§6.3

Singular Value Decomposition –

- Given any real m×n matrix A ∃ SVD of A:

$$\underset{m\times n}{A} = \underset{m\times m}{U}\,\underset{m\times n}{\Sigma}\,\underset{n\times n}{V^T} \qquad (*)$$

<u>Here</u>: U, V are orthogonal, and

$$\Sigma = \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & 0 & \\ & & \sigma_r & & \\ & 0 & & 0 & \\ & & & & \ddots\ 0 \\ & & & & \end{bmatrix}_{m\times n} \qquad \sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0$$

The σ_i are the "singular values" of A

- Cf: If $A_{n\times n}$ is symmetric, then it has an ON basis of e-vectors $\{v_1 \cdots v_n\}$ ⇒

$$\underset{n\times n}{V^{-1}}\,\underset{n\times n}{A}\,\underset{n\times n}{V} = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \quad \underline{\text{or}} \quad \boxed{A = VDV^T} \qquad (S)$$

(Since $V^T = V^{-1}$). Now (S) applies only to symmetric matrices, (*) to <u>any</u> m×n matrix, & $\sigma_i > 0$ while $\lambda_i < 0$ possible!

- For (*) $U_{m\times m}$ is different from $V_{n\times n}$

In fact:

① $U_{m\times m} = \begin{bmatrix} | & & | \\ u_1 & \cdots & u_m \\ | & & | \end{bmatrix}$ where u_i are an ON basis of e-vectors for the symmetr matrix $\underbrace{\overset{m\times n}{A}\,\overset{n\times m}{A^T}}_{m\times m}$

② $V_{n\times n} = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix}$ $\{v_i\}$ ON basis of e-vectors for $\underbrace{\overset{n\times m}{A^T}\,\overset{m\times n}{A}}_{n\times n}$ (symmetric)

③ $\Sigma_{m\times n} = \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}$ where $\sigma_1^2 \cdots \sigma_r^2$ are the nonzero eigenvalues of <u>both</u> AA^T and A^TA

- In case A complex, $A = U\Sigma\bar{V}^T$, U, V <u>unitary</u>

$\{u_i\}$ ON basis $A\bar{A}^T$, $\{v_i\}$ ON basis for $\bar{A}^TA$ . . .
We consider only real case ~ Strang

Thus:

$$\begin{array}{ccccc} & U & \Sigma & V^T \\ A_{m\times n} = & \begin{bmatrix} | & & | \\ u_1 & \cdots & u_m \\ | & & | \end{bmatrix} & \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \; 0 \end{bmatrix} & \begin{bmatrix} - & v_1 & - \\ & \vdots & \\ - & v_n & - \end{bmatrix} \\ & m\times m & m\times n & n\times n \end{array}$$

$\{u_i\}$ on basis in $\mathbb{R}^m$ $\{v_i\}$ on basis in $\mathbb{R}^n$

$$\begin{array}{ccc} A_{m\times n} = & \begin{bmatrix} | & & | \\ u_1 & \cdots & u_m \\ | & & | \end{bmatrix} & \begin{bmatrix} - & \sigma_1 v_1 & - \\ & \vdots & \\ - & \sigma_r v_r & - \\ & 0 & \\ & \vdots & \\ & 0 & \end{bmatrix} \\ & m\times m & m\times n \end{array}$$

Conclude:

$$A v_j = \begin{bmatrix} | & & | \\ u_1 & \cdots & u_n \\ | & & | \end{bmatrix} \begin{bmatrix} -\sigma_1 v_1- \\ \vdots \\ -\sigma_r v_r- \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{bmatrix} | \\ v_j \\ | \end{bmatrix} = \begin{bmatrix} | & & | \\ u_1 & \cdots & u_m \\ | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 v_1 \cdot v_j \\ \vdots \\ \sigma_r v_r \cdot v_j \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} | & & | \\ u_1 & \cdots & u_m \\ | & & | \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \sigma_j \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leftarrow j\text{th slot} = \begin{cases} \sigma_j u_j & j \le r \\ 0 & j > r \end{cases} \quad \begin{array}{l} v_i \cdot v_j = \delta_{ij} \\ \text{ON basis} \end{array}$$

If we let $\sigma_j = 0$ for $j > r$ we get

$$\boxed{A v_j = \sigma_j u_j}$$

- "σ_j tells the significance of the effect of v_j in matrix mult by A"

I.e. $\{v_1, \dots, v_n\}$ on basis for $\mathbb{R}^n \Rightarrow$

$$x = c_1 v_1 + \cdots + c_n v_n \qquad x \in \mathbb{R}^n$$

$$\langle v_i, x \rangle = \sum_{j=1}^{n} c_i \langle v_i, v_j \rangle = c_i$$

So

$$x = (x \cdot v_1) v_1 + \cdots + (x \cdot v_n) v_n$$

Thus

$$Ax = c_1 A v_1 + \cdots + c_n A v_n = c_1 \sigma_1 v_1 + \cdots + c_n \sigma_n v_n$$

$$= (x \cdot v_1) \sigma_1 v_1 + \cdots + (x \cdot v_n) \sigma_n v_n \qquad \sigma_k = 0 \;\; k > r$$

$$\boxed{Ax = \sigma_1 (x \cdot v_1) v_1 + \cdots + \sigma_r (x \cdot v_r) v_n}$$

Conclude: If $\sigma_k << 1$, v_k has little effect & can be ignored.

Moreover: because $\{u_i\}$ & $\{v_i\}$ are ON., (and hence as stable as possible under perturbation) they give the best measure of relative importance of basis elements. (Attempt to make this rigorous is the purpose of Trefethen's notes...)

Trefethen: "The most accurate method for finding an ON. basis for ~~the~~ range or nullspace of a matrix is via SVD. (QR factorization is faster but less accurate)... The SVD is an ingredient in robust algorithms for least squares fitting, intersection of subspaces, regularization & numerous others."

- Another way to view this —

$$A = \sigma_1 \underbrace{u_1 v_1^T}_{\substack{\text{rank-1} \\ \text{matrix}}} + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T$$

I.e.

$$u_1 v_1^T = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \begin{bmatrix} a_1 \cdots a_n \end{bmatrix} = \begin{bmatrix} (b_i a_j) \end{bmatrix}$$

$$(u_1 v_1^T)\, x = \underset{u_1}{\begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}} \underbrace{\begin{bmatrix} a_1, \ldots, a_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}}_{v_1 \cdot x} = (v_1 \cdot x)\, u_1$$

So

$$\left[\sum_{i=1}^{r} \sigma_i (u_i v_i^T) \right] x = \sum_{i=1}^{r} \sigma_i (v_i \cdot x)\, u_i = Ax \checkmark$$

Thus we can say: σ_i gives the magnitude of the effect of $u_i v_i^T$ in the matrix A.

Application: (Image Processing)

Image on 1000 x 1000 screen of pixels

Say each color assigned a number -

⇒ Image recorded by 1000 x 1000 matrix of #'s A

- Computer finds SVD of A:

$$A = \sum_{i=1}^{1000} \sigma_i (u_i v_i^T)$$

If $\sigma_i \ll 1$ for $i \geq 30$, image is encoded by $\sum_{i=1}^{30} \sigma_i u_i v_i^T$ much less information!

HW - Matlab Hand in for Extra Credit by Fri 12-1

<u>Note</u>: The SVD displays all of the Fundamental vector spaces associated with A:

$$A_{m\times n} = \overset{U}{\begin{bmatrix} | & & | \\ u_1 & \cdots & u_m \\ | & & | \end{bmatrix}} \overset{\Sigma}{\begin{bmatrix} \sigma_1 & & & & 0 \\ & \ddots & & & \\ 0 & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \; 0 \end{bmatrix}} \overset{V^T}{\begin{bmatrix} - & v_1 & - \\ & \vdots & \\ - & v_n & - \end{bmatrix}}$$

$$A = \begin{bmatrix} | & & | & | & & | \\ u_1 & \cdots & u_r & u_{r+1} & \cdots & u_m \\ | & & | & | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & 0 \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ 0 & & & & \ddots \; 0 \end{bmatrix} \begin{bmatrix} - & v_1 & - \\ & \vdots & \\ - & v_r & - \\ - & v_{r+1} & - \\ & \vdots & \\ - & v_n & - \end{bmatrix} \begin{matrix} \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} \text{Row}(A) \\ \oplus \\ \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} \begin{matrix} \text{Ker}(A) \\ \shortparallel \\ \text{Row}(A)^{\perp} \end{matrix} \end{matrix}$$

$$= \underbrace{\begin{bmatrix} | & & | \\ u_1 & \cdots & u_r \\ | & & | \end{bmatrix}}_{\text{Col}(A)} \!\oplus\! \underbrace{\begin{bmatrix} | & & | \\ u_{r+1} & \cdots & u_m \\ | & & | \end{bmatrix}}_{\begin{matrix} \text{Ker}(A^T) \\ \shortparallel \\ \text{Col}(A)^{\perp} \end{matrix}} \begin{bmatrix} - & \sigma_1 v_1 & - \\ & \vdots & \\ - & \sigma_r v_r & - \\ - & 0 & - \\ - & 0 & - \end{bmatrix} \begin{matrix} \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} \text{Row}(A) \\ \\ \\ \end{matrix}$$

☒ Derivation of SVD for $A_{m\times n}$:

- We have: Thm: $A^T = A \Rightarrow$ real evals & ON basis of eigenvectors so

$$A = SDS^T \qquad S = [u_1 \cdots u_n]$$

Only ~~holds~~ applies for $n\times n$ square symm. matrices

- Given $A_{m\times n}$, both $(A^TA)_{n\times n}$ & $(AA^T)_{m\times m}$ are symmetric.

Defn: an $n\times n$ symmetric matrix is positive semi-definite if $\lambda_i \geq 0$ ∀ eval λ_i

Thm: A^TA & AA^T are both pos semi-defi
with the same nonzero evals $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_r^2$

To prove this we need:

Lemma: $A_{n \times n}$ symmetric. Then A is pos semi-def iff

$$v^TAv \geq 0 \quad \forall\, v \in \mathbb{R}^n$$

Pf. (⇒) If evals of A satis $\lambda_i \geq 0$, then

$$A = SDS^T \qquad D = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \quad \lambda_i \geq 0$$

So

$$v^TAv = (v^TS)D(S^Tv) = (S^Tv)^TD(\underbrace{S^Tv}_{w})$$

$$= w^TDw = \sum_{i=1}^{n} w_i^2 \lambda_i \geq 0 \checkmark$$

($\Leftarrow$) If $v^T A v \ge 0 \ \forall \ v \in \mathbb{R}^n$, then choose $v = v_i$ evector for λ_i to get

$$v_i^T A v_i = \lambda_i v_i^T v_i = \lambda \|v_i\|^2 \ge 0$$

$\Rightarrow \lambda_i \ge 0 \ \forall \ i$. ✓

Cor: $A^T A$ & AA^T are always pos semidef

Pf. $$v^T A^T A v = (Av)^T \underset{\substack{\uparrow \\ \text{matrix}}}{\cdot} (Av) = (Av) \underset{\substack{\uparrow \\ \text{dot}}}{\cdot} (Av) = \|Av\|^2 \ge \quad \forall \ v$$

$$u^T A A^T u = (A^T u) \underset{\substack{\uparrow \\ \text{dot}}}{\cdot} (A^T u) = \|A^T u\|^2 \ge 0 \quad \forall \ w$$

$\Rightarrow$ both are symm pos semi-def.

Conclude:

$$A^TA = V\begin{bmatrix} \sigma_1^2 & & & & 0 \\ & \ddots & & & \\ & & \sigma_r^2 & & \\ & & & 0 & \\ 0 & & & & \ddots\, 0 \end{bmatrix} V^T \qquad V = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix} \quad \text{(v)}$$

$\{v_1 \cdots v_n\}$ o.n. basis of eigenvectors

(WLOG order $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_r^2 \geq 0$, $\sigma_j^2 = 0$ $j \geq r$)

and

$$AA^T = U\begin{bmatrix} \bar{\sigma}_1^2 & & & & 0 \\ & \ddots & & & \\ & & \bar{\sigma}_r^2 & & \\ & & & 0 & \\ 0 & & & & \ddots\, 0 \end{bmatrix} U^T \qquad U = \begin{bmatrix} | & & | \\ u_1 & \cdots & u_m \\ | & & | \end{bmatrix} \quad \text{(u}$$

$\{u_1, \ldots, u_m\}$ o.n. basis of eigenvectors.

Claim: For the right choice of $u_1, \ldots, u_M$, we have $\bar{r} = r$ & $\bar{\sigma}_i^2 = \sigma_i^2$ $i = 1, \ldots, r$.

Pf. Consider first $\{v_1, \ldots, v_r\}$ an ON. set of basis vectors $\in \mathbb{R}^n$ spanning the non-zero eigenspaces of A^TA. We show

$$\left\{u_1 = \frac{1}{\sigma_1} Av_1, \ldots, u_r = \frac{1}{\sigma_r} Av_r\right\}$$

is an ON. set of vectors $\in \mathbb{R}^m$ satisfying

$$AA^Tu_i = \sigma_i^2 u_i. \qquad (*)$$

This shows that with this choice of u_i's in (u), $\bar{\sigma}_i^2 = \sigma_i^2$ $i = 1, \ldots, r$, so in particular $\bar{r} \geq r$. Reversing the roles of u & v then shows $\bar{r} \leq r \Rightarrow \bar{r} = r$.

For (*), assume

$$A^T A v_i = \sigma_i^2 v_i$$

Mult by $A \Rightarrow$ $AA^T(Av_i) = \sigma_i^2 (Av_i)$

$\Rightarrow$ Av_i is an eigenvector of AA^T for e-val σ_i^2

Also

$$\|Av_i\|^2 = \langle Av_i, Av_i \rangle = v_i^T \underbrace{A^T A v_i}_{\sigma_i^2 v_i} = \sigma_i^2 v_i^T \cdot v_i$$

$$= \sigma_i^2 \|v_i\|^2$$

$$= \sigma_i^2$$

$\therefore$ $u_i = \frac{1}{\sigma_i} Av_i$ is a unit e-vector of AA^T.

for every σ_i^2 $i = 1, \ldots, r$.

Now AA^T symmetric $\Rightarrow$ eigenvectors from different e-vals are $\perp$. It remains to prove that if $v_i \perp v_j$ are e-vectors from the same $\sigma^2 > 0$ then $u_i \perp u_j$ as well.

For this, assume $v_i \perp v_j$ are e-vectors for the same e-val $\sigma^2 \neq 0$ of A^TA. Then $u_i = \frac{1}{\sigma} A v_i$ & $u_j = \frac{1}{\sigma} A v_j$ are unit and

$$\langle u_i, u_j \rangle = \langle \tfrac{1}{\sigma} A v_i, \tfrac{1}{\sigma} A v_j \rangle$$

$$= \frac{1}{\sigma^2} \langle A v_i, A v_j \rangle$$

$$= \frac{1}{\sigma^2} v_i^T \underbrace{A^T A v_j}_{\sigma^2 v_j} = \frac{1}{\cancel{\sigma^2}} \cancel{\sigma^2} \, v_i^T \cdot v_j$$

$$= v_i \overset{\text{dot}}{\overset{\downarrow}{\cdot}} v_j = 0 \checkmark \quad (\text{since } v_i \perp v_j)$$

Thus we can take ⓥ & ⓤ as valid for $\bar{\sigma}_i = \sigma_i$, $\bar{r} = r$ & $u_i = \frac{1}{\sigma_i} A v_i$ $i = 1, \dots, r$.

(Complete these to an ON basis $v_1, \dots, v_n$ and $u_1, \dots, u_m$ any way you like!)

☐ we can now prove that —

$$A_{m\times n} = \begin{bmatrix} | & & | \\ u_1 & \cdots & u_m \\ | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & 0 \\ & \ddots & & & \\ & & \sigma_r & & \\ 0 & & & 0 & \\ & & & & \ddots \; 0 \end{bmatrix} \begin{bmatrix} — & v_1 & — \\ & \vdots & \\ — & v_n & — \end{bmatrix}$$

For this it suffices to show they agree on a basis. But we have on RHS

$$(U \Sigma V^T) v_j = \sigma_j u_j$$

while we just showed that on LHS

$$A\left(\frac{1}{\sigma_j}\right) v_j = u_j$$

$$A v_j = \sigma_j u_j$$

Two matrices that agree on a basis must be equal!

- That completes the justification of the SVD of any real m×n matrix A ✓