SPECTRAL PERTURBATION THEORY

NOTES BY PARKER KNIGHT

1. Introduction

This set of notes will closely follow the second chapter of [1].

The general problem is defined as follows: suppose there exists some matrix of interest M^*, and we observe a perturbed version $M = M^* + E$ where E is a perturbation matrix. We are interested in characterizing how the spectral properties of M^* (i.e. the eigenspace or singular subspace) change in light of the perturbation E. To do so, we will describe the classic Davis-Kahan $\sin \Theta$ theorem for symmetric matrices, and Wedin's extension to general matrices. But first, we describe various metrics for describing the distance between subspaces.

Throughout the note set, we let $\mathcal{U}^*$ and $\mathcal{U}$ denote two r dimensional subspaces in $\mathbb{R}^n$. Let U^* and U be matrices in $\mathbb{R}^{n\times r}$ whose columns form an orthonormal basis for $\mathcal{U}^*$ and $\mathcal{U}$ respectively. For any matrix A, let $A_\perp$ denote its orthogonal compliment. Let $\mathcal{O}^{r\times r}$ denote the set of orthogonal matrices in $\mathbb{R}^{r\times r}$, and let $||.||_{op}$ denote the operator norm (largest singular value) of a matrix.

2. Distance between subspaces

A key challenge in describing the distance between subspaces is the notion of rotational ambiguity, namely for for any rotation matrix $R \in R^{r\times r}$, we have that UR is also an orthogonal basis for $\mathcal{U}$. So, even when $\mathcal{U}^* = \mathcal{U}$, we may have $|||U - U^*||| \neq 0$ for our matrix norm of choice $|||.|||$, depending on how the bases are rotated. Any useful distance metric on subspaces much account for this rotational ambiguity. Following the approach of [1], we describe a few difference useful choices of metric.

2.1. Distance with optimal rotation.

A natural approach to addressing the rotational invariance problem is to simply choose the rotation of U which is closest in norm to U^*. This yields the following distance metric:

$$\text{dist}(U, U^*) := \min_{R\in\mathcal{O}^{r\times r}} |||UR - U^*|||$$

2.2. Distance between projections.

Recall that the projection onto $\mathcal{U}$ is given by UU^T. A useful fact is that this projection matrix is unchanged by its rotation: for any $R \in \mathcal{O}^{r\times r}$, we have $UR(UR)^T = URR^TU^T = UU^T$. This motivates the following metric between subspaces:

$$\text{dist}_p(U, U^*) := |||UU^T - U^*U^{*T}|||$$

2.3. **Distance via principal angles.** Let $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_r \geq 0$ be the singular values of U^TU^* arranged in descending order. By the properties of the operator norm[1], we know that $||U^TU^*||_{op} \leq ||U||_{op}||U^*||_{op} = 1$. So, it follows that the singular values all fall with $[0.1]$. Using this fact, we define the principal angles between $\mathcal{U}$ and $\mathcal{U}^*$ as

$$\theta_i := \arccos(\sigma_i)$$

for $i = 1, ..., r$. These principal angles satisfy

$$0 \leq \theta_1 \leq \theta_r \leq \pi/2$$

Join these angles into a diagonal matrix Θ, and define the following metric

$$\text{dist}_{\sin}(U, U^*) := |||\sin\Theta|||$$

where $\sin$ is applied elementwise to Θ, and $|||.|||$ is a matrix norm of choice.

It can be shown (ommitted here) that these three distance metrics are equivalent up to a scaling factor of $\sqrt{2}$ when the norm of choice is the operator norm or the Frobenius norm. Following the example of [1], we will use the optimal rotation metric throughout these notes.

3. Davis-Kahan

Now we are able to develop a theory for understanding the effect of perturbation on eigenspaces.

Throughout this section, let M^* and $M = M^* + E$ be $n \times n$ real symmetric matrices, with eigendecompositions

$$M^* = \begin{bmatrix} U^* & U_\perp^* \end{bmatrix} \begin{bmatrix} \Lambda^* & 0 \\ 0 & \Lambda_\perp^* \end{bmatrix} \begin{bmatrix} U^{*T} \\ U_\perp^{*T} \end{bmatrix}$$
$$M = \begin{bmatrix} U & U_\perp \end{bmatrix} \begin{bmatrix} \Lambda & 0 \\ 0 & \Lambda_\perp \end{bmatrix} \begin{bmatrix} U^T \\ U_\perp^T \end{bmatrix}$$

where U^* and U lie in $\mathbb{R}^{n\times r}$. Now we present the Davis-Kahan theorem.

Theorem 1. *Suppose that*

$$diag(\Lambda^*) \subseteq [\alpha, \beta]$$
$$diag(\Lambda_\perp) \subseteq (-\infty, \alpha - \Delta] \cup [\beta + \Delta, \infty)$$

for some $\alpha, \beta \in \mathbb{R}$ and $\Delta > 0$. Then:

$$dist_{op}(U, U^*) \leq \sqrt{2}||\sin\Theta||_{op} \leq \frac{\sqrt{2}||EU^*||_{op}}{\Delta} \leq \frac{\sqrt{2}||E||_{op}}{\Delta}$$
$$dist_F(U, U^*) \leq \sqrt{2}||\sin\Theta||_F \leq \frac{\sqrt{2}||EU^*||_F}{\Delta} \leq \frac{\sqrt{2}||E||_{op}}{\Delta}$$

[1] This follows from the definition: Recall for any A and x with agreeing dimension, we have $||Ax||_2 \leq ||A||_{op}||x||_2$. It follows that $||BAx||_2 \leq ||B||_{op}||Ax||_2 \leq ||B||_{op}||A||_{op}||x||_2$. Thus, $||BA||_{op} \leq ||B||_{op}||A||_{op}$.

Proof. We prove the theorem by controlling the distance $|||U_\perp^T U^*|||$ where $|||.|||$ is unitarily invariant.

Without loss of generality, we constrain ourselves to the case where $\alpha = -\beta \leq 0$[2]. This assumption, combined with the assumption of the theorem, give us the following:

$$||\Lambda^*||_{op} \leq \beta$$

$$\sigma_{\min}(\Lambda^\perp) \geq \beta + \Delta$$

Now we can study $U_\perp^T U^*$. Using the properties of eigenvectors, we first observe the following identity:

$$U_\perp^T E U^* = U_\perp^T (M - M^*) U^* = \Lambda_\perp U_\perp^T U^* - U_\perp^T U^* \Lambda^*$$

Letting $R := EU^*$, we apply the triangle inequality to obtain:

$$\begin{aligned}
|||U_\perp^T R||| &\geq |||\Lambda_\perp U_\perp^T U^*||| - |||U_\perp^T U^* \Lambda^*||| \\
&\geq \sigma_{\min}(\Lambda_\perp)|||U_\perp^T U^*||| - ||\Lambda^*||_{op}|||U_\perp^T U^*||| \\
&\geq (\beta + \Delta - \beta)|||U_\perp^T U^*||| = \Delta|||U_\perp^T U^*|||
\end{aligned}$$

It immediately follows that

$$|||U_\perp^T U^*||| \leq \frac{|||U_\perp^T U^*}{\Delta} \leq \frac{|||R|||}{\Delta} = \frac{|||EU^*|||}{\Delta}$$

and we are done. □

4. Wedin

References

[1] Y. Chen, Y. Chi, J. Fan, and C. Ma, “Spectral methods for data science: A statistical perspective,” vol. 14, no. 5, pp. 566–806.

[2] If case is violated, we can ’center’ the matrices M and M^* and recover the same setting; see [1].