Perturbation Bounds and Spectral Clustering

Math 586 Final Project

Adrian Cao Yi Luo Zhichen Xu

Washington University in St. Louis
Department of Statistics and Data Science

December 4, 2023

Presentation Overview

Review

Suppose our $n \times n$ symmetric matrix can be expressed as:

$$\hat{\boldsymbol{X}} = \boldsymbol{X} + \boldsymbol{Z}$$

Question: Are eigenvalues/eigenvectors of $\hat{\boldsymbol{X}}$ close to $\boldsymbol{X}$?

Weyl's inequality

Suppose $\boldsymbol{A}$, $\boldsymbol{B}$ are real symmetric matrices with eigenvalues $\lambda_n \geq \cdots \geq \lambda_1$, and $\gamma_n \geq \cdots \geq \gamma_1$, respectively. Then

$$\max_{1 \leq i \leq n} |\lambda_i - \gamma_i| \leq \|\boldsymbol{A} - \boldsymbol{B}\|_{op}.$$

$$\max_{1 \leq i \leq n} |\lambda_i(\hat{\boldsymbol{X}}) - \lambda_i(\boldsymbol{X})| \leq \|\boldsymbol{Z}\|_{op}$$

Eigensubspace decomposition

$\boldsymbol{X} \in \mathbb{R}^{n \times n}$ be a symmetric matrix

$$\boldsymbol{X} = \boldsymbol{U \Sigma U}^T$$

$\boldsymbol{U} \in \mathbb{R}^{n \times n}$ is an orthogonal matrix, and $\boldsymbol{\Sigma} = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$
Given an integer $r < n$, define the eigensubspace decomposition as

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{U} & \boldsymbol{U}_\perp \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{U}^T \\ \boldsymbol{U}_\perp^T \end{bmatrix}$$

$\boldsymbol{Z} \in \mathbb{R}^{n \times n}$ be a symmetric perturbation

$$\hat{\boldsymbol{X}} = \boldsymbol{X} + \boldsymbol{Z} = \begin{bmatrix} \hat{\boldsymbol{U}} & \hat{\boldsymbol{U}}_\perp \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_1 & \mathbf{0} \\ \mathbf{0} & \hat{\boldsymbol{\Sigma}}_2 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{U}}^T \\ \hat{\boldsymbol{U}}_\perp^T \end{bmatrix}$$

$\sin\Theta$ distance

For two $n \times r$ orthogonal columns $\boldsymbol{U}$,$\hat{\boldsymbol{U}}$
Suppose singular values of $\boldsymbol{U}^T\hat{\boldsymbol{U}}$ are $\sigma_1 \geqslant \sigma_2 \geqslant ... \geqslant \sigma_r \geqslant 0$
Principle angles are:

$$\Theta(\boldsymbol{U}, \hat{\boldsymbol{U}}) = \operatorname{diag}\left(\cos^{-1}(\sigma_1), \cos^{-1}(\sigma_2), \cdots, \cos^{-1}(\sigma_r)\right)$$

Quantitative measure of distance between the column spaces of $\boldsymbol{U}$ and $\hat{\boldsymbol{U}}$:

$$\|\sin\Theta(\hat{\boldsymbol{U}}, \boldsymbol{U})\| \text{ or } \|\sin\Theta(\hat{\boldsymbol{U}}, \boldsymbol{U})\|_F$$

Davis-Kahan $\sin\Theta$ theorem

Theorem

Let $\hat{\Sigma}, \Sigma \in \mathbb{R}^{p\times p}$ be symmetric with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_p$, and $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_p$, receptively. Fix $1 \leq r \leq s \leq p$, and let $d = s - r + 1$ and $V = (v_r, v_{r+1}, \ldots, v_s)$, $\hat{V} = (\hat{v}_r, \ldots, \hat{v}_s) \in \mathbb{R}^{p\times d}$ satisfying $\Sigma v_j = \lambda_j v_j$ and $\hat{\Sigma}\hat{v}_j = \hat{\lambda}_j \hat{v}_j$. If $\delta = \inf\{|\hat{\lambda} - \lambda| : \lambda \in [\lambda_s, \lambda_r], \hat{\lambda} \in (-\infty, \hat{\lambda}_{s-1}] \bigcup [\hat{\lambda}_{r+1}, \infty)\} > 0$, then

$$\|\sin\Theta\|_F \leqslant \frac{\|\hat{\Sigma} - \Sigma\|_F}{\delta}.$$

$$\|sin\Theta(\hat{\boldsymbol{U}}, \boldsymbol{U})\|_F \leqslant \frac{\|\boldsymbol{Z}\|_F}{\delta}$$

Asymmetric cases

$\boldsymbol{X} \in \mathbb{R}^{p_1 \times p_2}$ is approximately rank- r with the SVD $X = \boldsymbol{U\Sigma V}^T$, where a significant gap exists between $\sigma_r(X)$ and $\sigma_{r+1}(X)$.
Decompose X as follows,

$$\boldsymbol{X} = \left[\begin{array}{cc} \boldsymbol{U} & \boldsymbol{U}_\perp \end{array} \right] \left[\begin{array}{cc} \boldsymbol{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2 \end{array} \right] \left[\begin{array}{c} \boldsymbol{V}^T \\ \boldsymbol{V}_\perp^T \end{array} \right]$$

where $\boldsymbol{U} \in \mathbb{O}_{p_1,r}, \boldsymbol{V} \in \mathbb{O}_{p_2,r}, \boldsymbol{\Sigma}_1 = \operatorname{diag}(\sigma_1(\boldsymbol{X}), \cdots, \sigma_r(\boldsymbol{X})) \in \mathbb{R}^{r\times r}, \boldsymbol{\Sigma}_2 = \operatorname{diag}(\sigma_{r+1}(\boldsymbol{X}), \cdots) \in \mathbb{R}^{(p_1-r)\times(p_2-r)}, [\boldsymbol{U}\ \boldsymbol{U}_\perp] \in \mathbb{O}_{p_1}, [\boldsymbol{V}\ \boldsymbol{V}_\perp] \in \mathbb{O}_{p_2}$ are orthogonal matrices.
$\boldsymbol{Z} \in \mathbb{R}^{p_1 \times p_2}$ be a perturbation

$$\hat{\boldsymbol{X}} = \boldsymbol{X} + \boldsymbol{Z} = \left[\begin{array}{cc} \hat{\boldsymbol{U}} & \hat{\boldsymbol{U}}_\perp \end{array} \right] \left[\begin{array}{cc} \hat{\boldsymbol{\Sigma}}_1 & \mathbf{0} \\ \mathbf{0} & \hat{\boldsymbol{\Sigma}}_2 \end{array} \right] \left[\begin{array}{c} \hat{\boldsymbol{V}}^T \\ \hat{\boldsymbol{V}}_\perp^T \end{array} \right]$$

Wedin's $\sin\Theta$ theorem

Wedin's $\sin\Theta$ Theorem states that if $\sigma_{\min}\left(\hat{\boldsymbol{\Sigma}}_1\right) - \sigma_{\max}\left(\boldsymbol{\Sigma}_2\right) = \delta > 0$, then

$$\max\{\|\sin\Theta(\boldsymbol{V},\hat{\boldsymbol{V}})\|, \|\sin\Theta(\boldsymbol{U},\hat{\boldsymbol{U}})\|\} \leq \frac{\max\left\{\|\boldsymbol{Z}\hat{\boldsymbol{V}}\|, \left\|\hat{\boldsymbol{U}}^T\boldsymbol{Z}\right\|\right\}}{\delta},$$

$$\max\left\{\|\sin\Theta(\boldsymbol{V},\hat{\boldsymbol{V}})\|_F, \|\sin\Theta(\boldsymbol{U},\hat{\boldsymbol{U}})\|_F\right\} \leq \frac{\max\left\{\|\boldsymbol{Z}\hat{\boldsymbol{V}}\|_F, \left\|\hat{\boldsymbol{U}}^T\boldsymbol{Z}\right\|_F\right\}}{\delta}.$$

Additional Notation

Decompose the perturbation Z into four blocks

$$Z = Z_{11} + Z_{12} + Z_{21} + Z_{22},$$

where

$$\begin{aligned} Z_{11} &= \mathbb{P}_U Z \mathbb{P}_V, \quad & Z_{21} &= \mathbb{P}_{U_\perp} Z \mathbb{P}_V, \\ Z_{12} &= \mathbb{P}_U Z \mathbb{P}_{V_\perp}, \quad & Z_{22} &= \mathbb{P}_{U_\perp} Z \mathbb{P}_{V_\perp}, \end{aligned}$$

Define

$$z_{ij} := \|Z_{ij}\| \quad \text{for} \quad i,j = 1,2.$$

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

Separate Perturbation Bounds

Theorem 1 (Perturbation bounds for singular subspaces)

Let $X, \hat{X}$, and Z be given as before. Denote $\alpha := \sigma_{\min}(U^\top \hat{X} V), \beta := \|U_\perp^\top \hat{X} V_\perp\|$. If $\alpha^2 > \beta^2 + z_{12}^2 \wedge z_{21}^2$, then

$$\|\sin\Theta(V, \hat{V})\| \leq \frac{\alpha z_{12} + \beta z_{21}}{\alpha^2 - \beta^2 - z_{21}^2 \wedge z_{12}^2} \wedge 1,$$

$$\|\sin\Theta(V, \hat{V})\|_F \leq \frac{\alpha \|Z_{12}\|_F + \beta \|Z_{21}\|_F}{\alpha^2 - \beta^2 - z_{21}^2 \wedge z_{12}^2} \wedge \sqrt{r}.$$

$$\|\sin\Theta(U, \hat{U})\| \leq \frac{\alpha z_{21} + \beta z_{12}}{\alpha^2 - \beta^2 - z_{21}^2 \wedge z_{12}^2} \wedge 1,$$

$$\|\sin\Theta(U, \hat{U})\|_F \leq \frac{\alpha \|Z_{21}\|_F + \beta \|Z_{12}\|_F}{\alpha^2 - \beta^2 - z_{21}^2 \wedge z_{12}^2} \wedge \sqrt{r}.$$

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

Separate Perturbation Bounds

Assumption in Theorem 1

$\alpha^2 > \beta^2 + z_{12}^2 \wedge z_{21}^2$:
It ensures the amplitude of $U^\top \hat{X} V = \Sigma + U^\top Z V$ dominates those of $U_\perp^\top \hat{X} V_\perp = \Sigma_2 + U_\perp^\top Z V_\perp$, $U^\top Z V_\perp$ and $U_\perp^\top Z V$, so that $\hat{U}$, $\hat{V}$ can be close to U, V respectively.

For lower bounds, we define the following class of (X, Z) pairs of $p_1 \times p_2$ matrices and perturbations,

$$\mathcal{F}_{r,\alpha,\beta,z_{21},z_{12}} = \{(X, Z) :$$
$$\sigma_{\min}\left(U^\top \hat{X} V\right) \geq \alpha, \left\|U_\perp^\top \hat{X} V_\perp\right\| \leq \beta, \|Z_{12}\| \leq z_{12}, \|Z_{21}\| \leq z_{21}\right\}.$$

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

Separate Perturbation Bounds

(part of) Theorem 2 (Perturbation Lower Bounds)

If $\alpha^2 \leq \beta^2 + z_{12}^2 \wedge z_{21}^2$ and $r \leq \frac{p_1 \wedge p_2}{2}$, then

$$\inf_{\tilde{V}} \sup_{(X,Z)\in\mathcal{F}} \|\sin\Theta(V, \tilde{V})\| \geq \frac{1}{2\sqrt{2}}.$$

Provided that $\alpha^2 > \beta^2 + z_{12}^2 + z_{21}^2$, $r \leq \frac{p_1 \wedge p_2}{2}$ we have the following lower bound for all estimate $\tilde{V} \in O_{p_2 \times r}$ based on the observations $\hat{X}$,

$$\inf_{\tilde{V}} \sup_{(X,Z)\in\mathcal{F}} \|\sin\Theta(V, \tilde{V})\| \geq \frac{1}{8\sqrt{10}} \left(\frac{\alpha z_{12} + \beta z_{21}}{\alpha^2 - \beta^2 - z_{12}^2 \wedge z_{21}^2} \wedge 1 \right).$$

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

Sketch of Proof

Proposition 1 (for proving Theorem 1)

Suppose $A \in \mathbb{R}^{p_1 \times p_2}$, $\tilde{V} = \begin{bmatrix} V & V_\perp \end{bmatrix} \in \mathbb{O}_{p_2}$ are right singular vectors of A, $V \in \mathbb{O}_{p_2,r}$, $V_\perp \in \mathbb{O}_{p_2,p_2-r}$ correspond to the first r and last $(p_2 - r)$ singular vectors respectively. $\tilde{W} = \begin{bmatrix} W & W_\perp \end{bmatrix} \in \mathbb{O}_{p_2}$ is any orthogonal matrix with $W \in \mathbb{O}_{p_2,r}$, $W_\perp \in \mathbb{O}_{p_2,p_2-r}$. Given that $\sigma_r(AW) > \sigma_{r+1}(A)$, we have

$$\|\sin\Theta(V,W)\| \le \frac{\sigma_r(AW)\left\|\mathbb{P}_{(AW)} A W_\perp\right\|}{\sigma_r^2(AW) - \sigma_{r+1}^2(A)} \wedge 1,$$

$$\|\sin\Theta(V,W)\|_F \le \frac{\sigma_r(AW)\left\|\mathbb{P}_{(AW)} A W_\perp\right\|_F}{\sigma_r^2(AW) - \sigma_{r+1}^2(A)} \wedge \sqrt{r}.$$

Set $A = \hat{X}$, $\bar{W} = \begin{bmatrix} V & V_\perp \end{bmatrix}$, $\tilde{V} = \begin{bmatrix} \hat{V} & \hat{V}_\perp \end{bmatrix}$ and Theorem 1 can be derived from these inequalities.

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

Sketch of Proof

Lemma 3 (SVD of 2-by-2 matrices: for proving Theorem 2)

1. $B = \begin{bmatrix} a & b \\ 0 & d \end{bmatrix}$, $\quad a, b, d \geq 0, \quad a^2 \leq b^2 + d^2$. Suppose $V = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix}$ is the right singular vectors of B, then
$$|v_{12}| = |v_{21}| \geq \frac{1}{\sqrt{2}}$$
2. $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, $\quad a, b, c, d \geq 0, \quad a^2 > d^2 + b^2 + c^2$. Suppose $V = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix}$ is the right singular vectors of A, then
$$|v_{12}| = |v_{21}| \geq \frac{1}{\sqrt{10}} \left(\frac{ab + cd}{a^2 - d^2 - b^2 \wedge c^2} \wedge 1 \right)$$

(Theorem 2): $$\inf_{\tilde{V}} \sup_{(X,Z)\in\mathcal{F}} \| \sin \Theta(V, \tilde{V}) \| \geq \frac{1}{2\sqrt{2}}$$
$$\inf_{\tilde{V}} \sup_{(X,Z)\in\mathcal{F}} \| \sin \Theta(V, \tilde{V}) \| \geq \frac{1}{8\sqrt{10}} \left(\frac{\alpha z_{12} + \beta z_{21}}{\alpha^2 - \beta^2 - z_{12}^2 \wedge z_{21}^2} \wedge 1 \right).$$

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

Comparison with Wedin's Theorem

Define the class of distributions $\mathcal{G}_\tau$ for some $\tau > 0$:

$$\text{If } Z \sim \mathcal{G}_\tau, \text{ then } \mathbb{E}Z = 0, \operatorname{Var}(Z) = 1, \mathbb{E}\exp(tZ) \le \exp(\tau t), \forall t \in \mathbb{R}.$$

The distribution of the entries of Z, Z_{ij}, is assumed to satisfy

$$Z_{ij} \overset{iid}{\sim} \mathcal{G}_\tau, \quad 1 \le i \le p_1, 1 \le j \le p_2.$$

Theorem 3 (Upper Bound in Low-rank Matrix Denoising)

Suppose $X = U\Sigma V^\top \in \mathbb{R}^{p_1 \times p_2}$ is of rank-r. There exists constants $C > 0$ that only depends on τ such that

$$\mathbb{E}\|\sin\Theta(V, \hat{V})\|^2 \le \frac{Cp_2\left(\sigma_r^2(X) + p_1\right)}{\sigma_r^4(X)} \wedge 1,$$

$$\mathbb{E}\|\sin\Theta(V, \hat{V})\|_F^2 \le \frac{Cp_2 r\left(\sigma_r^2(X) + p_1\right)}{\sigma_r^4(X)} \wedge r.$$

$$\mathbb{E}\|\sin\Theta(U, \hat{U})\|^2 \le \frac{Cp_1\left(\sigma_r^2(X) + p_2\right)}{\sigma_r^4(X)} \wedge 1,$$

$$\mathbb{E}\|\sin\Theta(U, \hat{U})\|_F^2 \le \frac{Cp_1 r\left(\sigma_r^2(X) + p_2\right)}{\sigma_r^4(X)} \wedge r.$$

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

Comparison with Wedin's Theorem

Consider $X \in \mathbb{R}^{p_1 \times p_2}$ is a rank-r matrix with $r \leq p_1 \ll p_2$, and $Z \in \mathbb{R}^{p_1 \times p_2}$ with i.i.d. standard normal entries. By RMT, $\alpha \geq \sigma_r(X) - \|Z_{11}\| \geq \sigma_r(X) - C(\sqrt{p_1} + \sqrt{p_2}), \quad \beta \leq C(\sqrt{p_1} + \sqrt{p_2}), \quad z_{12} \leq C\sqrt{p_2}, \quad z_{21} \leq C\sqrt{p_1}$ for some constant $C > 0$ with high probability.

Wedin's Theorem:	Theorem 3:
$\max\{\|\sin\Theta(V,\hat{V})\|, \|\sin\Theta(U,\hat{U})\|\} \leq \frac{C\max\{\sqrt{p_1},\sqrt{p_2}\}}{\sigma_r(X)}$	$\|\sin\Theta(V,\hat{V})\| \leq \frac{C\sqrt{p_2}}{\sigma_r(X)}$ $\|\sin\Theta(U,\hat{U})\| \leq \frac{C\sqrt{p_1}}{\sigma_r(X)}$

The bound by applying Wedin's theorem is sub-optimal for $\|\sin\Theta(U,\hat{U})\|$ if $p_2 \gg p_1$.

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

Spectral Clustering Algorithm

Algorithm

Input: Data points $X = \{x_1, x_2, \ldots, x_n\}$, number of clusters k.

1. Perform SVD on X to have

$$X = \sum_{i=1}^{p \wedge n} \hat{\lambda}_i \hat{u}_i \hat{v}_i^T$$

   where $\hat{\lambda}_1 \leq \hat{\lambda}_2 \leq \ldots \leq \hat{\lambda}_{p \wedge n} \leq 0$ and $\{\hat{u}_i\}_{i=1}^{p \wedge n} \in \mathbb{R}^p, \{\hat{v}_i\}_{i=1}^{p \wedge n} \in \mathbb{R}^p$
   Let $\hat{U}_{1:r} = (\hat{u}_1, \ldots, \hat{u}_r) \in \mathbb{R}^{p \times r}$

2. Perform clustering algorithm, say k-means, on the columns of $\hat{U}_{1:r}^T X$

Output: Clusters $C_1, C_2, \ldots, C_k$

Note, this is slightly different from the spectral clustering introduced in class as we have asymmetric X here.

Perturbation bound by Davis-Kahan Theorem (Example from Class)

Assume $p \to 0$ for a certain sparsity range, we have bound of order $\|E\|_{op} \leq c\sqrt{n}$ w.h.p..
Let $\delta = \min(\frac{\lambda_1}{n}, \frac{\lambda_1 - \lambda_2}{n})$. By Davis-Kahan, we have

$$\sin\Theta(\hat{V}_2, V_2) \leq \frac{\|E\|_{op}}{n \times \delta} \leq \frac{c\sqrt{n}}{n\delta} = \frac{c}{\sqrt{n}\delta}$$

Hence, the misspecification rate would be bounded

$$\mathcal{L}(z, z^*) = \min_{\phi \in \Phi} \frac{1}{n} \sum_{i \in [n]} \mathbb{I}\{z_i = \phi(z_i^*)\} \leq \frac{c^2}{\delta^2}$$

Here, z^* is the true cluster, and $\Phi = \{\phi : \phi$ is a bijection from [k] to [k] $\}$

Push it Further

Example we have in class would be for two cluster, and symmetric adjacency matrix. Now we would like to discuss a little more about asymmetric matrix with k clusters.
Let $\Delta = \min_{a,b\in[k]:a\neq b} \|\theta_a^* - \theta_b^*\|$, here θ_a^*, θ_b^* is the centers of the cluster.

Proposition 3.1 (Simplified)

$$\mathcal{L}(\hat{z}, z^*) \leq \frac{Ck\|\epsilon\|^2}{n\Delta^2}$$

Zhang and Zhou used the singular subspace perturbation by leave-one-out analysis to provide a sharper bound for this.

Singular Subspace Perturbation

We consider a mixture model with k centers $\theta_1^*, \theta_2^*, \ldots, \theta_k^* \in \mathbb{R}^P$ and a cluster assignment vector $z^* \in [k]^n$. Hence we have our observation

$$X_i = \theta_{z_i^*}^* + \epsilon_i$$

Denote X_{-i} be the submatrix of X with its ith column removed. And $\hat{U}_{1:r}$ and $\hat{U}_{-i,1:r}$ be the leading r left singular vector of X and X_{-i}, respectively. Define $\beta = \frac{1}{n/k} \min_{a \in [k]} |\{i : z_i^* = a\}|$ such that $\beta n/k$ is the smallest cluster size. κ as the rank of spectrum for signal.

Theorem 2.2

Assume $\beta n/k \geq 10$ and $\rho_0 = \frac{\lambda_\kappa}{\|\epsilon\|} > 16$

For any $i \in [n]$, we have

$$\|\hat{U}_{1:\kappa}\hat{U}_{1:\kappa}^T - \hat{U}_{-i,1:\kappa}\hat{U}_{-i,1:\kappa}^T\|_F \leq \frac{128}{\rho_0}\left(\sqrt{\frac{k\kappa}{\beta n}} + \frac{\|\hat{U}_{-i,1:\kappa}\hat{U}_{-i,1:\kappa}^T\epsilon_i\|}{\lambda_\kappa}\right)$$

Trace back to the $\sin\Theta$ theorem

Note by Lemma in Cai and Zhang, we have

Lemma 1(Simplified)

$$\|\hat{V}\hat{V}^T - VV^T\|_F = \sqrt{2}\|\sin\Theta(\hat{V}, V)\|_F$$

This also links to the $\sin\theta$ theorem previously.

Back to the Spectral Clustering

For the perturbation analysis, we cares about $\|\hat{U}_{1:r}\hat{U}_{1:r}^T\epsilon_i\|$ is small enough so that $\hat{U}_{1:r}\hat{U}_{1:r}^T X_i$ is close enough to $\hat{U}_{1:r}\hat{U}_{1:r}^T\theta^*_{z_i^*}$ and z_i^* is thus correctly recovered.
Note we have

$$\|\hat{U}_{1:\kappa}\hat{U}_{1:\kappa}^T\epsilon_i\| \leq \|\hat{U}_{-i,1:\kappa}\hat{U}_{-i,1:\kappa}^T\epsilon_i\| + \|U_{1:\kappa}\hat{U}_{1:\kappa}^T\epsilon_i - \hat{U}_{-i,1:\kappa}\hat{U}_{-i,1:\kappa}^T\epsilon_i\|_F\|\epsilon_i\| \tag{1}$$

Hence by the theorem 2.2 above, we have well-controlled perturbation for the leave-one-out singular space.

More for Spectral Analysis

Lemma 3.2 (Simplified)

With assumption as above,

$$\mathcal{I}(\hat{z}_i \neq \phi(z_i^*)) \leq \mathcal{I}\{(1 - C(\phi_0^{-1} + \rho_0^{-2}))\Delta \leq 2\|\hat{U}_{-i,1:\kappa}\hat{U}^T_{-i,1:\kappa}\epsilon_i\|\}$$

This lemma is crucial for that if we want a similar bound for this by Wedin's theorem, we would need to require the second term on the RHS of equation (1) to be much smaller than Δ. But with the Theorem 2.2, we build it more naturally.

References

Per-Åke Wedin (1972)
Perturbation bounds in connection with singular value decomposition
BIT Numerical Mathematics 12, pages99–111 (1972)

T. Tony Cai and Anru Zhang (2016)
Rate-Optimal Perturbation Bounds for Singular Subspaces with Applications to High-Dimensional Statistics
arXiv:1605.00353.

Anderson Y. Zhang and Harrison H. Zhou (2022)
Leave-One-Out Singular Subspace Perturbation Analysis for Spectral Clustering
arXiv:2205.14855.

The End

Questions? Comments?