Biometrics 64, 1–20 DOI: 10.1111/j.1541-0420.2005.00454.x
December 2008

Supporting Information for "An Iterative Penalized Least Squares Approach to Sparse Canonical Correlation Analysis" by

Qing Mai

Department of Statistics, Florida State University, Tallahassee, FL, USA

email: mai@stat.fsu.edu

and

Xin Zhang

Department of Statistics, Florida State University, Tallahassee, FL, USA

email: henry@stat.fsu.edu

This paper has been submitted for consideration for publication in *Biometrics*

Web Appendix A: Algorithm for structured variable selection

By Lemma 3, we can develop the following iterative algorithm to perform structured variable selection in SCCA, where the penalty can be taken as the group lasso penalty, fused lasso penalty, or the adaptive lasso penalty.

ALGORITHM 1: An iterative penalized least squares algorithm for structured variable selection in SCCA:

(1) Given $\mathbf{A}_{k-1}, \mathbf{B}_{k-1}$, compute $\mathbf{\Omega}_k$ as in Section 2.2;

(2) Initialize $\{\widehat{\boldsymbol{\alpha}}_k^{(0)}, \widehat{\boldsymbol{\beta}}_k^{(0)}\}$;

(3) For $m = 1, \ldots$, repeat the following two steps until convergence:

(a) Set $\widetilde{\mathbf{Y}}_k^{(m)} = \mathbf{\Omega}_k^{\mathrm{T}} \mathbf{Y} \widehat{\boldsymbol{\alpha}}_k^{(m)}$. Compute

$$\breve{\boldsymbol{\beta}}_k^{(m)} = \arg\min_{\boldsymbol{\beta}_k} \left\{ \frac{1}{2n} \|\widetilde{\mathbf{Y}}_k^{(m)} - \mathbf{X}\boldsymbol{\beta}_k\|_2^2 + P_{\mathbf{X}}(\boldsymbol{\beta}_k) \right\}, \tag{1}$$

and then set $\widehat{\boldsymbol{\beta}}_k^{(m)} = [\{\breve{\boldsymbol{\beta}}_k^{(m)}\}^{\mathrm{T}} \widehat{\mathbf{\Sigma}}_{\mathbf{XX}} \breve{\boldsymbol{\beta}}_k^{(m)}]^{-1/2} \cdot \breve{\boldsymbol{\beta}}_k^{(m)}$.

(b) Set $\widetilde{\mathbf{X}}_k^{(m)} = \mathbf{\Omega}_k \mathbf{X} \widehat{\boldsymbol{\beta}}_k^{(m)}$. Compute

$$\breve{\boldsymbol{\alpha}}_k^{(m)} = \arg\min_{\boldsymbol{\alpha}_k} \left\{ \frac{1}{2n} \|\widetilde{\mathbf{X}}_k^{(m)} - \mathbf{Y}\boldsymbol{\alpha}_k\|_2^2 + P_{\mathbf{Y}}(\boldsymbol{\alpha}_k) \right\}, \tag{2}$$

and then set $\widehat{\boldsymbol{\alpha}}_k^{(m)} = [\{\breve{\boldsymbol{\alpha}}_k^{(m)}\}^{\mathrm{T}} \widehat{\mathbf{\Sigma}}_{\mathbf{YY}} \breve{\boldsymbol{\alpha}}_k^{(m)}]^{-1/2} \cdot \breve{\boldsymbol{\alpha}}_k^{(m)}$.

(4) Output $(\widehat{\boldsymbol{\alpha}}_k, \widehat{\boldsymbol{\beta}}_k)$ at convergence.

Web Appendix B: Technical conditions for Theorem 1

For a generic symmetric matrix $\mathbf{N}$, $\lambda_{\max}(\mathbf{N})$ denotes its largest eigenvalue, while $\lambda_{\min}(\mathbf{N})$ denotes its smallest eigenvalue. We also assume that $K = O(1)$. To prove Theorem 1, we need the following regularity conditions:

(C1) There exists $c > 0$ such that $\dfrac{\lambda_{\min}(\mathbf{\Sigma}_{\mathbf{YY}})}{\lambda_{\max}(\mathbf{\Sigma}_{\mathbf{YY}})} > c$, $\dfrac{\lambda_{\min}(\mathbf{\Sigma}_{\mathbf{XX}})}{\lambda_{\max}(\mathbf{\Sigma}_{\mathbf{XX}})} > c$.

(C2) There exists σ_0^2 such that for $j = 1, \ldots, p, l = 1, \ldots, q$, we have $Ee^{tX_j} \leqslant e^{\sigma_0^2 t^2}, Ee^{tY_l} \leqslant e^{\sigma_0^2 t^2}$ for any $t > 0$.

(C3) For $k = 1, \ldots, K$, there exists $\Delta > 0$ such that $\rho_k^* - \rho_{k+1}^* > \Delta$.

(C4) $n, p \to \infty$, $\log(p+q) = o(n)$ and $\tau_0 = o\left\{\sqrt{\frac{n}{\log(p+q)}}\right\}$.

These four conditions are very mild. Condition (C1) requires that the covariance matrices are bounded away from singular. This is slightly weaker than what is commonly assumed in the literature. For example, Gao et al. (2017) assumed that the eigenvalues of $\mathbf{\Sigma_{XX}}$ and $\mathbf{\Sigma_{YY}}$ are all bounded above and below, which implies that the ratio between the smallest eigenvalues and the largest eigenvalues have to be bounded away from 0, as required by Condition (C1).

Second, Condition (C2) requires that the predictors are sub-Gaussian so that each element in the covariance matrix can be estimated accurately. This is also a commonly used assumption. In Gao et al. (2017) it is assumed that $\mathbf{X}$ and $\mathbf{Y}$ are both normal, which implies that Condition (C2) must be true. Third, Condition (C3) guarantees that all the directions are identifiable. Note that if Condition (C3) is violated and there exist ρ_{k-1}^*, ρ_k^* such that $\rho_{k-1}^* = \rho_k^*$, then $(\boldsymbol{\alpha}_{k-1}^*, \boldsymbol{\beta}_{k-1}^*)$ and $(\boldsymbol{\alpha}_k^*, \boldsymbol{\beta}_k^*)$ are not identifiable.

Finally, Condition (C4) restricts both the dimensions and the sparsity in the true canonical pairs. The assumption $\log(p+q) = o(n)$ is a restriction on the dimensions with respect to the sample size. It is easy to see that the dimension can grow at an exponential rate of the sample size. Hence, SCCA is expected to work well even in ultra-high dimensions. The assumption $\tau_0 = o\left\{\sqrt{\frac{n}{\log(p+q)}}\right\}$ can be viewed as a sparsity assumption. It requires that the ℓ_1 norms of $\{\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k\}_{k=1}^K$ do not grow too fast. In general, most high-dimensional methods require this type of conditions to be consistent.

Web Appendix C: Additional simulations with $p = q = 600$

As suggested by a referee, we also report the simulation results for Models 1–4 in Section 4 with dimensions $p = q = 600$. These dimensions are slightly larger than the sample size $n = 500$. We continue to compare our SCCA with PMD and COLAR, but CCA is no longer included, as now $p, q > n$ and CCA is no longer applicable. The results are listed in Table 1. SCCA continues to have competitive performance. It outperforms the two competitors except in Model 3, where the difference between SCCA and Colar is not statistically significant. However, SCCA is much more computationally efficient.

[Table 1 about here.]

Web Appendix D: Details on the computation cost comparison

In comparing the computation costs, all computation was done on the same Windows 7 desktop computer with Intel® Xeon® Processor (Quad Core, 3.7 GHz Turbo, 10 MB), 16.00 GB memory, 64-bit Operating System. The implementation of SCCA in `R` is available in `code.zip`. We used the implementation of PMD in the `R` package `PMA` available at

https://cran.r-project.org/web/packages/PMA/index.html.

The `Matlab` implementation of COLAR was obtained from

http://www-stat.wharton.upenn.edu/ zongming/research.html.

Outside the base `R` packages, the `R` packages `PMA` (version 1.0.11), `impute` (version 1.54.0, https://www.bioconductor.org/packages/release/bioc/html/impute.html), `glmnet` (version 2.0-16, https://cran.r-project.org/web/packages/glmnet/index.html), and `mvtnorm` (version 1.0-7, https://cran.r-project.org/web/packages/mvtnorm/index.html) were employed in the computation. Among these packages, `impute` is never called by the user. Instead, it is only required to use the `PMA` package. The package `glmnet` solves the ℓ_1 penalized least squares problem in SCCA, and `mvtnorm` generates multivariate normal random variables in our simulations.

Web Appendix E: Proofs for the formulation

We prove Lemmas 1–3 in this section. We need the following proposition for the proof of Lemma 1.

PROPOSITION 1: For $\mathbf{u} \in \mathbb{R}^d$, and orthogonal vectors $\mathbf{e}_1, \ldots, \mathbf{e}_K \in \mathbb{R}^d$, $K \leqslant d$, we have $\sum_{k=1}^K \cos^2\{\angle(\mathbf{u}, \mathbf{e}_k)\} \leqslant 1$.

Proof. [Proof of Proposition 1] Augment $\mathbf{e}_1, \ldots, \mathbf{e}_K$ to an orthogonal basis of $\mathbb{R}^d$, denoted as $\mathbf{e}_1, \ldots, \mathbf{e}_K, \mathbf{e}_{K+1}, \ldots, \mathbf{e}_d$. It is well-known that $\sum_{k=1}^d \cos^2\{\angle(\mathbf{u}, \mathbf{e}_k)\} = 1$. Because $\cos^2\{\angle(\mathbf{u}, \mathbf{e}_k)\} \geqslant 0$ for any k, the conclusion follows.

We proceed to the proof of Lemma 1

Proof. [Proof of Lemma 1] We prove the conclusion by induction. Note that (2) is equivalent to

$$\begin{aligned}(\widehat{\boldsymbol{\alpha}}'_k, \widehat{\boldsymbol{\beta}}'_k) &= \arg\max_{\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k} \boldsymbol{\alpha}_k^{\mathrm{T}} \left[\widehat{\boldsymbol{\Sigma}}_{\mathbf{YX}} - \widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}} \left\{ \sum_{l<k} \widehat{\rho}'_l \widehat{\boldsymbol{\alpha}}'_l (\widehat{\boldsymbol{\beta}}'_l)^{\mathrm{T}} \right\} \widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}} \right] \boldsymbol{\beta}_k, \\ &\quad \text{s.t. } \boldsymbol{\alpha}_k^{\mathrm{T}} \widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}} \boldsymbol{\alpha}_k = 1,\ \boldsymbol{\beta}_k^{\mathrm{T}} \widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}} \boldsymbol{\beta}_k = 1.\end{aligned}$$

Hence, it is easy to see that the conclusion holds when $k = 1$. Assume that the conclusion holds for $k \leqslant m-1$. For $k = m$, we note that $\widehat{\boldsymbol{\Sigma}}_{\mathbf{YX}} = \widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}} \left\{ \sum_{l=1}^{\min(p,q)} \widehat{\rho}_l \widehat{\boldsymbol{\alpha}}_l^{\mathrm{CCA}} (\widehat{\boldsymbol{\beta}}_l^{\mathrm{CCA}})^{\mathrm{T}} \right\} \widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}$, where $\widehat{\rho}_l = (\widehat{\boldsymbol{\alpha}}_l^{\mathrm{CCA}})^{\mathrm{T}} \widehat{\boldsymbol{\Sigma}}_{\mathbf{YX}} \widehat{\boldsymbol{\beta}}_l^{\mathrm{CCA}}$. We must have that $\widehat{\rho}_l$ are distinct with a probability of 1.

Also, $\widehat{\rho}_l = \widehat{\rho}'_l$ for $l < m$. It follows that, for any $\boldsymbol{\alpha}_m \in \mathbb{R}^q$, $\boldsymbol{\beta}_m \in \mathbb{R}^p$,

$$\begin{aligned}
&\boldsymbol{\alpha}_m^{\mathrm{T}}[\widehat{\boldsymbol{\Sigma}}_{\mathbf{YX}} - \widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}\{\sum_{l<k}\widehat{\rho}_l\widehat{\boldsymbol{\alpha}}_l^{\mathrm{CCA}}(\widehat{\boldsymbol{\beta}}_l^{\mathrm{CCA}})^{\mathrm{T}}\}\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}]\boldsymbol{\beta}_m \\
= \;& \boldsymbol{\alpha}_m^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}\left\{\sum_{l=m}^{\min(p,q)}\widehat{\rho}_l\widehat{\boldsymbol{\alpha}}_l^{\mathrm{CCA}}(\widehat{\boldsymbol{\beta}}_l^{\mathrm{CCA}})^{\mathrm{T}}\right\}\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}\boldsymbol{\beta}_m \\
= \;& \sum_{l=m}^{\min(p,q)}\widehat{\rho}_l\cos\{\angle(\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}^{1/2}\widehat{\boldsymbol{\alpha}}_l^{\mathrm{CCA}}, \widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}^{1/2}\boldsymbol{\alpha}_m)\}\cos\{\angle(\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}^{1/2}\widehat{\boldsymbol{\beta}}_l^{\mathrm{CCA}}, \widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}^{1/2}\boldsymbol{\beta}_m\} \\
\leqslant \;& \sum_{l=m}^{\min(p,q)}\widehat{\rho}_l[\cos^2\{\angle(\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}^{1/2}\widehat{\boldsymbol{\alpha}}_l^{\mathrm{CCA}}, \widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}^{1/2}\boldsymbol{\alpha}_m)\} + \cos^2\{\angle(\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}^{1/2}\widehat{\boldsymbol{\beta}}_l^{\mathrm{CCA}}, \widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}^{1/2}\boldsymbol{\beta}_m\}]/2 \\
\leqslant \;& \widehat{\rho}_m[\cos^2\{\angle(\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}^{1/2}\widehat{\boldsymbol{\alpha}}_m^{\mathrm{CCA}}, \widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}^{1/2}\boldsymbol{\alpha}_m)\} + \cos^2\{\angle(\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}^{1/2}\widehat{\boldsymbol{\beta}}_m^{\mathrm{CCA}}, \widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}^{1/2}\boldsymbol{\beta}_m\}]/2 \\
&+\widehat{\rho}_{m+1}\sum_{l=m+1}^{\min(p,q)}[\cos^2\{\angle(\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}^{1/2}\widehat{\boldsymbol{\alpha}}_l^{\mathrm{CCA}}, \widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}^{1/2}\boldsymbol{\alpha}_m)\} + \cos^2\{\angle(\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}^{1/2}\widehat{\boldsymbol{\beta}}_l^{\mathrm{CCA}}, \widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}^{1/2}\boldsymbol{\beta}_m\}]/2 \quad (3) \\
\leqslant \;& \widehat{\rho}_m\left[\cos^2\{\angle(\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}^{1/2}\widehat{\boldsymbol{\alpha}}_l^{\mathrm{CCA}}, \widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}^{1/2}\boldsymbol{\alpha}_m)\} + \cos^2\{\angle(\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}^{1/2}\widehat{\boldsymbol{\beta}}_l^{\mathrm{CCA}}, \widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}^{1/2}\boldsymbol{\beta}_m\}\right]/2 \\
&+\widehat{\rho}_{m+1}\left[2 - \cos^2\{\angle(\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}^{1/2}\widehat{\boldsymbol{\alpha}}_l^{\mathrm{CCA}}, \widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}^{1/2}\boldsymbol{\alpha}_m)\} - \cos^2\{\angle(\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}^{1/2}\widehat{\boldsymbol{\beta}}_l^{\mathrm{CCA}}, \widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}^{1/2}\boldsymbol{\beta}_m)\}\right]/2 \quad (4) \\
\leqslant \;& \widehat{\rho}_m,
\end{aligned}$$

where we use Proposition 1 to obtain (4) from (3) by noting that $\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}^{1/2}\widehat{\boldsymbol{\alpha}}_l^{\mathrm{CCA}}$ are orthogonal to each other and so are $\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}^{1/2}\widehat{\boldsymbol{\beta}}_l^{\mathrm{CCA}}$. Also note that the equality holds if and only if $\cos^2\left\{\angle(\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}^{1/2}\widehat{\boldsymbol{\alpha}}_m^{\mathrm{CCA}}, \widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}^{1/2}\boldsymbol{\alpha}_m)\right\} = \cos^2\left\{\angle(\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}^{1/2}\widehat{\boldsymbol{\beta}}_m^{\mathrm{CCA}}, \widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}^{1/2}\boldsymbol{\beta}_m)\right\} = 1$. The conclusion follows.

Because Lemma 2 is a special case of Lemma 3, we only prove Lemma 3. Lemma 3 is a consequence of the following Lemma 4.

LEMMA 4: *For any $\{\widetilde{Y}_i\}_{i=1}^n$ and a penalty function $P(\cdot)$ that satisfies Condition (C0), define $J(\boldsymbol{\beta}) = \dfrac{1}{2n}\sum_{i=1}^n\{\widetilde{Y}_i - (\mathbf{X}_i)^{\mathrm{T}}\boldsymbol{\beta}\}^2 + P(\boldsymbol{\beta})$.*

If $\breve{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}}\{J(\boldsymbol{\beta})\}$ is the unconstrained minimizer, and

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} J(\boldsymbol{\beta}), \;\; s.t. \; \boldsymbol{\beta}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}\boldsymbol{\beta} = 1, \tag{5}$$

then we must have $\widehat{\boldsymbol{\beta}} = \{\breve{\boldsymbol{\beta}}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}\breve{\boldsymbol{\beta}}\}^{-1/2}\breve{\boldsymbol{\beta}}$.

Proof. [Proof of Lemma 4] First, note that

$$\frac{1}{2n}\sum_{i=1}^{n}\{\widetilde{Y}_i - (\mathbf{X}_i)^{\mathrm{T}}\boldsymbol{\beta}\}^2 + P(\boldsymbol{\beta}) = \frac{1}{2n}\sum_{i=1}^{n}(\widetilde{Y}_i)^2 - \frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}_i(\mathbf{X}_i)^{\mathrm{T}}\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{\beta}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}\boldsymbol{\beta} + P(\boldsymbol{\beta}).$$

It follows that

$$\breve{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}}\left\{-\frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}_i(\mathbf{X}_i)^{\mathrm{T}}\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{\beta}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}\boldsymbol{\beta} + P(\boldsymbol{\beta})\right\}.$$

Define $c = \sqrt{\breve{\boldsymbol{\beta}}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}\breve{\boldsymbol{\beta}}}$. Then we have that $\breve{\boldsymbol{\beta}}$ also minimizes $-\frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}_i(\mathbf{X}_i)^{\mathrm{T}}\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{\beta}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}\boldsymbol{\beta} + \lambda\|\boldsymbol{\beta}\|_1$ over the set $\boldsymbol{\beta}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}\boldsymbol{\beta} = c^2$. Consequently,

$$\breve{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}}\{-\frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}_i(\mathbf{X}_i)^{\mathrm{T}}\boldsymbol{\beta} + P(\boldsymbol{\beta})\},\ \text{s.t. } \boldsymbol{\beta}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}\boldsymbol{\beta} = c^2.$$

Similarly, we can show that (5) is equivalent to

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}}\{-\frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}_i(\mathbf{X}_i)^{\mathrm{T}}\boldsymbol{\beta} + P(\boldsymbol{\beta})\},\ \text{s.t. } \boldsymbol{\beta}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}\boldsymbol{\beta} = 1.$$

Define $\widetilde{J}(\boldsymbol{\beta}) = -\frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}^i(\mathbf{X}^i)^{\mathrm{T}}\boldsymbol{\beta} + P(\boldsymbol{\beta})$. Then

$$\widetilde{J}(\widehat{\boldsymbol{\beta}}) \leqslant \widetilde{J}(c^{-1}\breve{\boldsymbol{\beta}}) = c^{-1}\widetilde{J}(\breve{\boldsymbol{\beta}}) \leqslant c^{-1}\widetilde{J}(c\widehat{\boldsymbol{\beta}}) = \widetilde{J}(\widehat{\boldsymbol{\beta}}),$$

where we use the fact $P(c\boldsymbol{\beta}) = cP(\boldsymbol{\beta})$. It follows that $\widetilde{J}(c^{-1}\breve{\boldsymbol{\beta}}) = \widetilde{J}(\widehat{\boldsymbol{\beta}})$. Combine this result with $(c^{-1}\breve{\boldsymbol{\beta}})^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}(c^{-1}\breve{\boldsymbol{\beta}}) = 1$ and we have the desired conclusion.

Now we turn to the proof of Lemma 3.

Proof. [Proof of Lemma 3] Note that

$$\begin{aligned} & \frac{1}{2n}\sum_{i=1}^{n}(\mathbf{Y}_i^{\mathrm{T}}\boldsymbol{\alpha}_k - \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}_k)^2 + \sum_{l<k}\widehat{\rho}_l\boldsymbol{\alpha}_k^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}\widehat{\boldsymbol{\alpha}}_l\cdot\boldsymbol{\beta}_k^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}\widehat{\boldsymbol{\beta}}_l \\ = & \frac{1}{2}\boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{Y}^{\mathrm{T}}(\mathbf{I} - \boldsymbol{\Omega}_k)^{\mathrm{T}}\mathbf{Y}\boldsymbol{\alpha}_k + \frac{1}{2n}\|\boldsymbol{\Omega}_k^{\mathrm{T}}\mathbf{Y}\boldsymbol{\alpha}_k - \mathbf{X}\boldsymbol{\beta}_k\|_2^2. \end{aligned}$$

Because the first term $\boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{Y}^{\mathrm{T}}(\mathbf{I} - \boldsymbol{\Omega}_k)^{\mathrm{T}}\mathbf{Y}\boldsymbol{\alpha}_k$ does not involve $\boldsymbol{\beta}_k$, we have that

$$\breve{\boldsymbol{\beta}}_k = \arg\min_{\boldsymbol{\beta}_k}\left\{\frac{1}{2n}\|\boldsymbol{\Omega}_k^{\mathrm{T}}\mathbf{Y}\boldsymbol{\alpha}_k - \mathbf{X}\boldsymbol{\beta}_k\|_2^2 + \lambda_{\boldsymbol{\beta}_k}P_{\mathbf{X}}(\boldsymbol{\beta}_k)\right\},\ \text{s.t. } \boldsymbol{\beta}_k^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}\boldsymbol{\beta}_k = 1.$$

Combine this fact with Lemma 4 and the first conclusion follows. The second conclusion can be proven similarly.

Web Appendix F: Technical lemmas

To show Theorem 1, we first collect some technical lemmas. In the following proofs, define $\mathbf{M}_k = \boldsymbol{\Sigma}_{\mathbf{YX}} - \boldsymbol{\Sigma}_{\mathbf{YY}}\{\sum_{i=1}^{k-1} \rho_i^* \boldsymbol{\alpha}_i^* (\boldsymbol{\beta}_i^*)^{\mathrm{T}}\}\boldsymbol{\Sigma}_{\mathbf{XX}}$. The following Lemma 5 guarantees that if a pair $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ produces a high canonical correlation, they have to have small angles with the true directions after some rotation.

Lemma 5: *Under Condition (C3), for any* $\boldsymbol{\alpha}, \boldsymbol{\beta}$ *such that* $\dfrac{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YX}}\boldsymbol{\beta}}{\sqrt{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}} \cdot \sqrt{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta}}} \geqslant \rho_1^* - \epsilon$, *then* $\cos^2(\angle \boldsymbol{\Sigma}_{\mathbf{YY}}^{1/2}\boldsymbol{\alpha}, \boldsymbol{\Sigma}_{\mathbf{YY}}^{1/2}\boldsymbol{\alpha}_1^*) \geqslant 1 - \frac{2\epsilon}{\Delta}$ *and* $\cos^2(\angle \boldsymbol{\Sigma}_{\mathbf{XX}}^{1/2}\boldsymbol{\beta}, \boldsymbol{\Sigma}_{\mathbf{XX}}^{1/2}\boldsymbol{\beta}_1^*) \geqslant 1 - \frac{2\epsilon}{\Delta}$.

Proof. [Proof of Lemma 5] We only show the conclusion for $\boldsymbol{\alpha}$, since that for $\boldsymbol{\beta}$ can be shown in the same way.

Note that

$$
\begin{aligned}
& \frac{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YX}}\boldsymbol{\beta}}{\sqrt{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}} \cdot \sqrt{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta}}} = \frac{\sum_{i=1}^{K} \rho_i^* (\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_i^*) \cdot (\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta}_i^*)}{\sqrt{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}} \cdot \sqrt{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta}}} \\
= & \sum_{i=1}^{K} \rho_i^* \cos\left\{\angle(\boldsymbol{\Sigma}_{\mathbf{YY}}^{1/2}\boldsymbol{\alpha}, \boldsymbol{\Sigma}_{\mathbf{YY}}^{1/2}\boldsymbol{\alpha}_i^*)\right\} \cdot \cos\left\{\angle(\boldsymbol{\Sigma}_{\mathbf{XX}}^{1/2}\boldsymbol{\beta}, \boldsymbol{\Sigma}_{\mathbf{XX}}^{1/2}\boldsymbol{\beta}_i^*)\right\} \\
\leqslant & \sum_{i=1}^{K} \frac{\rho_i^*}{2} \left[\cos^2\left\{\angle(\boldsymbol{\Sigma}_{\mathbf{YY}}^{1/2}\boldsymbol{\alpha}, \boldsymbol{\Sigma}_{\mathbf{YY}}^{1/2}\boldsymbol{\alpha}_i^*)\right\} + \cos^2\left\{\angle(\boldsymbol{\Sigma}_{\mathbf{XX}}^{1/2}\boldsymbol{\beta}, \boldsymbol{\Sigma}_{\mathbf{XX}}^{1/2}\boldsymbol{\beta}_i^*)\right\}\right] \\
\leqslant & \frac{\rho_1^*}{2} \left[\cos^2\left\{\angle(\boldsymbol{\Sigma}_{\mathbf{YY}}^{1/2}\boldsymbol{\alpha}, \boldsymbol{\Sigma}_{\mathbf{YY}}^{1/2}\boldsymbol{\alpha}_1^*\right\}\right] + \cos^2\left[\angle\{\boldsymbol{\Sigma}_{\mathbf{XX}}^{1/2}\boldsymbol{\beta}, \boldsymbol{\Sigma}_{\mathbf{XX}}^{1/2}\boldsymbol{\beta}_1^*)\right\}\Big] \\
& + \frac{\rho_1^* - \Delta}{2} \sum_{i=2}^{K} \left[\cos^2\left\{\angle(\boldsymbol{\Sigma}_{\mathbf{YY}}^{1/2}\boldsymbol{\alpha}, \boldsymbol{\Sigma}_{\mathbf{YY}}^{1/2}\boldsymbol{\alpha}_i^*)\right\} + \cos^2\left\{\angle(\boldsymbol{\Sigma}_{\mathbf{XX}}^{1/2}\boldsymbol{\beta}, \boldsymbol{\Sigma}_{\mathbf{XX}}^{1/2}\boldsymbol{\beta}_i^*)\right\}\right] \qquad (6) \\
\leqslant & \frac{\rho_1^*}{2} \left[\cos^2\left\{\angle(\boldsymbol{\Sigma}_{\mathbf{YY}}^{1/2}\boldsymbol{\alpha}, \boldsymbol{\Sigma}_{\mathbf{YY}}^{1/2}\boldsymbol{\alpha}_1^*)\right\} + \cos^2\left\{\angle(\boldsymbol{\Sigma}_{\mathbf{XX}}^{1/2}\boldsymbol{\beta}, \boldsymbol{\Sigma}_{\mathbf{XX}}^{1/2}\boldsymbol{\beta}_1^*)\right\}\right] \\
& + \frac{\rho_1^* - \Delta}{2} \left[1 - \cos^2\left\{\angle(\boldsymbol{\Sigma}_{\mathbf{YY}}^{1/2}\boldsymbol{\alpha}, \boldsymbol{\Sigma}_{\mathbf{YY}}^{1/2}\boldsymbol{\alpha}_1^*)\right\} + 1 - \cos^2\left\{\angle(\boldsymbol{\Sigma}_{\mathbf{XX}}^{1/2}\boldsymbol{\beta}, \boldsymbol{\Sigma}_{\mathbf{XX}}^{1/2}\boldsymbol{\beta}_1^*)\right\}\right] \qquad (7) \\
= & \rho_1^* - \frac{\Delta}{2} \left[1 - \cos^2\left\{\angle(\boldsymbol{\Sigma}_{\mathbf{YY}}^{1/2}\boldsymbol{\alpha}, \boldsymbol{\Sigma}_{\mathbf{YY}}^{1/2}\boldsymbol{\alpha}_1^*)\right\} + 1 - \cos^2\left\{\angle(\boldsymbol{\Sigma}_{\mathbf{XX}}^{1/2}\boldsymbol{\beta}, \boldsymbol{\Sigma}_{\mathbf{XX}}^{1/2}\boldsymbol{\beta}_1^*)\right\}\right] \\
\leqslant & \rho_1^* - \frac{\Delta}{2} \left[1 - \cos^2\left\{\angle(\boldsymbol{\Sigma}_{\mathbf{YY}}^{1/2}\boldsymbol{\alpha}, \boldsymbol{\Sigma}_{\mathbf{YY}}^{1/2}\boldsymbol{\alpha}_1^*)\right\}\right],
\end{aligned}
$$

where we use Proposition 1 to obtain (7) from (6). Now because $\dfrac{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YX}}\boldsymbol{\beta}}{\sqrt{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}} \cdot \sqrt{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta}}} \geqslant \rho_1^* - \epsilon$, we have $\cos^2\left(\angle(\boldsymbol{\Sigma}_{\mathbf{YY}}^{1/2}\boldsymbol{\alpha}, \boldsymbol{\Sigma}_{\mathbf{YY}}^{1/2}\boldsymbol{\alpha}_1^*)\right) \geqslant 1 - \frac{2\epsilon}{\Delta}$.

Lemma 5 concerns the angle between $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}_1^*$ after the multiplication by $\mathbf{\Sigma}_{\mathbf{YY}}^{1/2}$. Lemma 6 further translates such results to be about the vectors themselves.

LEMMA 6: *For two vectors* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$ *and a positive definite matrix* $\mathbf{\Sigma} \in \mathbb{R}^{d\times d}$*, define* $\mu_{\mathbf{\Sigma}} = \angle(\mathbf{\Sigma}^{1/2}\mathbf{u}, \mathbf{\Sigma}^{1/2}\mathbf{v})$ *and* $\mu_{\mathbf{I}} = \angle(\mathbf{u}, \mathbf{v})$*. If* $\mathbf{u}^{\mathrm{T}}\mathbf{\Sigma}\mathbf{u} = 1$,$\mathbf{v}^{\mathrm{T}}\mathbf{\Sigma}\mathbf{v} = 1$ *and* $\cos\mu_{\mathbf{\Sigma}} > 1 - \epsilon$*, then* $\cos\mu_{\mathbf{I}} > 1 - \frac{\lambda_{\max}(\mathbf{\Sigma})\epsilon}{\lambda_{\min}(\mathbf{\Sigma})}$.

Proof. [Proof of Lemma 6] Under our assumptions, we have

$$\begin{aligned} \lambda_{\min}(\mathbf{\Sigma})(\mathbf{u}-\mathbf{v})^{\mathrm{T}}(\mathbf{u}-\mathbf{v}) &\leqslant (\mathbf{u}-\mathbf{v})^{\mathrm{T}}\mathbf{\Sigma}(\mathbf{u}-\mathbf{v}) \\ &= \mathbf{u}^{\mathrm{T}}\mathbf{\Sigma}\mathbf{u} - 2\mathbf{u}^{\mathrm{T}}\mathbf{\Sigma}\mathbf{v} + \mathbf{v}^{\mathrm{T}}\mathbf{\Sigma}\mathbf{v} = 2 - 2\cos\mu_{\mathbf{\Sigma}} \leqslant 2\epsilon, \\ \lambda_{\max}(\mathbf{\Sigma})\mathbf{u}^{\mathrm{T}}\mathbf{u} &\geqslant \mathbf{u}^{\mathrm{T}}\mathbf{\Sigma}\mathbf{u} = 1, \quad \lambda_{\max}(\mathbf{\Sigma})\mathbf{v}^{\mathrm{T}}\mathbf{v} \geqslant \mathbf{v}^{\mathrm{T}}\mathbf{\Sigma}\mathbf{v} = 1. \end{aligned}$$

Consequently,

$$(\mathbf{u}-\mathbf{v})^{\mathrm{T}}(\mathbf{u}-\mathbf{v}) \leqslant \frac{2\epsilon}{\lambda_{\min}(\mathbf{\Sigma})}, \quad \mathbf{u}^{\mathrm{T}}\mathbf{u} \geqslant 1/\lambda_{\max}(\mathbf{\Sigma}), \quad \mathbf{v}^{\mathrm{T}}\mathbf{v} \geqslant 1/\lambda_{\max}(\mathbf{\Sigma}).$$

Now

$$\begin{aligned} \frac{2\lambda_{\max}(\mathbf{\Sigma})\epsilon}{\lambda_{\min}(\mathbf{\Sigma})} &\geqslant \frac{2\epsilon}{\lambda_{\min}(\mathbf{\Sigma})\sqrt{\mathbf{u}^{\mathrm{T}}\mathbf{u}}\sqrt{\mathbf{v}^{\mathrm{T}}\mathbf{v}}} \\ &\geqslant \frac{(\mathbf{u}-\mathbf{v})^{\mathrm{T}}(\mathbf{u}-\mathbf{v})}{\sqrt{\mathbf{u}^{\mathrm{T}}\mathbf{u}}\sqrt{\mathbf{v}^{\mathrm{T}}\mathbf{v}}} = \frac{\sqrt{\mathbf{u}^{\mathrm{T}}\mathbf{u}}}{\sqrt{\mathbf{v}^{\mathrm{T}}\mathbf{v}}} + \frac{\sqrt{\mathbf{v}^{\mathrm{T}}\mathbf{v}}}{\sqrt{\mathbf{u}^{\mathrm{T}}\mathbf{u}}} - 2\frac{\mathbf{u}^{\mathrm{T}}\mathbf{v}}{\sqrt{\mathbf{u}^{\mathrm{T}}\mathbf{u}}\sqrt{\mathbf{v}^{\mathrm{T}}\mathbf{v}}} \\ &\geqslant 2 - 2\cos\mu_{\mathbf{I}}. \end{aligned}$$

The conclusion follows.

Lemma 7 contains several intermediate inequalities that will facilitate our proof of the lemmas leading to Theorem 1.

LEMMA 7: *For any* $\boldsymbol{\alpha}, \boldsymbol{\beta}$ *such that* $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha} = 1, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta} = 1$*, define*

$$\min\{(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_k^*)^2, (\boldsymbol{\beta}^{\mathrm{T}}\mathbf{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta}_k^*)^2\} = 1 - a.$$

If $\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_k^* > 0$*, and* $\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta}^* > 0$*, under Condition (C3), we have*

$$\begin{aligned}
\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{M}_k\boldsymbol{\beta} &\leqslant \rho_k^* - \frac{\Delta}{2}a, && (8)\\
\max\{(\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^*)^2, (\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta}_l^*)^2\} &\leqslant 2a, \quad \textit{for } l \neq k, && (9)\\
\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YX}}\boldsymbol{\beta} &\geqslant \rho_k^* - (2K+1)\rho_1^* a. && (10)
\end{aligned}$$

Proof. [Proof of Lemma 7] We first show (8). Without loss of generality, we assume that $(\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_k^*)^2 = 1 - a \leqslant (\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta}_k^*)^2$. Also note that, because $\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha} = 1, \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta} = 1$, we have $(\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^*)^2 = \cos^2\{\angle(\boldsymbol{\Sigma}_{\mathbf{YY}}^{1/2}\boldsymbol{\alpha}, \boldsymbol{\Sigma}_{\mathbf{YY}}^{1/2}\boldsymbol{\alpha}_l^*)\}, (\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta}_l^*)^2 = \cos^2\{\angle(\boldsymbol{\Sigma}_{\mathbf{XX}}^{1/2}\boldsymbol{\beta}, \boldsymbol{\Sigma}_{\mathbf{XX}}^{1/2}\boldsymbol{\beta}_l^*)\}$. By Proposition 1, we have

$$\begin{aligned}
\sum_{l=k+1}^{K}\{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}(\boldsymbol{\alpha}_l^*)^{\mathrm{T}}\}^2 &\leqslant 1 - \{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}(\boldsymbol{\alpha}_k^*)^{\mathrm{T}}\}^2 = 1 - a,\\
\sum_{l=k}^{K}\{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}(\boldsymbol{\beta}_l^*)^{\mathrm{T}}\}^2 &\leqslant 1 - \{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}(\boldsymbol{\beta}_k^*)^{\mathrm{T}}\}^2 \leqslant 1.
\end{aligned}$$

Note that

$$\begin{aligned}
\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{M}_k\boldsymbol{\beta} &= \sum_{l=k}^{K}\rho_l^*\{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^* \cdot (\boldsymbol{\beta}_l^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta}\}\\
&\leqslant \frac{1}{2}\sum_{l=k}^{K}\rho_l^*\{(\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^*)^2 + (\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta}_l^*)^2\}\\
&\leqslant \frac{1}{2}\sum_{l=k}^{K}\rho_l^*(\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^*)^2 + \frac{1}{2}\rho_k^*\sum_{l=k}^{K}(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta}_l^*)^2\\
&\leqslant \frac{1}{2}\rho_k^*(\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_k^*)^2 + \frac{1}{2}\sum_{l=k+1}^{K}\rho_l^*(\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^*)^2 + \frac{1}{2}\rho_k^*\\
&\leqslant \frac{1}{2}\rho_k^*(1-a) + \frac{1}{2}(\rho_k^* - \Delta)a + \frac{1}{2}\rho_k^* = \rho_k^* - \frac{\Delta}{2}a.
\end{aligned}$$

Now we show (9). Combining the Cauchy-Schwarz inequality with the fact $(\boldsymbol{\alpha}_k^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^* = 0$ and $(\boldsymbol{\alpha}_l^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^* = 1$, we have, for $l \neq k$,

$$\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^* = (\boldsymbol{\alpha} - \boldsymbol{\alpha}_k^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^* \leqslant \sqrt{(\boldsymbol{\alpha} - \boldsymbol{\alpha}_k^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}(\boldsymbol{\alpha} - \boldsymbol{\alpha}_k^*)}.$$

Therefore,

$$
\begin{aligned}
(\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^*)^2 &\leqslant (\boldsymbol{\alpha}-\boldsymbol{\alpha}_k^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}(\boldsymbol{\alpha}-\boldsymbol{\alpha}_k^*)\\
&= \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha} - 2\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_k^* + (\boldsymbol{\alpha}_k^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_k^*\\
&\leqslant 2-2\sqrt{1-a} = \frac{4-4(1-a)}{2+2\sqrt{1-a}} \leqslant 2a.
\end{aligned}
$$

For (10), note that

$$
\begin{aligned}
\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YX}}\boldsymbol{\beta} &= \sum_{l=1}^{K}\rho_l^*\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^*\cdot(\boldsymbol{\beta}_l^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta}\\
&= \rho_k^*\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_k^*\cdot(\boldsymbol{\beta}_k^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta} + \sum_{l\neq k}\rho_l^*\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^*\cdot(\boldsymbol{\beta}_l^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta} \qquad (11)\\
&\geqslant \rho_k^*(1-a) - \sum_{l\neq k}\rho_l^*(2a) \qquad (12)\\
&\geqslant \rho_k^*(1-a) - 2K\rho_1^*a\\
&\geqslant \rho_k^* - (2K+1)\rho_1^*a,
\end{aligned}
$$

where we apply (9) to obtain (12) from (11).

Web Appendix G: Proofs for consistent estimation

For a matrix $\mathbf{N}\in\mathbb{R}^{p_1\times p_2}$, denote $\|\mathbf{N}\|_{\max} = \max_{i,j}|N_{ij}|$. Throughout the proof, we use C to denote a generic positive constant that could vary from line to line. We further assume that, for some $\epsilon>0$, we have $\|\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}-\boldsymbol{\Sigma}_{\mathbf{YY}}\|_{\max}\leqslant\epsilon$, $\|\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}-\boldsymbol{\Sigma}_{\mathbf{XX}}\|_{\max}\leqslant\epsilon$, $\|\widehat{\boldsymbol{\Sigma}}_{\mathbf{YX}}-\boldsymbol{\Sigma}_{\mathbf{YX}}\|_{\max}\leqslant\epsilon$. By Bickel and Levina (2008), we have that such an event happens with a probability no less than $1-(p+q)^2\exp(-Cn\epsilon^2)$ under Condition (C2).

As in Web Appendix E, we denote $\mathbf{M}_k = \boldsymbol{\Sigma}_{\mathbf{YX}} - \boldsymbol{\Sigma}_{\mathbf{YY}}\left\{\sum_{l<k}\rho_l^*\boldsymbol{\alpha}_l^*(\boldsymbol{\beta}_l^*)^{\mathrm{T}}\right\}\boldsymbol{\Sigma}_{\mathbf{XX}}$. We note that, strictly speaking, $(\boldsymbol{\alpha}_l^*,\boldsymbol{\beta}_l^*)$ are not identifiable even on the population level, because $(-\boldsymbol{\alpha}_l^*,-\boldsymbol{\beta}_l^*)$ can also be viewed as a canonical pair. For simplicity, we require the first nonzero element of $\boldsymbol{\alpha}_l^*$ to be positive. We also define

$$\widetilde{\boldsymbol{\alpha}}_k^* = \{(\boldsymbol{\alpha}_k^*)^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}\boldsymbol{\alpha}_k^*\}^{-1/2}\boldsymbol{\alpha}_k^*, \tag{13}$$

$$\widetilde{\boldsymbol{\beta}}_k^* = \{(\boldsymbol{\beta}_k^*)^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}\boldsymbol{\beta}_k^*\}^{-1/2}\boldsymbol{\beta}_k^*. \tag{14}$$

Denote $\tau_0 = \max_{k=1,\ldots,K}\{\|\boldsymbol{\alpha}_k^*\|_1, \|\boldsymbol{\beta}_k^*\|_1\}$. We have the following conclusions for $\widetilde{\boldsymbol{\alpha}}_k^*, \widetilde{\boldsymbol{\beta}}_k^*$.

LEMMA 8: *(1) We have*

$$|(\boldsymbol{\alpha}_k^*)^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}\boldsymbol{\alpha}_k^* - 1| \leqslant \tau_0^2\epsilon, \quad |(\boldsymbol{\beta}_k^*)^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}\boldsymbol{\beta}_k^* - 1| \leqslant \tau_0^2\epsilon. \tag{15}$$

(2) If $\epsilon\tau_0^2 < 3/4$, we further have

$$\|\widetilde{\boldsymbol{\alpha}}_k^*\|_1 \leqslant 2\tau_0, \quad \|\widetilde{\boldsymbol{\beta}}_k^*\|_1 \leqslant 2\tau_0, \tag{16}$$

$$(\widetilde{\boldsymbol{\alpha}}_k^*)^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{YX}}\widetilde{\boldsymbol{\beta}}_k^* \geqslant \rho_k^* - C\tau_0^2\epsilon. \tag{17}$$

Proof. [Proof of Lemma 8] For (15), note that

$$|(\boldsymbol{\alpha}_k^*)^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}\boldsymbol{\alpha}_k^* - 1| = |(\boldsymbol{\alpha}_k^*)^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}\boldsymbol{\alpha}_k^* - (\boldsymbol{\alpha}_k^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_k^*| = |(\boldsymbol{\alpha}_k^*)^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}} - \boldsymbol{\Sigma}_{\mathbf{YY}})\boldsymbol{\alpha}_k^*| \leqslant \tau_0^2\epsilon.$$

For (16), note that

$$(\boldsymbol{\alpha}_k^*)^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}\boldsymbol{\alpha}_k^* = (\boldsymbol{\alpha}_k^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_k^* + (\boldsymbol{\alpha}_k^*)^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}} - \boldsymbol{\Sigma}_{\mathbf{YY}})\boldsymbol{\alpha}_k^* \geqslant 1 - \tau_0^2\epsilon$$

where we use the first conclusion.

Therefore, if $\epsilon\tau_0^2 < 3/4$,

$$\|\widetilde{\boldsymbol{\alpha}}_k^*\|_1 \leqslant \frac{\tau_0}{\sqrt{1-\tau_0^2\epsilon}} \leqslant 2\tau_0.$$

Similarly, we have $\|\widetilde{\boldsymbol{\beta}}_k^*\|_1 \leqslant 2\tau_0$.

For (17), note that

$$(\widetilde{\boldsymbol{\alpha}}_k^*)^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{YX}}\widetilde{\boldsymbol{\beta}}_k^* = (\widetilde{\boldsymbol{\alpha}}_k^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YX}}\widetilde{\boldsymbol{\beta}}_k^* + (\widetilde{\boldsymbol{\alpha}}_k^*)^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{YX}} - \boldsymbol{\Sigma}_{\mathbf{YX}})\widetilde{\boldsymbol{\beta}}_k^* \geqslant \rho_k^*(1 - 4\tau_0^2\epsilon) - 4\tau_0^2\epsilon.$$

where we apply (16).

For ease of presentation, for some $N > 2$ that will be specified later, we define $\tau = N\tau_0$.

We also define

$$(\widehat{\boldsymbol{\alpha}}_k^\tau, \widehat{\boldsymbol{\beta}}_k^\tau) = \arg\min_{\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k} \left[\frac{1}{2n} \sum_{i=1}^n (\mathbf{Y}_i^{\mathrm{T}} \boldsymbol{\alpha}_k - \mathbf{X}_i^{\mathrm{T}} \boldsymbol{\beta}_k)^2 + \boldsymbol{\alpha}_k \left\{ \sum_{l<k} \widehat{\rho}_l \widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}} \widehat{\boldsymbol{\alpha}}_l^\tau \cdot (\widehat{\boldsymbol{\beta}}_l^\tau)^{\mathrm{T}} \widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}} \right\} \boldsymbol{\beta}_k \right]$$
$$+\lambda \|\boldsymbol{\alpha}_k\|_1 + \lambda \|\boldsymbol{\beta}_k\|_1, \tag{18}$$
$$\text{s.t. } \boldsymbol{\alpha}_k^{\mathrm{T}} \widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}} \boldsymbol{\alpha}_k = 1, \boldsymbol{\beta}_k^{\mathrm{T}} \widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}} \boldsymbol{\beta}_k = 1, \|\boldsymbol{\alpha}_k\|_1 \leqslant \tau, \|\boldsymbol{\beta}_k\|_1 \leqslant \tau, \boldsymbol{\alpha}_k^{\mathrm{T}} \boldsymbol{\Sigma}_{\mathbf{YY}} \boldsymbol{\alpha}_k^* \geqslant 0, \boldsymbol{\beta}_k^{\mathrm{T}} \boldsymbol{\Sigma}_{\mathbf{XX}} \boldsymbol{\beta}_k^* \geqslant 0,$$

where with a little abuse of notation we let $\widehat{\rho}_l = (\widehat{\boldsymbol{\alpha}}_l^\tau)^{\mathrm{T}} \widehat{\boldsymbol{\Sigma}}_{\mathbf{YX}} \widehat{\boldsymbol{\beta}}_l^\tau$. It is easy to see that $(\widehat{\boldsymbol{\alpha}}_k^\tau, \widehat{\boldsymbol{\beta}}_k^\tau)$ is a solution to the proposed SCCA problem over a smaller feasible set $\{(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k) : \|\boldsymbol{\alpha}_k\|_1 \leqslant \tau, \|\boldsymbol{\beta}_k\|_1 \leqslant \tau, \boldsymbol{\alpha}_k^{\mathrm{T}} \boldsymbol{\Sigma}_{\mathbf{YY}} \boldsymbol{\alpha}_k^* \geqslant 0, \boldsymbol{\beta}_k^{\mathrm{T}} \boldsymbol{\Sigma}_{\mathbf{XX}} \boldsymbol{\beta}_k^* \geqslant 0\}$. Note that $(\widehat{\boldsymbol{\alpha}}_k^\tau, \widehat{\boldsymbol{\beta}}_k^\tau)$ must exist, because they are the minimizer of a continuous function over a compact set. On the other hand, $(\widehat{\boldsymbol{\alpha}}_k^\tau, \widehat{\boldsymbol{\beta}}_k^\tau)$ can not be obtained in practice, as it requires information of $\boldsymbol{\alpha}_k^*, \boldsymbol{\beta}_k^*$. It only serves as an intermediate tool in our proof.

To show Theorem 1, we first show that $(\widehat{\boldsymbol{\alpha}}_k^\tau, \widehat{\boldsymbol{\beta}}_k^\tau)$ accurately estimates $(\boldsymbol{\alpha}_k^*, \boldsymbol{\beta}_k^*)$. Later we will show that $(\widehat{\boldsymbol{\alpha}}_k^\tau, \widehat{\boldsymbol{\beta}}_k^\tau)$ is a local minimizer to the original SCCA problem. Hence, there exists a local minimizer to SCCA that accurately estimates the true directions. In our proof, we use the shorthand notation $\widehat{\mathbf{M}}_k = \widehat{\boldsymbol{\Sigma}}_{\mathbf{YX}} - \widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}} \{\sum_{i=1}^{k-1} \widehat{\rho}_i \widehat{\boldsymbol{\alpha}}_i^\tau (\widehat{\boldsymbol{\beta}}_i^\tau)^{\mathrm{T}}\} \widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}$.

We start by showing that $(\widehat{\boldsymbol{\alpha}}_1^\tau, \widehat{\boldsymbol{\beta}}_1^\tau)$ produces a high canonical correlation in Lemmas 9–11. Lemma 9 indicates that $(\widehat{\boldsymbol{\alpha}}_1^\tau, \widehat{\boldsymbol{\beta}}_1^\tau)$ yields a high covariance on the population level.

Lemma 9: *If* $\tau_0^2 \epsilon < 3/4$, *we have* $(\widehat{\boldsymbol{\alpha}}_1^\tau)^{\mathrm{T}} \boldsymbol{\Sigma}_{\mathbf{YX}} \widehat{\boldsymbol{\beta}}_1^\tau \geqslant \rho_1^* - 4\tau_0^2 \epsilon - 4\lambda\tau_0 - \tau^2 \epsilon$.

Proof. [Proof of Lemma 9] First, note that

$$(\widehat{\boldsymbol{\alpha}}_1^\tau)^{\mathrm{T}} \boldsymbol{\Sigma}_{\mathbf{YX}} \widehat{\boldsymbol{\beta}}_1^\tau \geqslant (\widehat{\boldsymbol{\alpha}}_1^\tau)^{\mathrm{T}} \widehat{\boldsymbol{\Sigma}}_{\mathbf{YX}} \widehat{\boldsymbol{\beta}}_1^\tau - |(\widehat{\boldsymbol{\alpha}}_1^\tau)^{\mathrm{T}} (\boldsymbol{\Sigma}_{\mathbf{YX}} - \widehat{\boldsymbol{\Sigma}}_{\mathbf{YX}}) \widehat{\boldsymbol{\beta}}_1^\tau| \geqslant (\widehat{\boldsymbol{\alpha}}_1^\tau)^{\mathrm{T}} \widehat{\boldsymbol{\Sigma}}_{\mathbf{YX}} \widehat{\boldsymbol{\beta}}_1^\tau - \tau^2 \epsilon.$$

Now note that by Lemma 8, $\|\widetilde{\boldsymbol{\alpha}}_1^*\| \leqslant 2\tau_0, \|\widetilde{\boldsymbol{\beta}}_1^*\|_1 \leqslant 2\tau_0$, and hence $(\widetilde{\boldsymbol{\alpha}}_1^*, \widetilde{\boldsymbol{\beta}}_1^*)$ satisfies the constraints in (18). By the definition of $(\widehat{\boldsymbol{\alpha}}_1^\tau, \widehat{\boldsymbol{\beta}}_1^\tau)$, we have

$$(\widehat{\boldsymbol{\alpha}}_1^\tau)^{\mathrm{T}} \widehat{\boldsymbol{\Sigma}}_{\mathbf{YX}} \widehat{\boldsymbol{\beta}}_1^\tau - \lambda \|\widehat{\boldsymbol{\alpha}}_1^\tau\|_1 - \lambda \|\widehat{\boldsymbol{\beta}}_1^\tau\|_1 \geqslant (\widetilde{\boldsymbol{\alpha}}_1^*)^{\mathrm{T}} \widehat{\boldsymbol{\Sigma}}_{\mathbf{YX}} \widetilde{\boldsymbol{\beta}}_1^* - \lambda \|\widetilde{\boldsymbol{\alpha}}_1^*\|_1 - \lambda \|\widetilde{\boldsymbol{\beta}}_1^*\|_1.$$

It follows that

$$\begin{aligned}
(\widehat{\boldsymbol{\alpha}}_1^{\tau})^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{YX}}\widehat{\boldsymbol{\beta}}_1^{\tau} &\geqslant (\widetilde{\boldsymbol{\alpha}}_1^{*})^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{YX}}\widetilde{\boldsymbol{\beta}}_1^{*} - (\lambda\|\widetilde{\boldsymbol{\alpha}}_1^{*}\|_1 + \lambda\|\widetilde{\boldsymbol{\beta}}_1^{*}\|_1) \geqslant (\widetilde{\boldsymbol{\alpha}}_1^{*})^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{YX}}\widetilde{\boldsymbol{\beta}}_1^{*} - 4\lambda\tau_0 \\
&\geqslant (\widetilde{\boldsymbol{\alpha}}_1^{*})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YX}}\widetilde{\boldsymbol{\beta}}_1^{*} - 4\tau_0^2\epsilon - 4\lambda\tau_0 \\
&= \frac{\rho_1^*}{\sqrt{(\boldsymbol{\alpha}_1^{*})^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}\boldsymbol{\alpha}_1^{*} \cdot (\boldsymbol{\beta}_1^{*})^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}\boldsymbol{\beta}_1^{*}}} - 4\tau_0^2\epsilon - 4\lambda\tau_0 \\
&\geqslant \frac{\rho_1^*}{1-\tau_0^2\epsilon} - 4\tau_0^2\epsilon - 4\lambda\tau_0 \\
&\geqslant \rho_1^* - 4\tau_0^2\epsilon - 4\lambda\tau_0.
\end{aligned}$$

And we have the conclusion.

Next, in Lemma 10, we obtain some technical results concerning the lengths of $(\widehat{\boldsymbol{\alpha}}_1^{\tau}, \widehat{\boldsymbol{\beta}}_1^{\tau})$.

LEMMA 10: *For* $(\widehat{\boldsymbol{\alpha}}_1^{\tau}, \widehat{\boldsymbol{\beta}}_1^{\tau})$ *defined in* (18), *we have*

$$|(\widehat{\boldsymbol{\alpha}}_1^{\tau})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\widehat{\boldsymbol{\alpha}}_1^{\tau} - (\widehat{\boldsymbol{\alpha}}_1^{\tau})^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}\widehat{\boldsymbol{\alpha}}_1^{\tau}| \leqslant \tau^2\epsilon, \tag{19}$$

$$|(\widehat{\boldsymbol{\beta}}_1^{\tau})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\widehat{\boldsymbol{\beta}}_1^{\tau} - (\widehat{\boldsymbol{\beta}}_1^{\tau})^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}\widehat{\boldsymbol{\beta}}_1^{\tau}| \leqslant \tau^2\epsilon, \tag{20}$$

$$|(\widehat{\boldsymbol{\alpha}}_1^{\tau})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YX}}\widehat{\boldsymbol{\beta}}_1^{\tau} - (\widehat{\boldsymbol{\alpha}}_1^{\tau})^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{YX}}\widehat{\boldsymbol{\beta}}_1^{\tau}| \leqslant \tau^2\epsilon. \tag{21}$$

Proof. [Proof of Lemma 10] We first show (19). Note that

$$\begin{aligned}
&(\widehat{\boldsymbol{\alpha}}_1^{\tau})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\widehat{\boldsymbol{\alpha}}_1^{\tau} = (\widehat{\boldsymbol{\alpha}}_1^{\tau})^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}\widehat{\boldsymbol{\alpha}}_1^{\tau} + (\widehat{\boldsymbol{\alpha}}_1^{\tau})^{\mathrm{T}}(\boldsymbol{\Sigma}_{\mathbf{YY}} - \widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}})\widehat{\boldsymbol{\alpha}}_1^{\tau} \\
&= (\boldsymbol{\alpha}_1^{*})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_1^{*} + (\widehat{\boldsymbol{\alpha}}_1^{\tau})^{\mathrm{T}}(\boldsymbol{\Sigma}_{\mathbf{YY}} - \widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}})\widehat{\boldsymbol{\alpha}}_1^{\tau}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&|(\widehat{\boldsymbol{\alpha}}_1^{\tau})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\widehat{\boldsymbol{\alpha}}_1^{\tau} - (\boldsymbol{\alpha}_1^{*})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_1^{*}| = |(\widehat{\boldsymbol{\alpha}}_1^{\tau})^{\mathrm{T}}(\boldsymbol{\Sigma}_{\mathbf{YY}} - \widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}})\widehat{\boldsymbol{\alpha}}_1^{\tau}| \\
\leqslant\ &\|(\widehat{\boldsymbol{\alpha}}_1^{\tau})^{\mathrm{T}}\|_\infty\|(\boldsymbol{\Sigma}_{\mathbf{YY}} - \widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}})\widehat{\boldsymbol{\alpha}}_1^{\tau}\|_\infty \leqslant \|\boldsymbol{\Sigma}_{\mathbf{YY}} - \widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}\|_{\max}\|\widehat{\boldsymbol{\alpha}}_1^{\tau}\|_1^2 \leqslant \epsilon\tau^2.
\end{aligned}$$

Equations (20) & (21) can be proven similarly and the proofs are omitted here.

LEMMA 11: *For* $(\widehat{\boldsymbol{\alpha}}_1^{\tau}, \widehat{\boldsymbol{\beta}}_1^{\tau})$ *defined in* (18), *we have*

$$\frac{(\widehat{\boldsymbol{\alpha}}_1^{\tau})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YX}}\widehat{\boldsymbol{\beta}}_1^{\tau}}{\sqrt{(\widehat{\boldsymbol{\alpha}}_1^{\tau})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\widehat{\boldsymbol{\alpha}}_1^{\tau}} \cdot \sqrt{(\widehat{\boldsymbol{\beta}}_1^{\tau})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\widehat{\boldsymbol{\beta}}_1^{\tau}}} \geqslant \rho_1^* - C\tau^2\epsilon - C\lambda\tau_0.$$

Proof. [Proof of Lemma 11] Note that

$$\begin{aligned}
\frac{(\widehat{\boldsymbol{\alpha}}_1^{\tau})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YX}}\widehat{\boldsymbol{\beta}}_1^{\tau}}{\sqrt{(\widehat{\boldsymbol{\alpha}}_1^{\tau})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\widehat{\boldsymbol{\alpha}}_1^{\tau}}\cdot\sqrt{(\widehat{\boldsymbol{\beta}}_1^{\tau})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\widehat{\boldsymbol{\beta}}_1^{\tau}}} &\geqslant \frac{(\widehat{\boldsymbol{\alpha}}_1^{\tau})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YX}}\widehat{\boldsymbol{\beta}}_1^{\tau}}{1+\tau^2\epsilon} \geqslant (\widehat{\boldsymbol{\alpha}}_1^{\tau})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YX}}\widehat{\boldsymbol{\beta}}_1^{\tau}(1-\tau^2\epsilon)\\
&\geqslant (\rho_1^* - 4\tau_0^2\epsilon - 4\lambda\tau_0 - \tau^2\epsilon)(1-\tau^2\epsilon) \qquad (22)\\
&\geqslant \rho_1^* - C\tau^2\epsilon - C\lambda\tau_0.
\end{aligned}$$

where (22) follows from Lemma 9.

In Lemma 12, we further provide a few upper bounds for the estimation error, assuming that the first k canonical pairs are accurately estimated.

LEMMA 12: *If, for all $l < k$,*

$$\frac{\{(\widehat{\boldsymbol{\alpha}}_l^{\tau})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^*\}^2}{\{(\widehat{\boldsymbol{\alpha}}_l^{\tau})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\widehat{\boldsymbol{\alpha}}_l^{\tau}\}\{(\boldsymbol{\alpha}_l^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^*\}} \geqslant 1 - C\tau^2\epsilon - C\lambda\tau_0, \qquad (23)$$

$$\frac{\{(\widehat{\boldsymbol{\beta}}_l^{\tau})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta}_l^*\}^2}{\{(\widehat{\boldsymbol{\beta}}_l^{\tau})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\widehat{\boldsymbol{\beta}}_l^{\tau}\}\{(\boldsymbol{\beta}_l^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta}_l^*)\}} \geqslant 1 - C\tau^2\epsilon - C\lambda\tau_0, \qquad (24)$$

then for any $\boldsymbol{\alpha}, \boldsymbol{\beta}$ such that $\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha} = 1, \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta} = 1$, $\|\boldsymbol{\alpha}\|_1 \leqslant \tau, \|\boldsymbol{\beta}\|_1 \leqslant \tau$, we must have

$$|\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^* - \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\widehat{\boldsymbol{\alpha}}_l^{\tau}| \leqslant C\sqrt{\tau^2\epsilon + \lambda\tau_0}, \qquad (25)$$

$$|\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta}_l^* - \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\widehat{\boldsymbol{\beta}}_l^{\tau}| \leqslant C\sqrt{\tau^2\epsilon + \lambda\tau_0}, \qquad (26)$$

$$\boldsymbol{\alpha}^{\mathrm{T}}\widehat{\mathbf{M}}_k\boldsymbol{\beta} \leqslant \rho_k^* - \Delta a + C\sqrt{\tau^2\epsilon + \lambda\tau_0}\sqrt{a} + C\tau^2\epsilon + \lambda\tau_0, \qquad (27)$$

where $a = 1 - \min\{(\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_k^*)^2, (\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta}_k^*)^2\}$.

Proof. [Proof of Lemma 12] By (19), (23) implies that $\{(\widehat{\boldsymbol{\alpha}}_l^{\tau})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^*\}^2 \geqslant 1 - C\tau^2\epsilon - C\lambda\tau_0$. Hence, for (25), note that

$$\begin{aligned}
&|\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^* - \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\widehat{\boldsymbol{\alpha}}_l^{\tau}| = |\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}(\boldsymbol{\alpha}_l^* - \widehat{\boldsymbol{\alpha}}_l^{\tau})|\\
\leqslant\ & (\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha})^{1/2}\{(\boldsymbol{\alpha}_l^* - \widehat{\boldsymbol{\alpha}}_l^{\tau})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}(\boldsymbol{\alpha}_l^* - \widehat{\boldsymbol{\alpha}}_l^{\tau})\}^{1/2}\\
\leqslant\ & \{2 + \tau^2\epsilon - 2(\boldsymbol{\alpha}_l^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\widehat{\boldsymbol{\alpha}}_l^{\tau}\}^{1/2}\\
\leqslant\ & \{2 + \tau^2\epsilon - 2(1 - C\tau^2\epsilon - C\lambda\tau_0)^{1/2}\}^{1/2} \leqslant C(\tau^2\epsilon + \lambda\tau_0)^{1/2}.
\end{aligned}$$

Equation (26) can be shown similarly. Now we show (27). Note that

$$\begin{aligned}\boldsymbol{\alpha}^{\mathrm{T}}\widehat{\mathbf{M}}_k\boldsymbol{\beta} &= \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{M}_k\boldsymbol{\beta} + \boldsymbol{\alpha}^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{YX}} - \boldsymbol{\Sigma}_{\mathbf{YX}})\boldsymbol{\beta} + \sum_{l=1}^{k-1}\rho_l^*\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^* \cdot (\boldsymbol{\beta}_l^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta} \\ &\quad - \sum_{l=1}^{k-1}\widehat{\rho}_l\boldsymbol{\alpha}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}\widehat{\boldsymbol{\alpha}}_l^{\tau} \cdot (\widehat{\boldsymbol{\beta}}_l^{\tau})^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}\boldsymbol{\beta}. \end{aligned} \tag{28}$$

By (8), we have

$$\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{M}_k\boldsymbol{\beta} \leqslant \rho_k^* - \Delta a. \tag{29}$$

On the other hand,

$$\begin{aligned} & \sum_{l=1}^{k-1}\rho_l^*\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^* \cdot (\boldsymbol{\beta}_l^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta} - \sum_{l=1}^{k-1}\widehat{\rho}_l\boldsymbol{\alpha}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}\widehat{\boldsymbol{\alpha}}_l^{\tau} \cdot (\widehat{\boldsymbol{\beta}}_l^{\tau})^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}\boldsymbol{\beta} \\ = & \sum_{l=1}^{k-1}(\rho_l^* - \widehat{\rho}_l)\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^* \cdot (\boldsymbol{\beta}_l^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta} \\ & -\sum_{l=1}^{k-1}\widehat{\rho}_l\left\{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^* \cdot (\boldsymbol{\beta}_l^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta} - \boldsymbol{\alpha}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}\widehat{\boldsymbol{\alpha}}_l^{\tau} \cdot (\widehat{\boldsymbol{\beta}}_l^{\tau})^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}\boldsymbol{\beta}\right\} \\ \leqslant & \sum_{l=1}^{k-1}|\rho_l^* - \widehat{\rho}_l| + \sum_{l=1}^{k-1}|\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^* \cdot (\boldsymbol{\beta}_l^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta} - \boldsymbol{\alpha}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}\widehat{\boldsymbol{\alpha}}_l^{\tau} \cdot (\widehat{\boldsymbol{\beta}}_l^{\tau})^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}\boldsymbol{\beta}| \\ \equiv & S_1 + S_2. \end{aligned}$$

For S_1, we provide a bound for $|\widehat{\rho}_l - \rho_l^*|$. Note that

$$\begin{aligned} |\widehat{\rho}_l - \rho_l^*| &= |(\widehat{\boldsymbol{\alpha}}_l^{\tau})^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{YX}}\widehat{\boldsymbol{\beta}}_l^{\tau} - \rho_l^*| \\ &\leqslant |(\widehat{\boldsymbol{\alpha}}_l^{\tau})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YX}}\widehat{\boldsymbol{\beta}}_l^{\tau} - \rho_l^*| + |(\widehat{\boldsymbol{\alpha}}_l^{\tau})^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{YX}} - \boldsymbol{\Sigma}_{\mathbf{YX}})\widehat{\boldsymbol{\beta}}_l^{\tau}| \\ &\leqslant |(\widehat{\boldsymbol{\alpha}}_l^{\tau})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YX}}\widehat{\boldsymbol{\beta}}_l^{\tau} - \rho_l^*| + \tau^2\epsilon. \end{aligned}$$

Hence, it suffices to bound $|(\widehat{\boldsymbol{\alpha}}_l^{\tau})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YX}}\widehat{\boldsymbol{\beta}}_l^{\tau} - \rho_l^*|$. Note that

$$(\widehat{\boldsymbol{\alpha}}_l^{\tau})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YX}}\widehat{\boldsymbol{\beta}}_l^{\tau} = \sum_{i=1}^{K}\rho_i^*(\widehat{\boldsymbol{\alpha}}_l^{\tau})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_i^* \cdot (\widehat{\boldsymbol{\beta}}_l^{\tau})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta}_i^*.$$

For $i \neq l$,

$$\rho_i^*(\widehat{\boldsymbol{\alpha}}_l^\tau)^{\mathrm{T}}\boldsymbol{\Sigma}_{YY}\boldsymbol{\alpha}_i^* \cdot (\widehat{\boldsymbol{\beta}}_l^\tau)^{\mathrm{T}}\boldsymbol{\Sigma}_{XX}\boldsymbol{\beta}_i^* = (\widehat{\boldsymbol{\alpha}}_l^\tau - \boldsymbol{\alpha}_l^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{YY}\boldsymbol{\alpha}_i^* \cdot (\widehat{\boldsymbol{\beta}}_l^\tau - \boldsymbol{\beta}_l^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{XX}\boldsymbol{\beta}_i^* \leqslant C(\tau^2\epsilon + \lambda\tau_0)$$

where the last inequality follows from (25)–(26). For $i = l$,

$$\begin{aligned} &\rho_l^*(\widehat{\boldsymbol{\alpha}}_l^\tau)^{\mathrm{T}}\boldsymbol{\Sigma}_{YY}\boldsymbol{\alpha}_l^* \cdot (\widehat{\boldsymbol{\beta}}_l^\tau)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta}_l^* - \rho_l^* \leqslant \frac{\rho_l^*}{2}[\{(\widehat{\boldsymbol{\alpha}}_l^\tau)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^*\}^2 + \{(\widehat{\boldsymbol{\beta}}_l^\tau)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta}_l^*\}^2 - 2] \\ \leqslant\ & \frac{\rho_l^*}{2}[(\widehat{\boldsymbol{\alpha}}_l^\tau)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\widehat{\boldsymbol{\alpha}}_l^\tau + (\widehat{\boldsymbol{\beta}}_l^\tau)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\widehat{\boldsymbol{\beta}}_l^\tau - 2] \leqslant \rho_l^*\tau^2\epsilon, \end{aligned}$$

where the last inequality can be shown similarly to Lemma 10. In the meantime, because $(\widehat{\boldsymbol{\alpha}}_l^\tau)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^* \geqslant 0, (\widehat{\boldsymbol{\beta}}_l^\tau)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta}_l^* \geqslant 0$, by (23) &(24) we have

$$\begin{aligned} &\rho_l^*(\widehat{\boldsymbol{\alpha}}_l^\tau)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^* \cdot (\widehat{\boldsymbol{\beta}}_l^\tau)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta}_l^* - \rho_l^* \geqslant \rho_l^* \min[\{(\widehat{\boldsymbol{\alpha}}_l^\tau)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^*\}^2, \{(\widehat{\boldsymbol{\beta}}_l^\tau)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta}_l^*\}^2] - \rho_l^* \\ \geqslant\ & \rho_l^*(1 - C\tau^2\epsilon - C\lambda\tau_0)(1 - \tau^2\epsilon) - \rho_l^* \geqslant -C(\tau^2\epsilon + \lambda\tau_0) \end{aligned}$$

Hence, $|\widehat{\rho}_l - \rho_l^*| \leqslant C(\tau^2\epsilon + \lambda\tau_0)$ and thus

$$S_1 \leqslant C(\tau^2\epsilon + \lambda\tau_0). \tag{30}$$

For S_2, note that

$$\begin{aligned} & |\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^* \cdot (\boldsymbol{\beta}_l^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta} - \boldsymbol{\alpha}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}\widehat{\boldsymbol{\alpha}}_l^\tau \cdot (\widehat{\boldsymbol{\beta}}_l^\tau)^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}\boldsymbol{\beta}| \\ \leqslant\ & |\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^* - \boldsymbol{\alpha}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}\widehat{\boldsymbol{\alpha}}_l^\tau| \cdot |(\boldsymbol{\beta}_l^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta} - (\widehat{\boldsymbol{\beta}}_l^\tau)^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}\boldsymbol{\beta}| \\ & +|(\boldsymbol{\beta}_l^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta}| \cdot |\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^* - \boldsymbol{\alpha}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}\widehat{\boldsymbol{\alpha}}_l^\tau| \\ & +|\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^*| \cdot |(\boldsymbol{\beta}_l^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta} - (\widehat{\boldsymbol{\beta}}_l^\tau)^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}\boldsymbol{\beta}| \\ \leqslant\ & C(\tau^2\epsilon + \lambda\tau_0) + 2C(\tau^2\epsilon + \lambda\tau_0)^{1/2}\sqrt{2a}, \end{aligned} \tag{31}$$

where the last inequality follows from (25).

Hence, combining (28), (29), (30) with (31), we have the desired conclusion in (27).

Under the same conditions in Lemma 12, we derive a lower bound for $(\widetilde{\boldsymbol{\alpha}}_k^*, \widetilde{\boldsymbol{\beta}}_k^*)$.

Lemma 13: *Assume that* (23)–(24) *are true for* $l < k$. *Then for* $(\widetilde{\boldsymbol{\alpha}}_k^*, \widetilde{\boldsymbol{\beta}}_k^*)$ *defined in*

(13)–(14), *we have*

$$(\widetilde{\boldsymbol{\alpha}}_k^*)^{\mathrm{T}}\widehat{\mathbf{M}}_k\widetilde{\boldsymbol{\beta}}_k^* \geqslant \rho_k^* - C\tau^2\epsilon - C\lambda\tau_0. \tag{32}$$

Proof. [Proof of Lemma 13] By the definition of $(\widetilde{\boldsymbol{\alpha}}_k^*, \widetilde{\boldsymbol{\beta}}_k^*)$, we have $(\widetilde{\boldsymbol{\alpha}}_k^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^* = 0$, $(\widetilde{\boldsymbol{\beta}}_k^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta}_l^* = 0$ for $l \neq k$. Also recall that by Lemma 8 we have $(\widetilde{\boldsymbol{\alpha}}_k^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\widetilde{\boldsymbol{\alpha}}_k^* \leqslant 1 + \tau^2\epsilon$, $(\widetilde{\boldsymbol{\beta}}_k^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\widetilde{\boldsymbol{\beta}}_k^* \leqslant 1 + \tau^2\epsilon$. Consequently,

$$\begin{aligned}
&(\widetilde{\boldsymbol{\alpha}}_k^*)^{\mathrm{T}}\widehat{\mathbf{M}}_k\widetilde{\boldsymbol{\beta}}_k^* = (\widetilde{\boldsymbol{\alpha}}_k^*)^{\mathrm{T}}\mathbf{M}_k\widetilde{\boldsymbol{\beta}}_k^* + (\widetilde{\boldsymbol{\alpha}}_k^*)^{\mathrm{T}}(\widehat{\mathbf{M}}_k - \mathbf{M}_k)\widetilde{\boldsymbol{\beta}}_k^* \\
\geqslant\ & \rho_k^*(1 - C\tau_0^2\epsilon) + (\widetilde{\boldsymbol{\alpha}}_k^*)^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{YX}} - \boldsymbol{\Sigma}_{\mathbf{YX}})\widetilde{\boldsymbol{\beta}}_k^* - \sum_{l=1}^{k-1}\widehat{\rho}_l(\widetilde{\boldsymbol{\alpha}}_k^*)^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}\widehat{\boldsymbol{\alpha}}_l^{\tau} \cdot (\widetilde{\boldsymbol{\beta}}_k^*)^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}\widehat{\boldsymbol{\beta}}_l^{\tau} \\
\geqslant\ & \rho_k^* - C\tau^2\epsilon - \sum_{l=1}^{k-1}\widehat{\rho}_l(\widetilde{\boldsymbol{\alpha}}_k^*)^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}\widehat{\boldsymbol{\alpha}}_l^{\tau} \cdot (\widetilde{\boldsymbol{\beta}}_k^*)^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}\widehat{\boldsymbol{\beta}}_l^{\tau}.
\end{aligned}$$

For the last term, note that

$$\begin{aligned}
&|(\widetilde{\boldsymbol{\alpha}}_k^*)^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}\widehat{\boldsymbol{\alpha}}_l^{\tau}| = |(\widetilde{\boldsymbol{\alpha}}_k^*)^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}}\widehat{\boldsymbol{\alpha}}_l^{\tau} - (\widetilde{\boldsymbol{\alpha}}_k^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_l^*| \\
\leqslant\ & |(\widetilde{\boldsymbol{\alpha}}_k^*)^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{YY}} - \boldsymbol{\Sigma}_{\mathbf{YY}})\widehat{\boldsymbol{\alpha}}_l^{\tau}| + |(\widetilde{\boldsymbol{\alpha}}_k^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}(\widehat{\boldsymbol{\alpha}}_l^{\tau} - \boldsymbol{\alpha}_l^*)| \\
\leqslant\ & 2\tau\tau_0\epsilon + \sqrt{(1 + C\tau_0^2\epsilon)(\widehat{\boldsymbol{\alpha}}_l^{\tau} - \boldsymbol{\alpha}_l^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}(\widehat{\boldsymbol{\alpha}}_l^{\tau} - \boldsymbol{\alpha}_l^*)} \\
=\ & 2\tau\tau_0\epsilon + \sqrt{(1 + C\tau_0^2\epsilon)\{(\widehat{\boldsymbol{\alpha}}_l^{\tau})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\widehat{\boldsymbol{\alpha}}_l^{\tau} + 1 - 2(\boldsymbol{\alpha}_l^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\widehat{\boldsymbol{\alpha}}_l^{\tau}\}} \\
\leqslant\ & C(\tau^2\epsilon + \lambda\tau_0)^{1/2},
\end{aligned}$$

where we invoke (23). Similarly, $|(\widetilde{\boldsymbol{\beta}}_k^*)^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{XX}}\widehat{\boldsymbol{\beta}}_l^{\tau}| \leqslant C(\tau^2\epsilon + \lambda\tau_0)^{1/2}$. It follows that (32) is true.

With Lemmas 5–13, we prove the following Theorem 2 that with proper rotation, $(\widehat{\boldsymbol{\alpha}}_k^{\tau}, \widehat{\boldsymbol{\beta}}_k^{\tau})$ accurately estimates $(\boldsymbol{\alpha}_k^*, \boldsymbol{\beta}_k^*)$, for $k = 1, \ldots, K$.

THEOREM 2: *Define*

$$a_k = 1 - \min\left[\frac{\{(\widehat{\boldsymbol{\alpha}}_k^{\tau})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_k^*\}^2}{(\widehat{\boldsymbol{\alpha}}_k^{\tau})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\widehat{\boldsymbol{\alpha}}_k^{\tau} \cdot (\boldsymbol{\alpha}_k^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_k^*}, \frac{\{(\widehat{\boldsymbol{\beta}}_k^{\tau})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta}_k^*\}^2}{(\widehat{\boldsymbol{\beta}}_k^{\tau})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\widehat{\boldsymbol{\beta}}_k^{\tau} \cdot (\boldsymbol{\beta}_k^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta}_k^*}\right].$$

We have, for $k = 1, \ldots, K$,

$$a_k \leqslant C(\tau^2\epsilon + \lambda\tau_0), \tag{33}$$

$$\frac{(\widehat{\boldsymbol{\alpha}}_k^\tau)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YX}}\widehat{\boldsymbol{\beta}}_k^\tau}{\sqrt{(\widehat{\boldsymbol{\alpha}}_k^\tau)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\widehat{\boldsymbol{\alpha}}_k^\tau \cdot (\widehat{\boldsymbol{\beta}}_k^\tau)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\widehat{\boldsymbol{\beta}}_k^\tau}} \geqslant \rho_k^* - C\tau^2\epsilon - C\lambda\tau_0. \tag{34}$$

Proof. [Proof of Theorem 2] We show the conclusion by induction. By Lemma 11 we have (34) for $k = 1$. Combine (34) with Lemma 5 and we have (33) for $k = 1$.

Now assume that the desired conclusion holds for $k \leqslant m - 1$. For $k = m$, by (27), we have

$$(\widehat{\boldsymbol{\alpha}}_m^\tau)^{\mathrm{T}}\widehat{\mathbf{M}}_m\widehat{\boldsymbol{\beta}}_m^\tau \leqslant \rho_m^* - \Delta a_m - C\sqrt{\tau^2\epsilon + \lambda\tau_0}\sqrt{a_m} + C(\tau^2\epsilon + \lambda\tau_0).$$

By the definition of $\widehat{\boldsymbol{\alpha}}_m, \widehat{\boldsymbol{\beta}}_m$, we have

$$\rho_m^* - \Delta a_m + C(\tau^2\epsilon + \lambda\tau_0) + 2C(\tau^2\epsilon + \lambda\tau_0)^{1/2}\sqrt{2a_m} \geqslant (\widetilde{\boldsymbol{\alpha}}_m^*)^{\mathrm{T}}\widehat{\mathbf{M}}_m\widetilde{\boldsymbol{\beta}}_m^* - \frac{\lambda}{2}\|\widetilde{\boldsymbol{\alpha}}_m^*\|_1 - \frac{\lambda}{2}\|\widetilde{\boldsymbol{\beta}}_m^*\|_1.$$

By (32),

$$(\widetilde{\boldsymbol{\alpha}}_m^*)^{\mathrm{T}}\widehat{\mathbf{M}}_m\widetilde{\boldsymbol{\beta}}_m^* \geqslant \rho_m^* - C(\tau_0^2\epsilon + \lambda\tau_0).$$

Consequently,

$$\rho_m^* - \Delta a_m + C(\tau^2\epsilon + \lambda\tau_0) + 2C(\tau^2\epsilon + \lambda\tau_0)^{1/2}\sqrt{2a_m} \geqslant \rho_m^* - C(\tau^2\epsilon + \lambda\tau_0).$$

Solve this inequality and we have $\sqrt{a_m} \leqslant \sqrt{C(\tau^2\epsilon + 4\lambda\tau_0)}$.

Hence, we have $a_m \leqslant C(\tau^2\epsilon + \lambda\tau_0)$. Then (34) follows from Lemma 12.

Next, we show that $(\widehat{\boldsymbol{\alpha}}_k^\tau, \widehat{\boldsymbol{\beta}}_k^\tau)$ is a local minimizer to the original SCCA problem, which does not require any information on $\boldsymbol{\alpha}_k^*, \boldsymbol{\beta}_k^*$.

Lemma 14: *There exists $C_U > C_L > 0$, and $\epsilon_0 > 0$ such that, if for $\epsilon < \epsilon_0$,$C_L\tau_0\epsilon < \lambda < C_U\tau_0\epsilon$, we must have that $(\widehat{\boldsymbol{\alpha}}_k^\tau, \widehat{\boldsymbol{\beta}}_k^\tau)$ is a local minimizer to SCCA.*

Proof. [Proof of Lemma 14] We first show that, for any $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, if $\max\{\|\boldsymbol{\alpha}\|_1, \|\boldsymbol{\beta}\|_1\} = \tau$, then $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ cannot be a minimizer of the SCCA problem over $\{(\boldsymbol{\alpha}, \boldsymbol{\beta}) : \|\boldsymbol{\alpha}\|_1 \leqslant \tau, \|\boldsymbol{\beta}\|_1 \leqslant \tau, \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{YY}\boldsymbol{\alpha}_k^* \geqslant 0, \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_{XX}\boldsymbol{\beta}_k^* \geqslant 0\}$. To this end, let $a = 1 - \min\{(\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_k^*)^2, (\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta}_k^*)^2\}$.

By (27), we have

$$\boldsymbol{\alpha}^{\mathrm{T}}\widehat{\mathbf{M}}_k\boldsymbol{\beta} \leqslant \rho_k^* - \Delta a + C\sqrt{(\tau^2\epsilon + \lambda\tau_0)a} + C(\tau^2\epsilon + \lambda\tau_0) \leqslant \rho_k^* + C(\tau^2\epsilon + \lambda\tau_0).$$

Therefore, for sufficiently large N and $\boldsymbol{\alpha}, \boldsymbol{\beta}$ such that $\max\{\|\boldsymbol{\alpha}\|_1, \|\boldsymbol{\beta}\|_1\} = \tau$, we have

$$\boldsymbol{\alpha}^{\mathrm{T}}\widehat{\mathbf{M}}_k\boldsymbol{\beta} - \lambda/2(\|\boldsymbol{\alpha}\|_1 + \|\boldsymbol{\beta}\|_1) \leqslant \rho_k^* + C\tau^2\epsilon + C\lambda\tau_0 - \lambda\tau < \rho_k^* + C\tau^2\epsilon - 8\lambda\tau_0.$$

By (32),

$$(\widetilde{\boldsymbol{\alpha}}_k^*)^{\mathrm{T}}\widehat{\mathbf{M}}_k\widetilde{\boldsymbol{\beta}}_k^* - \lambda(\|\widetilde{\boldsymbol{\alpha}}_k^*\|_1 + \|\widetilde{\boldsymbol{\beta}}_k^*\|_1) \geqslant \rho_k^* - C\tau^2\epsilon - 4\lambda\tau_0.$$

Hence, there exists C_L such that, if $\lambda > C_L\tau_0\epsilon$, we have, for any $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ such that $\max\{\|\boldsymbol{\alpha}\|_1, \|\boldsymbol{\beta}\|_1\} = \tau$,

$$\boldsymbol{\alpha}^{\mathrm{T}}\widehat{\mathbf{M}}_k\boldsymbol{\beta} - \lambda(\|\boldsymbol{\alpha}\|_1 + \|\boldsymbol{\beta}\|_1) < (\widetilde{\boldsymbol{\alpha}}_k^*)^{\mathrm{T}}\widehat{\mathbf{M}}_k\widetilde{\boldsymbol{\beta}}_k^* - \lambda(\|\widetilde{\boldsymbol{\alpha}}_k^*\|_1 + \|\widetilde{\boldsymbol{\beta}}_k^*\|_1).$$

It follows that $\|\widehat{\boldsymbol{\alpha}}_k^\tau\|_1 < \tau, \|\widehat{\boldsymbol{\beta}}_k^\tau\|_1 < \tau$. Further, there exists ϵ_0, C_U such that for any $0 < \epsilon < \epsilon_0, \lambda < C_U\tau_0\epsilon$, the upper bounds in Theorem 2 are strictly less than 1, indicating that $(\widehat{\boldsymbol{\alpha}}_k^\tau)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{YY}}\boldsymbol{\alpha}_k^* > 0, (\widehat{\boldsymbol{\beta}}_k^\tau)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta}_k^* > 0$. It follows that $\widehat{\boldsymbol{\alpha}}_k^\tau, \widehat{\boldsymbol{\beta}}_k^\tau$ are local minimizers.

Now we are ready to prove Theorem 1.

Proof. [Proof of Theorem 1] By Lemma 14, there exists $\epsilon_0, C_U, C_L > 0$ such that if $C_L\tau_0\epsilon < \lambda < C_U\tau_0\epsilon$ and $0 < \epsilon < \epsilon_0$, we have $\{\widehat{\boldsymbol{\alpha}}_k^\tau, \widehat{\boldsymbol{\beta}}_k^\tau\}_{k=1}^K$ is a local minimizer to SCCA. By Theorem 2, if $\lambda < C_U\tau_0\epsilon$ for some C_U, then

$$\cos^2\left\{\angle(\boldsymbol{\Sigma}_{\mathbf{YY}}^{1/2}\widehat{\boldsymbol{\alpha}}_k^\tau, \boldsymbol{\Sigma}_{\mathbf{YY}}^{1/2}\boldsymbol{\alpha}_k^*)\right\} \geqslant 1 - C\tau_0^2\epsilon, \cos^2\left\{\angle(\boldsymbol{\Sigma}_{\mathbf{XX}}^{1/2}\widehat{\boldsymbol{\beta}}_k^\tau, \boldsymbol{\Sigma}_{\mathbf{XX}}^{1/2}\boldsymbol{\beta}_k^*)\right\} \geqslant 1 - C\tau_0^2\epsilon.$$

Then, under Condition (C1), Lemma 6 implies the desired conclusion.

References

Bickel, P. J. and Levina, E. (2008). Covariance regularization by thresholding. *Annals of Statistics* **36,** 2577–2604.

Gao, C., Ma, Z., Zhou, H. H., et al. (2017). Sparse cca: Adaptive estimation and computational barriers. *The Annals of Statistics* **45,** 2074–2101.

Received October 2007. *Revised February* 2008. *Accepted March* 2008.

	Model 1		Model 2		Model 3		Model 4	
$p, q = 600$	Err($\widehat{\mathbf{A}}$)	Err($\widehat{\mathbf{B}}$)	Err($\widehat{\mathbf{A}}$)	Err($\widehat{\mathbf{B}}$)	Err($\widehat{\mathbf{A}}$)	Err($\widehat{\mathbf{B}}$)	Err($\widehat{\mathbf{A}}$)	Err($\widehat{\mathbf{B}}$)
SCCA	0.1228	0.1262	0.1256	0.1325	0.2353	0.2325	0.1645	0.1709
	(0.0035)	(0.0045)	(0.0028)	(0.0053)	(0.0067)	(0.0042)	(0.0038)	(0.0045)
Colar	0.1401	0.1463	0.1397	0.1442	0.2270	0.2372	0.2041	0.2130
	(0.0027)	(0.0030)	(0.0027)	(0.0029)	(0.0050)	(0.0047)	(0.0033)	(0.0034)
PMD	1.4154	1.4154	1.0543	1.0453	1.1072	1.1182	1.7282	1.7158
	(0.0680)	(0.0691)	(0.0861)	(0.0879)	(0.0202)	(0.0263)	(0.0263)	(0.0269)

Table 1

Simulation results for Models 1–4 with dimensions $p = q = 600$. The reported numbers are the medians and standard errors (in parentheses) of Err($\widehat{\mathbf{A}}$) $= \|\mathbf{P}_{\widehat{\mathbf{A}}} - \mathbf{P}_{\mathbf{A}}\|_F$ and Err($\widehat{\mathbf{B}}$) $= \|\mathbf{P}_{\widehat{\mathbf{B}}} - \mathbf{P}_{\mathbf{B}}\|_F$, over 200 replicates.