1 Notes on matrix perturbation and Davis-Kahan $\sin(\Theta)$ theorem

In many situations, there is a symmetric matrix of interest $A \in \mathbb{R}^{n \times n}$, but one only has a perturbed version of it $\tilde{A} = A + H$ (H is a "small" symmetric matrix). How is $\tilde{A}$ affected by H?

Example: PCA. Let $A = \text{cov}(X)$ for some random vector X, and let $\tilde{A}$ be the sample covariance matrix on independent copies of X. If X is concentrated on a low dimensional subspace, then we can hope to discover this subspace from the principal components of $\tilde{A}$. How accurate is the subspace we find?

1.1 Spectral theorem

A non-zero vector v is an eigenvector of A if $Av = \lambda v$ for some scalar λ called the corresponding eigenvalue.

Theorem 1. *If $A \in \mathbb{R}^{n \times n}$ is symmetric, then there is an orthonormal basis $\{v_1, \ldots, v_n\}$ consisting of eigenvectors of A with real corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$:*

$$A = \lambda_1 v_1 {v_1}^* + \ldots + \lambda_n v_n {v_n}^*.$$

1.2 Eigenvalues

How are the eigenvalues of $\tilde{A}$ affected by H?

Let $\lambda_i(M)$ be the ith largest eigenvalue of a matrix M. Then

$$\begin{aligned} \lambda_1(\tilde{A}) &= \max_{\|u\|=1} u^*(A+H)u \\ &\leq \max_{\|u\|=1} u^*Au + \max_{\|u\|=1} u^*Hu \\ &= \lambda_1(A) + \lambda_1(H). \end{aligned}$$

Also, letting v_1 be the top eigenvector of A,

$$\begin{aligned} \lambda_1(\tilde{A}) &\geq {v_1}^*(A+H)v_1 \\ &= \lambda_1(A) + {v_1}^*Hv_1 \\ &\geq \lambda_1(A) + \lambda_n(H). \end{aligned}$$

Therefore

$$\lambda_1(A) + \lambda_n(H) \ \leq \ \lambda_1(\tilde{A}) \ \leq \ \lambda_1(A) + \lambda_1(H).$$

This can be extended to the 2nd, 3rd, *etc.* eigenvalues.

Theorem 2 (Weyl)**.** *For $i = 1, \ldots, n$:*

$$\lambda_i(A) + \lambda_n(H) \ \leq \ \lambda_i(\tilde{A}) \ \leq \ \lambda_i(A) + \lambda_1(H).$$

Therefore the (ordered) eigenvalues of a matrix are fairly stable with respect to a small perturbation.

1.3 Eigenvectors, eigenspaces

An eigenspace of A is the span of some eigenvectors of A. We can decompose A into its action on an eigenspace S and its action on the orthogonal complement $S^\perp$:

$$A = E_0 A_0 {E_0}^* + E_1 A_1 {E_1}^*$$

where E_0 is an orthonormal basis for S (*e.g.*, the eigenvectors of A that span S), and E_1 is an orthonormal basis for $S^\perp$ (this follows from the spectral theorem). We can similarly decompose $\tilde{A} = A + H$ with respect to a "corresponding" eigenspace $\tilde{S}$ (with $\dim \tilde{S} = \dim S$):

$$\tilde{A} = F_0 \Lambda_0 {F_0}^* + F_1 \Lambda_1 {F_1}^*.$$

How is close is $\tilde{S}$ to S?

A few things to consider:

1. How are we choosing the eigenspace S of A, and what is a suitable corresponding eigenspace $\tilde{S}$ of $\tilde{A}$? (Or vice versa.)

2. How do we measure the closeness between subspaces?

3. Under what conditions will the subspaces be close?

Suppose we find a few eigenvalues of $\tilde{A}$ that somehow stand out from the rest. For instance, as in PCA, we may find the first few eigenvalues to be much larger than the rest. Let $\tilde{S}$ be the corresponding eigenspace. If there is a similarly outstanding group of eigenvalues of A, then the hope is that the corresponding eigenspace S will be close to $\tilde{S}$ in some sense. For instance, we may be interested in how well $\tilde{S}$ approximates vectors in S. Any vector in S can be written as $E_0\alpha$ for some $\alpha \in \mathbb{R}^{\dim S}$; the projection of this vector onto $\tilde{S}$ is $F_0 {F_0}^* E_0 \alpha$. Then

$$\begin{aligned} \|E_0\alpha - F_0{F_0}^*E_0\alpha\| &= \|(I - F_0{F_0}^*)E_0\alpha\| \\ &= \|F_1{F_1}^*E_0\alpha\| \\ &= \|{F_1}^*E_0\alpha\|. \end{aligned}$$

Therefore vectors in S will be well-approximated by $\tilde{S}$ if ${F_1}^*E_0$ is "small".

The condition we will need is separation between the eigenvalues corresponding to S and those corresponding to $\tilde{S}^\perp$. Suppose the eigenvalues corresponding to S are all contained in an interval $[a, b]$. Then we will require that the eigenvalues corresponding to $\tilde{S}^\perp$ be excluded from the interval $(a - \delta, b + \delta)$ for some $\delta > 0$. To see why this is necessary, consider the following example:

$$A := \begin{bmatrix} 1+\delta & 0 \\ 0 & 1-\delta \end{bmatrix}, \quad H := \begin{bmatrix} -\delta & \delta \\ \delta & \delta \end{bmatrix}, \quad \tilde{A} := \begin{bmatrix} 1 & \delta \\ \delta & 1 \end{bmatrix}.$$

Here, the "size" of H is comparable to the gap between the relevant eigenvalues. Then the eigenvalues of A are $\lambda_1 = 1 + \delta$ and $\lambda_2 = 1 - \delta$, and its corresponding eigenvectors are $v_1 = (1, 0)$ and $v_2 = (0, 1)$. The eigenvalues of $\tilde{A}$ are also $\tilde{\lambda}_1 = 1 + \delta$ and $\tilde{\lambda}_2 = 1 - \delta$, but its corresponding eigenvectors are $\tilde{v}_1 = (1/\sqrt{2}, 1/\sqrt{2})$ and $\tilde{v}_2 = (-1/\sqrt{2}, 1/\sqrt{2})$; we have $\tilde{v}_2^* v_1 = 1/\sqrt{2}$, which can be arbitrarily large relative to δ.

Theorem 3 (Davis-Kahan $\sin(\Theta)$ theorem)**.** *Let $A = E_0 A_0 E_0{}^* + E_1 A_1 E_1{}^*$ and $A + H = F_0 \Lambda_0 F_0{}^* + F_1 \Lambda_1 F_1{}^*$ be symmetric matrices with $[E_0, E_1]$ and $[F_0, F_1]$ orthogonal. If the eigenvalues of A_0 are contained in an interval (a, b), and the eigenvalues of Λ_1 are excluded from the interval $(a - \delta, b + \delta)$ for some $\delta > 0$, then*

$$\|F_1{}^* E_0\| \leq \frac{\|F_1{}^* H E_0\|}{\delta}$$

for any unitarily invariant norm $\|\cdot\|$.

Proof. Since $AE_0 = E_0 A_0 E_0{}^* E_0 + E_1 A_1 E_1{}^* E_0 = E_0 A_0$, we have

$$\begin{aligned} HE_0 &= AE_0 + HE_0 - E_0 A_0 \\ &= (A + H)E_0 - E_0 A_0. \end{aligned}$$

Furthermore, $F_1{}^*(A + H) = \Lambda_1 F_1{}^*$, so

$$\begin{aligned} F_1{}^* H E_0 &= F_1{}^*(A + H)E_0 - F_1{}^* E_0 A_0 \\ &= \Lambda_1 F_1{}^* E_0 - F_1{}^* E_0 A_0. \end{aligned}$$

Let $c := (a + b)/2$ and $r := (b - a)/2 \geq 0$. By the triangle inequality, we have

$$\begin{aligned} \|F_1{}^* H E_0\| &= \|\Lambda_1 F_1{}^* E_0 - F_1{}^* E_0 A_0\| \\ &= \|(\Lambda_1 - cI) F_1{}^* E_0 - F_1{}^* E_0 (A_0 - cI)\| \\ &\geq \|(\Lambda_1 - cI) F_1{}^* E_0\| - \|F_1{}^* E_0 (A_0 - cI)\|. \end{aligned}$$

Here we have used a centering trick so that $A_0 - cI$ has eigenvalues contained in $[-r, r]$, and $\Lambda_1 - cI$ has eigenvalues excluded from $(-r - \delta, r + \delta)$. This implies that $\|A_1 - cI\|_2 \leq r$ and $\|(\Lambda_1 - cI)^{-1}\|_2 \leq (r + \delta)^{-1}$, respectively. Therefore

$$\begin{aligned} \|(\Lambda_1 - cI) F_1{}^* E_0\| &\geq \frac{1}{\|(\Lambda_1 - cI)^{-1}\|_2} \|F_1{}^* E_0\| \\ &\geq (r + \delta) \|F_1{}^* E_0\| \end{aligned}$$

and

$$\begin{aligned} \|F_1{}^* E_0 (A_0 - cI)\| &\leq \|A_0 - cI\|_2 \|F_1{}^* E_0\| \\ &\leq r \|F_1{}^* E_0\|. \end{aligned}$$

We conclude that $\|F_1{}^* H E_0\| \geq \delta \|F_1{}^* E_0\|$. □