Lecture 3: Spectral Method - basic pertubation theory

Lecturer: Yihong Wu — *Scribe: Matthias Löffler, Sep 12, 2018*

Our goal is to use the spectral method to pursue inference. As we will see, in planted clique and many related planted problems, the first few eigenvectors of the population matrix $\mathbb{E}X$ contain the information about the planted structures that we are interested in. Since we only have observations X at hand and do not know $\mathbb{E}X$, we compute the first few eigenvectors of X instead. Writing $X = \mathbb{E}X + (X - \mathbb{E}X)$, we expect that the error of estimating the first few eigenvectors of $\mathbb{E}X$ can be bounded by the size of the pertubation $X - \mathbb{E}X$.

3.1 Review of linear algebra

3.1.1 Eigendecomposition

Suppose that X is a symmetric real valued matrix in $\mathbb{R}^{n\times n}$.

Definition 3.1. The pair (λ, v) with $\lambda \in \mathbb{R}$ and $v \in \mathbb{R}^n$ is an eigenpair of X, consisting of an eigenvalue λ and an eigenvector v, if

$$Xv = \lambda v.$$

We order the eigenvalues of X by their sizes such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. The corresponding eigenvectors $[v_1, \ldots, v_n]$ form an orthonormal basis (ONB) of $\mathbb{R}^n$. Denote $V = [v_1, \ldots, v_n]$, $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. We can write the eigendecomposition of X as

$$X = V\Lambda V^\top = \sum_{i=1}^n \lambda_i v_i v_i^\top.$$

Also note that $\mathrm{rank}(X) = r \Leftrightarrow$ there exist exactly r nonzero λ_i's.

3.1.2 Singular value decomposition (SVD)

Now suppose that $X \in \mathbb{R}^{m\times n}$ is a real valued rectangular matrix. The singular value decomposition (SVD) of X is

$$X = U\Sigma V^\top = \sum \sigma_i U_i V_i^\top$$

where $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_r) \in \mathbb{R}^{r\times r}$, $\sigma_i \geq 0$, $U = [U_1, \ldots, U_r] \in \mathbb{R}^{m\times r}$ and $V = [V_1, \ldots, V_r] \in \mathbb{R}^{n\times r}$. The columns of U are orthonormal and we call them left singular vectors and likewise the columns of V are orthonormal too and we call them right singular vectors.
We can calculate Σ, U and V by taking eigendecompositions of $XX^\top$ and $X^\top X$. Indeed,

$$XX^\top = U\Sigma^2 U^\top \in \mathbb{R}^{m\times m} \qquad \text{and} \qquad X^\top X = V\Sigma^2 V^\top \in \mathbb{R}^{n\times n},$$

and

$$\sigma_i = \sqrt{\lambda_i(XX^\top)} = \sqrt{\lambda_i(X^\top X)}.$$

3.1.3 Matrix norms

Suppose again that $X \in \mathbb{R}^{m\times n}$. There are multiple ways to define a norm on X.

- We view X as a mn-dimensional vector with euclidean norm and define the Frobenius norm

$$\|X\|_F = \|\text{vec}(X)\|_2 = \sqrt{\sum_{i,j} X_{ij}^2}.$$

- We view X as a linear operator from $(\mathbb{R}^n, \|\cdot\|_p) \to (\mathbb{R}^m, \|\cdot\|_q)$ with operator norm

$$\|X\|_{p\to q} = \sup_{\|v\|_p=1} \|Av\|_q.$$

  For this course the most relevant matrix is the case of $p = q = 2$, where we equip $\mathbb{R}^n$ with the euclidean inner product. We denote

$$\|X\|_{2\to 2} =: \|X\|_{op},$$

  also known as the spectral norm.

We now prove that

$$\|X\|_{op} = \sigma_{\max}(X)$$

Using the SVD of X:

$$\|X\|_{op}^2 = \sup_{\|v\|_2=1} \|Xv\|_2^2 = \sup_{\|v\|_2=1} \left\|\sum \sigma_i U_i V_i^\top v\right\|_2^2 = \sup_{\|v\|_2=1} \sum \sigma_i^2 \langle V_i, v\rangle^2 = \sigma_{\max}(X)^2.$$

Remark 3.1.
- $\|\cdot\|_{op}$ is a norm and $\|X\|_{op} = \|X^\top\|_{op}$.
- $\|XY\| \le \|X\|_{op}\|Y\|_{op}$.
- If $X = x$ is a vector then $\|X\|_{op} = \|x\|_2$.
- $\|\cdot\|_{op}$ is orthogonal invariant, i.e. for any $R \in \mathbb{O}(n)$, $R' \in \mathbb{O}(m)$ we have $\|R'XR\|_{op} = \|X\|_{op}$.
- If $X = [X_1, \ldots, X_n]$ has orthonormal rows (columns), then $\|X\|_{op} = 1$.

Remark 3.2. Recall the matrix inner product: $\langle X, Y\rangle = \text{trace}(Y^\top X) = \sum_{i,j} X_{ij}Y_{ij}$. Using this we can write

$$\|X\|_{op} = \sigma_{\max}(X) = \sup_{\|A\|_F=1,\ \text{rank}(A)=1} \langle X, A\rangle = \sup_{\|u\|_2=\|v\|_2=1} \langle X, uv^\top\rangle.$$

Likewise, if X is real and symmetric we have that

$$\lambda_{\max}(X) = \sup_{\|v\|_2=1} \langle X, vv^\top\rangle, \quad \|X\|_{op} = \sigma_{\max}(X) = \sup_{\|v\|_2=1} |\langle X, vv^\top\rangle|.$$

Similar relations hold for $\sigma_{\min}$ and $\lambda_{\min}$ if one substitutes the sup's above for inf's.

3.2 Pertubation of eigenstructures

In this section we assume that we are given two matrices, X and $Y = X + Z$ where Z is a 'pertubation' of X. We are interested if eigenvectors and eigenvalues of X and Y are close when Z is 'small'. Unfortunately, in general this is not the case.

3.2.1 Negative results

Eigenvalues The eigenvalues λ_i are the roots of the polynomial $\det(\lambda I - X) = 0$, which is a polynomial in λ of degree n. Although the roots are continuous in the coefficients of the polynomial, in general the modulus of continuity is not Lipschitz and only $\frac{1}{\text{degree}}$-Hölder, and this is tight. Indeed, consider the two matrices

$$X = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad X_\varepsilon = \begin{bmatrix} 0 & 1 \\ \varepsilon & 0 \end{bmatrix}$$

Then $\lambda_1(X) = \lambda_2(X) = 0$, but $\lambda_1(X_\varepsilon) = \sqrt{\varepsilon}$ and $\lambda_2 = -\sqrt{\varepsilon}$. More generally consider

$$X = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & & \vdots \\ 0 & \dots & \dots & \dots & 1 \\ 0 & \dots & \dots & \dots & 0 \end{bmatrix} \quad \text{and} \quad X_\varepsilon = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & & \vdots \\ 0 & \dots & \dots & \dots & 1 \\ \varepsilon & \dots & \dots & \dots & 0. \end{bmatrix}$$

One can show that $\lambda_i(X) = 0$ but that $\lambda_i(X_\varepsilon) = \varepsilon^{1/n}$.

Therefore we need more assumptions on X to be able to obtain Lipschitz bounds, e.g. that X is a real and symmetric matrix.

Eigenvectors But even in the symmetric case eigenvector pertubations may fail dramatically. For $\varepsilon > 0$ consider

$$X = \begin{bmatrix} 1+\varepsilon & 0 \\ 0 & 1+\varepsilon \end{bmatrix} \quad \text{and} \quad Y = \begin{bmatrix} 1 & \varepsilon \\ \varepsilon & 1 \end{bmatrix}.$$

The eigenvalues of these two matrices are the same

$$\lambda_1(X) = \lambda_1(Y) = 1 + \varepsilon, \ \lambda_2(X) = \lambda_2(Y) = 1 - \varepsilon.$$

However, the eigenvectors are far apart:

$$v_1(X) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, v_2(X) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \textbf{but} \quad v_1(Y) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, v_2(Y) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

The lesson from this is that we need separation between the eigenvalues, a spectral (eigen) gap.

3.2.2 Pertubation bound for eigenvalues

Let X, Y, Z be real symmetric matrices in $\mathbb{R}^{n\times n}$ and suppose $Y = X + Z$. We have that

$$\begin{aligned}\lambda_1(X) + \lambda_n(Z) = \lambda_1(X) + \inf_{\|v\|_2=1} \langle Z, vv^\top \rangle \\ \leq \sup_{\|v\|_2=1} \langle X+Z, vv^\top \rangle = \lambda_1(Y) \\ \leq \lambda_1(X) + \sup_{\|v\|_2=1} \langle Z, vv^\top \rangle = \lambda_1(X) + \lambda_1(Z)\end{aligned}$$

and therefore

$$|\lambda_1(X) - \lambda_1(Y)| \leq \max(|\lambda_1(Z)|, |\lambda_n(Z)|) = \|Z\|_{op}.$$

More generally we have the following theorem (homework):

Theorem 3.1 (Weyl's inequality / Lidski's inequality)**.**

$$|\lambda_i(X) - \lambda_i(Y)| \leq \|Z\|_{op}.$$

3.2.3 Pertubation bounds for eigenspaces

Let X, Y, Z again be real symmetric matrices in $\mathbb{R}^{n\times n}$ and suppose $Y = X + Z$. Suppose that $X = \sum_i \lambda_i u_i u_i^\top$ and $Y = \sum \rho_i v_i v_i^\top$. We want to prove a pertubation bound for $u \triangleq u_1$ and $v \triangleq v_1$ and more generally for $U = [u_1, \ldots, u_r]$ and $V = [v_1, \ldots, v_r]$. However, considering $\|u - v\|_2$ makes no sense as u and v are only determined up to their sign, and similarly U and V are only defined up to orthogonal transformation. There are two possible workarounds:

- Consider the distance

  $$\min_{s\in\{\pm 1\}} \|u + sv\|_2 = \sqrt{2 - 2|\langle u, v\rangle|} = \sqrt{2 - 2\cos\theta} = \sqrt{2}\sin\frac{\theta}{2},$$

  and, more generally, $\inf_{R\in O(r)} \|U - VR\|$.

- Consider the distance between the linear subspaces spanned by u and v, defined through their respective projection matrices:

  $$\left\| uu^\top - vv^\top \right\|_F^2 = 2(1 - \langle u, v\rangle^2) = 2\sin^2(\theta),$$

  and in the general case $\|UU^\top - VV^\top\|_F$ or $\|UU^\top - VV^\top\|_{op}$.

Theorem 3.2 (Davis-Kahan)**.** *Let* $\cos\theta = |\langle u_1, v_1\rangle|$. *Then*

$$\sin\theta \leq \frac{\|Z\|_{op}}{\max(\rho_1 - \lambda_2, \lambda_1 - \rho_2)}.$$

Proof. Assume that $\rho_1 \geq \lambda_2$. Let us start from the eigenvalue equations:

$$Xu = \lambda_1 u \quad \text{and} \quad Yv = \rho_1 v.$$

Denote $U_\perp = [u_2, \dots, u_n] \in \mathbb{R}^{n \times n-1}$. Then

$$U_\perp^\top X = \begin{bmatrix} u_2^\top \\ \vdots \\ u_n^\top \end{bmatrix} X = \begin{bmatrix} \lambda_2 u_2^\top \\ \vdots \\ \lambda_n u_n^\top \end{bmatrix} = \begin{bmatrix} \lambda_2 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \begin{bmatrix} u_2^\top \\ \vdots \\ u_n^\top. \end{bmatrix}$$

Hence

$$\begin{aligned} U_\perp^\top (X+Z)v = \rho_1 U_\perp^\top v &\Leftrightarrow \begin{bmatrix} \lambda_2 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} U_\perp^\top v + U_\perp^\top Z V = \rho_1 U_\perp^\top v \\ &\Leftrightarrow \begin{bmatrix} \rho_1 - \lambda_2 & & \\ & \ddots & \\ & & \rho_1 - \lambda_n \end{bmatrix} U_\perp^\top v = U_\perp^\top Z V \\ &\Leftrightarrow U_\perp^\top v = \begin{bmatrix} \frac{1}{\rho_1 - \lambda_2} & & \\ & \ddots & \\ & & \frac{1}{\rho_1 - \lambda_n} \end{bmatrix} U_\perp^\top Z V \end{aligned}$$

Taking the $\|\cdot\|_2$-norm on both sides gives

$$\|U_\perp^\top v\|_2 \le \left\| \begin{bmatrix} \frac{1}{\rho_1 - \lambda_2} & & \\ & \ddots & \\ & & \frac{1}{\rho_1 - \lambda_n} \end{bmatrix} \right\|_{op} \left\| U_\perp^\top \right\|_{op} \|Z\|_{op} = \frac{\|Z\|_{op}}{|\rho_1 - \lambda_2|}.$$

Finally, note that

$$\|U_\perp^\top v\|_2^2 = v^\top U_\perp U_\perp^\top v = v^\top (I - uu^\top) v = 1 - \langle u, v\rangle^2 = \sin^2(\theta).$$

If $\rho_1 < \lambda_2$, then $\lambda_1 > \rho_2$. Exchanging the roles of X and Y we obtain the other statement. □

More generally, considering the first r eigenvectors we have for $U = [U_1, \dots, U_r]$ and $V = [V_1, \dots V_r]$ that for any unitarily invariant norm $\|\cdot\|$,

$$\|U_\perp^\top V\| \le \frac{\|Z\|}{\max(|\rho_r - \lambda_{r+1}|, |\lambda_r - \rho_{r+1}|)}.$$

One can generalize this to singular vectors by a technique sometimes called self-adjoint dilation:[1] For $X = U\Sigma V^\top \in \mathbb{R}^{m\times n}, Y = \tilde{U}\tilde{\Sigma}\tilde{V}^\top$ consider the matrix

$$\begin{bmatrix} 0 & X \\ X^\top & 0 \end{bmatrix} \in \mathbb{R}^{(m+n)\times(m+n)},$$

and likewise for Y. Observe that

$$\begin{bmatrix} 0 & X \\ X^\top & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ v_1 \end{bmatrix} = \sigma_1 \begin{bmatrix} u_1 \\ v_1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & X \\ X^\top & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ -v_1 \end{bmatrix} = -\sigma_1 \begin{bmatrix} u_1 \\ -v_1 \end{bmatrix}.$$

Now we can apply the Davis-Kahan Theorem (and $\sin\frac{\theta}{2} \le \sin\theta$) to obtain

$$\min_{s\in\{\pm 1\}} \left\| \begin{bmatrix} u_1 \\ v_1 \end{bmatrix} + s \begin{bmatrix} \tilde{u}_1 \\ \tilde{v}_1 \end{bmatrix} \right\|_2 \le \frac{2 \left\| \begin{bmatrix} 0 & Z \\ Z^\top & 0 \end{bmatrix} \right\|_{op}}{|\sigma_1(X) - \sigma_2(Y)|} = \frac{2\|Z\|_{op}}{|\sigma_1(X) - \sigma_2(Y)|}.$$

[1] Thanks for Cheng Mao for pointing this out.

References