Differential Network Analysis for Compositional Data

Chencheng Ma[1], Hao Chen[1], Yijuan Hu[2] and Xiang Zhan[3]

[1]*Shandong University,* [2]*Peking University,* [3]*Southeast University*

Supplementary material

In this supplementary material, we present 1) details of a computing algorithm to implement DNA-Coda; 2) complete proofs of theoretical results stated in the main text.

1. Implementation of DNA-Coda

For ease of reference, we re-write the objective function of DNA-Coda (i.e., equation (2.5) in the main text) in this supplementary file as:

$$\text{minimize } |\mathbf{\Delta}|_1 \quad \text{subject to} \quad \left|\left(\widehat{\mathbf{\Sigma}}_{Y,c} \otimes \widehat{\mathbf{\Sigma}}_{X,c}\right) \text{vec}(\mathbf{\Delta}) - \text{vec}\left(\widehat{\mathbf{\Sigma}}_{X,c} - \widehat{\mathbf{\Sigma}}_{Y,c}\right)\right|_\infty \leqslant \lambda. \quad (1.1)$$

The linear program (1.1) has a $p^2 \times p^2$ constraint matrix $\widehat{\mathbf{\Sigma}}_{Y,c} \otimes \widehat{\mathbf{\Sigma}}_{X,c}$ and can become computationally demanding for large p. To address this, we adopt the modified procedure proposed by Zhao et al. (2014), which is presented as follows.

Denote the (j,k) th entry of a matrix $\mathbf{\Delta}$ by δ_{jk}, and define $\boldsymbol{\beta}$ to be the $p(p+1)/2 \times 1$ vector with $\boldsymbol{\beta} = (\delta_{jk})_{1 \leqslant j \leqslant k \leqslant p}$. Estimating a symmetric $\mathbf{\Delta}$ is thus equivalent to estimating $\boldsymbol{\beta}$, which has only $p(p+1)/2$ parameters. Define the $p^2 \times p(p+1)/2$ matrix $\mathbf{S}$ with columns indexed by $1 \leqslant j \leqslant k \leqslant p$ and with rows indexed by $l = 1, \ldots, p$ and $m = 1, \ldots, p$, so that each entry is labeled as $S_{lm,jk}$. For $j \leqslant k$, let $S_{jk,jk} = S_{kj,jk} = 1$ and set all other entries of $\mathbf{S}$ equal to zero, an example can be seen in the online supplementary materials. Therefore,

the refined optimization problem can be written as:

$$\text{minimize } |\boldsymbol{\beta}|_1 \quad \text{subject to} \quad \left|\mathbf{S}^{\mathrm{T}}\left(\widehat{\boldsymbol{\Sigma}}_{Y,c}\otimes\widehat{\boldsymbol{\Sigma}}_{X,c}\right)\mathbf{S}\boldsymbol{\beta}-\mathbf{S}^{\mathrm{T}}\operatorname{vec}\left(\widehat{\boldsymbol{\Sigma}}_{X,c}-\widehat{\boldsymbol{\Sigma}}_{Y,c}\right)\right|_{\infty}\leqslant\lambda.$$

The inequality constraint can be further improved, that is,

$$\text{minimize } |\boldsymbol{\beta}|_1 \quad \text{subject to} \quad \begin{cases} \left|\mathbf{S}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_c\mathbf{S}\boldsymbol{\beta}-\mathbf{S}^{\mathrm{T}}\hat{\mathbf{b}}\right|_{\mathrm{O}\infty}\leqslant\lambda, \\ \left|\mathbf{S}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_c\mathbf{S}\boldsymbol{\beta}-\mathbf{S}^{\mathrm{T}}\hat{\mathbf{b}}\right|_{\mathrm{D}\infty}\leqslant\lambda/2, \end{cases} \tag{1.2}$$

where $\widehat{\boldsymbol{\Sigma}}_c=\widehat{\boldsymbol{\Sigma}}_{Y,c}\otimes\widehat{\boldsymbol{\Sigma}}_{X,c}, \hat{\mathbf{b}}=\operatorname{vec}\left(\widehat{\boldsymbol{\Sigma}}_{X,c}-\widehat{\boldsymbol{\Sigma}}_{Y,c}\right)$, and for a $p(p+1)/2\times 1$ vector $\mathbf{c}, |\mathbf{c}|_{O\infty}$ denotes the sup-norm of the entries of c corresponding to the off-diagonal elements of its matrix form, while $|\mathbf{c}|_{D\infty}$ denotes the sup-norm of the entries corresponding to the diagonal elements. The matrix form of $\hat{\boldsymbol{\beta}}$ will be denoted by $\widehat{\boldsymbol{\Delta}}$. Compared with (1.1), the estimator (1.2) requires only a $p(p+1)/2\times p(p+1)/2$ constraint matrix, but needs a stronger theoretical condition to guarantee support recovery and estimation consistency, which is discussed next.

In this section, we present a detailed exposition of the Alternating Direction Method of Multipliers (ADMM) (Boyd et al., 2011), a core optimization framework adopted in this work to address the challenges outlined in our methodology. We first give the example of $\mathbf{S}$ when $p=3$:

$$\mathbf{S}=\begin{pmatrix} S_{11,11} & S_{11,12} & S_{11,13} & S_{11,22} & S_{11,23} & S_{11,33} \\ S_{21,11} & S_{21,12} & S_{21,13} & S_{21,22} & S_{21,23} & S_{21,33} \\ S_{31,11} & S_{31,12} & S_{31,13} & S_{31,22} & S_{31,23} & S_{31,33} \\ S_{12,11} & S_{12,12} & S_{12,13} & S_{12,22} & S_{12,23} & S_{21,33} \\ S_{22,11} & S_{22,12} & S_{22,13} & S_{22,22} & S_{22,23} & S_{11,33} \\ S_{32,11} & S_{32,12} & S_{32,13} & S_{32,22} & S_{32,23} & S_{32,33} \\ S_{13,11} & S_{13,12} & S_{13,13} & S_{13,22} & S_{13,23} & S_{23,33} \\ S_{23,11} & S_{23,12} & S_{23,13} & S_{23,22} & S_{23,23} & S_{13,33} \\ S_{33,11} & S_{33,12} & S_{33,13} & S_{33,22} & S_{33,23} & S_{33,33} \end{pmatrix}=\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Next, to apply the ADMM algorithm, we reformulate the original optimization problem

$$\text{minimize } |\boldsymbol{\beta}|_1 \quad \text{subject to} \quad \begin{cases} \left|\mathbf{S}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_c\mathbf{S}\boldsymbol{\beta}-\mathbf{S}^{\mathrm{T}}\hat{\mathbf{b}}\right|_{\mathrm{O}\infty}\leqslant\lambda, \\ \left|\mathbf{S}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_c\mathbf{S}\boldsymbol{\beta}-\mathbf{S}^{\mathrm{T}}\hat{\mathbf{b}}\right|_{\mathrm{D}\infty}\leqslant\lambda/2, \end{cases}$$

as:

$$\widehat{\boldsymbol{\beta}} = \underset{\mathbf{r},\boldsymbol{\beta}}{\arg\min}\, f(\mathbf{r}) + |\boldsymbol{\beta}|_1 \quad \text{subject to} \quad \mathbf{r} + \mathbf{S}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_c\mathbf{S}\boldsymbol{\beta} = \mathbf{S}^{\mathrm{T}}\hat{\mathbf{b}},$$

where $f(\mathbf{r}) = +\infty$ if $\|\mathbf{r}\|_{\mathrm{O}_\infty} > \lambda$ or $\|\mathbf{r}\|_{\mathrm{D}_\infty} > \lambda/2$, $f(\mathbf{r}) = 0$ otherwise. The augmented Lagrangian for this new problem is defined as:

$$L_\rho(\mathbf{r}, \boldsymbol{\beta}, \mathbf{u}) = f(\mathbf{r}) + |\boldsymbol{\beta}|_1 + \mathbf{u}^{\mathrm{T}}\left(\mathbf{r} + \mathbf{S}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_c\mathbf{S}\boldsymbol{\beta} - \mathbf{S}^{\mathrm{T}}\hat{\mathbf{b}}\right) + \frac{\rho}{2}\left\|\mathbf{r} + \mathbf{S}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_c\mathbf{S}\boldsymbol{\beta} - \mathbf{S}^{\mathrm{T}}\hat{\mathbf{b}}\right\|_2^2,$$

where $\mathbf{u}$ is the Lagrange multiplier and $\rho > 0$ is a user-specified penalty parameter. The ADMM algorithm solves this by iterating through the following updates:

$$\begin{aligned}
\mathbf{r}^{(t+1)} &= \underset{\mathbf{r}}{\arg\min}\, \frac{1}{2}\left\|\frac{\mathbf{u}^{(t)}}{\rho} + \mathbf{S}^{\mathrm{T}}\hat{\mathbf{b}} - \mathbf{S}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_c\mathbf{S}\boldsymbol{\beta} - \mathbf{r}\right\|_2^2 + \frac{f(\mathbf{r})}{\rho}, \\
\boldsymbol{\beta}^{(t+1)} &= \underset{\boldsymbol{\beta}}{\arg\min}\, \frac{1}{2}\left\|\frac{\mathbf{u}^{(t)}}{\rho} - \mathbf{r}^{(t+1)} + \mathbf{S}^{\mathrm{T}}\hat{\mathbf{b}} - \mathbf{S}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_c\mathbf{S}\boldsymbol{\beta}\right\|_2^2 + \frac{|\boldsymbol{\beta}|_1}{\rho}, \\
\mathbf{u}^{(t+1)} &= \mathbf{u}^{(t)} + \rho\left(\mathbf{S}^{\mathrm{T}}\hat{\mathbf{b}} - \mathbf{r}^{(t+1)} - \mathbf{S}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_c\mathbf{S}\boldsymbol{\beta}^{(t+1)}\right),
\end{aligned}$$

where the updates are performed at each iteration t. The estimator can be computed by modifying code from the R package *flare*, which incorporates several strategies to accelerate convergence. These include using a closed-form expression for $\mathbf{r}^{(t+1)}$, employing a hybrid coordinate descent and linearization procedure for $\boldsymbol{\beta}^{(t+1)}$, and dynamically adjusting ρ at each iteration.

2. Proofs of theorems

2.1 Proof of Proposition 1

As established previously in the Coda literature (Cao et al., 2019; Zhang et al., 2025), the compositional covariance/precision matrix can be calculated from their basis counterparts

via the following connections:

$$\begin{aligned}
\boldsymbol{\Sigma}_{X,c} &= \boldsymbol{\Sigma}_{X,0} - \mathbf{v}_{X,p}\mathbf{v}_{X,p}^T\boldsymbol{\Sigma}_{X,0} - \boldsymbol{\Sigma}_{X,0}\mathbf{v}_{X,p}\mathbf{v}_{X,p}^T + \mathbf{v}_{X,p}\mathbf{v}_{X,p}^T\boldsymbol{\Sigma}_{X,0}\mathbf{v}_{X,p}\mathbf{v}_{X,p}^T, \\
\boldsymbol{\Omega}_{X,c} &= \boldsymbol{\Omega}_{X,0} - \frac{\boldsymbol{\Omega}_{X,0}\mathbf{1}_p\mathbf{1}_p^T\boldsymbol{\Omega}_{X,0}}{\mathbf{1}_p^T\boldsymbol{\Omega}_{X,0}\mathbf{1}_p},
\end{aligned}$$

which are useful to establish Proposition 1. In particular, from the Proposition 1(a) and 1(b) in Zhang et al. (2025), we get $\boldsymbol{\Sigma}_{X,c}\boldsymbol{\Omega}_{X,c} = \mathbf{G}$ and $\mathbf{G}\boldsymbol{\Omega}_{X,c} = \boldsymbol{\Omega}_{X,c}$. To show $\mathbf{G}\boldsymbol{\Sigma}_{X,c} = \boldsymbol{\Sigma}_{X,c}$, we consider the representation of $\boldsymbol{\Sigma}_{X,c}$ in terms of $\boldsymbol{\Sigma}_{X,0}$:

$$\boldsymbol{\Sigma}_{X,c} = \boldsymbol{\Sigma}_{X,0} - \frac{1}{p}\mathbf{1}_p\mathbf{1}_p^T\boldsymbol{\Sigma}_{X,0} - \frac{1}{p}\boldsymbol{\Sigma}_{X,0}\mathbf{1}_p\mathbf{1}_p^T + \frac{1}{p^2}\mathbf{1}_p\mathbf{1}_p^T\boldsymbol{\Sigma}_{X,0}\mathbf{1}_p\mathbf{1}_p^T,$$

Then,

$$\begin{aligned}
\mathbf{G}\boldsymbol{\Sigma}_{X,c} &= \left(\mathbf{I}_p - \frac{1}{p}\mathbf{1}_p\mathbf{1}_p^T\right)\left(\boldsymbol{\Sigma}_{X,0} - \frac{1}{p}\mathbf{1}_p\mathbf{1}_p^T\boldsymbol{\Sigma}_{X,0} - \frac{1}{p}\boldsymbol{\Sigma}_{X,0}\mathbf{1}_p\mathbf{1}_p^T + \frac{1}{p^2}\mathbf{1}_p\mathbf{1}_p^T\boldsymbol{\Sigma}_{X,0}\mathbf{1}_p\mathbf{1}_p^T\right) \\
&= \boldsymbol{\Sigma}_{X,c} - \left(\frac{1}{p}\mathbf{1}_p\mathbf{1}_p^T\boldsymbol{\Sigma}_{X,0} - \frac{1}{p^2}\mathbf{1}_p\mathbf{1}_p^T\mathbf{1}_p\mathbf{1}_p^T\boldsymbol{\Sigma}_{X,0} - \frac{1}{p^2}\mathbf{1}_p\mathbf{1}_p^T\boldsymbol{\Sigma}_{X,0}\mathbf{1}_p\mathbf{1}_p^T + \frac{1}{p^3}\mathbf{1}_p\mathbf{1}_p^T\mathbf{1}_p\mathbf{1}_p^T\boldsymbol{\Sigma}_{X,0}\mathbf{1}_p\mathbf{1}_p^T\right) \\
&= \boldsymbol{\Sigma}_{X,c}
\end{aligned}$$

Since $\boldsymbol{\Omega}_{X,c}$, $\boldsymbol{\Sigma}_{X,c}$ and $\mathbf{G}$ are all symmetrical, we have

$$\boldsymbol{\Omega}_{X,c}\boldsymbol{\Sigma}_{X,c} = \left(\boldsymbol{\Sigma}_{X,c}\boldsymbol{\Omega}_{X,c}\right)^\top = \mathbf{G}^\top = \mathbf{G},$$

and thus, $\boldsymbol{\Omega}_{X,c}\boldsymbol{\Sigma}_{X,c} = \boldsymbol{\Sigma}_{X,c}\boldsymbol{\Omega}_{X,c}$. Next ,we find that

$$\mathbf{G}\boldsymbol{\Omega}_{X,c} = \boldsymbol{\Omega}_{X,c}\boldsymbol{\Sigma}_{X,c}\boldsymbol{\Omega}_{X,c} = \boldsymbol{\Omega}_{X,c}\mathbf{G},$$

$$\mathbf{G}\boldsymbol{\Sigma}_{X,c} = \boldsymbol{\Sigma}_{X,c}\boldsymbol{\Omega}_{X,c}\boldsymbol{\Sigma}_{X,c} = \boldsymbol{\Sigma}_{X,c}\mathbf{G},$$

which completes the proof.

This proposition offers a more complete algebraic characterization of the relationship between $\boldsymbol{\Sigma}_{X,c}$ and $\boldsymbol{\Omega}_{X,c}$. The first equality establishes that $\boldsymbol{\Omega}_{X,c}$ acts as an approximate inverse of $\boldsymbol{\Sigma}_{X,c}$, with the centering matrix $\mathbf{G}$ serving as the identity operator in the constrained

space of compositional data. The subsequent equalities further delineate commutative and invariance properties with respect to $\mathbf{G}$, reflecting the fact that both $\mathbf{\Sigma}_{X,c}$ and $\mathbf{\Omega}_{X,c}$ reside in the CLR subspace defined by the compositional constraint $\mathbf{1}_p^\top \mathbf{Z}_X = 0$. Our result extends Proposition 1 of Zhang et al. (2025), which primarily highlights the identity $\mathbf{\Sigma}_{X,c}\mathbf{\Omega}_{X,c} = \mathbf{G}$ alongside spectral bounds. In contrast, our proposition derives additional symmetric commutation relations between these matrices and $\mathbf{G}$, that were not previously established.

2.2 Proof of Proposition 2

By Theorem 1 in Zhang et al. (2025),We have

$$\mathbf{\Delta}_0 - \mathbf{\Delta}_c = \frac{\mathbf{\Omega}_{Y,0}\mathbf{1}_p\mathbf{1}_p^\top\mathbf{\Omega}_{Y,0}}{\mathbf{1}_p^\top\mathbf{\Omega}_{Y,0}\mathbf{1}_p} - \frac{\mathbf{\Omega}_{X,0}\mathbf{1}_p\mathbf{1}_p^\top\mathbf{\Omega}_{X,0}}{\mathbf{1}_p^\top\mathbf{\Omega}_{X,0}\mathbf{1}_p} = \frac{A}{B} - \frac{C}{D},$$

where

$$A = \mathbf{\Omega}_{Y,0}\mathbf{1}_p\mathbf{1}_p^\top\mathbf{\Omega}_{Y,0}, \quad B = \mathbf{1}_p^\top\mathbf{\Omega}_{Y,0}\mathbf{1}_p, \quad C = \mathbf{\Omega}_{X,0}\mathbf{1}_p\mathbf{1}_p^\top\mathbf{\Omega}_{X,0}, \quad D = \mathbf{1}_p^\top\mathbf{\Omega}_{X,0}\mathbf{1}_p.$$

Putting them over a common denominator and expanding,

$$\mathbf{\Delta}_0 - \mathbf{\Delta}_c = \frac{AD - CB}{BD} = \frac{(A-C)D + C(D-B)}{BD}.$$

Write $\mathbf{\Omega}_{Y,0} = \mathbf{\Omega}_{X,0} + \mathbf{\Delta}_0$ and substitute into A:

$$A = \mathbf{\Omega}_{Y,0}\mathbf{1}_p\mathbf{1}_p^\top\mathbf{\Omega}_{Y,0} = (\mathbf{\Omega}_{X,0} + \mathbf{\Delta}_0)\mathbf{1}_p\mathbf{1}_p^\top(\mathbf{\Omega}_{X,0} + \mathbf{\Delta}_0).$$

Expand the product:

$$A = \mathbf{\Omega}_{X,0}\mathbf{1}_p\mathbf{1}_p^\top\mathbf{\Omega}_{X,0} + \mathbf{\Omega}_{X,0}\mathbf{1}_p\mathbf{1}_p^\top\mathbf{\Delta}_0 + \mathbf{\Delta}_0\mathbf{1}_p\mathbf{1}_p^\top\mathbf{\Omega}_{X,0} + \mathbf{\Delta}_0\mathbf{1}_p\mathbf{1}_p^\top\mathbf{\Delta}_0.$$

Hence

$$\begin{aligned} A - C &= \big(\mathbf{\Omega}_{X,0}\mathbf{1}_p\mathbf{1}_p^\top\mathbf{\Omega}_{X,0} + \mathbf{\Omega}_{X,0}\mathbf{1}_p\mathbf{1}_p^\top\mathbf{\Delta}_0 + \mathbf{\Delta}_0\mathbf{1}_p\mathbf{1}_p^\top\mathbf{\Omega}_{X,0} + \mathbf{\Delta}_0\mathbf{1}_p\mathbf{1}_p^\top\mathbf{\Delta}_0\big) - \mathbf{\Omega}_{X,0}\mathbf{1}_p\mathbf{1}_p^\top\mathbf{\Omega}_{X,0} \\ &= \mathbf{\Omega}_{X,0}\mathbf{1}_p\mathbf{1}_p^\top\mathbf{\Delta}_0 + \mathbf{\Delta}_0\mathbf{1}_p\mathbf{1}_p^\top\mathbf{\Omega}_{X,0} + \mathbf{\Delta}_0\mathbf{1}_p\mathbf{1}_p^\top\mathbf{\Delta}_0. \end{aligned}$$

Alternatively,

$$\boldsymbol{\Delta}_0\mathbf{1}_p\mathbf{1}_p^\top\boldsymbol{\Omega}_{Y,0} + \boldsymbol{\Omega}_{X,0}\mathbf{1}_p\mathbf{1}_p^\top\boldsymbol{\Delta}_0 = \boldsymbol{\Delta}_0\mathbf{1}_p\mathbf{1}_p^\top(\boldsymbol{\Omega}_{X,0} + \boldsymbol{\Delta}_0) + \boldsymbol{\Omega}_{X,0}\mathbf{1}_p\mathbf{1}_p^\top\boldsymbol{\Delta}_0,$$

which matches the expression above. Therefore

$$A - C = \boldsymbol{\Delta}_0\mathbf{1}_p\mathbf{1}_p^\top\boldsymbol{\Omega}_{Y,0} + \boldsymbol{\Omega}_{X,0}\mathbf{1}_p\mathbf{1}_p^\top\boldsymbol{\Delta}_0.$$

Using this identity we get the convenient decomposition

$$\begin{aligned}\boldsymbol{\Delta}_0 - \boldsymbol{\Delta}_c = &\frac{\boldsymbol{\Delta}_0\mathbf{1}_p\mathbf{1}_p^\top\boldsymbol{\Omega}_{Y,0} + \boldsymbol{\Omega}_{X,0}\mathbf{1}_p\mathbf{1}_p^\top\boldsymbol{\Delta}_0}{\mathbf{1}_p^\top\boldsymbol{\Omega}_{Y,0}\mathbf{1}_p} \\ &+ \frac{\boldsymbol{\Omega}_{X,0}\mathbf{1}_p\mathbf{1}_p^\top\boldsymbol{\Omega}_{X,0}}{\mathbf{1}_p^\top\boldsymbol{\Omega}_{Y,0}\mathbf{1}_p \cdot \mathbf{1}_p^\top\boldsymbol{\Omega}_{X,0}\mathbf{1}_p}\big(\mathbf{1}_p^\top\boldsymbol{\Omega}_{X,0}\mathbf{1}_p - \mathbf{1}_p^\top\boldsymbol{\Omega}_{Y,0}\mathbf{1}_p\big).\end{aligned}$$

We want to bound

$$|\boldsymbol{\Delta}_0 - \boldsymbol{\Delta}_c|_\infty.$$

First main term:

$$\left|\frac{\boldsymbol{\Delta}_0\mathbf{1}_p\mathbf{1}_p^\top\boldsymbol{\Omega}_{Y,0} + \boldsymbol{\Omega}_{X,0}\mathbf{1}_p\mathbf{1}_p^\top\boldsymbol{\Delta}_0}{\mathbf{1}_p^\top\boldsymbol{\Omega}_{Y,0}\mathbf{1}_p}\right|_\infty \le \frac{|\boldsymbol{\Delta}_0\mathbf{1}_p\mathbf{1}_p^\top\boldsymbol{\Omega}_{Y,0}|_\infty + |\boldsymbol{\Omega}_{X,0}\mathbf{1}_p\mathbf{1}_p^\top\boldsymbol{\Delta}_0|_\infty}{\mathbf{1}_p^\top\boldsymbol{\Omega}_{Y,0}\mathbf{1}_p}.$$

Note that for any matrices $\mathbf{A}, \mathbf{B}$, $\mathbf{A}\mathbf{1}_p\mathbf{1}_p^\top\mathbf{B} = (\mathbf{A}\mathbf{1}_p)(\mathbf{B}\mathbf{1}_p)^\top$. Thus

$$|\boldsymbol{\Delta}_0\mathbf{1}_p\mathbf{1}_p^\top\boldsymbol{\Omega}_{Y,0}|_\infty \le |\boldsymbol{\Delta}_0\mathbf{1}_p|_\infty\,|\boldsymbol{\Omega}_{Y,0}\mathbf{1}_p|_\infty \le \|\boldsymbol{\Delta}_0\|_1\,\|\boldsymbol{\Omega}_{Y,0}\|_1 \le \sqrt{p}\|\boldsymbol{\Delta}_0\|_1\,\|\boldsymbol{\Omega}_{Y,0}\|_2.$$

Similarly,

$$|\boldsymbol{\Omega}_{X,0}\mathbf{1}_p\mathbf{1}_p^\top\boldsymbol{\Delta}_0|_\infty \le \sqrt{p}\|\boldsymbol{\Delta}_0\|_1\,\|\boldsymbol{\Omega}_{X,0}\|_2.$$

Second term:

$$\left|\frac{\boldsymbol{\Omega}_{X,0}\mathbf{1}_p\mathbf{1}_p^\top\boldsymbol{\Omega}_{X,0}}{BD}(D - B)\right|_\infty \le \frac{|\boldsymbol{\Omega}_{X,0}\mathbf{1}_p\mathbf{1}_p^\top\boldsymbol{\Omega}_{X,0}|_\infty \cdot |\,\mathbf{1}_p^\top\boldsymbol{\Delta}_0\mathbf{1}_p\,|}{BD}.$$

Note that

$$|\boldsymbol{\Omega}_{X,0}\mathbf{1}_p\mathbf{1}_p^\top\boldsymbol{\Omega}_{X,0}|_\infty \le \|\boldsymbol{\Omega}_{X,0}\|_1^2 \le p\|\boldsymbol{\Omega}_{X,0}\|_2^2, \quad |\mathbf{1}_p^\top\boldsymbol{\Delta}_0\mathbf{1}_p| \le |\boldsymbol{\Delta}_0|_1.$$

Hence

$$\left|\frac{\mathbf{\Omega}_{X,0}\mathbf{1}_p\mathbf{1}_p^\top\mathbf{\Omega}_{X,0}}{BD}(D-B)\right|_\infty \leq \frac{p\|\mathbf{\Omega}_{X,0}\|_2^2\,|\mathbf{\Delta}_0|_1}{(\mathbf{1}_p^\top\mathbf{\Omega}_{Y,0}\mathbf{1}_p)(\mathbf{1}_p^\top\mathbf{\Omega}_{X,0}\mathbf{1}_p)}.$$

Final bound:

Using $\mathbf{1}_p^\top\mathbf{\Omega}_{Y,0}\mathbf{1}_p \geq p\lambda_{\min}(\mathbf{\Omega}_{Y,0}) \geq p/R_Y$, $\mathbf{1}_p^\top\mathbf{\Omega}_{X,0}\mathbf{1}_p \geq p\lambda_{\min}(\mathbf{\Omega}_{X,0}) \geq p/R_X$, $\|\mathbf{\Omega}_{X,0}\|_2 \leq R_X$, $\|\mathbf{\Omega}_{Y,0}\|_2 \leq R_Y$ and $\|\mathbf{\Delta}_0\|_1 \leq |\mathbf{\Delta}_0|_1 \leq M$, we obtain

$$\begin{aligned}
|\mathbf{\Delta}_0-\mathbf{\Delta}_c|_\infty \leq& \frac{R_Y}{p}\left(\sqrt{p}\|\mathbf{\Delta}_0\|_1\,\|\mathbf{\Omega}_{Y,0}\|_2+\sqrt{p}\|\mathbf{\Delta}_0\|_1\,\|\mathbf{\Omega}_{X,0}\|_2\right)+\frac{R_XR_Y}{p^2}p\|\mathbf{\Omega}_{X,0}\|_2^2\,|\mathbf{\Delta}_0|_1\\
\leq& \frac{M(R_Y^2+R_XR_Y)}{\sqrt{p}}+\frac{MR_X^3R_Y}{p}\\
=& O\left(\frac{M(R_Y^2+R_XR_Y)}{\sqrt{p}}\right)
\end{aligned}$$

Therefore, there exists some constant $C_0>0$ such that we can conclude that

$$|\mathbf{\Delta}_0-\mathbf{\Delta}_c|_\infty \leq \frac{C_0M(R_Y^2+R_YR_X)}{\sqrt{p}}.$$

The next lemma comes from the proofs of Lemma 1 in Zhang et al. (2025).

Lemma A1. *Suppose that* $\mathbf{\Omega}_{X,0}$, $\mathbf{\Omega}_{Y,0} \in \mathcal{U}(M,R_X,R_Y)$

(a) If Assumption 3 holds, then for sufficiently large p, *there exist some constants* $\eta_1>0$ *and* $K_1>0$ *such that*

$$E\left\{\exp\left(\eta_1 Z_{X,j}^2/\sigma_{jj}^{X,c}\right)\right\} \leq K_1$$

for all $j=1,\ldots,p$.

Then, for any $\xi>0$, *there exists some constant* $\alpha_X>0$ *such that*

$$|\widehat{\mathbf{\Sigma}}_{X,c}-\mathbf{\Sigma}_{X,c}|_\infty \leqslant \alpha_X(\log p/n_X)^{1/2}$$

with probability at least $1-O\left(p^{-\xi}\right)$, *where* $\alpha_X = 6\eta_X^{-1}R_X\left(2K_1+1\right)(\xi+2)$ *with* $\eta_X = \eta/\left(2R_X^2\right)$ *and* $K_1=K$. *The same properties above apply to* $\mathbf{\Sigma}_{Y,c}$.

(b) If Assumption 4 holds, then there exists some constant $K_1' > 0$ such that

$$E\left|Z_{X,j}\right|^{4\gamma+4+\varepsilon} \leq K_1'$$

for all $j = 1, \ldots, p$.

Then, for any $\xi > 0$, there exists some constant $\alpha_X' > 0$ such that

$$|\widehat{\mathbf{\Sigma}}_{X,c} - \mathbf{\Sigma}_{X,c}|_\infty \leqslant \alpha_X'(\log p/n_X)^{1/2}$$

with probability at least $1 - O\left(p^{-\xi/2} + n_X^{-\varepsilon/8}\right)$, where $\alpha_X' = 3R_X\sqrt{(\theta+1)(4+\xi)}$ with $\theta = \max_{1\leq i,j\leq p} E\left[Z_{X,i}Z_{X,j} - E\left(Z_{X,i}Z_{X,j}\right)\right]^2$. The same properties above apply to $\mathbf{\Sigma}_{Y,c}$.

Lemma A2. *Let $\mathbf{\Sigma}_c = \mathbf{\Sigma}_{Y,c} \otimes \mathbf{\Sigma}_{X,c}$. Label the entries of $\mathbf{S}^{\mathrm{T}}\mathbf{\Sigma}_c\mathbf{S}$ as $\mathbf{\Sigma}^{S,c}_{j'k',jk}(1 \leqslant j' \leqslant k' \leqslant p; 1 \leqslant j \leqslant k \leqslant p.)$. Then*

$$\begin{array}{ll}
\sigma^{S,c}_{j'k',jk} = \sigma^{Y,c}_{k'k}\sigma^{X,c}_{j'j} + \sigma^{Y,c}_{k'j}\sigma^{X,c}_{j'k} + \sigma^{Y,c}_{j'k}\sigma^{X,c}_{k'j} + \sigma^{Y,c}_{j'j}\sigma^{X,c}_{k'k}, & j' \neq k', j \neq k; \\
\sigma^{S,c}_{j'k',jj} = 2\sigma^{Y,c}_{k'j}\sigma^{X,c}_{j'j} + 2\sigma^{Y,c}_{j'j}\sigma^{X,c}_{k'j}, & j' \neq k', j = k; \\
\sigma^{S,c}_{j'j',jk} = 2\sigma^{Y,c}_{j'k}\sigma^{X,c}_{j'j} + 2\sigma^{Y,c}_{j'j}\sigma^{X,c}_{j'k}, & j' = k', j \neq k; \\
\sigma^{S,c}_{j'j',jj} = 4\sigma^{Y,c}_{j'j}\sigma^{X,c}_{j'j}, & j' = k', j = k.
\end{array}$$

Proof. Label the entries of $\mathbf{\Sigma}_c$ as $\sigma^c_{l'm',lm}, (l', m', l, m = 1, \ldots, p)$ and the entries of $\mathbf{S}$ as $S_{lm,jk}(l = 1, \ldots, p; m = 1, \ldots, p; 1 \leqslant j \leqslant k \leqslant p)$. According to the definition of the Kronecker product, we have the relationship $\sigma^c_{l'm',lm} = \sigma^{Y,c}_{m'm}\sigma^{X,c}_{l'l}$ and $\sigma^{S,c}_{j'k',jk} = \sum_{l',m',l,m} S_{l'm',j'k'}\sigma^c_{l'm',lm}S_{lm,jk}$. Thus, the lemma follows directly from the definition of the entries of S.

2.3 Proof of Lemma 1

Let the entries of $\mathbf{\Delta}_0$ be denoted by δ^0_{jk}, and define the $p(p+1)/2 \times 1$ vector $\boldsymbol{\beta}_0 = \left(\delta^0_{jk}\right)_{1\leqslant j\leqslant k\leqslant p}$. Define $\mathbf{\Sigma}_c$ as in Lemma A2, and let $\hat{\mathbf{b}}_c = \mathrm{vec}\left(\widehat{\mathbf{\Sigma}}_{X,c} - \widehat{\mathbf{\Sigma}}_{Y,c}\right)$ and $\mathbf{b}_c = \mathrm{vec}\left(\mathbf{\Sigma}_{X,c} - \mathbf{\Sigma}_{Y,c}\right)$. Since

$\mathbf{T}_X$ and $\mathbf{T}_Y$ satisfy the conditions of Lemma A1, we have

$$\left|\widehat{\mathbf{\Sigma}}_{X,c} - \mathbf{\Sigma}_{X,c}\right|_\infty, \ \left|\widehat{\mathbf{\Sigma}}_{Y,c} - \mathbf{\Sigma}_{Y,c}\right|_\infty \leqslant C\left\{\frac{\log p}{n_{\min}}\right\}^{1/2}$$

with probability at least $1 - O\left(p^{-\xi}\right)$ $\left(\text{or } 1 - O\left(p^{-\xi/2} + n_{\min}^{-\varepsilon/8}\right)\right)$, where $n_{\min} = \min(n_X, n_Y)$ and $C = \max\left(\alpha_X, \alpha_Y\right)$ (or $\max\left(\alpha'_X, \alpha'_Y\right)$) as defined in Lemma A1.

Let

$$\mathbf{E}_X = \widehat{\mathbf{\Sigma}}_{X,c} - \mathbf{\Sigma}_{X,c}, \qquad \mathbf{E}_Y = \widehat{\mathbf{\Sigma}}_{Y,c} - \mathbf{\Sigma}_{Y,c},$$

and define

$$\mathbf{R}_c = \mathbf{\Sigma}_{X,c}\mathbf{\Delta}_0\mathbf{\Sigma}_{Y,c} - \left(\mathbf{\Sigma}_{X,c} - \mathbf{\Sigma}_{Y,c}\right).$$

Since

$$\mathbf{\Sigma}_{X,c} = \mathbf{G}\mathbf{\Sigma}_{X,0}\mathbf{G}, \qquad \mathbf{\Sigma}_{Y,c} = \mathbf{G}\mathbf{\Sigma}_{Y,0}\mathbf{G},$$

and

$$\mathbf{\Sigma}_{X,0}\mathbf{\Delta}_0\mathbf{\Sigma}_{Y,0} = \mathbf{\Sigma}_{X,0} - \mathbf{\Sigma}_{Y,0},$$

we have

$$\mathbf{R}_c = \mathbf{G}\mathbf{\Sigma}_{X,0}\left(\mathbf{G}\mathbf{\Delta}_0\mathbf{G} - \mathbf{\Delta}_0\right)\mathbf{\Sigma}_{Y,0}\mathbf{G}.$$

Let $\mathbf{P} = p^{-1}\mathbf{1}_p\mathbf{1}_p^{\mathrm{T}}$, so that $\mathbf{G} = \mathbf{I}_p - \mathbf{P}$. Then

$$\mathbf{G}\mathbf{\Delta}_0\mathbf{G} - \mathbf{\Delta}_0 = -\mathbf{P}\mathbf{\Delta}_0 - \mathbf{\Delta}_0\mathbf{P} + \mathbf{P}\mathbf{\Delta}_0\mathbf{P}.$$

Therefore,

$$\begin{aligned}\mathbf{R}_c = & -\frac{1}{p}\left(\mathbf{G}\mathbf{\Sigma}_{X,0}\mathbf{1}_p\right)\left(\mathbf{1}_p^{\mathrm{T}}\mathbf{\Delta}_0\mathbf{\Sigma}_{Y,0}\mathbf{G}\right) \\ & -\frac{1}{p}\left(\mathbf{G}\mathbf{\Sigma}_{X,0}\mathbf{\Delta}_0\mathbf{1}_p\right)\left(\mathbf{1}_p^{\mathrm{T}}\mathbf{\Sigma}_{Y,0}\mathbf{G}\right) \\ & +\frac{1}{p^2}\left(\mathbf{G}\mathbf{\Sigma}_{X,0}\mathbf{1}_p\right)\left(\mathbf{1}_p^{\mathrm{T}}\mathbf{\Delta}_0\mathbf{1}_p\right)\left(\mathbf{1}_p^{\mathrm{T}}\mathbf{\Sigma}_{Y,0}\mathbf{G}\right).\end{aligned}$$

Using $\|\mathbf{G}\|_2 = 1$, $\|\mathbf{\Sigma}_{X,0}\|_2 \leqslant R_X$, $\|\mathbf{\Sigma}_{Y,0}\|_2 \leqslant R_Y$, $|\mathbf{\Delta}_0|_1 \leqslant M$, and $|\mathbf{1}_p|_2 = \sqrt{p}$, we first have

$$|\mathbf{G}\mathbf{\Sigma}_{X,0}\mathbf{1}_p|_\infty \leqslant |\mathbf{G}\mathbf{\Sigma}_{X,0}\mathbf{1}_p|_2 \leqslant \|\mathbf{G}\|_2\|\mathbf{\Sigma}_{X,0}\|_2|\mathbf{1}_p|_2 \leqslant R_X\sqrt{p},$$

$$\left|\mathbf{1}_p^{\mathrm{T}}\mathbf{\Sigma}_{Y,0}\mathbf{G}\right|_\infty = |\mathbf{G}\mathbf{\Sigma}_{Y,0}\mathbf{1}_p|_\infty \leqslant R_Y\sqrt{p}.$$

Moreover,

$$\left|\mathbf{1}_p^{\mathrm{T}}\mathbf{\Delta}_0\right|_1 = \sum_{j=1}^{p}\left|\sum_{k=1}^{p}(\Delta_0)_{kj}\right| \leqslant \sum_{j=1}^{p}\sum_{k=1}^{p}|(\Delta_0)_{kj}| = |\mathbf{\Delta}_0|_1 \leqslant M,$$

$$|\mathbf{\Delta}_0\mathbf{1}_p|_1 \leqslant |\mathbf{\Delta}_0|_1 \leqslant M,$$

and

$$\left|\mathbf{1}_p^{\mathrm{T}}\mathbf{\Delta}_0\mathbf{1}_p\right| \leqslant \sum_{j=1}^{p}\sum_{k=1}^{p}|(\Delta_0)_{jk}| = |\mathbf{\Delta}_0|_1 \leqslant M.$$

For the first term in $\mathbf{R}_c$,

$$\begin{aligned}
&\frac{1}{p}\left|(\mathbf{G}\mathbf{\Sigma}_{X,0}\mathbf{1}_p)\left(\mathbf{1}_p^{\mathrm{T}}\mathbf{\Delta}_0\mathbf{\Sigma}_{Y,0}\mathbf{G}\right)\right|_\infty \\
&= \frac{1}{p}|\mathbf{G}\mathbf{\Sigma}_{X,0}\mathbf{1}_p|_\infty\left|\mathbf{1}_p^{\mathrm{T}}\mathbf{\Delta}_0\mathbf{\Sigma}_{Y,0}\mathbf{G}\right|_\infty \\
&\leqslant \frac{1}{p}\left(R_X\sqrt{p}\right)\left|\mathbf{1}_p^{\mathrm{T}}\mathbf{\Delta}_0\right|_1|\mathbf{\Sigma}_{Y,0}\mathbf{G}|_\infty \\
&\leqslant \frac{1}{p}\left(R_X\sqrt{p}\right)MR_Y = \frac{R_XR_YM}{\sqrt{p}}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\frac{1}{p}\left|(\mathbf{G}\mathbf{\Sigma}_{X,0}\mathbf{\Delta}_0\mathbf{1}_p)\left(\mathbf{1}_p^{\mathrm{T}}\mathbf{\Sigma}_{Y,0}\mathbf{G}\right)\right|_\infty \\
&\leqslant \frac{1}{p}|\mathbf{G}\mathbf{\Sigma}_{X,0}|_\infty|\mathbf{\Delta}_0\mathbf{1}_p|_1\left|\mathbf{1}_p^{\mathrm{T}}\mathbf{\Sigma}_{Y,0}\mathbf{G}\right|_\infty \\
&\leqslant \frac{1}{p}R_XM\left(R_Y\sqrt{p}\right) = \frac{R_XR_YM}{\sqrt{p}}.
\end{aligned}$$

For the third term,

$$\begin{aligned}
&\frac{1}{p^2}\left|(\mathbf{G}\mathbf{\Sigma}_{X,0}\mathbf{1}_p)\left(\mathbf{1}_p^{\mathrm{T}}\mathbf{\Delta}_0\mathbf{1}_p\right)\left(\mathbf{1}_p^{\mathrm{T}}\mathbf{\Sigma}_{Y,0}\mathbf{G}\right)\right|_\infty \\
&\leqslant \frac{1}{p^2}\left(R_X\sqrt{p}\right)M\left(R_Y\sqrt{p}\right) = \frac{R_XR_YM}{p}.
\end{aligned}$$

Combining these bounds and noting that $p^{-1} \leqslant p^{-1/2}$ gives

$$|\mathbf{R}_c|_\infty \leqslant R_XR_YM\left(\frac{2}{\sqrt{p}} + \frac{1}{p}\right) \leqslant \frac{3R_XR_YM}{\sqrt{p}}.$$

Next, using $\mathrm{vec}(\boldsymbol{\Delta}_0) = \mathbf{S}\boldsymbol{\beta}_0$ and the identity $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^{\mathrm{T}} \otimes \mathbf{A})\,\mathrm{vec}(\mathbf{B})$, we have

$$\begin{aligned}
&\left|\widehat{\boldsymbol{\Sigma}}_{X,c}\boldsymbol{\Delta}_0\widehat{\boldsymbol{\Sigma}}_{Y,c} - \left(\widehat{\boldsymbol{\Sigma}}_{X,c} - \widehat{\boldsymbol{\Sigma}}_{Y,c}\right)\right|_\infty = \left|\widehat{\boldsymbol{\Sigma}}_c\mathbf{S}\boldsymbol{\beta}_0 - \hat{\mathbf{b}}_c\right|_\infty \\
&= \left|\mathbf{R}_c + \mathbf{E}_X\boldsymbol{\Delta}_0\boldsymbol{\Sigma}_{Y,c} + \boldsymbol{\Sigma}_{X,c}\boldsymbol{\Delta}_0\mathbf{E}_Y + \mathbf{E}_X\boldsymbol{\Delta}_0\mathbf{E}_Y - \mathbf{E}_X + \mathbf{E}_Y\right|_\infty .
\end{aligned}$$

For any conformable matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$,

$$|\mathbf{ABC}|_\infty \leqslant |\mathbf{A}|_\infty |\mathbf{B}|_1 |\mathbf{C}|_\infty .$$

Moreover,

$$|\boldsymbol{\Sigma}_{X,c}|_\infty \leqslant \|\boldsymbol{\Sigma}_{X,c}\|_2 \leqslant R_X, \qquad |\boldsymbol{\Sigma}_{Y,c}|_\infty \leqslant \|\boldsymbol{\Sigma}_{Y,c}\|_2 \leqslant R_Y .$$

Let

$$t = C\left\{\frac{\log p}{\min(n_X, n_Y)}\right\}^{1/2}.$$

Then

$$\begin{aligned}
|\mathbf{E}_X\boldsymbol{\Delta}_0\boldsymbol{\Sigma}_{Y,c}|_\infty &\leqslant MR_Y t, \\
|\boldsymbol{\Sigma}_{X,c}\boldsymbol{\Delta}_0\mathbf{E}_Y|_\infty &\leqslant MR_X t, \\
|\mathbf{E}_X\boldsymbol{\Delta}_0\mathbf{E}_Y|_\infty &\leqslant Mt^2, \\
|-\mathbf{E}_X + \mathbf{E}_Y|_\infty &\leqslant 2t.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\left|\widehat{\boldsymbol{\Sigma}}_{X,c}\boldsymbol{\Delta}_0\widehat{\boldsymbol{\Sigma}}_{Y,c} - \left(\widehat{\boldsymbol{\Sigma}}_{X,c} - \widehat{\boldsymbol{\Sigma}}_{Y,c}\right)\right|_\infty \\
&\leqslant \{M(R_X + R_Y) + 2\}\,t + Mt^2 + \frac{3R_X R_Y M}{\sqrt{p}}.
\end{aligned}$$

Under either Assumption 3 or Assumption 4,

$$\left\{\frac{\log p}{\min(n_X, n_Y)}\right\}^{1/2} = o(1).$$

Hence, for sufficiently large p, $t \leqslant 1$, and thus $t^2 \leqslant t$. Consequently,

$$\begin{aligned}
&\left|\widehat{\boldsymbol{\Sigma}}_{X,c}\boldsymbol{\Delta}_0\widehat{\boldsymbol{\Sigma}}_{Y,c} - \left(\widehat{\boldsymbol{\Sigma}}_{X,c} - \widehat{\boldsymbol{\Sigma}}_{Y,c}\right)\right|_\infty \\
&\leqslant C\left\{M(R_X + R_Y + 1) + 2\right\}\left\{\frac{\log p}{\min(n_X, n_Y)}\right\}^{1/2} + \frac{3R_X R_Y M}{\sqrt{p}} \\
&\leqslant 2C\left\{M(R_X + R_Y + 1) + 1\right\}\left\{\frac{\log p}{\min(n_X, n_Y)}\right\}^{1/2} + \frac{3R_X R_Y M}{\sqrt{p}}.
\end{aligned}$$

This proves Lemma 1.

2.4 Proof of Theorem 1

Define $\boldsymbol{\beta}_0$, $\hat{\mathbf{b}}_c$ and $\mathbf{b}_c$ as in the proof of Lemma 1, and define $\boldsymbol{\Sigma}_c$ as in Lemma A2. Let $\mathbf{h} = \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0$. The bound on $\left|\widehat{\boldsymbol{\Delta}} - \boldsymbol{\Delta}_0\right|_\infty = |\mathbf{h}|_\infty$ is obtained by following Lounici (2008). Denote the ath component of $\mathbf{S}^{\mathrm{T}}\boldsymbol{\Sigma}_c\mathbf{S}\mathbf{h}$ by $\left(\mathbf{S}^{\mathrm{T}}\boldsymbol{\Sigma}_c\mathbf{S}\mathbf{h}\right)_a$, the (a,b)th entry of $\mathbf{S}^{\mathrm{T}}\boldsymbol{\Sigma}_c\mathbf{S}$ by $\sigma_{ab}^{S,c}$, and the bth component of $\mathbf{h}$ by $\mathbf{h}_b$. Also let $\mu_c = \max_{a\neq b}\left|\sigma_{ab}^{S,c}\right|$. Then

$$\left(\mathbf{S}^{\mathrm{T}}\boldsymbol{\Sigma}_c\mathbf{S}\mathbf{h}\right)_a = \sum_{b=1}\sigma_{ab}^{S,c}\mathbf{h}_b = \sigma_{aa}^{S,c}\mathbf{h}_a + \sum_{b\neq a}\sigma_{ab}^{S,c}\mathbf{h}_b,$$

which implies that

$$\left|\sigma_{aa}^{S,c}\mathbf{h}_a\right| \leqslant \left|\mathbf{S}^{\mathrm{T}}\boldsymbol{\Sigma}_c\mathbf{S}\mathbf{h}\right|_\infty + \mu_c\sum_{b\neq a}|\mathbf{h}_b|. \tag{2.1}$$

The diagonal terms $\sigma_{aa}^{S,c}$ can be relabelled as $\sigma_{jk,jk}^{S,c}$, where it is possible that $j = k$. According to Lemma A2, these diagonal terms satisfy $\sigma_{jk,jk}^{S,c} \geqslant \sigma_{\min}^{S,c} > 0$, where $\sigma_{\min}^{S,c}$ is defined in Assumption 2. The off-diagonal terms $\sigma_{ab}^{S,c}$ with $a \neq b$ can be relabelled as $\sigma_{j'k',jk}^{S,c}$, where either $j' \neq j$ or $k' \neq k$. By Lemma A2, these off-diagonal terms satisfy

$$\left|\sigma_{j'k',jk}^{S,c}\right| \leqslant 4\max\left(\mu_{X,c}\sigma_{\max}^{Y,c}, \mu_{Y,c}\sigma_{\max}^{X,c}\right) = \mu_c.$$

Utilizing these facts together with Assumption 2, equation (2.1) becomes

$$|\mathbf{h}|_\infty \leqslant \frac{1}{\sigma_{\min}^{S,c}}\left(\left|\mathbf{S}^{\mathrm{T}}\boldsymbol{\Sigma}_c\mathbf{S}\mathbf{h}\right|_\infty + \frac{\sigma_{\min}^{S,c}}{2s}|\mathbf{h}|_1\right). \tag{2.2}$$

First, we show that $\boldsymbol{\beta}_0$ belongs to the feasible set with high probability. By Lemma 1,

$$\left|\widehat{\boldsymbol{\Sigma}}_c \mathbf{S}\boldsymbol{\beta}_0 - \hat{\mathbf{b}}_c\right|_\infty \leqslant B_{n,p},$$

where

$$B_{n,p} = 2C\left\{M(R_X + R_Y + 1) + 1\right\}\left\{\frac{\log p}{\min\left(n_X, n_Y\right)}\right\}^{1/2} + \frac{3R_X R_Y M}{\sqrt{p}}.$$

By the definition of $\mathbf{S}$,

$$\begin{aligned}\left|\mathbf{S}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_c \mathbf{S}\boldsymbol{\beta}_0 - \mathbf{S}^{\mathrm{T}}\hat{\mathbf{b}}_c\right|_\infty &\leqslant \left\|\mathbf{S}^{\mathrm{T}}\right\|_\infty \left|\widehat{\boldsymbol{\Sigma}}_c \mathbf{S}\boldsymbol{\beta}_0 - \hat{\mathbf{b}}_c\right|_\infty \\ &\leqslant 2B_{n,p} = \lambda,\end{aligned}$$

where $\left\|\mathbf{S}^{\mathrm{T}}\right\|_\infty = 2$. Hence, taking

$$\lambda = 4C\left\{M(R_X + R_Y + 1) + 1\right\}\left\{\frac{\log p}{\min\left(n_X, n_Y\right)}\right\}^{1/2} + \frac{6R_X R_Y M}{\sqrt{p}},$$

$\boldsymbol{\beta}_0$ is feasible with the stated probability. Since $\hat{\boldsymbol{\beta}}$ minimizes the ℓ_1 norm over the feasible set,

$$|\hat{\boldsymbol{\beta}}|_1 \leqslant |\boldsymbol{\beta}_0|_1 \leqslant |\boldsymbol{\Delta}_0|_1 \leqslant M.$$

Now, we derive a bound for $|\mathbf{h}|_1$. Let T_0 be the set of indices corresponding to the support of $\boldsymbol{\beta}_0$. For any $p(p+1)/2 \times 1$ vector $\mathbf{a}$, let $\mathbf{a}_{T_0}$ be the vector whose components outside T_0 are set to zero. Since $\boldsymbol{\beta}_0$ is supported on T_0 and $|\hat{\boldsymbol{\beta}}|_1 \leqslant |\boldsymbol{\beta}_0|_1$, we have

$$\left|\boldsymbol{\beta}_{0T_0}\right|_1 - \left(\left|\hat{\boldsymbol{\beta}}_{T_0}\right|_1 + \left|\hat{\boldsymbol{\beta}}_{T_0^c}\right|_1\right) \geqslant 0.$$

By the triangle inequality,

$$\left|\boldsymbol{\beta}_{0T_0} - \hat{\boldsymbol{\beta}}_{T_0}\right|_1 \geqslant \left|\boldsymbol{\beta}_{0T_0}\right|_1 - \left|\hat{\boldsymbol{\beta}}_{T_0}\right|_1 \geqslant \left|\hat{\boldsymbol{\beta}}_{T_0^c}\right|_1 = \left|\boldsymbol{\beta}_{0T_0^c} - \hat{\boldsymbol{\beta}}_{T_0^c}\right|_1.$$

That is,

$$\left|\mathbf{h}_{T_0^c}\right|_1 \leqslant \left|\mathbf{h}_{T_0}\right|_1.$$

Hence,

$$|\mathbf{h}|_1 \leqslant 2\,|\mathbf{h}_{T_0}|_1 \leqslant 2s^{1/2}\,|\mathbf{h}_{T_0}|_2\,.$$

To bound $|\mathbf{h}_{T_0}|_2$, note that, following Cai et al. (2009), for any s-sparse vector $\mathbf{d}$,

$$\begin{aligned}
\left|\mathbf{d}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\boldsymbol{\Sigma}_c\mathbf{S}\mathbf{d}\right| &\geqslant \sum_a \sigma_{aa}^{S,c}\mathbf{d}_a^2 - \left|\sum_{a\neq b}\sigma_{ab}^{S,c}\mathbf{d}_a\mathbf{d}_b\right| \\
&\geqslant \sigma_{\min}^{S,c}|\mathbf{d}|_2^2 - \mu_c\sum_{a\neq b}|\mathbf{d}_a\mathbf{d}_b| \\
&\geqslant \left\{\sigma_{\min}^{S,c} - (s-1)\mu_c\right\}|\mathbf{d}|_2^2.
\end{aligned}$$

This implies that

$$\begin{aligned}
\left|\mathbf{h}_{T_0}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\boldsymbol{\Sigma}_c\mathbf{S}\mathbf{h}\right| &\geqslant \left|\mathbf{h}_{T_0}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\boldsymbol{\Sigma}_c\mathbf{S}\mathbf{h}_{T_0}\right| - \left|\mathbf{h}_{T_0}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\boldsymbol{\Sigma}_c\mathbf{S}\mathbf{h}_{T_0^c}\right| \\
&\geq \left\{\sigma_{\min}^{S,c} - (s-1)\mu_c\right\}|\mathbf{h}_{T_0}|_2^2 - \left|\sum_{a,b}\sigma_{ab}^{S,c}\,\mathbf{h}_{T_0 a}\,\mathbf{h}_{T_0^c b}\right| \\
&\geqslant \left\{\sigma_{\min}^{S,c} - (s-1)\mu_c\right\}|\mathbf{h}_{T_0}|_2^2 - \mu_c\,|\mathbf{h}_{T_0}|_1\left|\mathbf{h}_{T_0^c}\right|_1 \\
&\geqslant \left\{\sigma_{\min}^{S,c} - (s-1)\mu_c\right\}|\mathbf{h}_{T_0}|_2^2 - \mu_c\,|\mathbf{h}_{T_0}|_1^2 \\
&\geqslant \left\{\sigma_{\min}^{S,c} - (2s-1)\mu_c\right\}|\mathbf{h}_{T_0}|_2^2\,.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\left|\mathbf{h}_{T_0}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\boldsymbol{\Sigma}_c\mathbf{S}\mathbf{h}\right| &\leqslant |\mathbf{h}_{T_0}|_1\left|\mathbf{S}^{\mathrm{T}}\boldsymbol{\Sigma}_c\mathbf{S}\mathbf{h}\right|_\infty \\
&\leqslant s^{1/2}\,|\mathbf{h}_{T_0}|_2\left|\mathbf{S}^{\mathrm{T}}\boldsymbol{\Sigma}_c\mathbf{S}\mathbf{h}\right|_\infty.
\end{aligned}$$

Consequently,

$$|\mathbf{h}|_1 \leq 2s^{1/2}\,|\mathbf{h}_{T_0}|_2 \leqslant \frac{2s\left|\mathbf{S}^{\mathrm{T}}\boldsymbol{\Sigma}_c\mathbf{S}\mathbf{h}\right|_\infty}{\sigma_{\min}^{S,c} - (2s-1)\mu_c}.$$

Substituting the above bound for $|\mathbf{h}|_1$ into (2.2) yields

$$|\mathbf{h}|_\infty \leqslant \frac{1}{\sigma_{\min}^{S,c}}\left\{1 + \frac{\sigma_{\min}^{S,c}}{\sigma_{\min}^{S,c} - (2s-1)\mu_c}\right\}\left|\mathbf{S}^{\mathrm{T}}\boldsymbol{\Sigma}_c\mathbf{S}\mathbf{h}\right|_\infty.$$

It remains to bound $\left|\mathbf{S}^{\mathrm{T}}\boldsymbol{\Sigma}_c\mathbf{S}\mathbf{h}\right|_\infty$. We have

$$\begin{aligned}
\mathbf{S}^{\mathrm{T}}\boldsymbol{\Sigma}_c\mathbf{S}\mathbf{h} &= \mathbf{S}^{\mathrm{T}}\boldsymbol{\Sigma}_c\mathbf{S}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)\\
&= \mathbf{S}^{\mathrm{T}}\boldsymbol{\Sigma}_c\mathbf{S}\hat{\boldsymbol{\beta}} - \mathbf{S}^{\mathrm{T}}\boldsymbol{\Sigma}_c\mathbf{S}\boldsymbol{\beta}_0\\
&= \mathbf{S}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_c\mathbf{S}\hat{\boldsymbol{\beta}} + \mathbf{S}^{\mathrm{T}}\left(\boldsymbol{\Sigma}_c-\widehat{\boldsymbol{\Sigma}}_c\right)\mathbf{S}\hat{\boldsymbol{\beta}} - \mathbf{S}^{\mathrm{T}}\boldsymbol{\Sigma}_c\mathbf{S}\boldsymbol{\beta}_0\\
&= \left(\mathbf{S}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_c\mathbf{S}\hat{\boldsymbol{\beta}} - \mathbf{S}^{\mathrm{T}}\hat{\mathbf{b}}_c\right) + \mathbf{S}^{\mathrm{T}}\left(\boldsymbol{\Sigma}_c-\widehat{\boldsymbol{\Sigma}}_c\right)\mathbf{S}\hat{\boldsymbol{\beta}}\\
&\quad + \mathbf{S}^{\mathrm{T}}\hat{\mathbf{b}}_c - \mathbf{S}^{\mathrm{T}}\boldsymbol{\Sigma}_c\mathbf{S}\boldsymbol{\beta}_0\\
&= \left(\mathbf{S}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_c\mathbf{S}\hat{\boldsymbol{\beta}} - \mathbf{S}^{\mathrm{T}}\hat{\mathbf{b}}_c\right) + \mathbf{S}^{\mathrm{T}}\left(\boldsymbol{\Sigma}_c-\widehat{\boldsymbol{\Sigma}}_c\right)\mathbf{S}\hat{\boldsymbol{\beta}}\\
&\quad + \mathbf{S}^{\mathrm{T}}\left(\hat{\mathbf{b}}_c-\mathbf{b}_c\right) + \mathbf{S}^{\mathrm{T}}\left(\mathbf{b}_c-\boldsymbol{\Sigma}_c\mathbf{S}\boldsymbol{\beta}_0\right).
\end{aligned}$$

Since $\hat{\boldsymbol{\beta}}$ is feasible,

$$\left|\mathbf{S}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_c\mathbf{S}\hat{\boldsymbol{\beta}} - \mathbf{S}^{\mathrm{T}}\hat{\mathbf{b}}_c\right|_\infty \leqslant \lambda.$$

Moreover, by the definition of $\mathbf{S}$,

$$\left\|\mathbf{S}^{\mathrm{T}}\right\|_\infty = 2, \qquad \|\mathbf{S}\|_1 = 2.$$

Let

$$t = C\left\{\frac{\log p}{\min\left(n_X, n_Y\right)}\right\}^{1/2}.$$

From the proof of Lemma 1,

$$|\mathbf{E}_X|_\infty \leqslant t, \qquad |\mathbf{E}_Y|_\infty \leqslant t, \qquad |\mathbf{R}_c|_\infty \leqslant \frac{3R_XR_YM}{\sqrt{p}}.$$

Furthermore,

$$\widehat{\boldsymbol{\Sigma}}_c - \boldsymbol{\Sigma}_c = \mathbf{E}_Y\otimes\boldsymbol{\Sigma}_{X,c} + \boldsymbol{\Sigma}_{Y,c}\otimes\mathbf{E}_X + \mathbf{E}_Y\otimes\mathbf{E}_X,$$

and hence

$$\begin{aligned}
\left|\widehat{\boldsymbol{\Sigma}}_c - \boldsymbol{\Sigma}_c\right|_\infty &\leqslant R_Xt + R_Yt + t^2\\
&\leqslant (R_X + R_Y + 1)t
\end{aligned}$$

for sufficiently large p, because $t = o(1)$.

Also,

$$\left|\hat{\mathbf{b}}_c - \mathbf{b}_c\right|_\infty \leqslant |\mathbf{E}_X|_\infty + |\mathbf{E}_Y|_\infty \leqslant 2t,$$

and

$$\mathbf{b}_c - \boldsymbol{\Sigma}_c \mathbf{S}\boldsymbol{\beta}_0 = -\operatorname{vec}(\mathbf{R}_c).$$

Therefore,

$$\begin{aligned}
\left|\mathbf{S}^{\mathrm{T}}\boldsymbol{\Sigma}_c\mathbf{S}\mathbf{h}\right|_\infty &\leqslant \lambda + 4M(R_X + R_Y + 1)t + 4t + \frac{6R_X R_Y M}{\sqrt{p}} \\
&= \lambda + 4C\left\{M(R_X + R_Y + 1) + 1\right\}\left\{\frac{\log p}{\min\left(n_X, n_Y\right)}\right\}^{1/2} + \frac{6R_X R_Y M}{\sqrt{p}} \\
&= 2\lambda.
\end{aligned}$$

It follows that

$$|\mathbf{h}|_\infty \leqslant \frac{1}{\sigma_{\min}^{S,c}}\left\{1 + \frac{\sigma_{\min}^{S,c}}{\sigma_{\min}^{S,c} - (2s-1)\mu_c}\right\} 2\lambda.$$

Next, we prove $\left\|\widehat{\boldsymbol{\Delta}} - \boldsymbol{\Delta}_0\right\|_F$. Solutions to this type of ℓ_1 constrained optimization problem satisfy $\left|\mathbf{h}_{T_0^c}\right|_1 \leqslant |\mathbf{h}_{T_0}|_1$. Cai et al. (2009) used this property of $\mathbf{h}$, along with their Lemma 3, to show that

$$\left|\mathbf{h}_{T_0^c \setminus T^*}\right|_2 \leqslant \left|\mathbf{h}_{T_0 \cup T^*}\right|_2,$$

where T^* is the set of indices corresponding to the $s/4$ largest components of $\mathbf{h}_{T_0^c}$. Thus,

$$|\mathbf{h}|_2 \leqslant 2\left|\mathbf{h}_{T_0 \cup T^*}\right|_2 \leqslant 2(1.25s)^{1/2}|\mathbf{h}|_\infty.$$

Therefore,

$$\begin{aligned}
\left\|\widehat{\boldsymbol{\Delta}} - \boldsymbol{\Delta}_0\right\|_F &\leqslant \sqrt{2}|\mathbf{h}|_2 \\
&\leqslant (10s)^{1/2}|\mathbf{h}|_\infty \\
&\leqslant \frac{(10s)^{1/2}}{\sigma_{\min}^{S,c}}\left\{1 + \frac{\sigma_{\min}^{S,c}}{\sigma_{\min}^{S,c} - (2s-1)\mu_c}\right\} 2\lambda.
\end{aligned}$$

This completes the proof of Theorem 1.

2.5 Proof of Corollary 1

Let $n_{\min} = \min(n_X, n_Y)$. Under the condition

$$n_{\min} = o(p \log p),$$

we have

$$\frac{p^{-1/2}}{\{\log p / n_{\min}\}^{1/2}} = \left\{ \frac{n_{\min}}{p \log p} \right\}^{1/2} = o(1).$$

Therefore,

$$\frac{1}{\sqrt{p}} = o\left\{ \left(\frac{\log p}{n_{\min}} \right)^{1/2} \right\}.$$

It follows that the $p^{-1/2}$ identification term in λ is asymptotically negligible relative to the sampling term. This proves Corollary 1.

2.6 Proof of Theorem 2

Let $\hat{\delta}_{jk}^{\tau_{n,p}}$ be the (j,k) th entry of $\hat{\mathbf{\Delta}}_{\tau_{n,p}}$. Then

$$\text{pr}\left\{ \mathcal{M}\left(\hat{\mathbf{\Delta}}_{\tau_{n,p}} \right) = \mathcal{M}\left(\mathbf{\Delta}_0 \right) \right\} = \text{pr}\left[\left\{ \max_{j,k:\delta_{jk}^0 = 0} \left| \hat{\delta}_{jk}^{\tau_{n,p}} \right| = 0 \right\} \cap \left\{ \min_{j,k:\delta_{jk}^0 > 0} \hat{\delta}_{jk}^{\tau_{n,p}} > 0 \right\} \cap \left\{ \max_{j,k:\delta_{jk}^0 < 0} \hat{\delta}_{jk}^{\tau_{n,p}} < 0 \right\} \right].$$

By the definition of $\hat{\delta}_{jk}^{\tau_{n,p}}$ and the assumption $\min_{j,k:\delta_{jk}^0 \neq 0} \left| \delta_{jk}^0 \right| > 2\tau_{n,p}$.

(a) Suppose $\delta_{jk}^0 > 0$, then

$$\hat{\delta}_{jk} = \delta_{jk}^0 - \left(\delta_{jk}^0 - \hat{\delta}_{jk} \right) > 2\tau_{n,p} - \tau_{n,p}$$

with probability going to 1 , by Theorem 1 , so $\hat{\delta}_{jk}^{\tau_{n,p}} = \hat{\delta}_{jk} > 0$.

(b) Suppose $\delta_{jk}^0 < 0$, then

$$\hat{\delta}_{jk} = \delta_{jk}^0 - \left(\delta_{jk}^0 - \hat{\delta}_{jk} \right) < -2\tau_{n,p} + \tau_{n,p}$$

with probability going to 1 , so $\hat{\delta}_{jk}^{\tau_{n,p}} = \hat{\delta}_{jk} < 0$.

(*c*) Suppose $\delta_{jk}^0 = 0$, then

$$\left|\hat{\delta}_{jk}\right| = \left|\hat{\delta}_{jk} - \delta_{jk}^0\right| \leqslant \tau_{n,p}$$

with probability going to 1 , so $\hat{\delta}_{jk}^{\tau_{n,p}} = 0$.

This proves Theorem 2.

References

Boyd, S., N. Parikh, E. Chu, B. Peleato, J. Eckstein, et al. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning 3*(1), 1–122.

Cai, T. T., L. Wang, and G. Xu (2009). Shifting inequality and recovery of sparse signals. *IEEE Transactions on Signal processing 58*(3), 1300–1308.

Cai, T. T., G. Xu, and J. Zhang (2009). On recovery of sparse signals via ℓ_1 minimization. *IEEE Transactions on Information Theory 55*(7), 3388–3397.

Cao, Y., W. Lin, and H. Li (2019). Large covariance estimation for compositional data via composition-adjusted thresholding. *Journal of the American Statistical Association 114*(526), 759–772.

Lounici, K. (2008). Sup-norm convergence rate and sign concentration property of lasso and dantzig estimators. *Electronic Journal of Statistics 2*, 90–102.

Zhang, S., H. Wang, and W. Lin (2025). Care: Large precision matrix estimation for compositional data. *Journal of the American Statistical Association 120*(549), 305–317.

Zhao, S. D., T. T. Cai, and H. Li (2014). Direct estimation of differential networks. *Biometrika 101*(2), 253–268.