Journal of Multivariate Analysis 156 (2017) 1–17

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

A statistical framework for pathway and gene identification from integrative analysis

Quefeng Li [a,b,*], Menggang Yu [c], Sijian Wang [c]

[a] Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27517, USA
[b] Statistical and Applied Mathematical Sciences Institute, Research Triangle Park, NC 27709, USA
[c] Department of Biostatistics and Medical Informatics, University of Wisconsin at Madison, Madison, WI 53792, USA

ARTICLE INFO

Article history:
Received 15 June 2016
Available online 21 January 2017

AMS subject classification:
92B15

Keywords:
Gene and pathway
High dimensional analysis
Integrative analysis
Variable selection

ABSTRACT

In the era of big data, integrative analyses that pool data from different sources are now extensively conducted in order to improve performance. Among many interesting applications, genomics research is an area where integrative methods become popular tools to identify prognostic biomarkers for various diseases. In this paper, we propose such a framework for pathway and gene identification. Our method employs a hierarchical decomposition on genes' effects followed by a proper regularization to identify important pathways and genes across multiple studies. Asymptotic theories are provided to show that our method is both pathway and gene selection consistent. More importantly, we explicitly show that pathway selection consistency needs milder statistical conditions than gene selection consistency, as it would allow false positives and negatives at the gene selection level. Finite-sample performance of our method is shown to be superior than other ad hoc methods in various simulation studies. We further apply our method to analyze five cardiovascular disease studies. Our method is intrinsically a general method on group-wise and element-wise selections from integrative analysis, which can have other applications beyond genomic research.

1. Introduction

High-throughput sequencing technology has become a prevalent tool in biological and medical research. A unique characteristic of the sequencing data is that they are high-dimensional-low-sample-size (HDLSS), i.e., the number of biomarkers is substantially greater than the number of participants in the studies. On the other hand, the decisive biomarkers regulating the phenotypes are usually sparse compared with the total number of biomarkers along the whole genome and their effects are often weak; making the results from individual studies unremarkable and hard to be reproduced [19]. For this reason, joint analysis of multiple genomic data has been used widely and proven to be essential for identifying decisive biomarkers. A good example is the discovery of the risk loci for type 2 diabetes [22,30].

There are two types of joint analysis. One is the classical meta-analysis that aggregates summary statistics from individual datasets to obtain an overall score, based on which statistical significance across all studies is assessed. A comprehensive review of meta-analysis and its applications in genomic studies can be found in [24]. The other is the integrative analysis using individual patient data (IPD) from multiple studies [14–16]. In the era of big data, IPD becomes more accessible from

* Corresponding author at: Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27517, USA.
E-mail address: quefeng@email.unc.edu (Q. Li).

http://dx.doi.org/10.1016/j.jmva.2016.12.005
0047-259X/

large genomic consortia such as Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA). The integrative analysis links phenotypes to gene expression via a joint model of multiple studies and employs variable selection techniques to select decisive genes. It explores the relationship between phenotypes and genes in a more direct way and aims to identify genes that can better predict the phenotypes. Compared with the traditional meta-analysis, it is more appealing as the raw data contain more information than the summary statistics.

Groups of genes, commonly referred to as pathways, are often involved in the same biological processes. The pathway information has been well documented in publicly available databases such as KEGG [20] and Gene Ontology [2]. Since pathways play more defined biological roles than individual genes, pathway analysis has been advocated for better comprehension of the biological mechanism of the disease process. [13] provided a review of pathway analysis approaches developed in the recent decade, including popular methods such as Gene Set Enrichment Analysis [23] and Over-Representation Analysis [10,12]. Some recent advance in pathway analysis, e.g., a Bayesian method proposed in [3], further considers within-pathway selections of genes. In addition to pathways, this method identifies a list of genes under selected pathways that are associated with the response. However, all the above methods only focus on the analysis of a single dataset.

To the best of our knowledge, there are no existing methods that conduct pathway and within-pathway selections based on joint analysis of multiple datasets. In this paper, we propose such a new integrative analysis that utilizes a novel decomposition on genes' effects to reflect the relationship between pathways and genes. This decomposition, together with sparsity-induced penalties, enables both pathway and gene selections. A main difficulty of pathway selection is that the effects of pathways are unobserved. We show that once a reasonable identifiability condition is assumed, whether a pathway has nonzero effect is definitive. Hence, the pathway selection is sensible.

Under our framework, the employment of pathway information offers substantial benefit in statistical properties of the proposed method. We *explicitly* show that pathway selection requires much weaker conditions to achieve desired asymptotic properties. The main reason is that correct selection of pathways can allow some false positives and negatives at the gene selection level (see Section 4). In particular, for the pathway selection, we are able to avoid the irrepresentable condition that has been shown to be necessary for variable selection consistency in HDLSS settings, when either a convex or folded-concave penalty is used [6,31]. The reason is that our penalty's subgradient is infinitely large at the origin, leading to the removal of unimportant pathways without the irrepresentable condition. In addition, we show that correct pathway selection requires weaker conditions on the minimal signal strength than correct gene selection.

The rest of the paper is organized as follows. In Section 2, we introduce our method. In Section 3, we discuss the estimation and computation of our method. In Section 4, we provide theoretical results regarding pathway and gene selection consistency, where conditions needed for the two types of consistency are explicitly compared. In Section 5, we discuss how our method deals with the issue of pathway overlapping. In Section 6, we examine the finite-sample performance of our method through simulation studies. Examples of both overlapping and non-overlapping pathways are simulated. In Section 7, our method is applied to an integrative analysis of five cardiovascular studies. The paper is completed with a discussion in Section 8. All technical proofs are relegated to the Appendix.

2. A hierarchical decomposition for pathway selection

We consider M independent studies with n_m subjects in study m. Let y_{mi} be a binary phenotype of the ith subject in the mth study and $\boldsymbol{x}_{mi} = (x_{mi,1}, \ldots, x_{mi,d})^\top$ be the corresponding expression of d genes. Our proposed methods can be applied to data with any type of outcomes. We focus on the binary phenotype in this paper, because our motivational studies (see Section 7) have binary outcomes/phenotypes. It is natural to assume the following logistic regression model:

$$\log \frac{\Pr(y_{mi}=1|\boldsymbol{x}_{mi})}{\Pr(y_{mi}=0|\boldsymbol{x}_{mi})} = \alpha_m + \sum_{k=1}^{K}\sum_{j=1}^{G_k} x_{mi,kj}\beta_{kjm}, \quad k=1,\ldots,K;\ j=1,\ldots,G_k,$$

where α_m is the intercept, β_{kjm} is the effect of the jth gene under the kth pathway in the mth study, and they both are allowed to vary with m. K is the number of pathways and G_k is the number of genes under the kth pathway. We assume pathways do not have overlapping genes until Section 5 in which we then discuss the case of overlapping pathways.

We decompose β_{kjm} into three components as

$$\beta_{kjm} = p_k g_{kj} \zeta_{kjm}. \tag{1}$$

The parameter p_k is the effect of the kth pathway among all studies; g_{kj} is the effect of the jth gene under the kth pathway; ζ_{kjm} is the gene's effect in the mth study. Such a decomposition is inspired by the hierarchical LASSO [32], which targets for group selection within a single dataset. In decomposition (1), the exact values of p_k, g_{kj} and ζ_{kjm} are not identifiable (since $cp_k \cdot g_{kj}\zeta_{kjm}/c = p_k g_{kj}\zeta_{kjm}$ for a nonzero constant c). However, as will be shown in Proposition 1, it is definitive whether they equal to zero once a reasonable identifiability condition is assumed.

A natural definition of an important pathway across multiple studies is that at least one of its genes has nonzero effect in at least one study. More explicitly, the set of important pathways is defined as

$$\mathcal{P} = \{k:\ \text{there exists at least one } (j, m) \text{ such that } \beta^*_{kjm} \neq 0\}, \tag{2}$$

where β^*_{kjm} denotes the true effect.

Next, we show that this definition is reasonable as it essentially requires an important pathway to have nonzero pathway effect, i.e., $p_k^* \neq 0$. In fact, the true effect β_{kjm}^* can always be decomposed hierarchically as $\beta_{kjm}^* = p_k^* g_{kj}^* \zeta_{kjm}^*$, where $(p_k^*, g_{kj}^*, \zeta_{kjm}^*)$ is a counterpart of $(p_k, g_{kj}, \zeta_{kjm})$. One possible decomposition is

$$p_k^* = \sum_{j=1}^{G_k}\sum_{m=1}^{M}|\beta_{kjm}^*|, \qquad g_{kj}^* = \Big(\sum_{m=1}^{M}|\beta_{kjm}^*|\Big)/\Big(\sum_{j=1}^{G_k}\sum_{m=1}^{M}|\beta_{kjm}^*|\Big), \qquad \zeta_{kjm}^* = \beta_{kjm}^*/\Big(\sum_{m=1}^{M}|\beta_{kjm}^*|\Big),$$

where 0/0 is defined as 1. Even though the decomposition is not unique and therefore the exact values of $(p_k^*, g_{kj}^*, \zeta_{kjm}^*)$ are not definitive, we show that $I(p_k^* \neq 0)$ is definitive once a reasonable identifiability condition is imposed. In this sense, we are clear whether a pathway has a nonzero pathway effect. Moreover, we show that $\mathcal{P}$ defined in (2) is the same as the support of $\boldsymbol{p}^*$ (the set of nonzero elements of $\boldsymbol{p}^*$), where $\boldsymbol{p}^* = (p_1^*, \ldots, p_K^*)^\top$. In other words, we require an important pathway to have nonzero pathway effect.

Proposition 1. *It holds that $\mathcal{P} = \{k : p_k^* \neq 0\}$, under the following condition*
(C1) *For any fixed (k, j), if $\beta_{kjm}^* = 0$ for all $m = 1, \ldots, M$, then $g_{kj}^* = 0$. In addition, for any fixed k, if $\beta_{kjm}^* = 0$ for all $j = 1, \ldots, G_k$ and $m = 1, \ldots, M$, then $p_k^* = 0$.*

Condition (C1) requires that if a gene's effects are 0 in every study, its total gene effect should be 0. In addition, if all its member genes' effects are 0 in every study, a pathway should have zero effect. Without this condition, p_k^* can be any number if $\beta_{kjm}^* = 0$ for all $j \in \{1, \ldots, G_k\}$ and $m \in \{1, \ldots, M\}$, as long as $g_{kj}^* = 0$ and $\zeta_{kjm}^* = 0$. In other words, a pathway's effect is still undefined even though all its genes have zero effects in all studies. Hence, Condition (C1) is imposed to avoid the identifiability issue under the above circumstance.

3. Estimation and computation

Let $\ell_m(\alpha_m, \boldsymbol{p}, \boldsymbol{g}, \boldsymbol{\zeta}_m)$ be the log-likelihood of the mth study (divided by n_m) such that

$$\ell_m(\alpha_m, \boldsymbol{p}, \boldsymbol{g}, \boldsymbol{\zeta}_m) = \frac{1}{n_m}\sum_{i=1}^{n_m}\left[y_{mi}\left(\alpha_m + \sum_{k=1}^{K}\sum_{j=1}^{G_k} x_{mi,kj} p_k g_{kj} \zeta_{kjm}\right) - \log\left\{1 + \exp\left(\alpha_m + \sum_{k=1}^{K}\sum_{j=1}^{G_k} x_{mi,kj} p_k g_{kj} \zeta_{kjm}\right)\right\}\right],$$

where $\boldsymbol{p} = (p_1, \ldots, p_K)^\top$, $\boldsymbol{g} = (g_{11}, \ldots, g_{KG_K})^\top$, and $\boldsymbol{\zeta}_m = (\zeta_{11m}, \ldots, \zeta_{KG_K m})^\top$. We impose L_1 penalties on $\boldsymbol{p}, \boldsymbol{g}, \boldsymbol{\zeta}_m$ and solve the following optimization problem:

$$(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{p}}, \hat{\boldsymbol{g}}, \hat{\boldsymbol{\zeta}}) = \underset{\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{g}, \boldsymbol{\zeta}}{\operatorname{argmin}} \sum_{m=1}^{M} -\ell_m(\alpha_m, \boldsymbol{p}, \boldsymbol{g}, \boldsymbol{\zeta}_m) + \sum_{k=1}^{K}|p_k| + \sum_{k=1}^{K}\sum_{j=1}^{G_k}|g_{kj}| + \chi_n \sum_{k=1}^{K}\sum_{j=1}^{G_k}\sum_{m=1}^{M}|\zeta_{kjm}|, \tag{3}$$

where χ_n is a positive tuning parameter, $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_M)^\top$ and $\hat{\boldsymbol{\zeta}} = (\hat{\boldsymbol{\zeta}}_1^\top, \ldots, \hat{\boldsymbol{\zeta}}_M^\top)^\top$. The coefficient β_{kjm} is then estimated by $\hat{\beta}_{kjm} = \hat{p}_k \hat{g}_{kj} \hat{\zeta}_{kjm}$. Nonzero $\hat{\beta}_{kjm}$'s are considered as selected. Selection properties will be studied in Section 4.

Remark 1. In problem (3), the tuning parameters in front of penalties on $\boldsymbol{p}$ and $\boldsymbol{g}$ can always be normalized to be 1, due to the fact that $cp_k \cdot dg_{kj} \cdot (cd)^{-1}\zeta_{kjm} = p_k g_{kj} \zeta_{kjm}$, for nonzero constants c and d.

Similar as (2), we define pathways selected by our method as

$$\hat{\mathcal{P}} = \{k : \text{ there exists at least one } (j, m) \text{ such that } \hat{\beta}_{kjm} \neq 0\}. \tag{4}$$

In Proposition 2, we show that $\hat{\mathcal{P}}$ is the same as the support of $\hat{\boldsymbol{p}}$. Theorem 1 in Section 4 shows that asymptotically $\hat{\mathcal{P}} = \mathcal{P}$ using definitions in (2) and (4). Therefore, it implies that $\{k : \hat{p}_k \neq 0\} = \{k : p_k^* \neq 0\}$. In other words, the solution $\hat{\boldsymbol{p}}$ given by our method can asymptotically recover the support of $\boldsymbol{p}^*$.

Proposition 2. $\hat{\mathcal{P}} = \{k : \hat{p}_k \neq 0\}$.

To solve (3), we propose fitting $\boldsymbol{p}, \boldsymbol{g}, \boldsymbol{\zeta}$ and $\boldsymbol{\alpha}$ iteratively. We fix three parameters and solve for the other and iterate between the steps until the algorithm converges. At each step, (3) is a convex problem, which can be efficiently solved by the coordinate descent algorithm [8,29]. The value of the objective function decreases after each iteration, hence the convergence is guaranteed. Let $\boldsymbol{X}_m \in \mathcal{R}^{n_m \times d}$ be the matrix containing data from the mth study. Denote $\boldsymbol{H} \in \mathcal{R}^{d \times K}$ as the indicator matrix of the membership of genes in different pathways such that its (j, k)th element $h_{jk} = 1$ if the jth gene belongs to the kth pathway and $h_{jk} = 0$ otherwise. Let $\hat{\boldsymbol{\alpha}}^{(s)}$, $\hat{\boldsymbol{p}}^{(s)}$, $\hat{\boldsymbol{g}}^{(s)}$, and $\hat{\boldsymbol{\zeta}}^{(s)}$ be the solutions of $\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{g}$ and $\boldsymbol{\zeta}$ at the sth step. We elaborate the algorithm as follows.

The algorithm of solving (3):

Step 1. For each dataset, standardize columns to have zero mean and unit variance. Initialize $\hat{\boldsymbol{g}}^{(0)}$ and $\boldsymbol{H}\hat{\boldsymbol{p}}^{(0)}$, e.g., let $\hat{\boldsymbol{g}}^{(0)} = \mathbf{1}$ and $\boldsymbol{H}\hat{\boldsymbol{p}}^{(0)} = \mathbf{1}$.

Step 2. In the sth iteration, let $\tilde{\boldsymbol{X}}_m = \boldsymbol{X}_m \cdot \hat{\boldsymbol{g}}^{(s-1)} \cdot \boldsymbol{H}\hat{\boldsymbol{p}}^{(s-1)}$, where $\cdot$ represents product by columns. Denote

$$\ell(\boldsymbol{\zeta}) = \sum_{m=1}^{M} \frac{1}{n_m} \sum_{i=1}^{n_m} \left[y_{mi} \left(\alpha_m^{(s-1)} + \sum_{k=1}^{K} \sum_{j=1}^{G_k} \tilde{x}_{mi,kj} \zeta_{kjm} \right) - \log \left\{ 1 + \exp\left(\alpha_m^{(s-1)} + \sum_{k=1}^{K} \sum_{j=1}^{G_k} \tilde{x}_{mi,kj} \zeta_{kjm} \right) \right\} \right].$$

Solve

$$\hat{\boldsymbol{\zeta}}^{(s)} = \underset{\zeta}{\operatorname{argmin}} -\ell(\boldsymbol{\zeta}) + \chi_n \sum_{k=1}^{K} \sum_{j=1}^{G_k} \sum_{m=1}^{M} |\zeta_{kjm}|.$$

Step 3. Update $\tilde{\boldsymbol{X}}_m$ by letting $\tilde{\boldsymbol{X}}_m = \boldsymbol{X}_m \cdot \hat{\boldsymbol{\zeta}}_m^{(s)} \cdot \boldsymbol{H}\hat{\boldsymbol{p}}^{(s-1)}$, where $\hat{\boldsymbol{\zeta}}_m^{(s)}$ is the subvector of $\hat{\boldsymbol{\zeta}}^{(s)}$ for the solutions in the mth dataset. Let

$$\ell(\boldsymbol{g}) = \sum_{m=1}^{M} \frac{1}{n_m} \sum_{i=1}^{n_m} \left[y_{mi} \left(\alpha_m^{(s-1)} + \sum_{k=1}^{K} \sum_{j=1}^{G_k} \tilde{x}_{mi,kj} g_{kj} \right) - \log \left\{ 1 + \exp\left(\alpha_m^{(s-1)} + \sum_{k=1}^{K} \sum_{j=1}^{G_k} \tilde{x}_{mi,kj} g_{kj} \right) \right\} \right].$$

Solve

$$\hat{\boldsymbol{g}}^{(s)} = \underset{\boldsymbol{g}}{\operatorname{argmin}} -\ell(\boldsymbol{g}) + \sum_{k=1}^{K} \sum_{j=1}^{G_k} |g_{kj}|.$$

Step 4. Let $\boldsymbol{Z}_m = (\boldsymbol{X}_m \cdot \hat{\boldsymbol{g}}^{(s)} \cdot \hat{\boldsymbol{\zeta}}_m^{(s)})\boldsymbol{H}$ and

$$\ell(\boldsymbol{p}) = \sum_{m=1}^{M} \frac{1}{n_m} \sum_{i=1}^{n_m} \left[y_{mi} \left(\alpha_m^{(s-1)} + \sum_{k=1}^{K} z_{mk} p_k \right) - \log \left\{ 1 + \exp\left(\alpha_m^{(s-1)} + \sum_{k=1}^{K} z_{mk} p_k \right) \right\} \right],$$

where z_{mk} is the (m, k)th element of $\boldsymbol{Z}_m$. Solve

$$\hat{\boldsymbol{p}}^{(s)} = \underset{\boldsymbol{p}}{\operatorname{argmin}} -\ell(\boldsymbol{p}) + \sum_{k=1}^{K} |p_k|.$$

Step 5. Update $\tilde{\boldsymbol{X}}_m$ by letting $\tilde{\boldsymbol{X}}_m = \boldsymbol{X}_m \cdot \boldsymbol{H}\hat{\boldsymbol{p}}^{(s)} \cdot \hat{\boldsymbol{g}}^{(s)} \cdot \hat{\boldsymbol{\zeta}}_m^{(s)}$. Let

$$\ell(\boldsymbol{\alpha}) = \sum_{m=1}^{M} \frac{1}{n_m} \sum_{i=1}^{n_m} \left[y_{mi} \left(\alpha_m + \sum_{t=1}^{d} \tilde{x}_{mi,t} \right) - \log \left\{ 1 + \exp\left(\alpha_m + \sum_{t=1}^{d} \tilde{x}_{mi,t} \right) \right\} \right].$$

Solve $\hat{\boldsymbol{\alpha}}^{(s)} = \operatorname{argmin}_{\boldsymbol{\alpha}} \ \ell(\boldsymbol{\alpha})$.

Step 6. Let $\hat{\beta}_{kjm}^{(s)} = \hat{p}_k^{(s)} \hat{g}_{kj}^{(s)} \hat{\zeta}_{kjm}^{(s)}$. Return to step 2 and iterate, unless the difference between the values of the objective functions in (3) at two consecutive steps is less than some predefined threshold.

Next, we show that (3) has an equivalent form in terms of the original coefficient β_{kjm}.

Lemma 1. *The problem (3) is equivalent to*

$$(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) = \underset{\boldsymbol{\alpha}, \boldsymbol{\beta}}{\operatorname{argmin}} \sum_{m=1}^{M} -\ell_m(\alpha_m, \boldsymbol{\beta}_m) + 3\chi_n^{1/6} \sum_{k=1}^{K} \left\{ \sum_{j=1}^{G_k} \left(\sum_{m=1}^{M} |\beta_{kjm}| \right)^{1/2} \right\}^{1/3}. \tag{5}$$

The equivalent form (5) describes our method in a different perspective. It shows that our method penalizes the negative log-likelihood with different penalties at different levels of the coefficients. Regarding $\sum_{m=1}^{M} |\beta_{kjm}|$ as the overall effect of the jth gene under the kth pathway across M studies, it penalizes the gene's overall effect with an $L_{1/2}$-penalty. At the pathway level, an $L_{1/3}$-penalty is imposed. Due to the singularities of $L_{1/2}$ and $L_{1/3}$ penalties at the origin point, our method can select pathways as well as its important members. In addition, the L_1-penalty on the genes' effects in individual studies enables across-study selection for each gene and adapts to the heterogeneity among the studies. Due to the equivalence, (5) will be used to establish the theoretical results in Section 4.

It is easy to see that (5) is a non-convex problem thus hard to solve directly. However, due to the equivalence, it can be transferred into a sequence of convex optimization problems by using (3). This is another benefit brought by the hierarchical decomposition.

4. Theoretical properties

We present two results regarding the pathway and gene selection accuracy of our method. The results are considered for fixed M, but d and K can diverge with n, as long as $\log d = o(n)$. To simplify the presentation, we assume each dataset has n observations and the intercept terms $\{\alpha^*_{0m}\}_{m=1}^M$ are all 0.

We begin with introducing some notation. For any vector $\boldsymbol{a} \in \mathcal{R}^d$, let $\|\boldsymbol{a}\|_1 = \sum_{j=1}^d |a_j|$ be the L_1-norm of $\boldsymbol{a}$ and $\|\boldsymbol{a}\|_\infty = \max_j |a_j|$ be the sup-norm of $\boldsymbol{a}$. Define the support of $\boldsymbol{a}$ as $\{j : a_j \neq 0\}$. For a symmetric matrix $\boldsymbol{A} \in \mathcal{R}^{d\times d}$, let $\|\boldsymbol{A}\|_\infty = \max_i \sum_{j=1}^d |a_{ij}|$ be the matrix sup-norm of $\boldsymbol{A}$. For two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, denote $\boldsymbol{A} \circ \boldsymbol{B}$ as the Hadamard (componentwise) product of $\boldsymbol{A}$ and $\boldsymbol{B}$. In addition, denote $\boldsymbol{a}_S$ as the sub-vector of $\boldsymbol{a}$ with indices in S and $\boldsymbol{A}_S$ as columns of $\boldsymbol{A}$ with indices in S. In particular, we denote $kj\cdot$ as the set of $\{kj1, \ldots, kjM\}$. For a set S, we denote $|S|$ as its cardinality.

4.1. Pathway selection consistency

According to (2), to ensure that an important pathway is selected, it suffices to select one of its members in at least one study. In fact, we show that the member with the *largest* overall effect can be selected with large probability, where the overall effect of a gene is defined as the L_1-norm of its effects in all studies. To this end, it only requires the *largest* overall effect not to be small, which is much weaker than the typical assumption on the minimal signal strength. On the other hand, as will be seen in the proof of Theorem 1, the unique property of the penalty term in (5) (its subgradient is infinitely large at the origin point) ensures that unimportant pathways are ruled out without additional assumptions. This is not like the L_1 or other folded-concave penalties, where irrepresentable conditions are needed [6,31].

To present the result, let $K^* = M|\mathcal{P}|$, the number of important pathways multiplied by M, and

$$J = \left\{(k, j, m) : k \in \mathcal{P}, j = \operatorname*{argmax}_{j\in G_k} \|\boldsymbol{\beta}^*_{kj\cdot}\|_1, m = 1, \ldots, M\right\},$$

the collection of $|\mathcal{P}|$ genes' effects in M studies, whose overall effect is the largest in their own pathways. Denote $B = \min_{k\in\mathcal{P}, j:kj\cdot\in J} \|\boldsymbol{\beta}^*_{kj\cdot}\|_1/2$, the minimal overall effects of these $|\mathcal{P}|$ genes. We write B and K^* in terms of orders of n as $B \asymp n^{-\alpha_B}$ and $K^* \asymp n^{-\alpha_{K^*}}$.

We present (5) in a matrix form as

$$\operatorname*{argmin}_{\boldsymbol{\beta}\in\mathcal{R}^{Md}} \frac{1}{n}\{-\boldsymbol{Y}^\top \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}^\top \boldsymbol{f}(\boldsymbol{X}\boldsymbol{\beta})\} + \lambda_n \rho(\boldsymbol{\beta}), \tag{6}$$

where $\boldsymbol{X} \in \mathcal{R}^{Mn\times Md}$ is the block-diagonal design matrix, whose mth block contains data from the mth study, $\boldsymbol{Y} = (y_{11}, y_{12}, \ldots, y_{Mn})^\top$, $\boldsymbol{f}(\boldsymbol{\theta}) = (f(\theta_1), \ldots, f(\theta_{Mn}))^\top$ with $f(\theta_i) = \log\{1 + \exp(\theta_i)\}$, $\boldsymbol{\mu}(\boldsymbol{\theta}) = (f'(\theta_1), \ldots, f'(\theta_{Mn}))^\top$, $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \operatorname{diag}(f''(\theta_1), \ldots, f''(\theta_{Mn}))$, $\boldsymbol{e}$ is a vector with all elements equal to 1, the penalty term $\rho(\boldsymbol{\beta}) = \sum_{k=1}^K \{\sum_{j=1}^{G_k} \|\boldsymbol{\beta}_{kj\cdot}\|_1^{1/2}\}^{1/3}$ and $\lambda_n = 3\chi_n^{1/6}$.

Theorem 1. *Assume the following conditions hold.*
(C2) $0 < \alpha_B < 1/2$;
(C3) $0 < \alpha_{K^*} < 1/2$;
(C4) $\|\{\boldsymbol{X}_J^\top \boldsymbol{\Sigma}(\boldsymbol{X}\boldsymbol{\beta}^*)\boldsymbol{X}_J\}^{-1}\|_\infty = O(n^{-1})$;
(C5) $\max_{\boldsymbol{\delta}\in\mathcal{N}} \max_{k,j,m} \lambda_{\max}\left(\boldsymbol{X}_J^\top \operatorname{diag}\{|\boldsymbol{X}_{kjm}| \circ |\boldsymbol{\mu}''(\boldsymbol{X}_J\boldsymbol{\delta})|\}\boldsymbol{X}_J\right) = O(n)$, *where* $\mathcal{N} = \{\boldsymbol{\delta} \in \mathcal{R}^{K^*} : \|\boldsymbol{\delta} - \boldsymbol{\beta}^*_J\|_\infty \le B\}$.
If we choose λ_n *such that* $\lambda_n = o(B^{5/6}n^{-1/2})$, $\lambda_n\kappa_n = o(\tau_n)$, *where*

$$\kappa_n = \max_{\boldsymbol{\delta}\in\mathcal{N}} (5/36)\|\boldsymbol{\delta}\|_1^{-11/6},$$

and $\tau_n = \min_{\boldsymbol{\delta}\in\mathcal{N}} \lambda_{\min}(n^{-1}\boldsymbol{X}_J^\top \boldsymbol{\Sigma}(\boldsymbol{X}_J\boldsymbol{\delta})\boldsymbol{X}_J)$, *then for sufficiently large n, with probability greater than* $1 - 2K^*/n$, *there exists a solution* $\hat{\boldsymbol{\beta}}$ *to* (5), *such that*
(a) $\hat{\mathcal{P}} = \mathcal{P}$;
(b) $\|\hat{\boldsymbol{\beta}}_J - \boldsymbol{\beta}^*_J\|_\infty \le cn^{-\gamma}$, *where c is a generic positive constant and* γ *is some number that is arbitrarily smaller than* $1/2$.

Remark 2. Theorem 1 implies that our method is asymptotically pathway selection consistent and the corresponding estimation of $\boldsymbol{\beta}^*_J$ is uniformly consistent. In general, conditions needed for pathway selection consistency is weaker than that for gene selection consistency (see Theorem 2). For example, condition (C2) only requires the minimum of the *largest* overall effects under important pathways not to be small, compared with a stronger condition (C6) needed for gene selection consistency. Moreover, Theorem 1 does not need the irrepresentable condition [6,31] to reach pathway selection consistency, while this condition (C9) is needed to achieve gene selection consistency. The intuition is that, pathway selection consistency is easier to be achieved than gene selection consistency. For important pathways, correct pathway selection can tolerate false positive and negative selections of their genes, as long as at least one gene under the important pathway is selected in at least one study. For unimportant pathways, our method is able to remove them without an irrepresentable condition due to the fact that the subgradient of our penalty at the origin is infinitely large.

4.2. Gene selection consistency

If stronger conditions are satisfied, we show that our method is further gene selection consistent in *every study* in the sense that the support of $\hat{\boldsymbol{\beta}}$ is the same as that of $\boldsymbol{\beta}^*$. The genes' effects are often heterogeneous among various studies due to many practical factors, such as various types of biospecimens, inconsistent experimental environments, etc. For this reason, we allow a gene's effect to be zero in some studies but nonzero in others. The nonzero effects are also allowed to vary. The challenge of achieving gene selection consistency is, for genes have zero/nonzero effects in different studies, to identify the studies where the gene has zero effect. The gene selection consistency is stronger than pathway selection consistency, since the former implies the latter.

To present the result, we divide elements of $\boldsymbol{\beta}^*$ into three subsets:

$$I = \{(k,j,m) : \beta^*_{kjm} \neq 0\}; \qquad II = \{(k,j,m) : \beta^*_{kjm} = 0, \|\boldsymbol{\beta}^*_{kj\cdot}\|_1 \neq 0\};$$
$$III = \{(k,j,m) : \beta^*_{kjm} = 0, \|\boldsymbol{\beta}^*_{kj\cdot}\|_1 = 0\}.$$

Denote $s = |I|$, the cardinality of set I and $b = (1/2)\min\{|\beta^*_{kjm}| : (k,j,m) \in I\}$, the minimal nonzero element of $\boldsymbol{\beta}^*$,

$$h_n = \min_{(k,j,m)\in I}\Big(\sum_{j=1}^{G_k}\|\boldsymbol{\beta}^*_{kj\cdot}\|_1^{1/2}\Big)^{2/3}\|\boldsymbol{\beta}^*_{kj\cdot}\|_1^{1/2}, \qquad u_n = \max_{(k,j,m)\in I}\Big(\sum_{j=1}^{G_k}\|\boldsymbol{\beta}^*_{kj\cdot}\|_1^{1/2}\Big)^{2/3}\|\boldsymbol{\beta}^*_{kj\cdot}\|_1^{1/2}.$$

The magnitudes of b, s, d are quantified in the order of n as $b \asymp n^{-\alpha_b}$, $s \asymp n^{\alpha_s}$, and $\log d \asymp n^{\alpha_d}$, where α_s, α_u and α_d are positive numbers.

Theorem 2. *Under the following conditions*
(C6) $0 < \alpha_b < 1/2$;
(C7) $0 < \alpha_s < 1/2$ *and* $0 < \alpha_d < 1$;
(C8) $\|\{\boldsymbol{X}_I^\top \boldsymbol{\Sigma}(\boldsymbol{X}\boldsymbol{\beta}^*)\boldsymbol{X}_I\}^{-1}\|_\infty = O(n^{-1})$;
(C9) $\|\boldsymbol{X}_{II}^\top \boldsymbol{\Sigma}(\boldsymbol{X}\boldsymbol{\beta}^*)\boldsymbol{X}_I\{\boldsymbol{X}_I^\top \boldsymbol{\Sigma}(\boldsymbol{X}\boldsymbol{\beta}^*)\boldsymbol{X}_I\}^{-1}\|_\infty < h_n/(10u_n)$;
(C10) $\max_{\boldsymbol{\delta}\in\mathcal{N}_0}\max_{k,j,m}\lambda_{\max}\big(\boldsymbol{X}_I^\top \mathrm{diag}\{|\boldsymbol{X}_{kjm}| \circ |\boldsymbol{\mu}''(\boldsymbol{X}_I\boldsymbol{\delta})|\}\boldsymbol{X}_I\big) = O(n)$, $\mathcal{N}_0 = \{\boldsymbol{\delta} \in \mathcal{R}^s : \|\boldsymbol{\delta} - \boldsymbol{\beta}^*_I\|_\infty \le b\}$.
If we choose the penalty λ_n *such that* $(n^{-1}\log d)^{1/2} = o(\lambda_n)$, $sn^{-2/3} = o(\lambda_n)$, $\lambda_n = o(h_n n^{-1/2})$, *and* $\lambda_n\kappa_{0n} = o(\tau_{0n})$, *where*

$$\kappa_{0n} = \max_{\boldsymbol{\delta}\in\mathcal{N}_0,(k,j,m)\in I}\frac{1}{18}\Big(\sum_{j=1}^{G_k}\|\boldsymbol{\delta}_{kj\cdot}\|_1^{\frac{1}{2}}\Big)^{-\frac{5}{3}}\|\boldsymbol{\delta}_{kj\cdot}\|_1^{-1} + \frac{1}{12}\Big(\sum_{j=1}^{G_k}\|\boldsymbol{\delta}_{kj\cdot}\|_1^{\frac{1}{2}}\Big)^{-\frac{2}{3}}\|\boldsymbol{\delta}_{kj\cdot}\|_1^{-\frac{3}{2}}$$

and $\tau_{0n} = \min_{\boldsymbol{\delta}\in\mathcal{N}_0}\lambda_{\min}(n^{-1}\boldsymbol{X}_I^\top \boldsymbol{\Sigma}(\boldsymbol{X}_I\boldsymbol{\delta})\boldsymbol{X}_I)$, *then for sufficiently large n, with probability greater than* $1 - 2\{s/n + (Md - s)/d^2\}$, *there exists a solution* $\hat{\boldsymbol{\beta}}$ *to* (5), *such that*

(c) $\hat{\boldsymbol{\beta}}_{II\cup III} = \boldsymbol{0}$;
(d) $\|\hat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}^*_I\|_\infty \le cn^{-\gamma}$, *where c is a generic positive constant and* γ *is some number that is arbitrarily smaller than* 1/2.

5. Pathway overlap

So far, we have assumed that pathways do not have overlapping genes. In practice, pathways can share some common genes and this makes the analysis and interpretation much more challenging than the non-overlap case. In general, one can always classify any pathway into the following three cases according to whether its important member genes are shared with some other pathways. For a pathway that has at least one exclusive important gene (a gene that only belongs to that pathway and its effect is nonzero in at least one study), it is important according to definition in (2). For a pathway whose genes all have zero effects, it is unimportant. For a pathway whose important genes are all shared with other pathways, its importance is unclear. The reason is that the shared genes' contribution in different pathways is undetermined even if its total effect is known. Hence, if a pathway's importance cannot be judged by its exclusive genes, its importance is not always identifiable.

To quantify the distribution of the shared gene's effects in different pathways, we use the following assumption. Suppose $\boldsymbol{X}_{k_1j_1m} = \boldsymbol{X}_{k_2j_2m} = \cdots = \boldsymbol{X}_{k_Tj_Tm}$ for all $1 \le m \le M$, i.e., the gene belongs to T pathways. We still adopt the decomposition (1) and assume

(C11) $g_{k_1j_1} = \cdots = g_{k_Tj_T} = g$ and $\zeta_{k_1j_1m} = \cdots = \zeta_{k'_Tj'_Tm} = \zeta_m$ for all $1 \le m \le M$.

Under condition (C11), this shared gene's effect can be decomposed as $(p_{k_1} + \cdots + p_{k_T})g\zeta_m$. Its contribution in pathway k_ℓ is proportional to the pathway effect p_{k_ℓ}. Once condition (C11) is assumed, our algorithm in Section 2 still works for the overlapping case and gives a pathway effect estimator $\hat{\boldsymbol{p}}$. We select pathways by referring to the nonzero elements of $\hat{\boldsymbol{p}}$. As for gene selection, when a shared gene is selected, our algorithm tells under which pathways the gene is selected (by referring to nonzero elements of $\hat{\boldsymbol{p}}$).

In reality, the overlapping scheme of pathways can be complex. The asymptotic properties are difficult to establish under this setting. We leave it as a topic for future research. Nevertheless, we investigate the empirical performance of our method under this setting by simulation studies, where overlapping pathways for the first two definitive cases are simulated.

6. Simulation studies

We simulate scenarios of non-overlapping and overlapping pathways to inspect our method's finite-sample performance.

6.1. Non-overlapping pathways

We simulate $M = 10$ studies with sample size $n_m = 50$ in each study. The expression of $d = 100$ genes from $K = 20$ pathways are simulated, with 5 genes in each pathway. The gene expression of the ith subject in the mth study $\boldsymbol{x}_{mi}$ is i.i.d. from $\mathcal{N}_{100}(\mathbf{0}, \boldsymbol{I})$. The corresponding phenotype y_{mi} is generated from the logistic model

$$\Pr(y_{mi} = 1|\boldsymbol{x}_{mi}) = \frac{\exp(\boldsymbol{x}_{mi}^\top \boldsymbol{\beta}_m^*)}{1 + \exp(\boldsymbol{x}_{mi}^\top \boldsymbol{\beta}_m^*)},$$

where $\boldsymbol{\beta}_m^* = (\beta_{1m}^*, \beta_{2m}^*, \ldots, \beta_{dm}^*)^\top$.

We let $\beta_{kjm}^* = a_k b_{kj} c_{kjm}$, where a_k is i.i.d. from $N(\nu_k, 0.5^2)$ with $\nu_1 = 8$, $\nu_2 = 8$, $\nu_3 = -4$, $\nu_4 = -4$, $\nu_5 = -8$, and $a_k = 0$ for $k > 5$; b_{kj} is i.i.d. from Bernoulli(π_g); and c_{kjm} is i.i.d. from Bernoulli(π_m). That is, only the first five pathways are important. Each gene under the important pathway has probability π_g to be important. For an important gene, it has probability π_m to be important in the mth study. The probabilities π_g and π_m control the heterogeneity among genes and studies, respectively. We choose two values (0.3 and 0.9) for both π_g and π_m to represent different levels of heterogeneity. Simulations were repeated for 100 runs.

Our method is compared with two ad-hoc selection methods:

Separate Group LASSO (seGLASSO): it runs study-by-study selections by treating genes under each pathway as a group and imposes a Group LASSO penalty therein, i.e., in each study it solves the problem

$$\operatorname*{argmin}_{\boldsymbol{\beta}} -\ell_m(\boldsymbol{\beta}_m) + \lambda_m \sum_{k=1}^{20} \Big(\sum_{j=1}^{5} \beta_{kjm}^2 \Big)^{1/2},$$

where λ_m is a positive tuning parameter.

Stack Group LASSO (stGLASSO): it stacks data in all studies and penalizes the total negative log-likelihood with a Group LASSO penalty, i.e., it solves the problem

$$\operatorname*{argmin}_{\boldsymbol{\beta}} \sum_{m=1}^{10} -\ell_m(\boldsymbol{\beta}_m) + \lambda \sum_{k=1}^{20} \Big(\sum_{j=1}^{5} \sum_{m=1}^{10} \beta_{kjm}^2 \Big)^{1/2},$$

where λ is a positive tuning parameter.

The optimal tuning parameters in the three methods are all selected by minimizing the Bayesian Information Criterion.

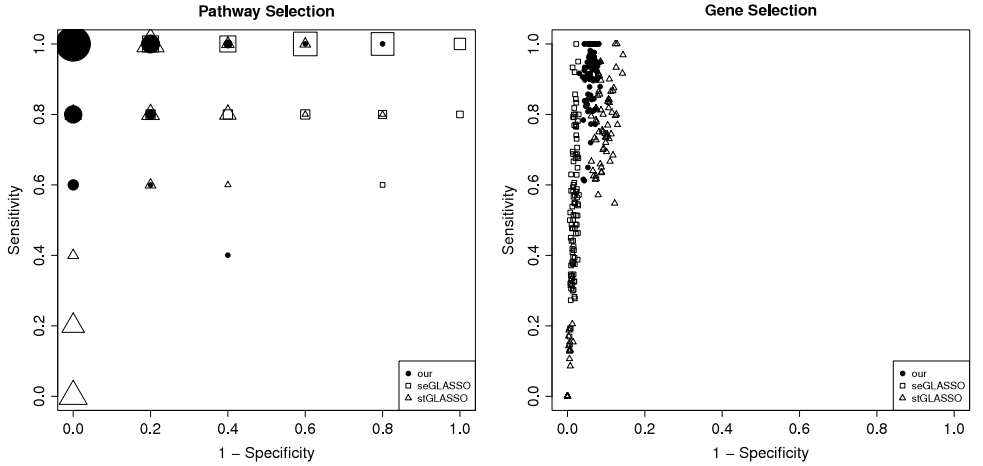
The three competitors' performance is assessed by their selection capability of pathways and genes. Two measurements will be presented: (a) Sensitivity: the proportion of important pathways/genes being selected; (b) Specificity: the proportion of unimportant pathways/genes not being selected. In particular, the set of important pathways is $\{1, \ldots, 5\}$. The set of important genes is defined as $\{(k, j, m) : \beta_{kjm}^* \neq 0\}$.

Fig. 1 gives the boxplots for pathway and gene selection among the 100 simulations. The overall performance is measured by both sensitivity and specificity. Tables 1 and 2 give the means of sensitivity and specificity. These results clearly show that the pathway selection by our method is the best in all settings. The gene selection is also the best when studies are heterogeneous ($\pi_m = 0.3$) and comparable to stGLASSO when studies are homogeneous ($\pi_m = 0.9$). The seGLASSO performs badly as it uses one data at a time and does not borrow strength across studies. The performance of stGLASSO changes drastically in different settings. When π_m is small, it performs badly as it does not take study heterogeneity into account. Even when π_m is large, its performance of pathway selection is still worse than ours. In conclusion, our method performs consistently well for various settings in terms of both pathway selection and gene selection accuracy.

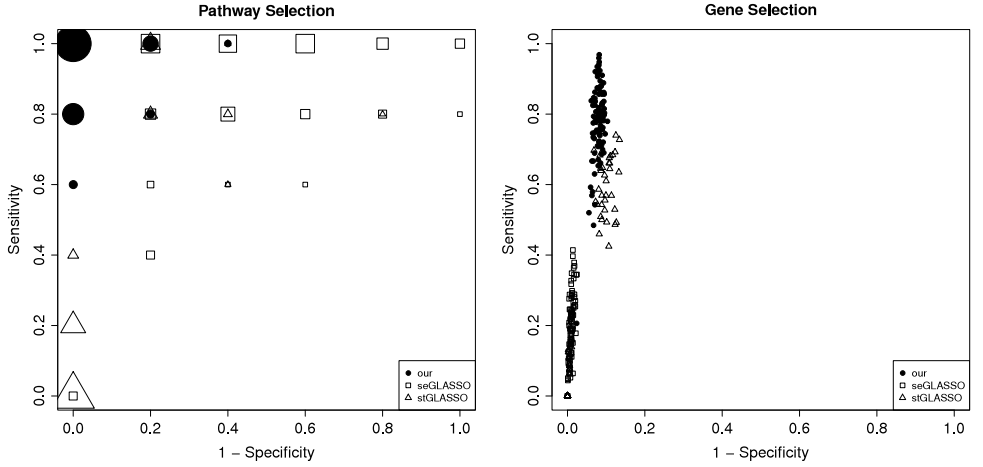
The computational speed of our estimator is comparable to that of seGLASSO and stGLASSO. From the algorithm in Section 3, each iteration of our method solves a convex optimization, similar to seGLASSO and stGLASSO. In the simulations, we noticed that our method usually took very few iterations to convergence.

6.2. Overlapping pathways

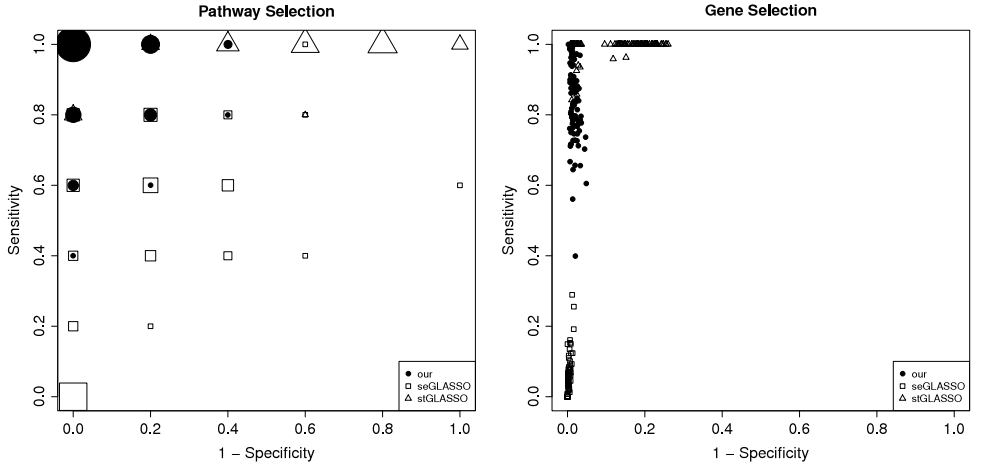
We simulate three more examples for pathways with overlapping genes. We simulate $M = 5$ studies with sample size $n_m = 30$ in each study, the number of pathways $K = 21$ and the number of genes $d = 100$. The gene expression values $\boldsymbol{x}_{mi}$ are generated in the same way as Section 6.1. The true effects of genes are generated by letting $\beta_{kjm}^* = p_k^* g_{kj}^* \zeta_{kjm}^*$ and the following three cases are considered. Table 3 gives the genes' effects for the first four pathways. The unlisted genes have zero effects.



(a) $\pi_d = 0.3$ and $\pi_m = 0.3$.

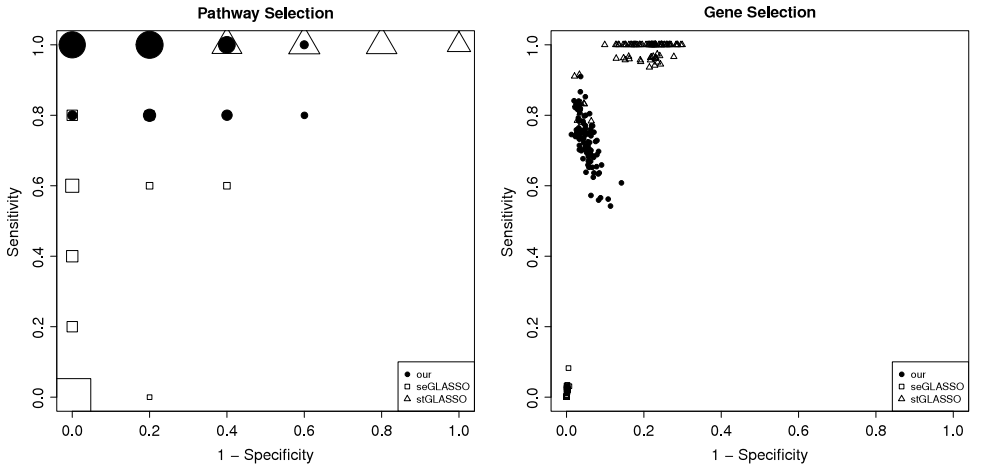


(b) $\pi_g = 0.9$ and $\pi_m = 0.3$.



(c) $\pi_d = 0.3$ and $\pi_m = 0.9$.

Fig. 1. ROC type of scatter plots for pathway and gene selections (the size of a point indicates its frequency among 100 simulations).



(d) $\pi_d = 0.9$ and $\pi_m = 0.9$.

Fig. 1. (*continued*)

Table 1
Pathway selection accuracy by the three methods.

		$\pi_m = 0.3$		$\pi_m = 0.9$	
		Sensitivity	Specificity	Sensitivity	Specificity
$\pi_g = 0.3$	Our	0.93	0.93	0.93	0.93
	seGLASSO	0.96	0.42	0.39	0.86
	stGLASSO	0.61	0.86	0.98	0.56
$\pi_g = 0.9$	Our	0.94	0.96	0.85	0.91
	seGLASSO	0.88	0.60	0.20	0.97
	stGLASSO	0.36	0.94	1.00	0.41

Table 2
Gene selection accuracy by the three methods.

		$\pi_m = 0.3$		$\pi_m = 0.9$	
		Sensitivity	Specificity	Sensitivity	Specificity
$\pi_g = 0.3$	Our	0.92	0.94	0.84	0.98
	seGLASSO	0.55	0.98	0.04	0.99
	stGLASSO	0.52	0.93	0.99	0.86
$\pi_g = 0.9$	Our	0.78	0.92	0.60	0.97
	seGLASSO	0.19	0.99	0.01	0.99
	stGLASSO	0.23	0.96	0.97	0.80

Example 1. $p_k^* = 1$ for $1 \le k \le 4$ and $p_k^* = 0$ otherwise; $g_{1\cdot}^* = (5, 5, 2.5, 2.5, 0)^\top$, $g_{2\cdot}^* = (2.5, 2.5, 5, 2.5, 2.5)^\top$, $g_{3\cdot}^* = (2.5, 2.5, 5, 5, 5)^\top$, $g_{4\cdot}^* = (-8, -8, -8, -8, -8)^\top$, and $g_{k\cdot}^* = (0, 0, 0, 0, 0)^\top$ otherwise; $\zeta_{kjm}^* = 1$ for $1 \le k \le 4$, $1 \le j \le 5$, $1 \le m \le 5$ and $\zeta_{kjm}^* = 0$ otherwise.

Example 2. $p_k^* = 1$ for $k = 1, 2, 4$ and $p_k^* = 0$ otherwise; $g_{1\cdot}^* = (5, 5, 2.5, 2.5, 0)^\top$, $g_{2\cdot}^* = (2.5, 2.5, 5, 0, 0)^\top$, $g_{4\cdot}^* = (-8, -8, -8, -8, -8)^\top$ and $g_{k\cdot}^* = (0, 0, 0, 0, 0)^\top$ otherwise; $\zeta_{kjm}^* = 1$ for $1 \le k \le 4$, $1 \le j \le 5$, $1 \le m \le 5$ and $\zeta_{kjm}^* = 0$ otherwise.

Example 3. $p_k^* = 1$ for $1 \le k \le 4$ and $p_k^* = 0$ otherwise; $g_{1\cdot}^* = (5, 5, 2, 2, 0)^\top$, $g_{2\cdot}^* = (2, 2, 5, 0, 1)^\top$, $g_{3\cdot}^* = (2, 2, 0, -2, 5, 5)^\top$, $g_{4\cdot}^* = (-1, -1, 0, -8, -8, -8)^\top$ and $g_{k\cdot}^* = (0, 0, 0, 0, 0)^\top$ otherwise; $\zeta_{kjm}^* = 1$ for $1 \le k \le 4$, $1 \le j \le 5$, $1 \le m \le 5$ and $\zeta_{kjm}^* = 0$ otherwise.

In Example 1, the first four pathways are important and three of them share some common genes. In Example 2, only pathways 1, 2, and 4 are important. Even though pathway 3 is unimportant, it shares two genes (gene 6 and 7) with an important pathway (pathway 2). We design such an example to see how well our method could remove an unimportant pathway, which overlaps with an important one. In Example 3, genes 3 and 4 are shared by all the four important pathways. Fig. 2 gives the proportion of each pathway being selected among 100 simulations. Table 4 gives the means of sensitivity/specificity of pathway and gene selection.

Fig. 2 shows that our method can correctly distinguish important and unimportant pathways. For the challenging case of Example 2, our method correctly selects pathway 2 and removes pathway 3, even though they share two common genes.

Table 3
The effects of genes in the first four pathways for the three simulated examples.

Example 1	Genes														
Pathway 1	5	5	2.5	2.5	0	–	–	–	–	–	–	–	–	–	–
Pathway 2	–	–	2.5	2.5	5	2.5	2.5	–	–	–	–	–	–	–	–
Pathway 3	–	–	–	–	–	2.5	2.5	5	5	5	–	–	–	–	–
Pathway 4	–	–	–	–	–	–	–	–	–	–	−8	−8	−8	−8	−8
Overall effect	5	5	5	5	5	5	5	5	5	5	−8	−8	−8	−8	−8
Example 2	Genes														
Pathway 1	5	5	2.5	2.5	0	–	–	–	–	–	–	–	–	–	–
Pathway 2	–	–	2.5	2.5	5	0	0	–	–	–	–	–	–	–	–
Pathway 3	–	–	–	–	–	0	0	0	0	0	–	–	–	–	–
Pathway 4	–	–	–	–	–	–	–	–	–	–	−8	−8	−8	−8	−8
Overall effect	5	5	5	5	5	0	0	0	0	0	−8	−8	−8	−8	−8
Example 3	Genes														
Pathway 1	5	5	2	2	0	–	–	–	–	–	–	–	–	–	–
Pathway 2	–	–	2	2	5	0	–	1	–	–	–	–	–	–	–
Pathway 3	–	–	2	2	–	–	0	–	−2	5	5	–	–	–	–
Pathway 4	–	–	−1	−1	–	–	–	–	–	–	–	0	−8	−8	−8
Overall effect	5	5	5	5	5	0	0	1	−2	5	5	0	−8	−8	−8

Table 4
Performance of pathway/gene selection of our method for overlapping pathways.

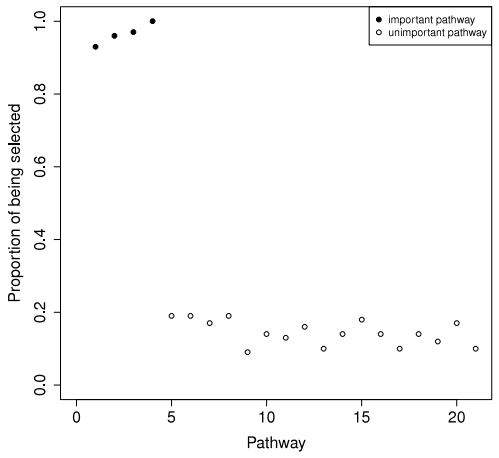
Pathway selection		Gene selection	
Sensitivity	Specificity	Sensitivity	Specificity
Example 1			
0.97	0.86	0.41	0.98
Example 2			
0.97	0.85	0.53	0.98
Example 3			
0.98	0.85	0.44	0.98

For the other challenging case of Example 3, our method correctly selects the first four pathways, even though two genes are shared by all of them.

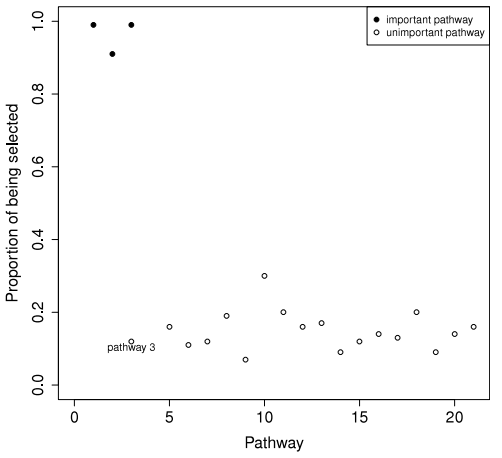
7. An integrative analysis of five cardiovascular disease studies

For further illustration, we applied our method to an integrative analysis of five cardiovascular disease (CVD) studies. These studies were aimed to identify biomarkers that are associated with immune response in the development of atherosclerosis. The phenotypes in these studies are binary. The case groups are patients showing certain atherosclerotic syndrome, e.g., having had ischemic strokes. The control groups are healthy people. All subjects were sequenced using microarrays. The raw data can be found on Gene Expression Omnibus with accession names "GSE12288", "GSE26561", "GSE20129", "GSE22255", and "GSE28829". Table 5 presents more details of the five studies. However, a careful inspection of their original findings revealed that these studies identified completely different sets of genes. Therefore, the underlying genomic mechanism is still largely unknown. It also indicated that the gene selections from case-by-case studies are hard to be reproduced. This motivated us to incorporate external pathway information and integrate the datasets. We cross-referenced the genes with the pathway information listed on the Kyoto Encyclopedia of Genes and Genomes (KEGG). In total, 4156 genes in 210 pathways from KEGG were involved in the analysis. We applied our method to select pathways as well as their important member genes.

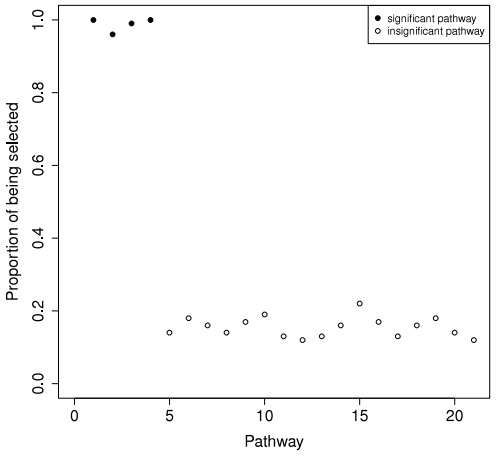
Eleven pathways were identified. Table 6 gives the selected pathways and genes. Table 7 gives the selected genes in each individual study. Among the selected pathways, "Antigen process and presentation", "Hedgehog signaling", "Osteoclast differentiation" and "Phagosome" were known to control key modules for activating human body's immune system [11,21,26,28]. These pathways can certainly be triggered by inflammation. Epidemiological and clinical studies have shown strong and consistent relationships between markers of inflammation and risk of cardiovascular events [27]. Moreover, "Ubiquitin mediated proteolysis", "mRNA surveillance" and "Protein processing in endoplasmic reticulum" play important roles in a broad array of basic cellular processes [4,9]. "Axon guidance" was previously identified to regulate molecules for the angiogenic growth of blood vessels [1]. The "peroxisome proliferator-activated receptor" (PPAR) was shown to be an important regulator of cardiac metabolism [7]. Elevated circulating levels of "cytokines and/or cytokine receptors" was also known to predict adverse outcomes in patients with heart failure [5]. These key pathways were identified by our method.



(a) Example 1.



(b) Example 2.



(c) Example 3.

Fig. 2. Proportion of each pathway being selected among 100 simulations.

In addition, Huntington's disease patients were known to have higher risk of CVD, presumably as a result of accumulation of toxic amyloid-like inclusions [17]. Our selection of "Huntington's disease" pathway indicates that Huntington's disease and CVD can share some common prognostic biomarkers.

Moreover, we used the pathway enrichment score proposed in [18] to assess how our selection of pathways is enriched. In particular, for each selected pathway, we created a 2×2 contingency table of the number of genes selected in and out of that pathway. Then, we tested whether that pathway is over-represented by the Fisher's exact test. The corresponding p-values were used as the enrichment scores, where smaller p-values indicate stronger evidence that the selected pathway is enriched. We calculated pathway enrichment scores in each individual study as well as over all five studies together. Table 6 gives the results of the enrichment scores. It clearly indicates that the selected pathways become more enriched when the integrative method is applied over multiple studies than their appearance in a single study.

In conclusion, our integrative method identified more key biomarkers than the original individual studies. The resulting pathway selections become more enriched. Both results necessitate our proposal of integrating data and pave the way for further studies of the interactions among the identified biomarkers (see Table 8).

8. Discussion

In this article, we have provided a general framework of enhancing the integrative analysis with pathway information, which selects pathways as well as their functioning genes. The incorporation of pathway information not only improves the biological interpretation of the results but also brings statistical advantage. We adopt a hierarchical decomposition on gene effects and show that whether a pathway has nonzero effect is determined once a reasonable identifiability condition is assumed. We show that our method can consistently select both pathways and genes. The pathway selection requires

Table 5
Summaries of datasets in the five cardiovascular studies.

	GSE12288	GSE16561	GSE20129	GSE22255	GSE28829
Total sample size	222	63	119	40	29
Case group size	110	39	48	20	16
Control group size	112	24	71	20	13
Microarray Platform	Affymetrix U133A	Illumina HumanRef8 V3.0	Illumina HumanRef8 V2.0	Affymetrix U133plus2	Affymetrix U133plus2

Table 6
Pathway and gene selections by our method and the p-values of pathway enrichment.

Names of selected pathways	Name of selected genes
Antigen processing and presentation	CANX, HLA-A, HLA-DRA
Axon guidance	SEMA6C
Cytokine–cytokine receptor interaction	BMP2, CXCR3, EDAR, IFNA4, IL21R, INHBA, INHBE
Hedgehog signaling	BMP2, BMP6, CSNK1D
Huntington's disease	ATP5B, DNAI2, NDUFA4L2
mRNA surveillance	NUDT21
Osteoclast differentiation	ACP5, GAB2, RELB, LILRB4
Phagosome	CANX, HLA-DRA
PPAR signaling	GK, RXRG
Protein processing in endoplasmic reticulum	CANX, CAPN1, DERL2, HSPH1
Ubiquitin mediated proteolysis	ANAPC13, ANAPC2, PPIL2, TRIM37, UBE3C

Table 7
Gene selections by our method in each dataset.

GSE12288	ACP5, ANAPC2, ATP5B, BMP2, BMP6, CANX, CAPN1, CSNK1D, CXCR3, DERL2, EDAR, HLA-A, HSPH1, LILRB4, NDUFA4L2, RELB, SEMA6C, TRIM37
GSE16561	ANAPC13, ATP5B, CANX, CSNK1D, DNAI2, RXRG
GSE20219	ACP5, ANAPC13, BMP6, CAPN1, CXCR3, EDAR, GAB2, GK, HLA-DRA, IFNA4, IL21R, INHBA, INHBE, NUDT21, UBE3C
GSE22255	ANAPC13, ATP5B, CANX, EDAR, PPIL2, RXRG
GSE28829	ACP5, ANAPC13, ANAPC2, PPIL2

Table 8
Pathway enrichment scores in each dataset and over all five datasets. The scores are not calculated when a pathway is not selected in a particular dataset.

	GSE12288	GSE16561	GSE20219	GSE22255	GSE28829	Overall
Antigen processing and presentation	0.02	–	0.19	–	–	0.008
Axon guidance	0.41	–	–	–	–	0.59
Cytokine–cytokine receptor interaction	0.08	–	0.01	0.46	–	0.001
Hedgehog signaling pathway	0.02	0.12	0.17	–	–	0.006
Huntington's disease	0.02	0.05	–	0.32	–	0.11
mRNA surveillance pathway	–	–	0.22	–	–	0.403
Osteoclast differentiation	0.11	–	0.08	–	0.30	0.017
Phagosome	0.45	–	–	0.29	–	0.08
PPAR signaling pathway	–	0.15	0.22	0.15	–	0.089
Protein processing in endoplasmic reticulum	0.02	0.30	0.42	–	–	0.021
Ubiquitin mediated proteolysis	0.09	0.26	0.07	0.03	0.003	0.001

weaker minimal signal strength condition which allows some false positives and negatives at the gene selection level, thus avoids the restrictive irrepresentable condition. Such advantages have been explicitly quantified in our theoretical results.

Our method can also be adapted to certain new scientific information. For example, for some pathways, their genes tend to be all upregulated or downregulated. Such information can be easily built into our method by requiring $g_{kj} \geq 0$ and $\zeta_{kjm} \geq 0$ for such a pathway k. These constraints can be added to Steps 2 and 3 of our algorithm at minimum extra computational cost. Once they are incorporated in the algorithm, our solution guarantees that the effects of genes under such pathways will be concordant.

The L_1 penalty on the decomposed parameters imposed in (3) is not the only choice. If the studies being integrated are more homogeneous, the L_1 penalty on $\boldsymbol{\zeta}$ can be replaced by the L_2 penalty. In this way, our method will select genes in an "all-in-or-all-out" fashion, meaning that a gene is either effective in all studies or not effective at all.

Acknowledgments

The authors would like to thank the editor-in-chief, the associate editor and two reviewers for reviewing the manuscript and provide valuable comments which have led to a substantial improvement from the original version. Li's work is partially

supported by NSF Grant DMS-1127914 and NIH Grants R01GM047845 and R01AI029168. Wang's work is partially supported by NIH Grant 5 R01 HG007377-02.

Appendix. Proofs

Proof of Proposition 1. First, if $p_k^* = 0$, then $\beta_{kjm}^* = 0$. Hence, it is obvious that $\{k : p_k^* \neq 0\}^c \subset \mathcal{P}^c$. On the other hand, by the definition of $\mathcal{P}$ and the identifiability condition (Condition 1), $\mathcal{P}^c \subset \{k : p_k^* = 0\}$.

Proof of Proposition 2. Since $\hat\beta_{kjm} = \hat p_k \hat g_{kj} \hat\zeta_{kjm}$, $\hat p_k = 0$ implies $\hat\beta_{kjm} = 0$ for all $j = 1, \ldots, G_k$ and $m = 1, \ldots, M$. Hence, $\{k : \hat p_k \neq 0\}^c \subset \hat{\mathcal{P}}^c$. On the other hand, if $\hat\beta_{kjm} = 0$ for all j and m, since $(\hat p_k, \hat g_{kj}, \hat\zeta_{kjm})$ minimizes (3), it must hold that $\hat p_k = \hat g_{kj} = \hat\zeta_{kjm} = 0$, otherwise $(0, 0, 0)$ will give a smaller value of the objective function in (3).

Proof of Lemma 1. Referring to the main body of the paper, under decomposition (2), (3) is equivalent to

$$\begin{aligned}\underset{\boldsymbol{\alpha},\boldsymbol{p},\boldsymbol{g},\boldsymbol{\zeta}}{\operatorname{argmin}} &- \sum_{m=1}^{M} \ell_m(\alpha_m, \boldsymbol{\beta}_m) + \sum_{k=1}^{K} |p_k| + \sum_{k=1}^{K}\sum_{j=1}^{G_k} |g_{kj}| + \chi_n \sum_{k=1}^{K}\sum_{j=1}^{G_k}\sum_{m=1}^{M} |\zeta_{kjm}|,\\ &\text{s.t. } \beta_{kjm} = p_k g_{kj} \zeta_{kjm}.\end{aligned}$$

Since the term $-\sum_{m=1}^{M} \ell_m(\alpha_m, \boldsymbol{\beta}_m)$ is irrelevant to $(\boldsymbol{p}, \boldsymbol{g}, \boldsymbol{\zeta})$, the above problem is further equivalent to

$$\begin{aligned}\min_{\boldsymbol{p},\boldsymbol{g},\boldsymbol{\zeta}} &\sum_{k=1}^{K} |p_k| + \sum_{k=1}^{K}\sum_{j=1}^{G_k} |g_{kj}| + \chi_n \sum_{k=1}^{K}\sum_{j=1}^{G_k}\sum_{m=1}^{M} |\zeta_{kjm}|,\\ &\text{s.t. } \beta_{kjm} = p_k g_{kj} \zeta_{kjm}.\end{aligned} \tag{A.1}$$

Next, we show that (A.1) is equivalent to

$$\begin{aligned}\min_{\boldsymbol{p},\boldsymbol{g},\boldsymbol{\zeta}} &\sum_{k=1}^{K} |p_k| + \sum_{k=1}^{K}\sum_{j=1}^{G_k} |g_{kj}| + \chi_n \sum_{k=1}^{K}\sum_{j=1}^{G_k}\sum_{m=1}^{M} |\zeta_{kjm}|,\\ &\text{s.t. } |\beta_{kjm}| = |p_k g_{kj} \zeta_{kjm}|.\end{aligned} \tag{A.2}$$

In fact, it holds trivially that any solution of (A.1) is also a solution of (A.2). On the contrary, if $(\hat p_k, \hat g_{kj}, \hat\zeta_{kjm})$ is a solution of (A.2), then

$$(\bar p_k, \bar g_{kj}, \bar\zeta_{kjm}) = \begin{cases} (\hat p_k, \hat g_{kj}, \hat\zeta_{kjm}) & \text{if } \beta_{kjm} \hat p_k \hat g_{kj} \hat\zeta_{kjm} \geq 0 \\ (\hat p_k, \hat g_{kj}, -\hat\zeta_{kjm}) & \text{if } \beta_{kjm} \hat p_k \hat g_{kj} \hat\zeta_{kjm} < 0 \end{cases}$$

is a solution of (A.1). Hence (A.1) and (A.2) are equivalent in the sense that there is a one-to-one correspondence between their solutions. By Lagrange multiplier, (A.2) is equivalent to minimizing

$$\mathcal{H}(\boldsymbol{p}, \boldsymbol{g}, \boldsymbol{\zeta}) := \sum_{k=1}^{K} |p_k| + \sum_{k=1}^{K}\sum_{j=1}^{G_k} |g_{kj}| + \chi_n \sum_{k=1}^{K}\sum_{j=1}^{G_k}\sum_{m=1}^{M} |\zeta_{kjm}| + \sum_{k=1}^{K}\sum_{j=1}^{G_k}\sum_{m=1}^{M} \alpha_{kjm}(|\beta_{kjm}| - |p_k g_{kj} \zeta_{kjm}|).$$

Then, we have

$$\frac{\partial \mathcal{H}}{\partial |p_k|} = 1 - \sum_{j=1}^{G_k}\sum_{m=1}^{M} \alpha_{kjm} |g_{kj} \zeta_{kjm}| = 0, \tag{A.3}$$

$$\frac{\partial \mathcal{H}}{\partial |g_{kj}|} = 1 - \sum_{m=1}^{M} \alpha_{kjm} |p_k \zeta_{kjm}| = 0, \tag{A.4}$$

$$\frac{\partial \mathcal{H}}{\partial |\zeta_{kjm}|} = \chi_n - \alpha_{kjm} |p_k g_{kj}| = 0, \tag{A.5}$$

(A.5) implies that $\alpha_{kjm} = \chi_n / |p_k g_{kj}|$, which together with (A.4) give $|g_{kj}| = \chi_n \sum_{m=1}^{M} |\zeta_{kjm}|$. Then, it follows from (A.3) that

$$|p_k| = \sum_{j=1}^{G_k} |g_{kj}| = \chi_n \sum_{j=1}^{G_k}\sum_{m=1}^{M} |\zeta_{kjm}|. \tag{A.6}$$

Since $|\beta_{kjm}| = |p_k g_{kj} \zeta_{kjm}|$, we have $\chi_n \sum_{m=1}^{M} |\beta_{kjm}| = |p_k||g_{kj}|^2$. Hence,

$$|g_{kj}| = \left(|p_k|^{-1} \chi_n \sum_{m=1}^{M} |\beta_{kjm}| \right)^{1/2}.$$

This together with $|p_k| = \sum_{j=1}^{G_k} |g_{kj}|$ gives that $|p_k| = \chi_n^{1/6} \{\sum_{j=1}^{G_k} (\sum_{m=1}^{M} |\beta_{kjm}|)^{1/2}\}^{1/3}$. Therefore,

$$\sum_{j=1}^{G_k} |g_{kj}| = \chi_n \sum_{j=1}^{G_k} \sum_{m=1}^{M} |\zeta_{kjm}| = \chi_n^{1/6} \left\{ \sum_{j=1}^{G_k} \left(\sum_{m=1}^{M} |\beta_{kjm}| \right)^{1/2} \right\}^{1/3}.$$

This together with (A.6) shows that

$$\sum_{k=1}^{K} |p_k| + \sum_{k=1}^{K} \sum_{j=1}^{G_k} |g_{kj}| + \chi_n \sum_{k=1}^{K} \sum_{j=1}^{G_k} \sum_{m=1}^{M} |\zeta_{kjm}| = 3\chi_n^{1/6} \sum_{k=1}^{K} \left\{ \sum_{j=1}^{G_k} \left(\sum_{m=1}^{M} |\beta_{kjm}| \right)^{1/2} \right\}^{1/3}.$$

Hence, (3) and (5) are equivalent.

Proof of Theorem 1. We show that there exists a solution $\hat{\boldsymbol{\beta}}$ to (5) such that its restriction on set J, i.e., $\hat{\boldsymbol{\beta}}_J$, satisfies (b), and all other elements are 0. Then, by definition, this solution also satisfies (a).

By optimization theory, the vector $\hat{\boldsymbol{\beta}}$ that satisfies the following Karush–Kuhn–Tucker (KKT) conditions is a solution to (6).

$$\boldsymbol{X}_J^\top \boldsymbol{Y} - \boldsymbol{X}_J^\top \boldsymbol{\mu}(\boldsymbol{X}\hat{\boldsymbol{\beta}}) = n\lambda_n \nabla \rho(\hat{\boldsymbol{\beta}}_J), \tag{A.7}$$

$$\boldsymbol{X}_{J^c}^\top \boldsymbol{Y} - \boldsymbol{X}_{J^c}^\top \boldsymbol{\mu}(\boldsymbol{X}\hat{\boldsymbol{\beta}}) \in n\lambda_n \partial \rho(\hat{\boldsymbol{\beta}}_{J^c}), \tag{A.8}$$

$$\lambda_{\min}(\boldsymbol{X}_J^\top \boldsymbol{\Sigma}(\boldsymbol{X}\hat{\boldsymbol{\beta}}) \boldsymbol{X}_J) > n\lambda_n \kappa(\hat{\boldsymbol{\beta}}_J). \tag{A.9}$$

In (A.7), $\nabla \rho(\hat{\beta}_{kjm}) = (1/6)\|\hat{\boldsymbol{\beta}}_{kj\cdot}\|_1^{-5/6} \mathrm{sign}(\hat{\beta}_{kjm})$ for $(k, j, m) \in J$. In (A.8), the subgradient $\partial \hat{\beta}_{kjm} = (-\|\hat{\boldsymbol{\beta}}_{kj\cdot}\|_1^{-5/6}/6, \|\hat{\boldsymbol{\beta}}_{kj\cdot}\|_1^{-5/6}/6)$ for $(k, j, m) \in J^c$. In (A.9),

$$\kappa(\hat{\boldsymbol{\beta}}_J) = \max_{(k,j,m) \in J} (5/36)\|\hat{\boldsymbol{\beta}}_{kj\cdot}\|_1^{-11/6}.$$

Let $\boldsymbol{\xi} = \boldsymbol{X}^\top \boldsymbol{Y} - \boldsymbol{X}^\top \boldsymbol{\mu}(\boldsymbol{X}\boldsymbol{\beta}^*)$. Consider the event $E = \{\|\boldsymbol{\xi}_J\|_\infty \le (2^{-1} n \log n)^{1/2}\}$. Since $y_{mi} \in \{0, 1\}$ and columns of $\boldsymbol{X}$ have been standardized to have norm $n^{1/2}$, by Proposition 4(a) of [6],

$$\Pr(|\xi_{kjm}| \ge t) \le 2\exp(-2t^2/n), \tag{A.10}$$

where ξ_{kjm} is the (k, j, m)th element of $\boldsymbol{\xi}$. Then, by the union bound,

$$\Pr(E) \ge 1 - \sum_{(k,j,m) \in J} P\{\xi_{kjm} \ge (2^{-1} n \log n)^{1/2}\} \ge 1 - 2K^*/n. \tag{A.11}$$

In the event E, we show two results: [1] within the hypercube $\mathcal{M} := \{\boldsymbol{\beta} : \|\boldsymbol{\beta} - \boldsymbol{\beta}_J\|_\infty \le cn^{-\gamma}\}$, there exists $\hat{\boldsymbol{\beta}}_J \in \mathcal{R}^{K^*}$ that satisfies (A.7) and (A.9); [2] $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_J, \boldsymbol{0})^\top$ satisfies (A.8). These two results together with the KKT conditions and (A.11) complete the proof.

[1] Let $\boldsymbol{\eta} = n\lambda_n \nabla \rho(\boldsymbol{\beta})$, where the element $\eta_{kjm} = (1/6) n\lambda_n \|\boldsymbol{\beta}_{kj\cdot}\|_1^{-5/6} \mathrm{sign}(\beta_{kjm})$. Under Condition 2, for sufficiently large n, we have

$$\|\boldsymbol{\beta}_{kj\cdot}^*\|_1 \ge 2B > 2cn^{-\gamma} \ge 2\|\boldsymbol{\beta}_{kj\cdot} - \boldsymbol{\beta}_{kj\cdot}^*\|_1, \tag{A.12}$$

for all $\boldsymbol{\beta} \in \mathcal{N}$. In the second inequality, we use the fact that γ is arbitrarily close to 1/2 so that $B > cn^{-\gamma}$. Therefore, $\|\boldsymbol{\beta}_{kj\cdot}\|_1 \ge \|\boldsymbol{\beta}_{kj\cdot}^*\|_1/2$. It further implies that $\|\boldsymbol{\eta}\|_\infty \le B^{-5/6} n\lambda_n$. Define

$$\boldsymbol{\Psi}(\boldsymbol{\beta}) = \boldsymbol{X}_J^\top \{\boldsymbol{\mu}(\boldsymbol{X}_J\boldsymbol{\beta}) - \boldsymbol{\mu}(\boldsymbol{X}_J\boldsymbol{\beta}_J^*)\} - (\boldsymbol{\xi}_J - \boldsymbol{\eta}). \tag{A.13}$$

We show that $\boldsymbol{\Psi}(\boldsymbol{\beta}) = 0$ has a solution $\hat{\boldsymbol{\beta}}_J$ within $\mathcal{M}$. Then, $\hat{\boldsymbol{\beta}}_J$ also solves (A.7). By Taylor expansion,

$$\boldsymbol{X}_J^\top \{\boldsymbol{\mu}(\boldsymbol{X}_J\boldsymbol{\beta}) - \boldsymbol{\mu}(\boldsymbol{X}_J\boldsymbol{\beta}_J^*)\} = \boldsymbol{X}_J^\top \boldsymbol{\Sigma}(\boldsymbol{X}\boldsymbol{\beta}^*) \boldsymbol{X}_J (\boldsymbol{\beta} - \boldsymbol{\beta}_J^*) + \boldsymbol{r},$$

where the (k, j, m)th element of $\boldsymbol{r}$ has $r_{kjm} = \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_J^*)^\top \boldsymbol{R}(\boldsymbol{\delta}^{(kjm)})(\boldsymbol{\beta} - \boldsymbol{\beta}_J^*)$, $\boldsymbol{R}(\boldsymbol{\delta}^{(kjm)}) = \boldsymbol{X}_J^\top\{\mathrm{diag}(\boldsymbol{X}_{kjm} \circ \boldsymbol{\mu}''(\boldsymbol{X}_J\boldsymbol{\delta}^{(kjm)}))\}\boldsymbol{X}_J$, $\boldsymbol{\delta}^{(kjm)}$ is a vector lying on the line segment joining $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_J^*$. By Conditions 2, 5, $\|\boldsymbol{r}\|_\infty = O\left(K^* n^{1-2\gamma}\right) = o(n^{1-\gamma})$, where in the last equality we use the fact that $K^* = o(n^\gamma)$ as γ is arbitrarily close to 1/2.

Let

$$\bar{\boldsymbol{\Psi}}(\boldsymbol{\beta}) = \{\boldsymbol{X}_J^\top \boldsymbol{\Sigma}(\boldsymbol{X}\boldsymbol{\beta}^*)\boldsymbol{X}_J\}^{-1}\boldsymbol{\Psi}(\boldsymbol{\beta}) = \boldsymbol{\beta} - \boldsymbol{\beta}_J^* + \boldsymbol{v},$$

where $\boldsymbol{v} = -\{\boldsymbol{X}_J^\top \boldsymbol{\Sigma}(\boldsymbol{X}\boldsymbol{\beta}^*)\boldsymbol{X}_J\}^{-1}(\boldsymbol{\xi}_J - \boldsymbol{\eta} - \boldsymbol{r})$. Then, we have

$$\begin{aligned}\|\boldsymbol{v}\|_\infty &\le \|\{\boldsymbol{X}_J^\top \boldsymbol{\Sigma}(\boldsymbol{X}\boldsymbol{\beta}^*)\boldsymbol{X}_J\}^{-1}\|_\infty(\|\boldsymbol{\xi}_J - \boldsymbol{\eta}\|_\infty + \|\boldsymbol{r}\|_\infty)\\ &= O(n^{-1}(\|\boldsymbol{\xi}_J\|_\infty + \|\boldsymbol{\eta}\|_\infty + \|\boldsymbol{r}\|_\infty))\\ &= O((n^{-1}\log n)^{1/2}) + o(n^{-1/2}) + o(n^{-\gamma}) = o(n^{-\gamma}).\end{aligned} \tag{A.14}$$

Hence, for sufficiently large n, if $(\boldsymbol{\beta} - \boldsymbol{\beta}_J^*)_{kjm} = n^{-\gamma}$, we have $\{\bar{\boldsymbol{\Psi}}(\boldsymbol{\beta})\}_{kjm} \ge n^{-\gamma} - \|\boldsymbol{v}\|_\infty \ge 0$, and if $(\boldsymbol{\beta} - \boldsymbol{\beta}_J^*)_{kjm} = -n^{-\gamma}$, we have $\{\bar{\boldsymbol{\Psi}}(\boldsymbol{\beta})\}_{kjm} \le -n^{-\gamma} + \|\boldsymbol{v}\|_\infty \le 0$. Since the function $\bar{\boldsymbol{\Psi}}(\boldsymbol{\beta})$ is continuous in $\mathcal{M}$, an application of Miranda's existence theorem (see, e.g., [25]) shows that equation $\bar{\boldsymbol{\Psi}}(\boldsymbol{\beta}) = \boldsymbol{0}$ has a solution $\hat{\boldsymbol{\beta}}_J$ in $\mathcal{M}$. Then, $\hat{\boldsymbol{\beta}}_J$ also solves $\boldsymbol{\Psi}(\boldsymbol{\beta}) = 0$ and further solves (A.7).

Since $B > cn^{-\gamma} \ge \|\hat{\boldsymbol{\beta}}_J - \boldsymbol{\beta}_J^*\|_\infty$, we have $\|\hat{\boldsymbol{\beta}}_{kj\cdot}\|_1 \ne \boldsymbol{0}$ for $(k, j, \cdot) \in J$. Therefore, for any $k \in \mathcal{P}$, there exists at least one (j, m) such that $\hat{\beta}_{kjm} \ne 0$. By definition, $\mathcal{P} \subset \hat{\mathcal{P}}$.

[2] Next, we show that $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_J, \boldsymbol{0})^\top$ satisfies (A.8), therefore solves (5). Thus, the solution $\hat{\boldsymbol{\beta}}$ admits the property that $\hat{\mathcal{P}} = \mathcal{P}$. To show that $\hat{\boldsymbol{\beta}}$ satisfies (A.8), recall that the subgradient $\partial\hat{\beta}_{kjm}$ is the interval $(-\|\hat{\boldsymbol{\beta}}_{kj\cdot}\|_1^{-5/6}/6, \|\hat{\boldsymbol{\beta}}_{kj\cdot}\|_1^{-5/6}/6)$. When $\hat{\boldsymbol{\beta}}_{J^c} = \boldsymbol{0}$, this interval is $(-\infty, +\infty)$. Hence, (A.8) always holds when $\hat{\boldsymbol{\beta}}_{J^c} = \boldsymbol{0}$.

Proof of Theorem 2. Again, by the KKT conditions, any vector $\hat{\boldsymbol{\beta}}$ satisfies (A.15)–(A.17) is a solution to (5).

$$\boldsymbol{X}_I^\top \boldsymbol{Y} - \boldsymbol{X}_I^\top \boldsymbol{\mu}(\boldsymbol{X}\hat{\boldsymbol{\beta}}) = n\lambda_n \nabla\rho(\hat{\boldsymbol{\beta}}_I), \tag{A.15}$$

$$\boldsymbol{X}_{II}^\top \boldsymbol{Y} - \boldsymbol{X}_{II}^\top \boldsymbol{\mu}(\boldsymbol{X}\hat{\boldsymbol{\beta}}) = n\lambda_n \partial\rho(\hat{\boldsymbol{\beta}}_{II}), \tag{A.16}$$

$$\lambda_{\min}(\boldsymbol{X}_I^\top \boldsymbol{\Sigma}(\boldsymbol{X}\hat{\boldsymbol{\beta}})\boldsymbol{X}_I) > n\lambda_n \kappa(\hat{\boldsymbol{\beta}}_I), \tag{A.17}$$

where $\nabla\rho(\hat{\beta}_{kjm}) = \tau_{kj}(\hat{\boldsymbol{\beta}}_{kj\cdot})\mathrm{sign}(\hat{\beta}_{kjm})/6$ for $(k, j, m) \in I$, $\partial\rho(\hat{\beta}_{kjm}) \in (-\tau_{kj}(\hat{\boldsymbol{\beta}}_{kj\cdot})/6, \tau_{kj}(\hat{\boldsymbol{\beta}}_{kj\cdot})/6)$, for $(k, j, m) \in II$,

$$\tau_{kj}(\hat{\boldsymbol{\beta}}_{kj\cdot}) = \left(\sum_{j=1}^{G_k} \|\hat{\boldsymbol{\beta}}_{kj\cdot}\|_1^{\frac{1}{2}}\right)^{-\frac{2}{3}} \|\hat{\boldsymbol{\beta}}_{kj\cdot}\|_1^{-\frac{1}{2}}, \tag{A.18}$$

$$\kappa(\hat{\boldsymbol{\beta}}_I) = \max_{(k,j,m)\in I} \frac{1}{18}\left(\sum_{j=1}^{G_k} \|\hat{\boldsymbol{\beta}}_{kj\cdot}\|_1^{\frac{1}{2}}\right)^{-\frac{5}{3}} \|\hat{\boldsymbol{\beta}}_{kj\cdot}\|_1^{-1} + \frac{1}{12}\left(\sum_{j=1}^{G_k} \|\hat{\boldsymbol{\beta}}_{kj\cdot}\|_1^{\frac{1}{2}}\right)^{-\frac{2}{3}} \|\hat{\boldsymbol{\beta}}_{kj\cdot}\|_1^{-\frac{3}{2}}. \tag{A.19}$$

Let $\boldsymbol{\xi} = \boldsymbol{X}^\top \boldsymbol{Y} - \boldsymbol{X}^\top \boldsymbol{\mu}(\boldsymbol{X}\boldsymbol{\beta}^*)$. Consider events

$$E_1 = \{\|\boldsymbol{\xi}_I\|_\infty \le (2^{-1}n\log n)^{1/2}\} \quad \text{and} \quad E_2 = \{\|\boldsymbol{\xi}_{I^c}\|_\infty \le (n\log p)^{1/2}\}.$$

By (A.10) and the union bound, we have

$$\begin{aligned}\Pr(E_1 \cap E_2) &\ge 1 - \sum_{(k,j,m)\in I} P\{|\xi_{kjm}| \ge (2^{-1}n\log n)^{1/2}\} - \sum_{(k,j,m)\in I^c} P\{|\xi_{kjm}| \ge (n\log p)^{1/2}\}\\ &\ge 1 - 2\{s/n + (Md - s)/d^2\}.\end{aligned}$$

In event $E_1 \cap E_2$, we show two results: [1] within the set $\mathcal{M}_0 := \{\boldsymbol{\beta} \in \mathcal{R}^s : \|\boldsymbol{\beta} - \boldsymbol{\beta}_I^*\|_\infty \le cn^{-\gamma}\}$, there exists a vector $\hat{\boldsymbol{\beta}}_I \in \mathcal{R}^s$ satisfying (A.15) and (A.17); [2] $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_I, \boldsymbol{0})^\top$ satisfies (A.16). These two steps together with the KKT conditions complete the proof.

[1] Let $\boldsymbol{\eta} = n\lambda_n \nabla\rho(\boldsymbol{\beta})$, where $\eta_{kjm} = n\lambda_n \tau_{kj}(\boldsymbol{\beta}_{kj\cdot})\mathrm{sign}(\beta_{kjm})/6$. Under Condition 7, Similarly as in (A.12) we have $\|\boldsymbol{\beta}_{kj\cdot}^*\|_1 > 2\|\boldsymbol{\beta}_{kj\cdot}^* - \boldsymbol{\beta}_{kj\cdot}\|_1$ and further $\|\boldsymbol{\beta}_{kj\cdot}\|_1 \ge \|\boldsymbol{\beta}_{kj\cdot}^*\|_1/2$. Therefore, $6|\eta_{kjm}| = n\lambda_n \tau_{kj}(\boldsymbol{\beta}_{kj\cdot}) \le n\lambda_n \tau_{kj}(\boldsymbol{\beta}_{kj\cdot}^*)$. Hence, $\|\boldsymbol{\eta}\|_\infty \le n\lambda_n h_n^{-1}$. It further implies that $\|\boldsymbol{\xi}_I - \boldsymbol{\eta}\|_\infty \le \|\boldsymbol{\xi}_I\|_\infty + \|\boldsymbol{\eta}\|_\infty \le (2^{-1}n\log n)^{1/2} + n\lambda_n h_n^{-1}$.

Define

$$\boldsymbol{\Psi}(\boldsymbol{\beta}) = \boldsymbol{X}_I^\top\{\boldsymbol{\mu}(\boldsymbol{X}_I\boldsymbol{\beta}) - \boldsymbol{\mu}(\boldsymbol{X}_I\boldsymbol{\beta}_I^*)\} - (\boldsymbol{\xi}_I - \boldsymbol{\eta}). \tag{A.20}$$

Again, we show that $\Psi(\boldsymbol{\beta}) = \mathbf{0}$ has a solution within $\mathcal{M}_0$. By a Taylor expansion, we have,

$$\boldsymbol{X}_l^\top\{\boldsymbol{\mu}(\boldsymbol{X}_l\boldsymbol{\beta}) - \boldsymbol{\mu}(\boldsymbol{X}_l\boldsymbol{\beta}_l^*)\} = \boldsymbol{X}_l^\top\boldsymbol{\Sigma}(\boldsymbol{X}\boldsymbol{\beta}^*)\boldsymbol{X}_l(\boldsymbol{\beta} - \boldsymbol{\beta}_l^*) + \boldsymbol{r},$$

where the Lagrange remainder $\boldsymbol{r} = (r_{kjm})^\top$ such that $r_{kjm} = (\boldsymbol{\beta} - \boldsymbol{\beta}_l^*)^\top\boldsymbol{R}(\tilde{\boldsymbol{\delta}}^{(kjm)})(\boldsymbol{\beta} - \boldsymbol{\beta}_l^*)/2$, $\boldsymbol{R}(\tilde{\boldsymbol{\delta}}^{(kjm)}) = \boldsymbol{X}_l^\top\{\mathrm{diag}(\boldsymbol{X}_{kjm} \circ \boldsymbol{\mu}''(\boldsymbol{X}_l\tilde{\boldsymbol{\delta}}^{(kjm)}))\}\boldsymbol{X}_l$ and $\tilde{\boldsymbol{\delta}}^{(kjm)}$ is a vector lying on the line segment joining $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_l^*$. By Condition 10, we have

$$\|\boldsymbol{r}\|_\infty = O(sn^{1-2\gamma}). \tag{A.21}$$

Let

$$\bar{\boldsymbol{\Psi}}(\boldsymbol{\beta}) = \{\boldsymbol{X}_l^\top\boldsymbol{\Sigma}(\boldsymbol{X}\boldsymbol{\beta}^*)\boldsymbol{X}_l\}^{-1}\boldsymbol{\Psi}(\boldsymbol{\beta}) = \boldsymbol{\beta} - \boldsymbol{\beta}_l^* + \boldsymbol{v},$$

where $\boldsymbol{v} = -\{\boldsymbol{X}_l^\top\boldsymbol{\Sigma}(\boldsymbol{X}\boldsymbol{\beta}^*)\boldsymbol{X}_l\}^{-1}(\boldsymbol{\xi}_l - \boldsymbol{\eta} - \boldsymbol{r})$. Then, it follows from Conditions 7, 8, and $\lambda_n = o(h_n n^{-\gamma})$ that

$$\begin{aligned}\|\boldsymbol{v}\|_\infty &\le \|\{\boldsymbol{X}_l^\top\boldsymbol{\Sigma}(\boldsymbol{X}\boldsymbol{\beta}^*)\boldsymbol{X}_l\}^{-1}\|_\infty(\|\boldsymbol{\xi}_l - \boldsymbol{\eta}\|_\infty + \|\boldsymbol{r}\|_\infty)\\ &= O((n^{-1}\log n)^{1/2} + \lambda_n h_n^{-1} + sn^{-2\gamma}) = o(n^{-\gamma}).\end{aligned} \tag{A.22}$$

By the same argument as in the proof of Theorem 1, there exists a vector $\hat{\boldsymbol{\beta}}_l$ within $\mathcal{M}_0$ such that $\bar{\boldsymbol{\Psi}}(\hat{\boldsymbol{\beta}}_l) = \mathbf{0}$. Hence, $\hat{\boldsymbol{\beta}}_l$ also solves (A.15). On the other hand, by the stated choice of λ_n, $\hat{\boldsymbol{\beta}}_l$ satisfies (A.17) for sufficiently large n.

[2] Let $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_l, \mathbf{0})^\top$. Next, we prove that $\hat{\boldsymbol{\beta}}$ satisfies (A.16) for the stated choice of λ_n. Indeed, (A.16) requires that

$$|\boldsymbol{X}_{kjm}^\top\boldsymbol{Y} - \boldsymbol{X}_{kjm}^\top\boldsymbol{\mu}(\boldsymbol{X}\hat{\boldsymbol{\beta}})| < \frac{1}{6}n\lambda_n\tau_{kj}(\hat{\boldsymbol{\beta}}_{kj\cdot}), \tag{A.23}$$

for any $(k, j, m) \in II$. As proved in [1], $\|\boldsymbol{\beta}_{kj\cdot}^*\|_1 > 2\|\boldsymbol{\beta}_{kj\cdot}^* - \hat{\boldsymbol{\beta}}_{kj\cdot}\|_1$, hence $\|\hat{\boldsymbol{\beta}}_{kj\cdot}\|_1 < 1.5\|\boldsymbol{\beta}_{kj\cdot}^*\|_1$. Then, we have $\tau_{kj}(\hat{\boldsymbol{\beta}}_{kj\cdot}) \ge 0.6\tau_{kj}(\boldsymbol{\beta}_{kj\cdot}^*)$. It implies that

$$\min_{k,j}\tau_{kj}(\hat{\boldsymbol{\beta}}_{kj\cdot}) \ge \min_{k,j}0.6\tau_{kj}(\boldsymbol{\beta}_{kj}^*) \ge 0.6u_n^{-1}.$$

To prove (A.16), by (A.23) and the above arguments, it suffices to show that

$$\|\boldsymbol{X}_{II}^\top\boldsymbol{Y} - \boldsymbol{X}_{II}^\top\boldsymbol{\mu}(\boldsymbol{X}\hat{\boldsymbol{\beta}})\|_\infty < 0.1n\lambda_n u_n^{-1}. \tag{A.24}$$

Observe that

$$\boldsymbol{X}_{II}^\top\boldsymbol{Y} - \boldsymbol{X}_{II}^\top\boldsymbol{\mu}(\boldsymbol{X}\hat{\boldsymbol{\beta}}) = \boldsymbol{X}_{II}^\top\{\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{X}\boldsymbol{\beta}^*)\} + \boldsymbol{X}_{II}^\top\{\boldsymbol{\mu}(\boldsymbol{X}\boldsymbol{\beta}^*) - \boldsymbol{\mu}(\boldsymbol{X}\hat{\boldsymbol{\beta}})\}. \tag{A.25}$$

In event E_2, $\|\boldsymbol{X}_{II}^\top\{\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{X}\boldsymbol{\beta}^*)\}\|_\infty = O((n\log d)^{1/2})$. Then, it follows from $(n^{-1}\log d)^{1/2} = o(\lambda_n)$ that

$$(n\lambda_n)^{-1}\|\boldsymbol{X}_{II}^\top\{\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{X}\boldsymbol{\beta}^*)\}\|_\infty = o(1). \tag{A.26}$$

For the second term on the right hand side of (A.25),

$$\boldsymbol{X}_{II}^\top\{\boldsymbol{\mu}(\boldsymbol{X}\hat{\boldsymbol{\beta}}) - \boldsymbol{\mu}(\boldsymbol{X}\boldsymbol{\beta}^*)\} = \boldsymbol{X}_{II}^\top\{\boldsymbol{\mu}(\boldsymbol{X}_l\hat{\boldsymbol{\beta}}_l) - \boldsymbol{\mu}(\boldsymbol{X}_l\boldsymbol{\beta}_l^*)\} = \boldsymbol{X}_{II}^\top\boldsymbol{\Sigma}(\boldsymbol{X}\boldsymbol{\beta}^*)\boldsymbol{X}_l(\hat{\boldsymbol{\beta}}_l - \boldsymbol{\beta}_l^*) + \boldsymbol{w},$$

where $\boldsymbol{w} = (w_{kjm})^\top$ such that $w_{kjm} = (\hat{\boldsymbol{\beta}}_l - \boldsymbol{\beta}_l^*)^\top\boldsymbol{R}(\bar{\boldsymbol{\delta}}^{(kjm)})(\hat{\boldsymbol{\beta}}_l - \boldsymbol{\beta}_l^*)/2$, where $\bar{\boldsymbol{\delta}}^{(kjm)}$ is a vector lying on the line segment joining $\hat{\boldsymbol{\beta}}_l$ and $\boldsymbol{\beta}_l^*$. Similarly as in (A.21),

$$\|\boldsymbol{w}\|_\infty = O(sn^{1-2\gamma}). \tag{A.27}$$

Since $\hat{\boldsymbol{\beta}}_l$ solves $\bar{\boldsymbol{\Psi}}(\boldsymbol{\delta}) = \mathbf{0}$, we have $\hat{\boldsymbol{\beta}}_l - \boldsymbol{\beta}_l^* = \{\boldsymbol{X}_l^\top\boldsymbol{\Sigma}(\boldsymbol{X}\boldsymbol{\beta}^*)\boldsymbol{X}_l\}^{-1}(\boldsymbol{\xi}_l - \boldsymbol{\eta} - \boldsymbol{r})$. Therefore,

$$\begin{aligned}&(n\lambda_n)^{-1}\boldsymbol{X}_{II}^\top\{\boldsymbol{\mu}(\boldsymbol{X}_l\hat{\boldsymbol{\beta}}_l) - \boldsymbol{\mu}(\boldsymbol{X}_l\boldsymbol{\beta}_l^*)\}\\ &\quad\le (n\lambda_n)^{-1}\|\boldsymbol{X}_{II}^\top\boldsymbol{\Sigma}(\boldsymbol{X}\boldsymbol{\beta}^*)\boldsymbol{X}_l\{\boldsymbol{X}_l^\top\boldsymbol{\Sigma}(\boldsymbol{X}\boldsymbol{\beta}^*)\boldsymbol{X}_l\}^{-1}\|_\infty\cdot(\|\boldsymbol{\xi}_l - \boldsymbol{\eta}\|_\infty + \|\boldsymbol{r}\|_\infty) + (n\lambda_n)^{-1}\|\boldsymbol{w}\|_\infty\\ &\quad\le (n\lambda_n)^{-1}\|\boldsymbol{X}_{II}^\top\boldsymbol{\Sigma}(\boldsymbol{X}\boldsymbol{\beta}^*)\boldsymbol{X}_l\{\boldsymbol{X}_l^\top\boldsymbol{\Sigma}(\boldsymbol{X}\boldsymbol{\beta}^*)\boldsymbol{X}_l\}^{-1}\|_\infty\|\boldsymbol{\eta}\|_\infty + (n\lambda_n)^{-1}O(\|\boldsymbol{\xi}_l\|_\infty + \|\boldsymbol{r}\|_\infty + \|\boldsymbol{w}\|_\infty),\end{aligned}$$

because by Condition 9, $\|\boldsymbol{X}_{II}^\top\boldsymbol{\Sigma}(\boldsymbol{X}\boldsymbol{\beta}^*)\boldsymbol{X}_l\{\boldsymbol{X}_l^\top\boldsymbol{\Sigma}(\boldsymbol{X}\boldsymbol{\beta}^*)\boldsymbol{X}_l\}^{-1}\|_\infty < 1$. It follows from (A.21) and $sn^{-2\gamma} = o(\lambda_n)$ that $(n\lambda_n)^{-1}O(\|\boldsymbol{\xi}_l\|_\infty + \|\boldsymbol{r}\|_\infty) = o(1)$. Meanwhile, by (A.27) and the choice of λ_n, $(n\lambda_n)^{-1}\|\boldsymbol{w}\|_\infty = o(1)$. Using Condition 9 and $\|\boldsymbol{\eta}\|_\infty \le n\lambda_n h_n^{-1}$, we have

$$(n\lambda_n)^{-1}\|\boldsymbol{X}_{II}^\top\boldsymbol{\Sigma}(\boldsymbol{X}\boldsymbol{\beta}^*)\boldsymbol{X}_l\{\boldsymbol{X}_l^\top\boldsymbol{\Sigma}(\boldsymbol{X}\boldsymbol{\beta}^*)\boldsymbol{X}_l\}^{-1}\|_\infty\cdot\|\boldsymbol{\eta}\|_\infty < 0.1u_n^{-1}.$$

Therefore, (A.24) holds.

References

[1] R.H. Adams, A. Eichmann, Axon guidance molecules in vascular patterning, Cold Spring Harbor. Perspect. Biol. 2 (5) (2010) a001875.
[2] M. Ashburner, C. Ball, J. Blake, D. Botstein, H. Butler, J. Cherry, A. Davis, K. Dolinski, S. Dwight, J. Eppig, et al., Gene Ontology: tool for the unification of biology, Nature Genet. 25 (1) (2000) 25–29.
[3] P. Carbonetto, M. Stephens, Integrated enrichment analysis of variants and pathways in genome-wide association studies indicates central role for IL-2 signaling genes in type 1 diabetes, and cytokine signaling genes in Crohn's disease, PLoS Genet. 9 (10) (2013) e1003770.
[4] L. Carrier, S. Schlossarek, M.S. Willis, T. Eschenhagen, The ubiquitin-proteasome system and nonsense-mediated mRNA decay in hypertrophic cardiomyopathy, Cardiovasc. Res. 85 (2) (2009) 330–338.
[5] A. Deswal, N.J. Petersen, A.M. Feldman, J.B. Young, B.G. White, D.L. Mann, Cytokines and cytokine receptors in advanced heart failure an analysis of the cytokine database from the Vesnarinone Trial (VEST), Circulation 103 (16) (2001) 2055–2059.
[6] J. Fan, J. Lv, Nonconcave penalized likelihood with NP-dimensionality, IEEE Trans. Inform. Theory 57 (8) (2011) 5467–5484.
[7] B. Finck, The PPAR regulatory system in cardiac physiology and disease, Cardiovasc. Res. 73 (2) (2007) 269–277.
[8] J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent, J. Stat. Softw. 33 (1) (2010) 1–22.
[9] J. Groenendyk, P.K. Sreenivasaiah, D.H. Kim, L.B. Agellon, M. Michalak, Biology of endoplasmic reticulum stress in the heart, Circ. Res. 107 (10) (2010) 1185–1197.
[10] D.W. Huang, B.T. Sherman, R.A. Lempicki, Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists, Nucleic Acids Res. 1 (2009) 1–13.
[11] J.C. Kagan, A. Iwasaki, Phagosome as the organelle linking innate and adaptive immunity, Traffic 13 (8) (2012) 1053–1061.
[12] P. Khatri, S. Drăghici, Ontological analysis of gene expression data: current tools, limitations, and open problems, Bioinformatics 21 (18) (2005) 3587–3595.
[13] P. Khatri, M. Sirota, A.J. Butte, Ten years of pathway analysis: current approaches and outstanding challenges, PLoS Comput. Biol. 8 (2) (2012) e1002375.
[14] Q. Li, S. Wang, C.-C. Huang, M. Yu, J. Shao, Meta-analysis based variable selection for gene expression data, Biometrics 70 (4) (2014) 872–880.
[15] J. Liu, J. Huang, S. Ma, Integrative analysis of multiple cancer genomic datasets under the heterogeneity model, Stat. Med. 32 (20) (2013) 3509–3521.
[16] J. Liu, J. Huang, Y. Zhang, Q. Lan, N. Rothman, T. Zheng, S. Ma, Integrative analysis of prognosis data on multiple cancer subtypes, Biometrics 70 (3) (2014) 480–488.
[17] G.C. Melkani, A.S. Trujillo, R. Ramos, R. Bodmer, S.I. Bernstein, K. Ocorr, Huntington's disease induced cardiac amyloidosis is reversed by modulating protein folding and oxidative stress pathways in the drosophila heart, PLoS Genet. 9 (12) (2013) e1004024-17.
[18] M.A. Newton, F.A. Quintana, J.A. Den Boon, S. Sengupta, P. Ahlquist, Random-set methods identify distinct aspects of the enrichment signal in gene-set analysis, Ann. Appl. Stat. (2007) 85–106.
[19] E. Ntzani, J. Ioannidis, et al., Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment, Lancet 362 (9394) (2003) 1439–1444.
[20] H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, M. Kanehisa, KEGG: Kyoto encyclopedia of genes and genomes, Nucleic Acids Res. 28 (1) (2000) 27–30.
[21] F. Rochais, K. Mesbah, R.G. Kelly, Signaling pathways controlling second heart field development, Circ. Res. 104 (8) (2009) 933–942.
[22] R. Saxena, B.F. Voight, V. Lyssenko, N.P. Burtt, P.I. de Bakker, H. Chen, J.J. Roix, S. Kathiresan, J.N. Hirschhorn, M.J. Daly, et al., Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels, Science 316 (5829) (2007) 1331–1336.
[23] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, PNAS 102 (43) (2005) 15545–15550.
[24] G.C. Tseng, D. Ghosh, E. Feingold, Comprehensive literature review and statistical considerations for microarray meta-analysis, Nucleic Acids Res. 40 (9) (2012) 3785–3799.
[25] M.N. Vrahatis, A short proof and a generalization of Mirandas existence theorem, Proc. Amer. Math. Soc. 107 (3) (1989) 701–703.
[26] J.M. Vyas, A.G. Van der Veen, H.L. Ploegh, The known unknowns of antigen processing and presentation, Nat. Rev. Immunol. 8 (8) (2008) 607–618.
[27] J.T. Willerson, P.M. Ridker, Inflammation as a cardiovascular risk factor, Circulation 109 (21 suppl 1) (2004) II-2–II-10.
[28] Y. Wu, M.B. Humphrey, M.C. Nakamura, Osteoclasts—the innate immune cells of the bone, Autoimmunity 41 (3) (2008) 183–194.
[29] T. Wu, K. Lange, Coordinate descent algorithms for lasso penalized regression, Ann. Appl. Stat. 2 (1) (2008) 224–244.
[30] E. Zeggini, M.N. Weedon, C.M. Lindgren, T.M. Frayling, K.S. Elliott, H. Lango, N.J. Timpson, J.R. Perry, N.W. Rayner, R.M. Freathy, et al., Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes, Science 316 (5829) (2007) 1336–1341.
[31] P. Zhao, B. Yu, On model selection consistency of lasso, J. Mach. Learn. Res. 7 (2006) 2541–2563.
[32] N. Zhou, J. Zhu, Group variable selection via a hierarchical lasso and its oracle property, Stat. Interface 3 (2010) 557–574.