EECS 16B

Spring 2022

Lecture 21

4/5/2022

LECTURE 21: more SVD

Finding a SVD for $A \in \mathbb{R}^{m \times n}$ (with rank $= r$) from evalues/evectors of:

$A^T A \in \mathbb{R}^{n \times n}$

Claims 1–3 (last lecture): Evalues of $A^T A$ are real and nonnegative. r of them are strictly positive; the remaining $n-r$ are zero.

Step 1: Find orthogonal matrix V diagonalizing $A^T A$:

$$V^T A^T A V = \begin{bmatrix} \lambda_1 & & & & \\ & \ddots & & & \\ & & \lambda_r & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix} \Big\} n-r$$

$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$

Step 2: For each $i = 1, \ldots r$ pick ith column $\vec{v}_i$ of V (which is evector for $A^T A$ for evalue λ_i). Let

$$\sigma_i = \sqrt{\lambda_i}, \quad \vec{u}_i = \frac{1}{\sigma_i} A \vec{v}_i.$$

$AA^T \in \mathbb{R}^{m \times m}$

$\rightarrow$ same claims can be adapted: real, nonnegative evalues, r of which are strictly positive, remaining $m-r$ are zero.

Step 1: Find orthogonal matrix $U \in \mathbb{R}^{m \times m}$ diagonalizing AA^T:

$$U^T A A^T U = \begin{bmatrix} \lambda_1 & & & & \\ & \ddots & & & \\ & & \lambda_r & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix} \Big\} m-r$$

$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$

Step 2: For each $i = 1 \ldots r$ pick ith column $\vec{u}_i$ of U (which is evector of AA^T for evalue λ_i). Let

$$\sigma_i = \sqrt{\lambda_i}, \quad \vec{v}_i = \frac{1}{\sigma_i} A^T \vec{u}_i.$$

Which procedure to use? Choose $A^T A$ or AA^T based on which one looks simpler for finding evalues/evectors. If $m < n$, AA^T $(m \times m)$ is smaller than $A^T A$ $(n \times n)$ and may be preferable.

<u>Example 2 (from Lecture 20):</u> $A = \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}$ $\quad A^T = \begin{bmatrix} 4 & -3 \\ 4 & 3 \end{bmatrix}$

Lecture 20: $A^T A = \begin{bmatrix} 25 & 7 \\ 7 & 25 \end{bmatrix}$. Today: $AA^T = \begin{bmatrix} 32 & 0 \\ 0 & 18 \end{bmatrix}$

Easier b/c diagonal: choose $U = I$: $\vec{u}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ $\vec{u}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

$\lambda_1 = 32$, $\lambda_2 = 18 \Rightarrow \sigma_2 = \sqrt{\lambda_2} = 3\sqrt{2}$

$\Downarrow$

$\sigma_1 = \sqrt{\lambda_1} = 4\sqrt{2}$

$$\vec{v}_2 = \frac{1}{\sigma_2} A^T \vec{u}_2 = \frac{1}{3\sqrt{2}} \begin{bmatrix} -3 \\ 3 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$\vec{v}_1 = \frac{1}{\sigma_1} A^T \vec{u}_1 = \frac{1}{4\sqrt{2}} \begin{bmatrix} 4 \\ 4 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Last time: same σ_1, σ_2. $\vec{u}_1, \vec{v}_1$ same as today.

$\vec{v}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$, $\vec{u}_2 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$:

Signs of each opposite to those above but same outer product $\vec{u}_2 \vec{v}_2^T$.

Example 3 (from Lecture 20):

$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ $\quad AA^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ $\quad$ $U = I$ works for step 1

$\vec{u}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ $\vec{u}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

$\Downarrow$

$\lambda_1 = \lambda_2 = 1$

$\Downarrow$

$\sigma_1 = \sigma_2 = 1$ $\quad$ $\vec{v}_1 = A^T \vec{u}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ $\quad \vec{v}_2 = A^T \vec{u}_2 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$

Since $AA^T = I$, any other orthonormal $\vec{u}_1, \vec{u}_2$ will work.

$\vec{u}_1 = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$ $\quad \vec{u}_2 = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$

$\vec{u}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \vec{u}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ above are a special case: $\theta = 0$

$\vec{v}_1 = A^T \vec{u}_1 = \begin{bmatrix} \cos\theta \\ -\sin\theta \end{bmatrix}$ $\quad \vec{v}_2 = A^T \vec{u}_2 = \begin{bmatrix} -\sin\theta \\ -\cos\theta \end{bmatrix}$. $\quad$ Conclusion:

Repeated evalues of A^TA or AA^T ($\lambda_1=\lambda_2=1$ in this example) are another source of nonuniqueness in SVD.

<u>Full SVD:</u> So far two forms studied:

1) $A=\sum_{i=1}^{r}\sigma_i\vec{u}_i\vec{v}_i^T$ (outer product form)

2) $$A=\underbrace{\begin{bmatrix}\vec{u}_1 & \cdots & \vec{u}_r\end{bmatrix}}_{\substack{U_r \\ m\times r}}\underbrace{\begin{bmatrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_r\end{bmatrix}}_{\substack{\Sigma_r \\ r\times r}}\underbrace{\begin{bmatrix}\vec{v}_1^T \\ \vdots \\ \vec{v}_r^T\end{bmatrix}}_{\substack{V_r^T \\ r\times n}}$$ (compact form)

Full SVD is another equivalent form:

3) Complete $\vec{u}_1 \cdots \vec{u}_r$ to an orthonormal basis for $\mathbb{R}^m$ with $\vec{u}_{r+1} \cdots \vec{u}_m$; complete $\vec{v}_1 \cdots \vec{v}_r$ to orthonormal basis for $\mathbb{R}^n$ with $\vec{v}_{r+1} \cdots \vec{v}_n$. Then:

$$\underbrace{\begin{bmatrix}\overbrace{\vec{u}_1 \cdots \vec{u}_r}^{U_r} & \overbrace{\vec{u}_{r+1} \cdots \vec{u}_m}^{U_{m-r}}\end{bmatrix}}_{m\times m}\underbrace{\left[\begin{array}{c|c}\overbrace{\begin{matrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_r\end{matrix}}^{\Sigma_r} & 0_{r\times(n-r)} \\ \hline 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)}\end{array}\right]}_{m\times n}\underbrace{\begin{bmatrix}\left.\begin{matrix}\vec{v}_1^T \\ \vdots \\ \vec{v}_r^T\end{matrix}\right\} V_r^T \\ \left.\begin{matrix}\vec{v}_{r+1}^T \\ \vdots \\ \vec{v}_n^T\end{matrix}\right\} V_{n-r}^T\end{bmatrix}}_{n\times n}$$

$=: U$ (orthogonal)

$=: \Sigma$

$=: V^T$ (orthogonal)

<u>Note:</u>

1) If A is wide and full row rank ($n>m=r$), then:

$$\Sigma = [\Sigma_r \quad 0_{r\times(n-r)}].$$

If A is tall and full column rank $(m > n = r)$, then

$$\Sigma = \begin{bmatrix} \Sigma_r \\ 0_{(m-r)\times r} \end{bmatrix}.$$

If A is square and full rank $(m = n = r)$:

$$\Sigma = \Sigma_r.$$

2) $\vec{v}_{r+1} \cdots \vec{v}_n$ are evectors of A^TA corresponding to 0 evalues or, equivalently, an orthonormal basis for

$$\text{Null}(A^TA) = \text{Null}(A)$$

↘ shown in proof of Claim 3 in Lec. 20

Since $\text{span}\{\vec{v}_{r+1}, \ldots \vec{v}_n\} = \text{col}(V_{n-r})$,

$$\boxed{\text{col}(V_{n-r}) = \text{Null}(A).}$$

$\vec{u}_{r+1} \cdots \vec{u}_m$ are evectors for 0 evalues of AA^T therefore an orthonormal basis for

$$\text{Null}(AA^T) = \text{Null}(A^T).$$

Thus,

$$\boxed{\text{Col}(U_{m-r}) = \text{Null}(A^T).}$$

$$A\vec{x} = \sum_{i=1}^{r} \sigma_i \vec{u}_i \underbrace{\vec{v}_i^T \vec{x}}_{\text{scalar}} = \sum_{i=1}^{r} \underbrace{\sigma_i (\vec{v}_i^T \vec{x})}_{\text{scalar}} \vec{u}_i$$

Therefore $A\vec{x}$ in the span of $\vec{u}_1 \cdots \vec{u}_r$ for any $\vec{x}$:

$$\boxed{\text{Col}(U_r) = \text{Col}(A).}$$

Similarly,

$$\boxed{\text{Col}(V_r) = \text{Col}(A^T).}$$

Example 1 (Lecture 20): $A = \begin{bmatrix} 4 & 4 \\ 3 & 3 \end{bmatrix}$

$$= \underbrace{5\sqrt{2}}_{\sigma_1} \underbrace{\left(\frac{1}{5}\begin{bmatrix} 4 \\ 3 \end{bmatrix}\right)}_{\vec{u}_1} \underbrace{\left(\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \end{bmatrix}\right)}_{\vec{v}_1^T}$$

Put in full SVD form:

$$\vec{u}_2 = \frac{1}{5}\begin{bmatrix} -3 \\ 4 \end{bmatrix} \qquad \vec{v}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$U = [\vec{u}_1 \ \vec{u}_2] = \frac{1}{5}\begin{bmatrix} 4 & -3 \\ 3 & 4 \end{bmatrix} \quad V = [\vec{v}_1 \ \vec{v}_2] = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$A = U\begin{bmatrix} 5\sqrt{2} & 0 \\ 0 & 0 \end{bmatrix}V^T$$

$\vec{u}_1$ spans $Col(A)$, $\vec{v}_1$ spans $Col(A^T)$

$\vec{u}_2$ spans $Null(A^T)$, $\vec{v}_2$ spans $Null(A)$.

Geometric Interpretation of SVD

$A = U\Sigma V^T$, U and V orthogonal matrices

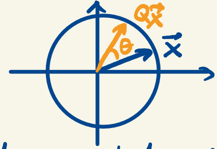
Note:

1) Multiplying a vector $\vec{x}$ by an orthogonal matrix Q does not change its length:

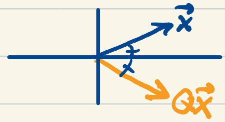
$$\|Q\vec{x}\| = \|\vec{x}\|$$

because $\|Q\vec{x}\|^2 = (Q\vec{x})^T(Q\vec{x}) = \vec{x}^T\underbrace{Q^TQ}_{=I}x = \vec{x}^T\vec{x} = \|\vec{x}\|^2.$

Examples: $Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ rotates by θ:



$Q = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ reflects relative to horizontal axis:



In either case $Q\vec{x}$ is a re-orientation of $\vec{x}$. No change in magnitude: $\|Q\vec{x}\| = \|\vec{x}\|$.

2) Multiplying a vector by $\Sigma_r = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}$ stretches the first entry by σ_1, second entry by σ_2, and so on.

Combining the observations above we can interpret multiplication of a vector $\vec{x}$ by $A = U\Sigma V^T$ as the composition of three operations:

i) $V^T\vec{x}$, which reorients $\vec{x}$ without changing its length;

ii) $\Sigma(V^T\vec{x})$, which stretches the vector $V^T\vec{x}$ along each axis with corresponding singular value;

iii) $U(\Sigma V^T\vec{x})$, which again reorients the resulting vector.

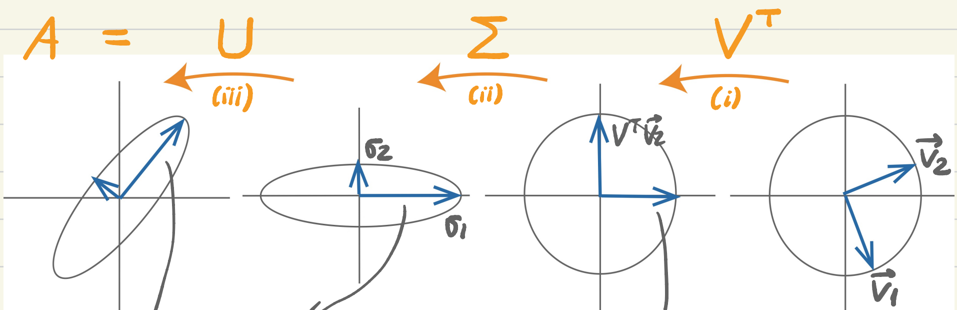


Illustration of multiplication $A\vec{x} = U\Sigma V^T\vec{x}$ when $\vec{x}$ is $\vec{v}_1$, first col. of V.
$\left\{ \begin{array}{l} U\Sigma V^T\vec{v}_1 \\ = A\vec{v}_1 \end{array} \right.$

$\Sigma V^T\vec{v}_1 = \begin{bmatrix} \sigma_1 & \\ & \sigma_2 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \sigma_1 \\ 0 \end{bmatrix}$

$V^T\vec{v}_1 = \begin{bmatrix} \vec{v}_1^T \\ \vec{v}_2^T \end{bmatrix}\vec{v}_1 = \begin{bmatrix} \vec{v}_1^T\vec{v}_1 \\ \vec{v}_2^T\vec{v}_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

<u>Note</u>: $\boxed{\|A\vec{x}\| \leq \sigma_1\|\vec{x}\|}$ because U and V^T do not change length and multiplication by Σ can amplify length by at most σ_1. This is not a conservative bound: it holds with equality when $\vec{x} = \vec{v}_1$ as shown above, or when $\vec{x} = \alpha\vec{v}_1$ for any $\alpha \in \mathbb{R}$.