Supplement to "Optimal Permutation Recovery in Permuted Monotone Matrix Model"

Rong Ma[1], T. Tony Cai[2] and Hongzhe Li[1]
Department of Biostatistics, Epidemiology and Informatics[1]
Department of Statistics[2]
University of Pennsylvania
Philadelphia, PA 19104

Abstract

In this Supplementary Material, we prove Theorem 3 & 4 and Proposition 1 to 4 in the main paper and the technical lemmas. The numerical comparisons of $\hat{\pi}$ and an SVD-based estimator $\tilde{\pi}$ are also included as an Appendix.

1 Proofs of Theorem 3 and 4

Proof of Theorem 3. As in the proof of Theorem 1, it suffices to bound the probability $P\big(\sum_{k=1}^{n}\hat{w}_k(X_{ki}-X_{k,i+1})\ge 0\big)$. Note that $\hat{w}$ is the first eigenvector of

$$A = X\left(I-\frac{1}{p}ee^\top\right)\left(I-\frac{1}{p}ee^\top\right)X^\top = TT^\top$$

where $T\in\mathbb{R}^{n\times p}$ and T admits the decomposition

$$T = \Theta\left(I-\frac{1}{p}ee^\top\right) + Z\left(I-\frac{1}{p}ee^\top\right) = \Theta' + E \in \mathbb{R}^{n\times p}$$

where $E_{ij}\sim N(0,(p-1)\sigma^2/p)$ and Θ' has SVD

$$\Theta' = U'D'V'^\top.$$

Define $w = u_1'\in\mathbb{R}^n$ as the first eigenvector of $\Theta'\Theta'^\top$. Following the same argument that leads to (13) of the main paper, we have, up to a change of sign for $\hat{w}$,

$$P\left(\sum_{k=1}^{n}\hat{w}_k(X_{ki}-X_{k,i+1})\ge 0\right) \le P\left(w^\top\phi_{i,i+1}\ge -\sqrt{2\delta}\|\phi_{i,i+1}\|_2\right) + P\left(|1-(\hat{w}^\top w)^2|>\delta\right).$$

The following lemmas parallel Lemma 1 and Lemma 2 in the proof of Theorem 1. In particular, Lemma 8 provides a perturbation bound for the leading eigenvector of approximate rank-one matrices, which could be of independent interest.

Lemma 7. *Under the conditions of Theorem 3, let* $\Xi_i = \|\Theta'_{.i} - \Theta'_{.i+1}\|_2$. *Then for any* $\delta > 0$, *we have*

$$P\bigg(w^\top \phi_{i,i+1} \geq -\sqrt{2\delta}\|\phi_{i,i+1}\|\bigg) \tag{1.1}$$

$$\leq \Phi\bigg(C\sqrt{\delta}\bigg[(\sqrt{n}+\sqrt{\log p})^2 + \frac{\Xi_i^2}{\sigma^2} + \sqrt{\log p}\frac{\Xi_i}{\sigma}\bigg]^{1/2} + \frac{w^\top(\Theta'_{.i} - \Theta'_{.i+1})}{\sigma}\bigg) + \frac{C}{p^c} \tag{1.2}$$

for some universal constant $C > 0$.

Lemma 8. *Suppose* $n \lesssim p$ *and* $\lambda_1^2(\Theta') \geq \lambda_2^2(\Theta') + C\sigma^2(n + \sqrt{np})$ *for some* $C > 0$, *it follows that*

$$P\bigg(|1 - (\hat{w}^\top w)^2| \leq \frac{C\sigma^2(\lambda_1^2(\Theta') + \sigma^2 p)(n + \log p)}{(\lambda_1^2(\Theta') - \lambda_2^2(\Theta'))^2}\bigg) \geq 1 - \frac{C}{p^c}.$$

Combining Lemma 7 and Lemma 8, since

$$\lambda_1^2(\Theta') - \lambda_2^2(\Theta') > C_0\sigma^2(n + \sqrt{np}) \quad \text{and let} \quad \delta = \frac{C_0\sigma^2(\lambda_1^2(\Theta') + \sigma^2 p)(n + \log p)}{(\lambda_1^2(\Theta') - \lambda_2^2(\Theta'))^2} \tag{1.3}$$

for some $C_0 > 0$, we have

$$P\bigg(\sum_{k=1}^n \hat{w}_k(X_{ki} - X_{k,i+1}) \geq 0\bigg)$$
$$\leq \Phi\bigg(C\sqrt{\delta}\bigg[(\sqrt{n}+\sqrt{\log p})^2 + \frac{\Xi_i^2}{\sigma^2} + \frac{\Xi_i}{\sigma}\sqrt{\log p}\bigg]^{1/2} + \frac{w^\top(\Theta'_{.i} - \Theta'_{.i+1})}{\sigma}\bigg) + \frac{C}{p^c}$$

for some $C, c > 0$. The rest of the analysis is divided into several cases.

Case 1. $\log p \lesssim n$**.** In this case, under (1.3), we have

$$P\bigg(\sum_{k=1}^n \hat{w}_k(X_{ki} - X_{k,i+1}) \geq 0\bigg)$$
$$\leq \Phi\bigg(C\sqrt{\delta}\bigg[n + \frac{\|\Theta'_{.i} - \Theta'_{.i+1}\|_2^2}{\sigma^2} + \frac{\|\Theta'_{.i} - \Theta'_{.i+1}\|_2}{\sigma}\sqrt{\log p}\bigg]^{1/2} + \frac{w^\top(\Theta'_{.i} - \Theta'_{.i+1})}{\sigma}\bigg) + \frac{C}{p^c}$$

then

1. if $\|\Theta'_{.i} - \Theta'_{.i+1}\|_2 \lesssim \sigma\sqrt{n}$, we have $P\big(\sum_{k=1}^n \hat{w}_k(X_{ki} - X_{k,i+1}) \geq 0\big) \leq \Phi\big(C\sqrt{\delta n} +$

$\frac{w^\top(\Theta'_{.i}-\Theta'_{.i+1})}{\sigma}\big) + \frac{C}{p^c} \le \frac{C'}{p^c}$, where the last inequality follows from

$$\sqrt{\log p} \lesssim \Gamma/\sigma, \qquad \lambda_1^2(\Theta') \gtrsim \lambda_2^2(\Theta') + \frac{\sigma^2 n(n+\log p)}{\Gamma^2/\sigma^2} + \frac{\sigma^2\sqrt{np(n+\log p)}}{\Gamma/\sigma}.$$

2. $\|\Theta'_{.i}-\Theta'_{.i+1}\|_2 \gtrsim \sigma\sqrt{n}$, we have $P\big(\sum_{k=1}^n \hat{w}_k(X_{ki}-X_{k,i+1}) \ge 0\big) \le \Phi\big(C\sqrt{\delta}\frac{\|\Theta'_{.i}-\Theta'_{.i+1}\|_2}{\sigma} + \frac{w^\top(\Theta'_{.i}-\Theta'_{.i+1})}{\sigma}\big) + \frac{C}{p^c} \le \frac{C'}{p^c}$, where the last inequality follows from

$$\Gamma/\sigma \gtrsim \sqrt{n} \gtrsim \sqrt{\log p}, \qquad \lambda_1^2(\Theta') \gtrsim \lambda_2^2(\Theta') + \frac{\sigma^2\Xi^2(n+\log p)}{\Gamma^2} + \frac{\sigma^2\Xi\sqrt{p(n+\log p)}}{\Gamma}.$$

This completes the proof of case 1.

Case 2. $\log p \gtrsim n$**.** In this case, under (1.3), we have

$$\begin{aligned} &P\bigg(\sum_{k=1}^n \hat{w}_k(X_{ki}-X_{k,i+1}) \ge 0\bigg) \\ &\le \Phi\Bigg(C\sqrt{\delta}\bigg[\log p + \frac{\|\Theta'_{.i}-\Theta'_{.i+1}\|_2^2}{\sigma^2} + \frac{\|\Theta'_{.i}-\Theta'_{.i+1}\|_2}{\sigma}\sqrt{\log p}\bigg]^{1/2} + \frac{w^\top(\Theta'_{.i}-\Theta'_{.i+1})}{\sigma}\Bigg) + \frac{C}{p^c} \end{aligned}$$

In addition, since $\|\Theta'_{.i}-\Theta'_{.i+1}\|_2 \gtrsim \sigma\sqrt{\log p}$, we have

$$P\bigg(\sum_{k=1}^n \hat{u}_k(X_{ki}-X_{k,i+1}) \ge 0\bigg) \le \Phi\bigg(\frac{C\sqrt{\delta}}{\sigma}\|\Theta'_{.i}-\Theta'_{.i+1}\|_2 + \frac{w^\top(\Theta'_{.i}-\Theta'_{.i+1})}{\sigma}\bigg) + \frac{C}{p^c} \le \frac{C'}{p^c}$$

where the last inequality follows from

$$\Gamma/\sigma \gtrsim \sqrt{\log p} \gtrsim \sqrt{n}, \qquad \lambda_1^2(\Theta') \gtrsim \lambda_2^2(\Theta') + \frac{\sigma^2\Xi^2(n+\log p)}{\Gamma^2} + \frac{\sigma^2\Xi\sqrt{p(n+\log p)}}{\Gamma}.$$

This completes the proof of case 2. As a result, it follows that, up to a change of sign of $\hat{w}$,

$$P((\mathfrak{r}(\hat{w}^\top X))^{-1} \ne id) \le \sum_{i=1}^{p-1} P\bigg(\sum_{k=1}^n \hat{w}_k(X_{ki}-X_{k,i+1}) \ge 0\bigg) \le \frac{C}{p^c}$$

for some constant $C, c > 0$. □

Proof of Theorem 4. Similar to the proof of Theorem 2, we write

$$\mathbb{E}[\tau_K(\hat{w}^\top Y, \pi)] = \frac{2}{p(p-1)}\sum_{i<j} P\bigg(\sum_{k=1}^n \hat{w}_k(X_{ki}-X_{kj}) \ge 0\bigg)$$

where

$$\sum_{i<j} P\bigg(\sum_{k=1}^n \hat{w}_k(X_{ki}-X_{kj}) \ge 0\bigg) = \sum_{(i,j):j=i+1} P\bigg(\sum_{k=1}^n \hat{w}_k(X_{ki}-X_{kj}) \ge 0\bigg) + \sum_{(i,j):j>i+1} P\bigg(\sum_{k=1}^n \hat{w}_k(X_{ki}-X_{kj}) \ge 0\bigg).$$

In the following, we first show

$$P\bigg(\sum_{k=1}^n \hat{w}_k(X_{ki}-X_{k,i+1}) \ge 0\bigg) \le \frac{ce^{-\Gamma^2/2\sigma^2}}{\Gamma/\sigma+\sqrt{\Gamma^2/\sigma^2+8/\pi}} + \frac{C}{p^c} \tag{1.4}$$

for some constant $c>0$ and therefore

$$\sum_{(i,j):j=i+1} P\bigg(\sum_{k=1}^n \hat{w}_k(X_{ki}-X_{kj}) \ge 0\bigg) \le \frac{cpe^{-\Gamma^2/2\sigma^2}}{\Gamma/\sigma+\sqrt{\Gamma^2/\sigma^2+8/\pi}} + \frac{C}{p^c}$$

Then we show that

$$\sum_{(i,j):j>i+1} P\bigg(\sum_{k=1}^n \hat{w}_k(X_{ki}-X_{kj}) \ge 0\bigg) \le C\frac{p\sigma}{\Gamma}\min\left\{1, e^{-\Gamma^2/2\sigma^2}\log\left(1+\frac{2\sigma^2}{\Gamma^2}\right)\right\} + \frac{C}{p^c} \tag{1.5}$$

and conclude that

$$\mathbb{E}[\tau_K(\hat{w}^\top Y,\pi)] \le \frac{C\sigma}{p\Gamma}\min\left\{1, e^{-\Gamma^2/2\sigma^2}\log\left(1+\frac{2\sigma^2}{\Gamma^2}\right)\right\} + \frac{Ce^{-\Gamma^2/2\sigma^2}}{p(\Gamma/\sigma+\sqrt{8/\pi})} + \frac{C}{p^{c+2}}.$$

The bound $\mathbb{E}[\tau_K(\hat{w}^\top Y,\pi)] \le 1$ is trivial.

Proof of (1.4). Following the same argument as in the proof of Theorem 3, we have for $1 \le i \le p-1$ and $\delta = \frac{\sigma^2(n+\log p)(\lambda_1^2(\Theta')+\sigma^2 p)}{(\lambda_1^2(\Theta')-\lambda_2^2(\Theta'))^2}$,

$$\begin{aligned} P\bigg(\sum_{k=1}^n \hat{w}_k(X_{ki}-X_{k,i+1}) \ge 0\bigg) &\le P\Big(w^\top\phi_{ij} \ge -\sqrt{2\delta}\|\phi_{i,i+1}\|_2\Big) + P\Big(|1-(\hat{w}^\top w)^2| > \delta\Big) \\ &\le P\Big(w^\top\phi_{i,i+1} \ge -\sqrt{2\delta}\|\phi_{i,i+1}\|_2\Big) + \frac{C}{p^c} \end{aligned}$$

where the last inequality follows from Lemma 8. For the first term, by Lemma 7, for $\delta < 1$,

$$P\bigg(\sum_{k=1}^{n} \hat{w}_k(X_{ki} - X_{k,i+1}) \geq 0\bigg) \leq \Phi\bigg(C\sqrt{\delta}\bigg((\sqrt{n} + \sqrt{\log p})^2 + \frac{\|\Theta'_{.i} - \Theta'_{.i+1}\|_2^2}{\sigma^2} + \frac{\|\Theta'_{.i} - \Theta'_{.i+1}\|_2}{\sigma}\sqrt{\log p}\bigg)^{1/2} + \frac{w^\top(\Theta'_{.i} - \Theta'_{.i+1})}{\sigma}\bigg) + \frac{C}{p^c}.$$

The rest of the analysis is divided into several cases.

Case 1. $\log p \leq n$**.** In this case, we have

$$P\bigg(\sum_{k=1}^{n} \hat{w}_k(X_{ki} - X_{k,i+1}) \geq 0\bigg) \leq \Phi\bigg(C\sqrt{\delta}\bigg(n + \frac{\|\Theta'_{.i} - \Theta'_{.i+1}\|_2^2}{\sigma^2} + \frac{\|\Theta'_{.i} - \Theta'_{.i+1}\|_2}{\sigma}\sqrt{\log p}\bigg)^{1/2} + \frac{w^\top(\Theta'_{.i} - \Theta'_{.i+1})}{\sigma}\bigg) + \frac{C}{p^c}.$$

In addition,

1. if $\|\Theta'_{.i} - \Theta'_{.i+1}\|_2 \lesssim \sigma\sqrt{n}$, we have

   $$\begin{aligned} P\bigg(\sum_{k=1}^{n} \hat{w}_k(X_{ki} - X_{k,i+1}) \geq 0\bigg) &\leq \Phi\bigg(C\sqrt{\delta n} + \frac{w^\top(\Theta'_{.i} - \Theta'_{.i+1})}{\sigma}\bigg) + \frac{C}{p^c} \\ &\leq \Phi\bigg(\frac{C\Gamma}{\sigma}\bigg) + \frac{C}{p^c} \end{aligned}$$

   where the last inequality follows from

   $$\lambda_1^2(\Theta') \gtrsim \lambda_2^2(\Theta') + \frac{\sigma^2 n(n + \log p)}{\Gamma^2/\sigma^2} + \frac{\sigma^2\sqrt{np(n + \log p)}}{\Gamma/\sigma}.$$

   Thus, using Formula 7.1.13 from Abramowitz and Stegun (1965),

   $$\Phi(-x) < \frac{2}{t + \sqrt{t^2 + 8/\pi}}\phi(x), \tag{1.6}$$

   for $1 \leq i \leq p - 1$, we have

   $$P\bigg(\sum_{k=1}^{n} \hat{w}_k(X_{ki} - X_{k,i+1}) \geq 0\bigg) \leq \frac{C}{\Gamma/\sigma + \sqrt{\Gamma^2/\sigma^2 + 8/\pi}} \exp(-\Gamma^2/(2\sigma^2)) + \frac{C}{p^c}.$$

2. if $\|\Theta'_{.i} - \Theta'_{.i+1}\|_2 \gtrsim \sigma\sqrt{n}$, we have

$$P\bigg(\sum_{k=1}^{n} \hat{w}_k(X_{ki} - X_{k,i+1}) \geq 0\bigg) \leq \Phi\bigg(\frac{C\sqrt{\delta}}{\sigma}\|\Theta'_{.i} - \Theta'_{.i+1}\|_2 + \frac{w^\top(\Theta'_{.i} - \Theta'_{.i+1})}{\sigma}\bigg) + \frac{C}{p^c}$$
$$\leq \frac{C}{\Gamma/\sigma + \sqrt{\Gamma^2/\sigma^2 + 8/\pi}} \exp(-\Gamma^2/(2\sigma^2)) + \frac{C}{p^c},$$

where the last inequality follows from (1.6) and

$$\lambda_1^2(\Theta') \gtrsim \lambda_2^2(\Theta') + \frac{\sigma^2\Xi^2(n + \log p)}{\Gamma^2} + \frac{\sigma^2\Xi\sqrt{p(n+\log p)}}{\Gamma}.$$

This completes the proof of case 1.

Case 2. $\log p > n$**.** In this case, we have

$$P\bigg(\sum_{k=1}^{n} \hat{w}_k(X_{ki} - X_{k,i+1}) \geq 0\bigg) \leq \Phi\bigg(C\sqrt{\delta}\bigg(\log p + \frac{\|\Theta'_{.i} - \Theta'_{.i+1}\|_2^2}{\sigma^2} + \frac{\|\Theta'_{.i} - \Theta'_{.i+1}\|_2}{\sigma}\sqrt{\log p}\bigg)^{1/2}$$
$$+ \frac{w^\top(\Theta'_{.i} - \Theta'_{.i+1})}{\sigma}\bigg) + \frac{C}{p^c}.$$

In addition,

1. if $\|\Theta'_{.i} - \Theta'_{.i+1}\|_2 \gtrsim \sigma\sqrt{\log p}$, we have

$$P\bigg(\sum_{k=1}^{n} \hat{u}_k(X_{ki} - X_{k,i+1}) \geq 0\bigg) \leq \Phi\bigg(\frac{C\sqrt{\delta}}{\sigma}\|\Theta'_{.i} - \Theta'_{.i+1}\|_2 + \frac{w^\top(\Theta'_{.i} - \Theta'_{.i+1})}{\sigma}\bigg) + \frac{C}{p^c}$$
$$\leq \frac{C}{\Gamma/\sigma + \sqrt{\Gamma^2/\sigma^2 + 8/\pi}} \exp(-\Gamma^2/(2\sigma^2)) + \frac{C}{p^c}.$$

2. if $\|\Theta'_{.i} - \Theta'_{.i+1}\|_2 \lesssim \sigma\sqrt{\log p}$, we have

$$P\bigg(\sum_{k=1}^{n} \hat{u}_k(X_{ki} - X_{k,i+1}) \geq 0\bigg) \leq \Phi\bigg(C\sqrt{\delta\log p} + \frac{w^\top(\Theta'_{.i} - \Theta'_{.i+1})}{\sigma}\bigg) + \frac{C}{p^c}$$
$$\leq \frac{\sqrt{2/\pi}}{\Gamma/\sigma + \sqrt{\Gamma^2/\sigma^2 + 8/\pi}} \exp(-\Gamma^2/(2\sigma^2)) + \frac{C}{p^c}$$

where the last inequality follows from

$$\lambda_1^2(\Theta') \gtrsim \lambda_2^2(\Theta') + \frac{\sigma^2\log p(n+\log p)}{\Gamma^2/\sigma^2} + \frac{\sigma^2\sqrt{p\log p(n+\log p)}}{\Gamma/\sigma}.$$

This completes the proof of Case 2 and in sum we have proven (1.4).

Proof of (1.5). Following the same construction of S_1 and S_2 in the proof of (18) of the main paper, we have

$$\sum_{(i,j):j>i+1} P\bigg(\sum_{k=1}^{n} \hat{w}_k(X_{ki}-X_{kj}) \geq 0\bigg) = \sum_{(i,j)\in S_1} P\bigg(\sum_{k=1}^{n} \hat{w}_k(X_{ki}-X_{kj}) \geq 0\bigg) + \sum_{(i,j)\in S_2} P\bigg(\sum_{k=1}^{n} \hat{w}_k(X_{ki}-X_{kj}) \geq 0\bigg) \tag{1.7}$$

For the first term, for any $(i,j) \in S_1$, we have

$$\begin{aligned} P\bigg(\sum_{k=1}^{n} \hat{w}_k(X_{ki}-X_{kj}) \geq 0\bigg) &\leq P\Big(w^\top \phi_{ij} \geq -\sqrt{2\delta}\|\phi_{ij}\|_2\Big) + P\Big(|1-(\hat{w}^\top w)^2| > \delta\Big) \\ &\leq P\Big(w^\top \phi_{ij} \geq -\sqrt{2\delta}\|\phi_{ij}\|_2\Big) + \frac{C}{p^c} \end{aligned}$$

where the last inequality follows from Lemma 8. To bound $P\Big(w^\top \phi_{ij} \geq -\sqrt{2\delta}\|\phi_{ij}\|_2\Big)$, similar argument as in Theorem 3 implies, for $\delta < 1$,

$$\begin{aligned} P\bigg(\sum_{k=1}^{n} \hat{w}_k(X_{ki}-X_{kj}) \geq 0\bigg) \leq \Phi\bigg(C\sqrt{\delta}\Big((\sqrt{n}+\sqrt{\log p})^2 + \frac{\|\Theta'_{.i}-\Theta'_{.j}\|_2^2}{\sigma^2} + \frac{\|\Theta'_{.i}-\Theta'_{.j}\|_2}{\sigma}\sqrt{\log p}\Big)^{1/2} \\ + \frac{w^\top(\Theta'_{.i}-\Theta'_{.i+1})}{\sigma}\bigg) + \frac{C}{p^c}. \end{aligned}$$

The rest of the analysis is divided into two cases.

Case 1. $\log p \lesssim n$. In this case, we have

$$\begin{aligned} P\bigg(\sum_{k=1}^{n} \hat{w}_k(X_{ki}-X_{kj}) \geq 0\bigg) \leq \Phi\bigg(C\sqrt{\delta}\Big(n + \frac{\|\Theta'_{.i}-\Theta'_{.j}\|_2^2}{\sigma^2} + \frac{\|\Theta'_{.i}-\Theta'_{.j}\|_2}{\sigma}\sqrt{\log p}\Big)^{1/2} \\ + \frac{w^\top(\Theta'_{.i}-\Theta'_{.j})}{\sigma}\bigg) + \frac{C}{p^c}. \end{aligned}$$

In addition, let $\Gamma_{ij} = \lambda_1(\Theta')|v'_{1i} - v'_{1j}| \geq |i-j|\Gamma$,

1. if $\|\Theta'_{.i}-\Theta'_{.j}\|_2 \lesssim \sigma\sqrt{n}$, we have

$$P\bigg(\sum_{k=1}^{n} \hat{w}_k(X_{ki}-X_{kj}) \geq 0\bigg) \leq \Phi\bigg(C\sqrt{\delta n} + \frac{w^\top(\Theta'_{.i}-\Theta'_{.j})}{\sigma}\bigg) + \frac{C}{p^c} \leq \frac{C}{p^c},$$

since

$$\lambda_1^2(\Theta') \gtrsim \lambda_2^2(\Theta') + \frac{\sigma^2 n(n+\log p)}{\Gamma^2/\sigma^2} + \frac{\sigma^2\sqrt{np(n+\log p)}}{\Gamma/\sigma}$$
$$\geq \lambda_2^2(\Theta') + \frac{\sigma^2 n(n+\log p)}{\Gamma_{ij}^2/\sigma^2} + \frac{\sigma^2\sqrt{np(n+\log p)}}{\Gamma_{ij}/\sigma},$$

and $\Gamma_{ij} \geq \sigma\sqrt{C\log p}$ for $(i,j)\in S_1$.

2. if $\|\Theta'_{.i} - \Theta'_{.j}\|_2 \gtrsim \sigma\sqrt{n}$, then let $\Xi_{ij} = \|\Theta'_{.i} - \Theta'_{.j}\|_2 \leq |i-j|\Xi$, we have

$$P\bigg(\sum_{k=1}^n \hat{w}_k(X_{ki} - X_{kj}) \geq 0\bigg) \leq \Phi\bigg(\frac{C\sqrt{\delta}}{\sigma}\|\Theta'_{.i} - \Theta'_{.j}\|_2 + \frac{w^\top(\Theta'_{.i} - \Theta'_{.j})}{\sigma}\bigg) + \frac{C}{p^c} \leq \frac{C}{p^c},$$

where the last inequality follows from

$$\lambda_1^2(\Theta') \gtrsim \lambda_2^2(\Theta') + \frac{\sigma^2\Xi^2(n+\log p)}{\Gamma^2} + \frac{\sigma^2\Xi\sqrt{p(n+\log p)}}{\Gamma}$$
$$\geq \lambda_2^2(\Theta') + \frac{\sigma^2\Xi_{ij}^2(n+\log p)}{\Gamma_{ij}^2} + \frac{\sigma^2\Xi_{ij}\sqrt{p(n+\log p)}}{\Gamma_{ij}},$$

and $\Gamma_{ij} \geq \sigma\sqrt{C\log p}$.

This completes the proof of case 1.

Case 2. $\log p \gtrsim n$**.** In this case, we have

$$P\bigg(\sum_{k=1}^n \hat{w}_k(X_{ki} - X_{kj}) \geq 0\bigg) \leq \Phi\bigg(C\sqrt{\delta}\bigg(\log p + \frac{\|\Theta'_{.i} - \Theta'_{.j}\|_2^2}{\sigma^2} + \frac{\|\Theta'_{.i} - \Theta'_{.j}\|_2}{\sigma}\sqrt{\log p}\bigg)^{1/2}$$
$$+ \frac{w^\top(\Theta'_{.i} - \Theta'_{.j})}{\sigma}\bigg) + \frac{C}{p^c}.$$

In addition, since $\Gamma_{ij} \geq \sigma\sqrt{C\log p}$, similar arguments yield

$$P\bigg(\sum_{k=1}^n \hat{u}_k(X_{ki} - X_{kj}) \geq 0\bigg) \leq \frac{C}{p^c}.$$

This completes the proof of Case 2. Combining Case 1 and Case 2, we have proven

$$\sum_{(i,j)\in S_1} P\bigg(\sum_{k=1}^n \hat{w}_k(X_{ki} - X_{kj}) \geq 0\bigg) \leq \frac{C|S_1|}{p^c} \leq \frac{C}{p^{c_0}}. \tag{1.8}$$

Now for the second term in (1.7), since $\Gamma|i-j| \le \Gamma_{ij}$, we have for $(i,j) \in S_2$,

$$P\left(\sum_{k=1}^{n} \hat{w}_k(X_{ki} - X_{kj}) \ge 0\right) \le \frac{ce^{-|i-j|^2\Gamma^2/(2\sigma^2)}}{|i-j|\Gamma/\sigma + \sqrt{|i-j|^2\Gamma^2/\sigma^2 + 8/\pi}} + \frac{C}{p^c}.$$

Note that, on the one hand,

$$\frac{ce^{-|i-j|^2\Gamma^2/(2\sigma^2)}}{|i-j|\Gamma/\sigma + \sqrt{|i-j|^2\Gamma^2/\sigma^2 + 8/\pi}} \le \frac{c\sigma e^{-|i-j|^2\Gamma^2/2\sigma^2}}{|i-j|\Gamma}.$$

Following similar argument that leads to (22) of the main paper, we have

$$\sum_{(i,j)\in S_2} P\left(\sum_{k=1}^{n} \hat{w}_k(X_{ki} - X_{kj}) \ge 0\right) \le \frac{C\sigma p}{\Gamma} e^{-\Gamma^2/2\sigma^2} \log\left(1 + \frac{2\sigma^2}{\Gamma^2}\right) + \frac{C}{p^c}. \tag{1.9}$$

On the other hand, note that

$$\frac{e^{-|i-j|^2\Gamma^2/(2\sigma^2)}}{|i-j|\Gamma/\sigma + \sqrt{|i-j|^2\Gamma^2/\sigma^2 + 8/\pi}} \le ce^{-|i-j|^2\Gamma^2/(2\sigma^2)},$$

we have

$$\sum_{(i,j)\in S_2} P\left(\sum_{k=1}^{n} \hat{w}_k(X_{ki} - X_{kj}) \ge 0\right) \le Cp\sigma/\Gamma + \frac{C'}{p^c}. \tag{1.10}$$

Combining (1.9) and (1.10), we have

$$\sum_{(i,j)\in S_2} P\left(\sum_{k=1}^{n} \hat{w}_k(X_{ki} - X_{kj}) \ge 0\right) \le C\frac{p\sigma}{\Gamma} \min\left\{1, e^{-\Gamma^2/2\sigma^2} \log\left(1 + \frac{2\sigma^2}{\Gamma^2}\right)\right\} + \frac{C}{p^c} \tag{1.11}$$

Combining (1.8) and (1.11), we have

$$\sum_{(i,j):j>i+1} P\left(\sum_{k=1}^{n} \hat{w}_k(X_{ki} - X_{kj}) \ge 0\right) \le C\frac{p\sigma}{\Gamma} \min\left\{1, e^{-\Gamma^2/2\sigma^2} \log\left(1 + \frac{2\sigma^2}{\Gamma^2}\right)\right\} + \frac{C}{p^c},$$

which leads to (1.5). □

2 Proofs of Proposition 1-4

Proof of Proposition 1. Since $\Theta \in \mathcal{D}$, for any nonnegative unit vector $w \in \mathbb{R}^n$, the vector $w^\top\Theta \in \mathbb{R}^p$ is monotonic increasing, so that $\mathfrak{r}(w^\top\Theta) = id$. It then follows that $\mathfrak{r}(w^\top\Theta\Pi) = \mathfrak{r}(w^\top\Theta) \circ \pi^{-1} = \pi^{-1}$. □

Proof of Proposition 2. The key observation is, for any Γ, we have

$$\Lambda \geq C\Gamma^2 p^3. \tag{2.1}$$

Then if $\Gamma \gtrsim \sigma\sqrt{n}$ holds, apparently GSS can be implied by (2.1) and MSG; if $\Gamma \lesssim \sigma\sqrt{n}$, then the results depend on the relative magnitude of $\sigma^2 n^2/\Gamma^2$ and p. Specifically, if $p \gtrsim \sigma^2 n^2/\Gamma^2$, then $p \gtrsim (\sigma^3 n/\Gamma^3)^{2/5}$ implies GSS; if $p \lesssim \sigma^2 n^2/\Gamma^2$, then $p \gtrsim (\sigma^4 n^2/\Gamma^4)^{1/3}$ implies GSS. $\square$

Proof of Proposition 3. Note that by SVD of Θ', we have

$$v_1' = \underset{\|x\|_2=1}{\arg\max}\, x^\top \Theta^\top \Theta x = \underset{\|x\|_2=1}{\arg\max} \sum_{i=1}^{n} \left(\sum_{j=1}^{p} x_j \theta_{ij}' \right)^2. \tag{2.2}$$

To prove that v_1' is monotone, we need the following rearrangement inequality.

Lemma 9 (Rearrangement Inequality)**.** *If $a_1 \geq a_2 \geq ... \geq a_n$ and $b_1 \geq b_2 \geq ... \geq b_n$, then*

$$a_n b_1 + ... + a_1 b_n \leq a_{\sigma(1)} b_1 + ... + a_{\sigma(n)} b_n \leq a_1 b_1 + ... + a_n b_n,$$

where σ is any permutation in S_n.

Lemma 10. *For any Θ' defined above, its first left singular vector u_1' is either nonpositive or nonnegative.*

Now since u_1' is either nonnegative or nonpositive, then we know that $\sum_{j=1}^{p} v_j \theta_{ij}'$ have the same sign for all $i = 1, ..., n$. By Lemma 9, we have that the components of v_1' are either in increasing order or in decreasing order. $\square$

Proof of Proposition 4. Note that

$$v_1' = \underset{\|x\|_2=1}{\arg\max}\, x^\top {\Theta'}^\top \Theta' x = \underset{\|x\|_2=1}{\arg\max} \sum_{i=1}^{n} (x^\top \Theta_{i.}')^2. \tag{2.3}$$

Let $f(x, \Theta', \lambda) = \sum_{i=1}^{n} (x^\top \Theta_{i.}')^2 + \lambda(\|x\|_2 - 1)$. By Lagrange's multiplier method, we have

$$\frac{\partial f}{\partial x_j} = \sum_{i=1}^{n} 2\theta_{ij}(x^\top \Theta_{i.}') + 2\lambda x_j = 0,$$

for $j = 1, ..., p$. It follows that

$$\frac{\sum_{i=1}^{n} \theta_{ij}' ({v'}_1^\top \Theta_{i.}')}{v_{1j}'} = -\lambda, \qquad \text{for all } j = 1, ..., p,$$

where

$$\lambda^2 = \sum_{j=1}^{p} \left[\sum_{i=1}^{n} \theta'_{ij} ({v'}_1^\top \Theta'_{i.}) \right]^2 = \|{v'}_1^\top {\Theta'}^\top \Theta'\|_2^2.$$

Thus, by Lemma 10, we have

$$\begin{aligned} |v'_{1,j+1} - v'_{1,j}| &= \frac{\sum_{i=1}^{n} |\theta'_{i,j+1} - \theta'_{i,j}| |{v'}_1^\top \Theta'_{i.}|}{|\lambda|} \\ &\geq \frac{\sum_{i=1}^{n} \delta \|\Theta_{i.}\|_2 |{v'}_1^\top \Theta'_{i.}|}{|\lambda|} \\ &\geq \delta, \end{aligned}$$

where the last inequality follows from

$$|\lambda| = \|{v'}_1^\top {\Theta'}^\top \Theta'\|_2 = \left\| \sum_{i=1}^{n} (v_1^\top \Theta'_{i.}) \Theta'_{i.} \right\|_2 \leq \sum_{i=1}^{n} \|({v'}_1^\top \Theta'_{i.}) \Theta'_{i.}\|_2 \leq \sum_{i=1}^{n} ({v'}_1^\top \Theta'_{i.}) \|\Theta'_{i.}\|_2.$$

This completes the proof. □

3 Proofs of Technical Lemmas

Proof of Lemma 1. Note that

$$\phi_{i,i+1}^\top w = \sum_{k=1}^{n} w_k (T_{ki} - T_{k,i+1}) \sim N\left(\|a\|_2 (\eta_i - \eta_{i+1}), \frac{2(p-1)\sigma^2}{p} \right),$$

and

$$\|\phi_{i,i+1}\|_2^2 = \sum_{k=1}^{n} (T_{ki} - T_{k,i+1})^2, \quad \text{where} \quad T_{ki} - T_{k,i+1} \sim N(a_k(\eta_i - \eta_{i+1}), 2(p-1)\sigma^2/p).$$

We construct the standardized chi-square statistic

$$Y_k^2 = \left[\frac{T_{k,i} - T_{k,i+1} - a_k(\eta_i - \eta_{i+1})}{\sigma \sqrt{2(p-1)/p}} \right]^2.$$

Define $Q = \frac{p-1}{p} 2\sigma^2 \cdot \sum_{k=1}^{n} (Y_k^2 - 1)$. Then

$$Q = \|\phi_{i,i+1}\|_2^2 + \|a\|_2^2 (\eta_i - \eta_{i+1})^2 - 2(\eta_i - \eta_{i+1}) \sum_{k=1}^{n} (T_{ki} - T_{k,i+1}) a_k - \frac{p-1}{p} \cdot 2n\sigma^2.$$

Our next result follows from an exponential inequality for chi-square random variables proved by Laurent and Massart (2000).

Lemma 11. *Let $(Y_1, ..., Y_n)$ be i.i.d. Gaussian variables with mean 0 and variance 1. Let $a_1, ..., a_n$ be nonnegative. We set $\|a\|_\infty = \sup_{1\le i\le n}|a_i|, \|a\|_2^2 = \sum_{i=1}^n a_i^2$. Let $Z = \sum_{i=1}^n a_i(Y_i^2 - 1)$. Then the following inequalities hold for any positive x:*

$$P(Z \geq 2\|a\|_2\sqrt{x} + 2\|a\|_\infty x) \leq \exp(-x), \tag{3.1}$$

$$P(Z \leq -2\|a\|_2\sqrt{x}) \leq \exp(-x). \tag{3.2}$$

By choosing $x = c\log p$, we have, with probability at least $1 - \frac{1}{p^c}$,

$$\begin{aligned}\|\phi_{i,i+1}\|_2^2 &\leq c\sigma^2\sqrt{n\log p} + c\sigma^2\log p - \|a\|_2^2(\eta_i - \eta_{i+1})^2 + 2(\eta_i - \eta_{i+1})\sum_{k=1}^n (T_{ki} - T_{k,i+1})a_k + \frac{p-1}{p}\cdot 2n\sigma^2 \\ &\leq c\sigma^2(\sqrt{n} + \sqrt{\log p})^2 - \|a\|_2^2(\eta_i - \eta_{i+1})^2 + 2(\eta_i - \eta_{i+1})\sum_{k=1}^n (T_{ki} - T_{k,i+1})a_k\end{aligned} \tag{3.3}$$

Note that the term

$$(\eta_i - \eta_{i+1})\sum_{k=1}^n (T_{ki} - T_{k,i+1})a_k \sim N\left(\|a\|_2^2(\eta_i - \eta_{i+1})^2, 2\|a\|_2^2(\eta_i - \eta_{i+1})^2 \cdot \frac{p-1}{p}\sigma^2\right)$$

It follows from the tail bound of standard Gaussian distribution $\Phi(-x) \leq \frac{1}{x}\phi(x)$ that

$$P\left((\eta_i - \eta_{i+1})\sum_{k=1}^n (T_{ki} - T_{k,i+1})a_k \geq \|a\|_2^2(\eta_i - \eta_{i+1})^2 + 2\sigma\sqrt{c\log p}\cdot\|a\|_2|\eta_i - \eta_{i+1}|\right) \leq \frac{1}{p^c}. \tag{3.4}$$

Combining (3.3) and (3.4), we have

$$P\left(\|\phi_{i,i+1}\|_2^2 \leq c'\sigma^2(\sqrt{n} + \sqrt{\log p})^2 + \|a\|_2^2(\eta_i - \eta_{i+1})^2 + c'\sigma\sqrt{\log p}\|a\|_2|\eta_i - \eta_{i+1}|\right) \geq 1 - \frac{C}{p^c}$$

for some constant $C, c, c' > 0$. Thus,

$$\begin{aligned}&P\left(w^\top\phi_{i,i+1} \geq -\sqrt{2\delta}\|\phi_{i,i+1}\|\right) \\ &\quad\leq \Phi\left(\frac{C\sqrt{\delta}}{\sigma}\left[\sigma^2(\sqrt{n} + \sqrt{\log p})^2 + \|a\|_2^2(\eta_i - \eta_{i+1})^2 + \|a\|_2|\eta_i - \eta_{i+1}|\sigma\sqrt{\log p}\right]^{1/2} + \frac{\|a\|_2(\eta_i - \eta_{i+1})}{\sigma}\right) \\ &\quad\quad + \frac{C}{p^c}\end{aligned}$$

□

Proof of Lemma 2. The proof follows from the following deterministic perturbation bound and concentration inequalities adapted from Cai and Zhang (2018).

Lemma 12. *Suppose $A \in \mathbb{R}^{n\times p}$, $\tilde{V} = \begin{bmatrix} V & V_\perp \end{bmatrix} \in \mathbb{O}_n$ are left singular vectors of A, $V \in \mathbb{O}_{n,r}, V_\perp \in \mathbb{O}_{n,n-r}$ correspond to the first r and last $(n-r)$ singular vectors respectively. $\tilde{W} = \begin{bmatrix} W & W_\perp \end{bmatrix} \in \mathbb{O}_n$ is any orthogonal matrix with $W \in \mathbb{O}_{n,r}$, $W_\perp \in \mathbb{O}_{n,n-r}$. Given that $\lambda_r(W^\top A) > \lambda_{r+1}(A)$, we have*

$$\|\sin\Theta(V,W)\| \leq \frac{\lambda_r(W^\top A)\|\mathbb{P}_{A^\top W}A^\top W_\perp\|^2}{\lambda_r^2(W^\top A) - \lambda_{r+1}^2(A)} \wedge 1.$$

Lemma 13. *Suppose $X \in \mathbb{R}^{n\times p}$ is a rank-r matrix with left singular space as $V \in \mathbb{O}_{n,r}$, $Z \in \mathbb{R}^{n\times p}$, Z is a i.i.d. sub-Gaussian random matrix with sub-Gaussian parameter τ. $Y = X + Z$. Then there exists constants C, c such that for any $x > 0$,*

$$P\big(\lambda_r^2(V^\top Y) \leq (\lambda_r^2(X) + \tau^2 p)(1-x)\big) \leq C\exp\left\{Cr - c\tau^{-2}(\lambda_r^2(X)+\tau^2 p)(x^2 \wedge x)\right\}, \quad (3.5)$$

$$P\big(\lambda_{r+1}^2(Y) \geq \tau^2 p(1+x)\big) \leq C\exp\left\{Cn - cp\cdot(x^2\wedge x)\right\} \quad (3.6)$$

Moreover, there exists C_0, C, c such that whenever $\lambda_r^2(X) \geq C_0\tau^2 n$, for any $x > 0$ we have

$$P(\|\mathbb{P}_{V^\top Y}V_\perp^\top Y\| \leq x) \geq 1 - C\exp\left\{Cn - c\tau^{-2}\min(x^2, x\sqrt{\lambda_r^2(X)+\tau^2 p})\right\} - C\exp\left\{-c\tau^{-2}(\lambda_r^2(X)+\tau^2 p)\right\}. \quad (3.7)$$

Note that $T = \Theta' + E$ where $\Theta' = \|a\|\cdot w\eta'^\top$ is rank 1, and $\hat{w}$ is the first left singular vector of T, by Lemma 12, we have

$$|1 - (\hat{w}^\top w)^2| = \|\sin\Theta(w,\hat{w})\|^2 \leq \frac{\lambda_1^2(w^\top T)\|\mathbb{P}_{w^\top T}w_\perp^\top T\|^2}{(\lambda_1^2(w^\top T) - \lambda_2^2(T))^2} \wedge 1. \quad (3.8)$$

To bound the quantity $\lambda_1^2(w^\top T)$, using Lemma 13, by choosing $x = \frac{\lambda^2}{3(\lambda^2+\sigma^2 p)}$ in (3.5), since $\lambda^2 \geq C\sigma^2(n + \sqrt{pn})$ for some sufficiently large C, or $\sigma^2(\lambda^2+\sigma^2 p)(\sigma^2 n + \log p)/\lambda^4 \leq 1$, we have

$$c\sigma^{-2}\min\left\{\frac{\lambda^4}{\lambda^2+\sigma^2 p}, \lambda^2\right\} - C \geq \frac{c_0\sigma^{-2}\lambda^4}{\lambda^2+\sigma^2 p} - C \geq c_1 \log p.$$

Thus

$$P\big(\lambda_1^2(w^\top T) \leq \frac{2\lambda^2}{3} + \sigma^2 p\big) \leq C\exp\left\{C - \frac{c_0\sigma^{-2}\lambda^4}{\lambda^2 + p}\right\} \leq C\exp\{-c_1\log p\}. \quad (3.9)$$

To bound $\lambda_2^2(T)$, if $p \le n + \log p$, by choosing $x = \frac{n+\log p}{p}$ in (3.6), we have

$$c \min\left\{\frac{(n+\log p)^2}{p}, n + \log p\right\} - Cn \ge c \log p$$

for some $C, c > 0$, so that

$$P\big(\lambda_2^2(T) \ge \sigma^2(p + n + \log p)\big) \le C \exp\left\{Cn - cp \cdot x^2 \wedge x\right\} \le C \exp\{-c \log p\} = \frac{C}{p^c}. \quad (3.10)$$

If $p > n + \log p$, by choosing $x = \sqrt{\frac{n+\log p}{p}}$ in (3.6), we have

$$c \min\left\{\sqrt{p(n+\log p)}, n + \log p\right\} - Cn \ge c \log p$$

for some $C, c > 0$, so that

$$P\big(\lambda_2^2(T) \ge \sigma^2(p + \sqrt{p(n+\log p)})\big) \le C \exp\left\{Cn - cp \cdot x^2 \wedge x\right\} \le C \exp\{-c \log p\} = \frac{C}{p^c}. \quad (3.11)$$

Lastly, to bound $\|\mathbb{P}_{w^\top T} w_\perp^\top T\|$, choosing $x = \sigma\sqrt{n + \log p}$ in (3.7), since $\lambda^2 \ge \sigma^2 n$, we have

$$\sigma^2(n + \log p) \le c(\lambda^2 + \sigma^2 p)$$

for sufficiently large $c > 0$ and therefore

$$c\sigma^{-2} \min\left\{x^2, x\sqrt{\lambda^2 + \sigma^2 p}\right\} - Cn \ge c \log p.$$

So that

$$P(\|\mathbb{P}_{w^\top T} w_\perp^\top T\| \le \sigma\sqrt{n + \log p}) \ge 1 - \frac{C}{p^c} - C \exp\big\{-c\sigma^{-2}(\lambda^2 + \sigma^2 p)\big\}. \quad (3.12)$$

Combining (3.9) (3.10) (3.11) and (3.12), by the fact that $x/(x-y)^2$ is decreasing in x and increasing in y for $x > y > 0$, we have

$$P\left(\frac{\lambda^2(w^\top T)\|\mathbb{P}_{w^\top T} w_\perp^\top T\|^2}{(\lambda_r^2(w^\top T) - \lambda_2^2(T))^2} \le C' \frac{\sigma^2(\lambda^2 + \sigma^2 p)(n + \log p)}{\lambda^4}\right) \ge 1 - Cp^{-c}.$$

$\square$

Proof of Lemma 7. Note that

$$\phi_{i,i+1}^{\top} w = \sum_{k=1}^{n} w_k (T_{ki} - T_{k,i+1}) \sim N\left(w^{\top} (\Theta'_{\cdot i} - \Theta'_{\cdot i+1}), \frac{2(p-1)\sigma^2}{p} \right),$$

and

$$\|\phi_{i,i+1}\|_2^2 = \sum_{k=1}^{n} (T_{ki} - T_{k,i+1})^2, \quad \text{where} \quad T_{ki} - T_{k,i+1} \sim N(\theta'_{ki} - \theta'_{k,i+1}, 2(p-1)\sigma^2/p).$$

We construct the standardized chi-square statistic

$$Y_k^2 = \left[\frac{T_{k,i} - T_{k,i+1} - (\theta'_{ki} - \theta'_{k,i+1})}{\sigma \sqrt{2(p-1)/p}} \right]^2 .$$

Define $Q = \frac{p-1}{p} 2\sigma^2 \cdot \sum_{k=1}^{n} (Y_k^2 - 1)$. Then

$$Q = \|\phi_{i,i+1}\|_2^2 + \|\Theta'_{\cdot i} - \Theta'_{\cdot i+1}\|_2^2 - 2 \sum_{k=1}^{n} (T_{ki} - T_{k,i+1})(\theta'_{ki} - \theta'_{k,i+1}) - \frac{p-1}{p} \cdot 2n\sigma^2 .$$

By Lemma 11, choosing $x = c \log p$, we have, with probability at least $1 - \frac{1}{p^c}$,

$$\begin{aligned} \|\phi_{i,i+1}\|_2^2 &\leq c\sigma^2 \sqrt{n \log p} + c\sigma^2 \log p - \|\Theta'_{\cdot i} - \Theta'_{\cdot i+1}\|_2^2 + 2 \sum_{k=1}^{n} (T_{ki} - T_{k,i+1})(\theta'_{ki} - \theta'_{k,i+1}) + \frac{p-1}{p} \cdot 2n\sigma^2 \\ &\leq c\sigma^2 (\sqrt{n} + \sqrt{\log p})^2 - \|\Theta'_{\cdot i} - \Theta'_{\cdot i+1}\|_2^2 + 2 \sum_{k=1}^{n} (T_{ki} - T_{k,i+1})(\theta'_{ki} - \theta'_{k,i+1}). \end{aligned} \tag{3.13}$$

Note that the term

$$\sum_{k=1}^{n} (T_{ki} - T_{k,i+1})(\theta'_{ki} - \theta'_{k,i+1}) \sim N\left(\|\Theta'_{\cdot i} - \Theta'_{\cdot i+1}\|_2^2, 2\|\Theta'_{\cdot i} - \Theta'_{\cdot i+1}\|_2^2 \cdot \frac{p-1}{p} \sigma^2 \right)$$

It follows from the tail bound of standard Gaussian distribution $\Phi(-x) \leq \frac{1}{x} \phi(x)$ that

$$P\left(\sum_{k=1}^{n} (T_{ki} - T_{k,i+1})(\theta'_{ki} - \theta'_{k,i+1}) \geq \|\Theta'_{\cdot i} - \Theta'_{\cdot i+1}\|_2^2 + 2\sigma \sqrt{c \log p} \cdot \|\Theta'_{\cdot i} - \Theta'_{\cdot i+1}\|_2 \right) \leq \frac{1}{p^c}. \tag{3.14}$$

Combining (3.13) and (3.14), we have

$$P\left(\|\phi_{i,i+1}\|_2^2 \leq c' \sigma^2 (\sqrt{n} + \sqrt{\log p})^2 + \|\Theta'_{\cdot i} - \Theta'_{\cdot i+1}\|_2^2 + c' \sigma \sqrt{\log p} \|\Theta'_{\cdot i} - \Theta'_{\cdot i+1}\|_2 \right) \geq 1 - \frac{C}{p^c}$$

for some constant $C, c, c' > 0$. Thus,

$$\begin{aligned}
&P\left(w^\top \phi_{i,i+1} \geq -\sqrt{2\delta}\|\phi_{i,i+1}\|\right) \\
&\quad \leq \Phi\left(C\sqrt{\delta}\left[(\sqrt{n}+\sqrt{\log p})^2 + \frac{\|\Theta'_{\cdot i} - \Theta'_{\cdot i+1}\|_2^2}{\sigma^2} + c'\sqrt{\log p}\frac{\|\Theta'_{\cdot i} - \Theta'_{\cdot i+1}\|_2}{\sigma}\right]^{1/2} + \frac{w^\top(\Theta'_{\cdot i} - \Theta'_{\cdot i+1})}{\sigma}\right) \\
&\quad\quad + \frac{C}{p^c}
\end{aligned}$$

□

Proof of Lemma 8. The proof of Lemma 8 depends on Lemma 12 as well as the following concentration inequalities.

Lemma 14. *Suppose $X \in \mathbb{R}^{n\times p}$ is a rank-r matrix with first left singular vector $v_1 \in \mathbb{R}^n$, Z is a i.i.d. sub-Gaussian random matrix with sub-Gaussian parameter τ. $Y = X + Z$. Then there exists constants C, c such that for any $x > 0$, $\lambda_1^2(X) \geq 3\sqrt{\log(n \vee p)}$*

$$P\left(\lambda_1^2(v_1^\top Y) \leq \lambda_1^2(X) + \tau^2(p - \sqrt{p\log(p\vee n)} - \log(p \vee n))\right) \leq \frac{C}{p^c}, \tag{3.15}$$

$$P(\lambda_2^2(Y) > (\lambda_2^2(X) + \tau^2 p)(1+t)) \leq C\exp(Cn - c\tau^{-2}(\lambda_n^2 + \tau^2 p)t^2 \wedge t). \tag{3.16}$$

Moreover, there exists C_0, C, c such that for any $x > 0$ we have

$$P(\|\mathbb{P}_{Y^\top v_1} Y^\top v_{1\perp}\| < t) \geq 1 - C\exp(Cn - c\tau^{-2}\min(t^2, t\sqrt{\lambda_1^2(X) + \tau^2 p})) - C\exp\left\{-c\tau^{-2}(\lambda_1^2(X) + \tau^2 p)\right\}. \tag{3.17}$$

Following a similar argument as the proof of Lemma 2, by taking $t = \sigma\sqrt{\frac{n+\log p}{\lambda_n^2 + \sigma^2 p}}$ in (3.16), since $n \lesssim p$, we have

$$P(\lambda_2^2(Y) > \lambda_2^2(\Theta') + \tau^2 p) \leq \frac{C}{p^c}.$$

Similarly, by taking $t = \sigma\sqrt{n + \log p}$, we have

$$P(\|\mathbb{P}_{Y^\top w} Y^\top w_\perp\|^2 < \sigma^2(n + \log p)) \geq 1 - \frac{C}{p^c}.$$

Then Lemma 12 implies, if $\lambda_1^2(\Theta') - \lambda_2^2(\Theta') \geq C\sigma^2(n + \sqrt{np})$ for some $C > 0$,

$$P\left(|1 - (\hat{w}^\top w)^2| \leq \frac{C\sigma^2(\lambda_1^2(\Theta') + \sigma^2 p)(n + \log p)}{(\lambda_1^2(\Theta') - \lambda_2^2(\Theta'))^2}\right) \geq 1 - \frac{C}{p^c}.$$

□

Proof of Lemma 10. By SVD of $\Theta' = (\theta'_{ij}) \in \mathbb{R}^{n\times p}$, its first left singular vector

$$u'_1 = \arg\max_{\|x\|_2=1} x^\top \Theta' {\Theta'}^\top x = \arg\max_{\|x\|_2=1} \sum_{j=1}^{p} \Big(\sum_{i=1}^{n} x_i \theta'_{ij} \Big)^2. \tag{3.18}$$

In the following, we show that, for any unit vector $x \in \mathbb{R}^n$,

$$\sum_{j=1}^{p} \Big(\sum_{i=1}^{n} x_i \theta'_{ij} \Big)^2 \le \sum_{j=1}^{p} \Big(\sum_{i=1}^{n} |x_i| \theta'_{ij} \Big)^2, \tag{3.19}$$

from which we conclude that u'_1 is either nonpositive or nonnegative. Toward this end, note that

$$\sum_{j=1}^{p} \Big(\sum_{i=1}^{n} x_i \theta'_{ij} \Big)^2 = \sum_{j=1}^{p} \sum_{i=1}^{n} x_i^2 {\theta'}^2_{ij} + \sum_{i\neq k} x_i x_k \Big(\sum_{j=1}^{p} {\theta'}_{ij} {\theta'}_{kj} \Big).$$

Then the inequality (3.19) follows from

$$\sum_{i\neq k} x_i x_k \Big(\sum_{j=1}^{p} {\theta'}_{ij} {\theta'}_{kj} \Big) \le \sum_{i\neq k} |x_i x_k| \Big(\sum_{j=1}^{p} {\theta'}_{ij} {\theta'}_{kj} \Big),$$

which is true as long as $\sum_{j=1}^{p} {\theta'}_{ij} {\theta'}_{kj} \ge 0$ for all pairs $i \neq k$. We conclude this proof by showing that

Fact. For any nondecreasing vectors $a, b \in \mathbb{R}^n$, such that $\sum_{i=1}^{n} a_i = 0$. Then it follows that $a^\top b \ge 0$.

To see this, note that since a and b are both nondecreasing, there exist a constant δ such that the components of $b + \delta \cdot \mathbf{1}$ has the same sign as a. Hence the claim follows from $0 \le a^\top (b + \delta \cdot \mathbf{1}) = a^\top b + \delta a^\top \mathbf{1} = a^\top b$. (Thank Rui Duan for suggesting this simple proof.) □

Proof of Lemma 14. Note that

$$\lambda_1^2(v_1^\top Y) = \|v_1^\top Y\|_2^2 = \lambda_1^2(X) + 2\lambda_1(X) v_1^\top Z u_1 + \|v_1^\top Z\|_2^2.$$

The linear term $v_1^\top Z u_1 = \sum_{i,j} v_{1i} u_{1j} Z_{ij}$ is subgaussian with parameter $c\|v_1 u_1^\top\|_F = c$ for some constant c only depending on τ. Therefore, by concentration inequality for subgaussian random variables

$$P(v_1^\top Z u_1 \ge t) \le \exp\Big\{ -\frac{t^2}{2c\tau^2} \Big\}. \tag{3.20}$$

On the other hand, the quadratic term $\|v_1^\top Z\|_2^2 = \sum_{i=1}^p (v_1^\top Z_i)^2$ where Z_i for $i = 1, ..., p$ are the i-th column of Z. Each of $(v_1^\top Z_i)^2$ is subexponential with mean

$$\mathbb{E}(v_1^\top Z_i)^2 = \mathbb{E}Z_{ij}^2.$$

Then concentration inequality for subexponential random variables yields

$$P\left(\left|\sum_{i=1}^p (v_1^\top Z_i)^2 - p\mathbb{E}Z_{ij}^2\right| \geq t\right) \leq 2\exp\left\{ - c\min\left(\frac{t^2}{\tau^4 p}, \frac{t}{\tau^2}\right)\right\} \tag{3.21}$$

for some constant $c > 0$. By taking $t = \tau\sqrt{\log(p \vee n)}$ in (3.20), we have

$$P(v_1^\top Z u_1 \geq \tau\sqrt{\log(p \vee n)}) \leq \frac{C}{p^c}.$$

Taking $t = \tau^2\sqrt{p\log(p \vee n)}$ if $\sqrt{\log(p \vee n)} \leq \sqrt{p}$ and $t = \tau^2\log(p \vee n)$ otherwise in (3.21), we have

$$P\left(\sum_{i=1}^p (v_1^\top Z_i)^2 \leq \tau^2 p - \tau^2\sqrt{p\log(p \vee n)} - \tau^2\log(p \vee n)\right) \leq \frac{C}{p^c}.$$

Combining these, we have, with probability at least $1 - O(p^{-c})$,

$$\begin{aligned}\lambda_1^2(v_1^\top Y) &= \lambda_1^2(X) + 2\tau\lambda_1(X)v_1^\top Z u_1 + \|v_1^\top Z\|_2^2 \\ &\geq \lambda_1^2(X) - 2\tau\lambda_1(X)\sqrt{\log(p \vee n)} + \tau^2 p - \tau^2\sqrt{p\log(p \vee n)} - \tau^2\log(p \vee n).\end{aligned}$$

If in addition $\lambda_1(X) \geq 3\tau\sqrt{\log(n \vee p)}$, then

$$P\left(\lambda_1^2(v_1^\top Y) \leq \lambda_1^2(X) + \tau^2 p - \tau^2\sqrt{p\log(p \vee n)} - \tau^2\log(p \vee n)\right) \leq \frac{C}{p^c}.$$

For $\lambda_2^2(Y)$, note that

$$\lambda_2(Y) = \min_{\mathrm{rank}(B) \leq 1} \|Y - B\| \leq \|Y - [v_1 \quad \mathbf{0}]^\top \cdot Y\| = \|v_{1\perp}^\top Y\|$$

for $v_{1\perp} \in \mathbb{R}^{n \times (n-1)}$. It suffices to obtain an upper bound for $\lambda_1(v_{1\perp}^\top Y)$ with high probability. Next,

$$\begin{aligned}\|v_{1\perp}^\top Y\|_2^2 &= \|v_{1\perp}^\top YY^\top v_{1\perp}\| \\ &\leq \|\mathbb{E}v_{1\perp}^\top YY^\top v_{1\perp}\| + \|v_{1\perp}^\top YY^\top v_{1\perp} - \mathbb{E}v_{1\perp}^\top YY^\top v_{1\perp}\| \\ &= \lambda_2^2(X) + \tau^2 p + \|v_{1\perp}^\top YY^\top v_{1\perp} - \mathbb{E}v_{1\perp}^\top YY^\top v_{1\perp}\|\end{aligned} \tag{3.22}$$

Define the normalization matrix $M \in \mathbb{R}^{(n-1)\times(n-1)}$ as

$$M = \begin{bmatrix} (\lambda_2^2(X)+\tau^2 p)^{-1/2} & & & & \\ & \ddots & & & \\ & & (\lambda_r^2(X)+\tau^2 p)^{-1/2} & & \\ & & & \ddots & \\ & & & & (\lambda_n^2(X)+\tau^2 p)^{-1/2} \end{bmatrix},$$

we have

$$M^\top v_{1\perp}^\top \mathbb{E}YY^\top v_{1\perp} M = I_{n-1}.$$

Let $Q = v_{1\perp}^\top YY^\top v_{1\perp} - \mathbb{E}v_{1\perp}^\top YY^\top v_{1\perp}$, we have

$$\|Q\| = \|(M^{-1})^\top M^\top QMM^{-1}\| \le \|M^\top QM\| \|M^{-1}\|^2.$$

By construction we have

$$\|M^{-1}\| = (\lambda_2^2(X)+\tau^2 p)^{1/2},$$

then

$$\|Q\| \le (\lambda_2^2(X)+\tau^2 p)\|M^\top v_{1\perp}^\top YY^\top v_{1\perp} M - M^\top v_{1\perp}^\top \mathbb{E}YY^\top v_{1\perp} M\|.$$

Now it suffices to obtain a concentration inequality for $\|M^\top v_{1\perp}^\top YY^\top v_{1\perp} M - I_{n-1}\|$. The main idea is to use the ϵ-net argument to split the spectral norm deviation to the deviations of single random variables, which can be further controlled by the Hanson-Wright inequality. Specifically, for any unit vector $u \in \mathbb{R}^{n-1}$, we have

$$\begin{aligned}
& u^\top M^\top v_{1\perp}^\top YY^\top v_{1\perp} Mu - u^\top I_{n-1} u \\
&= (u^\top M^\top v_{1\perp}^\top XX^\top v_{1\perp} Mu - \mathbb{E}u^\top M^\top v_{1\perp}^\top XX^\top v_{1\perp} Mu) \\
&\quad + 2(u^\top M^\top v_{1\perp}^\top XZ^\top v_{1\perp} Mu - \mathbb{E}u^\top M^\top v_{1\perp}^\top XZ^\top v_{1\perp} Mu) \\
&\quad + (u^\top M^\top v_{1\perp}^\top ZZ^\top v_{1\perp} Mu - \mathbb{E}u^\top M^\top v_{1\perp}^\top ZZ^\top v_{1\perp} Mu) \\
&= 2(X^\top v_{1\perp} Mu)^\top Z^\top (v_{1\perp} Mu) + (v_{1\perp} Mu)^\top (ZZ^\top - \tau^2 p I_n)(v_{1\perp} Mu).
\end{aligned}$$

In the following, we shall bound the two terms separately.

To bound the second term, for any fixed unit vector $u \in \mathbb{R}^{n-1}$, we vectorize $Z \in \mathbb{R}^{n\times p}$ into $\mathrm{vec}(Z) \in \mathbb{R}^{np}$ as

$$\mathrm{vec}(Z) = (Z_{11}, Z_{21}, ..., Z_{n1}, Z_{12}, ..., Z_{n2}, ..., Z_{1p}, Z_{np})^\top.$$

We also define the block diagonal matrix

$$D = \begin{bmatrix} (v_{1\perp}Mu)(v_{1\perp}Mu)^\top & & \\ & \ddots & \\ & & (v_{1\perp}Mu)(v_{1\perp}Mu)^\top \end{bmatrix} \in \mathbb{R}^{np\times np}.$$

It then follows that

$$(v_{1\perp}Mu)^\top(ZZ^\top - \tau^2 pI_n)(v_{1\perp}Mu) = \mathrm{vec}(Z)^\top D \cdot \mathrm{vec}(Z) - \mathbb{E}\mathrm{vec}(Z)^\top D \cdot \mathrm{vec}(Z).$$

Besides,

$$\|D\| = \|(v_{1\perp}Mu)(v_{1\perp}Mu)^\top\| = \|Mu\|_2^2 \le \|M\|^2 = (\lambda_n^2(X) + \tau^2 p)^{-1},$$

$$\|D\|_F^2 = p\|(v_{1\perp}Mu)(v_{1\perp}Mu)^\top\|_F^2 \le p\|Mu\|_2^4 \le p(\lambda_n^2(X) + \tau^2 p)^{-2}.$$

By Hansen-Wright inequality (Rudelson and Vershynin, 2013),

$$\begin{aligned} &P(|(v_{1\perp}Mu)^\top(ZZ^\top - \tau^2 pI_n)(v_{1\perp}Mu)| > t) \\ &= P(|\mathrm{vec}(Z)^\top D \cdot \mathrm{vec}(Z) - \mathbb{E}\mathrm{vec}(Z)^\top D \cdot \mathrm{vec}(Z)| > t) \\ &\le 2\exp\left\{ -c\min\left(\frac{t^2(\lambda_n^2(X) + \tau^2 p)^2}{\tau^4 p}, \frac{t(\lambda_n^2(X) + \tau^2 p)}{\tau^2}\right)\right\} \end{aligned} \tag{3.23}$$

for some $c > 0$.

Next, we bound the first term

$$\begin{aligned} (X^\top v_{1\perp}Mu)^\top Z^\top(v_{1\perp}Mu) &= \mathrm{tr}(Z^\top(v_{1\perp}Mu)(X^\top v_{1\perp}Mu)^\top) \\ &= \mathrm{vec}(Z)^\top \cdot \mathrm{vec}((v_{1\perp}Mu)(X^\top v_{1\perp}Mu)^\top). \end{aligned}$$

Since

$$X^\top v_{1\perp}M = U\begin{bmatrix} 0 & \dots & 0 & 0 \\ \lambda_2(X)(\lambda_2^2(X) + \tau^2 p)^{-1/2} & & & 0 \\ & \ddots & & \vdots \\ & & \lambda_n(X)(\lambda_n^2(X) + \tau^2 p)^{-1/2} & 0 \end{bmatrix},$$

we know $\|X^\top v_{1\perp} M\| \le 1$ and

$$\begin{aligned}\|\mathrm{vec}((v_{1\perp}Mu)(X^\top v_{1\perp}Mu)^\top)\|_2^2 &= \|(v_{1\perp}Mu)(X^\top v_{1\perp}Mu)^\top\|_F^2 \\ &= \|v_{1\perp}Mu\|_2^2 \cdot \|X^\top v_{1\perp}Mu\|_2^2 \\ &\le \|M\|^2 \le (\lambda_n^2 + \tau^2 p)^{-1}.\end{aligned}$$

By the concentration inequality for i.i.d. subgaussian random variables, we have

$$P(|(X^\top v_{1\perp}Mu)^\top Z^\top (v_{1\perp}Mu)| > t) \le C \exp\bigg(-c\frac{t^2(\lambda_n^2+\tau^2 p)}{\tau^2}\bigg) \tag{3.24}$$

for some constant $C, c > 0$. Combining (3.24) and (3.23), we have, for any fixed unitary $u \in \mathbb{R}^{n-1}$,

$$P(|u^\top M^\top v_{1\perp}^\top YY^\top v_{1\perp}Mu - u^\top I_{n-1}u| > t) \le C\exp(-c\tau^{-2}(\lambda_n^2+\tau^2 p)t^2 \wedge t)$$

for all $t > 0$. Next, we use the following lemma proved by Cai and Zhang (2018) concerning the ϵ-net argument for unit ball.

Lemma 15. *For any $p \ge 1$, denote $\mathbb{B}^p = \{x \in \mathbb{R}^p : \|x\|_2 \le 1\}$ as the p-dimensional unit ball in the Euclidean space. Suppose $K \in \mathbb{R}^{p_1 \times p_2}$ is a random matrix. Then we have for $t > 0$,*

$$P(\|K\| \ge 3t) \le 7^{p_1+p_2} \cdot \max_{u\in\mathbb{B}^{p_1}, v\in\mathbb{B}^{p_2}} P(|u^\top K v| \ge t).$$

It then follows that

$$P(\|M^\top v_{1\perp}^\top YY^\top v_{1\perp}M - I_{n-1}\| > t) \le C\exp(Cn - c\tau^{-2}(\lambda_n^2+\tau^2 p)t^2 \wedge t). \tag{3.25}$$

Recall (3.22), we have

$$P(\lambda_2^2(Y) > \lambda_2^2(X) + \tau^2 p + (\lambda_2^2(X)+\tau^2 p)t) \le C\exp(Cn - \tau^{-2}(\lambda_n^2+\tau^2 p)t^2 \wedge t).$$

Finally, we consider $\|\mathbb{P}_{v_1^\top Y} v_{1\perp}^\top Y\|$. Define the constant

$$m = (\lambda_1^2(X) + \tau^2 p)^{-1/2}.$$

It then follows that

$$m^2 v_1^\top \mathbb{E} YY^\top v_1 = 1.$$

Since

$$\begin{aligned}\|\mathbb{P}_{Y^\top v_1} Y^\top v_{1\perp}\| &= \|\mathbb{P}_{mY^\top v_1} Y^\top v_{1\perp}\| \\ &= \|mY^\top v_1((mY^\top v_1)^\top (mY^\top v_1))^{-1}(mY^\top v_1)^\top Y^\top v_{1\perp}\| \\ &\leq \|mY^\top v_1\|_2^{-1}\|(mY^\top v_1)^\top Y^\top v_{1\perp}\|.\end{aligned}$$

In the following we analyze $\|mY^\top v_1\|_2$ and $\|(mY^\top v_1)^\top Y^\top v_{1\perp}\|$ separately.

Since

$$\|mY^\top v_1\|^2 = m^2|v_1^\top YY^\top v_1| \geq 1 - |m^2 v_1^\top YY^\top v_1 - m^2 v_1^\top \mathbb{E}YY^\top v_1|,$$

following the same argument that leads to (3.25), we have

$$P(|m^2 v_1^\top YY^\top v_1| > 1-t) \geq 1 - C\exp(C - c\tau^{-2}(\lambda_1^2(X) + \tau^2 p)t^2 \wedge t).$$

Now set $t = 1/2$, we can choose C_0 large enough, such that $\lambda_1^2(X) \geq \tau^2 C_0$ and therefore $C - c\tau^{-2}(\lambda_1^2(X) + \tau^2 p) \leq -c'\tau^{-2}(\lambda_1^2(X) + \tau^2 p)$ for some $c, c' > 0$. In this case,

$$P(|m^2 v_1^\top YY^\top v_1| > 1/2) \geq 1 - C\exp(-c\tau^{-2}(\lambda_1^2(X) + \tau^2 p)).$$

For $\|(mY^\top v_1)^\top Y^\top v_{1\perp}\|$, note that $mv_1^\top XX^\top v_{1\perp} = 0$, we have

$$\begin{aligned}(mY^\top v_1)^\top Y^\top v_{1\perp} &= mv_1^\top (X+Z)(X+Z)^\top v_{1\perp} \\ &= mv_1^\top XZ^\top v_{1\perp} + mv_1^\top ZX^\top v_{1\perp} + mv_1^\top ZZ^\top v_{1\perp}.\end{aligned}$$

Following similar idea of the proof of (3.25), we can show for any unit vector $u \in \mathbb{R}^{n-1}$,

$$P(|mv_1^\top XZ^\top v_{1\perp}u| > t) \leq C\exp\left(\frac{-ct^2}{\tau^2\|(v_{1\perp}u)(mv_1^\top X)\|_F^2}\right) \leq C\exp(-c\tau^{-2}t^2),$$

$$P(|mv_1^\top ZZ^\top v_{1\perp}u| > t) = P(|mv_1^\top (ZZ^\top - \tau^2 pI_n)v_{1\perp}u| > t) \leq C\exp(-c\tau^{-2}\min(t^2, \sqrt{\lambda_1^2(X) + \tau^2 p}t)).$$

By the ϵ-net argument again (Lemma 15), we have

$$P(\|(mY^\top v_1)^\top Y^\top v_{1\perp}\| > t) \leq C\exp(Cn - c\tau^{-2}\min(t^2, t\sqrt{\lambda_1^2(X) + \tau^2 p})).$$

$\square$

A Comparison with an SVD-based Esitmator

In this section, we compare the empirical performance of our proposed estimator $\hat{\pi}$ to that of an alternative estimator $\tilde{\pi} = (\mathfrak{r}(\hat{v}_1))^{-1}$ where $\hat{v}_1$ is the first right singular vector of Y. This estimator is closely related to $\hat{\pi}$ except that it does not centralize the rows in Y before estimating its singular subspaces. However, this normalization step is essential in order for the resulting estimator to be invariant to the unknown intercepts of the growth models. The signal matrix $\Theta = (\theta_{ij}) \in \mathbb{R}^{n\times p}$ is generated under the following two regimes:

- $S_1(\alpha, n, p)$: For any $1 \le j \le p$, $\theta_{ij} = \log(1 + \alpha j + \beta_i)$ for $1 \le i \le n/2$ where as $\theta_{ij} = 0$ for $n/2 < i \le n$, $\beta_i \sim \text{Unif}(1, 3)$ for all $1 \le i \le n$;
- $S_2(\alpha, n, p)$: For any $1 \le j \le p$, $\theta_{ij} = \alpha j + \beta_i$ for $i = 1$ where as $\theta_{ij} = 0$ for $2 \le i \le n$, $\beta_i \sim \text{Unif}(1, 3)$ for all $1 \le i \le n$;

In each setting, we evaluate the empirical performance of each method over a range of n, p and α. Each setting is repeated for 200 times. The empirical normalized Kendall's tau is reported using boxplots, as shown in Figure 1 of our supplementary material. From Figure 1, our proposed estimator $\hat{\pi}$ performs better than $\tilde{\pi}$ in all the settings, especially in $S_2(\alpha, n, p)$ where the signals are concentrated at one row.

B Supplementary Figures and Tables

In Figure 2, the graphical representation of the weight vectors $\hat{w}$ for our proposed estimator $\hat{\pi}$, and the pseudo-weight vector $\tilde{w}$ for the estimator π_{max} based on 200 simulations under four different models in Section 6.1 of our main paper is given.

In Figure 3, the Taxonomic tree of 45 closely related species used in generating the shotgun metagenomic data used on in s Gao and Li (2018) as well as Section 6.2 of our main paper is given.

Table S.1 lists the p-values of 8 contig clusters from the Wilcoxon rank sum test of the ePTRs between the responser and non-responser groups, and the taxonomic annotations with lineage scores indicating the quality of each taxonomic classification (see Section 6.3 of our main paper).

References

Abramowitz, M. and I. A. Stegun (1965). *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*, Volume 55. Courier Corporation.

Cai, T. T. and A. Zhang (2018). Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics. *The Annals of Statistics 46*(1), 60–89.

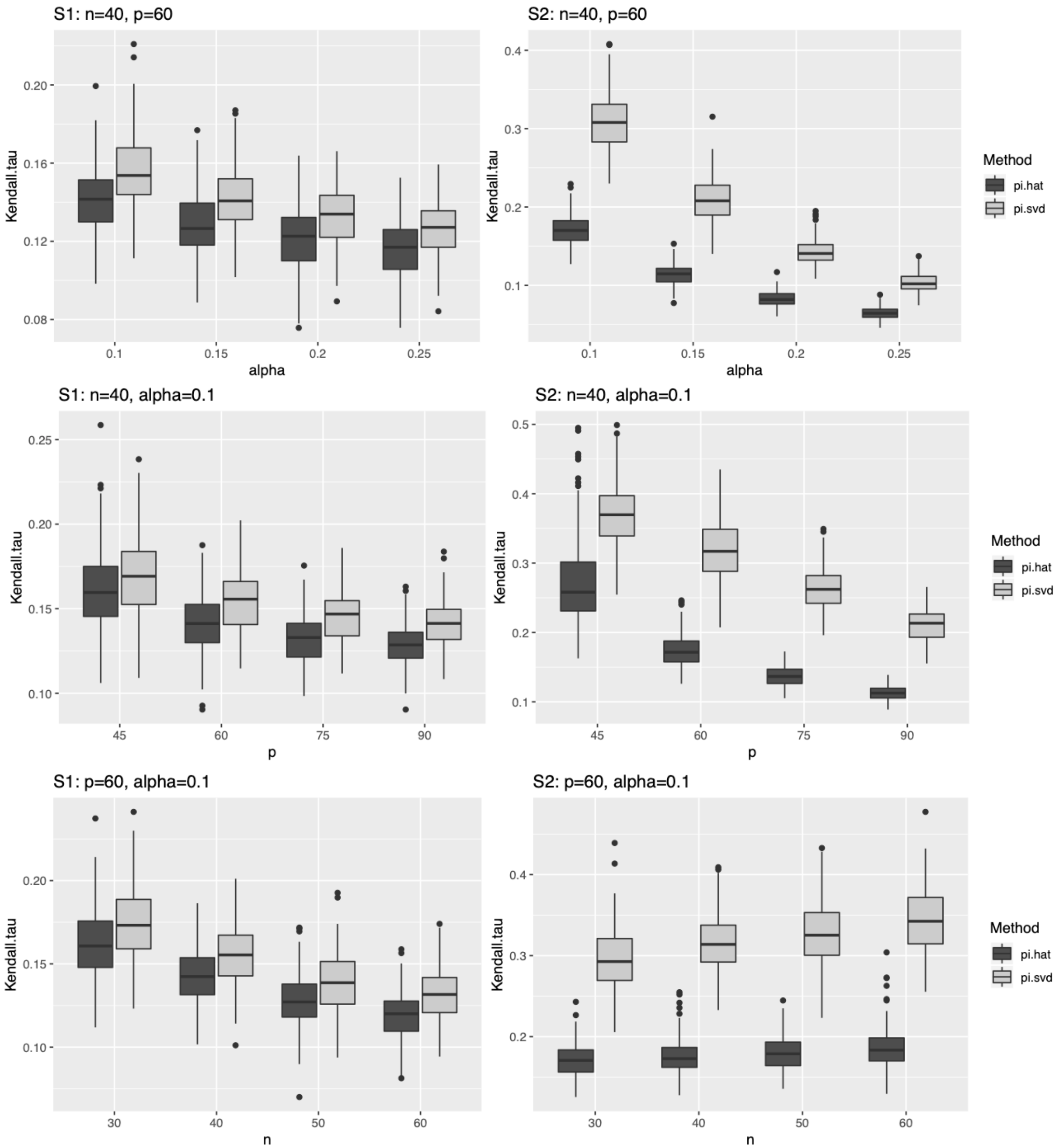

Figure 1: Boxplots of the empirical normalized Kendall's distance between the estimated permutation and true permutation under models $S_1(\alpha, p, n)$ and $S_2(\alpha, p, n)$. $\hat{\pi}$: proposed estimator; π_{svd}: estimator based on SVD.

Gao, Y. and H. Li (2018). Quantifying and comparing bacterial growth dynamics in multiple metagenomic samples. *Nature Methods 15*, 1041–1044.

Table S.1: The p-values of 8 contig clusters from the Wilcoxon rank sum test of the ePTRs between the responser and non-responser groups, and the taxonomic annotations with lineage scores indicating the quality of each taxonomic classification.

Contig Clusters (n_1, n_2)	P-values	Taxonomic Annotations with Lineage Scores
bin.004 (8,8)	0.9592	Firmicutes (phylum): 0.88; Clostridia (class): 0.78; Clostridiales (order): 0.78; Lachnospiraceae (family): 0.40
bin.007 (9,11)	0.4119	Firmicutes (phylum): 0.96; Clostridia (class): 0.92; Clostridiales (order): 0.92;
bin.016 (5,8)	0.3543	Firmicutes (phylum): 0.86; Clostridia (class): 0.74; Clostridiales (order): 0.74;
bin.017 (7,7)	0.2086	Firmicutes (phylum): 0.92; Erysipelotrichia (class): 0.47; Erysipelotrichales (order): 0.47; Erysipelotrichaceae (family): 0.47;
bin.026 (7,9)	0.0418	Firmicutes (phylum): 0.90; Clostridia (class): 0.88; Clostridiales (order): 0.88;
bin.041 (6,6)	0.2402	Firmicutes (phylum): 0.95; Clostridia (class): 0.91; Clostridiales (order): 0.91; Lachnospiraceae (family): 0.49; Roseburia (genus): 0.45;
bin.058 (8,15)	0.5063	Bacteroidetes (phylum): 0.89; Bacteroidia (class): 0.88; Bacteroidales (order): 0.88; Bacteroidaceae (family): 0.84; Bacteroides (genus): 0.84;
bin.065 (5,8)	0.0653	Bacteroidetes (phylum): 0.88; Bacteroidia (class): 0.88; Bacteroidales (order): 0.88; Bacteroidaceae (family): 0.85; Bacteroides (genus): 0.85;

Laurent, B. and P. Massart (2000). Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics 28*(5), 1302–1338.

Rudelson, M. and R. Vershynin (2013). Hanson-wright inequality and sub-gaussian concentration. *Electronic Communications in Probability 18*.

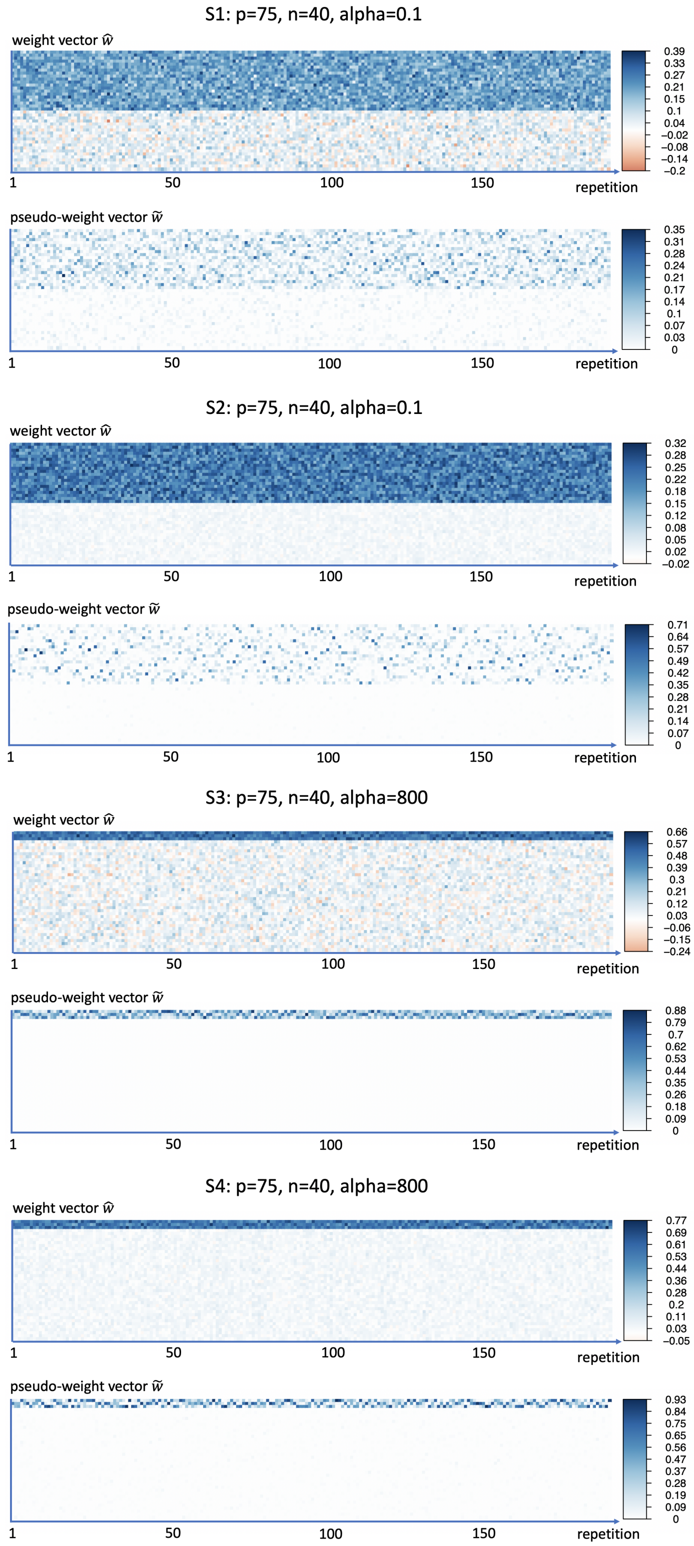

Figure 2: The graphical representation of the weight vectors $\hat{w}$ for our proposed estimator $\hat{\pi}$, and the pseudo-weight vector $\tilde{w}$ for the estimator π_{max} based on 200 simulations under four different models. Each column represents an n dimensional weight vector, and there are 200 columns in each plot.

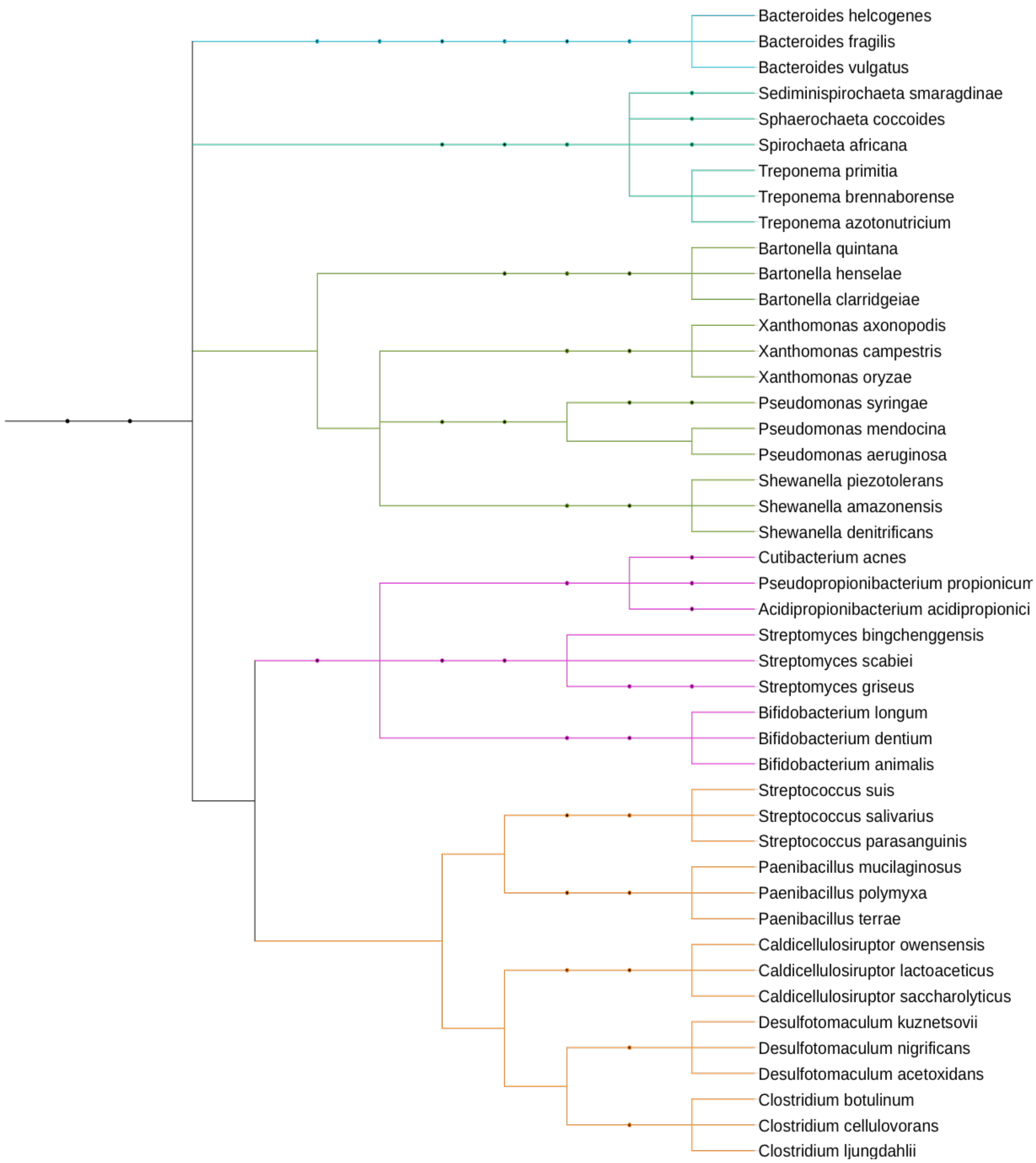

Figure 3: Taxonomic tree of 45 closely related species used in generating the shotgun metagenomic data used on in Gao and Li (2018).