EE290 Mathematics of Data Science Lecture 2 - 09/03/2019

Lecture 2: Matrix Decompositions, Norms, and Perturbation

Lecturer: Jiantao Jiao *Scribe: Ziye Ma*

This lecture is based on [1, Chapter 3].

1 Eigendecomposition

Suppose X is a real symmetric matrix $\in \mathbb{R}^{n\times n}$ (i.e., $X_{i,j} = X_{j,i}, \forall(i,j)$).

Definition 1. ***Eigenpair***: *The pair* $(\lambda, \mathbf{v})$*, where* $\lambda \in \mathbb{R}$, $\mathbf{v} \in \mathbb{R}^n, \mathbf{v} \neq 0$ *is an eigenpair of* X *if* $X\mathbf{v} = \lambda\mathbf{v}$.

Theorem 2. *If* $X \in \mathbb{R}^{n\times n}$ *is a real symmetric matrix. Then it has* n *real eigenvalues, sorted as* $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$. *The corresponding eigenvectors* $V = [\mathbf{v}_1\, \mathbf{v}_2 ... \mathbf{v}_n]$ *form a orthonormal basis of* $\mathbb{R}^n$. *Denote* $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$*, we have*

$$X = V\Lambda V^T = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^T \tag{1}$$

2 Singular Value Decomposition (SVD)

Theorem 3 (Singular Value Decomposition)**.** *Suppose* $X \in \mathbb{R}^{m\times n}$. *Then, the Singular Value Decomposition (SVD) of* X *is*

$$X = U\Sigma V^T = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T \tag{2}$$

where $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r) \in \mathbb{R}^{r\times r}, \sigma_i \geq 0, U = \begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2 & ... & \mathbf{u}_r\end{bmatrix} \in \mathbb{R}^{m\times r}, V = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & ... & \mathbf{v}_r\end{bmatrix} \in \mathbb{R}^{n\times r}$. *Here* $\mathbf{u}_i$ *are orthogonal to each other, and* $\|\mathbf{u}_i\| = 1\ \forall i$. *Similarly,* $\mathbf{v}_i$ *are orthogonal to each other, and* $\|\mathbf{v}_i\| = 1\ \forall i$.

Now, if we right-multiply X by a generic vector $\mathbf{v} \in \mathbb{R}^n$,

$$X\mathbf{v} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i (\mathbf{v}_i^T \mathbf{v}) \tag{3}$$

We can calculate Σ, U, V using the eigendecompositions of XX^T and X^TX. Indeed,

$$XX^T = U\Sigma V^T V\Sigma U^T = U\Sigma^2 U^T \tag{4}$$

$$X^TX = V\Sigma^2 V^T \tag{5}$$

$$\tag{6}$$

and

$$\sigma_i = \sqrt{\lambda_i(XX^T)} = \sqrt{\lambda_i(X^TX)} \tag{7}$$

3 Matrix Norm

Definition 4 (Vector p-norm)**.** ***p-norm:***
If $\mathbf{x} \in \mathbb{R}^d$*, then*

$$\|\mathbf{x}\|_p = \begin{cases} (\sum_{i=1}^d |x_i|^p)^{\frac{1}{p}} & p \geq 1 \\ \max_i |x_i| & p = \infty. \end{cases} \tag{8}$$

3.1 Frobenius Norm

Given $X \in \mathbb{R}^{m \times n}$,

$$\|X\|_F = \|vec(X)\|_2 = \sqrt{\sum_{i,j} X_{ij}^2} = \sqrt{\text{Tr}(XX^T)} \tag{9}$$

3.2 $p-q$ Norm

Note here $p \geq 1, q \geq 1$. Given $X \in \mathbb{R}^{m \times n}$,

$$\|X\|_{p \to q} = \sup_{\|\mathbf{v}\|_p = 1} \|X\mathbf{v}\|_q \tag{10}$$

The (2-2) norm is called operator norm $\|X\|_{op}$.

$$\|X\|_{op}^2 = \sup_{\|\mathbf{v}\|_2=1} \|X\mathbf{v}\|_2^2 = \sup_{\|\mathbf{v}\|_2=1} \left\| \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T \mathbf{v} \right\|_2^2 \tag{11}$$

$$= \sup_{\|\mathbf{v}\|_2=1} \sum_{i=1}^r \sigma_i^2 (\mathbf{v}_i^T \mathbf{v})^2 \tag{12}$$

$$= \sigma_1^2 \tag{13}$$

where σ_1 is the largest singular value of X.

3.3 Properties of Operator Norm

1. $\|X\|_{op} = \|X^T\|_{op}$
2. $\|XY\|_{op} \leq \|X\|_{op} \|Y\|_{op}$
3. R orthogonal then $\|XR\|_{op} = \|RX\|_{op} = \|X\|_{op}$ (orthogonally invariant).
4. If X is orthogonal, then $\|X\|_{op} = 1$
5. $X = \begin{bmatrix} \mathbf{x} & \mathbf{x}, \ldots, & \mathbf{x} \end{bmatrix} \in \mathbb{R}^{d \times m} \Rightarrow \|X\|_{op} = \sqrt{m} \|\mathbf{x}\|_2$

3.4 Further insights of Operator Norm

$$\|X\|_{op} = \sigma_1(X) = \sup_{\|\mathbf{v}\|_2 = \|\mathbf{u}\|_2 = 1} \langle X, \mathbf{u}\mathbf{v}^T \rangle \tag{14}$$

$$= \sup_{\|A\|_F = 1, r(A) = 1} \text{Tr}(X^T A) \tag{15}$$

where

$$\text{Tr}(X^T Y) = \sum_{i,j} X_{ij} Y_{ij} \tag{16}$$

4 Matrix Perturbation

First we observe that the eigenvalues of a matrix in general is not *Lipschitz* with respect to the perturbations measured in operator norm.

Example 5. Say originally,

$$X = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \tag{17}$$

then this matrix is perturbed minimally by a very small number ϵ, denoted as X_ϵ:

$$X_\epsilon = \begin{bmatrix} 0 & 1 \\ \epsilon & 0 \end{bmatrix} \tag{18}$$

However, in this case, X has eigenvalues $\lambda_1 = \lambda_2 = 0$, while X_ϵ has eigenvalues $\lambda_1 = \sqrt{\epsilon}, \lambda_2 = -\sqrt{\epsilon}$, which is not Lipschitz in ϵ.

However, for symmetric matrices we indeed can show the eigenvalues are Lipschitz with respect to the perturbations measured in operator norm. Consider symmetric matrices X, Y, Z, where

$$Y = X + Z \tag{19}$$

Then

$$\begin{aligned}
\lambda_1(X) + \lambda_n(Z) &= \lambda_1(X) + \inf_{\|\mathbf{v}\|_2=1} \mathrm{Tr}(Z\mathbf{v}\mathbf{v}^T) && (20)\\
&\le \sup_{\|\mathbf{v}\|_2=1} \mathrm{Tr}((X+Z)\mathbf{v}\mathbf{v}^T) && (21)\\
&= \lambda_1(Y) && (22)\\
&\le \lambda_1(X) + \lambda_1(Z) && (23)
\end{aligned}$$

$\Rightarrow$

$$\begin{aligned}
|\lambda_1(X) - \lambda_1(Y)| &\le \max(|\lambda_1(Z)|, |\lambda_n(Z)|) && (24)\\
&= \|Z\|_{op} && (25)
\end{aligned}$$

Theorem 6 (Weyl (Lidskii)'s Inequality)**.** *If X, Y are both symmetric matrices, then*

$$|\lambda_i(X) - \lambda_i(Y)| \le \|Z\|_{op}. \tag{26}$$

Now we want to know how much eigenvectors will be affected according to the Z perturbation. To compare the similarity of two eigenvectors, we will be using the angle θ between them. The smaller the angle, the closer the two eigenvectors since eigenvectors are all about directions, not magnitudes. Again, assuming that X, Y, Z are symmetric matrices,

$$X = \sum_{i=1}^{n} \lambda_i \mathbf{u}_i \mathbf{u}_i^T \tag{27}$$

$$Y = \sum_{i=1}^{n} \rho_i \mathbf{v}_i \mathbf{v}_i^T \tag{28}$$

We introduce some measures of the distances between eigenvectors

1. $\min_{s\in\{1,-1\}} \|\mathbf{u} + s\mathbf{v}\|_2 = \sqrt{2 - 2|\mathbf{u}^T\mathbf{v}|} = \sqrt{2 - 2\cos(\theta)} = 2\sin(\frac{\theta}{2})$
2. $\left\|\mathbf{u}\mathbf{u}^T - \mathbf{v}\mathbf{v}^T\right\|_F = 2(1 - |\mathbf{u}^T\mathbf{v}|^2) = 2\sin^2(\theta)$

Theorem 7 (Davis-Kahan)**.**

$$\sin(\theta) \leq \frac{\|Z\|_{op}}{\max(\rho_1 - \lambda_2, \lambda_1 - \rho_2)} \tag{29}$$

Proof We know

$$X\mathbf{u}_1 = \lambda_1 \mathbf{u}_1 \tag{30}$$

$$Y\mathbf{v}_1 = \rho_1 \mathbf{v}_1 \tag{31}$$

Then we define

$$U_\perp = \begin{bmatrix} \mathbf{u}_2 & \mathbf{u}_3 & ... & \mathbf{u}_n \end{bmatrix} \in \mathbb{R}^{n\times(n-1)}, \text{orthogonal complement of } \mathbf{u}_1 \tag{32}$$

then

$$U_\perp^T X = \begin{bmatrix} \mathbf{u}_2^T \\ \mathbf{u}_3^T \\ ... \\ \mathbf{u}_n^T \end{bmatrix} X = \begin{bmatrix} \lambda_2 & .. \\ .. & .. \\ .. & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{u}_2^T \\ ... \\ \mathbf{u}_n^T \end{bmatrix} \tag{33}$$

Since also

$$Y = X + Z \tag{34}$$

$\Rightarrow$

$$U_\perp^T (X + Z)\mathbf{v}_1 = \rho_1 U_\perp^T \mathbf{v}_1 \tag{35}$$

$$\begin{bmatrix} \lambda_2 & .. \\ .. & .. \\ .. & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{u}_2^T \\ ... \\ \mathbf{u}_n^T \end{bmatrix} \mathbf{v}_1 + U_\perp^T Z \mathbf{v}_1 = \rho_1 U_\perp^T \mathbf{v}_1 \tag{36}$$

$$\begin{bmatrix} \rho_1 - \lambda_2 & .. \\ .. & .. \\ .. & \rho_1 - \lambda_n \end{bmatrix} U_\perp^T \mathbf{v}_1 = U_\perp^T Z \mathbf{v}_1 \tag{37}$$

$$U_\perp^T \mathbf{v}_1 = \begin{bmatrix} \frac{1}{\rho_1 - \lambda_2} & .. \\ .. & .. \\ .. & \frac{1}{\rho_1 - \lambda_n} \end{bmatrix} U_\perp^T Z \mathbf{v}_1 \tag{38}$$

Therefore

$$\left\| U_\perp^T \mathbf{v}_1 \right\|_{op} \leq \left\| \begin{bmatrix} \frac{1}{\rho_1 - \lambda_2} & .. \\ .. & .. \\ .. & \frac{1}{\rho_1 - \lambda_n} \end{bmatrix} \right\|_{op} \left\| U_\perp^T \right\|_{op} \|Z\|_{op} \tag{39}$$

and

$$\left\| U_\perp^T \mathbf{v}_1 \right\|_2^2 = \mathbf{v}_1^T U_\perp U_\perp^T \mathbf{v}_1 = \mathbf{v}_1^T (1 - \mathbf{u}_1 \mathbf{u}_1^T) \mathbf{v}_1 = 1 - |\mathbf{v}_1^T \mathbf{u}_1|^2 = \sin^2(\theta) \tag{40}$$

□

References

[1] Yihong Wu and Jiaming Xu. Statistical inference on graphs: Selected topics, October 2019. http://www.stat.yale.edu/~yw562/teaching/stats-graphs.pdf.