Supplementary Material to "Direct covariance matrix estimation with compositional data"

Aaron J. Molstad[⋆,∗], Karl Oskar Ekvall[⋆,†], and Piotr M. Suder[‡]
Department of Statistics[⋆] and Genetics Institute[∗], University of Florida
Division of Biostatistics, Institute of Environmental Medicine,
Karolinska Institutet[†]
Department of Statistical Science, Duke University[‡]

1 Theorem proofs

1.1 Notation and key lemmas

For a tensor $\boldsymbol{\Delta} \in \mathbb{R}^{H\times p\times p}$, define the norm $|||\boldsymbol{\Delta}|||_{1,2} = \sum_{j,k} \|\boldsymbol{\Delta}_{\cdot jk}\|_2$ and for a set $\mathcal{S} \subset [p]\times[p]$ define $\boldsymbol{\Delta}_{\mathcal{S}}$ as the tensor whose (h,j,k)th entry equals $\boldsymbol{\Delta}_{(h)jk}$ if $(j,k)\in\mathcal{S}$ and zero otherwise. Similarly, $\boldsymbol{\Delta}_{\mathcal{S}^c}$ is the tensor whose (h,j,k) entry equals $\boldsymbol{\Delta}_{(h)jk}$ if $(j,k)\in\mathcal{S}^c$ and zero otherwise. Let $\boldsymbol{\Delta}^-$ be the tensor which has (h,j,j)th entry equal to zero for all $j\in[p]$ and $h\in[H]$, but is otherwise equal to $\boldsymbol{\Delta}$, and let $\boldsymbol{\Delta}^+ = \boldsymbol{\Delta} - \boldsymbol{\Delta}^-$. Define $\boldsymbol{\xi} = \max_{jk} \|\widehat{\boldsymbol{\Theta}}_{\cdot jk} - \boldsymbol{\tau}^*_{\cdot jk}\|_2$, $\boldsymbol{\eta} = \sqrt{\sum_{j=1}^p \|\sum_{k\neq j}(\widehat{\boldsymbol{\Theta}}_{\cdot jk} - \boldsymbol{\tau}^*_{\cdot jk})\|_2^2}$, and $\boldsymbol{\tau}^*_{(h)jk} = \omega^*_{(h)j} + \omega^*_{(h)k} - 2\Omega^*_{(h)jk}$. Let $\check{s} = \max_j \tilde{s}_j$ and note that, when $H=1$, $\tilde{s}_j = s_j$.

Let $\ell(\boldsymbol{\Omega}) = \sum_{h=1}^H \ell_{(h)}(\Omega_{(h)})$ where

$$\ell_{(h)}(\Omega) = \|\widehat{\Theta}_{(h)} - \omega\mathbb{1}_p^\top - \mathbb{1}_p\omega^\top + 2\Omega\|_F^2.$$

We denote the Hessian of $\mathrm{vec}(\Omega)\mapsto \ell_{(h)}(\Omega)$ by $\nabla^2\ell \in \mathbb{R}^{p^2\times p^2}$; it does not depend on Ω or h since $\ell_{(h)}$ is quadratic.

We will use the following lemmas. Proofs of the lemmas are in the subsequent section.

Lemma 1 (Quadratic lower bound)**.** *For any $a_1 > 0$ and $a_2 > 0$, it holds for every $\boldsymbol{\Delta} \in \mathbb{C}_a := \{\boldsymbol{\Delta}\in\mathbb{R}^{H\times p\times p} : \|\boldsymbol{\Delta}^-_{\mathcal{S}^c}\|_{1,2} \leq a_1\|\boldsymbol{\Delta}^-_{\mathcal{S}}\|_{1,2} + a_2, \boldsymbol{\Delta}_{(h)\cdot\cdot} = \boldsymbol{\Delta}^\top_{(h)\cdot\cdot}, h\in[H]\}$ that, when*

∗Correspondence: amolstad@ufl.edu

$p \geq 5$,

$$\sum_{h=1}^{H} \mathrm{vec}(\Delta_{(h)})^\top \nabla^2 \ell \, \mathrm{vec}(\Delta_{(h)}) \geq \left(4 - \frac{32\check{s}(1+a_1)^2}{p-4}\right) |||\boldsymbol{\Delta}^-|||_F^2 + p|||\boldsymbol{\Delta}^+|||_F^2 - \frac{32a_2^2}{p(p-4)}$$

Lemma 2. *For* $\widehat{\boldsymbol{\Delta}} = \widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}^*$*, it holds that*

$$\gamma(|||\widehat{\boldsymbol{\Delta}}_{\mathcal{S}}|||_{1,2} - |||\widehat{\boldsymbol{\Delta}}_{\mathcal{S}^c}|||_{1,2}) \geq \frac{1}{2}\sum_{h=1}^{H} \mathrm{vec}(\widehat{\Delta}_h)^\top \nabla^2 \ell \, \mathrm{vec}(\widehat{\Delta}_{(h)}) - 4\boldsymbol{\xi}|||\widehat{\boldsymbol{\Delta}}^-|||_{1,2} - 4\boldsymbol{\eta}|||\widehat{\boldsymbol{\Delta}}^+|||_F.$$

The following lemma is used only when analyzing the proposed estimator for one population, that is, when $H = 1$.

Lemma 3. *Suppose* $H = 1$ *and that the* $\log(W_{ij})$ *are independent over* i *and have sub-Gaussian norms bounded by* $K < \infty$*. Then, for* $j \neq k$*, a universal constant* $\nu > 0$*, and any* $c_1 > 0$*, if* $\lambda = \sqrt{c_1 \log(p)/n} \to 0$*, it holds for all large enough* n *and* p *that*

$$P\left(\max_{j,k} |\widehat{\Theta}_{jk} - \tau_{jk}^*| \geq \lambda\right) \leq 6p^{2-\nu c_1/K^4}.$$

When $H > 1$, the following lemma is used in place of the previous.

Lemma 4. *Under Assumption* **A4***, if* $\gamma = \{\sqrt{d_1 H L^4/n_{\min}} + \sqrt{k\log(p)/n_{\min}}\} \to 0$ *for constants* $k > 0$ *and* $d_1 > 0$ *sufficiently large, there exists constant* $d_2 > 0$ *such that*

$$P\left(\max_{l,m} \|\widehat{\boldsymbol{\Theta}}_{\cdot lm} - \boldsymbol{\tau}_{\cdot lm}^*\|_2 \geq \gamma\right) \leq p^{2-\frac{2kn_{\min}}{d_2^2 L^8 N}}$$

for N *sufficiently large.*

1.2 Proof of Theorem 1

Proof of Theorem 1. Specializing notation to the case with $H = 1$ and dropping the non-negative quadratic term, Lemma 2 implies

$$(\lambda - 4\xi)\|\widehat{\Delta}_{S^c}^-\|_1 \leq (\lambda + 4\xi)\|\widehat{\Delta}_S^-\|_1 + 4\eta\|\widehat{\Delta}^+\|_F \leq (\lambda + 4\xi)\|\widehat{\Delta}_S^-\|_1 + 4\xi p^{3/2}\|\widehat{\Delta}^+\|_F,$$

where the last inequality follows from $\eta \leq \xi p^{3/2}$. Thus, by Lemma 3, we can, for any $c_2 > 0$, pick c_1 sufficiently large so that with probability tending to one,

$$\|\widehat{\Delta}^-_{S^c}\|_1 \leq 2\|\widehat{\Delta}^-_S\|_1 + (p^{3/2}/c_2)\|\widehat{\Delta}^+\|_F.$$

Next, by Lemma 1, specializing notation to the case $H = 1$, with $a_1 = 2$ and $a_2 = (p^{3/2}/c_2)\|\widehat{\Delta}^+\|_F$,

$$\operatorname{vec}(\widehat{\Delta})^\top \nabla^2 \ell \operatorname{vec}(\widehat{\Delta}) \geq \left(4 - \frac{288\check{s}}{p-4}\right)\|\widehat{\Delta}^-\|_F^2 + p\|\widehat{\Delta}^+\|_F^2 - \frac{32p^3}{c_2^2 p(p-4)}\|\widehat{\Delta}^+\|_F^2.$$

Thus, for c_2 large enough, it holds with probability tending to one that

$$\operatorname{vec}(\widehat{\Delta})^\top \nabla^2 \ell \operatorname{vec}(\widehat{\Delta}) \geq 2\|\widehat{\Delta}^-\|_F^2 + \frac{p}{2}\|\widehat{\Delta}^+\|_F^2, \tag{1}$$

which follows from assumption **A2**. Next, by Lemma 2 and (1)

$$\begin{aligned}
0 &\geq \frac{1}{2}\operatorname{vec}(\widehat{\Delta})\nabla^2 \ell \operatorname{vec}(\widehat{\Delta}) - 4\xi\|\widehat{\Delta}^-\|_1 - 4p^{3/2}\xi\|\widehat{\Delta}^+\|_F + \lambda\|\widehat{\Delta}^-_{S^c}\|_1 - \lambda\|\widehat{\Delta}^-_S\|_1 \\
&\geq \|\widehat{\Delta}^-\|_F^2 + \frac{p}{4}\|\widehat{\Delta}^+\|_F^2 + (\lambda - 4\xi)\|\widehat{\Delta}^-_{S^c}\|_1 - (\lambda + 4\xi)\|\widehat{\Delta}^-_S\|_1 - 4p^{3/2}\xi\|\widehat{\Delta}^+\|_F \\
&\geq \|\widehat{\Delta}^-\|_F^2 + \frac{p}{4}\|\widehat{\Delta}^+\|_F^2 - (\lambda + 4\xi)\|\widehat{\Delta}^-_S\|_1 - 4p^{3/2}\xi\|\widehat{\Delta}^+\|_F \\
&\geq \|\widehat{\Delta}^-\|_F^2 + \frac{p}{4}\|\widehat{\Delta}^+\|_F^2 - (\lambda + 4\xi)\sqrt{s}\|\widehat{\Delta}^-\|_F - 4p^{3/2}\xi\|\widehat{\Delta}^+\|_F.
\end{aligned}$$

Thus, by Lemma 3, for $c_3 > 0$ and $c_4 > 0$ sufficiently large, $\lambda + 4\xi \leq c_3 \log(p)/n$, $4\xi \leq c_4 \log(p)/n$ and

$$0 \geq \|\widehat{\Delta}^-\|_F^2 - c_3\sqrt{\frac{s\log(p)}{n}}\|\widehat{\Delta}^-\|_F + \frac{p}{4}\|\widehat{\Delta}^+\|_F^2 - c_4\sqrt{\frac{p^3\log(p)}{n}}\|\widehat{\Delta}^+\|_F, \tag{2}$$

with probability tending to one, where $s = \sum_{j=1}^p s_j$.

To establish the error bound for $\|\widehat{\Delta}^+\|_F$, suppose for the sake of contradiction that $\|\widehat{\Delta}^+\|_F \geq 8c_4\sqrt{p\log(p)/n}$. Then the sum of the last two terms on the right-hand side of (2) is no smaller than $16c_4^2p^2\log(p)/n - 8c_4^2p^2\log(p)/n = 8c_4^2p^2\log(p)/n$. Additionally, by minimizing the quadratic in $\|\widehat{\Delta}^-\|_F$, one gets that the sum of the first two terms is no smaller than $-c_3^2 s\log(p)/(4n)$. Thus, since $\max_j s_j = o(p)$ and $s \leq p\max_j s_j$ the last right-hand side in (2) is positive for large enough p, so $\widehat{\Omega}$ is not a minimizer. This is the desired contradiction, so we conclude $\|\widehat{\Delta}^+\|_F < 8c_4\sqrt{p\log(p)/n}$ with probability tending to one.

To establish the main result, observe (2) implies

$$\|\widehat{\Delta}^-\|_F^2 - c_3\sqrt{\frac{s\log(p)}{n}}\left\|\widehat{\Delta}^- + \sqrt{\frac{p}{4}}\widehat{\Delta}^+\right\|_F + \frac{p}{4}\|\widehat{\Delta}^+\|_F^2 - c_4\sqrt{\frac{4p^2\log(p)}{n}}\left\|\widehat{\Delta}^- + \sqrt{\frac{p}{4}}\widehat{\Delta}^+\right\|_F \leq 0.$$

Rearranging terms,

$$\|\widehat{\Delta}^-\|_F^2 + \frac{p}{4}\|\widehat{\Delta}^+\|_F^2 \leq \left(c_3\sqrt{\frac{s\log(p)}{n}} + c_4\sqrt{\frac{4p^2\log(p)}{n}}\right)\left\|\widehat{\Delta}^- + \sqrt{\frac{p}{4}}\widehat{\Delta}^+\right\|_F,$$

so that by dividing both sides by $\|\widehat{\Delta}^- + \sqrt{\frac{p}{4}}\widehat{\Delta}^+\|_F$, we have

$$\left\|\widehat{\Delta}^- + \sqrt{\frac{p}{4}}\widehat{\Delta}^+\right\|_F \leq \left(c_3\sqrt{\frac{s\log(p)}{n}} + c_4\sqrt{\frac{4p^2\log(p)}{n}}\right). \tag{3}$$

□

2 Proof of Theorem 2

Proof of Theorem 2. In order to prove Theorem 2, we use a similar series of arguments as in the proof of Theorem 1. First, notice that Lemma 2 implies

$$(\gamma - 4\boldsymbol{\xi})|||\widehat{\boldsymbol{\Delta}}^-_{S^c}|||_{1,2} \leq (\gamma + 4\boldsymbol{\xi})|||\widehat{\boldsymbol{\Delta}}^-_S|||_{1,2} + 4\boldsymbol{\eta}|||\widehat{\boldsymbol{\Delta}}^+|||_F \leq (\gamma + 4\boldsymbol{\xi})|||\widehat{\boldsymbol{\Delta}}^-_S|||_{1,2} + 4\boldsymbol{\xi}p^{3/2}|||\widehat{\boldsymbol{\Delta}}^+|||_F,$$

where the last inequality follows from $\boldsymbol{\eta} \leq \boldsymbol{\xi}p^{3/2}$. By Lemma 4, we can, for any $c_2 > 0$, pick c_1 and k sufficiently large so that with probability tending to one,

$$|||\widehat{\boldsymbol{\Delta}}^-_{S^c}|||_{1,2} \leq 2|||\widehat{\boldsymbol{\Delta}}^-_S|||_{1,2} + (p^{3/2}/c_2)|||\widehat{\boldsymbol{\Delta}}^+|||_F.$$

Next, by Lemma 1 with $a_1 = 2$ and $a_2 = (p^{3/2}/c_2)|||\widehat{\boldsymbol{\Delta}}^+|||_F$,

$$\sum_{h=1}^H \mathrm{vec}(\widehat{\Delta}_{(h)})^\top \nabla^2\ell\,\mathrm{vec}(\widehat{\Delta}_{(h)}) \geq \left(4 - \frac{288\check{s}}{p-4}\right)\|\widehat{\boldsymbol{\Delta}}^-\|_F^2 + p\|\widehat{\boldsymbol{\Delta}}^+\|_F^2 - \frac{32p^3}{c_2^2 p(p-4)}\|\widehat{\boldsymbol{\Delta}}^+\|_F^2.$$

Thus, for c_2 large enough, it holds with probability tending to one that

$$\sum_{h=1}^{H} \mathrm{vec}(\widehat{\Delta}_{(h)})^\top \nabla^2 \ell \, \mathrm{vec}(\widehat{\Delta}_{(h)}) \geq 2 |||\widehat{\boldsymbol{\Delta}}^-|||_F^2 + \frac{p}{2} |||\widehat{\boldsymbol{\Delta}}^+|||_F^2, \tag{4}$$

which follows from assumption **A5**. Next, by Lemma 2,

$$\begin{aligned}
0 &\geq \frac{1}{2} \sum_{h=1}^{H} \mathrm{vec}(\widehat{\Delta}_{(h)})^\top \nabla^2 \ell \, \mathrm{vec}(\widehat{\Delta}_{(h)}) - 4\boldsymbol{\xi} |||\widehat{\boldsymbol{\Delta}}^-|||_{1,2} - 4p^{3/2}\boldsymbol{\xi} |||\widehat{\boldsymbol{\Delta}}^+|||_F + \gamma |||\widehat{\boldsymbol{\Delta}}^-_{S^c}|||_{1,2} - \gamma |||\widehat{\boldsymbol{\Delta}}^-_S|||_{1,2} \\
&\geq |||\widehat{\boldsymbol{\Delta}}^-|||_F^2 + \frac{p}{4} |||\widehat{\boldsymbol{\Delta}}^+|||_F^2 + (\gamma - 4\boldsymbol{\xi}) |||\widehat{\boldsymbol{\Delta}}^-_{S^c}|||_{1,2} - (\gamma + 4\boldsymbol{\xi}) |||\widehat{\boldsymbol{\Delta}}^-_S|||_{1,2} - 4p^{3/2}\boldsymbol{\xi} |||\widehat{\boldsymbol{\Delta}}^+|||_F \\
&\geq |||\widehat{\boldsymbol{\Delta}}^-|||_F^2 + \frac{p}{4} |||\widehat{\boldsymbol{\Delta}}^+|||_F^2 - (\gamma + 4\boldsymbol{\xi}) |||\widehat{\boldsymbol{\Delta}}^-_S|||_{1,2} - 4p^{3/2}\boldsymbol{\xi} |||\widehat{\boldsymbol{\Delta}}^+|||_F \\
&\geq |||\widehat{\boldsymbol{\Delta}}^-|||_F^2 + \frac{p}{4} |||\widehat{\boldsymbol{\Delta}}^+|||_F^2 - (\gamma + 4\boldsymbol{\xi}) \sqrt{\tilde{s}} |||\widehat{\boldsymbol{\Delta}}^-|||_F - 4p^{3/2}\boldsymbol{\xi} |||\widehat{\boldsymbol{\Delta}}^+|||_F
\end{aligned}$$

where $\tilde{s} = \sum_{j=1}^{p} \tilde{s}_j$. Then, with our specific choice of γ, there exists $c_5 > 0$ and $c_6 > 0$ such that $(\gamma + 4\boldsymbol{\xi}) \leq c_5[(HJ^4/n_{\min})^{1/2} + \{\log(p)/n\}^{1/2}]$ and $4\boldsymbol{\xi} \leq c_6[(HJ^4/n_{\min})^{1/2} + \{\log(p)/n\}^{1/2}]$ with probability tending to one by Lemma 4. Thus, with probability tending to one

$$\begin{aligned}
0 \geq |||\widehat{\boldsymbol{\Delta}}^-|||_F^2 &- c_5 \sqrt{\tilde{s}} \left(\sqrt{\frac{HL^4}{n_{\min}}} + \sqrt{\frac{\log(p)}{n_{\min}}} \right) |||\widehat{\boldsymbol{\Delta}}^-|||_F \\
&+ \frac{p}{4} |||\widehat{\boldsymbol{\Delta}}^+|||_F^2 - c_6 p \left(\sqrt{\frac{HL^4}{n_{\min}}} + \sqrt{\frac{\log(p)}{n_{\min}}} \right) |||\widehat{\boldsymbol{\Delta}}^+|||_F.
\end{aligned} \tag{5}$$

By an essentially identical argument as in the proof of Theorem 1, one can show this implies

$$|||\widehat{\boldsymbol{\Delta}}^+|||_F \leq 8c_6 p \left(\sqrt{\frac{HL^4}{n_{\min}}} + \sqrt{\frac{\log(p)}{n_{\min}}} \right),$$

and similarly, by the same arguments used to establish (3),

$$|||\widehat{\boldsymbol{\Delta}}^- + \sqrt{p/4}\,\widehat{\boldsymbol{\Delta}}^+|||_F^2 \leq \left(c_5 \sqrt{\tilde{s}} + c_6 p \right) \left(\sqrt{\frac{HL^4}{n_{\min}}} + \sqrt{\frac{\log(p)}{n_{\min}}} \right),$$

from which the conclusion follows. □

3 Proofs of Lemmas

Proof of Lemma 1. Inspecting the definition of $\ell_{(h)}$ shows

$$\sum_{h=1}^{H} \operatorname{vec}(\Delta_{(h)})^\top \nabla^2 \ell \operatorname{vec}(\Delta_{(h)}) = \sum_{i=1}^{p} \sum_{k \neq j} \sum_{h=1}^{H} \{\Delta_{(h)jj} + \Delta_{(h)kk} - 2\Delta_{(h)jk}\}^2.$$

Expanding the squares gives

$$\sum_{h=1}^{H} \sum_{j=1}^{p} \sum_{k \neq j} (\Delta_{(h)jj} + \Delta_{(h)kk})^2 - 4 \sum_{h=1}^{H} \sum_{j=1}^{p} \sum_{k \neq j} (\Delta_{(h)jj} + \Delta_{(h)kk}) \Delta_{(h)jk}$$
$$+ 4 \sum_{h=1}^{H} \sum_{j=1}^{p} \sum_{k \neq j} \Delta_{(h)jk}^2 = \mathrm{I} + \mathrm{II} + \mathrm{III}.$$

Note $\mathrm{III} = 4|||\boldsymbol{\Delta}^-|||_F^2$. Next, using that

$$\sum_{j=1}^{p} \sum_{k \neq j} \Delta_{(h)kk}^2 = \sum_{j=1}^{p} \sum_{k=1}^{p} \Delta_{(h)kk}^2 - \sum_{j=1}^{p} \Delta_{(h)jj}^2 = (p-1)\|\Delta_{(h)}^+\|_F^2$$

and

$$\sum_{j=1}^{p} \sum_{k \neq j} \Delta_{(h)jj} \Delta_{(h)kk} = \sum_{j=1}^{p} \left\{ \Delta_{(h)jj} \left(\sum_{k=1}^{p} \Delta_{(h)kk} - \Delta_{(h)jj} \right) \right\} = \left(\sum_{j=1}^{p} \Delta_{(h)jj} \right)^2 - \|\Delta_{(h)}^+\|_F^2.$$

we get

$$\begin{aligned}
\mathrm{I} &= \sum_{h=1}^{H} \sum_{j=1}^{p} \sum_{k \neq j} \Delta_{(h)jj}^2 + \sum_{h=1}^{H} \sum_{j=1}^{p} \sum_{k \neq j} \Delta_{(h)kk}^2 + 2 \sum_{h=1}^{H} \sum_{j=1}^{p} \sum_{k \neq j} \Delta_{(h)jj} \Delta_{(h)kk} \\
&= 2(p-2)|||\boldsymbol{\Delta}^+|||_F^2 + 2 \sum_{h=1}^{H} \left(\sum_{j=1}^{p} \Delta_{(h)jj} \right)^2 \\
&\geq 2(p-2)|||\boldsymbol{\Delta}^+|||_F^2.
\end{aligned}$$

Using this inequality and Cauchy–Schartz's gives

$$\begin{aligned}
\mathrm{I}+\mathrm{II}+\mathrm{III} \geq\ & 4|||\boldsymbol{\Delta}^-|||_F^2 + 2(p-2)|||\boldsymbol{\Delta}^+|||_F^2 - 4\sum_{h=1}^{H}\sum_{j=1}^{p}\sum_{k\neq j}\Delta_{(h)jk}\Delta_{(h)jj} \\
& - 4\sum_{h=1}^{H}\sum_{j=1}^{p}\sum_{k\neq j}\Delta_{(h)jk}\Delta_{(h)kk} \\
=\ & 4|||\boldsymbol{\Delta}^-|||_F^2 + 2(p-2)|||\boldsymbol{\Delta}^+|||_F^2 - 8\sum_{j=1}^{p}\sum_{k\neq j}\sum_{h=1}^{H}\Delta_{(h)jk}\Delta_{(h)jj} \\
\geq\ & 4|||\boldsymbol{\Delta}^-|||_F^2 + 2(p-2)|||\boldsymbol{\Delta}^+|||_F^2 - 8\sum_{j=1}^{p}\sum_{k\neq j}\|\boldsymbol{\Delta}_{\cdot jk}\|_2\|\boldsymbol{\Delta}_{\cdot jj}\|_2 \\
=\ & 4|||\boldsymbol{\Delta}^-|||_F^2 + p|||\boldsymbol{\Delta}^+|||_F^2 + \sum_{j=1}^{p}\left\{(p-4)\|\boldsymbol{\Delta}_{\cdot jj}\|_2^2 - 8\|\boldsymbol{\Delta}_{\cdot jj}\|_2\sum_{k\neq j}\|\boldsymbol{\Delta}_{\cdot jk}\|_2\right\}.
\end{aligned}$$

Here, the first equality used

$$\sum_{j=1}^{p}\sum_{k\neq j}\Delta_{(h)jk}\Delta_{(h)jj} = \sum_{j=1}^{p}\sum_{k=1}^{p}1(j\neq k)\Delta_{(h)jj}\Delta_{(h)jk} = \sum_{k=1}\sum_{j\neq k}\Delta_{(h)jj}\Delta_{(h)jk}.$$

To simplify notation, let $D_j = \sum_{k\neq j}\|\boldsymbol{\Delta}_{\cdot jk}\|_2$ be the sum of the Euclidean norms of the off-diagonal fibers in the jth row. Completing the square in the jth summand of the last lower bound of $\mathrm{I}+\mathrm{II}+\mathrm{III}$ gives

$$\left(\sqrt{(p-4)}\|\boldsymbol{\Delta}_{\cdot jj}\| - \sqrt{\frac{16}{p-4}D_j}\right)^2 - \frac{16}{p-4}D_j^2,$$

and hence

$$\mathrm{I}+\mathrm{II}+\mathrm{III} \geq 4|||\boldsymbol{\Delta}^-|||_F^2 + p|||\boldsymbol{\Delta}^+|||_F^2 - \frac{16}{p-4}\sum_{j=1}^{p}D_j^2.$$

To bound $\sum_{j=1}^{p}D_j^2$, fix an arbitrary $a_3>0$ and consider the optimization problem

$$\max_{\boldsymbol{\Delta}\in\mathbb{R}^{H\times p\times p}}\sum_{j=1}^{p}D_j^2 \quad \text{s.t.} \quad |||\boldsymbol{\Delta}^-_{\mathcal{S}^c}|||_{1,2} \leq a_1|||\boldsymbol{\Delta}^-_{\mathcal{S}}|||_{1,2} + a_2, \quad \Delta_{(h)} = \Delta_{(h)}^\top, \quad |||\boldsymbol{\Delta}|||_F^2 \leq a_3.$$

The maximum becomes no smaller if we drop the symmetry constraint and add a constraint that all elements be positive. Moreover, we may assume $H=1$ since only the Euclidean norms of fibers affect the objective function and constraints. Thus, $\boldsymbol{\Delta} = \Delta$ and $D_j = \|\Delta^j\|_1$, where Δ^j is Δ with all elements but the off-diagonal ones in the jth row set to zero.

We get the concave optimization problem

$$\max_{\Delta \in \mathbb{R}^{p\times p}} \sum_{j=1}^{p} \|\Delta^j\|_1^2 \quad \text{s.t.} \quad \sum_{j=1}^{p}\sum_{k=1}^{p} (\Delta^-_{\mathcal{S}^c})_{jk} \le a_1 \sum_{j=1}^{p}\sum_{k=1}^{p} (\Delta^-_{\mathcal{S}})_{jk} + a_2, \quad \|\Delta\|_F^2 \le a_3, \quad \Delta_{jk} \ge 0. \tag{6}$$

Note that, except for the definition of $\mathcal{S}$, the order of elements within rows does not matter in (6). Thus, we may, without affecting the maximum, assume $\mathcal{S} = \{(j,k) \in [p]\times[p] : k \le s_j\}$ if we also redefine Δ^j to be Δ with all elements but the first $p-1$ in the jth row set to zero, Δ^- to be Δ with the last column set to zero, and $\Delta^+ = \Delta - \Delta^-$. Effectively, this interchanges the diagonal element in each row with the last element in that row and puts the support indicated by $\mathcal{S}$ on the first s_j elements in the jth row. Now, the feasible set becomes no smaller if $s_j = \check{s}$ for each j, so we can assume this without decreasing the maximum.

With these definitions, pick a feasible Δ and note any Δ' obtained by permuting the rows of Δ is feasible and attains the same value. Thus, by concavity, the convex combination which puts equal weight $1/p!$ on all $p!$ row-permutations of Δ, say $\tilde{\Delta}$, is feasible and attains at least the same value as Δ. By construction, every row of $\tilde{\Delta}$ is the same. Thus, we have shown that for any feasible Δ there is a feasible $\tilde{\Delta}$ that attains at least the same value and has all rows equal.

Pick an arbitrary feasible Δ and a corresponding $\tilde{\Delta}$. Since all rows of $\tilde{\Delta}$ are the same, the constraint $\|\tilde{\Delta}^-_{\mathcal{S}^c}\|_1 \le a_1\|\tilde{\Delta}^-_{\mathcal{S}}\|_1 + a_2$ implies $\|\tilde{\Delta}^j_{\mathcal{S}^c}\|_1 \le a_1\|\tilde{\Delta}^j_{\mathcal{S}}\|_1 + a_2/p$ for every $j \in [p]$. Thus,

$$\sum_{j=1}^{p} \|\tilde{\Delta}^j\|_1^2 \le \sum_{j=1}^{p} \{(1+a_1)\|\tilde{\Delta}^j_{\mathcal{S}}\|_1 + a_2/p\}^2 \le \sum_{j=1}^{p} \{(1+a_1)\sqrt{\check{s}}\|\tilde{\Delta}^j_{\mathcal{S}}\|_F + a_2/p\}^2.$$

Moreover,

$$\{(1+a_1)\sqrt{\check{s}}\|\tilde{\Delta}^j_{\mathcal{S}}\|_F + a_2/p\}^2 \le 2(1+a_1)^2\check{s}\|\tilde{\Delta}^j\|_F^2 + 2a_2^2/p^2$$

and hence

$$\sum_{j=1}^{p} \|\Delta^j\|_1^2 \le \sum_{j=1}^{p} \|\tilde{\Delta}^j\|_1^2 \le \sum_{j=1}^{p} \{2(1+a_1)^2\check{s}\|\tilde{\Delta}^j\|_F^2 + 2a_2^2/p^2\} = 2(1+a_1)^2\check{s}\|\Delta^-\|_F^2 + 2a_2^2/p.$$

Thus,

$$\mathrm{I} + \mathrm{II} + \mathrm{III} \geq 4\|\Delta^-\|_F^2 + p\|\Delta^+\|_F^2 - \frac{16}{p-4}\{2(1+a_1)^2\check{s}\|\Delta^-\|_F^2 + 2a_2^2/p\}$$
$$= \left(4 - \frac{32\check{s}(1+a_1)^2}{p-4}\right)\|\Delta^-\|_F^2 + p\|\Delta^+\|_F^2 - \frac{32a_2^2}{p(p-4)},$$

which completes the proof. □

Proof of Lemma 2. By definition,

$$\ell(\widehat{\mathbf{\Omega}}) + \gamma|||\widehat{\mathbf{\Omega}}|||_{1,2} \leq \ell(\mathbf{\Omega}^*) + \gamma|||\mathbf{\Omega}^*|||_{1,2},$$

and hence

$$\ell(\widehat{\mathbf{\Omega}}) - \ell(\mathbf{\Omega}^*) \leq \gamma\left(|||\widehat{\mathbf{\Omega}}|||_{1,2} - |||\mathbf{\Omega}^*|||_{1,2}\right) \leq \gamma\left(|||\widehat{\mathbf{\Delta}}_{\mathcal{S}}|||_{1,2} - |||\widehat{\mathbf{\Delta}}_{\mathcal{S}^c}|||_{1,2}\right),$$

where the last line uses $\mathbf{\Omega}^*_{\mathcal{S}} = \mathbf{\Omega}^*$ and $\mathbf{\Omega}^*_{\mathcal{S}^c} = 0$. Additionally, for any $\mathbf{\Omega}$ and $\mathbf{\Delta} = \mathbf{\Omega} - \mathbf{\Omega}^*$,

$$\begin{aligned}
&\ell(\mathbf{\Omega}) - \ell(\mathbf{\Omega}^*) \\
&\quad= \sum_{h=1}^{H}\left\{\frac{1}{2}\operatorname{vec}(\Delta_{(h)})^\top\nabla^2\ell\operatorname{vec}(\Delta_{(h)}) + 4\sum_{j=1}^{p}\sum_{k\neq j}(\widehat{\Theta}_{(h)jk} - \tau^*_{(h)jk})(\mathbf{\Delta}_{(h)jk} - \mathbf{\Delta}_{(h)jj})\right\} \\
&\quad= \frac{1}{2}\sum_{h=1}^{H}\operatorname{vec}(\Delta_{(h)})^\top\nabla^2\ell\operatorname{vec}(\Delta_{(h)}) + 4\sum_{h=1}^{H}\sum_{j=1}^{p}\sum_{k\neq j}(\widehat{\Theta}_{(h)jk} - \tau^*_{(h)jk})\mathbf{\Delta}_{(h)jk} \\
&\qquad\quad - 4\sum_{h=1}^{H}\sum_{j=1}^{p}\mathbf{\Delta}_{(h)jj}\sum_{k\neq j}(\widehat{\Theta}_{(h)jk} - \tau^*_{(h)jk}) \\
&\quad= \frac{1}{2}\sum_{h=1}^{H}\operatorname{vec}(\Delta_{(h)})^\top\nabla^2\ell\operatorname{vec}(\Delta_{(h)}) + 4\sum_{j=1}^{p}\sum_{k\neq j}(\widehat{\mathbf{\Theta}}_{\cdot jk} - \tau^*_{\cdot jk})^\top\mathbf{\Delta}_{\cdot jk} - 4\sum_{j=1}^{p}\mathbf{\Delta}_{\cdot jj}^\top\left\{\sum_{k\neq j}(\widehat{\mathbf{\Theta}}_{\cdot jk} - \tau^*_{\cdot jk})\right\}.
\end{aligned}$$

Thus, using that $|(\widehat{\mathbf{\Theta}}_{\cdot jk} - \tau^*_{\cdot jk})^\top\mathbf{\Delta}_{\cdot jk}| \leq \|\widehat{\mathbf{\Theta}}_{\cdot jk} - \tau^*_{\cdot jk}\|_2\|\mathbf{\Delta}_{\cdot jk}\|_2 \leq \boldsymbol{\xi}\|\mathbf{\Delta}_{\cdot jk}\|_2$ and $|\mathbf{\Delta}_{\cdot jj}^\top\{\sum_{k\neq j}(\widehat{\mathbf{\Theta}}_{\cdot jk} - \tau^*_{\cdot jk})\}| \leq \|\mathbf{\Delta}_{\cdot jj}\|_2\|\sum_{k\neq j}(\widehat{\mathbf{\Theta}}_{\cdot jk} - \tau^*_{\cdot jk})\|_2$, we get

$$\ell(\mathbf{\Omega}) - \ell(\mathbf{\Omega}^*) \geq \frac{1}{2}\sum_{h=1}^{H}\operatorname{vec}(\Delta_{(h)})^\top\nabla^2\ell\operatorname{vec}(\Delta_{(h)}) - 4\boldsymbol{\xi}|||\mathbf{\Delta}^-|||_{1,2} - 4\boldsymbol{\eta}\|\mathbf{\Delta}^+\|_F.$$

Combining the upper and lower bounds of $\ell(\widehat{\mathbf{\Omega}}) - \ell(\mathbf{\Omega}^*)$ gives

$$\gamma\left(|||\widehat{\mathbf{\Delta}}_{\mathcal{S}}|||_{1,2} - |||\widehat{\mathbf{\Delta}}_{\mathcal{S}^c}|||_{1,2}\right) \geq \frac{1}{2}\sum_{h=1}^{H} \mathrm{vec}(\widehat{\Delta}_{(h)})^\top \nabla^2 \ell \, \mathrm{vec}(\widehat{\Delta}_{(h)}) - 4\boldsymbol{\xi}|||\widehat{\mathbf{\Delta}}^-|||_{1,2} - 4\boldsymbol{\eta}\|\widehat{\mathbf{\Delta}}^+\|_F,$$

which completes the proof. □

We will use the following to prove Lemma 3.

Lemma 5. *If $H = 1$ and the $\log(W_{ij})$ are independent over i and have sub-Gaussian norms bounded by $K < \infty$, then for $j \neq k$, for a universal constant $\nu > 0$,*

$$P\left(|\widehat{\Theta}_{jk} - \tau^*_{jk}| \geq \epsilon\right) \leq 6\exp\{-\nu n \min(\epsilon/K^2, \epsilon^2/K^4)\}.$$

Proof of Lemma 5. Suppose first $\widehat{\Theta}_{jk} = n^{-1}\sum_{i=1}^n \log(X_{ij}/X_{ik})^2$ and note this makes sense since $\mathbb{E}\{\log(X_{ij}/X_{ik})\} = \mathbb{E}\{\log(W_{ij}) - \log(W_{ik})\} = 0$. Now

$$n^{-1}\sum_{i=1}^n \log(X_{ij}/X_{ik})^2 = n^{-1}\sum_{i=1}^n \{\log(W_{ij})^2 + \log(W_{ik})^2 - 2\log(W_{ij})\log(W_{ik})\}.$$

We thus have, since $\mathbb{E}\{\log(W_{ij})\} = 0$,

$$P\left(|\widehat{\Theta}_{jk} - \omega^*_j - \omega^*_k + 2\Omega^*_{jk}| \geq \epsilon\right) \leq P\left(|\sum_{i=1}^n \{\log(W_{ij})^2 - \omega^*_j\}| \geq n\epsilon/3\right)$$
$$+ P\left(|\sum_{i=1}^n \{\log(W_{ik})^2 - \omega^*_k\}| \geq n\epsilon/3\right) + P\left(|\sum_{i=1}^n \{\log(W_{ij})\log(W_{ik}) - \Omega^*_{jk}\}| \geq n\epsilon/6\right).$$

Each of these terms enjoys sub-exponential concentration, meaning they are upper bounded by $2\exp\{-c\min(\epsilon/K^2, \epsilon^2/K^4)n\}$, where K^2 is the sub-Exponential norm bound of the $\log(W_{ij})^2$ and $\nu > 0$ a universal constant (Vershynin, 2018, Lemma 2.7.5 and Corollary 2.8.4).

Now, with $\widehat{\Theta}_{jk} = n^{-1}\sum_{i=1}^n \{\log(X_{ij}/X_{ik}) - n^{-1}\sum_{l=1}^n \log(X_{lj}/X_{lk})\}^2$, which is equal to $n^{-1}\sum_{i=1}\log(X_{ij}/X_{ik})^2 - \{n^{-1}\sum_{i=1}\log(X_{ij}/X_{ik})\}^2$, routine arguments for the concentration of the sample mean of sub-Gaussian random variables show $\{n^{-1}\sum_{i=1}\log(X_{ij}/X_{ik})\}^2$ is of smaller order than $n^{-1}\sum_{i=1}\log(X_{ij}/X_{ik})^2 - \tau^*_{jk}$ and so essentially the same proof applies; we omit the details for brevity.

□

Proof of Lemma 3. By Lemma 5 and a union bound,

$$P\left(\max_{j,k}|\widehat{\Theta}_{jk}-\tau^*_{jk}|\geq\epsilon\right)\leq 6p^2\exp(-\nu n\epsilon^2/K^4)=6\exp\{(2-\nu c_1/K^4)\log(p)\},$$

which completes the proof. $\square$

Proof of Lemma 2. By definition,

$$\ell(\widehat{\boldsymbol{\Omega}})+\gamma|||\widehat{\boldsymbol{\Omega}}|||_{1,2}\leq\ell(\boldsymbol{\Omega}^*)+\gamma|||\boldsymbol{\Omega}^*|||_{1,2},$$

and hence

$$\ell(\widehat{\boldsymbol{\Omega}})-\ell(\boldsymbol{\Omega}^*)\leq\gamma\left(|||\widehat{\boldsymbol{\Omega}}|||_{1,2}-|||\boldsymbol{\Omega}^*|||_{1,2}\right)\leq\gamma\left(|||\widehat{\boldsymbol{\Delta}}_{\mathcal{S}}|||_{1,2}-|||\widehat{\boldsymbol{\Delta}}_{\mathcal{S}^c}|||_{1,2}\right),$$

where the last line uses $\boldsymbol{\Omega}^*_{\mathcal{S}}=\boldsymbol{\Omega}^*$ and $\boldsymbol{\Omega}^*_{\mathcal{S}^c}=0$. On the other hand, by the triangle inequality, for any $\boldsymbol{\Omega}$ and $\boldsymbol{\Delta}=\boldsymbol{\Omega}-\boldsymbol{\Omega}^*$,

$$\begin{aligned}
&\ell(\boldsymbol{\Omega})-\ell(\boldsymbol{\Omega}^*)\\
&\quad=\sum_{h=1}^{H}\left\{\frac{1}{2}\operatorname{vec}(\Delta_{(h)})^\top\nabla^2\ell\operatorname{vec}(\Delta_{(h)})+4\sum_{j=1}^{p}\sum_{k\neq j}(\widehat{\Theta}_{(h)jk}-\tau^*_{(h)jk})(\boldsymbol{\Delta}_{(h)jk}-\boldsymbol{\Delta}_{(h)jj})\right\}\\
&\quad=\frac{1}{2}\sum_{h=1}^{H}\operatorname{vec}(\Delta_{(h)})^\top\nabla^2\ell\operatorname{vec}(\Delta_{(h)})+4\sum_{h=1}^{H}\sum_{j=1}^{p}\sum_{k\neq j}(\widehat{\Theta}_{(h)jk}-\tau^*_{(h)jk})\boldsymbol{\Delta}_{(h)jk}\\
&\qquad\quad-4\sum_{h=1}^{H}\sum_{j=1}^{p}\boldsymbol{\Delta}_{(h)jj}\sum_{k\neq j}(\widehat{\Theta}_{(h)jk}-\tau^*_{(h)jk})\\
&\quad=\frac{1}{2}\sum_{h=1}^{H}\operatorname{vec}(\Delta_{(h)})^\top\nabla^2\ell\operatorname{vec}(\Delta_{(h)})+4\sum_{j=1}^{p}\sum_{k\neq j}(\widehat{\boldsymbol{\Theta}}_{\cdot jk}-\tau^*_{\cdot jk})^\top\boldsymbol{\Delta}_{\cdot jk}-4\sum_{j=1}^{p}\boldsymbol{\Delta}^\top_{\cdot jj}\left\{\sum_{k\neq j}(\widehat{\boldsymbol{\Theta}}_{\cdot jk}-\tau^*_{\cdot jk})\right\}\\
&\quad\geq\frac{1}{2}\sum_{h=1}^{H}\operatorname{vec}(\Delta_{(h)})^\top\nabla^2\ell\operatorname{vec}(\Delta_{(h)})-4\boldsymbol{\xi}|||\boldsymbol{\Delta}^-|||_{1,2}-4\boldsymbol{\eta}\|\boldsymbol{\Delta}^+\|_F.
\end{aligned}$$

$\square$

Proof of Lemma 4. First, recall that we assume $\log(W_{(h)ij})\in[-L,L]$. Thus, defining $v_{(h)ilm}=\log(W_{(h)il}/W_{(h)im})$, it can be easily verified that $v^2_{(h)ilm}\leq 4L^2$.

To establish the desired concentration inequality, we will use McDiarmid's inequality (Vershynin, 2018, Theorem 2.9.1) to get a high-probability bound for $\|\widehat{\boldsymbol{\Theta}}_{\cdot lm}-\boldsymbol{\tau}^*_{\cdot lm}\|_2$ for an arbitrary pair (l,m) with $l \neq m$, and then the union bound to control $\max_{l,m} \|\widehat{\boldsymbol{\Theta}}_{\cdot lm} - \boldsymbol{\tau}^*_{\cdot lm}\|_2$.

Let $\widehat{\boldsymbol{\Theta}}$ be the tensor of sample variation matrices based on $\{W_{(h)1}, \ldots, W_{(h)n_{(h)}}\}_{h=1}^H$ and let $\widetilde{\boldsymbol{\Theta}}$ be the tensor of sample variation matrices based on $\{\widetilde{W}_{(h)1}, \ldots, \widetilde{W}_{(h)n_{(h)}}\}_{h=1}^H$ where $\widetilde{W}_{(h)i} = W_{(h)i}$ for all but a single pair $(h^\star, i^\star)$, i.e., $\widetilde{W}_{(h^\star)i^\star} \neq W_{(h^\star)i^\star}$. To apply McDiarmid's inequality, we need to find a $c_{h^\star,i^\star}$ such that

$$\big|\|\widehat{\boldsymbol{\Theta}}_{\cdot lm} - \boldsymbol{\tau}^*_{\cdot lm}\|_2 - \|\widetilde{\boldsymbol{\Theta}}_{\cdot lm} - \boldsymbol{\tau}^*_{\cdot lm}\|_2\big| \leq \|\widehat{\boldsymbol{\Theta}}_{\cdot lm} - \widetilde{\boldsymbol{\Theta}}_{\cdot lm}\|_2 \leq c_{h^\star,i^\star}$$

for each pair $(h^\star, i^\star)$. Notice that bounding $\|\widehat{\boldsymbol{\Theta}}_{\cdot lm} - \widetilde{\boldsymbol{\Theta}}_{\cdot lm}\|_2^2$ is a matter of bounding

$$\begin{aligned}\|\widehat{\boldsymbol{\Theta}}_{\cdot lm} - \widetilde{\boldsymbol{\Theta}}_{\cdot lm}\|_2^2 = \sum_{h=1}^H \Bigg(\frac{1}{n_{(h)}} \sum_{i=1}^{n_{(h)}} \Bigg[\left\{ \log\left(\frac{W_{(h)il}}{W_{(h)im}}\right) - n_{(h)}^{-1} \sum_{j=1}^{n_{(h)}} \log\left(\frac{W_{(h)jl}}{W_{(h)jm}}\right) \right\}^2 \\ - \left\{ \log\left(\frac{\widetilde{W}_{(h)il}}{\widetilde{W}_{(h)im}}\right) - n_{(h)}^{-1} \sum_{j=1}^{n_{(h)}} \log\left(\frac{\widetilde{W}_{(h)jl}}{\widetilde{W}_{(h)jm}}\right) \right\}^2 \Bigg] \Bigg)^2.\end{aligned}$$

Now, using that by definition $v_{(h)ilm} \log\left(\frac{W_{(h)il}}{W_{(h)im}}\right)$, and since $v_{(h)ilm} = \tilde{v}_{(h)ilm}$ for all $i \in [n_{(h)}]$ when $h \neq h^\star$, we have that

$$\begin{aligned}\|\widehat{\boldsymbol{\Theta}}_{\cdot lm} - \widetilde{\boldsymbol{\Theta}}_{\cdot lm}\|_2^2 &= \Bigg(\frac{1}{n_{(h^\star)}} \sum_{i=1}^{n_{(h^\star)}} \Bigg[\left\{ v_{(h^\star)ilm} - n_{(h^\star)}^{-1} \sum_{j=1}^{n_{(h^\star)}} v_{(h^\star)jlm} \right\}^2 \\ &\qquad - \left\{ \tilde{v}_{(h^\star)ilm} - n_{(h^\star)}^{-1} \sum_{j=1}^{n_{(h^\star)}} v_{(h^\star)jlm} - n_{(h^\star)}^{-1} \tilde{v}_{(h^\star)i^\star lm} + n_{(h^\star)}^{-1} v_{(h^\star)i^\star lm} \right\}^2 \Bigg] \Bigg)^2 \\ &= \left[\frac{1}{n_{(h^\star)}} \sum_{i=1}^{n_{(h^\star)}} \left\{ a_i^2 - (\tilde{a}_i + b)^2 \right\} \right]^2 = \left\{ \frac{1}{n_{(h^\star)}} \sum_{i=1}^{n_{(h^\star)}} \left(a_i^2 - \tilde{a}_i^2 - 2\tilde{a}_i b - b^2 \right) \right\}^2\end{aligned}$$

where $a_i = v_{(h^\star)ilm} - n_{(h^\star)}^{-1} \sum_{j=1}^{n_{(h^\star)}} v_{(h^\star)jlm}$, $\tilde{a}_i = \tilde{v}_{(h^\star)ilm} - n_{(h^\star)}^{-1} \sum_{j=1}^{n_{(h^\star)}} v_{(h^\star)jlm}$, and $b = n_{(h^\star)}^{-1} v_{(h^\star)i^\star lm} - n_{(h^\star)}^{-1} \tilde{v}_{(h^\star)i^\star lm}$. Because $a_i = \tilde{a}_i$ for all $i \neq i^\star$, there exists constant $d_1 > 0$ such that

$$= \left\{ \frac{(a_{i^\star}^2 - \tilde{a}_{i^\star}^2)}{n_{(h^\star)}} - \frac{1}{n_{(h^\star)}} \sum_{i=1}^{n_{(h^\star)}} \left(2\tilde{a}_i b + b^2\right) \right\}^2 = \left\{ \frac{(a_{i^\star}^2 - \tilde{a}_{i^\star}^2)}{n_{(h^\star)}} + \frac{2b}{n_{(h^\star)}} \underbrace{(v_{(h^\star)i^\star lm} - \tilde{v}_{(h^\star)i^\star lm})}_{=-\sum_{i=1}^{n_{(h^\star)}} \tilde{a}_i = b n_{(h^\star)}} - b^2 \right\}^2$$

$$= \left\{ \frac{(a_{i^\star}^2 - \tilde{a}_{i^\star}^2)}{n_{(h^\star)}} + b^2 \right\}^2 \leq \frac{4a_{i^\star}^4 + 4\tilde{a}_{i^\star}^4}{n_{(h^\star)}^2} + 2b^4 \leq \frac{d_1^2 L^4}{n_{(h^\star)}^2} \leq \frac{d_1^2 L^4}{n_{\min}^2}.$$

The second to last inequality above follows from the fact that

$$a_i^2 = \left(v_{(h^\star)ilm} - n_{(h^\star)}^{-1} \sum_{j=1}^{n_{(h^\star)}} v_{(h^\star)jlm)} \right)^2 \leq 2v_{(h^\star)ilm}^2 + 2\left(\frac{1}{n_{(h^\star)}} \sum_{j=1}^{n_{(h^\star)}} v_{(h^\star)jlm} \right)^2 \leq 8L^2 + 8L^2 = 16L^2,$$

and similarly for $\tilde{a}_i^2$, and also that

$$b^2 = \frac{1}{n_{(h^\star)}^2} (v_{(h^\star)i^\star lm} - \tilde{v}_{(h^\star)i^\star lm})^2 \leq \frac{2}{n_{(h^\star)}^2} (v_{(h^\star)i^\star lm}^2 + \tilde{v}_{(h^\star)i^\star lm}^2) \leq \frac{16L^2}{n_{(h^\star)}^2}.$$

Hence, we have shown there exists a finite $d_1 > 0$ such that

$$\|\widehat{\boldsymbol{\Theta}}_{\cdot lm} - \widetilde{\boldsymbol{\Theta}}_{\cdot lm}\|_2 \leq \frac{d_1 L^2}{n_{\min}}$$

for all pairs $(h^\star, i^\star)$. Thus applying McDiarmid's inequality, we have

$$P(\|\widehat{\boldsymbol{\Theta}}_{\cdot lm} - \boldsymbol{\tau}_{\cdot lm}^*\|_2 \geq \delta + \mathbb{E}\|\widehat{\boldsymbol{\Theta}}_{\cdot lm} - \boldsymbol{\tau}_{\cdot lm}^*\|_2) \leq \exp\left(-\frac{2\delta^2}{d_1^2 \sum_{h=1}^{H} \sum_{i=1}^{n_{(h)}} \frac{L^4}{n_{\min}^2}} \right) \leq \exp\left(-\frac{2n_{\min}^2 \delta^2}{d_1^2 L^4 N} \right).$$

All that remains is to bound $\mathbb{E}\|\widehat{\boldsymbol{\Theta}}_{\cdot lm} - \boldsymbol{\tau}^*_{\cdot lm}\|_2$. First applying Jensen's inequality,

$$\begin{aligned}
&\mathbb{E}\|\widehat{\boldsymbol{\Theta}}_{\cdot lm} - \boldsymbol{\tau}^*_{\cdot lm}\|_2 \\
&\leq \sqrt{\sum_{h=1}^{H} \mathbb{E}\left[n_{(h)}^{-1}\sum_{i=1}^{n_{(h)}}\left\{\left(v_{(h)ilm} - n_{(h)}^{-1}\sum_{j=1}^{n_{(h)}} v_{(h)jlm}\right)^2 - \boldsymbol{\tau}^*_{\cdot lm}\right\}\right]^2} \\
&= \sqrt{\sum_{h=1}^{H} \mathbb{E}\left[n_{(h)}^{-1}\sum_{i=1}^{n_{(h)}}\left\{\left(\underbrace{v_{(h)ilm} - \mathbb{E}(v_{(h)ilm})}_{\hat{v}_{(h)ilm}} - n_{(h)}^{-1}\sum_{j=1}^{n_{(h)}}\underbrace{\{v_{(h)jlm} - \mathbb{E}(v_{(h)ilm})\}}_{\hat{v}_{(h)ilm}}\right)^2 - \boldsymbol{\tau}^*_{\cdot lm}\right\}\right]^2} \\
&= \sqrt{\sum_{h=1}^{H} \mathbb{E}\left[n_{(h)}^{-1}\sum_{i=1}^{n_{(h)}}(\hat{v}^2_{(h)ilm} - \boldsymbol{\tau}^*_{\cdot lm}) - \left(n_{(h)}^{-1}\sum_{i=1}^{n_{(h)}}\hat{v}_{(h)ilm}\right)^2\right]^2} \\
&\leq \sqrt{2\sum_{h=1}^{H}\underbrace{\mathbb{E}\left\{n_{(h)}^{-1}\sum_{i=1}^{n_{(h)}}(\hat{v}^2_{(h)ilm} - \boldsymbol{\tau}^*_{\cdot lm})\right\}^2}_{A_1} + 2\sum_{h=1}^{H}\underbrace{\mathbb{E}\left(n_{(h)}^{-1}\sum_{i=1}^{n_{(h)}}\hat{v}_{(h)ilm}\right)^4}_{A_2}}.
\end{aligned}$$

Next, we bound A_1 and A_2 separately. Notice that because $v_{(h)ilm} \in [-2L, 2L]$, with $\hat{v}_{(h)ilm} = v_{(h)ilm} - \mathbb{E}(v_{(h)ilm})$ it follows that $\hat{v}_{(h)ilm} \in [-4L, 4L]$. Thus, for A_1,

$$A_1 = \mathbb{E}\left\{n_{(h)}^{-1}\sum_{i=1}^{n_{(h)}}(\hat{v}^2_{(h)ilm} - \boldsymbol{\tau}^*_{\cdot lm})\right\}^2 = \mathbb{V}\left\{n_{(h)}^{-1}\sum_{i=1}^{n_{(h)}}(\hat{v}^2_{(h)ilm} - \boldsymbol{\tau}^*_{\cdot lm})\right\} = \sum_{i=1}^{n_{(h)}}\frac{\mathbb{V}\left(\hat{v}^2_{(h)ilm}\right)}{n^2_{(h)}} \leq \frac{16^2L^4}{4n_{(h)}}$$

because for bounded random variable $Y \in [a, b]$, $\mathbb{V}(Y) \leq \frac{1}{4}(b-a)^2$. For A_2, notice that

$$\begin{aligned}
A_2 = \mathbb{E}\left(n_{(h)}^{-1}\sum_{i=1}^{n_{(h)}}\hat{v}_{(h)ilm}\right)^4 &= 256L^4\mathbb{E}\left(n_{(h)}^{-1}\sum_{i=1}^{n_{(h)}}\underbrace{(\hat{v}_{(h)ilm}/4L)}_{u_{(h)ilm}}\right)^4 \\
&\leq 256L^4\mathbb{E}\left(n_{(h)}^{-1}\sum_{i=1}^{n_{(h)}}u_{(h)ilm}\right)^2 \qquad \text{(since } u_{(h)il} \in [-1, 1] \text{ by construction)} \\
&= \frac{256L^4}{n^2_{(h)}}\sum_{i=1}^{n_{(h)}}\mathbb{V}(u_{(h)ilm}) \qquad \text{(by } \mathbb{E}(u_{(h)ilm}) = 0 \text{ and independence)} \\
&\leq \frac{256L^4}{n_{(h)}}.
\end{aligned}$$

Putting the pieces together, we have shown that there exists some constant $0 < d_2 < \infty$ such that

$$\mathbb{E}\|\widehat{\boldsymbol{\Theta}}_{\cdot lm} - \boldsymbol{\tau}^*_{\cdot lm}\|_2 \leq \sqrt{\frac{d_2 H L^4}{n_{\min}}}.$$

Thus, we conclude that

$$P\left(\|\widehat{\boldsymbol{\Theta}}_{\cdot lm} - \boldsymbol{\tau}^*_{\cdot lm}\|_2 \geq \delta + \sqrt{\frac{d_2 H L^4}{n_{\min}}}\right) \leq \exp\left(-\frac{2n_{\min}^2 \delta^2}{d_1^2 L^4 N}\right).$$

By the union bound, taking $\delta = \sqrt{k \log(p)/n}$,

$$\begin{aligned}
&P\left(\max_{l,m} \|\widehat{\boldsymbol{\Theta}}_{\cdot lm} - \boldsymbol{\tau}^*_{\cdot lm}\|_2 \geq \sqrt{\frac{d_2 H L^4}{n_{\min}}} + \sqrt{\frac{k \log(p)}{n_{\min}}}\right) \\
&\quad \leq p^2 \; P\left(\|\widehat{\boldsymbol{\Theta}}_{\cdot lm} - \boldsymbol{\tau}^*_{\cdot lm}\|_2 \leq \sqrt{\frac{d_2 H L^4}{n_{\min}}} + \sqrt{\frac{k \log(p)}{n_{\min}}}\right) \\
&\quad \leq p^2 \; \exp\left(-\frac{2n_{\min}^2}{d_1^2 L^4 N} \frac{k \log(p)}{n_{\min}}\right) = p^2 \; \exp\left(-\frac{2n_{\min} k \log(p)}{d_1^2 L^4 N}\right) = p^{2 - \frac{2kn_{\min}}{d_1^2 L^4 N}},
\end{aligned}$$

which can be made arbitrarily close to zero by taking $k > 0$ sufficiently large.

$\square$

4 Additional simulation study results

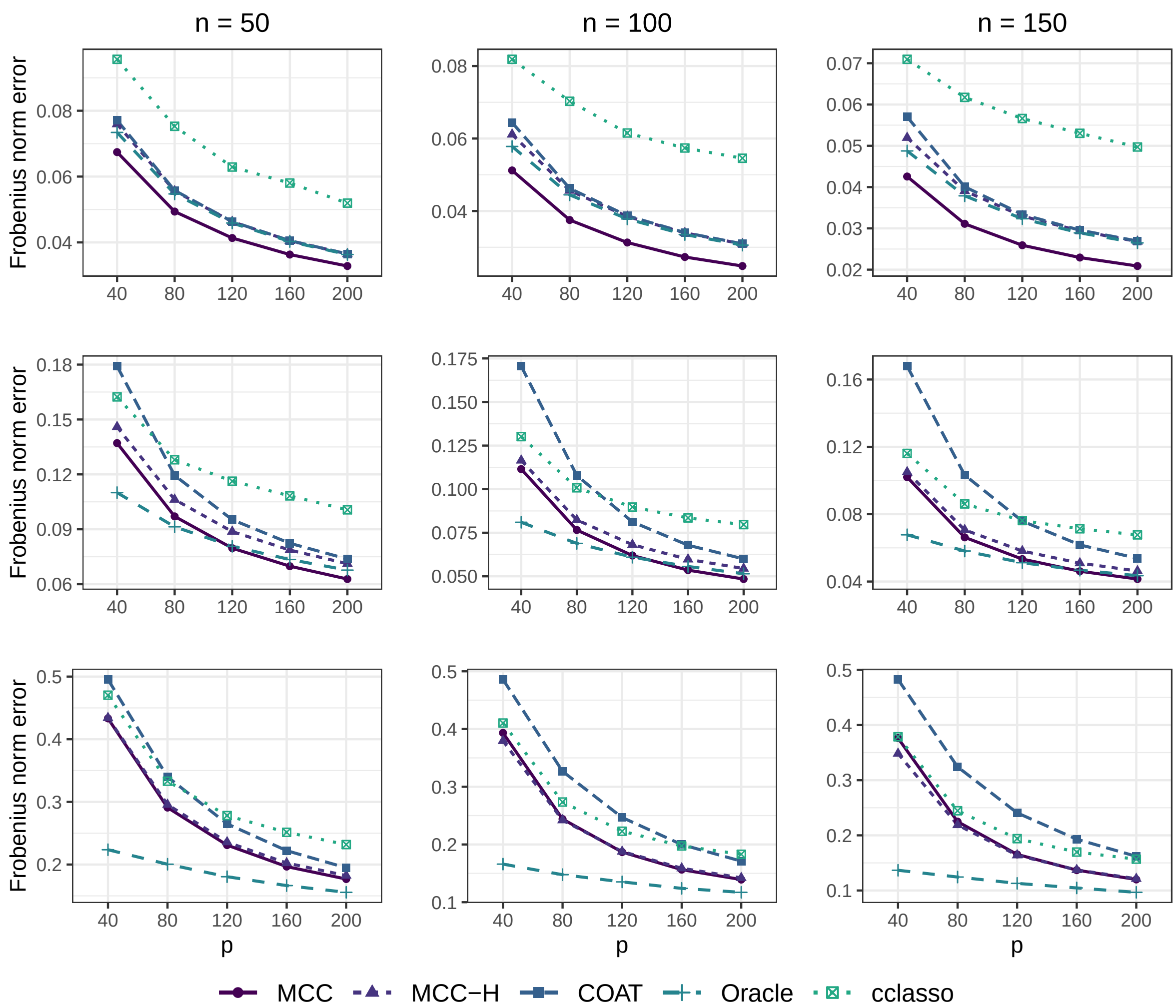

Figure 1: Average Frobenius norm error on the covariance scale (divided by p) over 50 independent replications under (top row) Model 1, (middle row) Model 2, and (bottom row) Model 3 with $(n, p) \in \{50, 100, 150\} \times \{40, 80, 120, 160, 200\}$.

References

Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press.

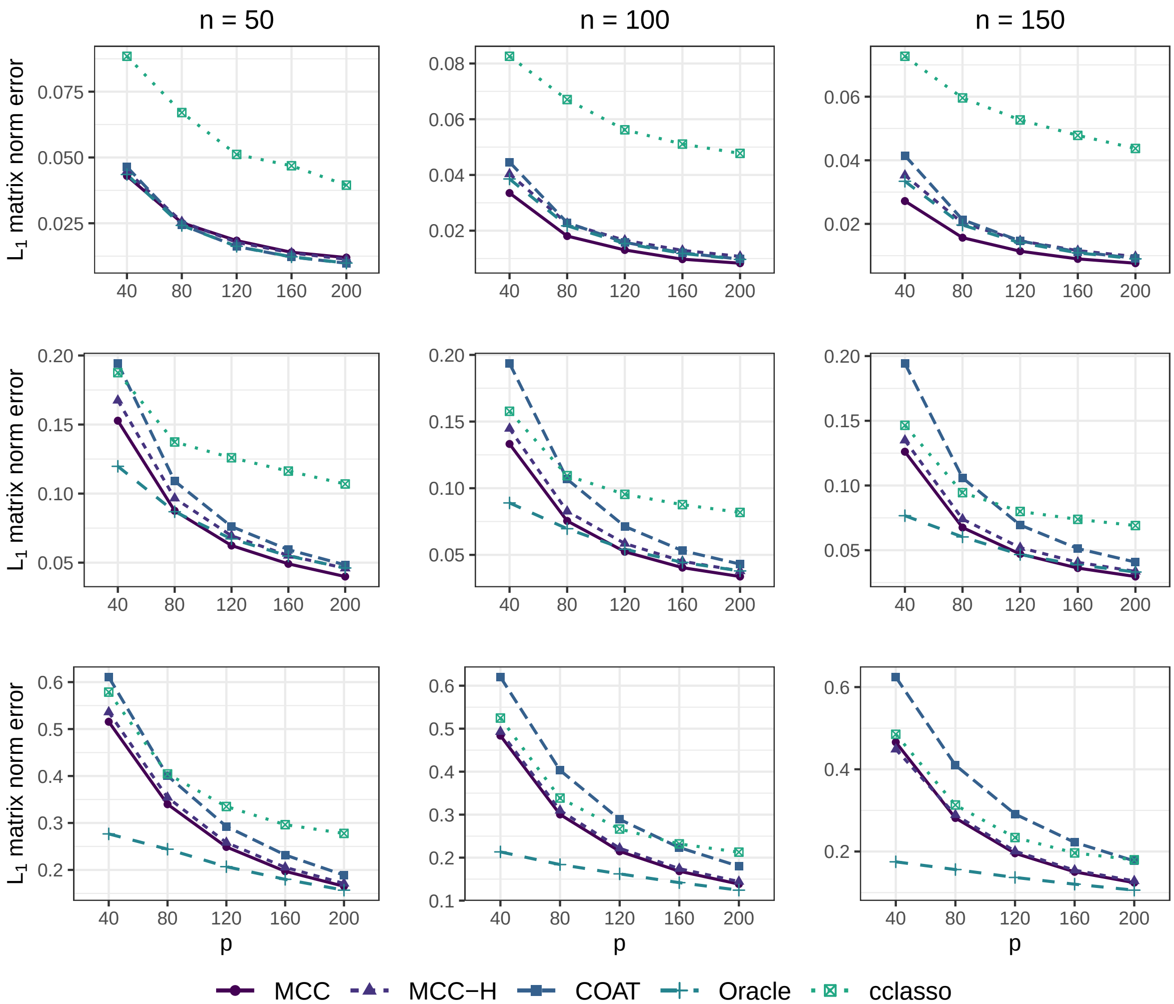

Figure 2: Average L_1 matrix norm error on the covariance scale over 50 independent replications under (top row) Model 1, (middle row) Model 2, and (bottom row) Model 3 with $(n, p) \in \{50, 100, 150\} \times \{40, 80, 120, 160, 200\}$.

Node	Phyla	Family	Genus	Species
1	Bacteroidetes	Porphyromonadaceae	Parabacteroides	
2	Bacteroidetes	Bacteroidaceae	Bacteroides	caccae
3	Bacteroidetes	Bacteroidaceae	Bacteroides	ovatus
4	Firmicutes	Lachnospiraceae	[Ruminococcus]	torques
5	Firmicutes	Lachnospiraceae	Blautia	
6	Firmicutes	Ruminococcaceae	Ruminococcus	
7	Firmicutes	Lachnospiraceae	Roseburia	faecis
8	Firmicutes	Lachnospiraceae	[Ruminococcus]	
9	Proteobacteria	Enterobacteriaceae	Escherichia	coli
10	Proteobacteria	Alcaligenaceae	Sutterella	
11	Firmicutes	Lachnospiraceae	Coprococcus	
12	Bacteroidetes	[Odoribacteraceae]	Odoribacter	
13	Firmicutes	Lachnospiraceae	Blautia	producta
14	Firmicutes	Veillonellaceae	Phascolarctobacterium	
15	Firmicutes	Lachnospiraceae		
16	Firmicutes	Ruminococcaceae	Ruminococcus	
17	Firmicutes	Lachnospiraceae		
18	Firmicutes	Lachnospiraceae	Coprococcus	
19	Firmicutes	Ruminococcaceae	Ruminococcus	bromii
20	Bacteroidetes	Porphyromonadaceae	Parabacteroides	distasonis
21	Actinobacteria	Bifidobacteriaceae	Bifidobacterium	adolescentis
22	Firmicutes	Lachnospiraceae	Coprococcus	
23	Proteobacteria	Pseudomonadaceae	Pseudomonas	fragi
24	Bacteroidetes	[Barnesiellaceae]		
25	Firmicutes	Ruminococcaceae	Ruminococcus	bromii
26	Firmicutes	Ruminococcaceae	Oscillospira	
27	Euryarchaeota	Methanobacteriaceae	Methanobrevibacter	
28	Firmicutes	Ruminococcaceae	Ruminococcus	
29	Proteobacteria	Enterobacteriaceae	Klebsiella	
30	Firmicutes	Lachnospiraceae	Lachnobacterium	
31	Verrucomicrobia	Verrucomicrobiaceae	Akkermansia	muciniphila
32	Firmicutes	Lachnospiraceae	Blautia	
33	Bacteroidetes	[Barnesiellaceae]		
34	Bacteroidetes	Porphyromonadaceae	Parabacteroides	distasonis
35	Firmicutes	Lachnospiraceae	Blautia	producta
36	Bacteroidetes	Bacteroidaceae	Bacteroides	uniformis