Supplementary material to "Eigen selection in spectral clustering: a theory guided practice"

Xiao Han[*], Xin Tong[†] and Yingying Fan[‡]

S.1 Toy examples

Here we present the setting of the toy examples that correspond to Tables 1-2.

Model A: $\boldsymbol{\mu}_1 = (\boldsymbol{\mu}_{11}^\top, \boldsymbol{\mu}_{12}^\top)^\top$, $\boldsymbol{\mu}_2 = -(\boldsymbol{\mu}_{31}^\top, \boldsymbol{\mu}_{11}^\top, \boldsymbol{\mu}_{32}^\top)^\top$, where $\boldsymbol{\mu}_{31}$ is an $(l/2)$-dimensional vector in which all entries are 0, $\boldsymbol{\mu}_{32}$ is a $(p - 3l/2)$-dimensional vector in which all entries are 0, $p \in \{100, 200, 400, 600, 800\}$, $l = 8$. The covariance matrix $\boldsymbol{\Sigma} = r^2\mathbf{I}$, $r = 2$. In this model, we also let $n_1 = n_2 = n/2 = 100$. In this model, it is easy to see that the entries of the second right singular vector of $\mathbb{E}\mathbf{X}$ are all equal and thus it does not have clustering power.

Model B: $\boldsymbol{\mu}_1 = 2(\boldsymbol{\mu}_{11}^\top, \boldsymbol{\mu}_{12}^\top)^\top$, $\boldsymbol{\mu}_2 = (\boldsymbol{\mu}_{12}^\top, \boldsymbol{\mu}_{11}^\top)^\top$, where $\boldsymbol{\mu}_{11}$ is an l-dimensional vector in which all entries are 1, $\boldsymbol{\mu}_{12}$ is a $(p-l)$-dimensional vector in which all entries are 0, $p \in \{100, 200, 400, 600, 800\}$, $l = 24$. The covariance matrix $\boldsymbol{\Sigma} = r^2\mathbf{I}$, $r = 2$. In this model, we also let $n_1 = n_2 = n/2 = 50$. Then we have $d_2 = d_1/2$.

With Models A and B, we compare the k-means approach that acts on $\hat{\mathbf{u}}_1$ with the k-means approach that acts on both $\hat{\mathbf{u}}_1$ and $\hat{\mathbf{u}}_2$, which are eigenvectors of $\mathbf{X}^\top\mathbf{X}$. We simulate for 100 times from these models and calculate the average misclustering rates and the corresponding standard error in Tables 1-2.

S.2 The properties of the spectrum of $\mathbf{H}$

Because

$$\mathrm{rank}((\mathbb{E}\mathbf{X})^\top) \leq \mathrm{rank}(\mathbf{a}_1\boldsymbol{\mu}_1^\top) + \mathrm{rank}(\mathbf{a}_2\boldsymbol{\mu}_2^\top) = 2\,, \tag{S.1}$$

there exist at most two n-dimensional orthogonal unit vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ such that

$$\mathbf{H} = d_1^2\mathbf{u}_1\mathbf{u}_1^\top + d_2^2\mathbf{u}_2\mathbf{u}_2^\top, \text{ where } d_1^2 \geq d_2^2 \geq 0\,. \tag{S.2}$$

Here, d_1^2 and d_2^2 are the top two eigenvalues of $\mathbf{H}$ and $\mathbf{u}_1$ and $\mathbf{u}_2$ are the corresponding (population) eigenvectors. Under our model setting, we have $d_1^2 > 0$ because otherwise $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \mathbf{0}$, contradicting with the model assumption. For simplicity, in the following, we use $\mathbf{u} = (\mathbf{u}(1), \ldots, \mathbf{u}(n))^\top$ to denote either $\mathbf{u}_1$ or $\mathbf{u}_2$ and d^2 to denote its corresponding eigenvalue. By the definition of eigenvalue,

$$\mathbf{H}\mathbf{u} = d^2\mathbf{u}\,. \tag{S.3}$$

[*]Xiao Han, International Institute of Finance, School of Management, University of Science and Technology of China, Hefei, Anhui, China, 230026(Email: xhan011@e.ntu.edu.sg).

[†]Xin Tong, Data Sciences and Operations Department, University of Southern California, Los Angeles, CA, 90089(Email: xint@marshall.usc.edu). Han, Tong and Fan have equal contribution. Tong is the corresponding author.

[‡]Yingying Fan, Data Sciences and Operations Department, University of Southern California, Los Angeles, CA, 90089(Email: fanyingy@marshall.usc.edu).

Note that $\mathbf{H}$ has a block structure by suitable permutation of rows and columns. For example, when $\mathbf{a}_1 = (1,0,1,0)^\top$, $\mathbf{a}_2 = (0,1,0,1)^\top$, substituting $\mathbf{a}_1$ and $\mathbf{a}_2$ into (2), we have

$$\mathbf{H} = \begin{pmatrix} c_{11} & c_{12} & c_{11} & c_{12} \\ c_{12} & c_{22} & c_{12} & c_{22} \\ c_{11} & c_{12} & c_{11} & c_{12} \\ c_{12} & c_{22} & c_{12} & c_{22} \end{pmatrix}.$$

By exchanging the 2nd and 3rd rows and columns of $\mathbf{H}$ simultaneously, we can get the following matrix with a clear block structure

$$\widetilde{\mathbf{H}} = \left(\begin{array}{cc:cc} c_{11} & c_{11} & c_{12} & c_{12} \\ c_{11} & c_{11} & c_{12} & c_{12} \\ \hdashline c_{12} & c_{12} & c_{22} & c_{22} \\ c_{12} & c_{12} & c_{22} & c_{22} \end{array}\right).$$

The eigenvalues of $\mathbf{H}$ and $\widetilde{\mathbf{H}}$ are the same and the eigenvectors are the same up to proper permutation of their coordinates. Inspired by the block structure of $\mathbf{H}$ after proper permutation, we can see that (2) and (S.3) imply

$$c_{11} \sum_{\mathbf{a}_1(i)=1} \mathbf{u}(i) + c_{12} \sum_{\mathbf{a}_1(i)=0} \mathbf{u}(i) = d^2 \mathbf{u}(j), \quad \text{for } j \text{ such that } \mathbf{a}_1(j) = 1\,, \tag{S.4}$$

$$c_{22} \sum_{\mathbf{a}_1(i)=0} \mathbf{u}(i) + c_{12} \sum_{\mathbf{a}_1(i)=1} \mathbf{u}(i) = d^2 \mathbf{u}(j), \quad \text{for } j \text{ such that } \mathbf{a}_1(j) = 0\,. \tag{S.5}$$

From (S.4) and (S.5), we conclude that if $d^2 > 0$, then

$$\mathbf{a}_1(i) = \mathbf{a}_1(j) \Longrightarrow \mathbf{u}(i) = \mathbf{u}(j)\,. \tag{S.6}$$

S.3 Proof of Theorem 1

We use $\mathbf{u} = (\mathbf{u}(1), \ldots, \mathbf{u}(n))^\top$ to denote either $\mathbf{u}_1$ or $\mathbf{u}_2$ and d^2 to denote its corresponding eigenvalue, unless specified otherwise.

Because $\mathbf{a}_1$ only takes two values, by (S.6), there are at most two values of $\mathbf{u}(i)$, $i = 1, \ldots, n$. We denote these values by v_1 and v_2. By (S.4) and (S.5), the number of v_1's in $\mathbf{u}$ is either n_1 or n_2. Without loss of generality, we assume the number of v_1's in $\mathbf{u}$ is n_1 and the number of v_2's in $\mathbf{u}$ is n_2.

Then it follows from (S.4) and (S.5) that

$$n_1 c_{11} v_1 + n_2 c_{12} v_2 = d^2 v_1\,, \text{ and } \quad n_1 c_{12} v_1 + n_2 c_{22} v_2 = d^2 v_2\,. \tag{S.7}$$

These equations are equivalent to

$$(d^2 - n_1 c_{11}) v_1 = n_2 c_{12} v_2\,, \tag{S.8}$$

$$n_1 c_{12} v_1 = (d^2 - n_2 c_{22}) v_2\,. \tag{S.9}$$

In view of (S.8) and (S.9), we have both d_1^2 and d_2^2 solve the equation

$$(d^2 - n_2 c_{22})(d^2 - n_1 c_{11}) = n_1 n_2 c_{12}^2\,. \tag{S.10}$$

Then (3) and (4) follows from (S.10) directly. Now let us prove (a)-(d) of Theorem 1 one by one.

(a) When $c_{12}^2 = c_{11}c_{22}$, by (3) and (4) we have $d_1^2 = n_1 c_{11} + n_2 c_{22}$ and $d_2^2 = 0$. Then $\mathbf{u}_2$ does not have clustering power. Substituting $d_1^2 = n_1 c_{11} + n_2 c_{22}$ into (S.8) and (S.9), we obtain that $\mathbf{u}_1 \propto \mathbf{1}$ if and only if $c_{11} = c_{12} = c_{22}$, which is equivalent to $\mu_1 = \mu_2$. This is a contradiction to the condition that $\mu_1 \neq \mu_2$ in this paper. Therefore $\mathbf{u}_1$ has clustering power.

(b) When $c_{12} = 0$, $c_{12}^2 \neq c_{11}c_{22}$ and $n_1 c_{11} = n_2 c_{22}$, by (3) and (4) we conclude that $d_1^2 = d_2^2 = n_1 c_{11}$. Since $\mathbf{u}_1^\top \mathbf{u}_2 = 0$, it is easy to see that at least one of $\mathbf{u}_1$ and $\mathbf{u}_2$ has clustering power.

(c) When $c_{12} = 0$, $c_{12}^2 \neq c_{11}c_{22}$ and $n_1 c_{11} \neq n_2 c_{22}$, then it follows from (3) and (4) that $d_1^2 = \max\{n_1 c_{11}, n_2 c_{22}\}$ and $d_2^2 = \min\{n_1 c_{11}, n_2 c_{22}\}$. Moreover, by $0 = c_{12}^2 \neq c_{11}c_{22}$ we have $c_{11}, c_{22} > 0$, which implies that $d_2^2 > 0$. Combining these with (S.8) and (S.9), we have both $\mathbf{u}_1$ and $\mathbf{u}_2$ have clustering power. Moreover, both $\mathbf{u}_1$ and $\mathbf{u}_2$ contain zero entries in view of (S.7).

(d) When $c_{12} \neq 0$ and $c_{12}^2 \neq c_{11}c_{22}$. By (3) and (4) we have $d_1^2, d_2^2 \neq n_1 c_{11} \neq 0$, by (S.8) we have

$$v_1 = \frac{n_2 c_{12}}{d^2 - n_1 c_{11}} v_2 \, . \tag{S.11}$$

Therefore if $n_2 c_{12}/(d^2 - n_1 c_{11}) \neq 1$, the corresponding eigenvector $\mathbf{u}$ has clustering power. Moreover, in case (d), $n_2 c_{12}/(d^2 - n_1 c_{11}) = 1$ is equivalent to $d^2 = n_1 c_{11} + n_2 c_{12} = n_1 c_{12} + n_2 c_{22}$ by (S.8) and (S.9). Moreover, the corresponding eigenvector $\mathbf{u}$ has all entries equal to the same value and thus has no clustering power. Since $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal, when $n_1 c_{11} + n_2 c_{12} = n_1 c_{12} + n_2 c_{22}$, exactly one of $\mathbf{u}_1$ and $\mathbf{u}_2$ has clustering power. If $n_1 c_{11} + n_2 c_{12} \neq n_1 c_{12} + n_2 c_{22}$, then $n_2 c_{12}/(d_1^2 - n_1 c_{11}) \neq 1$ and $n_2 c_{12}/(d_2^2 - n_1 c_{11}) \neq 1$ and thus both $\mathbf{u}_1$ and $\mathbf{u}_2$ have clustering power.

S.4 The upper bound of $\widehat{t}_1 - \widehat{t}_2$

Equations (15) and (18) imply that

$$\widehat{t}_1 - \widehat{t}_2 = \left\{ \left(g_{11}(t_1) + g_{22}(t_1)\right)^2 - 4\left(g_{11}(t_1) g_{22}(t_1) - g_{12}^2(t_1)\right) \right\}^{\frac{1}{2}} + o_p(1) \, . \tag{S.12}$$

To bound the main term in (S.12), we calculate the variance and covariance of $\mathbf{v}_i^\top \mathbf{W} \mathbf{v}_j$, $1 \leq i, j \leq 2$, as follows.

$$\begin{aligned} \mathrm{var}(\mathbf{v}_i^\top \mathbf{W} \mathbf{v}_i) &= 4 \mathbf{w}_i^\top \boldsymbol{\Sigma} \mathbf{w}_i, \ i = 1, 2 \, , \\ \mathrm{var}(\mathbf{v}_1^\top \mathbf{W} \mathbf{v}_2) &= \mathbf{w}_1^\top \boldsymbol{\Sigma} \mathbf{w}_1 + \mathbf{w}_2^\top \boldsymbol{\Sigma} \mathbf{w}_2, \ i = 1, 2 \, , \\ \mathrm{cov}(\mathbf{v}_i^\top \mathbf{W} \mathbf{v}_i, \mathbf{v}_1^\top \mathbf{W} \mathbf{v}_2) &= 2 \mathbf{w}_1^\top \boldsymbol{\Sigma} \mathbf{w}_2, \ i = 1, 2 \, , \end{aligned} \tag{S.13}$$

where $\mathbf{w}_i$ is the last p entries of $\mathbf{v}_i$. Also note that

$$\mathbb{E} \mathbf{W}^2 = \mathrm{diag}(n \boldsymbol{\Sigma}, (\mathrm{tr} \boldsymbol{\Sigma}) \mathbf{I}) \, . \tag{S.14}$$

Hence, $\mathbf{v}_1^\top \mathbb{E} \mathbf{W}^2 \mathbf{v}_1 - \mathbf{v}_2^\top \mathbb{E} \mathbf{W}^2 \mathbf{v}_2 = n(\mathbf{w}_1 \boldsymbol{\Sigma} \mathbf{w}_1 - \mathbf{w}_2 \boldsymbol{\Sigma} \mathbf{w}_2)$ and $\mathbf{v}_1^\top \mathbb{E} \mathbf{W}^2 \mathbf{v}_2 = n \mathbf{w}_1 \boldsymbol{\Sigma} \mathbf{w}_2$. By Lemma 3 in the Supplementary Material and (10), we have

$$\mathbf{v}_1^\top \mathbb{E} \mathbf{W}^2 \mathbf{v}_1 = \frac{1}{2}(n \mathbf{w}_1^\top \boldsymbol{\Sigma} \mathbf{w}_1 + tr \boldsymbol{\Sigma}) \sim \sigma_n^2 \, . \tag{S.15}$$

By (S.14) and Assumption 1 on $\boldsymbol{\Sigma}$, for $\mathbf{M}_1$ and $\mathbf{M}_2$ with finite columns and spectral norms, we have

$$\|\mathcal{R}(\mathbf{M}_1, \mathbf{M}_2, t_1) + \sum_{l=0,\, l \neq 1}^{2} t_1^{-(l+1)} \mathbf{M}_1^\top \mathbb{E} \mathbf{W}^l \mathbf{M}_2 \| = O\left(\frac{\sigma_n^3}{t_1^4}\right) . \tag{S.16}$$

Then (S.15), (S.16), Assumption 1 and the definition of $g(z)$ together imply that

$$\left| g_{ij}(t_1) - \frac{t_1^2}{d_i} + \mathbf{v}_i^T \mathbf{W} \mathbf{v}_j + t_1 + \frac{\mathbf{v}_i^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_j}{t_1} \right| = O\left(\frac{\sigma_n^3}{t_1^2}\right) \ll \frac{\mathbf{v}_1^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_1}{t_1} . \tag{S.17}$$

By Lemma 1 we have $t_1 = d_1 + O(\frac{\sigma_n^2}{d_2})$, (S.17) suggests that we have with probability tending to 1,

$$\left\{ \left(g_{11}(t_1) + g_{22}(t_1)\right)^2 - 4\left(g_{11}(t_1) g_{22}(t_1) - g_{12}^2(t_1)\right) \right\}^{\frac{1}{2}} \tag{S.18}$$

$$\leq \left\{ \left(\frac{t_1^2(d_1 - d_2)}{d_1 d_2} + \frac{\mathbf{v}_1^\top \mathbb{E} \mathbf{W}^2 \mathbf{v}_1 - \mathbf{v}_2^\top \mathbb{E} \mathbf{W}^2 \mathbf{v}_2}{t_1} + \mathbf{v}_1^\top \mathbf{W} \mathbf{v}_1 - \mathbf{v}_2^\top \mathbf{W} \mathbf{v}_2 \right)^2 + 4\left(\frac{\mathbf{v}_1^\top \mathbb{E} \mathbf{W}^2 \mathbf{v}_2}{t_1} + \mathbf{v}_1^\top \mathbf{W} \mathbf{v}_2 \right)^2 \right\}^{\frac{1}{2}} \tag{S.19}$$

$$+ \epsilon \frac{\mathbf{v}_1^\top \mathbb{E} \mathbf{W}^2 \mathbf{v}_1}{t_1} ,$$

for any positive constant ϵ. Through (S.13) and (S.15), we see that on both sides of (S.18), the information of $\mathbf{\Sigma}$ plays an important role. Therefore, a good thresholding procedure on $\widehat{t}_1 - \widehat{t}_2$ would involve an accurate estimate of $\mathbf{\Sigma}$, which is difficult to obtain in the absence of label information.

S.5 Proof of Proposition 1

The main idea for proving Proposition 1 is to carefully construct a matrix whose eigenvalue is $\widehat{t}_k - t_1$, then using similar idea for proving Lemma 1 by analysing the resolvent entries of the matrices such as $(\mathbf{W} - z\mathbf{I})^{-1}$, we can get the desired asymptotic expansions.

By the conditions in Proposition 1, for sufficiently large n, there exists some positive constant L such that

$$\frac{\sigma_n^L}{d_1^L} < \frac{1}{2d_1^4} , \tag{S.20}$$

and in the sequel we fix this L. Indeed, $\frac{\sigma_n^L}{d_1^{3L/4}} \ll 1$ and therefore (S.20) holds for $L = 16$.

Assumption (12) implies that

$$\frac{d_1}{d_2} = 1 + o(1). \tag{S.21}$$

It follows from $d_2 \gg \sigma_n$ and (S.21) that

$$\frac{a_n}{d_2} = 1 + o(1) \text{ and } \frac{b_n}{d_1} = 1 + o(1) . \tag{S.22}$$

Moreover, it follows from (S.21) and Assumption 1 that

$$\frac{\sigma_n}{a_n} \leq \frac{1}{2n^\epsilon}, \text{ for some positive constant } \epsilon . \tag{S.23}$$

Throughout the proof, (S.23) will be applied in every $O_p(\cdot)$, $o_p(\cdot)$, $O(\cdot)$ and $o(\cdot)$ terms without explicit quotation. We define a Green function of $\mathbf{W}$ (defined in (9)) by

$$\mathbf{G}(z) = (\mathbf{W} - z\mathbf{I})^{-1}, \ z \in \mathbb{C}, \ |z| > \|\mathbf{W}\| . \tag{S.24}$$

By Weyl's inequality, we have $|\widehat{t}_k - d_k| \leq \|\mathbf{W}\|$, $k = 1, 2$. Thus, by (S.22) and Lemma 4, with probability tending to 1,

$$\min\{\widehat{t}_2, a_n\} \gg \|\mathbf{W}\| . \tag{S.25}$$

Therefore, $\mathbf{G}(z)$, $z \in [a_n, b_n]$, $\mathbf{G}(\widehat{t}_1)$ and $\mathbf{G}(\widehat{t}_2)$ are well defined and nonsingular with probability tending to 1. Since we only need to show the conclusions of Proposition 1 hold with probability tending to 1, in the sequel of this proof, we will assume the existence and nonsingularity of $\mathbf{G}(\widehat{t}_k)$.

By the decomposition of $\mathbb{E}\mathcal{Z}$ in (8) and definition of $\mathbf{W}$ in (9), we have $\mathcal{Z} = \mathbf{VDV}^\top - \mathbf{V}_-\mathbf{DV}_-^\top + \mathbf{W}$. Then it can be calculated that

$$\begin{aligned}
0 &= \det\left(\mathcal{Z} - \widehat{t}_k\mathbf{I}\right)\\
&= \det\left(\mathbf{W} - \widehat{t}_k\mathbf{I} + \mathbf{VDV}^\top - \mathbf{V}_-\mathbf{DV}_-^\top\right)\\
&= \det\left(\mathbf{G}^{-1}(\widehat{t}_k) + (\mathbf{VDV}^\top - \mathbf{V}_-\mathbf{DV}_-^\top)\right)\\
&= \det\left(\mathbf{G}^{-1}(\widehat{t}_k)\right)\det\left(\mathbf{I} + \mathbf{G}(\widehat{t}_k)(\mathbf{VDV}^\top - \mathbf{V}_-\mathbf{DV}_-^\top)\right), \quad k = 1, 2\,.
\end{aligned}$$

Since $\mathbf{G}(\widehat{t}_k)$ is a nonsingular matrix, $\det[\mathbf{G}^{-1}(\widehat{t}_k)] \neq 0$, which leads to

$$\det\left(\mathbf{I} + \mathbf{G}(\widehat{t}_k)(\mathbf{VDV}^\top - \mathbf{V}_-\mathbf{DV}_-^\top)\right) = 0.$$

Notice that $(\mathbf{VDV}^\top - \mathbf{V}_-\mathbf{DV}_-^\top) = (\mathbf{V}, \mathbf{V}_-)\begin{pmatrix}\mathbf{D} & 0\\ 0 & -\mathbf{D}\end{pmatrix}(\mathbf{V}, \mathbf{V}_-)^\top$. Combining this with the identity $\det(\mathbf{I} + \mathbf{AB}) = \det(\mathbf{I} + \mathbf{BA})$ for any matrices $\mathbf{A}$ and $\mathbf{B}$, we have

$$0 = \det[\mathbf{I} + \mathbf{G}(\widehat{t}_k)(\mathbf{VDV}^\top - \mathbf{V}_-\mathbf{DV}_-^\top)] = \det\left[\mathbf{I} + \begin{pmatrix}\mathbf{D} & 0\\ 0 & -\mathbf{D}\end{pmatrix}(\mathbf{V}, -\mathbf{V}_-)^\top\mathbf{G}(\widehat{t}_k)(\mathbf{V}, -\mathbf{V}_-)\right].$$

Since $\mathbf{D} > 0$, it follows from the equation above that

$$\det\left[\begin{pmatrix}\mathbf{D}^{-1} & 0\\ 0 & -\mathbf{D}^{-1}\end{pmatrix} + (\mathbf{V}, -\mathbf{V}_-)^\top\mathbf{G}(\widehat{t}_k)(\mathbf{V}, -\mathbf{V}_-)\right] = 0, \quad \text{for } k = 1, 2\,. \tag{S.26}$$

To analyze (S.26), we prove some properties of $\mathbf{G}(z)$ and the related expressions. First of all, by Lemma 1, we have

$$t_k - d_k = O\left(\frac{\sigma_n^2}{a_n}\right), \ k = 1, 2\,. \tag{S.27}$$

Therefore the distance of t_k and d_k is well controlled and will be used later in this proof. Now we turn to analyse $\widehat{t}_k$, $k = 1, 2$. By (S.25), we have

$$\mathbf{G}(z) = (\mathbf{W} - z\mathbf{I})^{-1} = -\sum_{i=0}^{\infty}\frac{\mathbf{W}^i}{z^{i+1}}\,, \tag{S.28}$$

and

$$\mathbf{G}'(z) = -(\mathbf{W} - z\mathbf{I})^{-2} = \sum_{i=0}^{\infty}\frac{(i+1)\mathbf{W}^i}{z^{i+2}}, \ z \in [a_n, b_n]\,. \tag{S.29}$$

By (S.20), (S.28), (S.29), Lemmas 3 and 4, for any $z \in [a_n, b_n]$ we have

$$\begin{aligned}
\mathbf{M}_1^\top\mathbf{G}(z)\mathbf{M}_2 = \mathbf{M}_1^\top(\mathbf{W} - z\mathbf{I})^{-1}\mathbf{M}_2 &= -\sum_{i=0}^{\infty}\frac{1}{z^{i+1}}\mathbf{M}_1^\top\mathbf{W}^i\mathbf{M}_2\\
&= \mathcal{R}(\mathbf{M}_1, \mathbf{M}_2, z) - z^{-2}\mathbf{M}_1^\top\mathbf{W}\mathbf{M}_2 - \sum_{i=2}^{L}\frac{1}{z^{i+1}}\mathbf{M}_1^\top(\mathbf{W}^i - \mathbb{E}\mathbf{W}^i)\mathbf{M}_2 + \tilde{\Delta}_{n1}\\
&= \mathcal{R}(\mathbf{M}_1, \mathbf{M}_2, z) - z^{-2}\mathbf{M}_1^\top\mathbf{W}\mathbf{M}_2 + \Delta_{n1}\,,
\end{aligned} \tag{S.30}$$

and

$$\mathbf{M}_1^\top \mathbf{G}'(z)\mathbf{M}_2 = \mathbf{M}_1^\top(\mathbf{W}-z\mathbf{I})^{-2}\mathbf{M}_2 = \sum_{i=0}^{\infty}\frac{i+1}{z^{i+2}}\mathbf{M}_1^\top\mathbf{W}^i\mathbf{M}_2$$

$$= \mathcal{R}'(\mathbf{M}_1,\mathbf{M}_2,z) + 2z^{-3}\mathbf{M}_1^\top\mathbf{W}\mathbf{M}_2 + \sum_{i=2}^{L}\frac{i+1}{z^{i+2}}\mathbf{M}_1^\top(\mathbf{W}^i-\mathbb{E}\mathbf{W}^i)\mathbf{M}_2 + \tilde{\Delta}_n$$

$$= \mathcal{R}'(\mathbf{M}_1,\mathbf{M}_2,z) + 2z^{-3}\mathbf{M}_1^\top\mathbf{W}\mathbf{M}_2 + \Delta_n\,, \tag{S.31}$$

where $\|\Delta_{n1}\| = O_p(\frac{\sigma_n}{a_n^3})$, $\|\tilde{\Delta}_{n1}\| = O_p(\frac{1}{a_n^3})$, $\|\Delta_n\| = O_p(\frac{\sigma_n}{a_n^4})$ and $\|\tilde{\Delta}_n\| = O_p(\frac{1}{a_n^4})$. Notice that

$$\mathcal{R}'(\mathbf{M}_1,\mathbf{M}_2,z) = \frac{\mathbf{M}_1^\top\mathbf{M}_2}{z^2} + \frac{\mathbf{M}_1^\top\mathbb{E}\mathbf{W}^2\mathbf{M}_2}{z^4} + \sum_{i=3}^{L}\frac{i+1}{z^{i+2}}\mathbf{x}^\top\mathbb{E}\mathbf{W}^i\mathbf{y}\,.$$

It follows from Lemma 3 and (10) that for all $z\in[a_n,b_n]$

$$\left\|\mathcal{R}(\mathbf{M}_1,\mathbf{M}_2,z) + z^{-1}\mathbf{M}_1^\top\mathbf{M}_2\right\| = O(\sigma_n^2/a_n^3)\,, \tag{S.32}$$

and

$$\left\|\mathcal{R}'(\mathbf{M}_1,\mathbf{M}_2,z) - z^{-2}\mathbf{M}_1^\top\mathbf{M}_2\right\| = O(\sigma_n^2/a_n^4)\,. \tag{S.33}$$

By (S.30) and Lemma 3, we can conclude that for all $z\in[a_n,b_n]$

$$\left\|\mathbf{V}^\top\mathbf{G}(z)\mathbf{V}_-\right\| = a_n^{-2}O_p(1) + a_n^{-3}O_p(\sigma_n^2)\,, \tag{S.34}$$

and

$$\left\|\mathbf{M}_1^\top\mathbf{G}(z)\mathbf{M}_2 - \mathcal{R}(\mathbf{M}_1,\mathbf{M}_2,z)\right\| = \left\|z^{-2}\mathbf{M}_1^\top\mathbf{W}\mathbf{M}_2\right\| + O_p\left(\frac{\sigma_n}{a_n^3}\right) = O_p\left(\frac{1}{a_n^2}\right)\,. \tag{S.35}$$

By (S.32) and (S.35), we have

$$\left\|\left(-\mathbf{D}^{-1}+\mathbf{V}_-^\top\mathbf{G}(z)\mathbf{V}_-\right)^{-1} - (-\mathbf{D}+\mathcal{R}(\mathbf{V}_-,\mathbf{V}_-,z))^{-1}\right\|$$

$$\leq \left\|\mathbf{V}_-^\top\mathbf{G}(z)\mathbf{V}_- - \mathcal{R}(\mathbf{V}_-,\mathbf{V}_-,z)\right\|\left\|\left(-\mathbf{D}^{-1}+\mathbf{V}_-^\top\mathbf{G}(z)\mathbf{V}_-\right)^{-1}\right\|\left\|(-\mathbf{D}+\mathcal{R}(\mathbf{V}_-,\mathbf{V}_-,z))^{-1}\right\|$$

$$= O_p(1),\ \ z\in[a_n,b_n]\,. \tag{S.36}$$

Moreover, by (S.32), (S.33) and (S.35) we have

$$\left\|\left[\left(-\mathbf{D}^{-1}+\mathbf{V}_-^\top\mathbf{G}(z)\mathbf{V}_-\right)^{-1} - (-\mathbf{D}+\mathcal{R}(\mathbf{V}_-,\mathbf{V}_-,z))^{-1}\right]'\right\| \tag{S.37}$$

$$= \Big\|\left(-\mathbf{D}^{-1}+\mathbf{V}_-^\top\mathbf{G}(z)\mathbf{V}_-\right)^{-1}\mathbf{V}_-^\top\mathbf{G}'(z)\mathbf{V}_-\left(-\mathbf{D}^{-1}+\mathbf{V}_-^\top\mathbf{G}(z)\mathbf{V}_-\right)^{-1}$$

$$- (-\mathbf{D}+\mathcal{R}(\mathbf{V}_-,\mathbf{V}_-,z))^{-1}\mathcal{R}'(\mathbf{V}_-,\mathbf{V}_-,z)\,(-\mathbf{D}+\mathcal{R}(\mathbf{V}_-,\mathbf{V}_-,z))^{-1}\Big\|$$

$$= O\left\{\left\|\mathbf{V}_-^\top\mathbf{G}'(z)\mathbf{V}_- - \mathcal{R}'(\mathbf{V}_-,\mathbf{V}_-,z)\right\|\left\|\left(-\mathbf{D}^{-1}+\mathbf{V}_-^\top\mathbf{G}(z)\mathbf{V}_-\right)^{-1}\right\|^2\right\}$$

$$+ O\Big\{\left\|\left[-\mathbf{D}^{-1}+\mathbf{V}_-^\top\mathbf{G}(z)\mathbf{V}_-\right]^{-1} - (-\mathbf{D}+\mathcal{R}(\mathbf{V}_-,\mathbf{V}_-,z))^{-1}\right\|$$

$$\cdot\left(\left\|\left(-\mathbf{D}^{-1}+\mathbf{V}_-^\top\mathbf{G}(z)\mathbf{V}_-\right)^{-1}\right\| + \left\|\left(-\mathbf{D}^{-1}+\mathbf{V}_-^\top\mathbf{G}(z)\mathbf{V}_-\right)^{-1}\right\|\right)\|\mathcal{R}'(\mathbf{V}_-,\mathbf{V}_-,z)\|\Big\}$$

$$= O_p\left(\frac{1}{a_n}\right) + O_p\left(\frac{\sigma_n}{a_n^2}\right)\,,$$

and

$$\left\| \left\{ (-\mathbf{D} + \mathcal{R}(\mathbf{V}_-, \mathbf{V}_-, z))^{-1} \right\}' \right\| \tag{S.38}$$
$$= \left\| (-\mathbf{D} + \mathcal{R}(\mathbf{V}_-, \mathbf{V}_-, z))^{-1} \mathcal{R}'(\mathbf{V}_-, \mathbf{V}_-, z) (-\mathbf{D} + \mathcal{R}(\mathbf{V}_-, \mathbf{V}_-, z))^{-1} \right\|$$
$$= O(1), \ z \in [a_n, b_n] .$$

By (S.31)–(S.37), we have the following expansions

$$\begin{aligned} \mathbf{V}^\top \mathbf{F}(z)\mathbf{V} &= \mathbf{V}^\top \mathbf{G}(z)\mathbf{V}_- \left(-\mathbf{D}^{-1}\mathbf{I} + \mathbf{V}_-^\top \mathbf{G}(z)\mathbf{V}_-\right)^{-1} \mathbf{V}_-^\top \mathbf{G}(z)\mathbf{V} \\ &= \mathcal{R}(\mathbf{V}, \mathbf{V}_-, z) \left(-\mathbf{D}^{-1}\mathbf{I} + \mathcal{R}(\mathbf{V}_-, \mathbf{V}_-, z)\right)^{-1} \mathcal{R}(\mathbf{V}_-, \mathbf{V}, z) + \Delta_{n2} , \end{aligned} \tag{S.39}$$

and

$$\begin{aligned} \mathbf{V}^\top \mathbf{F}'(z)\mathbf{V} &= 2\mathbf{V}^\top \mathbf{G}'(z)\mathbf{V}_- \left(-\mathbf{D}^{-1} + \mathbf{V}_-^\top \mathbf{G}(z)\mathbf{V}_-\right)^{-1} \mathbf{V}_-^\top \mathbf{G}(z)\mathbf{V} \\ &\quad + \mathbf{V}^\top \mathbf{G}(z)\mathbf{V}_- \left\{ \left(-\mathbf{D}^{-1} + \mathbf{V}_-^\top \mathbf{G}(z)\mathbf{V}_-\right)^{-1} \right\}' \mathbf{V}_-^\top \mathbf{G}(z)\mathbf{V} \\ &= 2\mathcal{R}'(\mathbf{V}, \mathbf{V}_-, z) (-\mathbf{D} + \mathcal{R}(\mathbf{V}_-, \mathbf{V}_-, z))^{-1} \mathcal{R}(\mathbf{V}_-, \mathbf{V}, z) \\ &\quad + \mathcal{R}(\mathbf{V}, \mathbf{V}_-, z) \left\{ (-\mathbf{D} + \mathcal{R}(\mathbf{V}_-, \mathbf{V}_-, z))^{-1} \right\}' \mathcal{R}(\mathbf{V}_-, \mathbf{V}, z) \\ &\quad + \Delta_{n3} , \end{aligned} \tag{S.40}$$

where $\|\Delta_{n2}\| = O_p(\frac{\sigma_n^2}{a_n^4})$ and $\|\Delta_{n3}\| = O_p(\frac{1}{a_n^4}) + O_p(\frac{\sigma_n^3}{a_n^6})$.

Now we turn to (S.26). By (S.30), (S.32) and (S.35), we can see that $\|\mathbf{V}^\top \mathbf{G}(\widehat{t}_k)\mathbf{V}_-\| = O_p(\frac{1}{a_n^2})$, $|\mathbf{v}_1 \mathbf{G}(\widehat{t}_k)\mathbf{v}_2| = O_p(\frac{1}{a_n^2})$ and $|\mathbf{v}_{-1} \mathbf{G}(\widehat{t}_k)\mathbf{v}_{-2}| = O_p(\frac{1}{a_n^2})$. In other words, the off diagonal terms in the determinant (S.26) are all $O_p(\frac{1}{a_n^2})$.

The 3rd diagonal entry in the determinant (S.26) is $\mathbf{v}_{-1}^\top \mathbf{G}(\widehat{t}_k)\mathbf{v}_{-1} - \frac{1}{d_1}$. By (S.30), (S.32) and (S.35), we have $\mathbf{v}_{-1}^\top \mathbf{G}(\widehat{t}_k)\mathbf{v}_{-1} = -\frac{1}{d_k} + o_p(\frac{1}{a_n})$. i.e. $\mathbf{v}_{-1}^\top \mathbf{G}(\widehat{t}_k)\mathbf{v}_{-1} - \frac{1}{d_1} = -\frac{1}{d_k} - \frac{1}{d_1} + o_p(\frac{1}{a_n})$. Similarly, the 4th diagonal entry is $\mathbf{v}_{-2}^\top \mathbf{G}(\widehat{t}_k)\mathbf{v}_{-2} - \frac{1}{d_2} = -\frac{1}{d_k} - \frac{1}{d_2} + o_p(\frac{1}{a_n})$. Therefore the matrix $\mathbf{V}_-^\top \mathbf{G}(\widehat{t}_k)\mathbf{V}_- - \mathbf{D}^{-1}$ is invertible with probability tending to 1. Recalling the determinant formula for block structure matrix that

$$\det \begin{pmatrix} \mathbf{A} & \mathbf{B}^\top \\ \mathbf{B} & \mathbf{C} \end{pmatrix} = \det(\mathbf{C}) \det(\mathbf{A} - \mathbf{B}^\top \mathbf{C}^{-1}\mathbf{B}) ,$$

for any invertible matrix $\mathbf{C}$ and setting $\mathbf{C} = \mathbf{V}_-^\top \mathbf{G}(\widehat{t}_k)\mathbf{V}_- - \mathbf{D}$, we have with probability tending to 1,

$$\det(\mathbf{V}^\top (\mathbf{G}(\widehat{t}_k) - \mathbf{F}(\widehat{t}_k))\mathbf{V} + \mathbf{D}^{-1}) = 0 , \tag{S.41}$$

where $\mathbf{F}(z) = \mathbf{G}(z)\mathbf{V}_- \left(-\mathbf{D}^{-1} + \mathbf{V}_-^\top \mathbf{G}(z)\mathbf{V}_-\right)^{-1} \mathbf{V}_-^\top \mathbf{G}(z)$.

The three equations (S.31), (S.33) and (S.40) lead to

$$\|\mathbf{V}^\top (\mathbf{G}'(z) - \mathbf{F}'(z))\mathbf{V} - \frac{1}{z^2}\widetilde{\mathcal{P}}_z^{-1} - 2z^{-3}\mathbf{V}^\top \mathbf{W}\mathbf{V}\| = O_p\left(\frac{\sigma_n}{a_n^4}\right) , \tag{S.42}$$

for $z \in [a_n, b_n]$, where

$$\widetilde{\mathcal{P}}_z^{-1} = z^2 \left(\frac{A_{\mathbf{V},z}}{z}\right)' ,$$

and

$$A_{\mathbf{V},z} = \left\{ t\mathcal{R}(\mathbf{V}, \mathbf{V}, z) - z\mathcal{R}(\mathbf{V}, \mathbf{V}_-, z) (-\mathbf{D} + \mathcal{R}(\mathbf{V}_-, \mathbf{V}_-, z))^{-1} \mathcal{R}(\mathbf{V}_-, \mathbf{V}, z) \right\}^\top . \tag{S.43}$$

Further, recalling the definition in (S.43), it holds that

$$\frac{1}{z^2}\widetilde{\mathcal{P}}_z^{-1} = \left(\frac{A_{\mathbf{V},z}}{z}\right)' = \mathcal{R}'(\mathbf{V},\mathbf{V},z) - 2\mathcal{R}'(\mathbf{V},\mathbf{V}_-,z)\left(-\mathbf{D}+\mathcal{R}(\mathbf{V}_-,\mathbf{V}_-,z)\right)^{-1}$$
$$\times\mathcal{R}(\mathbf{V}_-,\mathbf{V},z) - \mathcal{R}(\mathbf{V},\mathbf{V}_-,z)\left\{\left(-\mathbf{D}+\mathcal{R}(\mathbf{V}_-,\mathbf{V}_-,z)\right)^{-1}\right\}'\mathcal{R}(\mathbf{V}_-,\mathbf{V},z)\,. \tag{S.44}$$

By (S.32), (S.33) and (S.38), we have

$$\|\widetilde{\mathcal{P}}_z^{-1} - \mathbf{I}\| = O\left(\frac{\sigma_n^2}{a_n^2}\right)\,.$$

Plugging this into (S.42) and by Lemmas 3, we have for all $z \in [a_n, b_n]$,

$$\|\mathbf{V}^\top\left(\mathbf{G}'(z) - \mathbf{F}'(z)\right)\mathbf{V} - z^{-2}\mathbf{I} - 2z^{-3}\mathbf{V}^\top\mathbf{W}\mathbf{V}\| = a_n^{-4}O_p(\sigma_n^2)\,. \tag{S.45}$$

Hence there exists a 2×2 random matrix $\mathbf{B}$ such that

$$\mathbf{V}^\top\left(\mathbf{G}'(z) - \mathbf{F}'(z)\right)\mathbf{V} = z^{-2}\mathbf{B}(z), \tag{S.46}$$

where $\|\mathbf{B}(z) - \mathbf{I}\| = O_p(a_n^{-1} + a_n^{-2}\sigma_n^2)$.

Further, in light of expressions (S.30) and (S.39), we can obtain the asymptotic expansion

$$\|\mathbf{I} + \mathbf{D}\mathbf{V}^\top\left(\mathbf{G}(z) - \mathbf{F}(z)\right)\mathbf{V} - f(z) + z^{-2}\mathbf{D}\mathbf{V}^\top\mathbf{W}\mathbf{V}\| = O_p(a_n^{-2}\sigma_n)\,, \tag{S.47}$$

for all $z \in [a_n, b_n]$, where $f(z)$ is defined in (11).

In view of (S.47) and the definition of t_k, we have

$$\left\|\mathbf{I} + \mathbf{D}\mathbf{V}^\top\left(\mathbf{G}(t_k) - \mathbf{F}(t_k)\right)\mathbf{V} - f(t_k) + t_k^{-2}\mathbf{D}\mathbf{V}^\top\mathbf{W}\mathbf{V}\right\| = O_p\left(\frac{\sigma_n}{a_n^2}\right)\,,\ k = 1,2\,. \tag{S.48}$$

By (S.41), (S.46) and (S.48), an application of the mean value theorem yields

$$0 = \det(\mathbf{I} + \mathbf{D}\mathbf{V}^\top\left(\mathbf{G}(\widehat{t}_k) - \mathbf{F}(\widehat{t}_k)\right)\mathbf{V}) = \det(\mathbf{I} + \mathbf{D}\mathbf{V}^\top\left(\mathbf{G}(t_1) - \mathbf{F}(t_1)\right)\mathbf{V}$$
$$+\mathbf{D}\tilde{\mathbf{B}}(\widehat{t}_k - t_1))\,,\ k = 1,2\,, \tag{S.49}$$

where $\tilde{\mathbf{B}} = (\tilde{B}_{ij}(\tilde{t}_{ij}))$, $\tilde{t}_{ij}^2\tilde{B}_{ij}(\tilde{t}_{ij}) = \delta_{ij} + O_p(a_n^{-1} + a_n^{-2}\sigma_n^2)$ by (S.46) and $\tilde{t}_{ij}$ is some number between t_1 and $\widehat{t}_k$. By (S.47), similar to (S.110)–(S.115), we can show that

$$|\widehat{t}_k - t_1| = O_p\left(1 + \frac{\sigma_n^2}{a_n}\right) + |d_1 - d_k|\,. \tag{S.50}$$

(S.49) can be rewritten as

$$0 = \det(\mathbf{I} + \mathbf{D}\mathbf{V}^\top\left(\mathbf{G}(\widehat{t}_k) - \mathbf{F}(\widehat{t}_k)\right)\mathbf{V}) = \det(\mathbf{I} + \mathbf{D}\mathbf{V}^\top\left(\mathbf{G}(t_1) - \mathbf{F}(t_1)\right)\mathbf{V}$$
$$+t_1^{-2}\mathbf{D}\mathbf{C}(\widehat{t}_k - t_1))\,,\ k = 1,2\,, \tag{S.51}$$

where

$$\|\mathbf{C} - \mathbf{I}\| = O_p\left(a_n^{-1} + a_n^{-2}\sigma_n^2 + \frac{d_1 - d_2}{a_n}\right)\,. \tag{S.52}$$

We know that $\widehat{t}_k - t_1$, $k = 1,2$ are the eigenvalues of $t_1^2\mathbf{C}^{-1}\mathbf{D}^{-1}\left(\mathbf{I} + \mathbf{D}\mathbf{V}^\top\left(\mathbf{G}(t_1) - \mathbf{F}(t_1)\right)\mathbf{V}\right)$. Combining (S.27) with the definition of $g(z)$ in (17), we have $g_{ij}(t_k) = O(\frac{\sigma_n^2}{a_n} + d_1 - d_2) + O_p(1)$, $1 \le i,j,k \le 2$. The asymptotic expansions in (S.48), (S.52) and Lemma 5 together with the condition (12) and (S.22) imply that

$$t_1^2\mathbf{C}^{-1}\mathbf{D}^{-1}\left(\mathbf{I} + \mathbf{D}\mathbf{V}^\top\left(\mathbf{G}(t_1) - \mathbf{F}(t_1)\right)\mathbf{V}\right) = g(t_1) + \Delta_{n4}\,, \tag{S.53}$$

where Δ_{n4} is a symmetric matrix with $\|\Delta_{n4}\| = o_p(1)$. By (S.53), we can rewrite (S.51) as follows,

$$\det(g(t_1) + \Delta_{n4} + (\widehat{t}_k - t_1)\mathbf{I}) = 0,\ k = 1, 2\,. \tag{S.54}$$

Moreover, by (17), the eigenvalues of $g(t_1)$ are

$$\frac{1}{2}\left[-g_{11}(t_1) - g_{22}(t_1) \pm \left\{(g_{11}(t_1) + g_{22}(t_1))^2 - 4\left(g_{11}(t_1)g_{22}(t_1) - g_{12}^2(t_1)\right)\right\}^{\frac{1}{2}}\right]. \tag{S.55}$$

Combining (S.54)–(S.55) with Weyl's inequality and noticing that $\widehat{t}_1 > \widehat{t}_2$, we have the following expansions

$$\widehat{t}_1 - t_1 = \frac{1}{2}\left[-g_{11}(t_1) - g_{22}(t_1) + \left\{(g_{11}(t_1) + g_{22}(t_1))^2 - 4\left(g_{11}(t_1)g_{22}(t_1) - g_{12}^2(t_1)\right)\right\}^{\frac{1}{2}}\right] + o_p(1)\,, \tag{S.56}$$

and

$$\widehat{t}_2 - t_1 = \frac{1}{2}\left[-g_{11}(t_1) - g_{22}(t_1) - \left\{(g_{11}(t_1) + g_{22}(t_1))^2 - 4\left(g_{11}(t_1)g_{22}(t_1) - g_{12}^2(t_1)\right)\right\}^{\frac{1}{2}}\right] + o_p(1). \tag{S.57}$$

Expanding the determinant at t_2 in (S.49) and repeating the process from (S.49)–(S.47), we also have

$$\widehat{t}_2 - t_2 = \frac{1}{2}\left[-g_{11}(t_2) - g_{22}(t_2) - \left\{(g_{11}(t_2) + g_{22}(t_2))^2 - 4\left(g_{11}(t_2)g_{22}(t_2) - g_{12}^2(t_2)\right)\right\}^{\frac{1}{2}}\right] + o_p(1). \tag{S.58}$$

S.6 More discussion of Proposition 1

In this section we show that the major terms at the right hand sides of (15) and (16) are meaningful, as shown in the following lemma.

Lemma 2.

$$\frac{1}{2}\left[-g_{11}(t_1) - g_{22}(t_1) + \left\{(g_{11}(t_1) + g_{22}(t_1))^2 - 4\left(g_{11}(t_1)g_{22}(t_1) - g_{12}^2(t_1)\right)\right\}^{\frac{1}{2}}\right] = O_p(1)\,, \tag{S.59}$$

and

$$\frac{1}{2}\left[-g_{11}(t_2) - g_{22}(t_2) - \left\{(g_{11}(t_2) + g_{22}(t_2))^2 - 4\left(g_{11}(t_1)g_{22}(t_2) - g_{12}^2(t_2)\right)\right\}^{\frac{1}{2}}\right] = O_p(1)\,. \tag{S.60}$$

Proof. The proofs of (S.59) and (S.60) are the same, so we only prove (S.59).

By Lemma 3, we have $g_{ij}(t_1) = \frac{t_1^2}{d_i} f_{ij}(t_1) + O_p(1)$. Therefore it suffices to show that

$$\frac{1}{2}\left[-\frac{t_1^2}{d_i}f_{11}(t_1) - \frac{t_1^2}{d_2}f_{22}(t_1) + \left\{(g_{11}(t_1) + g_{22}(t_1))^2 - 4\left(g_{11}(t_1)g_{22}(t_1) - g_{12}^2(t_1)\right)\right\}^{\frac{1}{2}}\right] = O_p(1)\,.$$

By Lemma 3, for any $\epsilon > 0$, there exists a constant M_0 such that

$$\mathbb{P}\left(\|\mathbf{V}^\top \mathbf{W}\mathbf{V}\| \geq M_0\right) \leq \epsilon\,.$$

Now we consider the inequality constraint on the event $\{\|\mathbf{V}^\top \mathbf{W}\mathbf{V}\| \leq M_0\}$. Let $h_1 = \frac{t_1^2}{d_1} f_{11}(t_1) + \frac{t_1^2}{d_2} f_{22}(t_1)$. It follows from the definition of t_1, (S.94), (S.109) and (S.110) that

$$f_{11}(t_1) \geq 0\,,\ \text{and}\ f_{22}(t_1) \geq 0\,.$$

Let

$$h_2 = 2h_1(\mathbf{v}_1^\top \mathbf{W}\mathbf{v}_1 + \mathbf{v}_2^\top \mathbf{W}\mathbf{v}_2) - 4\frac{t_1^2}{d_1} f_{11}(t_1)\mathbf{v}_2^\top \mathbf{W}\mathbf{v}_2 - 4\frac{t_1^2}{d_2} f_{22}(t_1)\mathbf{v}_1^\top \mathbf{W}\mathbf{v}_1 + 4t_1^2\left(\frac{f_{12}(t_1)}{d_1} + \frac{f_{21}(t_1)}{d_2}\right)\mathbf{v}_1^\top \mathbf{W}\mathbf{v}_2\,,$$

and

$$h_3 = (\mathbf{v}_1^\top \mathbf{W} \mathbf{v}_1 - \mathbf{v}_2^\top \mathbf{W} \mathbf{v}_2)^2 + 4(\mathbf{v}_1^\top \mathbf{W} \mathbf{v}_2)^2 \,.$$

By the definition of g and the above equations, we have

$$(g_{11}(t_1) + g_{22}(t_1))^2 - 4\left(g_{11}(t_1)g_{22}(t_1) - g_{12}^2(t_1)\right) = h_1^2 + h_2 + h_3 \,.$$

Note that $|h_2| \le M_1|h_1|$ and $|h_3| \le M_2$, where M_1 and M_2 are polynomial functions of M_0 (depending on M_0 only). Now we consider two cases:

1. $|h_3| \le |h_1|$, then we have $|h_2 + h_3| \le (M_2 + 1)|h_1|$. Then

$$\begin{aligned}&\left|-\frac{t_1^2}{d_1} f_{11}(t_1) - \frac{t_1^2}{d_2} f_{22}(t_1) + \left\{(g_{11}(t_1) + g_{22}(t_1))^2 - 4\left(g_{11}(t_1)g_{22}(t_1) - g_{12}^2(t_1)\right)\right\}^{\frac{1}{2}}\right| \\ &= | - h_1 + (h_1^2 + h_2 + h_3)^{\frac{1}{2}}| = \frac{|h_2 + h_3|}{h_1 + (h_1^2 + h_2 + h_3)^{\frac{1}{2}}} \le M_2 + 1 \,.\end{aligned}$$

2. $|h_3| \ge |h_1|$, then

$$\begin{aligned}&\left|-\frac{t_1^2}{d_1} f_{11}(t_1) - \frac{t_1^2}{d_2} f_{22}(t_1) + \left\{(g_{11}(t_1) + g_{22}(t_1))^2 - 4\left(g_{11}(t_1)g_{22}(t_1) - g_{12}^2(t_1)\right)\right\}^{\frac{1}{2}}\right| \\ &= | - h_1 + (h_1^2 + h_2 + h_3)^{\frac{1}{2}}| \le (M_2 + 1)^2 + M_1 M_2 \,.\end{aligned} \tag{S.61}$$

Combining the two cases, we have shown that given $\|\mathbf{V}^\top \mathbf{W} \mathbf{V}\| \le M_0$, there exists M_3 depending on M_0 only such that

$$\left|\frac{1}{2}\left[-g_{11}(t_1) - g_{22}(t_1) + \left\{(g_{11}(t_1) + g_{22}(t_1))^2 - 4\left(g_{11}(t_1)g_{22}(t_1) - g_{12}^2(t_1)\right)\right\}^{\frac{1}{2}}\right]\right| \le M_3 \,.$$

In other words,

$$\frac{1}{2}\left[-g_{11}(t_1) - g_{22}(t_1) + \left\{(g_{11}(t_1) + g_{22}(t_1))^2 - 4\left(g_{11}(t_1)g_{22}(t_1) - g_{12}^2(t_1)\right)\right\}^{\frac{1}{2}}\right] = O_p(1) \,.$$

This concludes the proof of Lemma 2. □

S.7 Proof of Theorem 2

By Lemma 4 and weyl's inequality $|\widehat{t}_k - d_k| \le \|\mathbf{W}\|$, $k = 1, 2$, we have

$$\mathbb{P}\left(\widehat{t}_2 \ge d_2 - C_0 \max\{n^{\frac{1}{2}}, p^{\frac{1}{2}}\}\right) \ge 1 - n^{-2} \,,$$

and

$$\mathbb{P}\left(\widehat{t}_1 \le d_1 + C_0 \max\{n^{\frac{1}{2}}, p^{\frac{1}{2}}\}\right) \ge 1 - n^{-2} \,,$$

for some positive constant C_0 and sufficiently large n. Combining the above two equations with $d_1 \gg \sigma_n$, and $d_1/d_2 \le 1 + n^{-c}$, we have

$$\mathbb{P}\left(\frac{\widehat{t}_1}{\widehat{t}_2} \ge 1 + C\left(\frac{\sigma_n}{d_1} + \frac{1}{n^c}\right)\right) \to 0 \,,$$

where C is some positive constant.

S.8 Proof of Theorem 3

By Lemma 4, there exists a constant $C > 0$ such that

$$\mathbb{P}\left(\|\mathbf{W}\| \ge C\max\{n^{\frac{1}{2}}, p^{\frac{1}{2}}\}\right) \le n^{-D}. \tag{S.62}$$

By Weyl's inequality, we have

$$\max_{i=1,2} |\widehat{t}_i - d_i| \le \|\mathbf{W}\|\,. \tag{S.63}$$

By (S.63) and the condition that $d_1 \ge (1+c)d_2$, we have

$$\frac{\widehat{t}_1}{\widehat{t}_2} \ge \frac{d_1 - \|\mathbf{W}\|}{d_2 + \|\mathbf{W}\|} \ge \frac{1 + c - \frac{\|\mathbf{W}\|}{d_2}}{1 + \frac{\|\mathbf{W}\|}{d_2}}\,. \tag{S.64}$$

If $d_2 \ge \frac{c}{c+4} C\max\{n^{\frac{1}{2}}, p^{\frac{1}{2}}\}$, by (S.62) and (S.64), we have

$$\mathbb{P}\left(\frac{\widehat{t}_1}{\widehat{t}_2} \le 1 + \frac{c}{2}\right) \le \mathbb{P}\left(\frac{1 + c - \frac{\|\mathbf{W}\|}{d_2}}{1 + \frac{\|\mathbf{W}\|}{d_2}} \le 1 + \frac{c}{2}\right) \le n^{-D}$$

If $d_2 < \frac{c}{c+4} C\max\{n^{\frac{1}{2}}, p^{\frac{1}{2}}\}$, by the condition that $d_1 \gg \sigma_n$, (S.62) and (S.64), for sufficiently large n we have

$$\mathbb{P}\left(\frac{\widehat{t}_1}{\widehat{t}_2} \le 1 + \frac{c}{2}\right) \le n^{-D}\,. \tag{S.65}$$

This together with the assumption that $d_1/d_2 \ge 1 + c$ implies (20). Now we turn to (21). Let $\widehat{\mathbf{u}}_1 = \sqrt{2}(\widehat{\mathbf{v}}_1(1), \ldots, \widehat{\mathbf{v}}_1(n))^\top$ and $\widehat{\mathrm{u}}_1 = \sqrt{2}(\widehat{\mathbf{v}}_1(n+1), \ldots, \widehat{\mathbf{v}}_1(n+p))^\top$. Notice that $\widehat{\mathbf{v}}_1$ is the unit eigenvector of $\mathcal{Z}$ corresponding to $\widehat{d}_1$. By the definition of $\mathcal{Z}$, we know that $\widehat{\mathbf{u}}_1$ is the unit eigenvector of $\mathbf{X}^\top\mathbf{X}$ corresponding to $\widehat{d}_1^2$ and $\widehat{\mathrm{u}}_1$ is the unit eigenvector of $\mathbf{X}\mathbf{X}^\top$ corresponding to $\widehat{d}_1^2$. Similarly, by the condition that all of the entries of $\mathbf{u}_1$ are equal, we imply that the first entries of $\mathbf{v}_1$ are equal to $(2n)^{-1/2}$. By the second inequality of Theorem 10 in the supplement of Cai et al. (2013), we obtain that

$$2 - 2(\mathbf{v}_1^\top \widehat{\mathbf{v}}_1)^2 \le \frac{\|\mathbf{W}\|}{d_1 - d_2 - \|\mathbf{W}\|}\,. \tag{S.66}$$

Since $d_1/d_2 \ge 1 + c$, we have

$$d_1 - d_2 \ge c(1+c)^{-1} d_1\,. \tag{S.67}$$

Let $C_0 = \max\{c(1+c)^{-1}, C\} - 1$, where C is given in (S.62). By (S.62), (S.66) and (S.67), we imply that

$$\mathbb{P}\Big(2 - 2(\mathbf{v}_1^\top \widehat{\mathbf{v}}_1)^2 \le \frac{(C_0 + 1)(\frac{\sigma_n}{d_1})^{2/3}}{C_0}\Big) \ge 1 - n^{-D}\,.$$

$$\mathbb{P}\Big(|\mathbf{v}_1^\top \widehat{\mathbf{v}}_1| \ge 1 - \sqrt{\frac{\sigma_n}{d_1}}\Big) \ge 1 - n^{-D}\,, \tag{S.68}$$

where $n \ge n_0(\epsilon, D)$. Notice that $\widehat{\mathbf{u}}_1$ is a unit vector, we have

$$|\mathbf{v}_1^\top \widehat{\mathbf{v}}_1| \le \frac{1}{\sqrt{2n}} |\mathbf{1}_n^\top \widehat{\mathbf{u}}_1| + \frac{1}{2}\,.$$

This together with (S.68) implies that

$$\mathbb{P}\left(\left|\left(\frac{1}{n}\right)^{\frac{1}{2}} |\mathbf{1}_n^\top \widehat{\mathbf{u}}_1| - \left(\frac{1}{2}\right)^{\frac{1}{2}}\right| \ge \sqrt{\frac{\sigma_n}{d_1}}\right) \le n^{-D}\,. \tag{S.69}$$

This completes the proof.

S.9 Proof of Theorem 4

Let $\widehat{\mathbf{u}}_k = \sqrt{2}(\widehat{\mathbf{v}}_k(1), \ldots, \widehat{\mathbf{v}}_k(n))^\top$ and $\widehat{\mathbf{u}}_k = \sqrt{2}(\widehat{\mathbf{v}}_k(n+1), \ldots, \widehat{\mathbf{v}}_k(n+p))^\top$. Notice that $\widehat{\mathbf{v}}_k$ is the unit eigenvector of $\mathcal{Z}$ corresponding to $\widehat{d}_k$. By the definition of $\mathcal{Z}$, we know that $\widehat{\mathbf{u}}_k$ is the unit eigenvector of $\mathbf{X}^\top\mathbf{X}$ corresponding to $\widehat{d}_k^2$ and $\widehat{\mathbf{u}}_k$ is the unit eigenvector of $\mathbf{X}\mathbf{X}^\top$ corresponding to $\widehat{d}_k^2$. By the second inequality of Theorem 10 in the supplement of Cai et al. (2013), we obtain that

$$\|\widehat{\mathbf{V}}\widehat{\mathbf{V}}^\top - \mathbf{V}\mathbf{V}^\top\|_F \le \frac{\sqrt{2}K_0\|\mathbf{W}\|}{d_{K_0} - \|\mathbf{W}\|}\,. \tag{S.70}$$

By Lemma 2.4 of Jin et al. (2016), there exists an orthogonal matrix $\mathbf{O} = (\mathbf{o}_1, \ldots, \mathbf{o}_{2K_0})$ such that

$$\|\widehat{\mathbf{V}} - \mathbf{V}\mathbf{O}\|_F \le \|\widehat{\mathbf{V}}\widehat{\mathbf{V}}^\top - \mathbf{V}\mathbf{V}^\top\|_F\,.$$

Combining this with (S.70), we have

$$\|\widehat{\mathbf{V}} - \mathbf{V}\mathbf{O}\|_F \le \frac{\sqrt{2}K_0\|\mathbf{W}\|}{d_{K_0} - \|\mathbf{W}\|}\,. \tag{S.71}$$

By Lemma 4 and (S.71), we have

$$\mathbb{P}\left(\max_{1\le k\le K_0} \|\mathbf{v}_k - \mathbf{V}\mathbf{o}_k\| \ge \sqrt{\frac{\sigma_n}{d_{K_0}}}\right) \le n^{-D}\,. \tag{S.72}$$

The proof is completed by Cauchy-Schwarz inequality that

$$|\mathbf{x}^\top(\mathbf{u}_k - \mathbf{U}\mathbf{o}_k)| \le \sqrt{2}\|\mathbf{x}\|\|\mathbf{v}_k - \mathbf{V}\mathbf{o}_k\| = \|\mathbf{v}_k - \mathbf{V}\mathbf{o}_k\|\,.$$

S.10 Proof of Theorem 5

This proof idea is similar to the proof of Theorem 6 in Fan et al. (2020), where $\widehat{\Sigma} = \frac{1}{n}\sum_{i=1}^n \mathbf{w}_i\mathbf{w}_i^\top$. First of all, by Assumptions 1(i), 2–3 and similar proof as Lemma 1 in Fan et al. (2020), we have

$$\|\mathrm{diag}(\widehat{\Phi})\mathrm{diag}(\Phi)^{-1} - \mathbf{I}\|_\infty = o_p(1)\,. \tag{S.73}$$

Therefore, similar to the proof of Theorem 2 in the supplement of Fan et al. (2020), we can show that

$$\max_{K_0+1\le j\le p} |\lambda_j(\widehat{\mathbf{R}}) - \lambda_j(\mathrm{diag}(\Phi)^{-1/2}\widehat{\Phi}\mathrm{diag}(\Phi)^{-1/2})| = o_p(1)\,. \tag{S.74}$$

Combining this with Weyl's inequality, we have

$$\lambda_{K_0+1}(\widehat{\mathbf{R}}) \le \lambda_1(\mathrm{diag}(\Phi)^{-1/2}\widehat{\Sigma}\mathrm{diag}(\Phi)^{-1/2}) + o_p(1)\,. \tag{S.75}$$

By similar arguments as Lemma S.6 in Fan et al. (2020), we can show that

$$\lambda_1(\mathrm{diag}(\Phi)^{-1/2}\widehat{\Sigma}\mathrm{diag}(\Phi)^{-1/2}) \le \lambda_1(\mathrm{diag}(\Phi)^{-1}\Sigma)(1+\sqrt{\frac{p}{n}})\psi(\lambda_1(\mathrm{diag}(\Phi)^{-1}\Sigma)(1+\sqrt{\frac{p}{n}})) + o_p(1)\,, \tag{S.76}$$

where $\psi(x) = 1 + \frac{p}{n}\int \frac{t}{x-t}dH(t)$. By (24) of Fan et al. (2020), (S.74), we have a similar result as the last formula on page S36 of Fan et al. (2020) that

$$\psi(\underline{m}(\lambda_j(\widehat{\mathbf{R}}))) - \psi(\underline{m}_{n,j}(\lambda_j(\widehat{\mathbf{R}}))) = o_p(1),\ j \in [K_0+1, K]\,. \tag{S.77}$$

It follows from (S.76), (S.77), and the monotonicity of $x\psi(x)$ $(x > 1 + \sqrt{p/n})$ that

$$\lambda^C_{K_0+1}(\widehat{\mathbf{R}}) \le 1 + \sqrt{p/n} + o_p(1)\,,$$

which implies that

$$\mathbb{P}(\widehat{K}_0 \le K_0) \to 1\,.$$

Moreover, by (S.73), we have

$$\lambda^C_{K_0}(\widehat{\mathbf{R}}) \ge (1+o_p(1))\lambda_{K_0}(\mathrm{diag}(\Phi)^{-1/2}\widehat{\Phi}\mathrm{diag}(\Phi)^{-1/2}). \tag{S.78}$$

Notice that we have the linearization matrices of $\mathrm{diag}(\Phi)^{-1/2}\widehat{\Phi}\mathrm{diag}(\Phi)^{-1/2}$ and $\mathbf{R}$, which are

$$\begin{pmatrix} 0 & \frac{1}{\sqrt{n}}\mathrm{diag}(\Phi)^{-1/2}\mathbf{X}^\top \\ \frac{1}{\sqrt{n}}\mathbf{X}\mathrm{diag}(\Phi)^{-1/2} & 0 \end{pmatrix} \text{ and } \begin{pmatrix} 0 & \frac{1}{\sqrt{n}}\mathrm{diag}(\Phi)^{-1/2}\mathbb{E}\mathbf{X}^\top \\ \frac{1}{\sqrt{n}}\mathbb{E}\mathbf{X}\mathrm{diag}(\Phi)^{-1/2} & 0 \end{pmatrix},$$

we have

$$\begin{aligned}&\begin{pmatrix} 0 & \frac{1}{\sqrt{n}}\mathrm{diag}(\Phi)^{-1/2}\mathbf{X}^\top \\ \frac{1}{\sqrt{n}}\mathbf{X}\mathrm{diag}(\Phi)^{-1/2} & 0 \end{pmatrix} - \begin{pmatrix} 0 & \frac{1}{\sqrt{n}}\mathrm{diag}(\Phi)^{-1/2}\mathbb{E}\mathbf{X}^\top \\ \frac{1}{\sqrt{n}}\mathbb{E}\mathbf{X}\mathrm{diag}(\Phi)^{-1/2} & 0 \end{pmatrix}\\ &= \begin{pmatrix} 0 & \frac{1}{\sqrt{n}}\mathrm{diag}(\Phi)^{-1/2}\mathbf{W}^\top \\ \frac{1}{\sqrt{n}}\mathbf{W}\mathrm{diag}(\Phi)^{-1/2} & 0 \end{pmatrix}.\end{aligned}$$

Therefore, by weyl's inequality and (S.76), we have

$$|\lambda_{K_0}(\mathrm{diag}(\Phi)^{-1/2}\widehat{\Phi}\mathrm{diag}(\Phi)^{-1/2}) - \lambda_{K_0}(\mathbf{R})| = O_p(\sqrt{\lambda_{K_0}(\mathbf{R})}).$$

This together with Assumption 2, we have

$$\left|\frac{\lambda_{K_0}(\mathrm{diag}(\Phi)^{-1/2}\widehat{\Phi}\mathrm{diag}(\Phi)^{-1/2})}{\lambda_{K_0}(\mathbf{R})} - 1\right| = o_p(1). \tag{S.79}$$

By (S.78) and (S.79), it is easy to see that with probability tending to 1,

$$\lambda^C_{K_0}(\widehat{\mathbf{R}}) \to \infty\,,$$

which implies that

$$\mathbb{P}(\widehat{K}_0 \ge K_0) \to 1\,.$$

Combining the above two probabilities together, the proof of the first statement of Theorem 5 is completed. Now we move on to the second statement. By the properties of conditional probability, we have

$$\begin{aligned}
&\mathbb{P}\left(\max_{1\le k\le \widehat{K}_0} |\mathbf{x}^\top(\widehat{\mathbf{u}}_k - \mathbf{U}\mathbf{o}_k)| \le \sqrt{\frac{2\sigma_n}{d_{\widehat{K}_0}}}\right) = \mathbb{P}\left(\max_{1\le k\le \widehat{K}_0} |\mathbf{x}^\top(\widehat{\mathbf{u}}_k - \mathbf{U}\mathbf{o}_k)| \le \sqrt{\frac{2\sigma_n}{d_{K_0}}}\,\middle|\,\widehat{K}_0 = K_0\right)\mathbb{P}\left(\widehat{K}_0 = K_0\right)\\
&+\mathbb{P}\left(\max_{1\le k\le \widehat{K}_0} |\mathbf{x}^\top(\widehat{\mathbf{u}}_k - \mathbf{U}\mathbf{o}_k)| \le \sqrt{\frac{2\sigma_n}{d_{\widehat{K}_0}}}\,\middle|\,\widehat{K}_0 \ne K_0\right)\mathbb{P}\left(\widehat{K}_0 \ne K_0\right)\\
&=\mathbb{P}\left(\max_{1\le k\le \widehat{K}_0} |\mathbf{x}^\top(\widehat{\mathbf{u}}_k - \mathbf{U}\mathbf{o}_k)| \le \sqrt{\frac{2\sigma_n}{d_{K_0}}}\,\middle|\,\widehat{K}_0 = K_0\right)\mathbb{P}\left(\widehat{K}_0 = K_0\right) + o(1)\\
&=\mathbb{P}\left(\max_{1\le k\le \widehat{K}_0} |\mathbf{x}^\top(\widehat{\mathbf{u}}_k - \mathbf{U}\mathbf{o}_k)| \le \sqrt{\frac{2\sigma_n}{d_{K_0}}}\,\middle|\,\widehat{K}_0 = K_0\right)\mathbb{P}\left(\widehat{K}_0 = K_0\right)\\
&+\mathbb{P}\left(\max_{1\le k\le \widehat{K}_0} |\mathbf{x}^\top(\widehat{\mathbf{u}}_k - \mathbf{U}\mathbf{o}_k)| \le \sqrt{\frac{2\sigma_n}{d_{K_0}}}\,\middle|\,\widehat{K}_0 \ne K_0\right)\mathbb{P}\left(\widehat{K}_0 \ne K_0\right) + o(1)\\
&=\mathbb{P}\left(\max_{1\le k\le \widehat{K}_0} |\mathbf{x}^\top(\widehat{\mathbf{u}}_k - \mathbf{U}\mathbf{o}_k)| \le \sqrt{\frac{2\sigma_n}{d_{K_0}}}\right) + o(1) = 1 - o(1)\,.
\end{aligned} \tag{S.80}$$

Then the proof of Theorem 5 is completed.

S.11 Technical Lemmas and their proofs

Lemma 3. *Take (i) in Assumption 1. For* $\mathbf{X}$ *we considered in this paper and any positive integer* l*, there exists a positive constant* C_l *(depending on* l*) such that*

$$\mathbb{E}|\mathbf{x}^\top(\mathbf{W}^l - \mathbb{E}\mathbf{W}^l)\mathbf{y}|^2 \le C_l\sigma_n^{l-1}\,, \tag{S.81}$$

and $\mathbb{E}\mathbf{x}^\top\mathbf{W}\mathbf{y} = 0$ *and*

$$|\mathbb{E}\mathbf{x}^\top\mathbf{W}^l\mathbf{y}| \le C_l\sigma_n^l\,,\ \textit{for}\ l \ge 2. \tag{S.82}$$

where $\mathbf{x}$ *and* $\mathbf{y}$ *are two unit vectors (random or not random) independent of* $\mathbf{W}$*.*

Proof. Let $\mathcal{Y} = \Sigma^{-\frac{1}{2}}(\mathbf{X} - \mathbb{E}\mathbf{X})$. Recall that $\mathbf{X} = (X_1, \dots, X_n)$ is defined in (1) by

$$X_i = Y_i\boldsymbol{\mu}_1 + (1 - Y_i)\boldsymbol{\mu}_2 + W_i,\ i = 1, \dots, n\,,$$

where $\{W_i\}_{i=1}^n$ are i.i.d. from $\mathcal{N}(0, \Sigma)$. The entries of $\mathcal{Y}$ are i.i.d. standard normal random variables. Moreover, we decompose $\mathbf{W}$ defined in (9) by

$$\mathbf{W} = \begin{pmatrix} \mathbf{I} & 0 \\ 0 & \Sigma^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} 0 & \mathcal{Y}^\top \\ \mathcal{Y} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{I} & 0 \\ 0 & \Sigma^{\frac{1}{2}} \end{pmatrix}.$$

Let the eigen decomposition of Σ be $\mathbf{U}\Lambda\mathbf{U}^\top$. Since the entries of $\mathcal{Y}$ are i.i.d. standard normal random variables, we have $\mathcal{Y} \overset{d}{=} \mathbf{U}\mathcal{Y}$. Then $\mathbf{W}$ can be written as

$$\mathbf{W} \overset{d}{=} \begin{pmatrix} \mathbf{I} & 0 \\ 0 & \mathbf{U} \end{pmatrix} \begin{pmatrix} 0 & \mathcal{Y}^\top\Lambda \\ \Lambda\mathcal{Y} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{I} & 0 \\ 0 & \mathbf{U}^\top \end{pmatrix}.$$

Therefore

$$\mathbf{x}^\top\mathbf{W}^l\mathbf{y} = \mathbf{x}^\top \begin{pmatrix} \mathbf{I} & 0 \\ 0 & \mathbf{U} \end{pmatrix} \begin{pmatrix} 0 & \mathcal{Y}^\top\Lambda \\ \Lambda\mathcal{Y} & 0 \end{pmatrix}^L \begin{pmatrix} \mathbf{I} & 0 \\ 0 & \mathbf{U}^\top \end{pmatrix} \mathbf{y}\,.$$

Let $\widetilde{\mathbf{x}} = \begin{pmatrix} \mathbf{I} & 0 \\ 0 & \mathbf{U}^\top \end{pmatrix}\mathbf{x}$, $\widetilde{\mathbf{y}} = \begin{pmatrix} \mathbf{I} & 0 \\ 0 & \mathbf{U}^\top \end{pmatrix}\mathbf{y}$ and $\widetilde{\mathbf{W}} = \begin{pmatrix} 0 & \mathcal{Y}^\top\Lambda \\ \Lambda\mathcal{Y} & 0 \end{pmatrix}$, then we have

$$\mathbf{x}^\top\mathbf{W}^l\mathbf{y} = \widetilde{\mathbf{x}}^\top\widetilde{\mathbf{W}}^l\widetilde{\mathbf{y}}\,, \tag{S.83}$$

where above diagonal entries of $\widetilde{\mathbf{W}} = (\widetilde{w}_{ij})_{1\le i,j\le n}$ are independent normal random variables such that for any positive integer r,

$$\max_{1\le i,j\le n} \mathbb{E}|\widetilde{w}_{ij}|^r \le \|\Sigma\|^r c_r\,, \tag{S.84}$$

where c_r is the r-th moment of standard normal distribution. Actually, if $\{\widetilde{w}_{ij}\}_{1\le i,j\le n}$ were bounded random variables with

$$\max_{1\le i,j\le n} |\widetilde{w}_{ij}| \le 1\,, \tag{S.85}$$

then Lemmas 4 and 5 of Fan et al. (2020+) imply that there exists a positive constant c_l depending on l such that

$$\mathbb{E}|\widetilde{\mathbf{x}}^\top(\widetilde{\mathbf{W}}^l - \mathbb{E}\widetilde{\mathbf{W}}^l)\widetilde{\mathbf{y}}|^2 \le c_l\sigma_n^{l-1}\,, \tag{S.86}$$

and

$$|\mathbb{E}\widetilde{\mathbf{x}}^\top\widetilde{\mathbf{W}}^l\widetilde{\mathbf{y}}| \le c_l\sigma_n^l\,. \tag{S.87}$$

To establish Lemma 3, it remains to relax the bounded restriction (S.85). In other words, we need to replace the condition (S.85) by the condition of $\widetilde{w}_{ij}$, $1 \le i, j \le n$ in (S.84). We highlight the difference of the proof. Expanding $\mathbb{E}(\widetilde{\mathbf{x}}^\top \widetilde{\mathbf{W}}^l \widetilde{\mathbf{y}} - \mathbb{E}\widetilde{\mathbf{x}}^\top \widetilde{\mathbf{W}}^l \widetilde{\mathbf{y}})^2$ yields

$$
\begin{aligned}
&\mathbb{E}|\mathbf{x}^\top(\mathbf{W}^l - \mathbb{E}\mathbf{W}^l)\mathbf{y}|^2 = \mathbb{E}(\widetilde{\mathbf{x}}^\top \widetilde{\mathbf{W}}^l \widetilde{\mathbf{y}} - \mathbb{E}\widetilde{\mathbf{x}}^\top \widetilde{\mathbf{W}}^l \widetilde{\mathbf{y}})^2 \\
&= \sum_{\substack{1\le i_1,\cdots,i_{l+1},j_1,\cdots,j_{l+1}\le n,\\ i_s\ne i_{s+1},\, j_s\ne j_{s+1},\, 1\le s\le l}} \mathbb{E}\Big((\widetilde{x}_{i_1}\widetilde{w}_{i_1i_2}\widetilde{w}_{i_2i_3}\cdots\widetilde{w}_{i_li_{l+1}}\widetilde{y}_{i_{l+1}} - \mathbb{E}\widetilde{x}_{i_1}\widetilde{w}_{i_1i_2}\widetilde{w}_{i_2i_3}\cdots\widetilde{w}_{i_li_{l+1}}\widetilde{y}_{i_{l+1}}) \\
&\quad \times (\widetilde{x}_{j_1}\widetilde{w}_{j_1j_2}\widetilde{w}_{j_2j_3}\cdots\widetilde{w}_{j_lj_{l+1}}\widetilde{y}_{j_{l+1}} - \mathbb{E}\widetilde{x}_{j_1}\widetilde{w}_{j_1j_2}\widetilde{w}_{j_2j_3}\cdots\widetilde{w}_{j_lj_{l+1}}\widetilde{y}_{j_{l+1}}) \Big).
\end{aligned} \tag{S.88}
$$

Let $\mathbf{i} = (i_1, \dots, i_{l+1})$ and $\mathbf{j} = (j_1, \dots, j_{l+1})$ with $1 \le i_1, \cdots, i_{l+1}, j_1, \cdots, j_{l+1} \le n$, $i_s \ne i_{s+1}$, $j_s \ne j_{s+1}$, $1 \le s \le l$. We define an undirected graph $\mathcal{G}_\mathbf{i}$ whose vertices represent $i_1, \dots, i_{l+1}$ in $\mathbf{i}$, and only i_s and i_{s+1}, for $s = 1, \dots, l$, are connected in $\mathcal{G}_\mathbf{i}$. Similarly we can define $\mathcal{G}_\mathbf{j}$. By the definitions of $\mathcal{G}_\mathbf{i}$ and $\mathcal{G}_\mathbf{j}$, for each term

$$
\begin{aligned}
&\mathbb{E}\Big((\widetilde{x}_{i_1}\widetilde{w}_{i_1i_2}\widetilde{w}_{i_2i_3}\cdots\widetilde{w}_{i_li_{l+1}}\widetilde{y}_{i_{l+1}} - \mathbb{E}\widetilde{x}_{i_1}\widetilde{w}_{i_1i_2}\widetilde{w}_{i_2i_3}\cdots\widetilde{w}_{i_li_{l+1}}\widetilde{y}_{i_{l+1}}) \\
&\quad \times (\widetilde{x}_{j_1}\widetilde{w}_{j_1j_2}\widetilde{w}_{j_2j_3}\cdots\widetilde{w}_{j_lj_{l+1}}\widetilde{y}_{j_{l+1}} - \mathbb{E}\widetilde{x}_{j_1}\widetilde{w}_{j_1j_2}\widetilde{w}_{j_2j_3}\cdots\widetilde{w}_{j_lj_{l+1}}\widetilde{y}_{j_{l+1}}) \Big),
\end{aligned}
$$

there exists a one to one corresponding graph $\mathcal{G}_\mathbf{i} \cup \mathcal{G}_\mathbf{j}$ for $\{\widetilde{w}_{i_si_{s+1}}\}_{s=1}^l \cup \{\widetilde{w}_{j_sj_{s+1}}\}_{s=1}^l$. If $\mathcal{G}_\mathbf{i}$ and $\mathcal{G}_\mathbf{j}$ are not connected, $\widetilde{w}_{i_1i_2}\widetilde{w}_{i_2i_3}\cdots\widetilde{w}_{i_li_{l+1}}$ and $\widetilde{w}_{j_1j_2}\widetilde{w}_{j_2j_3}\cdots\widetilde{w}_{j_lj_{l+1}}$ are independent, therefore we have

$$
\begin{aligned}
&\mathbb{E}\Big((\widetilde{x}_{i_1}\widetilde{w}_{i_1i_2}\widetilde{w}_{i_2i_3}\cdots\widetilde{w}_{i_li_{l+1}}\widetilde{y}_{i_{l+1}} - \mathbb{E}\widetilde{x}_{i_1}\widetilde{w}_{i_1i_2}\widetilde{w}_{i_2i_3}\cdots\widetilde{w}_{i_li_{l+1}}\widetilde{y}_{i_{l+1}}) \\
&\quad \times (\widetilde{x}_{j_1}\widetilde{w}_{j_1j_2}\widetilde{w}_{j_2j_3}\cdots\widetilde{w}_{j_lj_{l+1}}\widetilde{y}_{j_{l+1}} - \mathbb{E}\widetilde{x}_{j_1}\widetilde{w}_{j_1j_2}\widetilde{w}_{j_2j_3}\cdots\widetilde{w}_{j_lj_{l+1}}\widetilde{y}_{j_{l+1}}) \Big) = 0.
\end{aligned} \tag{S.89}
$$

Therefore we have

$$
\begin{aligned}
\text{L.H.S. of (S.81)} &= \sum_{\substack{\mathbf{i},\mathbf{j},\mathcal{G}_\mathbf{i} \text{ and } \mathcal{G}_\mathbf{j} \text{ are connected},\\ i_s\ne i_{s+1},\, j_s\ne j_{s+1},\, 1\le s\le l,}} \mathbb{E}\Big((\widetilde{x}_{i_1}\widetilde{w}_{i_1i_2}\widetilde{w}_{i_2i_3}\cdots\widetilde{w}_{i_li_{l+1}}\widetilde{y}_{i_{l+1}} - \mathbb{E}\widetilde{x}_{i_1}\widetilde{w}_{i_1i_2}\widetilde{w}_{i_2i_3}\cdots\widetilde{w}_{i_li_{l+1}}\widetilde{y}_{i_{l+1}}) \\
&\quad \times (\widetilde{x}_{j_1}\widetilde{w}_{j_1j_2}\widetilde{w}_{j_2j_3}\cdots\widetilde{w}_{j_lj_{l+1}}\widetilde{y}_{j_{l+1}} - \mathbb{E}\widetilde{x}_{j_1}\widetilde{w}_{j_1j_2}\widetilde{w}_{j_2j_3}\cdots\widetilde{w}_{j_lj_{l+1}}\widetilde{y}_{j_{l+1}}) \Big) \\
&\le \sum_{\substack{\mathbf{i},\mathbf{j},\mathcal{G}_\mathbf{i} \text{ and } \mathcal{G}_\mathbf{j} \text{ are connected},\\ i_s\ne i_{s+1},\, j_s\ne j_{s+1},\, 1\le s\le l,}} \mathbb{E}|\widetilde{x}_{i_1}\widetilde{w}_{i_1i_2}\widetilde{w}_{i_2i_3}\cdots\widetilde{w}_{i_li_{l+1}}\widetilde{y}_{i_{l+1}}\widetilde{x}_{j_1}\widetilde{w}_{j_1j_2}\widetilde{w}_{j_2j_3}\cdots\widetilde{w}_{j_lj_{l+1}}\widetilde{y}_{j_{l+1}}| \\
&\quad + \sum_{\substack{\mathbf{i},\mathbf{j},\mathcal{G}_\mathbf{i} \text{ and } \mathcal{G}_\mathbf{j} \text{ are connected},\\ i_s\ne i_{s+1},\, j_s\ne j_{s+1},\, 1\le s\le l,}} \mathbb{E}|\widetilde{x}_{i_1}\widetilde{w}_{i_1i_2}\widetilde{w}_{i_2i_3}\cdots\widetilde{w}_{i_li_{l+1}}\widetilde{y}_{i_{l+1}}|\mathbb{E}|\widetilde{x}_{j_1}\widetilde{w}_{j_1j_2}\widetilde{w}_{j_2j_3}\cdots\widetilde{w}_{j_lj_{l+1}}\widetilde{y}_{j_{l+1}}|.
\end{aligned} \tag{S.90}
$$

Notice that each expectation in the last two lines of (S.90) involves the product of independent random variables and the dependency of $\widetilde{w}_{i_1i_2}\widetilde{w}_{i_2i_3}\cdots\widetilde{w}_{i_li_{l+1}}$ and $\widetilde{w}_{j_1j_2}\widetilde{w}_{j_2j_3}\cdots\widetilde{w}_{j_lj_{l+1}}$ are from some shared factors, say $\widetilde{w}_{ab}^{m_1}$ and $\widetilde{w}_{ab}^{m_2}$ respectively, $m_1, m_2 \ge 1$. By Holder's inequality that

$$
\mathbb{E}|\widetilde{w}_{ab}|^{m_1}\mathbb{E}|\widetilde{w}_{ab}|^{m_2} \le \mathbb{E}|\widetilde{w}_{ab}|^{m_1+m_2},
$$

we have

$$
\text{(S.90)} \le 2 \sum_{\substack{\mathbf{i},\mathbf{j},\mathcal{G}_\mathbf{i} \text{ and } \mathcal{G}_\mathbf{j} \text{ are connected},\\ i_s\ne i_{s+1},\, j_s\ne j_{s+1},\, 1\le s\le l,}} \mathbb{E}|\widetilde{x}_{i_1}\widetilde{w}_{i_1i_2}\widetilde{w}_{i_2i_3}\cdots\widetilde{w}_{i_li_{l+1}}\widetilde{y}_{i_{l+1}}\widetilde{x}_{j_1}\widetilde{w}_{j_1j_2}\widetilde{w}_{j_2j_3}\cdots\widetilde{w}_{j_lj_{l+1}}\widetilde{y}_{j_{l+1}}|. \tag{S.91}
$$

By (S.91), to prove (S.81), it suffices to calculate the upper bound of the expectations at the right hand side of (S.91). By the independency of $\widetilde{w}_{ij}$, the upper bound of

$$
\mathbb{E}|\widetilde{x}_{i_1}\widetilde{w}_{i_1i_2}\widetilde{w}_{i_2i_3}\cdots\widetilde{w}_{i_li_{l+1}}\widetilde{y}_{i_{l+1}}\widetilde{x}_{j_1}\widetilde{w}_{j_1j_2}\widetilde{w}_{j_2j_3}\cdots\widetilde{w}_{j_lj_{l+1}}\widetilde{y}_{j_{l+1}}|
$$

is controlled by the r-th moments of $\widetilde{w}_{ij}$ with (S.84), $r = 1, \ldots, 2l$. The topology of $\mathcal{G}_{\mathbf{i}}$ and $\mathcal{G}_{\mathbf{j}}$ are the same as Lemma 4 of Fan et al. (2020+), the summation at the right hand side of (S.91) can be controlled by exactly the same steps as in the proof of Lemma 4 in Fan et al. (2020+). Hence (S.81) can be proved following the proof of Lemma 4 in Fan et al. (2020+). The proof of (S.82) is similar to that of Lemma 5 in Fan et al. (2020+) by the same modification. □

The next Lemma follows directly from Theorem 2.1 in Bloemendal et al. (2014).

Lemma 4. *Under Assumption 1, for any constant $c > 1$, we have for any ϵ, $D > 0$, there exists an integer $n_0(\epsilon, D)$ depending on ϵ and D, such that for all $n \geq n_0(\epsilon, D)$, it holds*

$$\mathbb{P}\left(\|\mathbf{W}\| \geq c \max\{\|\Sigma\|, 1\}(n^{\frac{1}{2}} + p^{\frac{1}{2}})\right) \leq n^{-D}.$$

Lemma 5. *Suppose that $c_{12} = 0$. If $n_1 c_{11} \geq n_2 c_{22}$, then we have*

$$d_1^2 = n_1 c_{11}, \ d_2^2 = n_2 c_{22},$$

otherwise

$$d_1^2 = n_2 c_{22}, \ d_2^2 = n_1 c_{11},$$

Proof. We prove this Lemma under the condition $n_1 c_{11} \geq n_2 c_{22}$. Recall the definition of $\mathbf{H}$ in (2), if $c_{12} = 0$, we have

$$\mathbf{H} = \mathbf{a}_1 \mathbf{a}_1^\top c_{11} + \mathbf{a}_2 \mathbf{a}_2^\top c_{22}.$$

Notice that $\mathbf{a}_1^\top \mathbf{a}_2 = 0$, $\|\mathbf{a}_1\|_2^2 = n_1$ and $\|\mathbf{a}_2\|_2^2 = n_2$, we imply that $\frac{\mathbf{a}_1}{\|\mathbf{a}_1\|_2}$ and $\frac{\mathbf{a}_2}{\|\mathbf{a}_2\|_2}$ are the two eigenvectors of $\mathbf{H}$ with corresponding eigenvalues $n_1 c_{11}$ and $n_2 c_{22}$. By the definition of d_1 and d_2 in (S.2) and the condition that $n_1 c_{11} \geq n_2 c_{22}$, we have

$$d_1^2 = n_1 c_{11}, \ d_2^2 = n_2 c_{22}.$$

□

Lemma 6. *Let $\mathbf{A}$ be a $p \times n$ matrix. Denote $\mathcal{A} = \begin{pmatrix} 0 & \mathbf{A}^\top \\ \mathbf{A} & 0 \end{pmatrix}$. If λ^2 is a non-zero eigenvalue of $\mathbf{A}^\top \mathbf{A}$, then $\pm\lambda$ ($\lambda > 0$) are the eigenvalues of $\mathcal{A}$. Moreover, assume that $\mathbf{a}$ and $\mathbf{b}$ are the unit eigenvectors of $\mathbf{A}^\top \mathbf{A}$ and $\mathbf{A}\mathbf{A}^\top$ respectively corresponding to λ^2, then*

$$\mathcal{A}\begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} = \lambda \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix}, \ \mathcal{A}\begin{pmatrix} \mathbf{a} \\ -\mathbf{b} \end{pmatrix} = -\lambda \begin{pmatrix} \mathbf{a} \\ -\mathbf{b} \end{pmatrix}. \tag{S.92}$$

Proof. By the definition of eigenvalue, any eigenvalue of $\mathcal{A}$ (denoted by x) satisfy the following formula

$$\det(\mathcal{A} - x\mathbf{I}) = \det\left(\begin{pmatrix} -x\mathbf{I} & \mathbf{A}^\top \\ \mathbf{A} & -x\mathbf{I} \end{pmatrix}\right) = 0. \tag{S.93}$$

If $x \neq 0$, then (S.93) is equivalent to

$$\det(\mathbf{A}^\top \mathbf{A} - x^2 \mathbf{I}) = 0.$$

Therefore the first conclusion that $\pm\lambda$ are the eigenvalues of $\mathcal{A}$. By the definition of $\mathbf{a}$ and $\mathbf{b}$, they are the right singular vector and left singular vector of $\mathbf{A}$ respectively corresponding to singular value λ. Then equations (S.92) follow. □

S.12 Proof of Lemma 1

The high level idea for proving (13) is to show that i) $\det(f(a_n)) > 0$ and $\det(f(b_n)) > 0$, ii) the function $\det(f(z))$ is strictly convex in $[a_n, b_n]$, and iii) there exists some $z \in (a_n, b_n)$ such that $\det(f(z)) \leq 0$. The result in (14) is then proved by carefully analyzing the behavior of the function $\det(f(z))$ around d_1 and d_2.

We prove (13) first. By the definition of $f(z)$ in (11), we have

$$\begin{aligned}
&\det(f(z)) = f_{11}(z)f_{22}(z) - f_{12}(z)f_{21}(z) \\
&= \Big(1 + d_1\Big(\mathcal{R}(\mathbf{v}_1,\mathbf{v}_1,z) - \mathcal{R}(\mathbf{v}_1,\mathbf{V}_-,z)\big(-\mathbf{D}+\mathcal{R}(\mathbf{V}_-,\mathbf{V}_-,z)\big)^{-1}\mathcal{R}(\mathbf{V}_-,\mathbf{v}_1,z)\Big)\Big) \\
&\times\Big(1 + d_2\Big(\mathcal{R}(\mathbf{v}_2,\mathbf{v}_2,z) - \mathcal{R}(\mathbf{v}_2,\mathbf{V}_-,z)\big(-\mathbf{D}+\mathcal{R}(\mathbf{V}_-,\mathbf{V}_-,z)\big)^{-1}\mathcal{R}(\mathbf{V}_-,\mathbf{v}_2,z)\Big)\Big) \\
&-d_1d_2\Big(\mathcal{R}(\mathbf{v}_1,\mathbf{v}_2,z) - \mathcal{R}(\mathbf{v}_1,\mathbf{V}_-,z)\big(-\mathbf{D}+\mathcal{R}(\mathbf{V}_-,\mathbf{V}_-,z)\big)^{-1}\mathcal{R}(\mathbf{V}_-,\mathbf{v}_2,z)\Big)^2 .
\end{aligned} \tag{S.94}$$

By Lemma 3 and the expansion (10), for any $\mathbf{M}_1$ and $\mathbf{M}_2$ with finite columns and spectral norms, we have

$$\big\|\mathcal{R}(\mathbf{M}_1,\mathbf{M}_2,z) + z^{-1}\mathbf{M}_1^\top\mathbf{M}_2\big\| = \| -\sum_{l=2}^{L} z^{-(l+1)}\mathbf{M}_1^\top\mathbb{E}\mathbf{W}^l\mathbf{M}_2\| = O(\sigma_n^2/a_n^3),\ \ z \in [a_n, b_n]\,, \tag{S.95}$$

and

$$\big\|\mathcal{R}'(\mathbf{M}_1,\mathbf{M}_2,z) - z^{-2}\mathbf{M}_1^\top\mathbf{M}_2\big\| = \| \sum_{l=2}^{L}(l+1) z^{-(l+2)}\mathbf{M}_1^\top\mathbb{E}\mathbf{W}^l\mathbf{M}_2\| = O(\sigma_n^2/a_n^4)\,. \tag{S.96}$$

Substituting $z = a_n$ into f, by (S.95), for large enough n we have

$$|\mathcal{R}(\mathbf{v}_1,\mathbf{v}_2,a_n)| = O\left(\frac{\sigma_n^2}{a_n^3}\right) \tag{S.97}$$

$$\|\big(-\mathbf{D}+\mathcal{R}(\mathbf{V}_-,\mathbf{V}_-,z)\big)^{-1}\| = O(b_n)\ z \in [a_n, b_n]\,. \tag{S.98}$$

By (S.97) and (S.98) we have

$$|\mathcal{R}(\mathbf{v}_i,\mathbf{V}_-,z)\big(-\mathbf{D}+\mathcal{R}(\mathbf{V}_-,\mathbf{V}_-,z)\big)^{-1}\mathcal{R}(\mathbf{V}_-,\mathbf{v}_j,z)| = O\left(\frac{\sigma_n^4}{a_n^5}\right),\ 1 \leq i,j \leq 2,\ \ z \in [a_n, b_n]\,. \tag{S.99}$$

By Assumption 1 on Σ, there exists a constant c such that $\Sigma \geq c\mathbf{I}$, therefore we have

$$\sigma_n^2 \geq \max\{\mathbf{v}_1^\top\mathbb{E}\mathbf{W}^2\mathbf{v}_1, \mathbf{v}_2^\top\mathbb{E}\mathbf{W}^2\mathbf{v}_2\} \geq \min\{\mathbf{v}_1^\top\mathbb{E}\mathbf{W}^2\mathbf{v}_1, \mathbf{v}_2^\top\mathbb{E}\mathbf{W}^2\mathbf{v}_2\} \geq c\sigma_n^2. \tag{S.100}$$

By (S.100) and Lemma 3, for large enough n we have

$$\begin{aligned}
1 + d_1\mathcal{R}(\mathbf{v}_1,\mathbf{v}_1,a_n) &= 1 - \frac{d_1}{a_n} - \sum_{i\geq 2}^{L}\frac{d_1\mathbf{v}_1^\top\mathbb{E}\mathbf{W}^i\mathbf{v}_1}{a_n^{i+1}} \\
&= 1 - \frac{d_1}{a_n} - \frac{d_1\mathbf{v}_1^\top\mathbb{E}\mathbf{W}^2\mathbf{v}_1}{a_n^3} + O(\frac{\sigma_n^3}{a_n^4}) \leq \frac{a_n - d_1}{2a_n} - \frac{c\sigma_n^2}{2a_n^2}\,,
\end{aligned}$$

and

$$1 + d_2\mathcal{R}(\mathbf{v}_2,\mathbf{v}_2,a_n) \leq \frac{a_n - d_2}{2a_n} - \frac{c\sigma_n^2}{2a_n^2}\,. \tag{S.101}$$

Substituting (S.97)–(S.101) into (S.94), we have

$$\det(f(a_n)) > 0\,. \tag{S.102}$$

Similar to the proof from (S.94) to (S.102), we imply that

$$\det(f(b_n)) > 0\,. \tag{S.103}$$

Moreover, by (S.94) and Lemma 3, we imply that

$$\Big(\det(f(z))\Big)'' = -\frac{2d_1}{z^3} - \frac{2d_2}{z^3} + \frac{6d_1d_2}{z^4} + o\left(\frac{d_1d_2}{a_n^4}\right) > 0,\ z \in [a_n, b_n]\,. \tag{S.104}$$

Therefore $\det(f(z))$ is a strictly convex function and has at most two solutions to the equation $\det(f(z)) = 0$, $z \in [a_n, b_n]$. By (S.95) and (S.96), we have

$$\begin{aligned}\frac{f_{11}'(z)}{d_1} &= \mathcal{R}'(\mathbf{v}_1, \mathbf{v}_1, z) - 2\mathcal{R}'(\mathbf{v}_1, \mathbf{V}_-, z)\big(-\mathbf{D} + \mathcal{R}(\mathbf{V}_-, \mathbf{V}_-, z)\big)^{-1}\mathcal{R}(\mathbf{V}_-, \mathbf{v}_1, z) \\ &- \mathcal{R}(\mathbf{v}_1, \mathbf{V}_-, z)\left(\big(-\mathbf{D} + \mathcal{R}(\mathbf{V}_-, \mathbf{V}_-, z)\big)^{-1}\right)'\mathcal{R}(\mathbf{V}_-, \mathbf{v}_1, z) > 0, z \in [a_n, b_n]\,.\end{aligned} \tag{S.105}$$

Therefore $f_{11}(z)$ is a monotonic function in $[a_n, b_n]$. Moreover, by the definitions of a_n, b_n, σ_n and Lemma 3, we have

$$f_{11}(a_n) < 0,\ f_{11}(b_n) > 0.$$

Hence we conclude that there is a unique point $\tilde{t}_1 \in [a_n, b_n]$ such that

$$f_{11}(\tilde{t}_1) = 0.$$

By similar arguments and

$$\begin{aligned}\frac{f_{22}'(z)}{d_2} &= \mathcal{R}'(\mathbf{v}_2, \mathbf{v}_2, z) - 2\mathcal{R}'(\mathbf{v}_2, \mathbf{V}_-, z)\big(-\mathbf{D} + \mathcal{R}(\mathbf{V}_-, \mathbf{V}_-, z)\big)^{-1}\mathcal{R}(\mathbf{V}_-, \mathbf{v}_2, z) \\ &- \mathcal{R}(\mathbf{v}_2, \mathbf{V}_-, z)\left(\big(-\mathbf{D} + \mathcal{R}(\mathbf{V}_-, \mathbf{V}_-, z)\big)^{-1}\right)'\mathcal{R}(\mathbf{V}_-, \mathbf{v}_2, z) > 0, z \in [a_n, b_n]\,,\end{aligned} \tag{S.106}$$

there exists $\tilde{t}_2 \in [a_n, b_n]$ such that

$$f_{22}(\tilde{t}_2) = 0.$$

Without loss of generality, we assume that

$$\tilde{t}_1 \geq \tilde{t}_2\,. \tag{S.107}$$

It follows from (S.94) that

$$\det(f(\tilde{t}_1)) \leq 0 \text{ and } \det(f(\tilde{t}_2)) \leq 0\,. \tag{S.108}$$

Therefore the existence of t_1 and t_2 are ensured by (S.102), (S.103), (S.108) and the convexity of $\det(f(z))$, $z \in [a_n, b_n]$ (t_1 is allowed to be equal to t_2). Furthermore, by the definition of t_1, t_2 and (S.107) we have

$$b_n \geq t_1 \geq \tilde{t}_1 \geq \tilde{t}_2 \geq t_2 \geq a_n\,. \tag{S.109}$$

Hence we complete the proof of (13) and now we turn to (14). Calculating the first derivative of f_{ii}, by Lemma 3, (S.105) and (S.106) we have

$$f_{ii}'(z) = \frac{d_i}{z^2} + O\left(\frac{\sigma_n^2}{d_i^2}\right) \sim \frac{1}{d_i}\,,\ z \in [a_n, b_n]\,, i = 1, 2\,. \tag{S.110}$$

Let $s_i = d_i + \frac{\mathbb{E}\mathbf{v}_1^\top \mathbf{W}^2 \mathbf{v}_1}{d_i}$, for f_{11}, by Lemma 3 we have

$$f_{11}(s_1) = 1 - d_1\left(\frac{1}{s_1} + \frac{\mathbf{v}_1^\top \mathbb{E}\mathbf{W}^2 \mathbf{v}_1}{s_1^3}\right) + O\left(\frac{\sigma_n^3}{d_1^3}\right) = O\left(\frac{\sigma_n^3}{d_1^3}\right)\,.$$

Combining this with (S.110), we imply that

$$\tilde{t}_1 = d_1 + \frac{\mathbf{v}_1^\top \mathbb{E}\mathbf{W}^2 \mathbf{v}_1}{d_1} + O\left(\frac{\sigma_n^3}{d_1^2}\right). \tag{S.111}$$

Similarly, we also have

$$\tilde{t}_2 = d_2 + \frac{\mathbf{v}_2^\top \mathbb{E}\mathbf{W}^2 \mathbf{v}_2}{d_2} + O\left(\frac{\sigma_n^3}{d_2^2}\right). \tag{S.112}$$

Finally, by Lemma 3 and (S.94), similar to the arguments of (S.102) and (S.103), we have

$$\det\left(f\left(d_1 + \frac{2\mathbf{v}_1^\top \mathbb{E}\mathbf{W}^2 \mathbf{v}_1}{d_1} + \frac{2\mathbf{v}_2^\top \mathbb{E}\mathbf{W}^2 \mathbf{v}_2}{d_2}\right)\right) > 0\,, \tag{S.113}$$

and

$$\det\left(f\left(d_2 - \frac{2\mathbf{v}_1^\top \mathbb{E}\mathbf{W}^2 \mathbf{v}_1}{d_1} - \frac{2\mathbf{v}_2^\top \mathbb{E}\mathbf{W}^2 \mathbf{v}_2}{d_2}\right)\right) > 0\,. \tag{S.114}$$

By (S.113) and (S.114) and the convexity of $\det(f(z))$, we have

$$d_2 - \frac{2\mathbf{v}_1^\top \mathbb{E}\mathbf{W}^2 \mathbf{v}_1}{d_1} - \frac{2\mathbf{v}_2^\top \mathbb{E}\mathbf{W}^2 \mathbf{v}_2}{d_2} \le t_2 \le t_1 \le d_1 + \frac{2\mathbf{v}_1^\top \mathbb{E}\mathbf{W}^2 \mathbf{v}_1}{d_1} + \frac{2\mathbf{v}_2^\top \mathbb{E}\mathbf{W}^2 \mathbf{v}_2}{d_2}$$

Combining this with (S.109), (S.111) and (S.112), we imply that

$$t_k - d_k = O\left(\frac{\sigma_n^2}{d_k}\right),\ k = 1, 2\,, \tag{S.115}$$

which implies Lemma 1 by (S.22).

S. DISCUSSION

In this section, we discuss two directions to generalize our model. One is to allow non-gaussian distribution random vectors, and the other is to discuss the clustering boundary of our model under some additional restrictions in the last two sections.

S.1 Non-Gaussian distribution

Checking the proof of our main theorem carefully, we can see that the key tool is Lemma 3. As long as Lemma 3 holds, then all of our theorems holds. Hence for non-gaussian distribution Z, it suffices to show Lemma 3 holds for non-gaussian distribution. The proof is expected to be more complicated than Lemmas 4 and 5 in Fan et al. (2020+) and is worthy for further investigation.

S.2 Clustering lower bound

In this section, we investigate the clustering lower bound for our model when $p \sim n$. In addition, we impose Prior distribution on Y_i – assume that $\{Y_i\}$ are i.i.d., $Y_i \sim \text{Bernoulli}(1/2)$, $i = 1, \ldots, n$. In addition, assume $\boldsymbol{\mu}_1 = -\boldsymbol{\mu}_2$. Let $l_i = 2Y_i - 1 \in \{-1, 1\}$ and $\widehat{l}_i$ be the estimator of l_i by some clustering algorithm. Similar to Jin et al. (2017), we introduce the Hamming distance to measure the performance of clustering:

$$\text{Hamm}_n = \frac{1}{n} \inf_{s \in \{-1,1\}} \left\{\sum_{i=1}^n \mathbb{P}(\widehat{l}_i \neq s l_i)\right\}. \tag{S.116}$$

The following theorem provides the clustering lower bound, below which clustering is impossible, regardless of what clustering method to use.

Theorem 5. *If $\boldsymbol{\mu}_1^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1 \to 0$, then for any clustering approach, we have*

$$\liminf_{n\to\infty} \text{Hamm}_n \ge \frac{1}{2}. \tag{S.117}$$

Proof. The main idea of this proof largely follows from Theorem 1.1 of Jin et al. (2017). Notice that under the conditions of this Theorem, the model (1) becomes

$$\mathbf{x}_i = l_i\boldsymbol{\mu}_1 + \mathbf{w}_i,\ i = 1, \ldots, n\,. \tag{S.118}$$

For any $1 \le i \le n$, we consider the testing problem that

$$H_{-1} : l_i = -1 \text{ vs } H_1 : l_i = 1\,.$$

Let $f_{\pm}^{(i)}$ be the joint density of $\mathbf{X}$ under $H_{\pm}$ respectively. By the property of total variation, it can be derived that

$$1 - \|f_1 - f_{-1}\|_{TV} \le \mathbb{P}(\widehat{l}_i \ne l_i | l_i = 1) + \mathbb{P}(\widehat{l}_i \ne l_i | l_i = -1)\,.$$

By the assumption that $Y_i \sim \text{Bernoulli}(1/2)$ and $\|f_1 - f_{-1}\|_{TV} = 1/2\|f_1 - f_{-1}\|_1$, we have

$$1/2 - \frac{1}{4}\|f_1^{(i)} - f_{-1}^{(i)}\|_1 \le \mathbb{P}(\widehat{l}_i \ne l_i)\,.$$

Therefore, in order to prove this theorem, it suffices to show that uniformly for all $1 \le i \le n$, we have

$$\|f_1^{(i)} - f_{-1}^{(i)}\|_1 \to 0\,.$$

Let $\mathbf{l} = (l_1, \ldots, l_n)^\top - l_i\mathbf{e}_i$. Then we have

$$\begin{aligned}\|f_1^{(i)} - f_{-1}^{(i)}\|_1 &= \mathbb{E}\Big|\int \sinh(\mathbf{x}_i^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1) e^{-\frac{\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}_1\|_2^2}{2}} e^{\mathbf{l}^\top\mathbf{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1 - (n-1)\frac{\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}_1\|_2^2}{2}} dF(\mathbf{l})\Big| \\ &\le \int \mathbb{E}\Big|\sinh(\mathbf{x}_i^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1) e^{-\frac{\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}_1\|_2^2}{2}} e^{\mathbf{l}^\top\mathbf{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1 - (n-1)\frac{\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}_1\|_2^2}{2}}\Big| dF(\mathbf{l})\,,\end{aligned} \tag{S.119}$$

where $\mathbb{E}$ is the expectation under the distribution of $\mathbf{X} = \mathbf{W}$. Therefore, it suffices for us to show that

$$\mathbb{E}\Big|\sinh(\mathbf{x}_i^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1) e^{-\frac{\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}_1\|_2^2}{2}} e^{\mathbf{l}^\top\mathbf{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1 - (n-1)\frac{\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}_1\|_2^2}{2}}\Big| \to 0\,. \tag{S.120}$$

Notice that $\mathbf{x}_i$ is independent of $\mathbf{l}^\top\mathbf{X}^\top$, we have

$$\begin{aligned}&\mathbb{E}\Big|\sinh(\mathbf{x}_i^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1) e^{-\frac{\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}_1\|_2^2}{2}} e^{\mathbf{l}^\top\mathbf{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1 - (n-1)\|\boldsymbol{\mu}_1\|_2^2/2}\Big| \\ =&\mathbb{E}\Big|\sinh(\mathbf{x}_i^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1) e^{-\frac{\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}_1\|_2^2}{2}}\Big| \mathbb{E}[e^{-\frac{\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}_1\|_2^2}{2}} e^{\mathbf{l}^\top\mathbf{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1 - (n-1)\frac{\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}_1\|_2^2}{2}}] \\ =&e^{-\frac{\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}_1\|_2^2}{2}}\mathbb{E}\Big|\sinh(\mathbf{x}_i^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1)\Big|\,.\end{aligned} \tag{S.121}$$

By the distribution of l_i we know that (S.121) is independent of i. Now we focus on $\mathbb{E}\Big|\sinh(\mathbf{x}_i^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1)\Big|$. Since the expectation is under the distribution that $\mathbf{x}_i = \mathbf{w}_i$, $\mathbf{x}_i^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1 \sim N(0, \|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}_1\|_2^2)$. For simplicity, let $z = \mathbf{x}_i^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1/\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}_1\|_2$ and $\sigma = \|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}_1\|_2$. Then

$$2\mathbb{E}\Big|\sinh(\mathbf{x}_i^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1)\Big| = 2\mathbb{E}\Big|\sinh(\sigma z)\Big| = 2\int_{z\ge 0}\frac{e^{\sigma z} - e^{-\sigma z}}{\sqrt{2\pi}} e^{-z^2/2} dz \tag{S.122}$$

$$\int_{z\ge 0}\frac{e^{\sigma z}}{\sqrt{2\pi}} e^{-z^2/2} dz = \frac{e^{\sigma^2/2}}{\sqrt{2\pi}}\int_{z\ge 0} e^{-(z-\sigma)^2/2} dz = e^{\sigma^2/2}\mathbb{P}(z \ge -\sigma)\,. \tag{S.123}$$

$$\int_{z\ge 0}\frac{e^{-\sigma z}}{\sqrt{2\pi}} e^{-z^2/2} dz = \frac{e^{\sigma^2/2}}{\sqrt{2\pi}}\int_{z\ge 0} e^{-(z+\sigma)^2/2} dz = e^{\sigma^2/2}\mathbb{P}(z \ge \sigma)\,. \tag{S.124}$$

By (S.123) and (S.124), we imply that

$$\mathbb{E}\Big|\sinh(\mathbf{x}_i^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1)\Big| = e^{\sigma^2/2}(\mathbb{P}(z \ge -\sigma) - \mathbb{P}(z \ge \sigma)) = e^{\sigma^2/2}(\mathbb{P}(-\sigma \le z < \sigma)\,. \tag{S.125}$$

By (S.121), (S.125) and the condition that $\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}_1\|_2 \to 0$, we finish our proof. we finish our proof. ☐

S.3 Exact recovery

In this section, we consider a special case that $\boldsymbol{\mu}_1 = -\boldsymbol{\mu}_2$. By Theorem 1, it is corresponding to the case that $d_2 = 0$ and $d_1^2 = n_1 c_{11} + n_2 c_{22} = nc_{11}$. We prove that for a little bigger $\|\boldsymbol{\mu}_1\|$, we have the following theorem and Corollary 1 for exact recovery.

Theorem 6. *Assume that $\Sigma = \mathbf{I}$, $\boldsymbol{\mu}_1 = -\boldsymbol{\mu}_2$, $\|\boldsymbol{\mu}_1\|_\infty = O(\frac{1}{n^{1/4}})$, $n = O(n_1) = O(n_2)$ and $p \sim n$, if there exists a positive constant ϵ such that $c_{11} \geq 2(1+\epsilon)\log n$, then there exists $s \in \{\pm 1\}$ such that with probability tending to 1, we have*

$$\sqrt{n} \min_{1\leq i\leq n} \{s l_i \widehat{\mathbf{u}}_1(i)\} \geq 1 - \frac{1}{\sqrt{1+\epsilon}} - \frac{C}{\sqrt{\log n}}\,, \tag{S.126}$$

for some positive constant $\mathbf{C}$.

Proof. We prove this theorem by considering the linearization matrix $\mathcal{Z}$ and $\widehat{\mathbf{v}}_1$. The idea of the proof follows from the proof of Theorem 3.1 of Abbe et al. (2020+). Concretely, we prove that **A1**–**A4** of Abbe et al. (2020+) hold and apply Theorem 1.1 of Abbe et al. (2020+) to show our result. Substituting $d_1^2 = nc_{11}$ and $c_{11} = c_{22} = -c_{12}$ into (S.8) and (S.9), without loss of generality, assume $\mathbf{u}_1$ has two different values v_1 and v_2 such that

$$v_1 = -v_2 = \frac{1}{\sqrt{n}}\,,$$

where v_1 is corresponding to $Y_i = 1$ and v_2 is corresponding to $Y_i = 0$. Then we have

$$l_i \mathbf{u}_1(i) = \frac{1}{\sqrt{n}},\ \ i = 1, \ldots, n\,. \tag{S.127}$$

By Lemma 4, for any positive constant $c > 1$, D and sufficiently large n we have

$$\mathbb{P}\left(\|\mathbf{W}\| \geq c(\sqrt{n} + \sqrt{p})\right) \leq n^{-D}\,.$$

Setting $\gamma = \max\{\frac{\|\boldsymbol{\mu}_1\|_\infty}{\sqrt{\log n}}, \frac{1}{\sqrt{n}}\} \to 0$, we have

$$\max\{\sqrt{c_{11}}, \|\boldsymbol{\mu}_1\|_\infty \sqrt{n}\} \leq \gamma d_1\,. \tag{S.128}$$

Notice that $\mathcal{Z}$ and $\mathbb{E}\mathcal{Z}$ are corresponding to $\mathbf{A}$ and $\mathbf{A}^*$ of Abbe et al. (2020+). Let $\Delta^* = d_1$, by (S.128), **A1** of Abbe et al. (2020+) holds. Moreover, **A2** follows from the assumption that $\Sigma = \mathbf{I}$. By Lemma 4, it is easy to see that **A3** of Abbe et al. (2020+) holds by (S.127). Similar to the proof of Theorem 3.1 in Abbe et al. (2020+), **A4** holds by setting $\phi(x) = x$. By Theorem 1.1 of Abbe et al. (2020+), with probability tending to 1, there exists a positive constant C such that

$$\min_{s\in\{\pm 1\}} \|s\widehat{\mathbf{v}}_1 - \frac{\mathcal{Z}\mathbf{v}_1}{d_1}\|_\infty = \min_{s\in\{\pm 1\}} \|s\widehat{\mathbf{v}}_1 - \mathbf{v}_1 - \frac{(\mathcal{Z} - \mathbb{E}\mathcal{Z})\mathbf{v}_1}{d_1}\|_\infty \leq C\gamma\|\mathbf{v}_1\|_\infty\,, \tag{S.129}$$

where $\mathbf{v}_1$ is the eigenvector of $\mathbb{E}\mathcal{Z}$ corresponding to d_1. By Lemma 6, we have $\mathbf{v}_1 = \frac{1}{\sqrt{2}}(\mathbf{u}_1^\top, \frac{\boldsymbol{\mu}_1^\top}{c_{11}})^\top$. Therefore by the conditions that $\|\boldsymbol{\mu}_1\|_\infty = O(\frac{1}{n^{1/4}})$ and $n = O(n_1) = O(n_2) = O(p)$, we have

$$\gamma\|\mathbf{v}_1\|_\infty = O(\frac{1}{\sqrt{n\log n}})\,. \tag{S.130}$$

Notice that each entry of $\sqrt{2}(\mathcal{Z} - \mathbb{E}\mathcal{Z})\mathbf{v}_1$ follows a standard gaussian distribution. This implies that

$$\mathbb{P}(\max_{1\leq i\leq n} |\mathbf{e}_i^\top(\mathcal{Z} - \mathbb{E}\mathcal{Z})\mathbf{v}_1| \geq \sqrt{\log n}) = O(\frac{1}{\sqrt{\log n}})\,. \tag{S.131}$$

By (S.129)–(S.131), with probability tending to 1, there exists $s \in \{\pm 1\}$ and some positive constant C such that

$$\sqrt{n} \max_{1\le i\le n} \{\|s\widehat{\mathbf{v}}_1(i) - \mathbf{v}_1(i)\|_\infty\} \le \frac{C}{\sqrt{2n\log n}} + \frac{\sqrt{\log n}}{\sqrt{2(1+\epsilon)n\log n}}\,. \tag{S.132}$$

Notice that $\mathbf{v}_1 = \frac{1}{\sqrt{2}}(\mathbf{u}_1^\top, \frac{\boldsymbol{\mu}_1^\top}{c_{11}})^\top$ and the first n entries of $\widehat{\mathbf{v}}_1$ is $\frac{1}{\sqrt{2}}\widehat{\mathbf{u}}_1$, by (S.127) and (S.132), we have

$$\sqrt{n} \min_{1\le i\le n} \{s l_i \widehat{\mathbf{u}}_1(i)\} \ge 1 - \frac{1}{\sqrt{1+\epsilon}} - \frac{C}{\sqrt{\log n}}\,. \tag{S.133}$$

□

By Theorem 6, we have the following corollary to ensure the existence of exact recovery for the model.

Corollary 1. *Under the conditions of Theorem 6, there exists one clustering approach such that*

$$\mathbb{P}(\widehat{Y}_i = Y_i, i = 1\ldots,n) = 1 - o(1)\,. \tag{S.134}$$

Proof. The following clustering procedure suffices.

1. Calculate the eigenvector of $\mathcal{Z}$ corresponding to the largest eigenvalue, which is $\widehat{\mathbf{v}}_1$ as we defined before.

2. $\widehat{Z}_i = \frac{sgn(\widehat{\mathbf{v}}_1(i))+1}{2}, i = 1,\ldots,n.$

If $\sum_{i=1}^n (2\widehat{Z}_i - 1)l_i > 0$, we let $\widehat{Y}_i = \widehat{Z}_i$, otherwise $\widehat{Y}_i = -(\widehat{Z}_i - 1)$. Without loss of generality, we assume that $\sum_{i=1}^n (2\widehat{Z}_i - 1)l_i > 0$ and therefore $\widehat{Y}_i = \widehat{Z}_i$. By the definition of $\widehat{Z}_i$ and the condition that $\sum_{i=1}^n (2\widehat{Z}_i - 1)l_i > 0$, Theorem 6 holds for $s = 1$. Hence

$$\mathbb{P}(\widehat{Y}_i = Y_i, i = 1\ldots,n | \sum_{i=1}^n (2\widehat{Z}_i - 1)l_i > 0) = 1 - o(1). \tag{S.135}$$

By almost the same arguments, we can prove similarly that

$$\mathbb{P}(\widehat{Y}_i = Y_i, i = 1\ldots,n | \sum_{i=1}^n (2\widehat{Z}_i - 1)l_i \le 0) = 1 - o(1). \tag{S.136}$$

Therefore, (S.134) follows from (S.135) and (S.136). □

REFERENCES

Emmanuel Abbe, Jianqing Fan, Kaizheng Wang, and Yiqiao Zhong. Entrywise eigenvector analysis of random matrices with low expected rank. *The Annals of Statistics*, page In print, 2020+.

Alex Bloemendal, László Erdos, Antti Knowles, Horng-Tzer Yau, and Jun Yin. Isotropic local laws for sample covariance and generalized wigner matrices. *Electronic Journal of Probability*, 19(33):1–53, 2014.

T. Tony Cai, Zongming Ma, and Yihong Wu. Sparse pca: Optimal rates and adaptive estimation. *The Annals of Statistics*, 41(6):3074–3110, 2013. doi: 10.1214/13-AOS1178. URL https://doi.org/10.1214/13-AOS1178.

Jianqing Fan, Jianhua Guo, and Shurong Zheng. Estimating number of factors by adjusted eigenvalues thresholding. *Journal of the American Statistical Association*, 0(0):1–10, 2020.

Jianqing Fan, Yingying Fan, Xiao Han, and Jinchi Lv. Asymptotic theory of eigenvectors for random matrices with diverging spikes. *Journal of the American Statistical Association*, 2020+. doi: 10.1080/01621459.2020.1840990.

Jiashun Jin, Wanjie Wang, et al. Influential features pca for high dimensional clustering. *The Annals of Statistics*, 44(6):2323–2359, 2016.

Jiashun Jin, Zheng Tracy Ke, and Wanjie Wang. Phase transitions for high dimensional clustering and related problems. *Annals of Statistics*, 45(5):2151–2189, 10 2017.