Web-based Appendix for Stable estimation in dimension reduction

Wenbo Wu and Xiangrong Yin*

February 14, 2014

Abstract

In this web appendix file, we provide the proofs of theorems, algorithms and additional simulation results and comparisons. Appendix A describes the proofs of Lemma 1 and Lemma 2 which are related to the Sparse Riesz Condition. The two lemmas will be used in proofs of Theorem 1 and Theorem 3. Appendix B provides the proof of Theorem 1, Appendix C includes the proof of Theorem 2. Appendix D provides a detailed explanation of the Grassmann Manifold and its relations with eigen-decomposition, and the link between a Grassmann Manifold and a dimension reduction matrix. Appendix E describes the proof of Theorem 3. Appendix F shows how the values of δ have little influence on the results of the proposed GMSE and SGMSE methods. Appendix G compares the proposed GMSE procedure to the sparse sufficient dimension reduction (SSDR) method (Li 2007), and discusses the differences and connections. Appendix H contains simulation results for showing the sensitivity of SED to the choice of the number of slices in sliced inverse dimension reduction methods. It also contains simulation results of the ensemble method proposed in Section 2.3.2 in the paper. In order to avoid confusing the equation numbers, all equations, tables, and figures in the this appendix file begin with "A.".

*204 Statistics Building, Department of Statistics, University of Georgia, Athens, GA 30602. Email: wuwenbo@uga.edu, xryin@uga.edu

Appendix A

The Sparse Riesz Condition controls the range of eigenvalues of covariance matrices of subsets of a fixed number of design vectors. We will show in Lemma 1 below that for design $\tilde{\mathbf{X}}_A^{\mathbf{w}}$ of rank m, in model (2.4), the Sparse Riesz Condition is satisfied, and in Lemma 2 that the Sparse Riesz Condition guarantees a finite bound.

Lemma 1 and Lemma 2 are needed to prove Theorem 1 and Theorem 3.

Lemma 1. For model (2.4), for any $\tilde{\mathbf{X}}_A^{\mathbf{w}}$ of rank $m \le p$, where $A \subseteq \{1, 2, \cdots, p\}$ with $|A| = p$, the Sparse Riesz Condition (2.10) is satisfied.

Proof: In model (2.4), for design $\tilde{\mathbf{X}}_A^{\mathbf{w}}$ of rank m, we have

$$\frac{\| \tilde{\mathbf{X}}_A^{\mathbf{w}} \mathbf{v} \|^2}{p^2 \| \mathbf{v} \|^2} = \frac{(\tilde{\mathbf{X}}_A^{\mathbf{w}} \mathbf{v})' \tilde{\mathbf{X}}_A^{\mathbf{w}} \mathbf{v}}{p^2 \mathbf{v}' \mathbf{v}} = \frac{\mathbf{v}' \mathcal{D}_A \mathbf{v}}{p^2 \mathbf{v}' \mathbf{v}} = \frac{\sum_{i=1}^{m} W_{Ai}^2 v_i^2}{p^2 \sum_{i=1}^{m} v_i^2}.$$

The Sparse Riesz Condition is satisfied, because

$$\phi_{min}(m) = \min_{|A|=m, \mathbf{v} \in \mathbb{R}^m} \frac{\| \tilde{\mathbf{X}}_A^{\mathbf{w}} \mathbf{v} \|^2}{p^2 \| \mathbf{v} \|^2} = \min_{|A|=m, \mathbf{v} \in \mathbb{R}^m} \frac{\sum_{i=1}^{m} W_{Ai}^2 v_i^2}{p^2 \sum_{i=1}^{m} v_i^2} \ge \min_{|A|=m, \mathbf{v} \in \mathbb{R}^m} \frac{\sum_{i=1}^{m} u^2 v_i^2}{p^2 \sum_{i=1}^{m} v_i^2} = \frac{u^2}{p^2},$$

since W_{Ai} are generated from $[u, 1]$ for some $u > 0$. For the same reason,

$$\phi_{max}(m) = \max_{|A|=m, \mathbf{v} \in \mathbb{R}^m} \frac{\| \tilde{\mathbf{X}}_A^{\mathbf{w}} \mathbf{v} \|^2}{p^2 \| \mathbf{v} \|^2} = \max_{|A|=m, \mathbf{v} \in \mathbb{R}^m} \frac{\sum_{i=1}^{m} W_{Ai}^2 v_i^2}{p^2 \sum_{i=1}^{m} v_i^2} \le \max_{|A|=m, \mathbf{v} \in \mathbb{R}^m} \frac{\sum_{i=1}^{m} v_i^2}{p^2 \sum_{i=1}^{m} v_i^2} = \frac{1}{p^2}.$$

Therefore, $0 < \frac{u^2}{p^2} \le \phi_{min}(m) \le \phi_{max}(m) \le \frac{1}{p^2} < \infty$ for all $m \le p$. $\square$

Lemma 2. Let $\phi_{min}(m)$ and $\phi_{max}(m)$ be defined as in (2.9) and the Sparse Riesz Condition (2.10) holds. Let $S_k \subset \{1, \cdots, p\}$, $\tilde{\mathbf{X}}_k^{\mathbf{w}} = (\tilde{\mathbf{x}}_j, j \in S_k)$ and $\mathbf{\Sigma}_{1k} = (\tilde{\mathbf{X}}_1^{\mathbf{w}})' \tilde{\mathbf{X}}_k^{\mathbf{w}} / n$, Then

$$\frac{\| \mathbf{v} \|^2}{\phi_{max}^Z(|S_1|)} \le \| \mathbf{\Sigma}_{11}^{-1/2} \mathbf{v} \|^2 \le \frac{\| \mathbf{v} \|^2}{\phi_{min}^Z(|S_1|)}, \tag{A.1}$$

for all $\mathbf{v}$ of proper dimension, and that

$$\| \boldsymbol{\beta}_k \|_1^2 \leq \frac{\| \tilde{\mathbf{X}}_k^{\mathbf{w}} \boldsymbol{\beta}_k \|^2 \, |S_k|}{n\phi_{min}^Z(|S_k|)}. \tag{A.2}$$

Proof: Let $\mathbf{v}, \mathbf{h} \in \mathbb{R}^{|S_1|}$ and $\mathbf{v} = \boldsymbol{\Sigma}_{11}^{1/2}\mathbf{h}$. Hence, $\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{v} = \mathbf{h}$. By Lemma 1, since the Sparse Riesz Condition (2.10) holds, we have that $\phi_{min}(|S_1|) \leq \frac{\|\tilde{\mathbf{X}}_1^{\mathbf{w}}\mathbf{h}\|^2}{n\|\mathbf{h}\|^2} \leq \phi_{max}(|S_1|)$.

Note that $\| \tilde{\mathbf{X}}_1^{\mathbf{w}}\mathbf{h} \|^2 = n \| \mathbf{v} \|^2$ and $\| \mathbf{h} \|^2 = \| \boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{v} \|^2$.

So $\phi_{min}(|S_1|) \leq \frac{\|\mathbf{v}\|^2}{\|\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{v}\|^2} \leq \phi_{max}(|S_1|)$, and $\frac{1}{\phi_{max}(|S_1|)} \leq \frac{\|\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{v}\|^2}{\|\mathbf{v}\|^2} \leq \frac{1}{\phi_{min}(|S_1|)}$, which yields (A.1). The Cauchy-Schwarz inequality implies that $\| \boldsymbol{\beta}_k \|_1^2 \leq \| \boldsymbol{\beta}_k \|^2 \, |S_k|$. By the Sparse Riesz Condition, $\phi_{min}(|S_k|) \leq \frac{\|\tilde{\mathbf{X}}_k^{\mathbf{w}}\boldsymbol{\beta}_k\|^2}{n\|\boldsymbol{\beta}_k\|^2} \Longrightarrow \| \boldsymbol{\beta}_k \|^2 \leq \frac{\|\tilde{\mathbf{X}}_k^{\mathbf{w}}\boldsymbol{\beta}_k\|^2}{n\phi_{min}(|S_k|)}$, which yields (A.2). □

Appendix B

Proof of Theorem 1. The goal of Theorem 1 is to find upper bounds of (2.12), (2.13), and (2.14). Since $S = \{k : \beta_k \neq 0\}$ is the set of important predictors in the true model, we define the following sets in Table A.1 to facilitate our proof.

	nonzero $\boldsymbol{\beta}_j : j \in S$	zero $\boldsymbol{\beta}_j : j \notin S$
S_1: selected j	S_3	S_4
$S_2 = S_1^c$	S_5	S_6

Table A.1: Definitions of sets

In our case, $\hat{q} = q_1 = |S_1|$. Define $\mathbf{Q}_{kj}$ to be the selection of variables in S_k from S_j:

$$\mathbf{Q}_{kj}\boldsymbol{\beta}_j = \boldsymbol{\beta}_k, \quad \boldsymbol{\beta}_1' = \boldsymbol{\beta}_3'\mathbf{Q}_{31}, \quad \boldsymbol{\beta}_k = \{\beta_j, j \in S_k\}.$$

Let $\tilde{\mathbf{X}}_i^{\mathbf{w}} = (\tilde{\mathbf{x}}_j, j \in S_i)$, define

$$\boldsymbol{\Sigma}_{jk} = \frac{1}{n}(\tilde{\mathbf{X}}_j^{\mathbf{w}})'\tilde{\mathbf{X}}_k^{\mathbf{w}}, \qquad \mathbf{f}_j = (\tilde{\mathbf{X}}_{S_j}^{\mathbf{w}})'(Y - \tilde{\mathbf{X}}^{\mathbf{w}}\boldsymbol{\beta})/\lambda, \;\; j = 1, 3, 4. \tag{A.3}$$

By (2.5), we have

$$\mathbf{\Sigma}_{jk} = \begin{cases} \mathcal{D}_i/p^2 & j = k, \\ \mathbf{0} & j \neq k. \end{cases} \tag{A.4}$$

where $\mathcal{D}_i = \text{diag}(W_{ij}^2)$, $j = 1, \cdots, q_i$. So

$$\mathbf{\Sigma}_{ii}^{-1} = p^2\mathcal{D}_i^{-1} = p^2\text{diag}(W_{ij}^{-2}). \tag{A.5}$$

With $\mathbf{P}_1$ be the projection from $\mathbb{R}^n$ to the span of $\{\tilde{\mathbf{x}}_j : j \in S_1\}$, we define,

$$\mathbf{v}_{1j} = \frac{\lambda}{\sqrt{n}}\mathbf{\Sigma}_{11}^{-1/2}\mathbf{Q}'_{j1}\mathbf{f}_j, \qquad \mathbf{w}_k = (\mathbf{I} - \mathbf{P}_1)\tilde{\mathbf{X}}_k^{\mathbf{w}}\boldsymbol{\beta}_k. \tag{A.6}$$

Since $\tilde{\mathbf{X}}^{\mathbf{w}}\boldsymbol{\beta} = \tilde{\mathbf{X}}_1^{\mathbf{w}}\boldsymbol{\beta}_1 + \tilde{\mathbf{X}}_2^{\mathbf{w}}\boldsymbol{\beta}_2$ and $(\mathbf{I} - \mathbf{P}_1)\tilde{\mathbf{X}}_1^{\mathbf{w}}\boldsymbol{\beta}_1 = \mathbf{0}$, then by (A.4),

$$\| \mathbf{w}_2 \|^2 = \| (\mathbf{I} - \mathbf{P}_1)\tilde{\mathbf{X}}^{\mathbf{w}}\boldsymbol{\beta} \|^2 = \| (\mathbf{I} - \mathbf{P}_1)\tilde{\mathbf{X}}_2^{\mathbf{w}}\boldsymbol{\beta}_2 \|^2 = \| \tilde{\mathbf{X}}_2^{\mathbf{w}}\boldsymbol{\beta}_2 \|^2 = \| \mathcal{W}_2\boldsymbol{\beta}_2 \|^2,$$

where $\mathcal{W}_i = \mathcal{D}_i^{1/2} = \text{diag}(W_{ij})$. The Karush-Kuhn-Tucker condition (KKT) states that a vector $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \cdots, \hat{\beta}_p)'$ is the solution of (2.6) if and only if

$$\begin{cases} \tilde{\mathbf{x}}'_j(Y - \tilde{\mathbf{X}}^{\mathbf{w}}\hat{\boldsymbol{\beta}}) = \text{sgn}(\hat{\beta}_j)\lambda, & |\hat{\beta}_j| > 0; \\ |\tilde{\mathbf{x}}'_j(Y - \tilde{\mathbf{X}}^{\mathbf{w}}\hat{\boldsymbol{\beta}})| \leq \lambda, & \hat{\beta}_j = 0. \end{cases} \tag{A.7}$$

In our case, the Karush-Kuhn-Tucker condition reduces to:

$$\begin{cases} \tilde{\mathbf{x}}'_j(Y - \tilde{\mathbf{X}}^{\mathbf{w}}\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}) = \lambda, & \hat{\beta}_j > 0; \\ |\tilde{\mathbf{x}}'_j(Y - \tilde{\mathbf{X}}^{\mathbf{w}}\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}})| \leq \lambda, & \hat{\beta}_j = 0, \end{cases}$$

because $\hat{\beta}_j$'s are eigenvalues which are non-negative. Since $S_4 \in S_1$ contains variables of nonzero estimates, by the Karush-Kuhn-Tucker condition and (A.3), each component of $|\mathbf{f}_4|$ is 1. Hence, $\| \mathbf{f}_4 \|^2 = |S_4| = q_4$. Since $|S| = q$, $S_3 = S_1 \cap S$, we have $|S_3| \leq |S| = q$. So $q_1 = |S_1| = |S_3| + |S_4| \leq q + \| \mathbf{f}_4 \|^2 \Longrightarrow \| \mathbf{f}_4 \|^2 \geq q_1 - q$. Then by (A.6) and the property of $\mathbf{Q}_{41}$,

$$\begin{aligned}\| \mathbf{v}_{14} \|^2 &= \frac{\lambda^2}{p^2}\mathbf{f}_4'\mathbf{Q}_{41}(\boldsymbol{\Sigma}_{11}^{-1/2})'\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{Q}_{41}'\mathbf{f}_4 = \frac{\lambda^2}{p^2} \| \boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{v} \|^2 \\ &\geq \frac{\lambda^2 \| \mathbf{v} \|^2}{n\phi_{max}(|S_1|)} = \frac{\lambda^2\mathbf{f}_4'\mathbf{Q}_{41}\mathbf{Q}_{41}'\mathbf{f}_4}{p^2\phi_{max}(|S_1|)} = \frac{\lambda^2 \| \mathbf{f}_4 \|^2}{p^2\phi_{max}(|S_1|)},\end{aligned}$$

where the inequality follows (A.1) by setting $\mathbf{v} = \mathbf{Q}_{41}'\mathbf{f}_4$. Hence, we have

$$\| \mathbf{v}_{14} \|^2 \geq \frac{\lambda^2(q_1 - q)}{p^2\phi_{max}(|S_1|)}. \tag{A.8}$$

Next, we will establish the results in Theorem 1 in three steps.

Step 1: Establish an upper bound for $\| \mathbf{v}_{14} \|^2 + \| \mathbf{w}_2 \|^2$.

Note that $S_2 = \{j : \hat{\beta}_j = 0\}$. Hence, $\hat{\boldsymbol{\beta}}_2 = 0$ implies that $\tilde{\mathbf{X}}^{\mathbf{w}}\hat{\boldsymbol{\beta}} = \tilde{\mathbf{X}}_1^{\mathbf{w}}\hat{\boldsymbol{\beta}}_1 + \tilde{\mathbf{X}}_2^{\mathbf{w}}\hat{\boldsymbol{\beta}}_2 = \tilde{\mathbf{X}}_1^{\mathbf{w}}\hat{\boldsymbol{\beta}}_1$. From (A.3) we have, $\mathbf{f}_1\lambda = (\tilde{\mathbf{X}}_1^{\mathbf{w}})'(Y - \tilde{\mathbf{X}}^{\mathbf{w}}\hat{\boldsymbol{\beta}}) = (\tilde{\mathbf{X}}_1^{\mathbf{w}})'(Y - \tilde{\mathbf{X}}_1^{\mathbf{w}}\hat{\boldsymbol{\beta}}_1) \Longrightarrow (\tilde{\mathbf{X}}_1^{\mathbf{w}})'\tilde{\mathbf{X}}_1^{\mathbf{w}}\hat{\boldsymbol{\beta}}_1 = (\tilde{\mathbf{X}}_1^{\mathbf{w}})'Y - \mathbf{f}_1\lambda$. Since $Y = \tilde{\mathbf{X}}^{\mathbf{w}}\boldsymbol{\beta} = \tilde{\mathbf{X}}_1^{\mathbf{w}}\boldsymbol{\beta}_1 + \tilde{\mathbf{X}}_2^{\mathbf{w}}\boldsymbol{\beta}_2$,

$$\begin{aligned}(\tilde{\mathbf{X}}_1^{\mathbf{w}})'\tilde{\mathbf{X}}_1^{\mathbf{w}}\hat{\boldsymbol{\beta}}_1 &= (\tilde{\mathbf{X}}_1^{\mathbf{w}})'\tilde{\mathbf{X}}_1^{\mathbf{w}}\boldsymbol{\beta}_1 + (\tilde{\mathbf{X}}_1^{\mathbf{w}})'\tilde{\mathbf{X}}_2^{\mathbf{w}}\boldsymbol{\beta}_2 - \mathbf{f}_1\lambda, \\ \mathcal{D}_1\hat{\boldsymbol{\beta}}_1 &= \mathcal{D}_1\boldsymbol{\beta}_1 - \mathbf{f}_1\lambda, \\ \mathcal{D}_1^{-1}\mathbf{f}_1\lambda &= \boldsymbol{\beta}_1 - \hat{\boldsymbol{\beta}}_1.\end{aligned}$$

Now, $(\tilde{\mathbf{X}}_2^{\mathbf{w}})'(Y - \tilde{\mathbf{X}}^{\mathbf{w}}\hat{\boldsymbol{\beta}}) = (\tilde{\mathbf{X}}_2^{\mathbf{w}})'\mathbf{y} - (\tilde{\mathbf{X}}_2^{\mathbf{w}})'\tilde{\mathbf{X}}^{\mathbf{w}}\hat{\boldsymbol{\beta}} = (\tilde{\mathbf{X}}_2^{\mathbf{w}})'\tilde{\mathbf{X}}_1^{\mathbf{w}}\boldsymbol{\beta}_1 + (\tilde{\mathbf{X}}_2^{\mathbf{w}})'\tilde{\mathbf{X}}_2^{\mathbf{w}}\boldsymbol{\beta}_2 - (\tilde{\mathbf{X}}_2^{\mathbf{w}})'\tilde{\mathbf{X}}_1^{\mathbf{w}}\hat{\boldsymbol{\beta}}_1 = \mathcal{D}_2\boldsymbol{\beta}_2$. By the Karush-Kuhn-Tucker condition, $|(\tilde{\mathbf{X}}_2^{\mathbf{w}})'(\mathbf{y} - \tilde{\mathbf{X}}^{\mathbf{w}}\hat{\boldsymbol{\beta}})|$ is bounded above componentwise by λ, then $|\mathcal{D}_2\boldsymbol{\beta}_2|$ is bounded above componentwise by λ. Moreover, since $\mathcal{D}_2\boldsymbol{\beta}_2$ is positive in our case, $\mathcal{D}_2\boldsymbol{\beta}_2$ is bounded above componentwise by λ. In other words, $W_i^2\beta_{2i} \leq \lambda$ for $i = 1, \cdots, q_2$. Therefore,

$$\| \mathbf{w}_2 \|^2 = \| \mathcal{W}_2\boldsymbol{\beta}_2 \|^2 = \boldsymbol{\beta}_2'\mathcal{W}_2'\mathcal{W}_2\boldsymbol{\beta}_2 = \boldsymbol{\beta}_2'\mathcal{D}_2\boldsymbol{\beta}_2 = \sum_{i=1}^{q_2} \beta_{2i}W_i^2\beta_{2i} \leq \sum_{i=1}^{q_2} \beta_{2i}\lambda = \| \boldsymbol{\beta}_2 \|_1 \lambda.$$

Next, we have,

$$\begin{aligned}
\mathbf{v}_{14}'(\mathbf{v}_{13}+\mathbf{v}_{14}) &= \frac{\lambda}{\sqrt{p^2}}\mathbf{f}_4'\mathbf{Q}_{41}(\mathbf{\Sigma}_{11}^{-1/2})'\left(\frac{\lambda}{\sqrt{p^2}}\mathbf{\Sigma}_{11}^{-1/2}\mathbf{Q}_{31}'\mathbf{f}_3+\frac{\lambda}{\sqrt{p^2}}\mathbf{\Sigma}_{11}^{-1/2}\mathbf{Q}_{41}'\mathbf{f}_4\right)\\
&= \frac{\lambda}{\sqrt{p^2}}\mathbf{f}_4'\mathbf{Q}_{41}(\mathbf{\Sigma}_{11}^{-1/2})'\left[\frac{\lambda}{\sqrt{p^2}}\mathbf{\Sigma}_{11}^{-1/2}\left(\mathbf{Q}_{31}'\mathbf{f}_3+\mathbf{Q}_{41}'\mathbf{f}_4\right)\right]\\
&= \frac{\lambda}{\sqrt{p^2}}\mathbf{f}_4'\mathbf{Q}_{41}(\mathbf{\Sigma}_{11}^{-1/2})'\left(\frac{\lambda}{\sqrt{p^2}}\mathbf{\Sigma}_{11}^{-1/2}\mathbf{f}_1\right)\\
&= \frac{\lambda^2}{p^2}\mathbf{f}_4'\mathbf{Q}_{41}\mathbf{\Sigma}_{11}^{-1}\mathbf{f}_1=\lambda^2\mathbf{f}_4'\mathbf{Q}_{41}\mathcal{D}_1^{-1}\mathbf{f}_1\\
&= \lambda\mathbf{f}_4'\mathbf{Q}_{41}(\boldsymbol{\beta}_1-\hat{\boldsymbol{\beta}}_1)=\lambda\mathbf{f}_4'(\boldsymbol{\beta}_4-\hat{\boldsymbol{\beta}}_4),
\end{aligned}$$

where the second to the last equality holds by $\mathcal{D}_1^{-1}\mathbf{f}_1\lambda=\boldsymbol{\beta}_1-\hat{\boldsymbol{\beta}}_1$. In our case, $\mathbf{f}_4$ is a vector of 1's and $\hat{\boldsymbol{\beta}}_4\geq 0$ componentwise. So $\mathbf{f}_4'\hat{\boldsymbol{\beta}}_4\geq 0$ implies that $\mathbf{v}_{14}'(\mathbf{v}_{13}+\mathbf{v}_{14})\leq \lambda\mathbf{f}_4'\boldsymbol{\beta}_4$. Combining $\mathbf{v}_{14}'(\mathbf{v}_{13}+\mathbf{v}_{14})$ and $\parallel \mathbf{w}_2\parallel^2$, we have

$$\parallel \mathbf{v}_{14}\parallel^2+\mathbf{v}_{14}'\mathbf{v}_{13}+\parallel \mathbf{w}_2\parallel^2\leq\parallel \boldsymbol{\beta}_2\parallel_1\lambda+\mathbf{f}_4'\boldsymbol{\beta}_4\lambda.$$

By the definitions of sets A_i in Table A.1, we have

$$\parallel \boldsymbol{\beta}_2\parallel_1+\mathbf{f}_4'\boldsymbol{\beta}_4=\parallel \boldsymbol{\beta}_2\parallel_1+\parallel \boldsymbol{\beta}_4\parallel_1=\parallel \boldsymbol{\beta}_5\parallel_1+\parallel \boldsymbol{\beta}_0\parallel_1\leq\parallel \boldsymbol{\beta}_5\parallel_1.$$

Hence, $\parallel \mathbf{v}_{14}\parallel^2+\parallel \mathbf{w}_2\parallel^2\leq\parallel \boldsymbol{\beta}_5\parallel_1\lambda+(-\mathbf{v}_{14})'\mathbf{v}_{13}\leq\parallel \boldsymbol{\beta}_5\parallel_1\lambda+\parallel \mathbf{v}_{14}\parallel\parallel \mathbf{v}_{13}\parallel$, where the last inequality is obtained by the Cauchy-Schwarz inequality.

Again by the Karush-Kuhn-Tucker condition, since $S_3\in S_1$ contains variables of nonzero estimates, each component of $|\mathbf{f}_3|$ is 1. So $\parallel \mathbf{f}_3\parallel^2=|S_3|=q_3$. By the property of $\mathbf{Q}_{31}$, we have

$$\begin{aligned}\| \mathbf{v}_{13} \|^2 &= \frac{\lambda^2}{p^2}\mathbf{f}_3'\mathbf{Q}_{31}\mathbf{\Sigma}_{11}^{-1}\mathbf{Q}_{31}'\mathbf{f}_3 = \frac{\lambda^2}{p^2} \| \mathbf{\Sigma}_{11}^{-1/2}\mathbf{v} \|^2 \\ &\leq \frac{\lambda^2 \| \mathbf{v} \|^2}{p^2\phi_{min}(|S_1|)} = \frac{\lambda^2\mathbf{f}_3'\mathbf{Q}_{31}\mathbf{Q}_{31}'\mathbf{f}_3}{p^2\phi_{min}(|S_1|)} = \frac{\lambda^2 \| \mathbf{f}_3 \|^2}{p^2\phi_{min}(|S_1|)} \\ &= \frac{\lambda^2|S_3|}{p^2\phi_{min}(|S_1|)},\end{aligned}$$

where the inequality follows (A.1) by setting $\mathbf{v} = \mathbf{Q}_{31}'\mathbf{s}_3$. Therefore, we have

$$\| \mathbf{v}_{14} \|^2 + \| \mathbf{w}_2 \|^2 \leq \| \boldsymbol{\beta}_5 \|_1 \lambda + \| \mathbf{v}_{14} \| \left(\frac{\lambda^2|S_3|}{p^2\phi_{min}(S_1)}\right)^{1/2}. \tag{A.9}$$

Define,

$$B_1 = \left(\frac{q\lambda^2}{p^2\phi^*}\right)^{1/2}, \qquad B_2 = \left(\frac{q\lambda^2}{p^2\phi_*}\right)^{1/2}, \qquad B_2^2 = CB_1^2,$$

where $\phi_* = \min_{m\leq p}\phi_{min}(m)$ and $\phi^* = \max_{m\leq p}\phi_{max}(m)$.

Step 2: Establish (2.12).

Assume S_1 contains all labels j for nonzero β_j:

$$S_1 = \{j : \hat{\beta}_j \neq 0 \text{ or } j \in S\}. \tag{A.10}$$

In this case, $S_5 = \varnothing$. So $\| \boldsymbol{\beta}_5 \|_1 = 0$, $S_3 = S$, and thus $|S_3| = q \leq q_1$. Because $\left(\frac{\lambda^2|S_3|}{p^2\phi_{min}(S_1)}\right)^{1/2} \leq B_2$ and $\| \mathbf{w}_2 \|^2 \geq 0$, together with (A.9), we have that $\| \mathbf{v}_{14} \|^2 + \| \mathbf{w}_2 \|^2 \leq \| \mathbf{v}_{14} \| B_2$, which implies $\| \mathbf{v}_{14} \|^2 \leq \| \mathbf{v}_{14} \| B_2$, and $\| \mathbf{v}_{14} \| \leq B_2$. Combining with (A.8), we have,

$$\begin{aligned}\frac{(q_1-q)\lambda^2}{p^2\phi^*} &\leq \frac{\lambda^2(q_1-q)}{p^2\phi_{max}(|S_1|)} \leq \| \mathbf{v}_{14} \|^2 \leq B_2^2, \\ (q_1-q) &\leq \frac{q\phi^*}{\phi_*}, \\ q_1 &\leq \frac{\phi^*}{\phi_*}q + q = Cq + q = r_1 q,\end{aligned}$$

where r_1 is defined in (2.11). Under assumption (A.10), S_1 is taken as the largest possible set which contains $\tilde{q}$ elements. In general, S_1 doesn't necessarily select all the variables with nonzero coefficients in the true model. Hence,

$$\hat{q} = q_1 = |S_1| \leq \tilde{q} = \#\{j : \hat{\beta}_j \neq 0 \text{ or } j \in S\} \leq r_1 q,$$

which is (2.12).

Step 3: Establish (2.13) and (2.14).

By Lemma 2, we have that

$$\| \boldsymbol{\beta}_5 \|_1^2 \leq \frac{\| \tilde{\mathbf{X}}_5^{\mathbf{w}} \boldsymbol{\beta}_5 \|^2 |S_5|}{p^2 \phi_{min}(|S_5|)} \leq \frac{\| \tilde{\mathbf{X}}_5^{\mathbf{w}} \boldsymbol{\beta}_5 \|^2 q}{p^2 \phi_*},$$

because by Table A.1 $|S_3| + |S_5| = |S| = q \implies |S_5| \leq q$ and $|S_3| \leq q$. Note that $S_5 \subseteq S_2$,

$$\| \tilde{\mathbf{X}}_5^{\mathbf{w}} \boldsymbol{\beta}_5 \|^2 \leq \| \tilde{\mathbf{X}}_2^{\mathbf{w}} \boldsymbol{\beta}_2 \|^2 = \| \mathbf{w}_2 \|^2 .$$

Combining the above two inequalities,

$$\| \boldsymbol{\beta}_5 \|_1 \lambda \leq \left(\frac{\| \tilde{\mathbf{X}}_5^{\mathbf{w}} \boldsymbol{\beta}_5 \|^2 q\lambda}{p^2 \phi_*} \right)^{\frac{1}{2}} \leq \left(\frac{\| \mathbf{w}_2 \|^2 q\lambda}{p^2 \phi_*} \right)^{\frac{1}{2}} \leq \| \mathbf{w}_2 \| B_2.$$

By the Cauchy-Schwarz inequality, $\| \mathbf{v}_{14} \| B_2 \leq \| \mathbf{v}_{14} \|^2 + \frac{B_2^2}{4}$. So based on (A.9) we have

$$\| \mathbf{v}_{14} \|^2 + \| \mathbf{w}_2 \|^2 \leq \| \mathbf{v}_{14} \|^2 + \frac{B_2^2}{4} + \| \mathbf{w}_2 \| B_2,$$

$$\| \mathbf{w}_2 \|^2 \leq \frac{B_2^2}{4} + \| \mathbf{w}_2 \| B_2.$$

One can easily show that $x^2 \leq c + 2bx$ implies $x^2 \leq (b + \sqrt{b^2 + c})^2 \leq 2c + 4b^2$. Setting $x = \| \mathbf{w}_2 \|, c = \frac{B_2^2}{4}, 2b = B_2$, we obtain the result in (2.13),

$$\| \mathbf{w}_2 \|^2 \leq \frac{B_2^2}{2} + B_2^2 = \frac{3B_2^2}{2} = \frac{3C}{2} B_1^2 = r_2 \left(\frac{q\lambda^2}{p^2 \phi^*} \right), \tag{A.11}$$

where r_2 is defined in (2.11).

By the Sparse Riesz Condition, $\phi_{min}(|S_5|) \leq \frac{\|\tilde{\mathbf{X}}_5^{\mathbf{w}}\boldsymbol{\beta}_5\|^2}{p^2\|\boldsymbol{\beta}_5\|^2} \Longrightarrow \| \ \boldsymbol{\beta}_5 \ \|^2 \leq \frac{\|\tilde{\mathbf{X}}_5^{\mathbf{w}}\boldsymbol{\beta}_5\|^2}{p^2\phi_{min}(|S_5|)}$. Since $\| \ \tilde{\mathbf{X}}_5^{\mathbf{w}}\boldsymbol{\beta}_5 \ \|^2 \leq \| \ \mathbf{w}_2 \ \|^2$, we have $\| \ \boldsymbol{\beta}_5 \ \|^2 \leq \frac{\|\mathbf{w}_2\|^2}{p^2\phi_*}$, which directly gives the result in (2.14) after combining with (A.11): $\| \ \boldsymbol{\beta}_5 \ \|^2 \leq r_2 \left(\frac{q\lambda^2}{p^4\phi^*\phi_*} \right)$. □

Appendix C

Proof of Theorem 2. Since we have shown that in model (2.4), the covariates $\mathbf{Z}$ satisfy the Sparse Riesz Condition, it follows directly from the result (2.14) of Theorem 1,

$$\sum_{j \in S} |\beta_j|^2 \mathbb{1}_{\{\hat{\beta}_j = 0\}} \leq r_2 \frac{q\lambda^2}{\phi^*\phi_* p^4} = 1.5 \frac{\phi^*}{\phi_*} \frac{q\lambda^2}{\phi^*\phi_* p^4} = \frac{1.5q\lambda^2}{\phi_*^2 p^4} \leq \frac{1.5q\lambda^2}{u^4},$$

because $0 < \frac{u^2}{p^2} \leq \phi_* \leq \phi^* \leq \frac{1}{p^2} < \infty$. Hence, $\forall j \in \hat{S}^\lambda$, $\beta_j > \sqrt{1.5q}\lambda/u^2$. By definition of S_{small}, we can conclude that $(S \setminus S_{small}) \subseteq \hat{S}^\lambda$, which verifies (2.16).

By Lemma 1 of Meinshausen and Bühlmann (2006), a variable $j \notin S$ is in the selected set $\hat{S}^\lambda$ only if

$$|\mathbf{z}_j'(Y - \tilde{\mathbf{X}}^{\mathbf{w}}\hat{\boldsymbol{\beta}}^{-j})| \geq \lambda, \tag{A.12}$$

where $\hat{\boldsymbol{\beta}}^{-j}$ is the solution to (2.6) under the constraint of $\hat{\beta}_j^{-j} = 0$. We can rewrite the left-hand side as

$$|\tilde{\mathbf{x}}_j' Y - \tilde{\mathbf{x}}_j'\tilde{\mathbf{X}}^{\mathbf{w}}\hat{\boldsymbol{\beta}}^{-j}| = |\tilde{\mathbf{x}}_j'\tilde{\mathbf{X}}^{\mathbf{w}}\boldsymbol{\beta} - \tilde{\mathbf{x}}_j'\tilde{\mathbf{X}}^{\mathbf{w}}\hat{\boldsymbol{\beta}}^{-j}| = |\tilde{\mathbf{x}}_j'\tilde{\mathbf{x}}_j\beta_j - \tilde{\mathbf{x}}_j'\tilde{\mathbf{x}}_j\hat{\beta}_j^{-j}| = 0.$$

The second equality is due to the orthogonal property of $\tilde{\mathbf{X}}^{\mathbf{w}}$ in (2.5). The last equality is because $\beta_j = 0$ and $\hat{\beta}_j^{-j} = 0$. Hence, the condition (A.12) will never be satisfied because $\lambda > 0$ which means that $\hat{S}^\lambda$ contains only variables in S. This completes the proof. □

Appendix D

In this section, we explain how a matrix in the Grassmann Manifold can be written as an eigen-decomposition solution, the link between the Grassmann Manifold and a dimension reduction matrix, and how to form an equivalent form for the purpose of obtaining a sparse estimate.

Grassmann Manifold

Suppose that a $p \times k$ matrix V is in the Grassmann Manifold with rank k.

First, we can extend it to a nonsingular $p \times p$ matrix $V^* = (V, V^\perp)$.

Second, applying singular value decomposition on V^*, we have $V^* = L\Lambda R'$ where the columns of $p \times p$ matrix L and $p \times p$ matrix R are corresponding left and right eigenvectors of V^*, and $\Lambda = (\Lambda_k, \Lambda_{p-k})$ is a diagonal matrix with non-zero singular values of V^* being its diagonal elements. Let

$$G = L\Lambda^{-2}L', \tag{A.13}$$

we have $V^{*'}GV^* = R\Lambda' L' L\Lambda^{-2}L'L\Lambda R' = \mathbf{I}_p$.

Let $\mathbf{M} = GV^*DV^{*'}G$, where G is found by (A.13) and D can be any diagonal matrix with diagonal terms being $\rho_1 >, \cdots, > \rho_p > 0$, then V^*, G, and $\mathbf{M}$ will satisfy the basic eigenvalue decomposition as

$$\mathbf{M}V^* = GV^*D. \tag{A.14}$$

Hence, for any $p \times k$ matrix V in the Grassmann Manifold with rank k, the columns of V are the eigenvectors of a symmetric and positive definite matrix $\mathbf{M}$, whose corresponding eigenvalues are $\rho_1, \cdots, \rho_k$, as long as $\rho_1 > \cdots > \rho_k > 0$. Typically, we would choose D so that $\rho_1 > \cdots, > \rho_p > 0$, and all the eigenvalues are bounded below by 0 and above by ∞.

Link between the Grassmann Manifold and a dimension reduction matrix

We assume that the CS is sparse, which means that only some variables are related to the response. Note that sparsity is not generally transformed from one scale to another scale. That is, if a model is sparse in the **X**-scale, it does not mean it is sparse in the **Z**-scale. Thus our discussion will focus on the original **X**-scale.

Suppose that the method specific kernel matrix $\mathbf{M}$ is obtained based on the original predictors $\mathbf{X}$. Basis directions are found by conducting the generalized eigenvalue problem of the form

$$\mathbf{M}V^* = GV^*D, \tag{A.15}$$

where columns of $V^* = (v_1, \cdots, v_p)$ are eigenvectors of $\mathbf{M}$ satisfying $V^{*'}GV^* = \mathbf{I}_p$ and D is a diagonal matrix with eigenvalues $\rho_1, \rho_2, \cdots, \rho_p$ of $\mathbf{M}$ in descending order. If the structural dimension of the CS is k, then the first k orthogonal eigenvectors, say, V, form an estimate of the central subspace.

Thus, in the Grassmann Manifold, we construct $\mathbf{M}$, D and G from V, while in dimension reduction, we have $\mathbf{M}$ and G to deduce V and D.

Our theoretical result in Section 2.2 in the paper requires that all eigenvalues are bounded below from 0 and above from ∞ to satisfy the Sparse Riesz Condition. This is not satisfied when we have a $p \times p$ dimension reduction matrix with k nonzero eigenvalues with $k < p$.

To fix this, from (A.15), we have $\mathbf{M} = \mathbf{M}V^*V^{*-1} = GV^*DV^{*'}G$. For some positive constant δ, let $\mathbf{M}_\delta = (\mathbf{M} + \delta G) = GV(D + \delta\mathbf{I}_p)V'G$. We will have similar eigenvalue decomposition on $\mathbf{M}_\delta$ as

$$\mathbf{M}_\delta V^* = GV^*D_\delta, \tag{A.16}$$

where D_δ is a diagonal matrix with eigenvalues $\rho_1 + \delta, \rho_2 + \delta, \cdots, \rho_p + \delta$. Since the eigenvectors of $\mathbf{M}_\delta$ are same as these of $\mathbf{M}$, we can work on $\mathbf{M}_\delta$ to estimate the basis directions of the central subspace. After some algebra, it requires the ratios of maximum and minimum eigenvalues for the matrix $\mathbf{M}_\delta$ to be bounded below from 0 and above from ∞. Let $m_1 \geq \cdots \geq m_p$ be the eigenvalues of $\mathbf{M}_\delta$, we have

$$m_1 = \rho_1 + \delta, \qquad m_p = \rho_p + \delta.$$

Therefore, by choosing $\delta > 0$, the eigenvalues of $\mathbf{M}_\delta$ are bounded below from 0 and above from ∞ so that the Sparse Riesz Condition will be satisfied. Under this condition, Theorem 3 applies to the dimension reduction matrix $\mathbf{M}_\delta$.

Appendix E

Proof of Theorem 3. Since the optimization problem (2.20) was developed based on the generalized eigenvalue problem (A.16), we again have an "error-free" model in the population: $\tilde{Y} = \tilde{\mathbf{X}}\boldsymbol{\beta}$. If the Sparse Riesz Condition is satisfied, following (5.8) in Zhang and Huang (2008), we have

$$\| \mathbf{v}_{14} \|^2 + \| \mathbf{w}_2 \|^2 \leq \| \boldsymbol{\beta}_5 \|_1 \lambda + \| \mathbf{v}_{14} \| \left(\frac{\lambda^2 |S_3|}{p\phi^Z_{min}(S_1)} \right)^{1/2}. \tag{A.17}$$

where $\mathbf{v}_{14}$, $\mathbf{w}_2$, S_1, and S_2 are defined in the same ways as in the proof of Theorem 1. Following the same steps as in the proof of Theorem 1, we are able to obtain the following two upper bounds:

$$\hat{q}(\lambda) \leq \tilde{q} = \#\{j : \hat{\beta}^\lambda_j \neq 0 \text{ or } j \in S\} \leq r_1 q, \tag{A.18}$$

and

$$\sum_{j \in S} |\beta_j|^2 \mathbb{1}_{\{\hat{\beta}^\lambda_j = 0\}} \leq r_2 \frac{q\lambda^2}{\phi^* \phi_* p^2}, \tag{A.19}$$

where r_1, r_2, ϕ^*, and ϕ_* are defined in (2.11).

The rest of the proof follows mostly from the steps of Meinshausen and Bühlmann (2010) in the proof of their Theorem 2.

Lemma 3. Define C by $(1 + C)q + 1 = \bar{C}q^2$ and assume $q \geq 3$. Let weights W_k be generated randomly in $[u, 1]$ as in (2.20), and let $\tilde{X}^{\mathbf{w}}_k = \tilde{X}^{\mathbf{w}} W_k$ for $k = 1, \cdots, p$ be the corresponding rescaled predictor variables. For $u^2 = \nu\phi_{min}(\bar{C}q^2)/\bar{C}q^2$ with $\nu > 0$, it holds under assumption (2.21) for all realizations W_k that

$$\frac{\phi^w_{max}(\bar{C}q^2)}{\phi^w_{min}(\bar{C}q^2)} \leq \frac{3C}{\kappa\sqrt{\nu}}. \tag{A.20}$$

Proof. We can follow exactly the same steps as in the proof of the Lemma 3 in Meinshausen and Bühlmann (2010). The only remark we need to make is that C in our notation is their $\bar{C}$ while our $\bar{C}$ is their C. Since the steps are similar, we omit the details.

Lemma 4. Let $\hat{S}^{\lambda,w}$ be the set $\{k : \hat{\beta}_k^{\lambda,w}\}$ of selected variables of the randomized lasso with $u \in (0,1]$ and randomly sampled weights $\mathbf{w}$. Suppose that $u^2 \geq (3/\kappa)^2\phi_{min}(\bar{C}q^2)/\bar{C}q^2$, we can show that

$$|\hat{S}^{\lambda,w} \cup S| \leq \bar{C}q^2 \text{ and } (S \setminus S_{small}) \subseteq \hat{S}^{\lambda,w}, \tag{A.21}$$

where $S_{small} = \{k : \beta_k \leq \sqrt{1.5}\bar{C}q^{3/2}\lambda\}$.

Proof. The proof of this lemma follows from Theorem 1 and Lemma 4 of Meinshausen and Bühlmann (2010). With Remark 2 in Zhang and Huang (2008), the equivalent condition of (2.21) requires the existence of some $C > 0$ such that

$$\frac{\phi_{max}((1+C)q+1)}{\phi_{min}((1+C)q+1)} < C,$$

where C is defined in (2.11). Hence, for all realizations W_i, if $u^2 \geq (3/\kappa)^2\phi_{min}(\bar{C}q^2)/\bar{C}q^2$, by Lemma 3, $\frac{\phi^w_{max}(\bar{C}q^2)}{\phi^w_{min}(\bar{C}q^2)}$ is bounded. Therefore, (A.18) and (A.19) hold which give us

$$|\hat{S}^{\lambda,w} \cup S| \leq (1+C)q \leq \bar{C}q^2,$$

and

$$\sum_{j\in S} |\beta_j|^2 \mathbb{1}_{\{\hat{\beta}_j^{\lambda}=0\}} \leq (1.5C^2)q\lambda^2 \leq (\sqrt{1.5}\bar{C}q^{3/2}\lambda)^2, \tag{A.22}$$

where the first inequality uses the fact that $1/\phi^*\phi_* \leq C$ and the second inequality follows from $C \leq \bar{C}q$. Accordingly, (A.22) is equivalent to the second part of (A.21).

Lemma 5 Let p_w be the probability of choosing weight u for each variable and $1-p_w$ the probability of choosing weight 1. Define $\tilde{p} = p_w(1-p_w)^{\bar{C}q^2}$ and let $\hat{\Pi}_k^\lambda$ be the probability of variable k being in the selected subset $\hat{S}^{\lambda,w}$ with respect to random sampling of the weights w. Under assumptions of Theorem 3, for any $\lambda \geq \inf\{\lambda : r_1 q + 1 \leq p\}$,

$$\max_{k \in N}(\hat{\Pi}_k^\lambda) < 1 - \tilde{p}, \tag{A.23}$$

$$\min_{k \in S \setminus S_{small}}(\hat{\Pi}_k^\lambda) \geq 1 - \tilde{p}, \tag{A.24}$$

where $S_{small} = \{k : \beta_k \leq \sqrt{1.5}\bar{C}q^{3/2}\lambda\}$.

Proof. Following Meinshausen and Bühlmann (2006), a variable $j \notin S$ is in the selected set $\hat{S}^{\lambda,w}$ only if

$$|\tilde{\mathbf{x}}_j'(Y - \tilde{\mathbf{X}}^w \hat{\boldsymbol{\beta}}^{-j})| \geq \lambda, \tag{A.25}$$

where $\hat{\boldsymbol{\beta}}^{-j}$ is the solution to (2.20) under the constraint of $\hat{\beta}_j^{-j} = 0$. Using Lemma 5 of Meinshausen and Bühlmann (2010) and Lemma 4 above, we can show that the left-hand side of (A.25) is bounded by $\| \; ((\tilde{\mathbf{X}}_B^{\mathbf{w}})'\tilde{\mathbf{X}}_B^{\mathbf{w}})^{-1}(\tilde{\mathbf{X}}_B^{\mathbf{w}})'\tilde{\mathbf{X}}_j^{\mathbf{w}} \; \|_1 \; \lambda \leq 2^{-1/4}\lambda < \lambda$ with probability greater than or equal to $p_w(1-p_w)^{\bar{C}q^2}$, where set $B = \hat{S}^{\lambda,w} \cup S$ and the first inequality is based on Lemma 5 of Meinshausen and Bühlmann (2010). This leads to the result (A.23). The consequence of Lemma 4 directly yields (A.24). Since our Lemma 5 is equivalent to Theorem 3, the proof of Theorem 3 is complete. $\square$

Appendix F

In Section 2.2.2 in the paper, we used $\delta > 0$ to have $\mathbf{M}_\delta = (\mathbf{M} + \delta G)$ for GMSE and SGMSE in order to satisfy the Sparse Riesz Condition. In this section, we investigate the choice of δ in GMSE and SGMSE. Our empirical evidences show that the choices of the positive constant δ have little effect on the final estimates. We ran simulations on the three models in Section 3.2.1 in the paper, with different choices of $\delta = 0.001, 0.01, 0.1, 0.5$. For the same model and method, varying δ does not greatly change the results. In addition, for all δ values, SGMSE is improved over GMSE. It seems that a smaller value of δ is preferable because it results in a lower false positive rate. Hence, a rule of thumb for appropriate δ to use in GMSE is between 0.001 to 0.01.

		$\delta = 0.001$		$\delta = 0.01$		$\delta = 0.1$		$\delta = 0.5$	
		TPR	**FPR**	**TPR**	**FPR**	**TPR**	**FPR**	**TPR**	**FPR**
SIR	GMSE SIR	1.000	0.073	1.000	0.023	1.000	0.018	1.000	0.025
	SGMSE SIR	1.000	0.003	1.000	0.000	1.000	0.000	1.000	0.003
PHD	GMSE PHD	1.000	0.268	1.000	0.274	1.000	0.380	1.000	0.350
	SGMSE PHD	1.000	0.085	1.000	0.095	1.000	0.140	1.000	0.154
SAVE	GMSE SAVE	1.000	0.283	1.000	0.263	1.000	0.304	1.000	0.451
	SGMSE SAVE	1.000	0.063	1.000	0.058	1.000	0.069	1.000	0.151

Table A.2: For different δ, TPR and FPR are computed among 100 replicates for SIR, PHD and SAVE models using GMSE and SGMSE in Section 3.2.1.

Appendix G

In this appendix, we compare GMSE to the sparse sufficient dimension reduction (SSDR) method (Li 2007), and SGMSE to stable SSDR (SSSDR).

Li's SSDR starts with an equivalent formulation of eigen-decomposition as

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \sum_{i=1}^{p} \| G^{-1}m_i - \boldsymbol{\alpha}\boldsymbol{\beta}^T m_i \|_G^2 + \lambda_2 tr(\boldsymbol{\beta}^T G \boldsymbol{\beta}) + \sum_{j=1}^{k} \lambda_{1j} \sum_{h=1}^{p} |\beta_{jh}|, \tag{A.26}$$

subject to $\boldsymbol{\alpha}^T G \boldsymbol{\alpha} = \mathrm{I}$, where the norm is the inner product with with respect to G. In (A.26), G takes the form of the covariance matrix Σ_x of $\mathbf{X}$, the values of m_i are columns of the square root of the method-specific dimension reduction matrix $\mathbf{M}$ and $\boldsymbol{\beta}$ is a $p \times k$ matrix of which the columns are the basis directions of the central subspace. The λ_2 and λ_{1j}'s are the tuning parameters corresponding to the L^1 and L^2 penalties. Then, Li (2007) showed that the optimization problem (A.26) can be solved in an alternative way by solving k independent LASSO problems for a given $\boldsymbol{\alpha}$ as:

$$\hat{\boldsymbol{\beta}}_j = \min_{\boldsymbol{\beta}_j} \left\{ \boldsymbol{\beta}_j^T (\mathbf{M} + \lambda_2 G) \boldsymbol{\beta}_j - 2\alpha_j^T M \boldsymbol{\beta}_j + \lambda_{1j} \sum_{h=1}^{p} |\beta_{jh}| \right\}, \tag{A.27}$$

subject to $\boldsymbol{\alpha}^T G \boldsymbol{\alpha} = \mathrm{I}$. For given $\boldsymbol{\beta}_j$'s, solving $\boldsymbol{\alpha}$ is just a usual OLS problem. Li (2007) also showed that (A.27) can be transformed into an equivalent problem as

$$\hat{\boldsymbol{\beta}}_j = \min_{\boldsymbol{\beta}_j} \left\{ \| \; u^* - m^* \boldsymbol{\beta}_j \; \|^2 + \lambda_{1j} \sum_{h=1}^{p} |\beta_{jh}| \right\}, \tag{A.28}$$

where,

$$m^* = \begin{pmatrix} \mathbf{M}^{1/2} \\ \sqrt{\lambda_2} G^{1/2} \end{pmatrix}, \qquad u^* = \begin{pmatrix} \mathbf{M}^{1/2} \alpha_j \\ 0 \end{pmatrix}.$$

In the above SSDR method, by introducing subsampling and random weight we also develop a Stable SSDR, which we call SSSDR.

However, the introduction of λ_2 in SSDR is only for the uniqueness of the eigenvectors. For this reason, Li's algorithm gives an invariant result for any $\lambda_2 > 0$. However, under this formulation, the Sparse Riesz Condition is not always satisfied, even though λ_2 is a nuisance parameter. Thus, we are not able to prove the theoretical result, even if we believe the result holds. In addition, with an information criterion to select λ_2, it slows down the computing speed. Nevertheless, the table below shows that in our simulations the two approaches have very comparable results.

	Original		SSDR		GMSE		SSSDR		SGMSE	
	TPR	**FPR**	**TPR**	**FPR**	**TPR**	**FPR**	**TPR**	**FPR**	**TPR**	**FPR**
SIR	1.000	1.000	1.000	0.012	1.000	0.022	1.000	0.001	1.000	0.002
PHD	1.000	1.000	1.000	0.171	1.000	0.249	1.000	0.044	1.000	0.058
SAVE	1.000	1.000	1.000	0.264	1.000	0.191	1.000	0.115	1.000	0.005

Table A.3: TPR and FPR are computed over 100 replicates for SSIR, SPHD and SSAVE models using Li's (2007) SSDR, SSSDR, GMSE and SSGMSE in Section 2.2.

Appendix H

In this section, we include two simulation studies: one simulation shows the sensitivity of sliced inverse methods to the choices of H; another simulation illustrates the stability of the results using the ensemble method proposed in Section 2.3.2 in the paper. We used model (3.1) for SIR and model (3.5) for SAVE as in Section 3.1 in the paper.

Figure A.1 shows that for each fixed $H = 5, 10, 15, 20$, the results for SIR do vary but not as much as these of SAVE (left column); stable procedures show significant improvement. While results for SIR do vary, but results vary more for SAVE (right column). We now use these four different numbers of slices to develop one aggregated dimension reduction matrix as proposed in Section 2.3.2. Figure A.2 shows that the ensemble method gives better and more stable results (left column), which are further improved by our newly developed stable procedure (right column).

References

Li, L. (2007), "Sparse sufficient dimension reduction", *Biometrika*, 94, 603-613.

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

Meinshausen, N. and Bühlmann, P. (2010), "Stability selection", *Journal of the Royal Statistical Society, Series B*, 72, 417-473.

Zhang, C. H. and Huang, J. (2008), "The sparsity and bias of the LASSO selection in High-Dimensional linear regression", *The Annals of Statistics*, 36, 1567-1594.

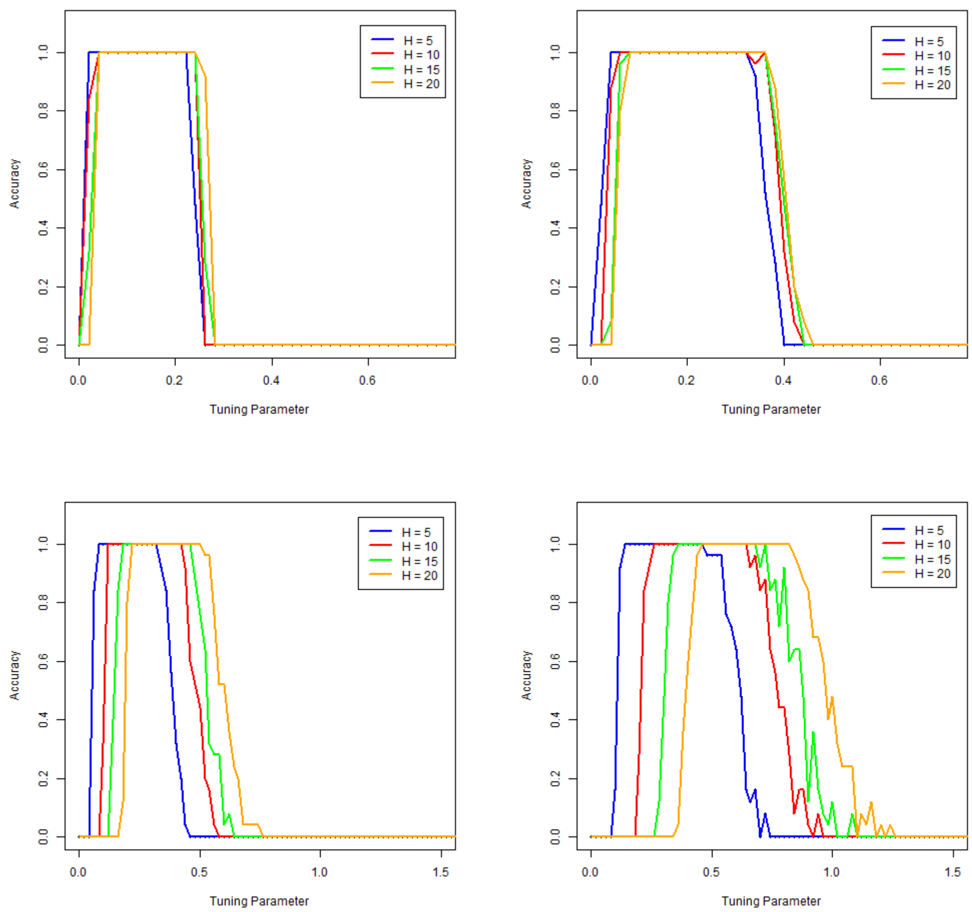

Figure A.1: Top row: SIR model; Bottom row: SAVE model. The accuracies of SED estimate (left column) and the stable SED estimate (right column) are plotted *vs* tuning parameter values for different choices of H = 5, 10, 15, 20.

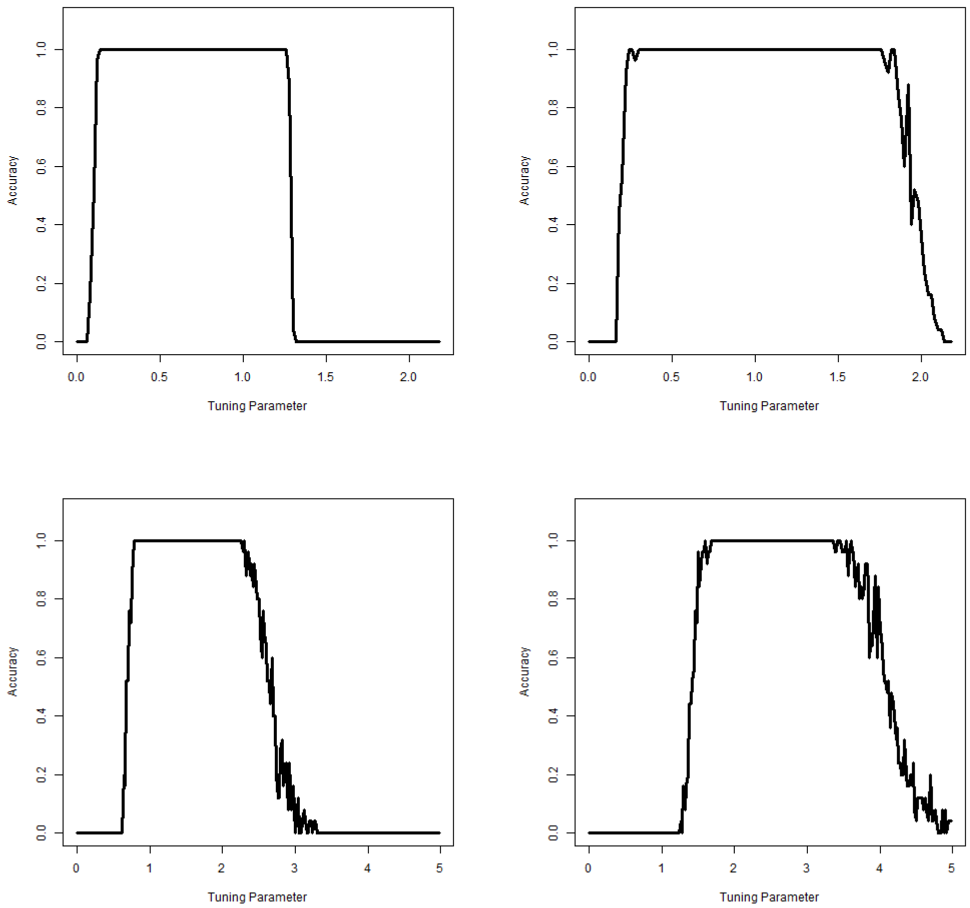

Figure A.2: Top row: SIR models (3.1); Bottom row: SAVE models (3.5). The accuracies of ensemble SED estimate (left column) and the stable ensemble SED estimate (right column) are plotted *vs* tuning parameter values.