $a > 2$. [20] utilized the local linear approximation (LLA; Zou and Li 51) for estimating the precision estimate with nonconcave penalized method. By a linear function, the LLA algorithm iteratively and locally approximates the penalty function as

$$P_\lambda(|\theta_{ab}|) \approx P_\lambda(|\theta_{ab}^{(0)}|) + P'_\lambda(|\theta_{ab}^{(0)}|)(|\theta_{ab}| - |\theta_{ab}^{(0)}|), \text{ for } \theta_{ab} \approx \theta_{ab}^{(0)},$$

where $\theta_{ab}^{(0)}$ is the unpenalized maximum likelihood estimate and $P_\lambda(\theta)$ denotes the penalty function of SCAD. It can be represented by

$$l(\Theta) \approx \text{trace}(S\Theta) - \log\det\Theta + \sum_{a\neq b}[P_\lambda(|\theta_{ab}^{(0)}|) + P'_\lambda(|\theta_{ab}^{(0)}|)(|\theta_{ab}| - |\theta_{ab}^{(0)}|)].$$

For $m = 1, 2, \cdots$, iteratively solve until the sequence converges

$$\Theta^{(m+1)} = \arg\min_{\Theta\succ 0}\Big\{\text{trace}(S\Theta) - \log\det\Theta + \sum_{a\neq b} P'_\lambda(|\theta_{ab}^{(m)}|)|\theta_{ab}|\Big\}.$$

The SCAD penalty provides a balanced approach between hard and soft thresholding. Although the L_1 penalty is known for its efficacy in variable selection, it shrinks all coefficients uniformly from time to time and this induces a possibility of reducing the influence of significant variables. The SCAD penalty, however, adjusts the level of regularization based on the absolute values of coefficients with a nonlinear penalty function. For detailed advantages of using SCAD, see [13].

3 The proposed Methods

3.1 Basic formulation

In high-dimensional settings, we encounter two major challenges in estimating the CS. Typically, standardization is applied in SDR. However, when the number of predictors (p) exceeds the sample size (n), the covariance matrix becomes singular, and standardization is no longer feasible. Furthermore, we can obtain only $H - 1$ sufficient predictors with H slices, since the rank of the objective matrix depends on the number of slices for SIR. In particular, the CS cannot be well estimated when the response variable is binary and the true dimensionality is two or greater. To address these problems, we propose algorithms for graphical model-based SIR, SAVE, and DR incorporating the SCAD penalty, referred to as SCAD-SIR, SCAD-SAVE, and SCAD-DR, respectively. We further extend the ideas to Glasso based DR (GraphDR).

Firstly, we compute the sample covariance matrix, represented as $\hat{S}_{\mathbf{X}} = (1/n)\sum_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^{\mathsf{T}}$. The next step involves using the SCAD or Glasso to estimate the inverse covariance matrix, denoted by $\hat{\Theta}_{\mathbf{X}} = \hat{\Sigma}^{-1}$. The inverse covariance matrix of the SCAD version can be estimated by

$$\hat{\Theta}_{\mathbf{X}} = \arg\min_{\Theta_{\mathbf{X}}\succ 0}\Big\{\text{trace}(\hat{S}_{\mathbf{X}}\Theta_{\mathbf{X}}) - \log\det\Theta_{\mathbf{X}} + \sum_{a\neq b} p_\lambda(|\theta_{X;ab}|)\Big\}, \tag{2}$$

where $\theta_{X;ab}$ denotes the (a, b)-th element of $\Theta_{\mathbf{X}}$. The inverse covariance matrix of the Glasso version can be estimated by

$$\hat{\Theta}_{\mathbf{X}} = \arg\min_{\Theta_{\mathbf{X}}\succ 0}\Big\{\text{trace}(\hat{S}_{\mathbf{X}}\Theta_{\mathbf{X}}) - \log\det\Theta_{\mathbf{X}} + \lambda_{n1}\sum_{a\neq b}|\theta_{ab}|\Big\}. \tag{3}$$

Next, we determine the inverse square root of the estimated precision matrix $\hat{\Sigma}^{-1/2}$ and can standardize the random vectors in the high-dimensional situation. Secondly, we calculate the objective matrices of a variety of SDR methods, by computing the empirical objective matrix $\hat{S}_{M_l}$, and applying the SCAD or Glasso. The inverse covariance matrix of the SCAD version can be estimated by

$$\hat{\Omega}_{M_l} = \arg\min_{\Omega_{M_l}\succ 0}\Big\{\text{trace}(\hat{S}_{M_l}\Omega_{M_l}) - \log\det\Omega_{M_l} + \sum_{a\neq b} p_\lambda(|\omega_{M_l;ab}|)\Big\}, \tag{4}$$

where $\omega_{M_l;ab}$ is (a, b)-th element of Ω_{M_l}. The inverse covariance matrix of the Glasso version can be estimated by

$$\hat{\Omega}_{M_l} = \arg\min_{\Omega_{M_l}\succ 0}\Big\{\text{trace}(\hat{S}_{M_l}\Omega_{M_l}) - \log\det\Omega_{M_l} + \lambda_{n2}\sum_{a\neq b}|\omega_{M_l;ab}|\Big\}. \tag{5}$$

By estimating the objective matrix from $\hat{\Omega}^{-1}$, we can obtain the estimated covariance matrix $\hat{\Lambda} = \hat{\Omega}^{-1}$, which has full rank. Ultimately, we can approximate the CS, considering more directions than the number of slices H. Note that the detailed procedures are provided in the following section.

3.2 Sample level estimation

The SCAD-SIR, SCAD-SAVE, SCAD-DR, and GraphDR methods share a common structure for estimating the CS.

They combine dimension reduction methods with regularization either via the SCAD penalty or Glasso. However, they are differ in how they model objective matrix.

We first compute the sample mean $\bar{\mathbf{X}}$ and obtain the sample covariance matrix from observations, represented as $\hat{S}_{\mathbf{X}} = (1/n)\sum_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^{\mathsf{T}}$. Next, the SCAD penalty (2) is applied to estimate the inverse covariance matrix of SCAD-SIR, SCAD-SAVE, and SCAD-DR, which is denoted by $\hat{\Theta}_{\mathbf{X}} = \hat{\Sigma}^{-1}$. In our implementation, we adapt $a = 3.7$, as suggested by [13] and compute the inverse square root of the estimated precision matrix $\hat{\Sigma}^{-1/2}$. GraphDR used Glasso penalty (3) to obtain the precision matrix and $\hat{\Sigma}^{-1/2}$. The standardized random vectors $\hat{\mathbf{Z}}$ are subsequently obtained.

The next step involves partitioning the range of Y into h non-overlapping intervals. Consequently, $\tilde{Y}$ takes the value l if Y falls in the l-th interval, $l = 1, \ldots, h$. This procedure is unnecessary if the value of Y is categorical. After slicing the response variable Y, we compute the empirical objective matrix for SCAD-SIR, represented as $\hat{S}_{M_l} = (n_h/n)\sum_{l=1}^{h}\hat{M}_l\hat{M}_l^{\mathsf{T}}$ where $\hat{M}_l = E_n(\hat{\mathbf{Z}}|Y \in K_l)$. The candidate matrix of SCAD-SAVE can be constructed as $\hat{S}_{M_{2l}} = \widehat{\mathrm{cov}}(\mathbf{Z}|Y \in K_l)$. In SCAD-DR, the combinations of $\hat{S}_{M_l}$ and $\hat{S}_{M_{2l}}$ are used. The next procedure involves finding an inverse of these objective matrices. This can be accomplished by applying (4) to $\hat{S}_{M_l}$ for SCAD-SIR, $\hat{S}_{M_{2l}}$ for SCAD-SAVE, and both of them to SCAD-DR. Again, we use $a = 3.7$, as recommended by [13]. GraphDR is similar to SCAD-DR but we use Glasso in (5) instead of the SCAD penalty in (4). Our objective matrix is an inverse of these results. Similar to [34], we address a major limitation of classical SIR, where the number of non-zero eigenvalues is restricted to $H - 1$, making it problematic for a binary response variable. Our method ensures a full-rank matrix and allows more directions to be identified than the number of slices. Moreover, this algorithm guarantees the positive definiteness of the objective matrix, a property not always secured in classical SIR.

Finally, for all SCAD-SIR, SCAD-SAVE, SCAD-DR, and GraphDR, we obtain the eigenvalues and the eigenvectors for via eigendecomposition of these objective matrices . Letting the first r eigenvectors be $\hat{v}_r$, where $k = 1, \cdots, r$, the estimate of $\mathcal{S}_{Y|\mathbf{X}}$ is $\hat{\Sigma}^{-1/2}(\hat{v}_1, \cdots, \hat{v}_r)$.

To solve SCAD penalty, we utilize the coordinate descent algorithm, which is well-known for handling nonconvex penalties. This approach iteratively updates each element of the precision matrix while keeping the others fixed, solving the penalized optimization problem for that specific element. To address the piecewise nature of the SCAD penalty, a specialized proximal operator is used. This induces efficient updates and sparsity in the resulting precision matrix. The algorithm typically initializes the precision matrix using a model-based method, such as Glasso, to provide a stable starting point. The detailed procedures for constructing the CS on sample level data are in Algorithms 1, 2, 3, and 4.

Algorithm 1 SCAD-SIR algorithm

1. Let the sample mean as $\bar{\mathbf{X}}$ and the sample covariance as $\hat{S}_{\mathbf{X}} = (1/n)\sum_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^{\mathsf{T}}$. Compute the estimated value for $\hat{\Theta}_{\mathbf{X}}$ with the SCAD:

$$\hat{\Theta}_{\mathbf{X}} = \arg\min_{\Theta_{\mathbf{X}}}\{\mathrm{trace}(\hat{S}_{\mathbf{X}}\Theta_{\mathbf{X}}) - \log\det\Theta_{\mathbf{X}} + \sum_{a\neq b} p_\lambda(|\theta_{X;ab}|)\},$$

where $\theta_{X;ab}$ denotes the (a, b)-th element of $\Theta_{\mathbf{X}}$, $p_\lambda(\cdot)$ is from (1). We can obtain $\hat{\Sigma}_{\mathbf{X}}^{-1/2}$ using $\hat{\Theta}_{\mathbf{X}}$.

2. Compute the standardized random vectors as

$$\mathbf{Z} = \hat{\Sigma}^{-1/2}(\mathbf{X}_i - \bar{\mathbf{X}}), i = 1, \cdots, n.$$

3. Divide the range of the response variable into h slices and replace each Y with $\tilde{Y}_l$ for $Y \in K_l, l = 1, \cdots, h$. Let the number of observations of $\tilde{Y}_l$ in the slice h be n_l.

4. Approximate $E(\mathbf{Z}|Y \in K_l)$ by

$$\hat{M}_l = E_n(\mathbf{Z}|Y \in K_l) = \frac{E_n[\mathbf{Z}I(Y \in K_l)]}{E_n[I(Y \in K_l)]}, l = 1, \cdots, h.$$

5. Compute the empirical covariance matrix $\hat{S}_{M_l} = \sum_{l=1}^{h}(n_l/n)\hat{M}_l\hat{M}_l^{\mathsf{T}}$ and obtain the estimated value for $\hat{\Omega}_{M_l}$ with the SCAD:

$$\hat{\Omega}_{M_l} = \arg\min_{\Omega_{M_l}}\{\mathrm{trace}(\hat{S}_{M_l}\Omega_{M_l}) - \log\det\Omega_{M_l} + \sum_{a\neq b} p_\lambda(|\omega_{M_l;ab}|)\},$$

where $\omega_{M_l;ab}$ is (a, b)-th element of Ω_{M_l} and $p_\lambda(\cdot)$ is from (1).

6. Let $\hat{v}_1, \ldots, \hat{v}_r$ be the first r eigenvectors of the estimated covariance matrix $\hat{\Lambda}_{\mathrm{SIR}} \equiv \hat{\Omega}_{M_l}^{-1}$ and let $\hat{\beta}_k = \hat{\Sigma}^{-1/2}\hat{v}_k, k = 1, \ldots, r$. The sufficient predictors are $\hat{\beta}_k^{\mathsf{T}}(\mathbf{X}_1 - \bar{\mathbf{X}}), \ldots, \hat{\beta}_k^{\mathsf{T}}(\mathbf{X}_n - \bar{\mathbf{X}}), k = 1, \ldots, r$.

3.3 Theoretical results

Here, we present the detailed theoretical outcomes of SCAD-SIR, SCAD-SAVE, and SCAD-DR. Most of the results are similar to [34] and we modify the previous results since we use the SCAD penalty instead of the L_1 penalty. Denote the number of variables, the number of nonzero elements in the precision matrix, and the number of slices, all of which are allowed to depend on the sample size n, as p_n, s_n, and H_n, respectively.

3.3.1 Definitions

We first introduce notation and definitions that will be used to state our regularity conditions clearly.

(D.1) Let $\Theta_0 = \Sigma_0^{-1}$, where $\theta_{0;ab}$ denotes the (a, b)-th element of Θ_0. Define the set of nonzero elements as

$$S_1 = \{(a, b) : \theta_{0;ab} \neq 0\}.$$

Algorithm 2 SCAD-SAVE algorithm

Steps 1 and 2 are identical to those in Algorithm 1.

3. Divide the range of the response variable into h slices and replace each Y with $\tilde{Y}_l$ for $Y \in K_l, l = 1, \cdots, h$. Let the number of observations of $\tilde{Y}_l$ in the slice h be n_l.
4. Approximate $\widehat{\text{cov}}(\mathbf{Z}|Y \in K_l)$ by

$$\hat{S}_{M_{2l}} = \widehat{\text{cov}}(\mathbf{Z}|Y \in K_l) = \frac{E_n[\mathbf{Z}\mathbf{Z}^{\mathsf{T}} I(Y \in K_l)]}{E_n[I(Y \in K_l)]}, \ l = 1, \cdots, h,$$

and obtain the estimated value for $\hat{\Omega}_{M_{2l}}$ with the SCAD:

$$\hat{\Omega}_{M_{2l}} = \arg\min_{\Omega_{M_{2l}}} \Big\{ \text{trace}(\hat{S}_{M_{2l}} \Omega_{M_{2l}}) - \log\det \Omega_{M_{2l}} + \sum_{a \neq b} p_\lambda(|\omega_{M_{2l};ab}|) \Big\},$$

where $\omega_{M_{2l};ab}$ is (a, b)-th element of $\Omega_{M_{2l}}$ and $p_\lambda(\cdot)$ is from (1).
5. Compute the estimated covariance matrix:

$$\hat{\Lambda}_{\text{SAVE}} = \sum_{l=1}^{h} (n_l/n)[I_p - \hat{\Omega}_{M_{2l}}^{-1}][I_p - \hat{\Omega}_{M_{2l}}^{-1}]^{\mathsf{T}}.$$

6. Let $\hat{v}_1, \ldots, \hat{v}_r$ be the first r eigenvectors of the estimated covariance matrix $\hat{\Lambda}_{\text{SAVE}}$ and let $\hat{\beta}_k = \hat{\Sigma}^{-1/2}\hat{v}_k, k = 1, \ldots, r$. The sufficient predictors are $\hat{\beta}_k^{\mathsf{T}}(X_1 - \bar{\mathbf{X}}), \ldots, \hat{\beta}_k^{\mathsf{T}}(X_n - \bar{\mathbf{X}}), k = 1, \ldots, r$.

We define

$$a_{n1} = \max_{(a,b)\in S_1} p'_{\lambda_{n1}}(|\theta_{0;ab}|),$$
$$b_{n1} = \max_{(a,b)\in S_1} p''_{\lambda_{n1}}(|\theta_{0;ab}|),$$

where $p'_\lambda(\cdot)$ and $p''_\lambda(\cdot)$ denote the first and second derivatives of the penalty function $p_\lambda(\cdot)$, respectively. The number of nonzero off-diagonal elements in Θ_0 is denoted by s_{n1}.
(D.2) Let $\Omega_0 = \Lambda_0^{-1}$, where $\omega_{0;cd}$ denotes the (c, d)-th element of Ω_0. Define the set of nonzero elements as

$$S_2 = \{(c, d) : \omega_{0;cd} \neq 0\}.$$

We define

$$a_{n2} = \max_{(c,d)\in S_2} p'_{\lambda_{n2}}(|\omega_{0;cd}|),$$
$$b_{n2} = \max_{(c,d)\in S_2} p''_{\lambda_{n2}}(|\omega_{0;cd}|),$$

and denote by s_{n2} the number of nonzero off-diagonal elements in Ω_0.

3.3.2 Regularity Conditions

We now state the assumptions required for our theoretical results.

Algorithm 3 SCAD-DR algorithm

Steps 1 and 2 are identical to those in Algorithm 1.

3. Divide the range of the response variable into h slices and replace each Y with $\tilde{Y}_l$ for $Y \in K_l, l = 1, \cdots, h$. Let the number of observations of $\tilde{Y}_l$ in the slice h be n_l.
4. Approximate $E(\mathbf{Z}|Y \in K_l), \widehat{\text{cov}}(\mathbf{Z}|Y \in K_l)$ by

$$\hat{M}_{1l} = E_n(\mathbf{Z}|Y \in K_l) = \frac{E_n[\mathbf{Z}I|Y \in K_l]}{E_n[I(Y \in K_l)]}, \ l = 1, \cdots, h,$$
$$\hat{S}_{M_{2l}} = \widehat{\text{cov}}(\mathbf{Z}|Y \in K_l) = \frac{E_n[\mathbf{Z}\mathbf{Z}^{\mathsf{T}} I(Y \in K_l]}{E_n[I(Y \in K_l)]}, \ l = 1, \cdots, h.$$

5. Compute the empirical covariance matrix $\hat{S}_{M_{1l}} = (n_l/n)\sum_{l=1}^{h} \hat{M}_{1l}\hat{M}_{1l}^{\mathsf{T}}$ and the estimated values for $\hat{\Omega}_{M_{1l}}, \hat{\Omega}_{M_{2l}}$ with the SCAD:

$$\hat{\Omega}_{M_{1l}} = \arg\min_{\Omega_{M_{1l}}} \Big\{ \text{trace}(\hat{S}_{M_{1l}} \Omega_{M_{1l}}) - \log\det \Omega_{M_{1l}} + \sum_{a \neq b} p_\lambda(|\omega_{M_{1l};ab}|) \Big\},$$
$$\hat{\Omega}_{M_{2l}} = \arg\min_{\Omega_{M_{2l}}} \Big\{ \text{trace}(\hat{S}_{M_{2l}} \Omega_{M_{2l}}) - \log\det \Omega_{M_{2l}} + \sum_{a \neq b} p_\lambda(|\omega_{M_l;ab}|) \Big\},$$

where $\omega_{M_{1l};ab}$ is (a, b)-th element of $\Omega_{M_{1l}}$, $\omega_{M_{2l};ab}$ is (a, b)-th element of $\Omega_{M_{2l}}$ and $p_\lambda(\cdot)$ is from (1).
6. From matrices $\hat{\Lambda}_{M_{1l}} \equiv \hat{\Omega}_{M_{1l}}^{-1}, \hat{\Lambda}_{M_{2l}} \equiv \hat{\Omega}_{M_{2l}}^{-1}$, compute

$$\hat{\Gamma}_1 = \sum_{l=1}^{h} E_n[I(Y \in K_l)]\hat{\Lambda}_{M_{2l}}^2, \qquad \hat{\Gamma}_2 = \sum_{l=1}^{h} E_n[I(Y \in K_l)](\hat{\Lambda}_{M_{1l}}\hat{\Lambda}_{M_{1l}}^{\mathsf{T}})^2,$$
$$\hat{\Gamma}_3 = (\sum_{l=1}^{h} \hat{\Lambda}_{M_{1l}}^{\mathsf{T}}\hat{\Lambda}_{M_{1l}})(\sum_{l=1}^{h} \hat{\Lambda}_{M_{1l}}\hat{\Lambda}_{M_{1l}}^{\mathsf{T}}),$$

and obtain $\hat{\Lambda}_{\text{DR}} = 2\hat{\Gamma}_1 + 2\hat{\Gamma}_2 + 2\hat{\Gamma}_3 - 2I_p$.
7. Let $\hat{v}_1, \ldots, \hat{v}_r$ be the first r eigenvectors of $\hat{\Lambda}_{\text{DR}}$ and let $\hat{\beta}_k = \hat{\Sigma}^{-1/2}\hat{v}_k, k = 1, \ldots, r$. The sufficient predictors are $\hat{\beta}_k^{\mathsf{T}}(X_1 - \bar{\mathbf{X}}), \ldots, \hat{\beta}_k^{\mathsf{T}}(X_n - \bar{\mathbf{X}}), k = 1, \ldots, r$.

(C.1) The minimum and maximum eigenvalues of the true covariance matrices Σ_0 and Λ_0 satisfy:

$$0 < \kappa_{1,\min} \leq \phi_{\min}(\Sigma_0) \leq \phi_{\max}(\Sigma_0) \leq \kappa_{1,\max} < \infty;$$
$$0 < \kappa_{2,\min} \leq \phi_{\min}(\Lambda_0) \leq \phi_{\max}(\Lambda_0) \leq \kappa_{2,\max} < \infty,$$

where $\phi_{\min}(\cdot)$ and $\phi_{\max}(\cdot)$ denote the minimum and maximum eigenvalues, respectively, and $\Lambda_0 = \text{cov}(E[X \mid Y])$. Here, $\kappa_{1,\min}$, $\kappa_{1,\max}$, $\kappa_{2,\min}$, and $\kappa_{2,\max}$ are positive constants.
(C.2) For some positive constants C_1 and C_2, the following rate conditions hold:

$$a_{n1} = O\big((1 + p_n/(s_{n1} + 1))\sqrt{\log(p_n)/n}\big),$$
$$b_{n1} = o(1), \quad \min_{(a,b)\in S_1} \frac{|\theta_{0;ab}|}{\lambda_{n1}} > C_1,$$

Algorithm 4 GraphDR algorithm

1. Let the sample mean as $\bar{\mathbf{X}}$ and the sample covariance as $\hat{S}_{\mathbf{X}} = (1/n)\sum_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^{\mathsf{T}}$. Compute the estimated value for $\hat{\Theta}_{\mathbf{X}}$ with the Glasso:

$$\hat{\Theta}_{\mathbf{X}} = \arg\min_{\Theta_{\mathbf{X}}}\{\text{trace}(\hat{S}_{\mathbf{X}}\Theta_{\mathbf{X}}) - \log\det\Theta_{\mathbf{X}} + \lambda_{n1}\sum_{a\neq b}|\theta_{ab}|\},$$

and obtain $\hat{\Sigma}_{\mathbf{X}}^{-1/2}$ using $\hat{\Theta}_{\mathbf{X}}$.
2. Compute the standardized random vectors

$$\hat{\mathbf{Z}} = \Sigma^{-1/2}(\mathbf{X}_i - \bar{\mathbf{X}}), i = 1, \cdots, n.$$

3. Divide the range of the response variable into h slices and replace each Y with $\tilde{Y}_l$ for $Y \in K_l, l = 1, \cdots, h$. Let the number of observations of $\tilde{Y}_l$ in the slice h be n_l.
4. Compute $\hat{M}1l$ and $\hat{S}M_{2l}$ following the same procedure in Algorithm 3.
5. Compute the empirical covariance matrix $\hat{S}_{M_{1l}} = (n_l/n)\sum_{l=1}^{h}\hat{M}_{1l}\hat{M}_{1l}^{\mathsf{T}}$ and the estimated values for $\hat{\Omega}_{M_{1l}}, \hat{\Omega}_{M_{2l}}$ with the Glasso:

$$\hat{\Omega}_{M_{1l}} = \arg\min_{\Omega_{M_{1l}}}\{\text{trace}(\hat{S}_{M_{1l}}\Omega_{M_{1l}}) - \log\det\Omega_{M_{1l}} + \lambda_{n2}\sum_{a\neq b}|\omega_{M_{1l};ab}|\},$$

$$\hat{\Omega}_{M_{2l}} = \arg\min_{\Omega_{M_{2l}}}\{\text{trace}(\hat{S}_{M_{2l}}\Omega_{M_{2l}}) - \log\det\Omega_{M_{2l}} + \lambda_{n2}\sum_{a\neq b}|\omega_{M_{2l};ab}|\}.$$

h
6. From matrices $\hat{\Lambda}_{M_{1l}} \equiv \hat{\Omega}_{M_{1l}}^{-1}$, $\hat{\Lambda}_{M_{2l}} \equiv \hat{\Omega}_{M_{2l}}^{-1}$, compute

$$\hat{\Gamma}_1 = \sum_{l=1}^{h} E_n[I(Y \in K_l)]\hat{\Lambda}_{M_{2l}}^2,$$

$$\hat{\Gamma}_2 = \sum_{l=1}^{h} E_n[I(Y \in K_l)](\hat{\Lambda}_{M_{1l}}\hat{\Lambda}_{M_{1l}}^{\mathsf{T}})^2, \text{ and}$$

$$\hat{\Gamma}_3 = (\sum_{l=1}^{h}\hat{\Lambda}_{M_{1l}}^{\mathsf{T}}\hat{\Lambda}_{M_{1l}})(\sum_{l=1}^{h}\hat{\Lambda}_{M_{1l}}\hat{\Lambda}_{M_{1l}}^{\mathsf{T}}),$$

and obtain $\hat{\Lambda}_{\text{DR}} = 2\hat{\Gamma}_1 + 2\hat{\Gamma}_2 + 2\hat{\Gamma}_3 - 2I_p$.
7. Let $\hat{\nu}_1, \ldots, \hat{\nu}_r$ be the first r eigenvectors of $\hat{\Lambda}_{\text{DR}}$ and let $\hat{\beta}_k = \hat{\Sigma}^{-1/2}\hat{\nu}_k, k = 1, \ldots, r$. The sufficient predictors are $\hat{\beta}_k^{\mathsf{T}}(X_1 - \bar{\mathbf{X}}), \ldots, \hat{\beta}_k^{\mathsf{T}}(X_n - \bar{\mathbf{X}}), k = 1, \ldots, r$.

$$a_{n2} = O\big((1 + p_n/(s_{n2}+1))\sqrt{\log(p_n)/H_n}\big),$$
$$b_{n2} = o(1), \quad \min_{(c,d)\in S_2}\frac{|\omega_{0;cd}|}{\lambda_{n2}} > C_2.$$

(C.3) The penalty function $p_\lambda(\cdot)$ is singular at the origin and satisfies:

$$\lim_{t\to 0}\frac{p_\lambda(t)}{\lambda t} = K > 0.$$

(C.4) There exist constants U_1, U_2, V_1 and V_2 such that for all $\theta_1, \theta_2 > U_1\lambda_{n1}$,

$$|p''_{\lambda_{n1}}(\theta_1) - p''_{\lambda_{n1}}(\theta_2)| \le V_1|\theta_1 - \theta_2|,$$

and similarly, for all $\theta_1, \theta_2 > U_2\lambda_{n2}$,

$$|p''_{\lambda_{n2}}(\theta_1) - p''_{\lambda_{n2}}(\theta_2)| \le V_2|\theta_1 - \theta_2|.$$

(C.1) gives uniform bounds to eigenvalues of Σ_0 and Λ_0. Condition (C.2) is utilized to prove consistency, where large values of a_{n1}, b_{n1}, a_{n2}, and b_{n2} indicate that the variance associated with the likelihood is influenced by the bias introduced through the penalty. The third condition induces sparsity and the fourth condition is for a smoothing of the penalty function.

3.3.3 Convergence rate

The following propositions are key components of our method.

Proposition 1 *Under regularity conditions (C.1)-(C.4), if*

$$(p_n + s_{n1})log\,p_n/n = O(\lambda_{n1}^2),$$
$$(p_n + s_{n1})(log\,p_n)^k/n = O(1),$$
$$(p_n + s_{n2})log\,p_n/H_n = O(\lambda_{n2}^2),$$
$$(p_n + s_{n2})(log\,p_n)^k/H_n = O(1)$$

for some $k > 1$, there exist local minimizers $\hat{\Theta}_X$ and $\hat{\Omega}$ such that

$$\| \hat{\Sigma}^{-1} - \Sigma_0^{-1} \|_F^2 = O_p\Big(\frac{(p_n + s_{n1})log\ p_n}{n}\Big),$$
$$\| \hat{\Lambda}^{-1} - \Lambda_0^{-1} \|_F^2 = O_p\Big(\frac{(p_n + s_{n2})log\ p_n}{H_n}\Big),$$

where $\|\cdot\|_F$ is the Frobenious norm. For these minimization results, we recommend referring to the proof of Theorem 1 of [20].

Proposition 2 *Under regularity conditions (C.1)-(C.4), we have*

$$||\hat{\Theta}_X\hat{\Lambda} - \Theta_{X;0}\Lambda_0||_2 = O_p\Big(\sqrt{\log p_n\big(\frac{p_n + s_{n1}}{n} + \frac{p_n + s_{n2}}{H_n}\big)}\Big).$$

Proof

$$\begin{aligned}
&\|\hat{\Sigma}^{-1}\hat{\Lambda} - \Sigma_0^{-1}\Lambda_0\|_2 \\
&= \| \hat{\Sigma}^{-1}\hat{\Lambda} - \hat{\Sigma}^{-1}\Lambda_0 + \hat{\Sigma}^{-1}\Lambda_0 - \Sigma_0^{-1}\Lambda_0 \|_2 \\
&\le \| \hat{\Sigma}^{-1}\hat{\Lambda} - \hat{\Sigma}^{-1}\Lambda_0 \|_2 + \| \hat{\Sigma}^{-1}\Lambda_0 - \Sigma_0^{-1}\Lambda_0 \|_2 \\
&\le \| \hat{\Sigma}^{-1} \|_2 \| \hat{\Lambda} - \Lambda_0 \|_2 \\
&\quad + \| \hat{\Sigma}^{-1} - \Sigma_0^{-1} \|_2 \| \Lambda_0 \|_2 .
\end{aligned}$$

where $\|\cdot\|_2$ is the operator norm. It is trivial that $\|\hat{\Sigma}^{-1}\|_2 = O_p(1)$ and $\|\Lambda_0\|_2 = O(1)$. $\|\hat{\Lambda} - \Lambda_0\|_2 = O_p\left(\sqrt{\frac{(p_n+s_{n2})\log p_n}{H_n}}\right)$ is from [34]. By combining these results with Proposition 1, we can achieve the desired result. □

4 Numerical experiments

In this section, we demonstrate the competitiveness of our method using a variety of numerical experiments by comparing linear SDR methods (SIR, SAVE, and DR), the graph informed SDR estimation (GraphSIR and GraphSAVE) from [34] with our method (SCAD-SIR, SCAD-SAVE, SCAD-DR, and GraphDR). Additionally, we illustrate our approach with two cases: the first involves a continuous response variable, while the second addresses a binary response variable in a classification problem. Finally, we conduct a comparison of our proposed methods with BCov-SDR (Zhang and Chen 49), a recently developed and robust SDR method that leverages ball covariance to address challenges such as outliers, heavy-tailed distributions, and minimal model assumptions. By including BCov-SDR in our analysis, we aim to highlight the competitiveness of our methods against a state-of-the-art approach in the field. The code used to conduct these numerical analyses is available on GitHub at https://github.com/kyongwonkim/DR-SCAD.

We use the following measure to evaluate the distance between the true vector β and the estimated vector $\hat{\beta}$. This measure is computed by summing the squares of the differences between each corresponding element of the vectors and taking the square root of this sum. This approach provides a straightforward and intuitive representation of the distance between β and $\hat{\beta}$ as

$$d(\beta, \hat{\beta}) = \| \hat{\beta}(\hat{\beta}^{\mathsf{T}}\hat{\beta})^{-1}\hat{\beta}^{\mathsf{T}} - \beta(\beta^{\mathsf{T}}\beta)^{-1}\beta^{\mathsf{T}} \|_F,$$

where $\|\cdot\|_F$ denotes the Froebenious norm.

For the tuning parameters for SCAD-SIR, SCAD-SAVE, and SCAD-DR methods, the selection of λ and the initialization of the precision matrix follow a process to achieve both stability and accuracy. The initial value for the precision matrix is obtained using the Glasso, where the covariance matrix is estimated by minimizing the log-likelihood function, subject to a penalization term. The penalization strength for Glasso is determined through cross-validation, where a grid of λ values is tested to identify the value that minimizes the cross-validated error. This process ensures that the resulting precision matrix provides a sparse and stable estimate, which serves as the starting point for the SCAD-penalized optimization. The selection of the tuning parameter λ for the SCAD penalty, which controls the degree of sparsity in the estimated precision matrix, is carried out through cross-validation. A sequence of λ values is generated, starting from a maximum value, $\lambda_{\max}$, which shrinks all coefficients to zero, and decreasing incrementally. This range ensures that both highly sparse and less sparse models are explored. For each λ in the grid, the model's performance is evaluated using cross-validation, with the goal of selecting the λ that minimizes the cross-validated error. By combining a robust initialization using Glasso with a carefully tuned λ selection via cross-validation, the SCAD-SIR, SCAD-SAVE, and SCAD-DR methods achieve both reliable convergence and accurate, interpretable results.

4.1 Case 1: Continuous response variable

We compare the performance of the proposed method on generated data. We use the following five models:

$$\textbf{Model 1}: Y = X_1 + X_2 + X_3 + X_4 + \sigma\epsilon, \text{ where } r = 1$$
$$\textbf{Model 2}: Y = \frac{(2+X_1^2)}{4} + X_2 + \sigma\epsilon, \quad \text{where} r = 2$$
$$\textbf{Model 3}: Y = X_1 + 0.3 \cdot \frac{X_2}{X_3 - 0.7} + \sigma\epsilon, \quad \text{where} r = 3$$
$$\textbf{Model 4}: Y = X_1 + \frac{\sin(X_2)}{X_3} + \sigma\epsilon, \quad \text{where} r = 3$$
$$\textbf{Model 5}: Y = X_1 + X_2 + \log(X_3^2) + \sigma\epsilon, \quad \text{where } r = 2$$

where $\epsilon \sim N(0,1)$, $\sigma = 0.2$, the predictor X is generated from $N(0, I_p)$, and r is a structural dimension which refers to the number of linear combinations of the predictors that are sufficient to capture all the information about the response variable. We take $n \in \{50, 100\}$ and $p \in \{10, 50, 100\}$ to investigate the effect of the sample size and dimension in estimation. We set the true structural dimension r from 1, 2, or 3.

Results computed from 100 repetitions are summarized in Tables 1 to 5 for each model. Here, H is the number of slices, n is the sample size, and p is the predictors' dimension. Table 1 provides the average and standard deviation of the results distance for Model 1, where SIR outperforms other methods when $p = 10$. Furthermore, when p is set to 50 or 100, GraphSIR excels relative to competing methods for a sample size of $n = 50$, and GraphDR demonstrates superior performance at $n = 100$. Table 2 details the results for Model 2 and we can observe that SCAD-SIR showing optimal performance. For Model 3, as illustrated in Table 3, SCAD-SIR achieves the smallest distance between the true and estimated directions. Table 4 shows the performance in Model 4, where GraphSAVE slightly outperforms our proposed method when the sample size n is 50, and SCAD-SAVE remains competitive with all of the existing methods across

Table 1 Comparisons of the average and standard deviation of the distance between $\hat{\beta}$ and β across various methods for Model 1.

		Classical method			**Graph**			**SCAD**		
n=50	$(H=3)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p=10$	0.36 (0.11)	1.27 (0.26)	0.37 (0.13)	0.54 (0.19)	1.19 (0.18)	0.50 (0.19)	0.53 (0.17)	1.36 (0.11)	0.75 (0.26)
	$p=50$	-	-	-	0.78 (0.17)	1.41 (0.01)	0.79 (0.16)	0.84 (0.20)	1.41 (0.01)	1.26 (0.17)
	$p=100$	-	-	-	1.03 (0.17)	1.41 (0.00)	1.07 (0.16)	1.06 (0.17)	1.41 (0.00)	1.28 (0.11)
	$(H=5)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p=10$	0.26 (0.08)	1.40 (0.03)	0.33 (0.11)	0.43 (0.18)	1.38 (0.14)	0.42 (0.20)	0.45 (0.16)	1.41 (0.02)	0.75 (0.29)
	$p=50$	-	-	-	0.72 (0.19)	1.41 (0.01)	0.72 (0.16)	0.79 (0.22)	1.41 (0.00)	1.29 (0.16)
	$p=100$	-	-	-	0.94 (0.20)	1.41 (0.00)	1.07 (0.17)	0.98 (0.21)	1.41 (0.00)	1.35 (0.08)
	$(H=7)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p=10$	0.22 (0.07)	1.40 (0.02)	0.40 (0.18)	0.42 (0.18)	1.40 (0.09)	0.42 (0.17)	0.47 (0.17)	1.41 (0.01)	0.77 (0.28)
	$p=50$	-	-	-	0.70 (0.20)	1.41 (0.00)	0.76 (0.17)	0.79 (0.23)	1.41 (0.00)	1.35 (0.10)
	$p=100$	-	-	-	0.98 (0.24)	1.41 (0.00)	1.18 (0.17)	1.02 (0.23)	1.41 (0.00)	1.39 (0.05)
n=100	$(H=3)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p=10$	0.24 (0.06)	0.41 (0.30)	0.22 (0.06)	0.38 (0.16)	0.41 (0.34)	0.37 (0.17)	0.30 (0.10)	0.64 (0.42)	0.46 (0.17)
	$p=50$	0.66 (0.10)	1.41 (0.01)	0.71 (0.10)	0.49 (0.10)	1.41 (0.01)	0.46 (0.10)	0.41 (0.12)	1.41 (0.00)	0.86 (0.23)
	$p=100$	-	-	-	0.70 (0.16)	1.41 (0.00)	0.65 (0.14)	0.58 (0.16)	1.41 (0.00)	1.07 (0.22)
	$(H=5)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p=10$	0.17 (0.05)	0.77 (0.47)	0.18 (0.05)	0.32 (0.14)	0.83 (0.51)	0.30 (0.15)	0.26 (0.09)	1.23 (0.33)	0.40 (0.13)
	$p=50$	0.49 (0.08)	1.41 (0.01)	0.61 (0.10)	0.42 (0.07)	1.41 (0.00)	0.40 (0.08)	0.36 (0.10)	1.41 (0.00)	0.98 (0.25)
	$p=100$	-	-	-	0.59 (0.11)	1.41 (0.00)	0.56 (0.09)	0.49 (0.13)	1.41 (0.00)	1.19 (0.21)
	$(H=10)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p=10$	0.11 (0.03)	1.40 (0.02)	0.21 (0.05)	0.28 (0.13)	1.39 (0.11)	0.30 (0.14)	0.24 (0.09)	1.41 (0.01)	0.42 (0.13)
	$p=50$	0.35 (0.06)	1.41 (0.01)	0.66 (0.13)	0.40 (0.08)	1.41 (0.00)	0.39 (0.09)	0.37 (0.12)	1.41 (0.00)	1.11 (0.23)
	$p=100$	-	-	-	0.56 (0.11)	1.41 (0.00)	0.60 (0.10)	0.49 (0.13)	1.41 (0.00)	1.35 (0.09)

all settings. Finally, Table 5 reports metrics for Model 5, where SCAD-SIR consistently outperforms other methods. From these results, we can observe that the performance of SCAD-SIR is superior in more complex models.

We generate **Model 5′**, which shares the same structure as **Model 5**, but with the predictors $X \in \mathbb{R}^{40}$ generated from a multivariate normal distribution $N(0, \Theta^{-1})$, where Θ is a 40×40 precision matrix. The diagonal entries of Θ are 1, 1, 1, 1.333, 3.010, 3.203, 1.543, 1.270, 1.554, 3, 1, 1, 1.2, 1, 1, 1, 1, 3, 2, 1, 1, 1, 1, 1, 1, 1.4, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1.2, 2, 1, and the off-diagonal entries are specified as follows: $\theta_{3,5} = \theta_{5,3} = 1.418$, $\theta_{4,10} = \theta_{10,4} = 0.744$, $\theta_{5,9} = \theta_{9,5} = 0.519$, $\theta_{5,10} = \theta_{10,5} = 0.577$, $\theta_{13,17} = \theta_{17,13} = 0.287$, $\theta_{17,20} = \theta_{20,17} = 0.542$, $\theta_{14,15} = \theta_{15,14} = 0.998$, $\theta_{21,23} = \theta_{23,21} = 0.864$, $\theta_{28,30} = \theta_{30,28} = 0.143$, $\theta_{38,40} = \theta_{40,38} = 0.142$, and $\theta_{33,37} = \theta_{37,33} = 0.247$, where θ_{ij} represents the (i, j)-th element of Θ. As shown in Table 6, the SCAD based methods, including SCAD-SIR, SCAD-SAVE, and SCAD-DR, demonstrate strong competitiveness compared to classical methods such as GraphSIR, GraphSAVE, and GraphDR when the predictors exhibit a correlated structure.

4.2 Case 2: Binary response variable

In case 2, we follow the setup of the example introduced in [22] and [32], twist problem, to investigate the binary response classification performance. There are two classes, with one corresponding to each curve.

Model 6 :

Class 1 : $X_1 = 20\cos\theta + U_1 + 1, X_2 = 20\sin\theta + U_2$, where U_1, U_2 and θ are independent generated from $N(0, 1)$, $N(0, 1)$ and $N(\pi, (0.25\pi)^2)$, respectively; $X_3, \ldots, X_p$ are independent generated from $N(0, 1)$.

Class 2 : $X_1 = 20\cos\theta + U_1, X_2 = 20\sin\theta + U_2$, where U_1, U_2 and θ are independent generated from $N(0, 1)$, $N(0, 1)$ and $N(\pi, (0.25\pi)^2)$, respectively; X_3, $\ldots$, X_p are independent generated from $N(0, 1)$.

In our study, we generate samples of size 300 for each class and perform standardization on the sample data. The CS is spanned by $((1, 0, 0, \ldots, 0, 0), (0, 1, 0, \ldots, 0, 0))^{\mathsf{T}}$. Figure 1 presents the results of the classification performance when $p = 20$. The left panel of Figure 1 represents the true shape with different colors indicating the two classes. In addition, the second and third panels illustrate the two directions estimated by GraphSIR and SCAD-SIR, respectively. We can see that SIR estimates only one direction, even though the true dimension of the CS is two. On the other hand, both GraphSIR and SCAD-SIR can effectively estimate the CS with $p = 20$. Next, we increase the number of the variables, p, to investigate the effect of the sample size and dimension in estimation. Figure 2 shows the results of the performance when $p = 300$. The left panel of Figure 2 shows the true shape with different colors indicating the two classes. In this case, since the response is binary, SIR is limited to estimating only one direction. The second and third panels of Figure 2 for $p = 300$ display the two directions estimated by GraphSIR and SCAD-SIR. We can observe that both GraphSIR and SCAD-SIR estimate the true shape well. To be more specific, SCAD-SIR performs better than GraphSIR in estimating the CS when $p = 300$. Furthermore, we increase the number of the variables, p, from 300 to 1000. In this case, SIR cannot estimate the CS, since the number of variables (p) is larger than the number of observations (n). Figure 3 illustrates the results of the performance when $p = 1000$. The left panel of Figure 2 shows the true shape with different colors indicating the two classes. The second and third panels of Figure 3 for $p = 1000$ display the two directions estimated by GraphSIR and SCAD-SIR. As the number of variables increases to 300 and 1000, we can see that a performance discrepancy between the two methods becomes apparent. Notably, the accuracy of GraphSIR decreases with a growing number of features, while SCAD-SIR continues to provide accurate estimates. In summary, this experiment highlights the competitiveness of SCAD-SIR to handle the high-dimensional datasets with strong robustness.

4.3 Case 3: Comparisons with the recent method

BCov-SDR (Zhang and Chen 49) is a SDR method designed to enhance robustness and flexibility by leveraging ball covariance, a dependence measure that effectively captures both linear and nonlinear relationships between variables. Unlike traditional methods, BCov-SDR does not impose stringent conditions such as linearity, constant variance, or continuity on predictors and responses. This makes it highly adaptable, even in the presence of outliers or heavy-tailed distributions. The method defines the central subspace by maximizing ball covariance subject to constraints. It is capable of handling multivariate responses and discrete or categorical predictors, expanding its applicability across diverse datasets.

To compare our methods with BCov-SDR, we consider the following model.

Model 7 : $Y = \beta^{\mathsf{T}} X + \sigma\epsilon$, where $r = 1$

where $\epsilon \sim N(0, 1)$, $\sigma = 0.2$, the predictor X is generated from $N(0, I_p)$ and $\beta = (1, -1, 0.5, 1, 5, 0, 0, 0, \cdots, 0)$. We take $n \in \{50, 100\}$ and $p \in \{50, 100, 200\}$ to investigate the effect of the sample size and dimension in estimation. We set the true structural dimension r at 1.

Table 7 compares the performance of our proposed methods-SCAD-SIR, SCAD-SAVE, SCAD-DR, and GraphDR-

Table 2 Comparisons of the average and standard deviation of the distance between $\hat{\beta}$ and β across various methods for Model 2.

		Classical method			Graph			SCAD		
n=50	$(H=3)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p=10$	1.40 (0.09)	1.66 (0.19)	1.36 (0.14)	1.36 (0.26)	1.34 (0.22)	1.32 (0.15)	1.35 (0.29)	1.43 (0.08)	1.41 (0.12)
	$p=50$	-	-	-	1.43 (0.17)	1.97 (0.03)	1.48 (0.04)	1.43 (0.18)	1.97 (0.03)	1.69 (0.11)
	$p=100$	-	-	-	1.44 (0.17)	1.99 (0.01)	1.55 (0.05)	1.44 (0.15)	1.99 (0.01)	1.73 (0.09)
	$(H=5)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p=10$	1.37 (0.09)	1.79 (0.12)	1.34 (0.16)	1.27 (0.14)	1.67 (0.24)	1.30 (0.16)	1.24 (0.17)	1.81 (0.13)	1.39 (0.11)
	$p=50$	-	-	-	1.45 (0.10)	1.97 (0.04)	1.51 (0.05)	1.44 (0.11)	1.97 (0.03)	1.75 (0.10)
	$p=100$	-	-	-	1.47 (0.09)	1.98 (0.03)	1.59 (0.06)	1.45 (0.09)	1.99 (0.01)	1.83 (0.09)
	$(H=7)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p=10$	1.35 (0.12)	1.82 (0.11)	1.39 (0.11)	1.28 (0.12)	1.77 (0.16)	1.34 (0.15)	1.27 (0.12)	1.84 (0.10)	1.40 (0.12)
	$p=50$	-	-	-	1.44 (0.12)	1.98 (0.03)	1.53 (0.05)	1.43 (0.13)	1.98 (0.02)	1.83 (0.09)
	$p=100$	-	-	-	1.46 (0.12)	1.99 (0.02)	1.65 (0.07)	1.44 (0.14)	1.99 (0.01)	1.92 (0.07)
n=100	$(H=3)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p=10$	1.36 (0.11)	1.27 (0.22)	1.23 (0.23)	1.24 (0.15)	1.06 (0.34)	1.25 (0.20)	1.27 (0.09)	1.07 (0.09)	1.34 (0.12)
	$p=50$	1.59 (0.05)	1.97 (0.02)	1.63 (0.06)	1.43 (0.14)	1.95 (0.05)	1.41 (0.04)	1.42 (0.12)	1.96 (0.05)	1.52 (0.05)
	$p=100$	-	-	-	1.46 (0.01)	1.98 (0.02)	1.44 (0.01)	1.44 (0.01)	1.98 (0.02)	1.55 (0.04)
	$(H=5)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p=10$	1.37 (0.09)	1.48 (0.19)	1.25 (0.21)	1.34 (0.30)	1.14 (0.31)	1.28 (0.21)	1.35 (0.20)	1.30 (0.10)	1.33 (0.13)
	$p=50$	1.54 (0.04)	1.97 (0.02)	1.58 (0.05)	1.42 (0.14)	1.96 (0.04)	1.42 (0.04)	1.42 (0.12)	1.97 (0.03)	1.55 (0.05)
	$p=100$	-	-	-	1.45 (0.10)	1.98 (0.03)	1.45 (0.02)	1.43 (0.11)	1.98 (0.01)	1.61 (0.05)
	$(H=10)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p=10$	1.34 (0.11)	1.76 (0.15)	1.26 (0.20)	1.26 (0.20)	1.63 (0.26)	1.28 (0.21)	1.27 (0.15)	1.78 (0.14)	1.32 (0.13)
	$p=50$	1.53 (0.04)	1.97 (0.03)	1.61 (0.06)	1.41 (0.17)	1.97 (0.03)	1.43 (0.04)	1.39 (0.22)	1.97 (0.02)	1.63 (0.07)
	$p=100$	-	-	-	1.46 (0.01)	1.98 (0.03)	1.47 (0.02)	1.44 (0.01)	1.99 (0.01)	1.81 (0.08)

Table 3 Comparisons of the average and standard deviation of the distance between $\hat{\beta}$ and β across various methods for Model 3.

		Classical method			Graph			SCAD		
n=50	$(H=3)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p=10$	1.68 (0.19)	1.98 (0.17)	1.80 (0.16)	1.58 (0.16)	1.84 (0.19)	1.76 (0.15)	1.65 (0.17)	1.85 (0.10)	1.85 (0.12)
	$p=50$	-	-	-	2.00 (0.13)	2.39 (0.03)	2.06 (0.07)	1.99 (0.13)	2.39 (0.04)	2.24 (0.09)
	$p=100$	-	-	-	1.98 (0.16)	2.42 (0.03)	2.05 (0.08)	1.98 (0.15)	2.42 (0.02)	2.27 (0.07)
	$(H=5)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p=10$	1.80 (0.15)	2.08 (0.14)	1.81 (0.17)	1.68 (0.09)	1.99 (0.19)	1.77 (0.16)	1.69 (0.19)	2.08 (0.16)	1.85 (0.10)
	$p=50$	-	-	-	1.99 (0.19)	2.39 (0.04)	2.09 (0.07)	2.00 (0.14)	2.39 (0.03)	2.27 (0.08)
	$p=100$	-	-	-	2.01 (0.14)	2.42 (0.03)	2.21 (0.07)	1.99 (0.16)	2.42 (0.02)	2.34 (0.05)
	$(H=7)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p=10$	1.78 (0.15)	2.10 (0.14)	1.84 (0.14)	1.64 (0.06)	2.05 (0.18)	1.77 (0.16)	1.67 (0.11)	2.10 (0.14)	1.84 (0.12)
	$p=50$	-	-	-	2.00 (0.12)	2.39 (0.04)	2.11 (0.07)	2.01 (0.12)	2.39 (0.03)	2.31 (0.07)
	$p=100$	-	-	-	2.03 (0.12)	2.42 (0.03)	2.26 (0.08)	2.02 (0.13)	2.42 (0.02)	2.30 (0.03)
n=100	$(H=3)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p=10$	1.63 (0.18)	1.81 (0.17)	1.75 (0.14)	1.49 (0.18)	1.71 (0.20)	1.70 (0.17)	1.52 (0.14)	1.69 (0.21)	1.80 (0.10)
	$p=50$	1.95 (0.09)	2.38 (0.03)	2.18 (0.06)	1.88 (0.26)	2.38 (0.04)	1.97 (0.04)	1.91 (0.21)	2.39 (0.04)	2.10 (0.07)
	$p=100$	-	-	-	1.97 (0.18)	2.42 (0.02)	2.01 (0.03)	1.98 (0.15)	2.42 (0.02)	2.17 (0.08)
	$(H=5)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p=10$	1.76 (0.16)	1.87 (0.18)	1.74 (0.14)	1.51 (0.15)	1.74 (0.20)	1.70 (0.18)	1.57 (0.09)	1.79 (0.18)	1.79 (0.11)
	$p=50$	2.14 (0.05)	2.38 (0.03)	2.17 (0.06)	1.91 (0.22)	2.38 (0.03)	1.97 (0.05)	1.93 (0.20)	2.39 (0.03)	2.16 (0.08)
	$p=100$	-	-	-	1.97 (0.18)	2.42 (0.01)	2.03 (0.04)	1.95 (0.18)	2.42 (0.01)	2.27 (0.08)
	$(H=10)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p=10$	1.74 (0.15)	2.06 (0.15)	1.75 (0.15)	1.57 (0.07)	1.93 (0.20)	1.70 (0.18)	1.63 (0.08)	2.05 (0.12)	1.80 (0.11)
	$p=50$	2.16 (0.05)	2.38 (0.03)	2.2 (0.06)	1.95 (0.18)	2.39 (0.03)	2.01 (0.04)	1.94 (0.19)	2.39 (0.03)	2.24 (0.07)
	$p=100$	-	-	-	1.94 (0.20)	2.42 (0.02)	2.08 (0.04)	1.93 (0.21)	2.42 (0.02)	2.36 (0.04)

Table 4 Comparisons of the average and standard deviation of the distance between $\hat{\beta}$ and β across various methods for Model 4.

		Classical method			**Graph**			**SCAD**		
n=50	$(H=3)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p=10$	1.72 (0.20)	1.93 (0.17)	1.83 (0.20)	1.70 (0.16)	1.85 (0.23)	1.71 (0.25)	1.70 (0.20)	1.87 (0.24)	1.91 (0.14)
	$p=50$	-	-	-	2.01 (0.23)	2.37 (0.05)	2.18 (0.13)	2.01 (0.23)	2.37 (0.05)	2.33 (0.05)
	$p=100$	-	-	-	2.10 (0.19)	2.41 (0.03)	2.22 (0.15)	2.08 (0.21)	2.41 (0.03)	2.37 (0.05)
	$(H=5)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p=10$	1.88 (0.14)	1.97 (0.15)	1.87 (0.18)	1.74 (0.17)	1.94 (0.18)	1.78 (0.17)	1.73 (0.07)	1.98 (0.15)	1.91 (0.13)
	$p=50$	-	-	-	2.08 (0.21)	2.37 (0.05)	2.25 (0.09)	2.06 (0.22)	2.37 (0.05)	2.35 (0.04)
	$p=100$	-	-	-	2.15 (0.23)	2.41 (0.04)	2.35 (0.06)	2.14 (0.24)	2.41 (0.02)	2.40 (0.03)
	$(H=7)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p=10$	1.89 (0.15)	2.02 (0.15)	1.89 (0.19)	1.75 (0.13)	1.98 (0.17)	1.79 (0.19)	1.73 (0.13)	2.03 (0.14)	1.93 (0.14)
	$p=50$	-	-	-	2.13 (0.21)	2.37 (0.06)	2.26 (0.08)	2.07 (0.23)	2.38 (0.04)	2.36 (0.04)
	$p=100$	-	-	-	2.20 (0.18)	2.41 (0.03)	2.37 (0.05)	2.15 (0.22)	2.41 (0.02)	2.40 (0.02)
n=100	$(H=3)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p=10$	1.70 (0.21)	1.78 (0.19)	1.62 (0.21)	1.78 (0.17)	1.65 (0.25)	1.64 (0.23)	1.81 (0.08)	1.59 (0.28)	1.82 (0.15)
	$p=50$	2.05 (0.09)	2.37 (0.04)	2.30 (0.06)	1.98 (0.14)	2.35 (0.05)	1.99 (0.14)	1.96 (0.16)	2.35 (0.06)	2.31 (0.06)
	$p=100$	-	-	-	2.01 (0.10)	2.41 (0.02)	2.03 (0.10)	2.00 (0.10)	2.41 (0.03)	2.39 (0.03)
	$(H=5)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p=10$	1.83 (0.13)	1.81 (0.18)	1.64 (0.19)	1.78 (0.18)	1.71 (0.23)	1.64 (0.20)	1.77 (0.09)	1.77 (0.19)	1.81 (0.12)
	$p=50$	2.28 (0.06)	2.37 (0.03)	2.31 (0.05)	1.98 (0.16)	2.36 (0.05)	2.02 (0.15)	1.98 (0.15)	2.37 (0.04)	2.34 (0.05)
	$p=100$	-	-	-	2.03 (0.10)	2.41 (0.03)	2.09 (0.13)	2.00 (0.10)	2.41 (0.02)	2.40 (0.02)
	$(H=10)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p=10$	1.84 (0.16)	1.90 (0.17)	1.74 (0.19)	1.76 (0.10)	1.85 (0.18)	1.69 (0.24)	1.77 (0.19)	1.92 (0.14)	1.83 (0.13)
	$p=50$	2.30 (0.05)	2.38 (0.03)	2.31 (0.06)	1.99 (0.15)	2.36 (0.05)	2.11 (0.09)	1.99 (0.17)	2.38 (0.03)	2.35 (0.04)
	$p=100$	-	-	-	2.03 (0.19)	2.41 (0.02)	2.25 (0.12)	2.00 (0.21)	2.42 (0.01)	2.41 (0.02)

Table 5 Comparisons of the average and standard deviation of the distance between $\hat{\beta}$ and β across various methods for Model 5.

		Classical method			Graph			SCAD		
n=50	$(H=3)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p=10$	1.55 (0.11)	1.58 (0.17)	1.47 (0.24)	1.52 (0.18)	1.48 (0.19)	1.43 (0.26)	1.51 (0.18)	1.53 (0.17)	1.62 (0.16)
	$p=50$	-	-	-	1.74 (0.19)	1.92 (0.08)	1.82 (0.12)	1.72 (0.20)	1.94 (0.05)	1.93 (0.06)
	$p=100$	-	-	-	1.79 (0.16)	1.97 (0.05)	1.86 (0.11)	1.79 (0.18)	1.98 (0.02)	1.95 (0.05)
	$(H=5)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p=10$	1.52 (0.11)	1.66 (0.15)	1.52 (0.22)	1.48 (0.27)	1.60 (0.17)	1.41 (0.28)	1.46 (0.15)	1.69 (0.15)	1.62 (0.17)
	$p=50$	-	-	-	1.72 (0.21)	1.91 (0.07)	1.82 (0.10)	1.70 (0.23)	1.95 (0.05)	1.93 (0.04)
	$p=100$	-	-	-	1.81 (0.19)	1.96 (0.05)	1.92 (0.07)	1.81 (0.17)	1.97 (0.02)	1.97 (0.02)
	$(H=7)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p=10$	1.54 (0.13)	1.71 (0.14)	1.59 (0.19)	1.50 (0.22)	1.64 (0.16)	1.49 (0.21)	1.49 (0.14)	1.73 (0.13)	1.67 (0.16)
	$p=50$	-	-	-	1.77 (0.19)	1.94 (0.06)	1.85 (0.10)	1.74 (0.22)	1.96 (0.03)	1.95 (0.03)
	$p=100$	-	-	-	1.84 (0.20)	1.98 (0.02)	1.95 (0.05)	1.83 (0.23)	1.98 (0.01)	1.98 (0.02)
n=100	$(H=3)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p=10$	1.44 (0.10)	1.39 (0.13)	1.08 (0.26)	1.42 (0.09)	1.26 (0.23)	1.12 (0.34)	1.39 (0.21)	1.29 (0.24)	1.39 (0.21)
	$p=50$	1.82 (0.06)	1.95 (0.04)	1.89 (0.07)	1.51 (0.13)	1.80 (0.12)	1.62 (0.13)	1.50 (0.13)	1.84 (0.09)	1.89 (0.09)
	$p=100$	-	-	-	1.56 (0.15)	1.96 (0.04)	1.64 (0.14)	1.51 (0.14)	1.96 (0.04)	1.95 (0.05)
	$(H=5)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p=10$	1.44 (0.11)	1.48 (0.11)	1.14 (0.24)	1.41 (0.16)	1.40 (0.11)	1.11 (0.31)	1.31 (0.12)	1.47 (0.10)	1.43 (0.18)
	$p=50$	1.81 (0.06)	1.95 (0.04)	1.85 (0.07)	1.51 (0.17)	1.90 (0.07)	1.62 (0.13)	1.49 (0.13)	1.93 (0.05)	1.92 (0.05)
	$p=100$	-	-	-	1.59 (0.15)	1.95 (0.05)	1.70 (0.09)	1.53 (0.20)	1.97 (0.03)	1.96 (0.03)
	$(H=10)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p=10$	1.43 (0.14)	1.58 (0.13)	1.32 (0.23)	1.39 (0.26)	1.53 (0.13)	1.24 (0.27)	1.32 (0.12)	1.62 (0.13)	1.48 (0.18)
	$p=50$	1.85 (0.07)	1.96 (0.03)	1.88 (0.06)	1.48 (0.23)	1.91 (0.08)	1.67 (0.13)	1.46 (0.26)	1.95 (0.04)	1.93 (0.05)
	$p=100$	-	-	-	1.57 (0.22)	1.96 (0.04)	1.79 (0.11)	1.51 (0.27)	1.98 (0.02)	1.97 (0.02)

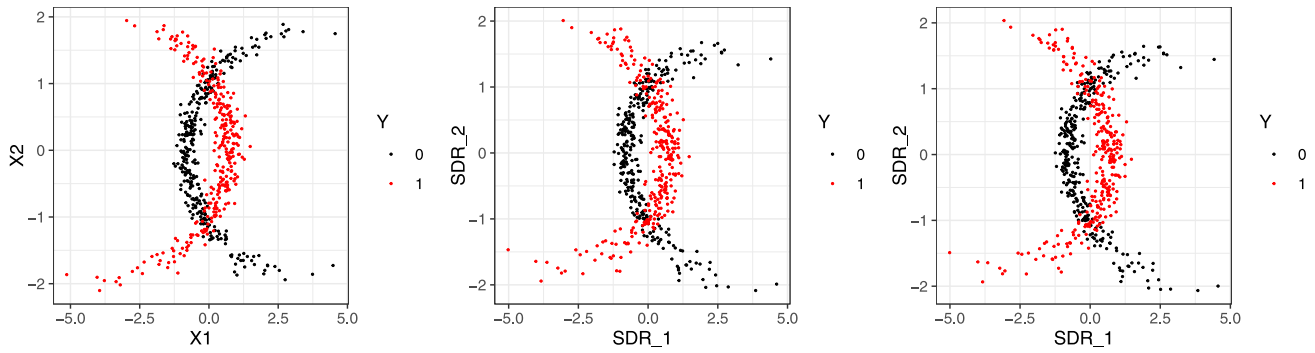

Fig. 1 Visualization of results for Model 6 ($p = 20, n = 600$). The left panel illustrates the true shape, the middle panel shows the two directions obtained from GraphSIR, and the right panel displays the two directions estimated from SCAD-SIR

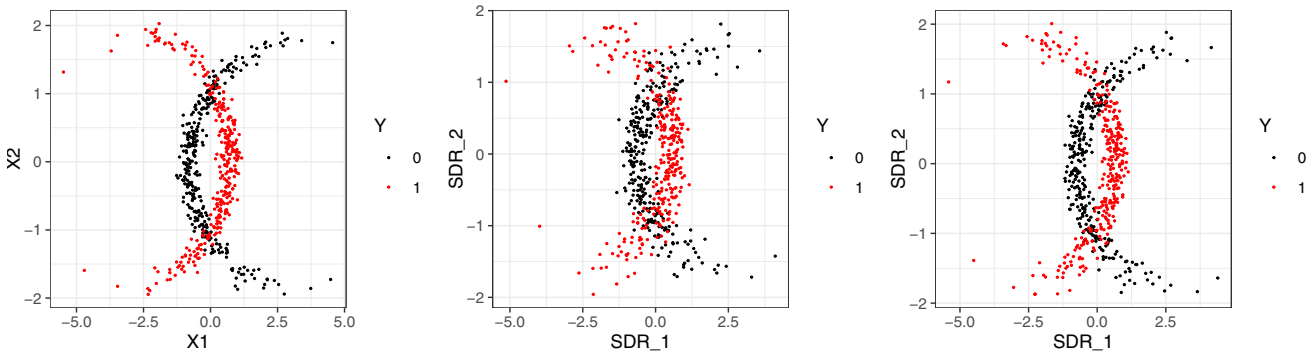

Fig. 2 Visualization of results for Model 6 ($p = 300, n = 600$). The left panel illustrates the true shape, the middle panel shows the two directions obtained from GraphSIR, and the right panel displays the two directions estimated from SCAD-SIR

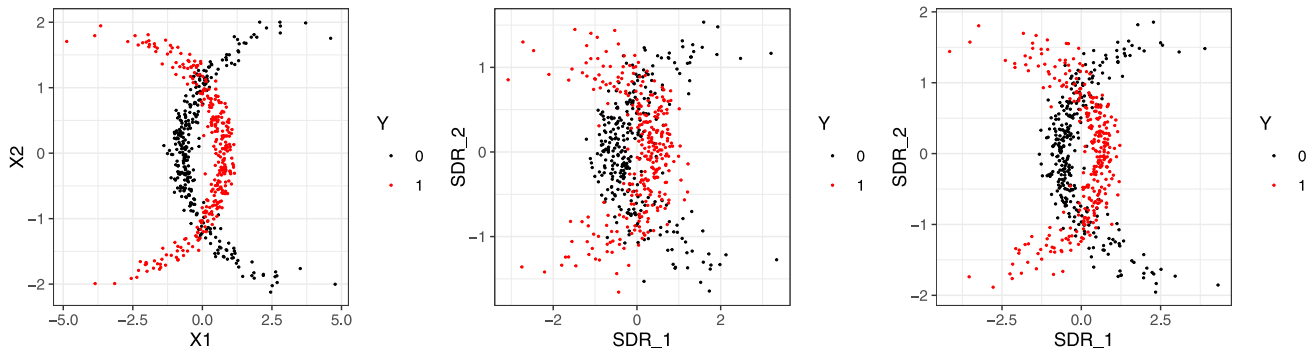

Fig. 3 Visualization of results for Model 6 ($p = 1000, n = 600$). The left panel illustrates the true shape, the middle panel shows the two directions obtained from GraphSIR, and the right panel displays the two directions estimated from SCAD-SIR

with BCov-SDR, GraphSIR, and GraphSAVE, using the distance between β and $\hat{\beta}$ as the performance metric with 100 repetitions. The results clearly demonstrate that SCAD-SIR, SCAD-DR, and GraphDR outperform BCov-SDR, achieving smaller distances and thus more accurate estimates of the SDR directions. In contrast, the SCAD-SAVE method performs comparably to BCov-SDR, with distances that are either on par with or slightly larger than those of BCov-SDR.

5 Real Data Analysis

We now apply SCAD-SIR, SCAD-SAVE, SCAD-DR, and GraphDR to the eye gene dataset (Scheetz et al. 37), breast cancer dataset (Augugliaro et al. 1), and Duke breast cancer dataset (Wang et al. 42) to compare the performance of our method with GraphSIR and GraphSAVE.

First, we consider the eye gene dataset (Scheetz et al. 37). This is gene expression data collected from the eyes of 120 rats, focusing on understanding the genetic regulation of eye function, with a particular interest in genes related to eye

Table 6 Comparisons of the average and standard deviation of the distance between $\hat{\beta}$ and β across various methods for Model 5′ ($p = 40$).

	Classical method			**Graph**			**SCAD**		
(n=50)	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
$H = 3$	1.91 (0.07)	1.95 (0.04)	1.91 (0.07)	1.73 (0.18)	1.59 (0.21)	1.72 (0.15)	1.67 (0.18)	1.91 (0.09)	1.87 (0.10)
$H = 5$	1.92 (0.05)	1.93 (0.05)	1.91 (0.05)	1.82 (0.16)	1.58 (0.22)	1.82 (0.15)	1.66 (0.21)	1.94 (0.05)	1.90 (0.06)
$H = 7$	1.93 (0.06)	1.95 (0.05)	1.94 (0.05)	1.86 (0.17)	1.65 (0.22)	1.86 (0.14)	1.72 (0.21)	1.94 (0.04)	1.91 (0.06)
(n=100)	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
$H = 3$	1.83 (0.08)	1.91 (0.08)	1.88 (0.07)	1.57 (0.18)	1.61 (0.22)	1.58 (0.15)	1.46 (0.14)	1.81 (0.14)	1.82 (0.10)
$H = 5$	1.84 (0.09)	1.92 (0.07)	1.88 (0.08)	1.57 (0.17)	1.49 (0.10)	1.58 (0.14)	1.47 (0.10)	1.87 (0.10)	1.86 (0.10)
$H = 10$	1.83 (0.08)	1.94 (0.05)	1.87 (0.09)	1.74 (0.18)	1.54 (0.17)	1.73 (0.18)	1.50 (0.12)	1.90 (0.06)	1.85 (0.08)

Table 7 Comparisons of the average and standard deviation of the distance between $\hat{\beta}$ and β across various methods for Model 7.

		Modern method		**Graph**			**SCAD**	
n=50	($p = 50$)	BCov-SDR	SIR	SAVE	DR	SIR	SAVE	DR
	$H = 3$	1.40 (0.04)	0.59 (0.06)	1.46 (0.27)	0.57 (0.08)	0.54 (0.08)	1.65 (0.18)	1.02 (0.20)
	$H = 5$		0.57 (0.06)	1.65 (0.20)	0.59 (0.07)	0.51 (0.08)	1.57 (0.15)	1.13 (0.18)
	$H = 7$		0.54 (0.07)	1.57 (0.10)	0.62 (0.08)	0.50 (0.07)	1.47 (0.09)	1.27 (0.15)
	($p = 100$)	BCov-SDR	SIR	SAVE	DR	SIR	SAVE	DR
	$H = 3$	1.41 (0.01)	0.63 (0.04)	1.54 (0.08)	0.71 (0.09)	0.58 (0.06)	1.69 (0.16)	1.10 (0.15)
	$H = 5$		0.63 (0.05)	1.66 (0.15)	0.79 (0.11)	0.56 (0.06)	1.60 (0.08)	1.26 (0.12)
	$H = 7$		0.62 (0.06)	1.57 (0.14)	0.90 (0.14)	0.55 (0.06)	1.47 (0.01)	1.37 (0.07)
	($p = 200$)	BCov-SDR	SIR	SAVE	DR	SIR	SAVE	DR
	$H = 3$	1.41 (0.04)	0.63 (0.03)	0.69 (0.20)	0.93 (0.23)	0.58 (0.04)	1.50 (0.02)	1.2 (0.11)
	$H = 5$		0.62 (0.04)	1.52 (0.08)	1.14 (0.23)	0.57 (0.05)	1.64 (0.48)	1.31 (0.08)
	$H = 7$		0.62 (0.03)	1.64 (0.04)	1.33 (0.15)	0.57 (0.04)	1.60 (0.10)	1.38 (0.05)
n=100	($p = 50$)	BCov-SDR	SIR	SAVE	DR	SIR	SAVE	DR
	$H = 3$	1.27 (0.18)	0.56 (0.05)	1.54 (0.04)	0.45 (0.03)	0.42 (0.07)	1.49 (0.15)	0.64 (0.09)
	$H = 5$		0.55 (0.05)	1.51 (0.09)	0.43 (0.04)	0.40 (0.07)	1.49 (0.06)	0.72 (0.11)
	$H = 10$		0.49 (0.05)	1.69 (0.09)	0.39 (0.04)	0.39 (0.07)	1.60 (0.14)	0.90 (0.16)
	($p = 100$)	BCov-SDR	SIR	SAVE	DR	SIR	SAVE	DR
	$H = 3$	1.35 (0.11)	0.60 (0.04)	1.41 (0.15)	0.50 (0.03)	0.48 (0.06)	1.54 (0.17)	0.72 (0.09)
	$H = 5$		0.59 (0.04)	1.52 (0.18)	0.47 (0.03)	0.45 (0.06)	1.55 (0.09)	0.84 (0.10)
	$H = 10$		0.56 (0.05)	1.56 (0.11)	0.49 (0.05)	0.44 (0.06)	1.45 (0.04)	1.20 (0.15)
	($p = 200$)	BCov-SDR	SIR	SAVE	DR	SIR	SAVE	DR
	$H = 3$	1.39 (0.04)	0.63 (0.03)	1.64 (0.09)	0.55 (0.03)	0.52 (0.05)	1.55 (0.17)	0.75 (0.09)
	$H = 5$		0.63 (0.03)	1.65 (0.03)	0.57 (0.03)	0.50 (0.05)	1.56 (0.08)	0.96 (0.11)
	$H = 10$		0.62 (0.04)	1.55 (0.19)	0.69 (0.05)	0.50 (0.05)	1.45 (0.04)	1.35 (0.09)

diseases. The dataset includes expression levels for 18,976 genes. This dataset is relevant for studying genetic factors related to eye function and disease. Here, we use the expression level of TRIM32, a gene involved in retinal development, as the response variable, and select 200 other genes as predictor variables, resulting in $n = 119$ and $p = 200$. The competitiveness of our method can be demonstrated by this dataset because of the large number of predictors relative to the sample size.

First, we screened out one outlier in the eye dataset. We chose the number of slices to ensure that each slice contains a sufficient amount of data and fixed the number of slices H to 3 and 10. We also change the estimated structural dimension $\hat{r}$ at six different levels $\hat{r} = 1, 2, 3, 4, 5, 6$. For each value of $\hat{r}$, the dataset is initially projected onto a $\hat{r}$-dimensional subspace, followed by fitting a simple linear regression model to this reduced dataset. We evaluate the performance of each method by using two metrics, the mean squared error (MSE) and the leave-one-out cross-validation (LOO-CV). MSE is

a widely used measure of model accuracy that evaluates the average squared difference between the predicted values and the actual values. MSE is calculated as

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2,$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, and n is the number of observations. The smaller the MSE, the better the model fits the data, as it indicates smaller deviations between the predicted and actual values. LOO-CV is a model validation method used to evaluate a model's predictive performance. In LOO-CV, one observation from the dataset is removed, the model is trained on the remaining data. The model's prediction is tested on the left out observation. This process is repeated for each observation in the dataset. The prediction error is averaged to provide an estimate of the model's performance. The LOO-CV error can be computed as

$$\text{LOO-CV} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_{(-i)})^2,$$

where y_i is the actual value and $\hat{y}_{(-i)}$ is the predicted value from the model trained without the i-th observation.

The result, summarized in Tables 8 and 9, shows the MSE and LOO-CV for various structural dimensions and the number of slices ($H = 3, 10$). It is evident from Tables 8 and 9 that SCAD-DR and GraphDR consistently outperform both GraphSIR and GraphSAVE across the different structural dimensions. We can see that the SCAD-DR provides the best performance when the structural dimension is set at $\hat{r} = 5$, highlighting the effectiveness of our approach. Furthermore, Tables 8 and 9 also indicate that the number of slices is not highly sensitive.

The strong performance of our SCAD-SIR, SCAD-SAVE, and SCAD-DR methods on this dataset, as evidenced by both low MSE and favorable LOO-CV results, indicates that these approaches are highly effective at identifying the key predictors of TRIM32 gene expression. This suggests that our dimension reduction methods, combined with the SCAD penalty, successfully capture the underlying relationships, while maintaining model simplicity and reducing overfitting.

The breast cancer dataset by [1] contains gene expression data collected from breast cancer patients. This study aimed at identifying genetic markers associated with the disease. This high-dimensional dataset includes a large number of gene expression profiles as predictor variables ($p = 287$) and clinical outcome (binary) as a response variable. This dataset contains 52 participants which make this dataset as high-dimension and low sample size dataset. This is typically used to study the relationship between gene expression profiles and cancer outcomes.

For the breast cancer dataset, we generate receiver operating characteristic (ROC) curves for the six models in Figure 4. These ROC curves are useful in evaluating performance of the classification model by examining the area under the curve (AUC), which reflects the trade-off between sensitivity (the true positive rate, calculated as $\frac{\text{TP}}{\text{TP+FN}}$) and $1 -$ specificity (the false positive rate, computed as $\frac{\text{FP}}{\text{TN+FP}}$). There are several methods for determining the structural dimension in SDR. One common approach is the sequential test, as comprehensively reviewed in [24], where hypotheses are tested sequentially to identify the number of significant dimensions. Another approach involves the ladle estimator (Luo and Li 26). However, they are quite challenging to find a structural dimension in $p > n$ situation. We recommend cross-validation combined with penalized SDR method where different structural dimensions are tested and the one with the lowest prediction error on a validation set is chosen. Here, we choose the structural dimension based on the cross-validation method. As shown in Figure 4, we can observe that SCAD-DR and GraphDR outperform the other methods. Especially, SCAD-DR achieves a higher classification accuracy with an AUC of 0.965, surpassing all other models. In addition, the SCAD penalty-based methods perform competitively in accurately identifying true positives while simultaneously maintaining a low false positive rate. Consequently, Figure 4 indicates the effectiveness of our methods in breast cancer prediction. The strong performance of our SCAD-SIR and SCAD-DR methods on the breast cancer dataset suggests that these approaches are highly effective at identifying relevant gene expressions for distinguishing between different outcomes. This result implies that our methods work well in capturing key features from high-dimensional genomic data and are well-suited for improving the accuracy of classification in complex biological datasets.

Finally, we consider the Duke breast cancer dataset (Wang et al. 42). This dataset comprises 46 patient samples with 7129 gene expression measurements. The primary objective of this study is to identify a small subset of genes that can be used as prognostic or predictive markers for breast cancer. The samples are categorized into two groups estrogen receptor-positive (ER+) and estrogen receptor-negative (ER-).

To refine the analysis, we first select 917 genes correlated with the COL2A1, MMP-7, CD24, PGK1, ESR1, and NAT1 expressions that have been found to be associated with breast cancer (Fogel et al. 11; Duan et al. 9; Gold et al. 15; Zhang et al. 53; Holst et al. 17; Bertram and Hass 3; Wakefield et al. 44; Bucan et al. 4). This results in a dataset with $n = 46$ and $p = 917$. Subsequently, we apply the ladle estimator (Luo

Table 8 MSE values, with the numbers in brackets representing the corresponding LOO-CV errors, for various methods applied to the eye gene dataset ($p = 200, n = 119, H = 3$). $\hat{r}$ denotes the structural dimension

	Graphical method			SCAD penalty method		
$\hat{r}$	GraphSIR	GraphSAVE	GraphDR	SCAD-SIR	SCAD-SAVE	SCAD-DR
1	1.153(1.192)	1.141(1.213)	0.741(0.773)	1.147(1.192)	1.135(1.198)	0.668(0.697)
2	1.023(1.086)	1.139(1.278)	0.741(0.790)	1.026(1.094)	1.133(1.254)	0.646(0.684)
3	0.987(1.073)	1.138(1.298)	0.731(0.793)	0.974(1.058)	1.131(1.277)	0.472(0.522)
4	0.961(1.057)	1.107(1.285)	0.730(0.811)	0.951(1.051)	1.128(1.296)	0.452(0.521)
5	0.904(1.012)	1.107(1.313)	0.729(0.830)	0.917(1.027)	1.097(1.308)	0.451(0.530)
6	0.898(1.021)	1.097(1.361)	0.702(0.805)	0.861(0.980)	1.091(1.335)	0.451(0.539)

Table 9 MSE values, with the numbers in brackets representing the corresponding LOO-CV errors, for various methods applied to the eye gene dataset ($p = 200, n = 119, H = 10$). $\hat{r}$ denotes the structural dimension.

	Graphical method			SCAD penalty method		
$\hat{r}$	GraphSIR	GraphSAVE	GraphDR	SCAD-SIR	SCAD-SAVE	SCAD-DR
1	1.223(1.268)	1.139(1.215)	1.221(1.264)	1.222(1.267)	1.145(1.242)	0.977(1.016)
2	1.068(1.138)	1.139(1.308)	0.857(0.913)	1.141(1.209)	1.136(1.269)	0.970(1.109)
3	1.040(1.126)	1.136(1.367)	0.721(0.784)	1.023(1.115)	1.121(1.298)	0.757(0.913)
4	1.011(1.118)	1.136(1.434)	0.714(0.800)	1.006(1.115)	1.120(1.343)	0.551(0.713)
5	0.958(1.084)	1.115(1.459)	0.699(0.792)	0.981(1.101)	1.113(1.376)	0.551(0.750)
6	0.955(1.105)	1.110(1.517)	0.687(0.795)	0.925(1.060)	1.099(1.417)	0.519(0.730)

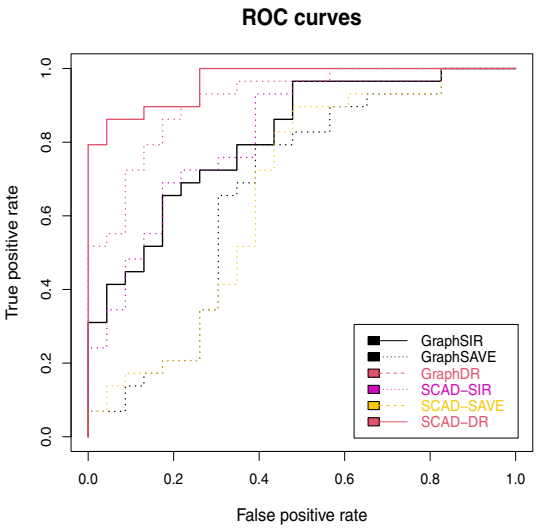

Fig. 4 The ROC curves on breast cancer dataset ($p = 287, n = 52$)

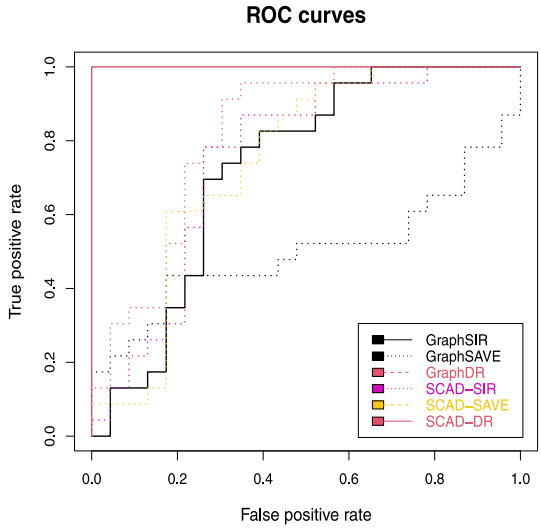

Fig. 5 The ROC curves on Duke breast cancer dataset ($p = 917, n = 46$)

and Li 26) to determine the structure dimension of the data. The chosen structural dimensions are as follows.

- GraphSIR: $\hat{r} = 1$
- GraphSAVE: $\hat{r} = 5$
- GraphDR: $\hat{r} = 1$
- SCAD-SIR: $\hat{r} = 1$
- SCAD-SAVE: $\hat{r} = 5$
- SCAD-DR: $\hat{r} = 1$

As illustrated in Figure 5, SCAD-DR exhibits superior performance compared to other methods in identifying the SDR directions. This is particularly evident in its ability to effectively discriminate between ER+ and ER- cases, which are critical in classifying breast cancer subtypes based on estrogen receptor status. The reduced directions identified by SCAD-DR capture the key variations in the data, allowing for a clear separation between the two groups. These results emphasize the practical utility of SCAD-DR in high-dimensional classification tasks in biomedical applications such as distinguishing between ER+ and ER- cases.

6 Discussion

In this paper, we introduced a graphical model based SDR method with the SCAD penalty to overcome the difficulties from traditional SDR methods for analyzing high dimension, low sample size dataset. We further presented Glasso based DR. We have provided compelling evidence of our approach's effectiveness, not only in achieving statistical accuracy in high-dimensional settings but also in successfully classifying binary response data. When applied to high-dimensional datasets, our approach consistently outperformed established baselines, including SIR, SAVE, and DR. Our approach demonstrated comparable or superior performance when compared to the graph informed method presented in [34], specifically GraphSIR and GraphSAVE.

Based on the empirical results, SCAD-SIR is best suited for problems where the relationship between predictors and response is primarily linear. SCAD-SAVE is preferable for detecting nonlinear relationships and capturing variance changes. SCAD-DR and GraphDR are the most appropriate when dealing with highly nonlinear relationships and complex interactions, offering greater flexibility at the cost of increased complexity and computational demand. However, to the best of our knowledge, it is important to note that there is no strict criterion for choosing the most suitable SDR method. The choice largely depends on the specific structure of the data and the underlying assumptions about the relationships between predictors and the response. Practitioners should consider the expected data structure, the nature of the relationships they aim to capture, and computational efficiency when selecting a method.

Recent advancements have been made in developing methods for SDR that incorporate penalties. [43] proposes a method for sparse Fréchet SDR in high-dimensional regression for non-Euclidean responses. They introduce a multitask regression framework incorporating a nonconvex penalty to identify sparse and low-dimensional representations of predictors. also introduces a new optimization algorithm called the double approximation shrinkage-thresholding algorithm to solve the nonconvex optimization problem.

SCAD (Fan and Li 13) simultaneously performs variable selection and coefficient estimation, leading to more precise and efficient modeling. The penalty functions it uses are symmetric and nonconcave with singularities at the origin. These functions minimize bias in large coefficients and guarantee continuous solutions, unlike LASSO, which can introduce bias. The SCAD approach is versatile and can be applied to various model types, including parametric, nonparametric, and generalized linear models. Although, we choose SCAD because it is well-established in practical use with simpler iterative algorithms, in future research, graphical model based SDR method found on some alternative penalities can be investigated. For instance, [50] introduces MC+, a method for combining the minimax concave penalty and the penalized linear unbiased selection algorithm. The method ensures selection consistency without relying on the strong assumptions required by LASSO, even in high-dimensional settings where $p \gg n$. Moreover, MC+ achieves certain minimax convergence rates and unbiasedness.

Appendix

Tables 10 through 19 summarize the computational times required for simulations conducted across Models 1 to 6, as well as empirical analyses involving the eye gene dataset, breast cancer dataset, and Duke breast cancer dataset. These tables provide a detailed quantitative assessment of the computational efficiency of the proposed SCAD-based methods relative to existing Graph-based approaches. Our results indicate that SCAD-SIR, SCAD-SAVE, and SCAD-DR exhibit comparable computational performance to GraphSIR, GraphSAVE, and GraphDR, despite the additional computational burden introduced by the SCAD penalty.

We evaluated the peak memory usage for each method using Models 5 and 6 to assess computational efficiency in terms of memory requirements in Tables 20 and 21. The results indicate that the peak memory consumption of the SCAD based methods is very similar to that of the graph-based methods. This similarity further emphasize the practicality of SCAD based approaches, as they achieve competitive performance without imposing significant additional memory demands, even in complex high-dimensional settings. We conducted our computations on an Apple MacBook Pro with an M2 Max chip and 32GB of RAM.

Table 10 Comparison of the average and standard deviation of computation times (in seconds) for Model 1.

		Classical method			Graph			SCAD		
n=50	$(H = 3)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	<0.01 (0.00)	<0.01 (0.00)	<0.01 (0.00)	0.90 (0.03)	1.79 (0.07)	2.24 (0.08)	0.90 (0.05)	1.79 (0.09)	2.23 (0.09)
	$p = 50$	-	-	-	0.90 (0.04)	1.80 (0.08)	2.24 (0.10)	0.91 (0.04)	1.82 (0.08)	2.25 (0.10)
	$p = 100$	-	-	-	0.93 (0.04)	1.89 (0.07)	2.35 (0.10)	0.95 (0.05)	1.95 (0.07)	2.40 (0.08)
	$(H = 5)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	<0.01 (0.00)	<0.01 (0.00)	<0.01 (0.00)	0.89 (0.04)	2.65 (0.09)	3.08 (0.13)	0.89 (0.04)	2.65 (0.10)	3.09 (0.12)
	$p = 50$	-	-	-	0.90 (0.05)	2.70 (0.13)	3.14 (0.12)	0.90 (0.04)	2.70 (0.10)	3.16 (0.12)
	$p = 100$	-	-	-	0.93 (0.04)	2.81 (0.10)	3.25 (0.10)	0.94 (0.04)	2.88 (0.12)	3.32 (0.10)
	$(H = 7)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	<0.01 (0.00)	<0.01 (0.00)	<0.01 (0.00)	0.90 (0.05)	3.60 (0.16)	4.04 (0.19)	0.90 (0.04)	3.58 (0.10)	4.03 (0.10)
	$p = 50$	-	-	-	0.91 (0.03)	3.67 (0.10)	4.13 (0.14)	0.91 (0.03)	3.66 (0.09)	4.12 (0.12)
	$p = 100$	-	-	-	0.94 (0.05)	3.79 (0.14)	4.22 (0.14)	0.94 (0.05)	3.81 (0.13)	4.26 (0.13)
n=100	$(H = 3)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	<0.01 (0.00)	<0.01 (0.00)	<0.01 (0.00)	0.88 (0.03)	1.77 (0.06)	2.20 (0.07)	0.89 (0.04)	1.77 (0.07)	2.21 (0.08)
	$p = 50$	<0.01 (0.00)	<0.01 (0.00)	<0.01 (0.00)	0.90 (0.03)	1.81 (0.06)	2.25 (0.08)	0.90 (0.04)	1.81 (0.07)	2.25 (0.08)
	$p = 100$	-	-	-	0.92 (0.04)	1.90 (0.08)	2.34 (0.10)	0.93 (0.04)	1.95 (0.08)	2.39 (0.10)
	$(H = 5)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	<0.01 (0.00)	<0.01 (0.00)	<0.01 (0.00)	0.88 (0.04)	2.65 (0.12)	3.07 (0.13)	0.88 (0.04)	2.63 (0.12)	3.08 (0.14)
	$p = 50$	<0.01 (0.00)	<0.01 (0.00)	<0.01 (0.00)	0.90 (0.04)	2.70 (0.11)	3.14 (0.13)	0.89 (0.04)	2.70 (0.12)	3.14 (0.14)
	$p = 100$	-	-	-	0.94 (0.04)	2.87 (0.13)	3.31 (0.13)	0.93 (0.04)	2.91 (0.11)	3.37 (0.14)
	$(H = 10)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	<0.01 (0.00)	<0.01 (0.01)	<0.01 (0.01)	0.97 (0.34)	5.35 (1.88)	5.81 (2.03)	0.94 (0.29)	5.33 (1.86)	5.81 (2.02)
	$p = 50$	<0.01 (0.00)	<0.01 (0.00)	<0.01 (0.00)	0.88 (0.05)	4.86 (0.25)	5.27 (0.26)	0.88 (0.05)	4.86 (0.25)	5.30 (0.27)
	$p = 100$	-	-	-	0.92 (0.05)	5.10 (0.23)	5.53 (0.26)	0.92 (0.05)	5.19 (0.23)	5.63 (0.25)

Table 11 Comparison of the average and standard deviation of computation times (in seconds) for Model 2.

		Classical method			Graph			SCAD		
n=50	$(H = 3)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	<0.01 (0.00)	0.01 (0.01)	<0.01 (0.01)	0.88 (0.04)	1.75 (0.10)	2.17 (0.11)	0.87 (0.05)	1.74 (0.08)	2.18 (0.10)
	$p = 50$	-	-	-	0.89 (0.05)	1.80 (0.09)	2.24 (0.11)	0.89 (0.05)	1.80 (0.08)	2.26 (0.11)
	$p = 100$	-	-	-	0.95 (0.04)	1.90 (0.08)	2.35 (0.12)	0.95 (0.05)	1.94 (0.10)	2.37 (0.13)
	$(H = 5)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	<0.01 (0.00)	0.01 (0.01)	<0.01 (0.01)	0.92 (0.05)	2.73 (0.15)	3.19 (0.20)	0.91 (0.06)	2.74 (0.17)	3.18 (0.18)
	$p = 50$	-	-	-	0.93 (0.05)	2.80 (0.14)	3.27 (0.18)	0.94 (0.05)	2.81 (0.16)	3.28 (0.19)
	$p = 100$	-	-	-	1.00 (0.04)	2.97 (0.14)	3.43 (0.18)	0.99 (0.06)	3.03 (0.15)	3.49 (0.17)
	$(H = 7)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	<0.01 (0.00)	0.01 (0.01)	<0.01 (0.01)	0.97 (0.05)	3.82 (0.21)	4.30 (0.23)	0.96 (0.06)	3.82 (0.23)	4.30 (0.27)
	$p = 50$	-	-	-	0.98 (0.05)	3.89 (0.22)	4.42 (0.23)	0.99 (0.05)	3.93 (0.22)	4.40 (0.24)
	$p = 100$	-	-	-	1.03 (0.05)	4.12 (0.19)	4.59 (0.25)	1.04 (0.06)	4.20 (0.20)	4.69 (0.23)
n=100	$(H = 3)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	<0.01 (0.00)	0.01 (0.01)	<0.01 (0.01)	0.89 (0.04)	1.76 (0.05)	2.20 (0.07)	0.89 (0.04)	1.76 (0.05)	2.21 (0.07)
	$p = 50$	<0.01 (0.00)	0.01 (0.01)	0.01 (0.01)	0.89 (0.05)	1.77 (0.10)	2.21 (0.13)	0.89 (0.05)	1.78 (0.11)	2.20 (0.14)
	$p = 100$	-	-	-	0.93 (0.05)	1.89 (0.10)	2.33 (0.13)	0.94 (0.05)	1.92 (0.10)	2.38 (0.13)
	$(H = 5)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	<0.01 (0.00)	<0.01 (0.01)	<0.01 (0.01)	0.91 (0.05)	2.72 (0.11)	3.17 (0.19)	0.91 (0.06)	2.73 (0.16)	3.19 (0.18)
	$p = 50$	0.01 (0.00)	0.01 (0.01)	0.01 (0.01)	0.93 (0.05)	2.80 (0.15)	3.26 (0.19)	0.93 (0.06)	2.81 (0.16)	3.25 (0.18)
	$p = 100$	-	-	-	0.99 (0.05)	2.99 (0.14)	3.44 (0.19)	0.98 (0.05)	3.04 (0.15)	3.52 (0.18)
	$(H = 10)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	<0.01 (0.00)	<0.01 (0.01)	<0.01 (0.01)	0.96 (0.05)	5.25 (0.24)	5.72 (0.33)	0.96 (0.06)	5.20 (0.28)	5.69 (0.34)
	$p = 50$	0.01 (0.00)	0.01 (0.01)	0.01 (0.01)	0.97 (0.05)	5.40 (0.26)	5.86 (0.32)	0.97 (0.06)	5.39 (0.28)	6.04 (0.65)
	$p = 100$	-	-	-	1.03 (0.05)	5.66 (0.26)	6.14 (0.31)	1.03 (0.05)	5.80 (0.27)	6.44 (0.27)

Table 12 Comparison of the average and standard deviation of computation times (in seconds) for Model 3.

		Classical method			Graph			SCAD		
n=50	$(H = 3)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	<0.01 (0.00)	<0.01 (0.01)	0.01 (0.01)	0.88 (0.04)	1.75 (0.09)	2.18 (0.12)	0.87 (0.05)	1.76 (0.09)	2.19 (0.13)
	$p = 50$	-	-	-	0.91 (0.04)	1.80 (0.09)	2.25 (0.12)	0.89 (0.05)	1.79 (0.10)	2.25 (0.13)
	$p = 100$	-	-	-	0.94 (0.05)	1.90 (0.10)	2.35 (0.13)	0.95 (0.06)	1.95 (0.10)	2.39 (0.13)
	$(H = 5)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	<0.01 (0.00)	<0.01 (0.01)	<0.01 (0.01)	0.92 (0.05)	2.74 (0.16)	3.20 (0.19)	0.92 (0.05)	2.75 (0.17)	3.18 (0.19)
	$p = 50$	-	-	-	0.97 (0.11)	2.85 (0.16)	3.29 (0.18)	0.94 (0.06)	2.83 (0.16)	3.32 (0.23)
	$p = 100$	-	-	-	0.99 (0.05)	2.97 (0.15)	3.44 (0.18)	1.00 (0.05)	3.03 (0.14)	3.52 (0.14)
	$(H = 7)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	<0.01 (0.00)	<0.01 (0.01)	0.01 (0.01)	0.97 (0.05)	3.85 (0.19)	4.31 (0.25)	0.96 (0.06)	3.82 (0.21)	4.32 (0.26)
	$p = 50$	-	-	-	0.97 (0.05)	3.92 (0.22)	4.41 (0.26)	0.98 (0.06)	3.95 (0.23)	4.43 (0.24)
	$p = 100$	-	-	-	1.03 (0.05)	4.11 (0.20)	4.61 (0.22)	1.04 (0.06)	4.19 (0.20)	4.73 (0.23)
n=100	$(H = 3)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	<0.01 (0.00)	<0.01 (0.01)	0.01 (0.01)	0.89 (0.04)	1.76 (0.09)	2.21 (0.13)	0.88 (0.05)	1.75 (0.11)	2.24 (0.16)
	$p = 50$	0.01 (0.00)	0.01 (0.01)	0.01 (0.01)	0.88 (0.05)	1.78 (0.10)	2.23 (0.12)	0.89 (0.05)	1.79 (0.11)	2.22 (0.14)
	$p = 100$	-	-	-	0.94 (0.05)	1.90 (0.09)	2.37 (0.11)	0.94 (0.05)	1.96 (0.11)	2.41 (0.14)
	$(H = 5)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	<0.01 (0.00)	<0.01 (0.01)	0.01 (0.01)	0.91 (0.05)	2.74 (0.14)	3.18 (0.20)	0.91 (0.06)	2.73 (0.17)	3.18 (0.20)
	$p = 50$	0.01 (0.00)	0.01 (0.01)	0.01 (0.01)	0.93 (0.05)	2.81 (0.15)	3.26 (0.18)	0.94 (0.05)	2.84 (0.16)	3.29 (0.20)
	$p = 100$	-	-	-	0.99 (0.05)	3.02 (0.14)	3.48 (0.17)	0.99 (0.05)	3.05 (0.17)	3.54 (0.19)
	$(H = 10)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	<0.01 (0.00)	<0.01 (0.01)	0.01 (0.01)	0.96 (0.05)	5.24 (0.29)	5.72 (0.33)	0.96 (0.06)	5.24 (0.29)	5.73 (0.34)
	$p = 50$	0.01 (0.00)	0.01 (0.01)	0.01 (0.01)	0.98 (0.05)	5.38 (0.28)	5.87 (0.31)	0.98 (0.05)	5.41 (0.30)	5.93 (0.31)
	$p = 100$	-	-	-	1.05 (0.05)	5.68 (0.21)	6.19 (0.26)	1.03 (0.06)	5.83 (0.24)	6.53 (0.33)

Table 13 Comparison of the average and standard deviation of computation times (in seconds) for Model 4.

		Classical method			Graph			SCAD		
n=50	$(H = 3)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	<0.01 (0.00)	<0.01 (0.01)	0.01 (0.01)	0.88 (0.05)	1.76 (0.08)	2.20 (0.11)	0.88 (0.05)	1.77 (0.09)	2.20 (0.12)
	$p = 50$	-	-	-	0.90 (0.06)	1.79 (0.11)	2.24 (0.14)	0.90 (0.06)	1.80 (0.11)	2.26 (0.13)
	$p = 100$	-	-	-	0.95 (0.04)	1.90 (0.10)	2.36 (0.13)	0.95 (0.05)	1.96 (0.10)	2.41 (0.12)
	$(H = 5)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	<0.01 (0.00)	<0.01 (0.01)	<0.01 (0.01)	0.92 (0.05)	2.76 (0.15)	3.22 (0.18)	0.92 (0.06)	2.76 (0.17)	3.20 (0.20)
	$p = 50$	-	-	-	0.96 (0.05)	2.86 (0.13)	3.30 (0.18)	0.94 (0.05)	2.83 (0.15)	3.31 (0.19)
	$p = 100$	-	-	-	1.00 (0.05)	3.01 (0.15)	3.47 (0.18)	1.00 (0.05)	3.03 (0.15)	3.50 (0.20)
	$(H = 7)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	<0.01 (0.00)	<0.01 (0.01)	0.01 (0.01)	0.96 (0.05)	3.84 (0.21)	4.33 (0.23)	0.96 (0.06)	3.85 (0.20)	4.32 (0.23)
	$p = 50$	-	-	-	0.98 (0.05)	3.92 (0.20)	4.41 (0.24)	0.99 (0.05)	3.95 (0.21)	4.45 (0.23)
	$p = 100$	-	-	-	1.04 (0.04)	4.14 (0.16)	4.64 (0.24)	1.04 (0.06)	4.20 (0.19)	4.73 (0.18)
n=100	$(H = 3)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	<0.01 (0.00)	<0.01 (0.01)	0.01 (0.01)	0.88 (0.05)	1.75 (0.10)	2.18 (0.11)	0.88 (0.05)	1.74 (0.11)	2.18 (0.14)
	$p = 50$	0.01 (0.00)	0.01 (0.01)	0.01 (0.01)	0.90 (0.05)	1.79 (0.10)	2.23 (0.13)	0.89 (0.05)	1.80 (0.10)	2.24 (0.14)
	$p = 100$	-	-	-	0.94 (0.04)	1.94 (0.09)	2.38 (0.14)	0.94 (0.05)	1.97 (0.10)	2.43 (0.12)
	$(H = 5)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	<0.01 (0.00)	<0.01 (0.01)	0.01 (0.01)	0.92 (0.05)	2.75 (0.15)	3.22 (0.17)	0.92 (0.05)	2.74 (0.16)	3.19 (0.20)
	$p = 50$	0.01 (0.00)	0.01 (0.01)	0.01 (0.01)	0.94 (0.05)	2.83 (0.15)	3.29 (0.19)	0.94 (0.05)	2.84 (0.16)	3.27 (0.19)
	$p = 100$	-	-	-	1.00 (0.05)	3.03 (0.13)	3.48 (0.18)	0.99 (0.06)	3.08 (0.15)	3.55 (0.16)
	$(H = 10)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	<0.01 (0.00)	<0.01 (0.01)	0.01 (0.01)	0.95 (0.05)	5.28 (0.30)	5.74 (0.30)	0.96 (0.06)	5.30 (0.28)	5.73 (0.31)
	$p = 50$	0.01 (0.00)	0.01 (0.01)	0.01 (0.01)	0.98 (0.05)	5.42 (0.26)	5.91 (0.32)	0.98 (0.06)	5.43 (0.29)	5.96 (0.32)
	$p = 100$	-	-	-	1.04 (0.05)	5.64 (0.23)	6.18 (0.32)	1.03 (0.05)	5.83 (0.27)	6.54 (0.39)

Table 14 Comparison of the average and standard deviation of computation times (in seconds) for Model 5.

		Classical method			Graph			SCAD		
n=50	$(H = 3)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	<0.01 (0.00)	<0.01 (0.01)	<0.01 (0.01)	0.89 (0.05)	1.77 (0.10)	2.20 (0.11)	0.88 (0.05)	1.77 (0.09)	2.21 (0.11)
	$p = 50$	-	-	-	0.90 (0.05)	1.81 (0.10)	2.25 (0.13)	0.90 (0.05)	1.81 (0.11)	2.25 (0.14)
	$p = 100$	-	-	-	0.95 (0.05)	1.91 (0.10)	2.37 (0.13)	0.96 (0.06)	1.96 (0.10)	2.42 (0.12)
	$(H = 5)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	<0.01 (0.00)	<0.01 (0.01)	<0.01 (0.01)	0.92 (0.05)	2.76 (0.14)	3.24 (0.19)	0.93 (0.06)	2.77 (0.17)	3.22 (0.18)
	$p = 50$	-	-	-	0.95 (0.05)	2.86 (0.14)	3.30 (0.18)	0.95 (0.06)	2.85 (0.17)	3.32 (0.20)
	$p = 100$	-	-	-	1.01 (0.05)	3.00 (0.15)	3.47 (0.17)	1.01 (0.05)	3.07 (0.14)	3.52 (0.16)
	$(H = 7)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	<0.01 (0.01)	<0.01 (0.01)	<0.01 (0.01)	0.97 (0.05)	3.86 (0.20)	4.34 (0.25)	0.96 (0.06)	3.87 (0.22)	4.35 (0.25)
	$p = 50$	-	-	-	0.99 (0.05)	3.95 (0.20)	4.45 (0.25)	0.99 (0.05)	3.98 (0.22)	4.44 (0.25)
	$p = 100$	-	-	-	1.05 (0.04)	4.14 (0.20)	4.65 (0.22)	1.05 (0.05)	4.22 (0.19)	4.75 (0.23)
n=100	$(H = 3)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	<0.01 (0.00)	<0.01 (0.01)	<0.01 (0.01)	0.88 (0.05)	1.75 (0.11)	2.19 (0.13)	0.88 (0.06)	1.74 (0.12)	2.17 (0.13)
	$p = 50$	<0.01 (0.00)	0.01 (0.01)	0.01 (0.01)	0.90 (0.05)	1.80 (0.10)	2.24 (0.13)	0.90 (0.05)	1.80 (0.11)	2.24 (0.15)
	$p = 100$	-	-	-	0.95 (0.04)	1.92 (0.10)	2.37 (0.12)	0.95 (0.05)	1.96 (0.10)	2.41 (0.12)
	$(H = 5)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	<0.01 (0.00)	<0.01 (0.01)	<0.01 (0.01)	0.92 (0.05)	2.75 (0.15)	3.22 (0.19)	0.92 (0.06)	2.76 (0.17)	3.21 (0.18)
	$p = 50$	0.01 (0.00)	0.01 (0.01)	0.01 (0.01)	0.95 (0.05)	2.84 (0.14)	3.31 (0.19)	0.94 (0.05)	2.84 (0.17)	3.31 (0.19)
	$p = 100$	-	-	-	1.00 (0.05)	3.01 (0.14)	3.50 (0.20)	0.99 (0.06)	3.08 (0.16)	3.57 (0.18)
	$(H = 10)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	<0.01 (0.00)	<0.01 (0.01)	<0.01 (0.01)	0.97 (0.05)	5.31 (0.28)	5.78 (0.35)	0.97 (0.06)	5.30 (0.32)	5.74 (0.31)
	$p = 50$	0.01 (0.00)	0.01 (0.01)	0.01 (0.01)	0.98 (0.05)	5.47 (0.27)	5.91 (0.32)	0.99 (0.06)	5.46 (0.30)	5.93 (0.30)
	$p = 100$	-	-	-	1.04 (0.04)	5.71 (0.29)	6.19 (0.29)	1.03 (0.06)	5.85 (0.25)	6.47 (0.29)

Table 15 Comparison of the average and standard deviation of computation times (in seconds) for Model 6.

		Classical method			**Graph**			**SCAD**		
n=600	$(H = 2)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 20$	<0.01 (0.00)	<0.01 (0.00)	<0.01 (0.01)	1.59 (0.20)	2.39 (0.32)	3.19 (0.39)	1.59 (0.21)	2.37 (0.29)	3.12 (0.41)
	$p = 300$	<0.01 (0.00)	<0.01 (0.00)	<0.01 (0.00)	2.34 (0.14)	6.00 (0.28)	6.51 (0.29)	2.40 (0.14)	5.97 (0.24)	6.50 (0.31)
	$p = 1000$	-	-	-	49.31 (9.97)	79.52 (3.27)	80.11 (3.27)	44.89 (2.89)	128.26 (5.99)	128.65 (5.86)

Table 16 Comparison of the computation times (in seconds) for various methods applied to the eye gene dataset ($p = 200$, $n = 119$, $H = 3$).

	Graphical method			SCAD penalty method		
$\hat{r}$	GraphSIR	GraphSAVE	GraphDR	SCAD-SIR	SCAD-SAVE	SCAD-DR
1	2.23	3.84	4.51	2.31	4.03	4.84
2	2.18	3.82	4.51	2.35	4.01	4.70
3	2.23	3.79	4.53	2.29	4.04	4.65
4	2.25	3.79	4.54	2.28	4.03	4.67
5	2.23	3.80	4.54	2.29	4.02	4.68
6	2.19	3.79	4.54	2.30	4.08	4.75

Table 17 Comparison of the computation times (in seconds) for various methods applied to the eye gene dataset ($p = 200$, $n = 119$, $H = 10$).

	Graphical method			SCAD penalty method		
$\hat{r}$	GraphSIR	GraphSAVE	GraphDR	SCAD-SIR	SCAD-SAVE	SCAD-DR
1	2.16	9.44	10.40	2.27	10.02	10.75
2	2.20	9.41	10.31	2.28	9.95	10.76
3	2.23	9.50	10.28	2.27	9.98	10.91
4	2.23	9.60	10.17	2.26	9.96	10.91
5	2.20	9.65	10.32	2.27	9.97	10.86
6	2.21	9.62	10.61	2.31	10.05	10.90

Table 18 Comparison of the computation times (in seconds) for various methods applied to breast cancer dataset ($p = 287$, $n = 52$).

Graphical method			SCAD penalty method		
GraphSIR	GraphSAVE	GraphDR	SCAD-SIR	SCAD-SAVE	SCAD-DR
2.43	3.35	4.05	2.57	3.68	4.46

Table 19 Comparison of the average of computation times (in seconds) for various methods applied to Duke breast cancer dataset ($p = 917$, $n = 46$).

Graphical method			SCAD penalty method		
GraphSIR	GraphSAVE	GraphDR	SCAD-SIR	SCAD-SAVE	SCAD-DR
20.92	27.57	28.53	23.1	34.2	35.36

Table 20 Comparison of the average and standard deviation of peak memory usage (GB) for Model 5.

		Classical method			Graph			SCAD		
n=50	$(H = 3)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	63.77 (62.25)	68.10 (69.48)	76.78 (82.22)	51.60 (45.25)	60.23 (54.59)	68.79 (69.59)	57.87 (47.68)	52.56 (52.90)	62.85 (74.91)
	$p = 50$	-	-	-	247.41 (0.79)	323.80 (0.93)	345.44 (1.24)	347.69 (0.96)	423.18 (1.13)	444.86 (1.51)
	$p = 100$	-	-	-	927.00 (0.80)	1229.49 (0.94)	1312.63 (1.25)	1324.40 (0.97)	1625.09 (1.13)	1708.26 (1.52)
	$(H = 5)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	39.02 (33.30)	44.46 (41.72)	54.03 (56.91)	36.94 (34.37)	46.45 (45.44)	55.63 (64.09)	44.08 (40.06)	53.96 (52.29)	64.48 (74.21)
	$p = 50$	-	-	-	247.58 (0.79)	358.79 (0.93)	381.32 (1.24)	347.86 (0.96)	458.57 (1.13)	481.14 (1.51)
	$p = 100$	-	-	-	927.27 (0.80)	1369.28 (0.94)	1454.21 (1.25)	1324.67 (0.97)	1765.68 (1.13)	1850.64 (1.52)
	$(H = 7)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	39.61 (33.24)	46.62 (41.72)	57.10 (56.91)	36.96 (34.36)	47.78 (44.87)	57.18 (63.42)	44.10 (40.05)	55.31 (51.71)	66.05 (73.54)
	$p = 50$	-	-	-	247.68 (0.79)	394.01 (0.93)	417.43 (1.24)	347.96 (0.96)	493.89 (1.13)	517.34 (1.51)
	$p = 100$	-	-	-	927.47 (0.80)	1509.60 (0.94)	1596.31 (1.25)	1324.87 (0.97)	1906.20 (1.13)	1992.95 (1.52)
n=100	$(H = 3)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	62.65 (49.38)	56.93 (43.48)	58.05 (56.91)	55.34 (43.60)	56.27 (45.24)	58.31 (63.15)	63.97 (49.63)	63.63 (51.81)	66.96 (73.36)
	$p = 50$	221.56 (33.26)	252.08 (41.72)	288.74 (56.91)	256.15 (0.79)	333.34 (0.93)	354.99 (1.24)	356.43 (0.96)	431.92 (1.13)	453.60 (1.51)
	$p = 100$	-	-	-	940.74 (0.80)	1244.83 (0.94)	1327.80 (1.05)	1338.14 (0.97)	1638.83 (1.13)	1722.01 (1.52)
	$(H = 5)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	46.06 (33.28)	51.46 (41.72)	61.82 (56.91)	41.05 (32.81)	50.64 (43.85)	59.82 (62.50)	48.19 (38.50)	58.07 (50.74)	68.59 (72.66)
	$p = 50$	225.79 (33.25)	288.78 (41.72)	332.53 (56.91)	256.33 (0.79)	368.04 (0.93)	390.57 (1.24)	356.60 (0.96)	467.32 (1.13)	489.88 (1.51)
	$p = 100$	-	-	-	941.01 (0.80)	1384.02 (0.94)	1468.95 (1.25)	1338.24 (0.85)	1779.42 (1.13)	1864.39 (1.52)
	$(H = 10)$	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR
	$p = 10$	48.80 (33.24)	58.22 (41.72)	71.21 (56.91)	41.21 (32.74)	54.09 (42.36)	63.78 (60.79)	48.36 (38.43)	61.69 (49.18)	72.64 (70.91)
	$p = 50$	232.93 (33.25)	380.39 (41.72)	442.66 (56.92)	256.72 (0.79)	455.98 (0.93)	480.73 (1.24)	356.99 (0.96)	555.76 (1.13)	580.54 (1.51)
	$p = 100$	-	-	-	941.66 (0.80)	1734.46 (0.94)	1823.86 (1.25)	1338.83 (0.94)	2130.87 (1.13)	2220.30 (1.52)

Table 21 Comparison of the average and standard deviation of peak memory usage (GB) for Model 6.

		Classical method			Graph			SCAD		
n=600 ($H = 2$)	SIR	SAVE	DR	SIR	SAVE	DR	SIR	SAVE	DR	
$p = 20$	168.68 (51.06)	174.64 (56.73)	198.41 (67.73)	133.37 (34.31)	157.74 (32.34)	163.21 (44.16)	152.23 (38.56)	176.10 (34.66)	178.58 (51.26)	
$p = 300$	5716.20 (51.01)	6396.04 (28.65)	6348.15 (15.00)	8455.89 (34.56)	10617.09 (34.57)	11343.54 (34.53)	12018.55 (38.86)	14089.76 (38.87)	14816.24 (38.82)	
$p = 1000$	-	-	-	88038.37 (2455.77)	110020.50 (3566.07)	114335.52 (6333.79)	90326.21 (5346.74)	111795.51 (5219.07)	114653.69 (6307.80)	

Acknowledgements We extend our gratitude to the Editor, the Associate Editor, and the three referees for their insightful reviews, which have greatly helped improve this paper.

Author Contributions The manuscript conceptualization, writing, reviewing, and editing were equally shared by Y.P, K.K, and J.K.Y.

Funding For Jae Keun Yoo, this work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Korean Ministry of Education (RS-2023-00240564 and RS-2023-00217022).
For Kyongwon Kim, this work is supported by the National Research Foundation of Korea (NRF) Grant funded by the Korean government (MSIT) (RS-2023-00219212, RS-2025-00513476) and the High-Potential Individuals Global Training Program (RS-2022-00154879) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

Declarations

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

References

Augugliaro, L., Mineo, A.M., Wit, E.C.: Differential geometric least angle regression: a differential geometric approach to sparse generalized linear models. Journal of the Royal Statistical Society Series B: Statistical Methodology **75**(3), 471–498 (2013). https://doi.org/10.1111/rssb.12000

Benouareth, A.: An efficient face recognition approach combining likelihood-based sufficient dimension reduction and lda. Multimedia Tools and Applications **80**(1), 1457–1486 (2021). https://doi.org/10.1007/s11042-020-09527-9

Bertram, C., Hass, R.: Mmp-7 is involved in the aging of primary human mammary epithelial cells (hmec). Experimental Gerontology **43**(3), 209–217 (2008)

Bucan, V., Reimers, K., Choi, C.Y., Eddy, M.-T., Vogt, P.M.: The anti-apoptotic protein lifeguard is expressed in breast cancer cells and tissues. Cellular & molecular biology letters **15**, 296–310 (2010)

Cheng, D., Li, J., Liu, L., Le, T.D., Liu, J., Yu, K.: Sufficient dimension reduction for average causal effect estimation. Data Mining and Knowledge Discovery **36**(3), 1174–1196 (2022). https://doi.org/10.1007/s10618-022-00832-5

Cook, R.D.: Save: a method for dimension reduction and graphics in regression. Communications in statistics-Theory and methods **29**(9–10), 2109–2121 (2000). https://doi.org/10.1080/03610920008832598

Cook, R.D.: An Introduction to Envelopes: Dimension Reduction for Efficient Estimation in Multivariate Statistics. John Wiley & Sons, New York (2018)

Cook, R.D., Weisberg, S.: Discussion of sliced inverse regression for dimension reduction. Journal of the American Statistical Association **86**(414), 328–332 (1991). https://doi.org/10.2307/2290563

Duan, Z., Lamendola, D., Yusuf, R., Penson, R., Preffer, F., Seiden, M.: Overexpression of human phosphoglycerate kinase 1 (pgk1) induces a multidrug resistance phenotype. Anticancer research **22**(4), 1933–1941 (2002)

Eaton, M.L.: A characterization of spherical distributions. Journal of Multivariate Analysis **20**(2), 272–276 (1986). https://doi.org/10.1016/0047-259X(86)90083-7

Fogel, M., Friederichs, J., Zeller, Y., Husar, M., Smirnov, A., Roitman, L., Altevogt, P., Sthoeger, Z.M.: Cd24 is a marker for human breast carcinoma. Cancer letters **143**(1), 87–94 (1999)

Friedman, J., Hastie, T., Tibshirani, R.: Sparse inverse covariance estimation with the graphical lasso. Biostatistics **9**(3), 432–441 (2008). https://doi.org/10.1093/biostatistics/kxm045

Fan, J., Li, R.: Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association **96**(456), 1348–1360 (2001). https://doi.org/10.1198/016214501753382273

Fan, J., Liao, Y., Liu, H.: An overview of the estimation of large covariance and precision matrices. The Econometrics Journal **19**(1), 1–32 (2016). https://doi.org/10.1111/ectj.12061

Gold, B., Kalush, F., Bergeron, J., Scott, K., Mitra, N., Wilson, K., Ellis, N., Huang, H., Chen, M., Lippert, R.: Estrogen receptor genotypes and haplotypes associated with breast cancer risk. Cancer research **64**(24), 8891–8900 (2004)

Hilafu, H., Safo, S.E.: Sparse sliced inverse regression for high dimensional data analysis. BMC bioinformatics **23**(1), 1–19 (2022). https://doi.org/10.1186/s12859-022-04700-3

Holst, F., Stahl, P.R., Ruiz, C., Hellwinkel, O., Jehan, Z., Wendland, M., Lebeau, A., Terracciano, L., Al-Kuraya, K., Jänicke, F.: Estrogen receptor alpha (esr1) gene amplification is frequent in breast cancer. Nature genetics **39**(5), 655–660 (2007)

Kim, K.: On principal graphical models with application to gene network. Computational Statistics & Data Analysis **166**, 107344 (2022)

Kim, K., Yoo, J.K.: More on directional regression. Communications for Statistical Applications and Methods **28**(5), 553–562 (2021). https://doi.org/10.29220/CSAM.2021.28.5.553

Lam, C., Fan, J.: Sparsistency and rates of convergence in large covariance matrix estimation. The Annals of Statistics **37**(6B), 4254–4278 (2009). https://doi.org/10.1214/09-AOS720

Li, K.C.: Sliced inverse regression for dimension reduction. Journal of the American Statistical Association **86**(414), 316–327 (1991). https://doi.org/10.2307/2290563

Li, K.C.: High dimensional data analysis via the sir/phd approach **5** (2000) http://www.stat.ucla.edu/~kcli/sir-PHD.pdf

Li, L.: Sparse sufficient dimension reduction. Biometrika **94**(3), 603–613 (2007). https://doi.org/10.1093/biomet/asm044

Li, B.: Sufficient Dimension Reduction: Methods and Applications with R. Chapman and Hall/CRC, New York (2018)

Li, B., Kim, K.: On sufficient graphical models. Journal of Machine Learning Research **25**(17), 1–64 (2024)

Luo, W., Li, B.: Combining eigenvalues and variation of eigenvectors for order determination. Biometrika **103**(4), 875–887 (2016)

Li, B., Wang, S.: On directional regression for dimension reduction. Journal of the American Statistical Association **102**(479), 997–1008 (2007). https://doi.org/10.1198/016214507000000536

Li, B., Zha, H., Chiaromonte, F.: Contour regression: a general approach to dimension reduction. The Annals of Statistics **33**(4), 1580–1616 (2005). https://doi.org/10.1214/009053605000000192

Lin, Q., Zhao, Z., Liu, J.S.: On consistency and sparsity for sliced inverse regression in high dimensions. The Annals of Statistics **46**(2), 580–610 (2018). https://doi.org/10.1214/17-AOS1561

Meinshausen, N., Bühlmann, P.: High-dimensional graphs and variable selection with the lasso. The Annals of Statistics **34**(3), 1436–1462 (2006). https://doi.org/10.1214/009053606000000281

Mazumder, R., Hastie, T.: The graphical lasso: New insights and alternatives. Electronic journal of statistics **6**, 2125–2149 (2012). https://doi.org/10.1214/12-EJS740

Meng, C., Yu, J., Zhang, J., Ma, P., Zhong, W.: Sufficient dimension reduction for classification using principal optimal transport direction. Advances in Neural Information Processing Systems **33**, 4015–4028 (2020)

Ma, S., Zhu, L., Zhang, Z., Tsai, C.-L., Carroll, R.J.: A robust and efficient approach to causal inference based on sparse sufficient dimension reduction. The Annals of Statistics **47**(3), 1505 (2019). https://doi.org/10.1214/18-AOS1722

Pircalabelu, E., Artemiou, A.: Graph informed sliced inverse regression. Computational Statistics & Data Analysis **164**, 107302 (2021). https://doi.org/10.1016/j.csda.2021.107302

Qian, W., Ding, S., Cook, R.D.: Sparse minimum discrepancy approach to sufficient dimension reduction with simultaneous variable selection in ultrahigh dimension. Journal of the American Statistical Association **114**(527), 1277–1290 (2019). https://doi.org/10.1080/01621459.2018.1497498

Rothman, A.J., Bickel, P.J., Levina, E., Zhu, J.: Sparse permutation invariant covariance estimation. Electronic Journal of Statistics **2**, 494–515 (2008). https://doi.org/10.1214/08-EJS176

Scheetz, T.E., Kim, K.-Y.A., Swiderski, R.E., Philp, A.R., Braun, T.A., Knudtson, K.L., Dorrance, A.M., DiBona, G.F., Huang, J., Casavant, T.L.: Regulation of gene expression in the mammalian eye and its relevance to eye disease. Proceedings of the National Academy of Sciences **103**(39), 14429–14434 (2006). https://doi.org/10.1073/pnas.0602562103

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

Tomassi, D., Forzani, L., Duarte, S., Pfeiffer, R.M.: Sufficient dimension reduction for compositional data. Biostatistics **22**(4), 687–705 (2021). https://doi.org/10.1093/biostatistics/kxz060

Tan, K., Shi, L., Yu, Z.: Sparse sir: Optimal rates and adaptive estimation. The Annals of Statistics **48**(1), 64–85 (2020). https://doi.org/10.1214/18-AOS1791

Tan, K.M., Wang, Z., Zhang, T., Liu, H., Cook, R.D.: A convex formulation for high-dimensional sparse sliced inverse regression. Biometrika **105**(4), 769–782 (2018). https://doi.org/10.1093/biomet/asy049

Wang, J., Higuchi, R., Modugno, F., Li, J., Umblas, N., Lee, J., Lui, L.-Y., Ziv, E., Tice, J.A., Cummings, S.R.: Estrogen receptor alpha haplotypes and breast cancer risk in older caucasian women. Breast cancer research and treatment **106**, 273–280 (2007)

Weng, J., Ke, C., Wang, P.: Sparse fréchet sufficient dimension reduction via nonconvex optimization. In: Conference on Parsimony and Learning, pp. 39–53 (2024). PMLR

Wakefield, L., Robinson, J., Long, H., Ibbitt, J.C., Cooke, S., Hurst, H.C., Sim, E.: Arylamine n-acetyltransferase 1 expression in breast cancer cell lines: a potential marker in estrogen receptor-positive tumors. Genes, Chromosomes and Cancer **47**(2), 118–126 (2008)

Yin, X., Hilafu, H.: Sequential sufficient dimension reduction for large p, small n problems. Journal of the Royal Statistical Society Series B: Statistical Methodology **77**(4), 879–892 (2015). https://doi.org/10.1111/rssb.12093

Yuan, M., Lin, Y.: Model selection and estimation in the gaussian graphical model. Biometrika **94**(1), 19–35 (2007). https://doi.org/10.1093/biomet/asm018

Yoo, J.K.: Analysis of microarray right-censored data through fused sliced inverse regression. Scientific Reports **9**(1), 15094 (2019). https://doi.org/10.1038/s41598-019-51441-0

Yuan, M.: High dimensional inverse covariance matrix estimation via linear programming. The Journal of Machine Learning Research **11**, 2261–2286 (2010)

Zhang, J., Chen, X.: Robust sufficient dimension reduction via ball covariance. Computational Statistics & Data Analysis **140**, 144–154 (2019)

Zhang, C.-H.: Nearly unbiased variable selection under minimax concave penalty. The Annals of Statistics **38**(2), 894–942 (2010). https://doi.org/10.1214/09-AOS729

Zou, H., Li, R.: One-step sparse estimates in nonconcave penalized likelihood models. The Annals of Statistics **36**(4), 1509–1533 (2008). https://doi.org/10.1214/009053607000000802

Zhao, T., Liu, H.: Calibrated precision matrix estimation for high-dimensional elliptical distributions. IEEE transactions on Information Theory **60**(12), 7874–7887 (2014). https://doi.org/10.1109/TIT.2014.2360980

Zhang, D., Tai, L.K., Wong, L.L., Chiu, L.-L., Sethi, S.K., Koay, E.S.: Proteomic study reveals that proteins involved in metabolic and detoxification pathways are highly expressed in her-2/neu-positive breast cancer. Molecular & Cellular Proteomics **4**(11), 1686–1696 (2005)