Supplement to "Subspace Perspective on Canonical Correlation Analysis: Dimension Reduction and Minimax Rates"

To establish the minimax lower bounds of CCA estimates for our proposed losses, we follow the analytical frameworks in the literature of PCA and CCA, e.g., [5, 1, 2], where the calculation is focused on the construction of the hypothesis class to which the packing lemma and Fano's inequality are applied. However, since we fix both λ_k and λ_{k+1} in the localized parameter spaces, new technical challenges arise and consequently we construct hypothesis classes based on the equality (0.1). In this section we also denote $\Delta := \lambda_k - \lambda_{k+1}$.

0.1. On Kullback-Leibler Divergence

The following lemma can be viewed as an extension of Lemma 14 in [2] from $\lambda_{k+1} = 0$ to arbitrary λ_{k+1}. The proof of the lemma can be found in Section 0.4.

Lemma 0.1. *For $i = 1, 2$ and $p_2 \geqslant p_1 \geqslant k$, let $\left[\boldsymbol{U}_{(i)},\ \boldsymbol{W}_{(i)}\right] \in \mathcal{O}(p_1, p_1), \left[\boldsymbol{V}_{(i)},\ \boldsymbol{Z}_{(i)}\right] \in \mathcal{O}(p_2, p_1)$ where $\boldsymbol{U}_{(i)} \in \mathbb{R}^{p_1 \times k}, \boldsymbol{V}_{(i)} \in \mathbb{R}^{p_2 \times k}$. For $0 \leqslant \lambda_2 < \lambda_1 < 1$, let $\Delta = \lambda_1 - \lambda_2$ and define*

$$\boldsymbol{\Sigma}_{(i)} = \begin{bmatrix} \boldsymbol{\Sigma}_x & \boldsymbol{\Sigma}_x^{1/2}(\lambda_1 \boldsymbol{U}_{(i)} \boldsymbol{V}_{(i)}^\top + \lambda_2 \boldsymbol{W}_{(i)} \boldsymbol{Z}_{(i)}^\top)\boldsymbol{\Sigma}_y^{1/2} \\ \boldsymbol{\Sigma}_y^{1/2}(\lambda_1 \boldsymbol{V}_{(i)} \boldsymbol{U}_{(i)}^\top + \lambda_2 \boldsymbol{Z}_{(i)} \boldsymbol{W}_{(i)}^\top)\boldsymbol{\Sigma}_x^{1/2} & \boldsymbol{\Sigma}_y \end{bmatrix} \quad i = 1, 2,$$

Let $\mathbb{P}_{(i)}$ denote the distribution of a random i.i.d. sample of size n from $N(0, \boldsymbol{\Sigma}_{(i)})$. If we further assume

$$[\boldsymbol{U}_{(1)}, \boldsymbol{W}_{(1)}] \begin{bmatrix} \boldsymbol{V}_{(1)}^\top \\ \boldsymbol{Z}_{(1)}^\top \end{bmatrix} = [\boldsymbol{U}_{(2)}, \boldsymbol{W}_{(2)}] \begin{bmatrix} \boldsymbol{V}_{(2)}^\top \\ \boldsymbol{Z}_{(2)}^\top \end{bmatrix}, \tag{0.1}$$

Then one can show that

$$D(\mathbb{P}_{(1)} || \mathbb{P}_{(2)}) = \frac{n\Delta^2(1 + \lambda_1\lambda_2)}{2(1-\lambda_1^2)(1-\lambda_2^2)} \|\boldsymbol{U}_{(1)}\boldsymbol{V}_{(1)}^\top - \boldsymbol{U}_{(2)}\boldsymbol{V}_{(2)}^\top\|_F^2.$$

Remark 0.2. *The conditon in (0.1) is crucial for obtaining the eigen-gap factor $1/\Delta^2$ in the lower bound and is the key insight behind the construction of the hypothesis class in the proof. [2] has a similar lemma but only deals with the case that the residual canonical correlations are zero. To the best of our knowledge, the proof techniques in [2, 3] cannot be directly used to obtain our results.*

0.2. Packing Number and Fano's Lemma

The following result on the packing number is based on the metric entropy of the Grassmannian manifold $G(k, r)$ due to [4]. We use the version adapted from Lemma 1 of [1] which is also used in [2].

Lemma 0.3. *For any fixed $\boldsymbol{U}_0 \in \mathcal{O}(p, k)$ and $\mathcal{B}_{\epsilon_0} = \{\boldsymbol{U} \in \mathcal{O}(p, k) : \|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}_0\boldsymbol{U}_0^\top\|_F \leqslant \epsilon_0\}$ with $\epsilon_0 \in (0, \sqrt{2[k \wedge (p-k)]}\,)$. Define the semi-metric $\rho(\cdot, \cdot)$ on $\mathcal{B}_{\epsilon_0}$ by*

$$\rho(\boldsymbol{U}_1, \boldsymbol{U}_2) = \|\boldsymbol{U}_1\boldsymbol{U}_1^\top - \boldsymbol{U}_2\boldsymbol{U}_2^\top\|_F.$$

Then there exists universal constant C such that for any $\alpha \in (0,1)$, the packing number $\mathcal{M}(\mathcal{B}_{\epsilon_0}, \rho, \alpha\epsilon_0)$ satisfies

$$\mathcal{M}(\mathcal{B}_{\epsilon_0}, \rho, \alpha\epsilon_0) \geqslant \left(\frac{1}{C\alpha}\right)^{k(p-k)}.$$

The following corollary is used to prove the lower bound.

Corollary 0.4. *If we change the set in Lemma 0.3 to $\widetilde{\mathcal{B}}_{\epsilon_0} = \{\boldsymbol{U} \in \mathcal{O}(p,k) : \|\boldsymbol{U} - \boldsymbol{U}_0\|_F \leqslant \epsilon_0\}$, then we still have*

$$\mathcal{M}(\widetilde{\mathcal{B}}_{\epsilon_0}, \rho, \alpha\epsilon_0) \geqslant \left(\frac{1}{C\alpha}\right)^{k(p-k)}.$$

Proof. Apply Lemma 0.3 to $\mathcal{B}_{\epsilon_0}$, there exists $\boldsymbol{U}_1, \cdots, \boldsymbol{U}_n$ with $n \geqslant (1/C\alpha)^{k(p-k)}$ such that

$$\|\boldsymbol{U}_i\boldsymbol{U}_i^\top - \boldsymbol{U}_0\boldsymbol{U}_0^\top\|_F \leqslant \epsilon_0,\ 1 \leqslant i \leqslant n,\ \|\boldsymbol{U}_i\boldsymbol{U}_i^\top - \boldsymbol{U}_j\boldsymbol{U}_j^\top\|_F \geqslant \alpha\epsilon_0, 1 \leqslant i \leqslant j \leqslant n.$$

Define $\widetilde{\boldsymbol{U}}_i = \arg\min_{\boldsymbol{U} \in \{\boldsymbol{U}_i\boldsymbol{Q},\ \boldsymbol{Q} \in \mathcal{O}(k)\}} \|\boldsymbol{U} - \boldsymbol{U}_0\|_F$, by Lemma 0.5,

$$\|\widetilde{\boldsymbol{U}}_i - \boldsymbol{U}_0\|_F \leqslant \|\widetilde{\boldsymbol{U}}_i\widetilde{\boldsymbol{U}}_i^\top - \boldsymbol{U}_0\boldsymbol{U}_0^\top\|_F \leqslant \epsilon_0.$$

Therefore, $\widetilde{\boldsymbol{U}}_1, \cdots, \widetilde{\boldsymbol{U}}_n \in \widetilde{\mathcal{B}}_{\epsilon_0}$ and

$$\|\widetilde{\boldsymbol{U}}_i\widetilde{\boldsymbol{U}}_i^\top - \widetilde{\boldsymbol{U}}_j\widetilde{\boldsymbol{U}}_j^\top\|_F = \|\boldsymbol{U}_i\boldsymbol{U}_i^\top - \boldsymbol{U}_j\boldsymbol{U}_j^\top\|_F \geqslant \alpha\epsilon_0.$$

which implies,

$$\mathcal{M}(\widetilde{\mathcal{B}}_{\epsilon_0}, \rho, \alpha\epsilon_0) \geqslant n \geqslant \left(\frac{1}{C\alpha}\right)^{k(p-k)}.$$

□

Lemma 0.5. *For any matrices $\boldsymbol{U}_1, \boldsymbol{U}_2 \in \mathcal{O}(p,k)$,*

$$\inf_{\boldsymbol{Q} \in \mathcal{O}(k,k)} \|\boldsymbol{U}_1 - \boldsymbol{U}_2\boldsymbol{Q}\|_F \leqslant \|\boldsymbol{P}_{\boldsymbol{U}_1} - \boldsymbol{P}_{\boldsymbol{U}_2}\|_F$$

Proof. By definition

$$\|\boldsymbol{U}_1 - \boldsymbol{U}_2\boldsymbol{Q}\|_F^2 = 2k - 2tr(\boldsymbol{U}_1^\top\boldsymbol{U}_2\boldsymbol{Q})$$

Let $\boldsymbol{U}_1^\top\boldsymbol{U}_2 = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top$ be the singular value decomposition. Then $\boldsymbol{V}\boldsymbol{U}^\top \in O(k,k)$ and

$$\begin{aligned}\inf_{\boldsymbol{Q} \in O(k,k)} \|\boldsymbol{U}_1 - \boldsymbol{U}_2\boldsymbol{Q}\|_F^2 &\leqslant 2k - 2tr(\boldsymbol{U}_1^\top\boldsymbol{U}_2\boldsymbol{V}\boldsymbol{U}^\top)\\ &= 2k - 2tr(\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^\top)\\ &= 2k - 2tr(\boldsymbol{D}).\end{aligned}$$

On the other hand,

$$\begin{aligned}\|\boldsymbol{P}_{\boldsymbol{U}_1} - \boldsymbol{P}_{\boldsymbol{U}_2}\|_F^2 &= \|\boldsymbol{U}_1\boldsymbol{U}_1^\top - \boldsymbol{U}_2\boldsymbol{U}_2^\top\|_F^2\\ &= 2k - 2tr(\boldsymbol{U}_1\boldsymbol{U}_1^\top\boldsymbol{U}_2\boldsymbol{U}_2^\top)\\ &= 2k - 2tr(\boldsymbol{U}_1^\top\boldsymbol{U}_2\boldsymbol{U}_2^\top\boldsymbol{U}_1)\\ &= 2k - 2tr(\boldsymbol{D}^2).\end{aligned}$$

Since $\boldsymbol{U}_1, \boldsymbol{U}_2 \in O(p,k)$, $\|\boldsymbol{U}_1^\top \boldsymbol{U}_2\| \leqslant 1$ and therefore all the diagonal elements of $\boldsymbol{D}$ is less than 1, which implies that $tr(\boldsymbol{D}) \geqslant tr(\boldsymbol{D}^2)$ and

$$\inf_{\boldsymbol{Q} \in O(k,k)} \|\boldsymbol{U}_1 - \boldsymbol{U}_2 \boldsymbol{Q}\|_F^2 \leqslant \|\boldsymbol{P}_{\boldsymbol{U}_1} - \boldsymbol{P}_{\boldsymbol{U}_2}\|_F^2.$$

□

Lemma 0.6 (Fano's Lemma [6]). *Let (Θ, ρ) be a (semi)metric space and $\{\mathbb{P}_\theta : \theta \in \Theta\}$ a collection of probability measures. For any totally bounded $T \subset \Theta$, denote $\mathcal{M}(T, \rho, \epsilon)$ the ϵ-packing number of T with respect to the metric ρ, i.e. , the maximal number of points in T whoese pairwise minimum distance in ρ is at least ϵ. Define the Kullback-Leibler diameter of T by*

$$d_{KL}(T) = \sup_{\theta, \theta' \in T} D(\mathbb{P}_\theta || \mathbb{P}_{\theta'}).$$

Then,

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}_\theta \left[\rho^2(\hat{\theta}, \theta)\right] \geqslant \sup_{T \subset \Theta} \sup_{\epsilon > 0} \frac{\epsilon^2}{4} \Big(1 - \frac{d_{KL}(T) + log\ 2}{log\ \mathcal{M}(T, \rho, \epsilon)}\Big)$$

0.3. Proof of Theorem 2.2: Lower Bound

For any fixed $\left[\boldsymbol{U}_{(0)},\ \boldsymbol{W}_{(0)}\right] \in \mathcal{O}(p_1, p_1)$ and $\left[\boldsymbol{V}_{(0)},\ \boldsymbol{Z}_{(0)}\right] \in \mathcal{O}(p_2, p_1)$ where $\boldsymbol{U}_{(0)} \in \mathbb{R}^{p_1 \times k}$, $\boldsymbol{V}_{(0)} \in \mathbb{R}^{p_2 \times k}$, $\boldsymbol{W}_{(0)} \in \mathbb{R}^{p_1 \times (p_1 - k)}$, $\boldsymbol{Z}_{(0)} \in \mathbb{R}^{p_2 \times (p_2 - k)}$, define

$$\mathcal{H}_{\epsilon_0} = \Big\{\big(\boldsymbol{U}, \boldsymbol{W}, \boldsymbol{V}, \boldsymbol{Z}\big) : \ \big[\boldsymbol{U},\ \boldsymbol{W}\big] \in \mathcal{O}(p_1, p_1) \text{ with } \boldsymbol{U} \in \mathbb{R}^{p_1 \times k},\ \big[\boldsymbol{V},\ \boldsymbol{Z}\big] \in \mathcal{O}(p_2, p_1)$$
$$\text{with } \boldsymbol{V} \in \mathbb{R}^{p_2 \times k}, \|\boldsymbol{U} - \boldsymbol{U}_{(0)}\|_F \leqslant \epsilon_0,\ [\boldsymbol{U}, \boldsymbol{W}] \begin{bmatrix} \boldsymbol{V}^\top \\ \boldsymbol{Z}^\top \end{bmatrix} = [\boldsymbol{U}_{(0)}, \boldsymbol{W}_{(0)}] \begin{bmatrix} \boldsymbol{V}_{(0)}^\top \\ \boldsymbol{Z}_{(0)}^\top \end{bmatrix} \Big\}.$$

For any fixed $\boldsymbol{\Sigma}_x \in \mathbb{S}_+^{p_1}, \boldsymbol{\Sigma}_y \in \mathbb{S}_+^{p_2}$ with $\kappa(\boldsymbol{\Sigma}_x) = \kappa_x, \kappa(\boldsymbol{\Sigma}_y) = \kappa_y$, consider the parametrization $\boldsymbol{\Sigma}_{xy} = \boldsymbol{\Sigma}_x \boldsymbol{\Phi} \Lambda \boldsymbol{\Psi}^\top \boldsymbol{\Sigma}_y$, for $0 \leqslant \lambda_{k+1} < \lambda_k < 1$, define

$$\mathcal{T}_{\epsilon_0} = \Big\{\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_x & \boldsymbol{\Sigma}_x^{1/2}(\lambda_k \boldsymbol{U}\boldsymbol{V}^\top + \lambda_{k+1} \boldsymbol{W}\boldsymbol{Z}^\top)\boldsymbol{\Sigma}_y^{1/2} \\ \boldsymbol{\Sigma}_y^{1/2}(\lambda_k \boldsymbol{V}\boldsymbol{U}^\top + \lambda_{k+1} \boldsymbol{Z}\boldsymbol{W}^\top)\boldsymbol{\Sigma}_x^{1/2} & \boldsymbol{\Sigma}_y \end{bmatrix},$$
$$\boldsymbol{\Phi} = \boldsymbol{\Sigma}_x^{-1/2}[\boldsymbol{U}, \boldsymbol{W}], \boldsymbol{\Psi} = \boldsymbol{\Sigma}_y^{-1/2}[\boldsymbol{V}, \boldsymbol{Z}], \big(\boldsymbol{U}, \boldsymbol{W}, \boldsymbol{V}, \boldsymbol{Z}\big) \in \mathcal{H}_{\epsilon_0}\Big\}.$$

It is straightforward to verify that $\mathcal{T}_{\epsilon_0} \subset \mathcal{F}(p_1, p_2, k, \lambda_k, \lambda_{k+1}, \kappa_x, \kappa_y)$. For any $\boldsymbol{\Sigma}_{(i)} \in \mathcal{T}_{\epsilon_0},\ i = 1, 2$, they yield to the parametrization,

$$\boldsymbol{\Sigma}_{(i)} = \begin{bmatrix} \boldsymbol{\Sigma}_x & \boldsymbol{\Sigma}_x^{1/2}(\lambda_k \boldsymbol{U}_{(i)}\boldsymbol{V}_{(i)}^\top + \lambda_{k+1} \boldsymbol{W}_{(i)}\boldsymbol{Z}_{(i)}^\top)\boldsymbol{\Sigma}_y^{1/2} \\ \boldsymbol{\Sigma}_y^{1/2}(\lambda_k \boldsymbol{V}_{(i)}\boldsymbol{U}_{(i)}^\top + \lambda_{k+1} \boldsymbol{Z}_{(i)}\boldsymbol{W}_{(i)}^\top)\boldsymbol{\Sigma}_x^{1/2} & \boldsymbol{\Sigma}_y \end{bmatrix},$$

where $\big(\boldsymbol{U}_{(i)}, \boldsymbol{W}_{(i)}, \boldsymbol{V}_{(i)}, \boldsymbol{Z}_{(i)}\big) \in \mathcal{H}_{\epsilon_0}$ and the leading-k canonical vectors are $\boldsymbol{\Phi}_{1:k}^{(i)} = \boldsymbol{\Sigma}_x^{-1/2}\boldsymbol{U}_{(i)}, \boldsymbol{\Psi}_{1:k}^{(i)} = \boldsymbol{\Sigma}_y^{-1/2}\boldsymbol{V}_{(i)}$. We define a semi-metric on $\mathcal{T}_{\epsilon_0}$ as

$$\rho(\boldsymbol{\Sigma}_{(1)}, \boldsymbol{\Sigma}_{(2)}) = \left\|\boldsymbol{P}_{\boldsymbol{\Sigma}_x^{1/2}\boldsymbol{\Phi}_{1:k}^{(1)}} - \boldsymbol{P}_{\boldsymbol{\Sigma}_x^{1/2}\boldsymbol{\Phi}_{1:k}^{(2)}}\right\|_F = \left\|\boldsymbol{P}_{\boldsymbol{U}_{(1)}} - \boldsymbol{P}_{\boldsymbol{U}_{(2)}}\right\|_F.$$

By Lemma 0.1,

$$D(\mathbb{P}_{\boldsymbol{\Sigma}_1} || \mathbb{P}_{\boldsymbol{\Sigma}_2}) = \frac{n\Delta^2(1 + \lambda_k\lambda_{k+1})}{2(1 - \lambda_k^2)(1 - \lambda_{k+1}^2)} \|\boldsymbol{U}_{(1)}\boldsymbol{V}_{(1)}^\top - \boldsymbol{U}_{(2)}\boldsymbol{V}_{(2)}^\top\|_F^2.$$

Further by the definition of $d_{KL}(T)$,

$$d_{KL}(T) = \frac{n\Delta^2(1+\lambda_k\lambda_{k+1})}{2(1-\lambda_k^2)(1-\lambda_{k+1}^2)} \sup_{\boldsymbol{\Sigma}_{(1)},\boldsymbol{\Sigma}_{(2)}\in\mathcal{T}_{\epsilon_0}} \|\boldsymbol{U}_{(1)}\boldsymbol{V}_{(1)}^\top - \boldsymbol{U}_{(2)}\boldsymbol{V}_{(2)}^\top\|_F^2. \tag{0.2}$$

To bound the Kullback-Leibler diameter, for any $\boldsymbol{\Sigma}_{(1)},\boldsymbol{\Sigma}_{(2)} \in \mathcal{T}_{\epsilon_0}$, by definition,

$$[\boldsymbol{U}_{(1)}, \boldsymbol{W}_{(1)}]\begin{bmatrix}\boldsymbol{V}_{(1)}^\top\\ \boldsymbol{Z}_{(1)}^\top\end{bmatrix} = [\boldsymbol{U}_{(2)}, \boldsymbol{W}_{(2)}]\begin{bmatrix}\boldsymbol{V}_{(2)}^\top\\ \boldsymbol{Z}_{(2)}^\top\end{bmatrix},$$

which implies that they are singular value decompositions of the same matrix. Therefore, there exists $\boldsymbol{Q} \in \mathcal{O}(p_1, p_1)$ such that

$$[\boldsymbol{U}_{(2)}, \boldsymbol{W}_{(2)}] = [\boldsymbol{U}_{(1)}, \boldsymbol{W}_{(1)}]\boldsymbol{Q}\ ,\ [\boldsymbol{V}_{(2)}, \boldsymbol{Z}_{(2)}] = [\boldsymbol{V}_{(1)}, \boldsymbol{Z}_{(1)}]\boldsymbol{Q}. \tag{0.3}$$

Decompose $\boldsymbol{Q}$ into four blocks such that

$$\boldsymbol{Q} = \begin{bmatrix}\boldsymbol{Q}_{11} & \boldsymbol{Q}_{12}\\ \boldsymbol{Q}_{21} & \boldsymbol{Q}_{22}\end{bmatrix}.$$

Substitute into (0.3),

$$\boldsymbol{U}_{(2)} = \boldsymbol{U}_{(1)}\boldsymbol{Q}_{11} + \boldsymbol{W}_{(1)}\boldsymbol{Q}_{21},\ \boldsymbol{V}_{(2)} = \boldsymbol{V}_{(1)}\boldsymbol{Q}_{11} + \boldsymbol{Z}_{(1)}\boldsymbol{Q}_{21}.$$

Then,

$$\begin{aligned}\|\boldsymbol{U}_{(2)} - \boldsymbol{U}_{(1)}\|_F^2 &= \|\boldsymbol{U}_{(1)}(\boldsymbol{Q}_{11} - \boldsymbol{I}_k) + \boldsymbol{W}_{(1)}\boldsymbol{Q}_{21}\|_F^2\\ &= \|\boldsymbol{U}_{(1)}(\boldsymbol{Q}_{11} - \boldsymbol{I}_k)\|_F^2 + \|\boldsymbol{W}_{(1)}\boldsymbol{Q}_{21}\|_F^2\\ &= \|\boldsymbol{Q}_{11} - \boldsymbol{I}_k\|_F^2 + \|\boldsymbol{Q}_{21}\|_F^2.\end{aligned}$$

The second equality is due to the fact that $\boldsymbol{U}_{(1)}$ and $\boldsymbol{W}_{(1)}$ have orthogonal column space and the third equality is valid because $\boldsymbol{U}_{(1)}, \boldsymbol{W}_{(1)} \in \mathcal{O}(p_1, k)$. By the same argument, we will have

$$\|\boldsymbol{V}_{(2)} - \boldsymbol{V}_{(1)}\|_F^2 = \|\boldsymbol{Q}_{11} - \boldsymbol{I}_k\|_F^2 + \|\boldsymbol{Q}_{21}\|_F^2.$$

Notice that

$$\begin{aligned}\|\boldsymbol{U}_{(1)}\boldsymbol{V}_{(1)}^\top - \boldsymbol{U}_{(2)}\boldsymbol{V}_{(2)}^\top\|_F^2 &= \|(\boldsymbol{U}_{(1)} - \boldsymbol{U}_{(2)})\boldsymbol{V}_{(1)} + \boldsymbol{U}_{(2)}(\boldsymbol{V}_{(1)} - \boldsymbol{V}_{(2)})\|_F^2\\ &\leqslant 2\|\boldsymbol{U}_{(1)} - \boldsymbol{U}_{(2)}\|_F^2 + 2\|\boldsymbol{V}_{(1)} - \boldsymbol{V}_{(2)}\|_F^2\\ &= 4\|(\boldsymbol{U}_{(1)} - \boldsymbol{U}_{(2)})\|_F^2\\ &\leqslant 8\left(\|(\boldsymbol{U}_{(1)} - \boldsymbol{U}_{(0)})\|_F^2 + \|(\boldsymbol{U}_{(0)} - \boldsymbol{U}_{(2)})\|_F^2\right)\\ &\leqslant 16\epsilon_0^2.\end{aligned}$$

Then, substitute into (0.2)

$$d_{KL}(T) \leqslant \frac{8n\Delta^2(1+\lambda_k\lambda_{k+1})}{(1-\lambda_k^2)(1-\lambda_{k+1}^2)}\epsilon_0^2. \tag{0.4}$$

Let $\mathcal{B}_{\epsilon_0} = \{\boldsymbol{U} \in O(p_1, k) : \|\boldsymbol{U} - \boldsymbol{U}_{(0)}\|_F \leqslant \epsilon_0\}$. Under the semi-metric $\widetilde{\rho}(\boldsymbol{U}_{(1)}, \boldsymbol{U}_{(2)}) = \|\boldsymbol{U}_{(1)}\boldsymbol{U}_{(1)}^\top - \boldsymbol{U}_{(2)}\boldsymbol{U}_{(2)}^\top\|_F$, we claim that the packing number of $\mathcal{H}_{\epsilon_0}$ is lower bounded by the packing number of $\mathcal{B}_{\epsilon_0}$. To prove this claim, it suffices to show that for any $\boldsymbol{U} \in \mathcal{B}_{\epsilon_0}$, there exists corresponding $\boldsymbol{W}, \boldsymbol{V}, \boldsymbol{Z}$ such that $(\boldsymbol{U}, \boldsymbol{W}, \boldsymbol{V}, \boldsymbol{Z}) \in \mathcal{H}_{\epsilon_0}$.

First of all, by definition, $\|\boldsymbol{U}-\boldsymbol{U}_0\|_F \leqslant \epsilon_0$. Let $\boldsymbol{W} \in \mathcal{O}(p_1, p_1-k)$ be the orthogonal complement of $\boldsymbol{U}$. Then $[\boldsymbol{U}, \boldsymbol{W}] \in \mathcal{O}(p_1, p_1)$ and therefore there exists $\boldsymbol{Q} \in \mathcal{O}(p_1, p_1)$ such that

$$[\boldsymbol{U}, \boldsymbol{W}] = [\boldsymbol{U}_{(0)}, \boldsymbol{W}_{(0)}]\boldsymbol{Q}.$$

Set $[\boldsymbol{V}, \boldsymbol{Z}] = [\boldsymbol{V}_{(0)}, \boldsymbol{Z}_{(0)}]\boldsymbol{Q} \in \mathcal{O}(p_2, p_1)$, then

$$[\boldsymbol{U}, \boldsymbol{W}]\begin{bmatrix}\boldsymbol{V}^\top \\ \boldsymbol{Z}^\top\end{bmatrix} = [\boldsymbol{U}_{(0)}, \boldsymbol{W}_{(0)}]\begin{bmatrix}\boldsymbol{V}_{(0)}^\top \\ \boldsymbol{Z}_{(0)}^\top\end{bmatrix},$$

which implies $(\boldsymbol{U}, \boldsymbol{W}, \boldsymbol{V}, \boldsymbol{Z}) \in \mathcal{H}_{\epsilon_0}$. Let

$$\epsilon = \alpha\epsilon_0 = c\left(\sqrt{k \wedge (p_1-k)} \wedge \sqrt{\frac{(1-\lambda_k^2)(1-\lambda_{k+1}^2)}{n\Delta^2(1+\lambda_k\lambda_{k+1})}k(p_1-k)}\right),$$

where $c \in (0,1)$ depends on α and is chosen small enough such that $\epsilon_0 = \epsilon/\alpha \in (0, \sqrt{2[k \wedge (p_1-k)]}\,)$. By Corollary 0.4,

$$\mathcal{M}(\mathcal{T}_{\epsilon_0}, \rho, \alpha\epsilon_0) = \mathcal{M}(\mathcal{H}_{\epsilon_0}, \widetilde{\rho}, \alpha\epsilon_0) \geqslant \mathcal{M}(\mathcal{B}_{\epsilon_0}, \widetilde{\rho}, \alpha\epsilon_0) \geqslant \left(\frac{1}{C\alpha}\right)^{k(p_1-k)}.$$

Apply Lemma 0.6 with $\mathcal{T}_{\epsilon_0}, \rho, \epsilon$,

$$\inf_{\widehat{\boldsymbol{\Phi}}_{1:k}} \sup_{\boldsymbol{\Sigma}\in\mathcal{F}} \mathbb{E}\left[\ \left\|\boldsymbol{P}_{\boldsymbol{\Sigma}_x^{1/2}\widehat{\boldsymbol{\Phi}}_{1:k}} - \boldsymbol{P}_{\boldsymbol{\Sigma}_x^{1/2}\boldsymbol{\Phi}_{1:k}}\right\|_F^2\right] \geqslant \sup_{T\subset\Theta}\sup_{\epsilon>0}\frac{\epsilon^2}{4}\left(1 - \frac{8c^2k(p_1-k)+\log 2}{k(p_1-k)\log\frac{1}{C\alpha}}\right).$$

Choose α small enough such that

$$1 - \frac{8c^2k(p_1-k)+\log 2}{k(p_1-k)\log\frac{1}{C\alpha}} \geqslant \frac{1}{2}.$$

Then the lower bound is reduced to

$$\begin{aligned}\inf_{\widehat{\boldsymbol{\Phi}}_{1:k}} \sup_{\boldsymbol{\Sigma}\in\mathcal{F}} \mathbb{E}\left[\ \left\|\boldsymbol{P}_{\boldsymbol{\Sigma}_x^{1/2}\widehat{\boldsymbol{\Phi}}_{1:k}} - \boldsymbol{P}_{\boldsymbol{\Sigma}_x^{1/2}\boldsymbol{\Phi}_{1:k}}\right\|_F^2\right] &\geqslant \frac{c^2}{8}\left\{\frac{(1-\lambda_k^2)(1-\lambda_{k+1}^2)}{n\Delta^2(1+\lambda_k\lambda_{k+1})}k(p_1-k) \wedge k \wedge (p_1-k)\right\} \\ &\geqslant C^2k\left\{\left(\frac{(1-\lambda_k^2)(1-\lambda_{k+1}^2)}{\Delta^2}\frac{p_1-k}{n}\right) \wedge 1 \wedge \frac{p_1-k}{k}\right\}\end{aligned}$$

The lower bound for operator norm error can be immediately obtained by noticing that $\boldsymbol{P}_{\boldsymbol{\Sigma}_y^{1/2}\widehat{\boldsymbol{\Psi}}_{1:k}} - \boldsymbol{P}_{\boldsymbol{\Sigma}_y^{1/2}\boldsymbol{\Psi}_{1:k}}$ has at most rank $2k$ and

$$\left\|\boldsymbol{P}_{\boldsymbol{\Sigma}_x^{1/2}\widehat{\boldsymbol{\Phi}}_{1:k}} - \boldsymbol{P}_{\boldsymbol{\Sigma}_x^{1/2}\boldsymbol{\Phi}_{1:k}}\right\|^2 \geqslant \frac{1}{2k}\left\|\boldsymbol{P}_{\boldsymbol{\Sigma}_x^{1/2}\widehat{\boldsymbol{\Phi}}_{1:k}} - \boldsymbol{P}_{\boldsymbol{\Sigma}_x^{1/2}\boldsymbol{\Phi}_{1:k}}\right\|_F^2.$$

To prove the results for $\widehat{\boldsymbol{\Psi}}_{1:k}$, one just need to change the definition of $\mathcal{H}_{\epsilon_0}$ to

$$\begin{aligned}\mathcal{H}_{\epsilon_0} = \Big\{(\boldsymbol{U}, \boldsymbol{W}, \boldsymbol{V}, \boldsymbol{Z}) :\ &\big[\boldsymbol{U},\ \boldsymbol{W}\big] \in \mathcal{O}(p_1, p_1) \text{ with } \boldsymbol{U} \in \mathbb{R}^{p_1\times k},\ \big[\boldsymbol{V},\ \boldsymbol{Z}\big] \in \mathcal{O}(p_2, p_1) \\ &\text{with } \boldsymbol{V} \in \mathbb{R}^{p_2\times k}, \|\boldsymbol{V}-\boldsymbol{V}_{(0)}\|_F \leqslant \epsilon_0,\ [\boldsymbol{U}, \boldsymbol{W}]\begin{bmatrix}\boldsymbol{V}^\top \\ \boldsymbol{Z}^\top\end{bmatrix} = [\boldsymbol{U}_{(0)}, \boldsymbol{W}_{(0)}]\begin{bmatrix}\boldsymbol{V}_{(0)}^\top \\ \boldsymbol{Z}_{(0)}^\top\end{bmatrix}\Big\}.\end{aligned}$$

By Corollary 0.4, the packing number now satisfies

$$\mathcal{M}(\mathcal{T}_{\epsilon_0}, \rho, \alpha\epsilon_0) = \mathcal{M}(\mathcal{H}_{\epsilon_0}, \widetilde{\rho}, \alpha\epsilon_0) \geqslant \mathcal{M}(\mathcal{B}_{\epsilon_0}, \widetilde{\rho}, \alpha\epsilon_0) \geqslant \left(\frac{1}{C\alpha}\right)^{k(p_2-k)}.$$

Following the same arguments for $\widehat{\mathbf{\Phi}}_{1:k}$, one can show that

$$\inf_{\widehat{\mathbf{\Psi}}_{1:k}} \sup_{\mathbf{\Sigma}\in\mathcal{F}} \mathbb{E}\left[\left\| \boldsymbol{P}_{\mathbf{\Sigma}_y^{1/2}\widehat{\mathbf{\Psi}}_{1:k}} - \boldsymbol{P}_{\mathbf{\Sigma}_y^{1/2}\mathbf{\Psi}_{1:k}} \right\|_F^2 \right] \geqslant C^2 k \left\{ \left(\frac{(1-\lambda_k^2)(1-\lambda_{k+1}^2)}{\Delta^2} \frac{p_2-k}{n} \right) \wedge 1 \wedge \frac{p_2-k}{k} \right\}$$

and

$$\left\| \boldsymbol{P}_{\mathbf{\Sigma}_y^{1/2}\widehat{\mathbf{\Psi}}_{1:k}} - \boldsymbol{P}_{\mathbf{\Sigma}_y^{1/2}\mathbf{\Psi}_{1:k}} \right\|^2 \geqslant \frac{1}{2k} \left\| \boldsymbol{P}_{\mathbf{\Sigma}_y^{1/2}\widehat{\mathbf{\Psi}}_{1:k}} - \boldsymbol{P}_{\mathbf{\Sigma}_y^{1/2}\mathbf{\Psi}_{1:k}} \right\|_F^2$$

0.4. Proof of Lemma 0.1

By simple algebra, the Kullback-Leibler divergence between two multivariate gaussian distributions satisfies

$$D(\mathbb{P}_{\mathbf{\Sigma}_{(1)}}||\mathbb{P}_{\mathbf{\Sigma}_{(2)}}) = \frac{n}{2}\Big\{ \mathrm{Tr}\left(\mathbf{\Sigma}_{(2)}^{-1}(\mathbf{\Sigma}_{(1)} - \mathbf{\Sigma}_{(2)}) \right) - \log\det(\mathbf{\Sigma}_{(2)}^{-1}\mathbf{\Sigma}_{(1)}) \Big\}.$$

Notice that

$$\mathbf{\Sigma}_{(i)} = \begin{bmatrix} \mathbf{\Sigma}_x^{1/2} & \\ & \mathbf{\Sigma}_y^{1/2} \end{bmatrix} \mathbf{\Omega}_{(i)} \begin{bmatrix} \mathbf{\Sigma}_x^{1/2} & \\ & \mathbf{\Sigma}_y^{1/2} \end{bmatrix},$$

where

$$\mathbf{\Omega}_{(i)} = \begin{bmatrix} \boldsymbol{I}_{p_1} & \lambda_1 \boldsymbol{U}_{(i)}\boldsymbol{V}_{(i)}^\top + \lambda_2 \boldsymbol{W}_{(i)}\boldsymbol{Z}_{(i)}^\top \\ \lambda_1 \boldsymbol{V}_{(i)}\boldsymbol{U}_{(i)}^\top + \lambda_2 \boldsymbol{Z}_{(i)}\boldsymbol{W}_{(i)}^\top & \boldsymbol{I}_{p_2} \end{bmatrix}.$$

Then,

$$D(\mathbb{P}_{\mathbf{\Sigma}_{(1)}}||\mathbb{P}_{\mathbf{\Sigma}_{(2)}}) = \frac{n}{2}\Big\{ \mathrm{Tr}(\mathbf{\Omega}_{(2)}^{-1}\mathbf{\Omega}_{(1)}) - (p_1+p_2) - \log\det(\mathbf{\Omega}_{(2)}^{-1}\mathbf{\Omega}_{(1)}) \Big\}.$$

Also notice that

$$\begin{aligned} \mathbf{\Omega}_{(i)} = \begin{bmatrix} \boldsymbol{I}_{p_1} & \\ & \boldsymbol{I}_{p_2} \end{bmatrix} &+ \frac{\lambda_1}{2} \begin{bmatrix} \boldsymbol{U}_{(i)} \\ \boldsymbol{V}_{(i)} \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_{(i)}^\top & \boldsymbol{V}_{(i)}^\top \end{bmatrix} - \frac{\lambda_1}{2} \begin{bmatrix} \boldsymbol{U}_{(i)} \\ -\boldsymbol{V}_{(i)} \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_{(i)}^\top & -\boldsymbol{V}_{(i)}^\top \end{bmatrix} \\ &+ \frac{\lambda_2}{2} \begin{bmatrix} \boldsymbol{W}_{(i)} \\ \boldsymbol{Z}_{(i)} \end{bmatrix} \begin{bmatrix} \boldsymbol{W}_{(i)}^\top & \boldsymbol{Z}_{(i)}^\top \end{bmatrix} - \frac{\lambda_2}{2} \begin{bmatrix} \boldsymbol{W}_{(i)} \\ -\boldsymbol{Z}_{(i)} \end{bmatrix} \begin{bmatrix} \boldsymbol{W}_{(i)}^\top & -\boldsymbol{Z}_{(i)}^\top \end{bmatrix}. \end{aligned}$$

Therefore $\mathbf{\Omega}_{(1)}, \mathbf{\Omega}_{(2)}$ share the same set of eigenvalues: $1+\lambda_1$ with multiplicity k, $1-\lambda_1$ with multiplicity k, $1+\lambda_2$ with multiplicity p_1-k, $1-\lambda_2$ with multiplicity p_1-k and 1 with multiplicity $2(p_2-p_1)$. This implies $\log\det(\mathbf{\Omega}_{(2)}^{-1}\mathbf{\Omega}_{(1)})) = 0$. On the other hand, by block inversion formula, we can compute

$$\mathbf{\Omega}_{(2)}^{-1} = \begin{bmatrix} \boldsymbol{I}_{p_1} + \frac{\lambda_1^2}{1-\lambda_1^2}\boldsymbol{U}_{(2)}\boldsymbol{U}_{(2)}^\top + \frac{\lambda_2^2}{1-\lambda_2}\boldsymbol{W}_{(2)}\boldsymbol{W}_{(2)}^\top & -\frac{\lambda_1}{1-\lambda_1^2}\boldsymbol{U}_{(2)}\boldsymbol{V}_{(2)}^\top - \frac{\lambda_2}{1-\lambda_2}\boldsymbol{W}_{(2)}\boldsymbol{Z}_{(2)}^\top \\ -\frac{\lambda_1}{1-\lambda_1^2}\boldsymbol{V}_{(2)}\boldsymbol{U}_{(2)}^\top - \frac{\lambda_2}{1-\lambda_2}\boldsymbol{Z}_{(2)}\boldsymbol{W}_{(2)}^\top & \boldsymbol{I}_{p_2} + \frac{\lambda_1^2}{1-\lambda_1^2}\boldsymbol{V}_{(2)}\boldsymbol{V}_{(2)}^\top + \frac{\lambda_2^2}{1-\lambda_2}\boldsymbol{Z}_{(2)}\boldsymbol{Z}_{(2)}^\top \end{bmatrix}.$$

Divide $\mathbf{\Omega}_{(2)}^{-1}\mathbf{\Omega}_{(1)}$ into blocks such that

$$\mathbf{\Omega}_{(2)}^{-1}\mathbf{\Omega}_{(1)} = \begin{bmatrix} \boldsymbol{J}_{11} & \boldsymbol{J}_{12} \\ \boldsymbol{J}_{21} & \boldsymbol{J}_{22} \end{bmatrix} \quad \textit{where} \quad \boldsymbol{J}_{11} \in \mathbb{R}^{p_1\times p_1},\ \boldsymbol{J}_{22} \in \mathbb{R}^{p_2\times p_2},$$

and

$$\begin{aligned} \boldsymbol{J}_{11} &= \frac{\lambda_1^2}{1-\lambda_1^2}(\boldsymbol{U}_{(2)}\boldsymbol{U}_{(2)}^\top - \boldsymbol{U}_{(2)}\boldsymbol{V}_{(2)}^\top\boldsymbol{V}_{(1)}\boldsymbol{U}_{(1)}^\top) + \frac{\lambda_2^2}{1-\lambda_2^2}(\boldsymbol{W}_{(2)}\boldsymbol{W}_{(2)} - \boldsymbol{W}_{(2)}\boldsymbol{Z}_{(2)}^\top\boldsymbol{Z}_{(1)}\boldsymbol{W}_{(1)}^\top) \\ &\quad - \frac{\lambda_1\lambda_2}{1-\lambda_1^2}(\boldsymbol{U}_{(2)}\boldsymbol{V}_{(2)}^\top\boldsymbol{Z}_{(1)}\boldsymbol{W}_{(1)}^\top) - \frac{\lambda_1\lambda_2}{1-\lambda_2^2}(\boldsymbol{W}_{(2)}\boldsymbol{Z}_{(2)}^\top\boldsymbol{V}_{(1)}\boldsymbol{U}_{(1)}^\top) \\ \boldsymbol{J}_{22} &= \frac{\lambda_1^2}{1-\lambda_1^2}(\boldsymbol{V}_{(2)}\boldsymbol{V}_{(2)}^\top - \boldsymbol{V}_{(2)}\boldsymbol{U}_{(2)}^\top\boldsymbol{U}_{(1)}\boldsymbol{V}_{(1)}^\top) + \frac{\lambda_2^2}{1-\lambda_2^2}(\boldsymbol{Z}_{(2)}\boldsymbol{Z}_{(2)} - \boldsymbol{Z}_{(2)}\boldsymbol{W}_{(2)}^\top\boldsymbol{W}_{(1)}\boldsymbol{Z}_{(1)}^\top) \\ &\quad - \frac{\lambda_1\lambda_2}{1-\lambda_1^2}(\boldsymbol{V}_{(2)}\boldsymbol{U}_{(2)}^\top\boldsymbol{W}_{(1)}\boldsymbol{Z}_{(1)}^\top) - \frac{\lambda_1\lambda_2}{1-\lambda_2^2}(\boldsymbol{Z}_{(2)}\boldsymbol{W}_{(2)}^\top\boldsymbol{U}_{(1)}\boldsymbol{V}_{(1)}^\top). \end{aligned}$$

We spell out the algebra for $tr(\boldsymbol{J}_{11})$, and $tr(\boldsymbol{J}_{22})$ can be computed in exactly the same fashion.

$$\begin{aligned} tr(\boldsymbol{U}_{(2)}\boldsymbol{U}_{(2)}^\top - \boldsymbol{U}_{(2)}\boldsymbol{V}_{(2)}^\top\boldsymbol{V}_{(1)}\boldsymbol{U}_{(1)}^\top) &= \frac{1}{2}tr(\boldsymbol{U}_{(2)}\boldsymbol{V}_{(2)}^\top\boldsymbol{V}_{(2)}\boldsymbol{U}_{(2)}^\top + \boldsymbol{U}_{(1)}\boldsymbol{V}_{(1)}^\top\boldsymbol{V}_{(1)}\boldsymbol{U}_{(1)}^\top - 2\boldsymbol{U}_{(2)}\boldsymbol{V}_{(2)}^\top\boldsymbol{V}_{(1)}\boldsymbol{U}_{(1)}^\top) \\ &= \frac{1}{2}\|\boldsymbol{U}_{(1)}\boldsymbol{V}_{(1)}^\top - \boldsymbol{U}_{(2)}\boldsymbol{V}_{(2)}\|_F^2. \end{aligned}$$

Similarly,

$$tr(\boldsymbol{W}_{(2)}\boldsymbol{W}_{(2)} - \boldsymbol{W}_{(2)}\boldsymbol{Z}_{(2)}^\top\boldsymbol{Z}_{(1)}\boldsymbol{W}_{(1)}^\top) = \frac{1}{2}\|\boldsymbol{W}_{(1)}\boldsymbol{Z}_{(1)}^\top - \boldsymbol{W}_{(2)}\boldsymbol{Z}_{(2)}\|_F^2.$$

By the assumption (0.1), i.e., $\boldsymbol{U}_{(1)}\boldsymbol{V}_{(1)}^\top + \boldsymbol{W}_{(1)}\boldsymbol{Z}_{(1)}^\top = \boldsymbol{U}_{(2)}\boldsymbol{V}_{(2)}^\top + \boldsymbol{W}_{(2)}\boldsymbol{Z}_{(2)}^\top$, we have

$$tr(\boldsymbol{W}_{(2)}\boldsymbol{W}_{(2)} - \boldsymbol{W}_{(2)}\boldsymbol{Z}_{(2)}^\top\boldsymbol{Z}_{(1)}\boldsymbol{W}_{(1)}^\top) = \frac{1}{2}\|\boldsymbol{U}_{(1)}\boldsymbol{V}_{(1)}^\top - \boldsymbol{U}_{(2)}\boldsymbol{V}_{(2)}\|_F^2.$$

Further,

$$\begin{aligned} tr(\boldsymbol{U}_{(2)}\boldsymbol{V}_{(2)}^\top\boldsymbol{Z}_{(1)}\boldsymbol{W}_{(1)}^\top) &= tr\left(\boldsymbol{U}_{(2)}\boldsymbol{V}_{(2)}^\top(\boldsymbol{U}_{(2)}\boldsymbol{V}_{(2)}^\top + \boldsymbol{W}_{(2)}\boldsymbol{Z}_{(2)}^\top - \boldsymbol{U}_{(1)}\boldsymbol{V}_{(1)}^\top)^\top\right) \\ &= tr\left(\boldsymbol{U}_{(2)}\boldsymbol{V}_{(2)}^\top(\boldsymbol{U}_{(2)}\boldsymbol{V}_{(2)}^\top - \boldsymbol{U}_{(1)}\boldsymbol{V}_{(1)}^\top)^\top\right) \\ &= \frac{1}{2}\|\boldsymbol{U}_{(1)}\boldsymbol{V}_{(1)}^\top - \boldsymbol{U}_{(2)}\boldsymbol{V}_{(2)}\|_F^2, \end{aligned}$$

and by the same argument,

$$tr(\boldsymbol{W}_{(2)}\boldsymbol{Z}_{(2)}^\top\boldsymbol{V}_{(1)}\boldsymbol{U}_{(1)}^\top) = \frac{1}{2}\|\boldsymbol{U}_{(1)}\boldsymbol{V}_{(1)}^\top - \boldsymbol{U}_{(2)}\boldsymbol{V}_{(2)}\|_F^2.$$

Sum these equations,

$$\begin{aligned} \mathrm{tr}(\boldsymbol{J}_{11}) &= \frac{1}{2}\left\{\frac{\lambda_1^2}{1-\lambda_1^2} + \frac{\lambda_2^2}{1-\lambda_2^2} - \frac{\lambda_1\lambda_2}{1-\lambda_1^2} - \frac{\lambda_1\lambda_2}{1-\lambda_2^2}\right\}\|\boldsymbol{U}_{(1)}\boldsymbol{V}_{(1)}^\top - \boldsymbol{U}_{(2)}\boldsymbol{V}_{(2)}\|_F^2 \\ &= \frac{\Delta^2(1+\lambda_1\lambda_2)}{2(1-\lambda_1^2)(1-\lambda_2^2)}\|\boldsymbol{U}_{(1)}\boldsymbol{V}_{(1)}^\top - \boldsymbol{U}_{(2)}\boldsymbol{V}_{(2)}\|_F^2. \end{aligned}$$

Repeat the argument for $\boldsymbol{J}_{22}$, one can show that

$$\mathrm{tr}(\boldsymbol{J}_{22}) = \mathrm{tr}(\boldsymbol{J}_{11}) = \frac{\Delta^2(1+\lambda_1\lambda_2)}{2(1-\lambda_1^2)(1-\lambda_2^2)}\|\boldsymbol{U}_{(1)}\boldsymbol{V}_{(1)}^\top - \boldsymbol{U}_{(2)}\boldsymbol{V}_{(2)}\|_F^2.$$

Therefore,

$$\begin{aligned} D(\mathbb{P}_{\boldsymbol{\Sigma}_{(1)}}||\mathbb{P}_{\boldsymbol{\Sigma}_{(2)}}) &= \frac{n}{2}tr(\boldsymbol{\Omega}_{(2)}^{-1}\boldsymbol{\Omega}_{(1)}) = \frac{n}{2}\left(tr(\boldsymbol{J}_{11}) + tr(\boldsymbol{J}_{22})\right) \\ &= \frac{n\Delta^2(1+\lambda_1\lambda_2)}{2(1-\lambda_1^2)(1-\lambda_2^2)}\|\boldsymbol{U}_{(1)}\boldsymbol{V}_{(1)}^\top - \boldsymbol{U}_{(2)}\boldsymbol{V}_{(2)}\|_F^2. \end{aligned}$$

References

[1] Cai, T. T., Z. Ma, Y. Wu, et al. (2013). Sparse pca: Optimal rates and adaptive estimation. *The Annals of Statistics 41*(6), 3074–3110.

[2] Gao, C., Z. Ma, Z. Ren, H. H. Zhou, et al. (2015). Minimax estimation in sparse canonical correlation analysis. *The Annals of Statistics 43*(5), 2168–2197.

[3] Gao, C., Z. Ma, and H. H. Zhou (2017). Sparse cca: Adaptive estimation and computational barriers. *The Annals of Statistics, to appear*.

[4] Szarek, S. J. (1982). Nets of grassmann manifold and orthogonal group. In *Proceedings of research workshop on Banach space theory (Iowa City, Iowa, 1981)*, Volume 169, pp. 185.

[5] Vu, V. Q., J. Lei, et al. (2013). Minimax sparse principal subspace estimation in high dimensions. *The Annals of Statistics 41*(6), 2905–2947.

[6] Yu, B. (1997). Assouad, fano, and le cam. In *Festschrift for Lucien Le Cam*, pp. 423–435. Springer.