Sparse Group Sufficient Dimension Reduction with Applications to Gene Pathway Data

Bilin Zeng[1], Xuerong Meggie Wen[2], and Lixing Zhu[3]

[1,2]*Department of Mathematics and Statistics, Missouri University of Science and Technology, MO 65409, U.S.A.*

[3]*Department of Mathematics, Hong Kong Baptist University, Hong Kong, China*

Abstract

For regression problems with grouped covariates, we adopt the idea of sparse group lasso (Friedman et al., 2010) to the framework of the sufficient dimension reduction. We propose a method called the *sparse group sufficient dimension reduction* (sgSDR) to conduct group and withingroup variable selections simultaneously without assuming a specific model structure on the regression function. Simulation studies show that our method is comparable to the sparse group lasso under the regular linear model setting, and outperforms sparse group lasso with higher true positive rates and substantially lower false positive rates when the regression function is nonlinear or (and) the error distributions are non-Gaussian. One immediate application of our method is to the gene pathway data analysis where genes naturally fall into groups (pathways). An analysis of a glioblastoma microarray data is included for illustration of our method.

KEY WORDS: Sparse Group Lasso; Gene Pathway Analysis; Sufficient Dimension Reduction.

1 Introduction

1.1 Sufficient Dimension Reduction

For a typical regression problem with a univariate random response Y and a p-dimensional random vector $\mathbf{X}$, sufficient dimension reduction (SDR: Li, 1991; Cook and Weisberg,

1991; Cook, 1998) aims to reduce the dimension of $\mathbf{X}$ without loss of information on the regression and without requiring a pre-specified parametric model. The basic idea of sufficient dimension reduction is to replace the predictors $\mathbf{X} \in \mathbb{R}^p$ with a lower dimensional projection $P_{\mathcal{S}}\mathbf{X}$ onto a subspace $\mathcal{S} \subseteq \mathbb{R}^p$ without the loss of information on the original regression of $Y|\mathbf{X}$, i.e., $Y \perp\!\!\!\perp \mathbf{X}|P_{\mathcal{S}}\mathbf{X}$, where $\perp\!\!\!\perp$ indicates independence and $P_{(.)}$ stands for a projection operator with respect to the standard inner product. Such an $\mathcal{S}$ is defined as a dimension reduction subspace, and the smallest one is called the *central subspace* $\mathcal{S}_{Y|\mathbf{X}}$ (Cook, 1998), which exists under very mild conditions (Cook, 1998; Yin et al., 2008). We assume the existence of $\mathcal{S}_{Y|\mathbf{X}}$ throughout this article. $\dim(\mathcal{S}_{Y|\mathbf{X}}) = d$ is called the structural dimension of the regression. The goal of sufficient dimension reduction is to estimate and make statistical inferences about $\mathcal{S}_{Y|\mathbf{X}}$ and d. Subsequent modeling and prediction can then be built upon the reduced dimensional projection.

Many methods have been developed to estimate $\mathcal{S}_{Y|\mathbf{X}}$. Among them, sliced inverse regression (SIR; Li, 1991), sliced average variance estimation (SAVE; Cook and Weisberg, 1991), minimum average variance estimation (MAVE; Xia et al., 2002), directional regression (DR; Li and Wang, 2007) and are the most widely investigated methods in the literature. Cook and Li (2002) proposed the central mean subspace where the interest of dimension reduction is restricted to the conditional mean function $\mathrm{E}(Y|\mathbf{X})$. Recently, Li et al. (2010) proposed a groupwise dimension reduction which incorporates the prior group information when the predictors under investigation fall naturally into several groups.

As pointed by Bondell and Li (2011), the general framework of sufficient dimension reduction is also useful for variable selection since no pre-specified underlying models between Y and $\mathbf{X}$ are required. Instead, usually a so-called "linearity condition" (Hall and Li, 1993; Wen and Cook, 2007) on the marginal distribution of $\mathbf{X}$ is assumed. This is a mild condition and holds approximately true when p goes to infinity. Ni, Cook and Tsai (2005), Li and Nachtsheim (2006) and Li and Yin (2008) proposed model-free variable selections by reformulating SDR as a penalized regression problem. Li (2007) proposed a unified approach combining SDR and shrinkage estimation to produce sparse estimators

of the central subspace. Wang et al. (2012) proposed a distribution-weighted lasso method for the single-index model. However, none of those model-free variable selections take the prior group (predictor network) information into account. Such situations do arise in the gene pathway analysis where genes naturally fall into groups (pathways/ gene networks; see the following subsection for more discussions). In this paper, we propose a method called the sparse group sufficient dimension reduction (sgSDR), which conducts both group and within-group variable selections simultaneously under the framework of sufficient dimension reduction. We then apply our method to a survival analysis for glioblastoma patients (Horvath et al., 2006) using gene expression profiles with about 1500 genes and 33 pathways.

1.2 Gene Pathway Analysis

Genetic association studies aim to detect the associations between gene expressions and the occurrence or progression of disease phenotypes. Recent developments in microarray techniques make it possible to profile gene expressions on a whole genome scale, simultaneously measuring expressions of thousands or tens of thousands of genes. New challenges arise for the analysis of microarray data due to the large number of genes surveyed and often the relatively small sample sizes. A large amount of existing approaches (to list a few: Alon et al., 1999; Dudoit et al., 2002; Nguyen and Rocke, 2002; Rosenwald et al., 2003) has been developed to identify a small subset of genes or linear combinations of genes which are often referred to as super genes, that have influential effects on diseases. Such studies can lead to better understanding of the genetic causation of diseases and better predictive models. However, since the presence of cluster structure of genes (gene pathways) was ignored, these methods are insufficient to dissect the complex genetic structure of many common diseases. Here the clusters are composed of co-regulated genes with coordinated functions. Gene annotation databases, such as KEGG (Ogata et al., 2000), Reactome (Matthews et al., 2008), PID (http://pid.nci.nih.gov/) and BioCyc (Karp et al., 2005), group functionally relevant genes into biological pathways. Since it

is commonly believed that genes carry out their functions through intricate pathways of reactions and interactions, intuitively, pathway-based analysis can offer an attractive alternative to improve the power of gene (or SNP)-based methods, and may help us to identify relevant subsets of genes in meaningful biological pathways underlying complex diseases.

There are considerable interests in pathway-based analysis (to list a few: Manoli et al., 2006; Wang et al., 2007; Li and Li, 2008; Wei and Pan, 2008; Ma and Kosorok, 2009; Pan et al., 2010; Zhu and Li, 2011). Pathway-based approaches in microarray data analysis often yield biological insights that are otherwise undetectable by focusing only on genes with the strongest evidence of differential expressions. Most pathway-based methods focus on identifying meaningful biological pathways underlying complex diseases, assuming that if a pathway (cluster) is strongly associated with the phenotype, then all genes within that pathway are associated with the phenotype. However, if only a subset of genes within a pathway contributes to the outcome, then these methods may result in loss of power. Our sparse group sufficient dimension reduction is developed to address this problem, where pathway selection and within pathway gene selection can be achieved simultaneously. Details of our method will be presented in Section 2.

The remainder of this article is organized as follows. Section 2 describes our statistical approach. It first reviews the sparse group lasso (Friedman et al., 2010), and then shows how it can be extended within the context of sufficient dimension reduction. The SLEP package (Liu et al., 2009) is adopted for the implementation of our method. Five-fold cross-validation is used to select the related tuning parameters. Section 3 reports simulation studies comparing the finite-sample performances of our method with the sparse group lasso. A real data example on glioblastoma study (Horvath et al., 2006) is discussed in Section 4. Conclusions and a brief discussion on future research directions are given in Section 5.

2 Sparse Group Sufficient Dimension Reduction

The lasso-penalized linear regression (Tibshirani, 1996) is applied to high-dimensional regression problems with tens to hundreds of thousands of predictors. It finds a solution with few nonzero entries by minimizing:

$$\frac{1}{2}||\mathbf{y} - \mathcal{X}\boldsymbol{\beta}||_2^2 + \lambda||\boldsymbol{\beta}||_1, \tag{2.1}$$

where $\mathbf{y} = (y_1, \ldots, y_n)^T$ is the observed centered response vector, $\mathcal{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$ is the centered design matrix with $\mathbf{x}_i = (x_1^i, \ldots, x_p^i)^T$ being the predictor values for the ith observed subject, $\boldsymbol{\beta} \in \mathbb{R}^p$ the vector of regression coefficients, $||\mathbf{z}||_2^2 = (\sum_j z_j^2)^{\frac{1}{2}}$ the Euclidean (l_2) norm and $||\mathbf{z}||_1 = \sum_j |z_j|$ the l_1 norm. The first term in (2.1) represents the loss function minimized in the ordinary least squares, the second term is the lasso penalty function while the multiplier $\lambda > 0$ is the penalty constant. Large value of λ will set some components β_j exactly to 0. The lasso has become a popular model selection and shrinkage estimation method since it is capable of producing sparse models and is computationally feasible. In some applications, it is natural to group predictors (Yuan and Lin, 2006). This raises the question of how to penalize a group of parameters. The group lasso proposed by Yuan and Lin (2006) overcomes that problem by minimizing the following penalized least squares regression:

$$\frac{1}{2}||\mathbf{y} - \sum_{g=1}^{G} \mathcal{X}^{(g)}\boldsymbol{\beta}^{(g)}||_2^2 + \lambda \sum_{g=1}^{G} \sqrt{p_g}||\boldsymbol{\beta}^{(g)}||_2, \tag{2.2}$$

where $\mathcal{X}^{(g)}$ is the submatrix of $\mathcal{X}$ with columns corresponding to the predictors in the gth group, $\boldsymbol{\beta}^{(g)}$ the coefficient vector of that group with p_g as its length. The rescaling factor p_g makes the penalty level proportional to the group size, which ensures that small groups are not overwhelmed by large groups in group selections. The group lasso penalty has been investigated in multiple studies (Bakin, 1999; Meier et al., 2008; Huang et al., 2009). The sparsity of the solution is determined by the tuning parameter λ. When the group size $p_g = 1$, group lasso is reduced to the regular lasso. While the group lasso can identify important groups, it is not capable of selecting important predictors within each group, which will be an issue when p_g is large.

Friedman et al. (2010) proposed the sparse group lasso (SGL) which could achieve sparsity of both groups and within each group by minimizing the following penalized least squares regression:

$$\frac{1}{2}||\mathbf{y} - \sum_{g=1}^{G} \mathcal{X}^{(g)}\boldsymbol{\beta}^{(g)}||_2^2 + \lambda_1 \sum_{g=1}^{G} \sqrt{p_g}||\boldsymbol{\beta}^{(g)}||_2 + \lambda_2||\boldsymbol{\beta}||_1. \tag{2.3}$$

It is reduced to the group lasso when $\lambda_2 = 0$, and the lasso when $\lambda_1 = 0$. Sparse group lasso is capable of selecting important groups and important predictors within the selected groups simultaneously. It might lead to better predictions since it takes the cluster structure into consideration; and also, its within-group variable selection aspect can lead to more parsimonious models and hence interpretable results. However, all the above lasso-based methods assume a **linear** relationship between the response and the predictors, and may **not** be robust to non-Gaussian errors. We propose a sparse group sufficient dimension reduction method to overcome these limitations.

Li et al. (2010) proposed the groupwise dimension reduction which incorporates the prior grouping information into the estimation of the central mean subspace. Simulation studies and real data analyses showed that the groupwise dimension reduction approach can substantially increase the estimation accuracy and enhance the estimates interpretability. However, their method is limited to the dimension reduction of the conditional mean ($\mathrm{E}(Y|\mathbf{X})$), and is not capable of variable selections. The *sparse group sufficient dimension reduction* (sgSDR) method we propose in this article can conduct variable selection in the general dimension reduction context (not limited only to the conditional mean) while incorporating the group knowledge, and also can be applied to the $n << p$ setting.

We focus on the following general single-index model:

$$Y = g(\boldsymbol{\beta}^T\mathbf{X}, \epsilon) \tag{2.4}$$

Without loss of generality, we assume that $\mathbf{X}$ is centered with $\mathrm{E}(\mathbf{X}) = 0$, and also suppose that $\mathbf{X}$ can be splitted into G groups, $\mathbf{X}^T = (\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \cdots, \mathbf{X}^{(G)})$, where $\mathbf{X}^{(g)}$ is a p_g-dimensional row vector, for $g = 1, \ldots, G$, and $\sum_{g=1}^{G} p_g = p$. Following Wang et al. (2012),

we consider the following minimization problem:

$$\frac{1}{2}||\mathbf{F}_n(\mathbf{y}) - \sum_{g=1}^{G} \mathcal{X}^{(g)}\boldsymbol{\beta}^{(g)}||_2^2 + \lambda_1 \sum_{g=1}^{G} \sqrt{p_g}||\boldsymbol{\beta}^{(g)}||_2 + \lambda_2||\boldsymbol{\beta}||_1, \tag{2.5}$$

where $\mathbf{F}_n(\mathbf{y}) = (F_n(y_1), \ldots, F_n(y_n))^T$ and $\mathcal{X}$ are all centered, and $F_n(.)$ is the empirical distribution function. We call the solution $(\boldsymbol{\beta}^{(g)})$ of (2.5) the sparse group sufficient dimension reduction estimator (sgSDR). Equation (2.5) is based on the following observation. The proof is similar to that of Proposition 2.1 of Wang et al. (2012) and is hence omitted.

Proposition 1 *Under the linearity condition, and assume that $\boldsymbol{\Sigma}_s$, the marginal covariance matrix of all the significant predictors (denoted by $\mathbf{X}_s$ here for easy of exposition) is invertible, then*

$$\boldsymbol{\Sigma}_s^{-1}\mathrm{Cov}\{\mathbf{X}_s, F(Y)\} = c\boldsymbol{\beta}_s,$$

where $\boldsymbol{\beta}_s$ consists all non-zero coefficients of $\boldsymbol{\beta}$ from (2.4)*, $c \in \mathbb{R}^1$ is a constant, $F(Y)$ is the cumulative distribution function of Y.*

We adopt the SLEP package (Liu et al., 2009) to implement our method. To select the two tuning parameters, λ_1 and λ_2, we employ the commonly used five-fold cross validation.

3 Simulation Studies

In this section, we compare the performance of our method with the sparse group lasso. We considered linear models, nonlinear models and generalized linear models with Gaussian and non-Gaussian errors. We use the average true positive rate (TPR = the ratio of the number of correctly declared active variables to the number of truly active variables); and the average false positive rate (FPR = the ratio of the number of falsely declared

Table 3.1: Linear model I with Gaussian error

	$l=1$		$l=2$		$l=3$	
	TPR	FPR	TPR	FPR	TPR	FPR
sgSDR	0.75	0.13	0.64	0.32	0.58	0.35
SGL	0.75	0.10	0.64	0.31	0.56	0.32

active variables to the total number of truly inactive variables) as evaluation measurements to summarize variable selection results from 100 simulation runs. We used the SLEP package (Liu et al., 2009) and Matlab for all our numerical studies.

Model I: For a fair comparison, we first consider a regular linear model as Simon et al. (2012) discussed in their paper. The predictor $\mathbf{X}$ is generated from $N(0, I_p)$, ϵ is standard normal and independent of $\mathbf{X}$, the univariate response Y is constructed as:

$$Y = \sum_{g=1}^{G} (\boldsymbol{\beta}^{(g)})^T \mathbf{X}^{(g)} + \sigma\epsilon, \tag{3.6}$$

where $G = 10$, σ is set to make the signal to noise ratio as 2. And the coefficients for the first l group are $\boldsymbol{\beta}^{(g)} = (1, 2, 3, 4, 5, 0, \ldots, 0)^T$, for $g = 1, \ldots, l$, with l varying from 1 to 3; and all zeros for the rest of $G - l$ groups. Following Simon et al. (2012), we took $n = 60$, $p = 1500$. Table 3.1 provides the average true positive and false positive rates. As shown in Table 3.1, the performances of sgSDR and SGL are comparable in the sense that the average TPRs and FPRs are very close to each other.

Model II: We now consider a variation of Model I. We take $p = 2000$, $G = 10$, Y is still generated as in (3.6), however, the predictors now are mildly correlated, ϵ follows $Cauchy(1)$ distribution, and $\boldsymbol{\beta}^{(g)} = (-2, 3, 0, \ldots, 0)^T$, for $g = 1, \ldots, l$, with l varying from 1 to 3; and zeros otherwise. Specifically, within each group, $\mathbf{X}^{(g)} = (X_1^{(g)}, \ldots, X_{200}^{(g)})$ are all generated as independent standard normal random variables except $X_3^{(g)}$, which is generated to be correlated with X_1^g and X_2^g by:

$$X_3^{(g)} = \frac{2}{3} X_1^{(g)} + \frac{2}{3} X_2^{(g)} + \frac{1}{3} e_g, \tag{3.7}$$

Table 3.2: Linear model with Cauchy error and correlated predictors

	$l=1$		$l=2$		$l=3$	
	TPR	FPR	TPR	FPR	TPR	FPR
sgSDR	1.00	0.02	0.98	0.04	0.92	0.04
SGL	0.80	0.42	0.71	0.41	0.74	0.42

Table 3.3: Linear model three with Cauchy error

	$l=5$		$l=10$		$l=15$	
	TPR	FPR	TPR	FPR	TPR	FPR
sgSDR	0.98	0.03	0.88	0.04	0.77	0.06
SGL	0.72	0.26	0.62	0.19	0.59	0.28

where e_g follows standard normal distribution. Different version of Model II was considered by Wang et al. (2012).

The simulation results with $n = 60$ from 100 simulation runs are shown on Table 3.2. We can see that our method (sgSDR) is more robust under the Cauchy random error, which tends to yield relatively higher TPR and significantly lower FPR, compared with SGL with respect to variable selections. With $p = 2000$ in our setting, SGL provides a FPR about 40% higher than our method, which means that about 800 more inactivate variables are mistakenly selected as significant variables by SGL.

Model III: In this example, the linear model (3.6) is reconsidered with larger sample size, larger dimension p and more groups, that is, $n = 200$, $p = 5000$ and $G = 50$. The predictors are generated by $N(0, \mathbf{\Sigma})$, where $\mathbf{\Sigma} = \left(.5^{|i-j|}\right)$, $i, j = 1, \ldots, p$. We consider l $(5, 10, 15)$ significant groups, with $\boldsymbol{\beta}^{(g)} = (3, 1.5, 2, \ldots, 0)^T$, $g = 1, \ldots, l$. The results are shown on Table 3.3. Similar conclusions as Model II can be drawn here.

Table 3.4: Nonlinear model with Gaussian and Cauchy error

		$l=1$		$l=2$		$l=3$	
	Method	TPR	FPR	TPR	FPR	TPR	FPR
Gaussian Error	sgSDR	0.99	0.03	0.99	0.02	0.97	0.03
	SGL	0.98	0.87	0.97	0.95	0.96	0.98
Cauchy Error	sgSDR	1.00	0.01	0.98	0.04	0.92	0.02
	SGL	1.00	1.00	1.00	0.99	1.00	1.00

Model IV: We now compare the performances of sgSDR and SGL for nonliner models under the standard normal and Cauchy errors. We consider the following model:

$$Y = \exp\big(\sum_{g=1}^{G} \mathbf{X}^{(g)}\boldsymbol{\beta}^{(g)} + 3\epsilon\big) \tag{3.8}$$

The predictors $\mathbf{X}$ and the coefficients $\boldsymbol{\beta}$ are set up exactly the same as those of Model II, and $\epsilon \sim N(0,1)$. As shown on Table 3.4, our method outperforms SGL with significantly lower FPR and slightly higher TPR. For models with non-nonlinear regression function and Cauchy errors, SGL fails completely, the average FPR for SGL is above 99%, which implies that it mistakenly selected over 1900 inactive predictors as significant ones.

4 A Real Data Analysis

We demonstrate our method by analyzing a microarray gene expression data with glioblastoma patients by Horvath et al. (2006). Glioblastoma is the most common and aggressive malignant brain tumor in humans. Patients with this disease have a median survival time of approximately 15 months from the time of diagnosis despite various treatments such as surgery, radiation and chemotherapy. Consisting of two independent sets of clinical tumor samples of $n = 55$ and $n = 65$, the dataset was obtained by Affymetrix HG-U133A arrays, and processed by the RMA method (Irizarry et al., 2003). As Pan et al. (2010) pointed out, the two datasets were somewhat different from each other, and they only used dataset one in their analysis. Following Pan et al. (2010), we also focus on

the 50 patients with observed survival times from dataset one, and took the log survival time (in days) as the response variable in our analysis and the gene expression profiles as predictors. Our goal is to simultaneously identify significant pathways and genes within those pathways that are strongly associated with the survival time from glioblastoma.

We merged the gene-expression data with the 33 regulatory pathways recorded in the KEGG database. Among the 1668-node of the 33 pathways, 1507 (Entrez ID) out of 22283 genes (Probe ID) are identified on the HG-U133A chip. Following Li and Li (2008), Pan et al. (2010), and Zhu and Li (2011), we only use these 1507 genes in our following analysis. When there are multiple probe set ids corresponding to a single Entrez KEGG id, we took the average expression levels of those probe ids.

We compared our result with Li and Li (2008). As reported on Table 4.1, our pathway selection is similar to that of Li and Li (2008) except for pathway 6, 13, 18, 17 and 27 (Cell cycle, Extracellular matrix-receptor interaction, Gap junction, Complement and coagulation cascades, Type I diabetes mellitus). Among those five pathways, the first three pathways were selected by our method but not by Li and Li (2008), while the latter two were selected by Li and Li (2008) only. As reported in Sun, et al. (2012), the entire tumor growth profile in brain cancer is a collective behavior of cells regulated by the cell cycle pathway (pathway 6). The study result from Phillips laboratory (UCSF) shows that heparan sulfate proteoglycans (HSPGs) in extracellular matrix (pathway 13) can change tumor cell behavior including proliferation, invasion and recruitment of inflammatory cells. Zhu and Li (2011) ranked all the 33 pathways according to their significances, pathway 17 and 27 which were only selected by Li and Li (2008), ranked 30th and 28th respectively, suggesting that they are not very important pathways.

MAPK signaling pathway (pathway 1), Cytokine-cytokine receptor interaction pathway (pathway 3), Neuroactive ligand-receptor interaction pathway (pathway 5), and Complement and coagulation cascades (pathway 18) were ranked as the top 4 significant pathways related to the brain cancer by Zhu and Li (2011) using a nonlinear dimension reduction method. Our pathway selection is consistent with Zhu and Li (2011), since

Table 4.1: Pathway Selections for Glioblastoma Data

Group	Pathway Name	sgSDR	Li and Li
1	MAPK signaling pathway	✓	✓
2	Calcium signaling pathway	✓	✓
3	Cytokine-cytokine receptor interaction	✓	✓
4	Phospatidylinositol signaling system	✓	✓
5	Neuroactive ligand-receptor interaction	✓	✓
6	Cell cycle	✓	
7	Ubiquitin mediated proteolysis	✓	✓
8	Apopttosis	✓	✓
9	Wnt signaling pathway	✓	✓
10	Transforming growth factor-beta signaling pathway	✓	✓
11	Axon guidance	✓	✓
12	Focal adhesion	✓	✓
13	Extracellular matrix-receptor interaction	✓	
14	Cell adhension molecules	✓	✓
15	Adherens junction	✓	✓
16	Tight junction	✓	✓
17	Gap junction		✓
18	Complement and coagulation cascades	✓	
19	Toll-like receptor signaling pathway	✓	✓
20	Jak-STAT signaling pathway	✓	✓
21	Natural killer cell mediated cytotoxicity	✓	✓
22	Circadian rhythm		
23	Regulation of actin cytotoxicity	✓	✓
24	Insulin signaling pathway	✓	✓
25	Adipocytokine signaling pathway	✓	✓
26	Type II diabetes mellitus	✓	✓
27	Type I diabetes mellitus		✓
28	Alzheimer's disease		
29	Prion diseases		
30	Cocaine addition		
31	Unknown		
32	Unknown		
33	Unknown		

all these 4 pathways are selected by sgSDR. For the within pathway gene selection, our method selected 85 unique genes. Among them, 10 genes are the same as that of Li and Li (2008), i.e., MAP3K7, CX3CL1, SYNJ2, UBE2E1, SMURF2, CLDN6, IRF3, IL21R, PCK1, FOXO1A. And FOXO1A was also identified by Pan et al. (2010) as one of the significant transcription factors associated with glioblastoma.

5 Conclusions and Discussion

We propose a method called sgSDR within the framework of sufficient dimension reduction which could conduct group and within group variable selection simultaneously. Our method is comparable to the sparse group lasso (Friedman et al., 2010; Simon et al., 2012) for the linear models, and outperform it when the regression function is nonlinear. Also, our method is robust to the error distributions. A glioblastoma data is used to illustrate the applications of our method to the gene pathway analysis. The consistency of our group and variable selections deserves further investigation.

References

[1] Alon, U., Barkai, N., Notterman, D., Gish, K., Mack, S. and Levine, J. (1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *PNAS*, **96**, 6745–6750.

[2] Bakin, S. (1999). Adaptive regression and model selection in data mining problems. *PhD Thesis*, Australian National University, Canberra.

[3] Bondell, H. D. and Li, L. (2009). Shrinkage inverse regression estimation for model-free variable selection. *Journal of the Royal Statistical Society, Ser. B*, **71**, 287–299.

[4] Cook, R. D. (1998). *Regression Graphics.* Wiley, New York.

[5] Cook, R. D. and Weisberg, S. (1991). Discussion of "Sliced inverse regression for dimension reduction" by Li. *Journal of the American Statistical Association*, **86**, 328–332.

[6] Dudoit, S., Fridyland J. F. and Speed, T. P. (2002). Comparison of discrimination methods for tumor classification based on microarray data. *Journal of the American Statistical Association*, **97**, 77–87.

[7] Friedman, J., Hastie, T. and Tibshirani, R. (2010). A note on the group lasso and a sparse group lasso. *Technical Report*, Statistics Department, Stanford University.

[8] Hall, P. and Li, K. C. (1993). On almost linearity of low dimensional projections from high dimensional data. *The Annals of Statistics*, **21**, 867–889.

[9] Horvath, et al. (2006). Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. *PNAS*, **103**, 17402–17407.

[10] Huang, J., Ma, S., Xie, H. and Zhang, C.-H. (2009). A group bridge approach for variable selection. Biometrika *Biometrika*, **96**, 339–355.

[11] Irizarry, et al. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, **4**, 249–264.

[12] Karp, P. D., Ouzounis, C. A. and others. (2005). Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Research*, **19**, 6083–6089.

[13] Li, B. and Wang, S.(2007). On directional regression for dimension reduction. *Journal of the American Statistical Association*, **102**, 997–1008.

[14] Li, C. and Li, H. (2008). Network-constrained regularization and variable selection for analysis of genomic data. *Bioinformatics*, **24**, 1175–1182.

[15] Li, K. C.(1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, **86**, 316–327.

[16] Li, L. (2007). Sparse sufficient dimension reduction. *Biometrika*, **94**, 603–613.

[17] Li, L., li, B. and Zhu, L.X. (2010). Groupwise dimension reduction. *Journal of the American Statistical Association*, **105**, 1188–1201.

[18] Li, L. and Nachtsheim, C. (2006). Sparse sliced inverse regression *Technometrics*, **48**, 503–510.

[19] Li, L. and Yin, X. (2008). Sliced inverse regression with regularizations. *Biometrics*, **64**, 124–131.

[20] Liu, J., Ji, S. and Ye, J. (2009). SLEP: sparse learning with efficient projections. Arizona State University, 2009.

[21] Ma, S. and Kosorok, M. R. (2009). Identification of differential gene pathways with principal component analysis. *Bioinformatics*, **25**, 882–889.

[22] Manoli, T., Gretz, N. and others. (2006). Group testing for pathway analysis improves comparability of different microarray datasets. *Bioinformatics*, **22**, 2500–2506.

[23] Matthews, L., Gopinath, G., Gillesphie, M. and others. (2008). Reactome knowledgebases of biological pathways and processes. *Nucleic Acids Research*, **37**, 619–622.

[24] Meier, L., van de Geer, S. and Bühlmann. (2008). The group lasso for logistic regression. *Journal of the Royal Statistical Society, Ser. B*, **70**, 53–71.

[25] Nguyen, D. and Rocke, D. M. (2002). Partial least squares proportional hazard regression for application to DNA microarray data. B *Bioinformatics*, **18**, 1625–1632.

[26] Ni, L., Cook, R. D. and Tsai, C. (2005). A note on shrinkage sliced inverse regression. *Biometrika*, **92**, 242–247.

[27] Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H. and Kanehisa, M. (1999). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, **27**, 29–34.

[28] Pan, W., Xie, B. and Shen, X. (2010). Incorporating predictor network in penalized regression with application to microarray data. *Biometrics*, **66**, 474–484.

[29] Rosenwald A, Wright G and others. (2003). The proliferation gene expression signature is a quantitative integrator of oncogenic events that predicts survival in mantle cell lymphoma. *Cancer Cell*, **3**, 185–197.

[30] Simon, N., Friedman, J., Hastie, T., and Tibshirani, R. (2012). The sparse group lasso. *Journal of Computational and Graphical Statistics*, in press.

[31] Sun, X., Zhang, L., and others. (2012). Multi-scale agent-based brain cancer modeling and prediction of TKI treatment response: Incorporating EGFR signaling pathway and angiogenesis. *BMC Bioinformatics*, **13**, 218.

[32] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Ser. B*, **58**, 267–288.

[33] Wang, K., Li, M. and Bucan, M. (2007). Pathway-based approaches for analysis of genomewide association studies. *American Journal of Human Genetics*, **81**, 1278–1283.

[34] Wang, T., Xu, P. and Zhu, L.-X. (2012). Non-convex penalized estimation in high-dimensional models with single-index structure. *Journal of Multivariate Analysis*, **109**, 221–235.

[35] Wei, P. and Pan, W. (2008). Incorporating gene networks into statistical tests for genomic data via a spatially correlated mixture model. *Bioinformatics*, **24**, 404–411.

[36] Wen, X. and Cook, R. D. (2007). Optimal sufficient dimension reduction in regressions with categorical predictors. *Journal of Statistical Planning and Inference*, **137**, 1931–1978.

[37] Xia et al. (2002). An adaptive estimation of dimension reduction space. *Journal of the Royal Statistical Society, Ser. B*, **64**, 363–410

[38] Yin, X., Li, B., and Cook, R. D. (2008). Successive direction extraction for estimating the central subspace in a multiple-index regression. *Journal of Multivariate Analysis*, **99**, 1733–1757.

[39] Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society, Ser. B*, **68**, 49–67.

[40] Zhu, H. and Li, L. (2011). Biological pathway selection through nonlinear dimension reduction. *Biostatistics*, **12**, 429–444.