Singular Value Decompositions

A – $m \times n$ matrix

Aim: Find orthonormal bases $\mathcal{B} \subset \mathbb{R}^n$ and $\mathcal{C} \subset \mathbb{R}^m$ such that

$A_{\mathcal{B},\mathcal{C}}$ is as simple as possible.

Matrix of T_A with respect to $\mathcal{B}$ and $\mathcal{C}$

Important Observation: $(A^T A)^T = A^T (A^T)^T = A^T A \Rightarrow A^T A$ symmetric

$n \times n$ matrix

$A^T A$ symmetric $\Rightarrow$ $A^T A$ orthogonally diagonalizable

Let $\mathcal{B} = \{\underline{v}_1, \dots, \underline{v}_n\} \subset \mathbb{R}^n$ an orthogonal basis of eigenvectors of $A^T A$.

eigenvalue

$\Rightarrow$ $A^T A \underline{v}_i = \lambda_i \underline{v}_i$ for all i

$\Rightarrow$ $\underline{v}_i^T A^T A \underline{v}_i = \underline{v}_i^T \lambda_i \underline{v}_i$ (recall $\|\underline{v}_i\| = 1 \Rightarrow \underline{v}_i^T \underline{v}_i = 1$)

$\Rightarrow$ $(A\underline{v}_i)^T (A\underline{v}_i) = \lambda_i$

$\Rightarrow$ $\|A\underline{v}_i\|^2 = \lambda_i \Rightarrow \lambda_i \geq 0$ for all i

Perhaps after reordering, assume $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$

Definition The singular values of A are the numbers

$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$ where $\sigma_i = \sqrt{\lambda_i}$.

Key Fact: $\sigma_i = \|A\underline{v}_i\|$

Theorem Assume $\sigma_1, \dots, \sigma_r \neq 0$ and $\sigma_{r+1}, \dots, \sigma_n = 0$.
Then $\mathrm{Rank}(A) = r$ and $\{A\underline{v}_1, \dots, A\underline{v}_r\}$ is an orthogonal basis for $\mathrm{Col}(A)$.

$\underline{\text{Proof}}$ $\{\underline{v}_1, \dots, \underline{v}_n\} \subset \mathbb{R}^n$ a basis $\Rightarrow$ $\{A\underline{v}_1, \dots, A\underline{v}_n\}$ spans $\mathrm{Col}(A)$ $\overset{\text{Range}(T_A)}{\|}$

$\|A\underline{v}_i\| = \sigma_i \quad \Rightarrow \quad A\underline{v}_i = \underline{0} \Leftrightarrow \sigma_i = 0$

$\Rightarrow$ $\{A\underline{v}_1, \dots, A\underline{v}_r\}$ spans $\mathrm{Col}(A)$ and all elements are $\underline{\text{non-zero}}$.

Let $i \neq j$

$$(A\underline{v}_i) \cdot (A\underline{v}_j) = (A\underline{v}_i)^T (A\underline{v}_j) = \underline{v}_i^T A^T A \underline{v}_j = \underline{v}_i^T \lambda_j \underline{v}_j$$
$$= \lambda_j \underline{v}_i^T \underline{v}_j = \lambda_j (\underline{v}_i \cdot \underline{v}_j) = 0$$

($\underline{v}_i, \underline{v}_j$ orthogonal if $i \neq j$)

$\Rightarrow$ $\{A\underline{v}_1, \dots, A\underline{v}_r\}$ is orthogonal set with non-zero vectors $\Rightarrow$ L.I.

$\Rightarrow$ $\{A\underline{v}_1, \dots, A\underline{v}_r\}$ is an orthogonal basis for $\mathrm{Col}(A)$

$\Rightarrow$ $\mathrm{Rank}(A) = r$ $\square$

$\underline{\text{Example}}$

$$A = \begin{pmatrix} \sqrt{2} & -1/\sqrt{2} & 0 \\ \sqrt{2} & 1/\sqrt{2} & 0 \end{pmatrix} \Rightarrow A^T A = \begin{pmatrix} \sqrt{2} & \sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \sqrt{2} & -1/\sqrt{2} & 0 \\ \sqrt{2} & 1/\sqrt{2} & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \leftarrow \text{already diagonal}$$

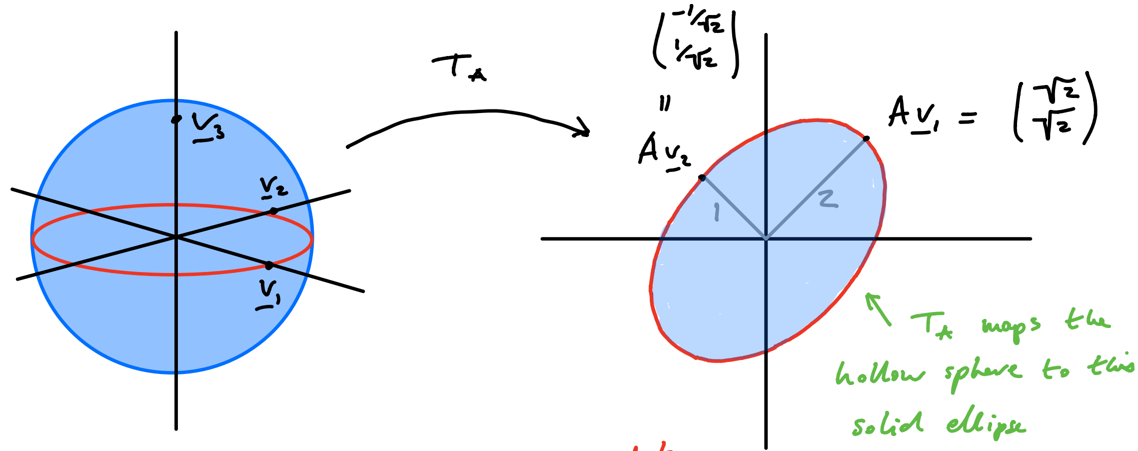
Choose $\underline{v}_1 = \underline{e}_1$, $\underline{v}_2 = \underline{e}_2$, $\underline{v}_3 = \underline{e}_3$

$A^T A \underline{v}_1 = 4\underline{v}_1$, $A^T A \underline{v}_2 = \underline{v}_2$, $A^T A \underline{v}_3 = \underline{0}$ $\Rightarrow$ $4, 1, 0$ are eigenvalues

$\Rightarrow$ $2, 1, 0$ are $\underline{\text{singular values}}$

$$\{A\underline{v}_1, A\underline{v}_2\} = \left\{ \begin{pmatrix} \sqrt{2} \\ \sqrt{2} \end{pmatrix}, \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \right\} = \text{basis for } \mathrm{Col}(A)$$

very obvious in this case as $\underline{v}_1 = \underline{e}_1$, $\underline{v}_2 = \underline{e}_2$



<u>Important General Fact:</u>

$\underline{x}$ restricted to "sphere" of radius 1 center $\underline{0}$ in $\mathbb{R}^n$

$$\text{Max}\ \{\, \|A\underline{x}\| \text{ where } \|\underline{x}\| = 1 \,\} = \text{Max}\ \{\sigma_1, \ldots, \sigma_n\} = \sigma_1$$

$$\Rightarrow \text{Max}\ \{\, \|A\underline{x}\| \text{ where } \|\underline{x}\| = d \,\} = d\,\sigma_1$$

$\underline{x}$ restricted to "sphere" of radius d center $\underline{0}$ in $\mathbb{R}^n$

$$\text{Min}\ \{\, \|A\underline{x}\| \text{ where } \|\underline{x}\| = 1 \,\} = \text{Min}\ \{\sigma_1, \ldots, \sigma_n\} = \sigma_n$$

$$\Rightarrow \text{Min}\ \{\, \|A\underline{x}\| \text{ where } \|\underline{x}\| = d \,\} = d\,\sigma_n$$

Back to general case:

Let $\underline{u}_i = \dfrac{1}{\|A\underline{v}_i\|} A\underline{v}_i = \dfrac{1}{\sigma_i} A\underline{v}_i$ for $1 \le i \le r$

Key Facts

1/ $\{\underline{u}_1, \ldots, \underline{u}_r\}$ is an <u>orthonormal basis</u> for $Col(A)$

2/ $A\underline{v}_i = \sigma_i \underline{u}_i$ for $1 \le i \le r$

<u>Observation</u> We can extend $\{\underline{u}_1, \ldots, \underline{u}_r\}$ is an <u>orthonormal basis</u> for all $\mathbb{R}^m$ as follows:

a) Find basis for $Nul(A^T) = (Col(A))^{\perp}$

b) Apply Gram-Schmidt and normalize to get orthonormal basis

$\{\underline{u}_{r+1}, \ldots, \underline{u}_m\}$ $\leftarrow$ must have size $m-r$ as $\dim(\mathrm{Col}(A)) + \dim(\mathrm{Col}(A))^{\perp}) = m$

c) Take union to give $\{\underline{u}_1, \ldots \underline{u}_r, \underline{u}_{r+1}, \ldots, \underline{u}_m\}$ an orthonormal basis for $\mathbb{R}^m$.

Singular Value Decompositions

Let A be an $m \times n$ matrix together with

1/ $\mathcal{B} = \{\underline{v}_1, \ldots \underline{v}_n\} \subset \mathbb{R}^n$ orthonormal basis of eigenvectors of A^TA with singular values

$$\sigma_1 \geq \sigma_2 \cdots \geq \sigma_r \geq \underbrace{\sigma_{r+1} \geq \cdots \geq \sigma_n}_{\text{zero}} \qquad (A^TA\underline{v}_i = \sigma_i^2 \underline{v}_i)$$

2/ $C = \{\underline{u}_1, \ldots, \underline{u}_m\} \subset \mathbb{R}^m$ orthonormal basis with $A\underline{v}_i = \sigma_i \underline{u}_i$ for all $1 \leq i \leq n$

$$\Rightarrow A_{\mathcal{B},C} = \left((A\underline{v}_1)_C \cdots (A\underline{v}_r)_C \; (A\underline{v}_{r+1})_C \cdots (A\underline{v}_n)_C\right)$$

Matrix associated to T_A with respect to $\mathcal{B}$ and C

$$= \left((\sigma_1\underline{u}_1)_C \cdots (\sigma_r\underline{u}_r)_C \; (\underline{0})_C \cdots (\underline{0})_C\right)$$

$$= \left(\begin{array}{ccc|c} \sigma_1 & & 0 & \\ & \ddots & & 0 \\ 0 & & \sigma_r & \\ \hline & 0 & & 0 \end{array}\right) \leftarrow \text{often denoted } \Sigma$$

Important Consequence: If $U = (\underline{u}_1 \cdots \underline{u}_m)$, $V = (\underline{v}_1 \cdots \underline{v}_n)$

Orthogonal Matrices

$U^T = U^{-1}$

$V^T = V^{-1}$

$$\Rightarrow U^{-1}AV = \left(\begin{array}{ccc|c} \sigma_1 & & 0 & \\ & \ddots & & 0 \\ 0 & & \sigma_r & \\ \hline & 0 & & 0 \end{array}\right)$$

$$\Rightarrow A = U\left(\begin{array}{ccc|c} \sigma_1 & & 0 & \\ & \ddots & & 0 \\ 0 & & \sigma_r & \\ \hline & 0 & & 0 \end{array}\right)V^{-1} = U\left(\begin{array}{ccc|c} \sigma_1 & & 0 & \\ & \ddots & & 0 \\ 0 & & \sigma_r & \\ \hline & 0 & & 0 \end{array}\right)V^T$$

Called a singular value decomposition of A

Remarks

$$\mathbb{R}^n \xrightarrow{\;T_A\;} \mathbb{R}^m$$

$$\underline{x} \longmapsto A\underline{x}$$

$$\underline{v}_i \longmapsto \begin{cases} \overset{\text{non-zero}}{\sigma_i}\, \underline{u}_i & i \le r \\ \underline{0} & \text{otherwise} \end{cases}$$

$$\{\underbrace{\underline{v}_1,\dots,\underline{v}_r}_{Col(A^T)},\ \underbrace{\underline{v}_{r+1},\dots,\underline{v}_n}_{Nul(A)}\} \qquad \{\underbrace{\underline{u}_1,\dots,\underline{u}_r}_{Col(A)},\ \underbrace{\underline{u}_{r+1},\dots,\underline{u}_m}_{Nul(A^T)}\}$$

Overview of Finding SVD :

1. Orthogonally diagonalize A^TA giving orthonormal basis $\{\underline{v}_1,\dots,\underline{v}_n$
2. Reorder so that eigenvalues are decreasing $\lambda_1 \ge \lambda_2 \ge \dots \ge \lambda_n \ge 0$
3. Take square roots to give $\overbrace{\sigma_1 \ge \sigma_2 \ge \dots \ge \sigma_r}^{\text{non-zero}} \ge \overbrace{\sigma_{r+1} \ge \dots \ge \sigma_n}^{\text{zero}}$
4. Define $\underline{u}_i = \dfrac{1}{\sigma_i} A\underline{v}_i$ for $1 \le i \le r$.
5. Extend $\{\underline{u}_1,\dots,\underline{u}_r\}$ to orthonormal basis $\{\underline{u}_1,\dots,\underline{u}_r,\underline{u}_{r+1},\dots,\underline{u}_m\}$
6. $U = (\underline{u}_1 \dots \underline{u}_m)$
   $V = (\underline{v}_1 \dots \underline{v}_m)$
   $\Sigma = \left(\begin{array}{c|c} \begin{matrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r \end{matrix} & 0 \\ \hline 0 & 0 \end{array}\right)$ $\Rightarrow$ $A = U\Sigma V^T$

S.V.D.

Terminology : $\{\underline{u}_1,\dots,\underline{u}_m\}$ = <u>left</u> singular vectors of A

$\{\underline{v}_1,\dots,\underline{v}_n\}$ = <u>right</u> singular vectors of A

SVD have <u>many</u> important applications

Example: Image Processing

Imagine we have a greyscale image such as the following:

Original Lena Image

512×512 resolution

We can encode this data with a 512×512 matrix A where each entry represents the brightness of the corresponding pixel.

SVD: $A = U \Sigma V^T$.

For any $1 \le k \le 512$ let
$$\Sigma_k := \left(\begin{array}{ccc|c} \sigma_1 & \cdots & 0 & \\ & \ddots & & 0 \\ 0 & & \sigma_k & \\ \hline & 0 & & 0 \end{array}\right)$$

And $A_k := U \Sigma_k V^T$

If $k \le \text{Rank}(A) \Rightarrow \text{Rank}(A_k) = k$

Important Fact:

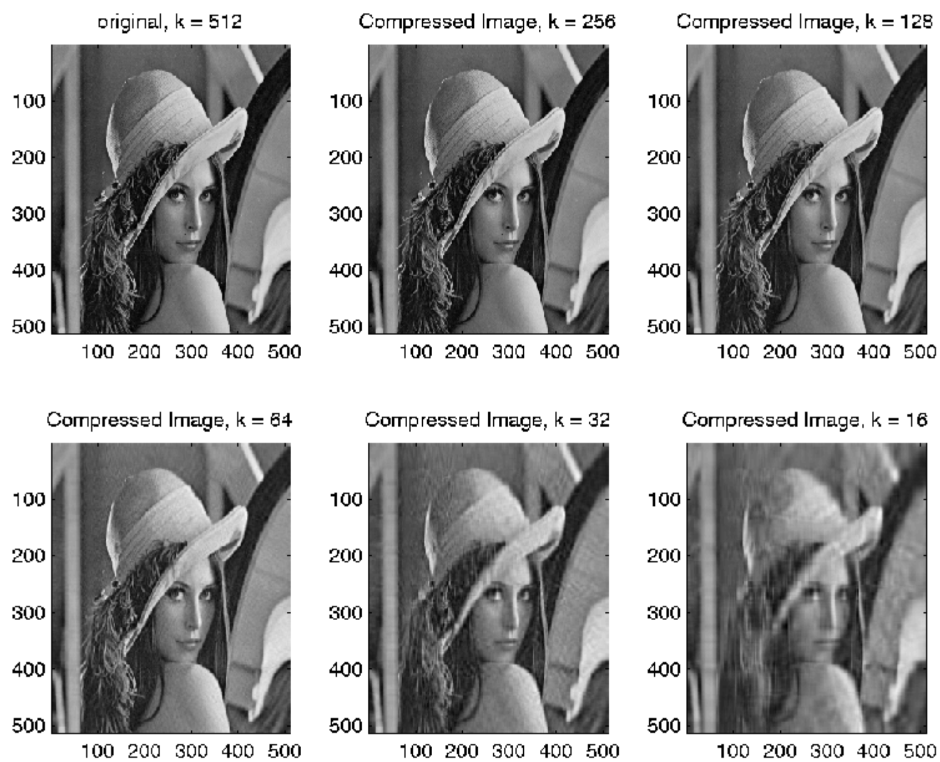
A_k is a "best possible" rank k approximation of A.

entries are as close as possible

$A_k = \sigma_1 \underline{u}_1 \underline{v}_1^T + \sigma_2 \underline{u}_2 \underline{v}_2^T + \cdots + \sigma_k \underline{u}_k \underline{v}_k^T$

Need $k + 2k \cdot 512$ numbers to determine

Each A_k corresponds to a "compressed" version of original image A.



This has all sorts of applications in A.I., cognitive science and machine learning, not to mention letting you store more pictures on your phone.