Supplementary Materials to “Distributed Sufficient Dimension Reduction for Heterogeneous Massive Data”

Kelin Xu, Liping Zhu, Jianqing Fan

Fudan University, Renmin University of China, Princeton University

Abstract: In the Supplement, we study three additional issues. Firstly, we propose a distributed algorithm when the central subspaces at the local nodes are distinctive from each other, which takes the advantages of the low-rank structure of the kernel matrices. Secondly, we demonstrate the distributed algorithm may outperform its pooled version in the presence of heterogeneity. A bootstrap procedure is also introduced to make a fair comparison. Finally, we provide technical details, such as proofs of theorems and some useful lemmas, in this Supplement.

S1 When the Local Central Subspaces are Distinctive

Throughout the main context, we assume the central subspaces at all local nodes are identical. To be specific, in model (1.2), we assume a common
S1. WHEN THE LOCAL CENTRAL SUBSPACES ARE DISTINCTIVE

basis, $B$, is shared by all $m$ local nodes. In some situations, this is perhaps unrealistic. In this Supplement, we allow the central subspaces at the local nodes to be distinctive from each other. In symbols, we assume that:

$$F_j(Y_{i,j} | x_{i,j}) = F_j(Y_{i,j} | B_j^T x_{i,j})$$

for $i = 1, \ldots, n$, $j = 1, \ldots, m$. (S1.1)

Here, $B_j$ is a $p \times d_j$ matrix. By the very purposes of dimension reduction, $d_j$ is much smaller than $p$. Recall the definition of $\Omega_j$ in Section 2 of the main context. It follows from Li (1991) and Zhu et al. (2010) that $\text{span}(\Omega_j) \subseteq \text{span}(B_j)$ under the linearity condition. An important observation is that $\Omega_j$ is a low-rank matrix, which has at most $d_j$ nonzero eigenvalues. Taking the advantage of this low-rank structure, we can recover $\Omega_j$ from its principal eigenvectors and the associate eigenvalues. With slight abuse of notations, we let $B_j \in \mathbb{R}^{p \times d_j}$ be the $d_j$ principal eigenvectors. In addition, let $\Lambda_j \in \mathbb{R}^{d_j \times d_j}$ be a diagonal matrix with its diagonal elements being the nonzero eigenvalues of $\Omega_j$. It follows immediately that $\Omega_j = B_j \Lambda_j B_j^T$. Therefore, it suffices for a distributed algorithm to transfer the principal eigenvectors and the nonzero eigenvalues to the central node.

Our proposed distributed algorithm proceeds as follows.

Algorithm 3

1. Estimate $\Omega_j$ at the $j$th local node, which yields $\hat{\Omega}_j$. We can simply use the dense estimate $\hat{\Omega}_j$ in Algorithm 1, if all covariance matrices
are invertible, or the sparse estimate $\hat{\Omega}_j$ in Algorithm 2 otherwise.

2. Apply a certain criterion to decide the rank of $\hat{\Omega}_j$ at the $j$th local node, which yields $\hat{d}_j$. In our subsequent illustration, we simply use the maximum eigenvalue ratio criterion of [Luo et al. (2009)].

3. Apply singular value decomposition to $\hat{\Omega}_j$ to obtain its top $\hat{d}_j$ eigenvectors $\hat{B}_{a3,j} \in \mathbb{R}^{p \times \hat{d}_j}$ and the associated eigenvalues, which are the diagonal elements of $\hat{\Lambda}_{a3,j} \in \mathbb{R}^{\hat{d}_j \times \hat{d}_j}$. We approximate $\hat{\Omega}_j$ with $\hat{B}_{a3,j} \hat{\Lambda}_{a3,j} \hat{B}_{a3,j}^T$.

4. Recall that $\hat{\Lambda}_{a3,j}$ is a diagonal matrix. Pass $\hat{B}_{a3,j}$ and the diagonal elements of $\hat{\Lambda}_{a3,j}$ to the central node to form

$$\hat{T}_{a3} \overset{\text{def}}{=} m^{-1} \sum_{j=1}^{m} \hat{B}_{a3,j} \hat{\Lambda}_{a3,j} \hat{B}_{a3,j}^T \quad (S1.2)$$

The communication cost in this step is

$$(p + 1) \sum_{j=1}^{m} \hat{d}_j.$$ 

5. Apply singular value decomposition to $\hat{T}_{a3}$ to obtain the first $d_0$ top eigenvectors, which yields $\hat{B}_{a3}$. If $d_0$ is unknown, we can again apply a certain criterion, say, [Luo et al. (2009)], to decide the rank of $\hat{T}_{a3}$.

We demonstrate the finite-sample performance of the above distributed algorithm through simulations.
Example 4: We illustrate the performance of Algorithm 3 in this simulated example. We fix $p = 200$, and draw $\mathbf{x} = (X_1, \ldots, X_p)^T \in \mathbb{R}^p$ from multivariate normal distribution with mean zero and covariance matrix $\Sigma = (\rho^{k-l})_{p \times p}$. We set $\rho = 0.5$. At each local node, we generate $Y$ from the following models with equal probability $1/2$:

$$
Y = \sin(\beta_1^T \mathbf{x}) + \varepsilon,
$$

$$
Y = \exp(\beta_2^T \mathbf{x}) + \varepsilon,
$$

We generate the error term $\varepsilon$ from standard normal distribution. We set $\beta_1 = (1, 1, 0, \ldots, 0)^T \in \mathbb{R}^p$ and $\beta_2 = (0, 0, 1, 1, 0, \ldots, 0)^T \in \mathbb{R}^p$. Let $m = \{2^5, 2^6, 2^7, 2^8\}$ and $N = \{2^9p, 2^{10}p, 2^{11}p, 2^{12}p\}$. All the $N$ observations are scattered uniformly across $m$ nodes each of size $n$.

We implement Algorithm 3 for both sliced inverse regression and cumulative slicing estimation under the case when all sample covariance matrices are invertible. In other words, we use the dense estimate $\hat{\Omega}_J$.

Example 5: We generate the observations in the same way as in Example 6, except for $\rho = 0.8$ in $\Sigma = (\rho^{k-l})_{p \times p}$. We implement Algorithm 3 for both sliced inverse regression and cumulative slicing estimation under the case when not all sample covariance matrices are invertible. In other words, we use the sparse estimate $\hat{\Omega}_J$.

In the above two examples, $\mathbf{B} = (\beta_1, \beta_2) \in \mathbb{R}^{p \times 2}$. We repeat each
simulation 1000 times, and report \( \text{dist}(\hat{B}, B^*) \), \( \text{dist}(B^*, B) \) and \( \text{dist}(\hat{B}, B) \) to evaluate the performance of distributed estimates. The simulation results are summarized in Figures 4 and 5 for Examples 4 and 5, respectively.

Figures 4 and 5 deliver similar messages. In both examples, cumulative slicing estimation outperforms sliced inverse regression, and the performance of the latter depends on the number of slices.

S2  American Gut Project Revisited

We revisit the American Gut Project in Section 4 of the main context. We explore how to implement a bootstrap procedure in the presence of heterogeneity. We compare two versions of bootstrap methods. There are three steps for both procedures. In the first step, we bootstrap new observations at each local node. To implement the first version of bootstrap, we perform dimension reduction on the locally bootstrapped observations in the second step, and aggregate the dimension reduction results at all local nodes in the third step. To implement the second version of bootstrap, we pool the locally bootstrapped observations together to form a complete bootstrap sample in the second step, and perform dimension reduction on the complete bootstrap sample in the third step. We replicate the above procedures 100 times, and compare the distances (1.4) of the central sub-
spaces obtained from the bootstrap samples and the original observations. The simulation results are summarized in Figure 6 in this Supplementary Material. It can be clearly seen that, the first version of bootstrap, which in spirit corresponds to the distributed dimension reduction, is much more stable than the second version of bootstrap, which indeed yields a pooled dimension reduction. This indicates that, in the presence of heterogeneity, the distributed dimension reduction methods are perhaps more advantageous than the pooled ones.

S3 Some Useful Lemmas

We first provide some lemmas that pave the road to prove Theorems 1 and 2. Define the spectral norm $\|A\| \overset{\text{def}}{=} \lambda_1^{1/2}(A^TA)$, where $\lambda_{\text{max}}$ stands for the maximum eigenvalue.

**Lemma 5.** In addition to Conditions (C1)-(C3), we assume the sample covariance matrices $\hat{\Sigma}_j^{-1}$ are all invertible. Then there exists an absolute positive constant $C$ such that

$$\|\hat{\Omega}_{a1,j} - \Omega_j\|_{\psi_1} \leq C(p/n)^{1/2}, \text{ for } j = 1, \ldots, m.$$ 

Proof of Lemma 5. Recall that $\hat{\Omega}_{a1,j} \overset{\text{def}}{=} \hat{\Sigma}_j^{-1} \hat{M}_j \hat{\Sigma}_j^{-1}$ and $\Omega_j \overset{\text{def}}{=} \Sigma_j^{-1} M_j \Sigma_j^{-1}$. We further define $Q_1 \overset{\text{def}}{=} \|\hat{\Omega}_{a1,j} - \Sigma_j^{-1} \hat{M}_j \hat{\Sigma}_j^{-1}\|$, $Q_2 \overset{\text{def}}{=} \|\Sigma_j^{-1} (\hat{M}_j - M_j) \Sigma_j^{-1}\|$, 


and $Q_3 \overset{\text{def}}{=} \| \hat{\Sigma}_j^{-1} M_j \Sigma_j^{-1} - \Omega_j \|$. It follows immediately that $Q_1 = \| \hat{\Sigma}_j^{-1} \hat{M}_j \Sigma_j^{-1} (\Sigma_j - \hat{\Sigma}_j) \hat{\Sigma}_j^{-1} \|$ and $Q_3 = \| \Sigma_j^{-1} (\Sigma_j - \hat{\Sigma}_j) \hat{\Sigma}_j^{-1} M_j \Sigma_j^{-1} \|$. By the triangular inequality, $\| \hat{\Omega}_{a1,j} - \Omega_j \| \leq Q_1 + Q_2 + Q_3$. The Cauchy-Schwartz inequality implies immediately that $\| \hat{\Omega}_{a1,j} - \Omega_j \| \leq Q_1 + Q_2 + Q_3$.

Define $r(\Sigma_j) \overset{\text{def}}{=} \text{trace}(\Sigma_j)/\{n\lambda_{\max}(\Sigma_j)\}$, which is not greater than $p/n$. Koltchinskii and Lounici (2017) showed that, under Conditions (C1)-(C2), for any $t \geq 1$, there exists a generic constant $C_3 \geq 1$ such that

\[ \text{pr} \left( \| \hat{\Sigma}_j - \Sigma_j \| \geq C_3 \| \Sigma_j \| \max \left\{ \{r(\Sigma_j)\}^{1/2}, r(\Sigma_j), (t/n)^{1/2}, t/n \right\} \right) \leq \exp(-t). \]

This, together with Lemma 2.2.1 in Van Der Vaart and Wellner (1996), entails that

\[ \| \| \hat{\Sigma}_j - \Sigma_j \| \psi_1 \| \leq C_4 (p/n)^{1/2}. \] (A.2)

Accordingly, $C_4$ in (A.2) is related to $C_3$, which has been proved as an absolute constant in Koltchinskii and Lounici (2017). Next we study the convergence rate of $\| \hat{M}_j - M_j \|$. Both $\hat{M}_{j,a}$ (5) and $\hat{M}_{j,c}$ (6) have the same
form. On account of $|\hat{p}_{h,j} - p_{h,j}| = O_p(n^{-1/2})$, the technical details for processing $\|\hat{M}_{j,a} - M_{j,a}\|$ and $\|\hat{M}_{j,c} - M_{j,c}\|$ are thus very similar. To avoid redundancy, we only provide the details for cumulative slicing estimation in what follows. Define

$$\tilde{M}_{j,c} \equiv n^{-1} \sum_{i=1}^{n} m_{j,c}(Y_{i,j}) m_{j,c}'(Y_{i,j}).$$

By triangular inequality, $\|\hat{M}_{j,c} - M_{j,c}\| \leq \|\hat{M}_{j,c} - \tilde{M}_{j,c}\| + \|\tilde{M}_{j,c} - M_{j,c}\|$. Following similar arguments for proving (A.2), we can show that

$$\|\|\hat{M}_{j,c} - M_{j,c}\|\|_{\psi_1} \leq C_5 (p/n)^{1/2},$$

(A.3)

where $C_5$ is an absolute constant. We turn to $\|\hat{M}_{j,c} - \tilde{M}_{j,c}\|$. By Lemma 5 in Wang et al. (2021), we have

$$\Pr(\|\hat{M}_{j,c} - \tilde{M}_{j,c}\| \geq t^2 + 2c_1^{1/2} t) \leq \exp(2 + \log n + p \log 5 - Cn t^2),$$

(A.4)

where $c_1$ and $C$ are absolute constants induced by Proposition 5.10 of Vershynin (2010). Following similar arguments for proving (A.2), we can also show that

$$\|\|\hat{M}_{j,c} - \tilde{M}_{j,c}\|\|_{\psi_1} \leq C_6 / n.$$  

(A.5)

Similarly, constant $C_6$ in (A.5) is constructed from constants in (A.4), thus $C_6$ is independent of $j$. By definition, $\|\|\hat{M}_{j,c} - M_{j,c}\|\|_{\psi_1} \leq \|\|\hat{M}_{j,c} - \tilde{M}_{j,c}\|\|_{\psi_1} + \|\|\tilde{M}_{j,c} - M_{j,c}\|\|_{\psi_1}$. This, together with (A.3) and (A.5), yields
S3. SOME USEFUL LEMMAS

that \( \| \hat{M}_{j,c} - M_{j,c} \|_{\psi_1} \leq C_7 (p/n)^{1/2} \), where \( C_7 \) is an absolute positive constant. Thus far we complete the proof for cumulative slicing estimation.

With similar arguments we can deal with sliced inverse regression. In other words, \( \| \hat{M}_j - M_j \|_{\psi_1} \leq C_8 (p/n)^{1/2} \) and \( C_8 \) is independent of \( j \). This, together with (A.1) and (A.2), completes the proof of Lemma 5.

The following lemma is a direct consequence of Lemma 5.

**Lemma 6.** In addition to Conditions (C1)-(C3), we assume the sample covariance matrices \( \hat{\Sigma}_j^{-1} \)’s are all invertible. Then there exists an absolute positive constant \( C \) such that

\[
\| \text{dist}(\hat{B}_{a,1,j}, B) \|_{\psi_1} \leq C (d_0 p/n)^{1/2}, \quad \text{for } j = 1, \ldots, m.
\]

**Proof of Lemma 6** Denote the column space of \( B \) by \( \text{span}(B) \). Further denote the principal angles between \( \text{span}(\hat{B}_j) \) and \( \text{span}(B) \) by \( \Theta(\hat{B}_j, B) \overset{\text{def}}{=} (\theta_{1,j}, \theta_{2,j}, \ldots, \theta_{d_0,j})^T \). In other words, the singular values of \( \hat{B}_j^T B \) are \( \cos(\theta_{1,j}), \cos(\theta_{2,j}), \ldots, \cos(\theta_{d_0,j}) \). Then

\[
\{ \text{dist}(\hat{B}_{a,1,j}, B) \}^2 = 2 \sum_{k=1}^{d_0} \{ 1 - \cos^2(\theta_{k,j}) \} = 2 \| \sin(\Theta(\hat{B}_j, B)) \|_F^2 \quad \text{(A.6)}
\]

By Conditions (C2)-(C3) and Davis-Kahan-Theorem [Yu et al. 2015, Theorem 2], the right hand side of the above display is bounded by \( C_0 d_0^{1/2} \| \hat{\Omega}_{a,1,j} - \Omega_j \|_F \), and \( C_0 \) is an absolute constant induced by Conditions (C2) and (C3).

Therefore, \( \text{dist}(\hat{B}_{a,1,j}, B) \leq C_0 d_0^{1/2} \| \hat{\Omega}_{a,1,j} - \Omega_j \|_F \). The proof is completed by
invoking Lemma \[5\] \(\Box\)

**Lemma 7.** In addition to Conditions (C1)-(C3), we assume the sample covariance matrices \(\hat{\Sigma}_j^{-1}\)'s are all invertible. Then there exists an absolute positive constant \(C\) such that

\[
\| E(\hat{\Omega}_{a1,j} - \Omega_j) \|_F \leq C \frac{p}{n}, \text{ for } j = 1, \ldots, m.
\]

Proof of Lemma 7: As defined in Lemma 5,

\[
\hat{\Omega}_{a1,j} - \Omega_j = \hat{\Sigma}_j^{-1} \hat{M}_j \hat{\Sigma}_j^{-1} - \Sigma_j^{-1} M_j \Sigma_j^{-1}
\]

\[
= (\hat{\Sigma}_j^{-1} - \Sigma_j^{-1} + \Sigma_j^{-1})(\hat{M}_j - M_j + M_j)(\hat{\Sigma}_j^{-1} - \Sigma_j^{-1} + \Sigma_j^{-1}) - \Sigma_j^{-1} M_j \Sigma_j^{-1}
\]

\[
= \hat{\Sigma}_j^{-1} (\Sigma_j - \hat{\Sigma}_j) \Sigma_j^{-1} (\hat{M}_j - M_j + M_j)(\hat{\Sigma}_j^{-1} - \Sigma_j^{-1} + \Sigma_j^{-1}) + 2 \Sigma_j^{-1} (\hat{M}_j - M_j) \hat{\Sigma}_j^{-1} (\Sigma_j - \hat{\Sigma}_j) \Sigma_j^{-1}
\]

\[
+ \hat{\Sigma}_j^{-1} (\Sigma_j - \hat{\Sigma}_j) \Sigma_j^{-1} M_j \hat{\Sigma}_j^{-1} (\Sigma_j - \hat{\Sigma}_j) \Sigma_j^{-1}
\]

Therefore, by Conditions (C2) and (C3), we can find an absolute positive constant \(C_1\), such that \(\| E(\hat{\Omega}_{a1,j} - \Omega_j) \|_F \leq C_1 \| E(\hat{\Sigma}_j - \Sigma_j) \|_F^2\). The proof is completed by invoking (A.2) \(\Box\).

Next we study the non-asymptotic error bound of the penalized estimates \(\hat{\Omega}_{a2,j}\)'s, which do not require all the sample covariance matrices are invertible. Define

\[
\| A \|_\infty \overset{def}{=} \max_{1 \leq k, l \leq p} |a_{kl}| \text{ and } D_j \overset{def}{=} \| \Gamma_{S_j, S_j,j}^{-1} \|_\infty.
\]

10
Lemma 8. In addition to Conditions (C1)-(C3), we assume there exist
generic constants $C_1$ and $C_2$ such that $\|\Sigma_j\|_\infty \geq C_1 \{\log(p)/n\}^{1/2}$ and
$C_2 s_j D_j \|\Sigma_j\|_\infty \{\log(p)/n\}^{1/2} < \kappa_j$ for all $j = 1, \ldots, m$. Then there exists
an absolute positive constant $C$, such that,

$$\|\|\hat{\Omega}_{a,2,j} - \Omega_j\|_\infty\|_\psi_1 \leq C \kappa_j^{-1} D_j \{\log(p)/n\}^{1/2}. $$

Proof of Lemma 8: We re-present an equivalent form of (2.7) to stack all $p$
columns into a vector. For notational clarity, we define $\hat{\Gamma}_j \coloneqq \hat{\Sigma}_j \otimes \hat{\Sigma}_j$. It
follows that,

$$\text{vec}(\hat{\Omega}_{a,2,j}) = \arg \min_{\text{vec}(\Phi_j)} \left[ \text{vec}(\Phi_j)^T \hat{\Gamma}_j \text{vec}(\Phi_j) - 2 \text{vec}(\hat{M}_j)^T \text{vec}(\Phi_j) + \lambda_{n,j} \|\text{vec}(\Phi_j)\|_1 \right].$$

Setting $\hat{A} = 2 \hat{\Gamma}_j$, $\bar{a} = 2 \text{vec}(\hat{M}_j)$, $A = 2 \Gamma_j$ and $a = 2 \text{vec}(M_j)$ in Lemma
7 of [Wang et al. (2021)], we can derive the non-asymptotic error bound of
vec$(\hat{\Omega}_{a,2,j})$. Towards this goal, we need to verify the conditions required by
Lemma 7 of [Wang et al. (2021)].

By definition, $\|\hat{\Gamma}_j - \Gamma_j\|_\infty = \|\hat{\Sigma}_j \otimes (\hat{\Sigma}_j - \Sigma_j) + (\hat{\Sigma}_j - \Sigma_j) \otimes \Sigma_j\|_\infty$,
which is not greater than $(\|\hat{\Sigma}_j\|_\infty + \|\Sigma_j\|_\infty) \|\hat{\Sigma}_j - \Sigma_j\|_\infty$. It follows that

$$\|\Gamma_{S_j,s,j} \Gamma_{S_j,s,j}^{-1} \|_\infty + 2 s_j \|\Gamma_{S_j,s,j}^{-1} \|_\infty \|\hat{\Gamma}_j - \Gamma_j\|_\infty \leq 1 - \kappa_j + 2 s_j D_j \{\|\hat{\Sigma}_j\|_\infty + \|\Sigma_j\|_\infty\} \|\hat{\Sigma}_j - \Sigma_j\|_\infty.$$ 

By Lemma 5 of [Wang et al. (2021)], we have $\|\|\hat{\Sigma}_j - \Sigma_j\|_\infty\|_\psi_1 \leq C_1 \{\log(p)/n\}^{1/2}$,
and $C_1$ is a general constant. Accordingly, $\|\|\hat{\Sigma}_j\|_\infty\|_\psi_1 \leq \|\Sigma_j\|_\infty + C_1 \{\log(p)/n\}^{1/2}$. 

11

S3. SOME USEFUL LEMMAS

Lemma 8. In addition to Conditions (C1)-(C3), we assume there exist
generic constants $C_1$ and $C_2$ such that $\|\Sigma_j\|_\infty \geq C_1 \{\log(p)/n\}^{1/2}$ and
$C_2 s_j D_j \|\Sigma_j\|_\infty \{\log(p)/n\}^{1/2} < \kappa_j$ for all $j = 1, \ldots, m$. Then there exists
an absolute positive constant $C$, such that,

$$\|\|\hat{\Omega}_{a,2,j} - \Omega_j\|_\infty\|_\psi_1 \leq C \kappa_j^{-1} D_j \{\log(p)/n\}^{1/2}. $$

Proof of Lemma 8: We re-present an equivalent form of (2.7) to stack all $p$
columns into a vector. For notational clarity, we define $\hat{\Gamma}_j \coloneqq \hat{\Sigma}_j \otimes \hat{\Sigma}_j$. It
follows that,

$$\text{vec}(\hat{\Omega}_{a,2,j}) = \arg \min_{\text{vec}(\Phi_j)} \left[ \text{vec}(\Phi_j)^T \hat{\Gamma}_j \text{vec}(\Phi_j) - 2 \text{vec}(\hat{M}_j)^T \text{vec}(\Phi_j) + \lambda_{n,j} \|\text{vec}(\Phi_j)\|_1 \right].$$

Setting $\hat{A} = 2 \hat{\Gamma}_j$, $\bar{a} = 2 \text{vec}(\hat{M}_j)$, $A = 2 \Gamma_j$ and $a = 2 \text{vec}(M_j)$ in Lemma
7 of [Wang et al. (2021)], we can derive the non-asymptotic error bound of
vec$(\hat{\Omega}_{a,2,j})$. Towards this goal, we need to verify the conditions required by
Lemma 7 of [Wang et al. (2021)].

By definition, $\|\hat{\Gamma}_j - \Gamma_j\|_\infty = \|\hat{\Sigma}_j \otimes (\hat{\Sigma}_j - \Sigma_j) + (\hat{\Sigma}_j - \Sigma_j) \otimes \Sigma_j\|_\infty$,
which is not greater than $(\|\hat{\Sigma}_j\|_\infty + \|\Sigma_j\|_\infty) \|\hat{\Sigma}_j - \Sigma_j\|_\infty$. It follows that

$$\|\Gamma_{S_j,s,j} \Gamma_{S_j,s,j}^{-1} \|_\infty + 2 s_j \|\Gamma_{S_j,s,j}^{-1} \|_\infty \|\hat{\Gamma}_j - \Gamma_j\|_\infty \leq 1 - \kappa_j + 2 s_j D_j \{\|\hat{\Sigma}_j\|_\infty + \|\Sigma_j\|_\infty\} \|\hat{\Sigma}_j - \Sigma_j\|_\infty.$$ 

By Lemma 5 of [Wang et al. (2021)], we have $\|\|\hat{\Sigma}_j - \Sigma_j\|_\infty\|_\psi_1 \leq C_1 \{\log(p)/n\}^{1/2}$,
and $C_1$ is a general constant. Accordingly, $\|\|\hat{\Sigma}_j\|_\infty\|_\psi_1 \leq \|\Sigma_j\|_\infty + C_1 \{\log(p)/n\}^{1/2}$. 

11
Consequently,

\[
\begin{align*}
\|\|\Gamma_{s_j, s_j}^{-1} - \Gamma_{s_j, s_j}\|\|_{\infty} &+ 2s_j\|\|\Gamma_{s_j, s_j}^{-1}\|\|\hat{\Gamma}_j - \Gamma_j\|\|_{\psi_1} \\
&\leq 1 - \kappa_j + 2s_jD_j(\|\|\hat{\Sigma}_j\|\|_{\psi_1} + \|\|\Sigma_j\|\|_{\infty})\|\|\hat{\Sigma}_j - \Sigma_j\|\|_{\psi_1}.
\end{align*}
\]

The right hand side is smaller than or equal to \(1 - \kappa_j + C_2 s_j D_j \|\|\Sigma_j\|\|_{\infty} \{\log(p)/n\}^{1/2}\), which, by the assumption we imposed, is strictly smaller than 1. Thus the first set of condition required by Lemma 7 of Wang et al. (2021) is satisfied.

Next we study the property of \(\Delta \overset{\text{def}}{=} 2\|\|\vec{\hat{M}}_j - \vec{M}_j\|\|_{\infty} + 2\|\|\hat{\Gamma}_j - \Gamma_j\|\|\vec{\Omega}_j\|\|_{\infty}\). We only process cumulative slicing estimation here. Let \(e_k\) be a unit length \(p\)-vector with its \(k\)th entry being one. By condition (C1) and Proposition 2.5.2 in Vershynin (2018), we can show that \(e_k^T(x_{i,j} - x_j)\) is sub-Gaussian for all \(k, j = 1, \ldots, p\). Using the general Hoeffding’s inequality (Vershynin, 2018, Theorem 2.6.3), \(e_k^T\hat{\mathbf{m}}_{j,c}(y)\) is also sub-Gaussian for all \(k, j = 1, \ldots, p\). Therefore, it follows immediately from Lemma 2.7.6 and Bernstein’s inequality in Vershynin (2018) that there exists general constants \(c_1\) and \(c_2\), such that \(\Pr\{||e_k(\hat{M}_j - M_j)e_l|| \geq t\} \leq 2\exp\{-n\min(c_1 t^2, c_2 t)\}\) for \(k, l = 1, \ldots, p\). Setting \(t = c_3 \{\log(p)/n\}^{1/2}\), we can find an absolute constant \(C\), such that \(\|\|\vec{\hat{M}}_j - \vec{M}_j\|\|_{\psi_1} = \)
S3. SOME USEFUL LEMMAS

\[ \|\hat{M}_j - M_j\|_\infty \|\psi_1 \leq C\{\log(p)/n\}^{1/2}. \]

In addition,

\[ \|(\hat{\Gamma}_j - \Gamma_j)\text{vec}(\Omega_j)\|_\infty = \|\hat{\Sigma}_j \Omega_j \hat{\Sigma}_j - \Sigma_j \Omega_j \Sigma_j\|_\infty \]

\[ \leq \|(\hat{\Sigma}_j - \Sigma_j)\Omega_j (\hat{\Sigma}_j - \Sigma_j)\|_\infty + 2\|\Sigma_j \Omega_j (\hat{\Sigma}_j - \Sigma_j)\|_\infty \]

\[ \leq s_j \|

which, by invoking \[ \|\hat{\Sigma}_j - \Sigma_j\|_\infty \|\psi_1 \leq C_1\{\log(p)/n\}^{1/2}, \]

implies immediately that \[ \|(\hat{\Gamma}_j - \Gamma_j)\text{vec}(\Omega_j)\|_\infty \|\psi_1 \leq C_2\{\log(p)/n\}^{1/2}, \]

and \( C_2 \) is independent of \( j \). It follows that \( \|\Delta\|_\psi_1 \leq C_3\{\log(p)/n\}^{1/2}. \) \( C_3 \) is also an absolute constant since it is induced by \( C \) and \( C_2 \). Set \( \lambda_{n,j} = 3C_3\kappa_j^{-1}\{\log(p)/n\}^{1/2}. \) Thus the second set of condition required by Lemma 7 of Wang et al. (2021) is satisfied. Thus we are enabled to complete the proof of Lemma 8 by applying Lemma 7 in Wang et al. (2021) directly. □

**Lemma 9.** Assume the conditions of Lemma 8 hold true. Then there exists an absolute positive constant \( C \) such that

\[ \|\text{dist}(\hat{B}_{a2,j}, B)\|_\psi_1 \leq C\kappa_j^{-1}D_j\{d_0s_j \log(p)/n\}^{1/2}. \]

Proof of Lemma 9. By Davis-Kahan-Theorem (Yu et al., 2015, Theorem 2), the result in (A.6), and conditions (C2) and (C3), there exist absolute constant \( C_0 \) such that

\[ \text{dist}(\hat{B}_{a2,j}, B) \leq C_0(8d_0)^{1/2}\|

This, together with Lemma 8, completes the proof of Lemma 9. □
S4. Proof of Lemma 1

By Jensen’s inequality,

\[ \|\Omega^*_a - BB^T\|_F \leq m^{-1} \sum_{j=1}^m \mathbb{E} \| \hat{B}_{a1,j} \hat{B}_{a1,j}^T - BB^T \|_F \leq \max_{1 \leq j \leq m} \| \text{dist}(\hat{B}_{a1,j}, B) \|_{\psi_1}. \]

For sufficiently large \( n \) such that \( n \geq 2d_0pC^2 \), Lemma 6 ensures that \( \| \text{dist}(\hat{B}_{a1,j}, B) \|_{\psi_1} < 1/4 \). Because \( BB^T \) is a projection matrix, \( \lambda_{d_0}(BB^T) = 1 \) and \( \lambda_{d_0+1}(BB^T) = 0 \). This, together with Weyl’s inequality, indicates that \( \lambda_{d_0}(\Omega^*) > 3/4 \) and \( \lambda_{d_0+1}(\Omega^*) < 1/4 \). It follows from Davis-Kahan-Theorem (Yu et al., 2015, Theorem 2) that

\[ \text{dist}(\hat{B}_{a1}, B_{a1}^*) \leq 4\| \hat{T}_{a1} - \Omega_{a1}^* \|_F. \] (A.7)

In addition, by Lemma 4 in Fan et al. (2019), we have

\[ \| \| \hat{T}_{a1} - \Omega_{a1}^* \|_F \|_{\psi_1} \leq Cm^{-1/2} \| \| \hat{B}_{a1,j} \hat{B}_{a1,j}^T - \mathbb{E}(\hat{B}_{a1,j} \hat{B}_{a1,j}^T) \|_F \|_{\psi_1}. \] (A.8)

It remains to bound \( \| \hat{B}_{a1,j} \hat{B}_{a1,j}^T - \mathbb{E}(\hat{B}_{a1,j} \hat{B}_{a1,j}^T) \|_F \) from above. By Jensen’s inequality, \( \| \mathbb{E}(\hat{B}_{a1,j} \hat{B}_{a1,j}^T) - BB^T \|_F \leq \) \( E\{\text{dist}(\hat{B}_{a1,j}, B)\} \leq \| \text{dist}(\hat{B}_{a1,j}, B) \|_{\psi_1} \). We apply triangle inequality to obtain that \( \| \hat{B}_{a1,j} \hat{B}_{a1,j}^T - \mathbb{E}(\hat{B}_{a1,j} \hat{B}_{a1,j}^T) \|_F \leq \text{dist}(\hat{B}_{a1,j}, B) + \| \mathbb{E}(\hat{B}_{a1,j} \hat{B}_{a1,j}^T) - BB^T \|_F \), which, by definition, is not greater than \( \| \text{dist}(\hat{B}_{a1,j}, B) \|_{\psi_1} + \| \text{dist}(\hat{B}_{a1,j}, B) \|_{\psi_1} \).

This implies that

\[ \| \| \hat{B}_{a1,j} \hat{B}_{a1,j}^T - \mathbb{E}(\hat{B}_{a1,j} \hat{B}_{a1,j}^T) \|_F \|_{\psi_1} \leq 2\| \text{dist}(\hat{B}_{a1,j}, B) \|_{\psi_1}. \] (A.9)
Invoking (A.7) - (A.9), we have \( \| \text{dist}(\hat{B}_{a1}, B^*_a) \|_{\psi_1} \leq 8Cm^{-1/2}\| \text{dist}(\hat{B}_{a1,j}, B) \|_{\psi_1} \). The proof is completed by invoking Lemma 6. □

S5  Proof of Lemma 2

With similar arguments for proving (A.6), we have \( \text{dist}(B^*_a, B) = 2^{1/2}\| \sin(\Theta(B^*_a, B)) \|_F \), where \( \Theta(B^*_a, B) \) are the principal angles between \( \text{span}(B^*_a) \) and \( \text{span}(B) \).

Invoking Davis-Kahan-Theorem (Yu et al. 2015, Theorem 2) again, we have

\[
\text{dist}(B^*_a, B) \leq 8^{1/2} \| \Omega_a1 - BB^T \|_F \leq 8^{1/2} \max_{1 \leq j \leq m} \| E(\hat{B}_{a1,j} \hat{B}_{a1,j}^T - BB^T) \|_F
\]

For notational clarity, we define \( E_{a1,j} \equiv \hat{\Omega}_{a1,j} - \Omega_j \). Let \( (\hat{b}_{1,j}, \hat{b}_{2,j}, \ldots, \hat{b}_{p,j}) \) be the eigenvectors of \( \hat{\Omega}_{a1,j} \), and \( (b_1, b_2, \ldots, b_p) \) be the eigenvectors of \( \Omega \).

For any fixed \( s \in \{0, 1, \ldots, p - d_0\} \), we define \( S \equiv \{s + 1, \ldots, s + d_0\} \) and

\[
G_k \equiv \sum_{l \notin S} (\lambda_l(\Omega_j) - \lambda_{s+k}(\Omega_j))^{-1} b_l b_l^T.
\]

Let \( f \) be a linear map \( f : \mathbb{R}^{p \times d_0} \mapsto \mathbb{R}^{p \times d_0}, (v_1, \ldots, v_{d_0}) \mapsto (-G_1v_1, \ldots, -G_{d_0}v_{d_0}) \).

By the linearity of \( f \) and the triangular inequality, we have

\[
\| E\left[\hat{B}_{a1,j} \hat{B}_{a1,j}^* - \{BB^T + f(E_{a1,j}B)B^T + Bf(E_{a1,j}B)^T\}\right] \|_F
\]
\[
\geq \| E\hat{B}_{a1,j} \hat{B}_{a1,j}^* - BB^T \|_F - \| f\{E(E_{a1,j})\}BB^T \|_F - \| Bf\{E(E_{a1,j})\}B^T \|_F.
\]
Define $\epsilon_{a_1} \equiv \|E_{a_1,j}\|/\min\{\lambda_s(\Omega_j) - \lambda_{s+1}(\Omega_j), \lambda_{s+d_0}(\Omega_j) - \lambda_{s+d_0+1}(\Omega_j)\}$. By Lemma 2 in Fan et al. (2019) and Jensen’s inequality,

$$\|E\left[\hat{B}_{a_1,j}\hat{B}_{a_1,j}^T - \{BB^T + f(E_{a_1,j}B)B^T + Bf(E_{a_1,j}B)^T\}\right]\|_F \leq 24d_0^{1/2}E(\epsilon_{a_1}^2).$$

In addition, $\|f\{E(E_{a_1,j})\}BB^T\|_F \leq \|f\{E(E_{a_1,j})\}\|_F \leq C\|E(E_{a_1,j})\|_F$. The first inequality follows from Lemma A1 in Yu et al. (2015), and the second is a direct application of Jensen’s inequality. Similarly, $\|BF\{E(E_{a_1,j})\}B^T\|_F \leq C\|E(E_{a_1,j})\|_F$. Lemma 7 proves that $\|E(E_{a_1,j})\|_F \leq C_1p/n$. Besides, Lemma 5 indicates that $\|E(\epsilon_{a_1}^2)\|_{\psi_1} \leq C_2p/n$.

Summarizing the above results, we obtain that there exists an absolute positive constant $C$ such that $\text{dist}(B_{a_1}^*, B) \leq Cd_0^{1/2}p/n$, for $1 \leq j \leq m$. The proof is now completed.

**S6 Proof of Lemma 3**

Following similar arguments for proving Lemma 1, we can prove this lemma by using Lemma 9. Details are omitted from the present context.
S7 Proof of Lemma 4

Invoking Davis-Kahan-Theorem [Yu et al. 2015, Theorem 2] again, we have

$$\text{dist}(B_{a2}^*, B) \leq 8^{1/2}\|\Omega_{a2}^* - BB^T\|_F \leq 8^{1/2} \max_{1 \leq j \leq m} \|E(\hat{\Omega}_{a2,j}^* - BB^T)\|_F$$

In parallel to Lemma 2, we define $E_{a2,j} \overset{\text{def}}{=} \hat{\Omega}_{a2,j} - \Omega_j$ and let $(\hat{b}_{1,j}, \hat{b}_{2,j}, \ldots, \hat{b}_{p,j})$ be the eigenvectors of $\hat{\Omega}_{a2,j}$. Recall that $(b_1, b_2, \ldots, b_p)$ are the eigenvectors of $\Omega$, and $f$ is a linear map defined in Lemma 2. By triangular inequality again, we have

$$\|E \left[ \hat{\Omega}_{a2,j}^* - \{BB^T + f(E_{a2,j}B)B^T + Bf(E_{a2,j}B)^T\} \right]\|_F \geq \|E \hat{\Omega}_{a2,j}^* - BB^T\|_F - \|f\{E(E_{a2,j})\}BB^T\|_F - \|Bf\{E(E_{a2,j})\}B^T\|_F.$$

Define $\epsilon_{a2} \overset{\text{def}}{=} \|E_{a2,j}\|/\min\{\lambda_s(\Omega) - \lambda_{s+1}(\Omega), \lambda_{s+d_0}(\Omega) - \lambda_{s+d_0+1}(\Omega)\}$. By Lemma 2 in Fan et al. (2019) and Jensen’s inequality,

$$\|E \left[ \hat{\Omega}_{a2,j}^* - \{BB^T + f(E_{a2,j}B)B^T + Bf(E_{a2,j}B)^T\} \right]\|_F \leq 24d_0^{1/2}E(\epsilon_{a2}^2).$$

In addition, $\|f\{E(E_{a2,j})\}BB^T\|_F \leq \|f\{E(E_{a2,j})\}\|_F \leq C\|E(E_{a2,j})\|_F$, Lemma 8 shows that $\|E_{a2,j}\|_\infty \psi_1 \leq C\kappa_j^{-1}D_j \{\log(p)/n\}^{1/2}$. Therefore, $\|E(E_{a2,j})\|_F \leq E\|E_{a2,j}\|_F \leq s_j^{1/2}E\|E_{a2,j}\|_\infty \leq s_j^{1/2}\|E_{a2,j}\|_\infty \psi_1$. Summarizing the above results, we obtain that $\text{dist}(B_{a2}^*, B) \leq C\kappa_j^{-1}D_j \{s_j \log(p)/n\}^{1/2}$ for $1 \leq j \leq m$. This completes the proof. □
BIBLIOGRAPHY

Bibliography


Figure 4: The horizontal axis stands for the log(2)-transformed value of the total sample size $N$, and the vertical axis stands for $\text{dist}(\hat{B}, B^*)$ in (A) and (D), $\text{dist}(B^*, B)$ in (B) and (E), and $\text{dist}(\hat{B}, B)$ in (C) and (F). All the distributed estimates of $B$ are obtained through Algorithm 3. The distributed estimates of sliced inverse regression are displayed in the subplots (A)-(C) and those of cumulative slicing estimation are displayed in the subplots (D)-(F).
Figure 5: The horizontal axis stands for the log(2)-transformed value of the total sample size $N$, and the vertical axis stands for $\text{dist}(\hat{\mathbf{B}}, \mathbf{B}^*)$ in (A) and (D), $\text{dist}(\mathbf{B}^*, \mathbf{B})$ in (B) and (E), and $\text{dist}(\hat{\mathbf{B}}, \mathbf{B})$ in (C) and (F). All the distributed estimates of $\mathbf{B}$ are obtained through Algorithm 2. The distributed estimates of sliced inverse regression are displayed in the subplots (A)-(C) and those of cumulative slicing estimation are displayed in the subplots (D)-(F).
Figure 6: The distances between the estimates obtained by using the bootstrapped samples and those obtained by using the original observations. The left boxplot stands for the distributed estimates and the right one corresponds to the pooled estimates.