

# CS724: Topics in Algorithms

## Singular Values

### Slide Set 8

Prof. Dan A. Simovici



## 1 Introduction

## 2 Singular Values and Singular Vectors

## 3 MATLAB Computations



The singular value decomposition has been described as the “Swiss Army knife of matrix decompositions” due to its many applications in the study of matrices; from our point of view, singular value decomposition is relevant for dimensionality reduction techniques in data mining.



Let  $A \in \mathbb{C}^{n \times n}$  be a square matrix which is unitarily diagonalizable. There exists a diagonal matrix  $D = \text{diag}(d_1, \dots, d_n) \in \mathbb{C}^{n \times n}$  and a unitary matrix  $X \in \mathbb{C}^{n \times n}$  such that  $A = XDX^H$ ; equivalently, we have  $AX = XD$ . If we denote the columns of  $X$  by  $\mathbf{x}_1, \dots, \mathbf{x}_n$ , then  $A\mathbf{x}_i = d_i\mathbf{x}_i$ , which shows that  $\mathbf{x}_i$  is a unit eigenvector that corresponds to the eigenvalue  $d_i$  for  $1 \leq i \leq n$ . Also, we have

$$\begin{aligned} A &= (\mathbf{x}_1 \cdots \mathbf{x}_n) \text{diag}(d_1, \dots, d_n) \begin{pmatrix} \mathbf{x}_1^H \\ \vdots \\ \mathbf{x}_n^H \end{pmatrix} \\ &= d_1 \mathbf{x}_1 \mathbf{x}_1^H + \cdots + d_n \mathbf{x}_n \mathbf{x}_n^H. \end{aligned}$$

This is the *spectral decomposition of  $A$* . Note that  $\text{rank}(\mathbf{x}_i \mathbf{x}_i^H) = 1$  for  $1 \leq i \leq n$ .



The SVD theorem extends this decomposition to rectangular matrices.

## Theorem

**(SVD Theorem)** If  $A \in \mathbb{C}^{m \times n}$  is a matrix and  $\text{rank}(A) = r$ , then  $A$  can be factored as  $A = UDV^H$ , where  $U \in \mathbb{C}^{m \times m}$  and  $V \in \mathbb{C}^{n \times n}$  are unitary matrices,

$$D = \begin{pmatrix} \sigma_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \sigma_2 & 0 & & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_r & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \end{pmatrix} \in \mathbb{C}^{m \times n},$$

and  $\sigma_1 \geq \dots \geq \sigma_r$  are real positive numbers.



# Proof

The square matrix  $A^H A \in \mathbb{C}^{n \times n}$  has the same rank as the matrix  $A$  and is positive semidefinite. Therefore, there are  $r$  positive eigenvalues of this matrix, denoted by  $\sigma_1^2, \dots, \sigma_r^2$ , where  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ .

Let  $\mathbf{v}_1, \dots, \mathbf{v}_r$  be the corresponding pairwise orthogonal, unit eigenvectors in  $\mathbb{C}^n$  and let

$$V_1 = (\mathbf{v}_1 \ \cdots \ \mathbf{v}_r),$$

$V = (\mathbf{v}_1 \ \cdots \ \mathbf{v}_r \ \mathbf{v}_{r+1} \ \cdots \ \mathbf{v}_n)$  be the matrix obtained by completing the set  $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$  to an orthogonal basis for  $\mathbb{C}^n$ , and let  $V_2 = (\mathbf{v}_{r+1} \ \cdots \ \mathbf{v}_n)$ .

We can write  $V = (V_1 \ V_2)$ .

Note that we have  $A^H A \mathbf{v}_i = \sigma_i^2 \mathbf{v}_i$  for  $1 \leq i \leq r$ .



## Proof cont'd

The equalities  $A^H A \mathbf{v}_i = \sigma_i^2 \mathbf{v}_i$  for  $1 \leq i \leq r$  involving the eigenvectors can now be written as

$$A^H A V_1 = V_1 E^2,$$

where  $E = \text{diag}(\sigma_1, \dots, \sigma_r)$ .

Define  $U_1 = A V_1 E^{-1} \in \mathbb{C}^{m \times r}$ . We have  $U_1^H = E^{-1} V_1^H A^H$ , so

$$U_1^H U_1 = E^{-1} V_1^H A^H A V_1 E^{-1} = E^{-1} V_1^H V_1 E^2 E^{-1} = I_r,$$

which shows that the columns of  $U_1$  are pairwise orthogonal unit vectors. Consequently,  $U_1^H A V_1 E^{-1} = I_r$ , so  $U_1^H A V_1 = E$ .



## Proof cont'd

If  $U_1 = (\mathbf{u}_1 \cdots, \mathbf{u}_r)$ , let  $U_2 = (\mathbf{u}_{r+1}, \dots, \mathbf{u}_m)$  be the matrix whose columns constitute the extension of the set  $\{\mathbf{u}_1 \cdots, \mathbf{u}_r\}$  to an orthogonal basis of  $\mathbb{C}^m$ . Define  $U \in \mathbb{C}^{m \times m}$  as  $U = (U_1 \ U_2)$ . Note that

$$\begin{aligned} U^H A V &= \begin{pmatrix} U_1^H \\ U_2^H \end{pmatrix} A (V_1 \ V_2) \\ &= \begin{pmatrix} U_1^H A V_1 & U_1^H A V_2 \\ U_2^H A V_1 & U_2^H A V_2 \end{pmatrix} = \begin{pmatrix} U_1^H A V_1 & U_1^H A V_2 \\ U_2^H A V_1 & U_2^H A V_2 \end{pmatrix} \\ &= \begin{pmatrix} U_1^H A V_1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} E & 0 \\ 0 & 0 \end{pmatrix}, \end{aligned}$$

which is the desired decomposition.



Observe that in the SVD described above known as the *full SVD* of  $A$ , the diagonal matrix  $D$  has the same format as  $A$ , while both  $U$  and  $V$  are square unitary matrices.



## Definition

Let  $A \in \mathbb{C}^{m \times n}$  be a matrix. A number  $\sigma \in \mathbb{R}_{>0}$  is a *singular value* of  $A$  if there exists a pair of vectors  $(\mathbf{u}, \mathbf{v}) \in \mathbb{C}^m \times \mathbb{C}^n$  such that

$$A\mathbf{v} = \sigma\mathbf{u} \text{ and } A^H\mathbf{u} = \sigma\mathbf{v}. \quad (1)$$

The vector  $\mathbf{u}$  is the *left singular vector* and  $\mathbf{v}$  is the *right singular vector* associated to the singular value  $\sigma$ .



Note that if  $(\mathbf{u}, \mathbf{v})$  is a pair of vectors associated to  $\sigma$ , then  $(a\mathbf{u}, a\mathbf{v})$  is also a pair of vectors associated with  $\sigma$  for every  $a \in \mathbb{C}$ .



Let  $A \in \mathbb{C}^{m \times n}$  and let  $A = UDV^H$ , where  $U \in \mathbb{C}^{m \times m}$ ,  $D = \text{diag}(\sigma_1, \dots, \sigma_r, 0, \dots, 0) \in \mathbb{C}^{m \times n}$  and  $V \in \mathbb{C}^{n \times n}$ .

Note that

$$\begin{aligned} A\mathbf{v}_j &= UDV^H\mathbf{v}_j = UDe_j \\ &\quad (\text{because } V \text{ is a unitary matrix}) \\ &= \sigma_j U\mathbf{e}_j = \sigma_j \mathbf{u}_j \end{aligned}$$

and

$$\begin{aligned} A^H\mathbf{u}_j &= VD^H U^H\mathbf{u}_j = VDU^H\mathbf{u}_j \\ &\quad VDe_j \\ &\quad (\text{because } U \text{ is a unitary matrix}) \\ &= \sigma_j V\mathbf{e}_j = \sigma_j \mathbf{v}_j. \end{aligned}$$

Thus, the  $j^{\text{th}}$  column of the matrix  $U$ ,  $\mathbf{u}_j$  and the  $j^{\text{th}}$  column of the matrix  $V$ ,  $\mathbf{v}_j$  are left and right singular vectors, respectively, associated to the singular value  $\sigma_j$ .





## Definition

Two matrices  $A, B \in \mathbb{C}^{m \times n}$  are *unitarily equivalent* (denoted by  $A \equiv_u B$ ) if there exist two unitary matrices  $W_1$  and  $W_2$  such that  $A = W_1^H B W_2$ . Clearly, if  $A \sim_u B$ , then  $A \equiv_u B$ .

## Theorem

*Let  $A$  and  $B$  be two matrices in  $\mathbb{C}^{m \times n}$ . If  $A$  and  $B$  are unitarily equivalent, then they have the same singular values.*



# Proof

Suppose that  $A \equiv_u B$ , that is,  $A = W_1^H B W_2$  for some unitary matrices  $W_1$  and  $W_2$ . If  $A$  has the SVD  $A = U^H \text{diag}(\sigma_1, \dots, \sigma_r, 0, \dots, 0) V$ , then

$$B = W_1 A W_2^H = (W_1 U^H) \text{diag}(\sigma_1, \dots, \sigma_r, 0, \dots, 0) (V W_2^H).$$

Since  $W_1 U^H$  and  $V W_2^H$  are both unitary matrices, it follows that the singular values of  $B$ .



Let  $\mathbf{v} \in \mathbb{C}^n$  be an eigenvector of the matrix  $A^H A$  that corresponds to a non-zero, positive eigenvalue  $\sigma^2$ , that is,  $A^H A \mathbf{v} = \sigma^2 \mathbf{v}$ . Define  $\mathbf{u} = \frac{1}{\sigma} A \mathbf{v}$ . We have  $A \mathbf{v} = \sigma \mathbf{u}$ . Also,

$$A^H \mathbf{u} = A^H \left( \frac{1}{\sigma} A \mathbf{v} \right) = \sigma \mathbf{v}.$$

This implies  $AA^H \mathbf{u} = \sigma^2 \mathbf{u}$ , so  $\mathbf{u}$  is an eigenvector of  $AA^H$  that corresponds to the same eigenvalue  $\sigma^2$ .

Conversely, if  $\mathbf{u} \in \mathbb{C}^m$  is an eigenvector of the matrix  $AA^H$  that corresponds to a non-zero, positive eigenvalue  $\sigma^2$ , we have  $AA^H \mathbf{u} = \sigma^2 \mathbf{u}$ . Thus, if  $\mathbf{v} = \frac{1}{\sigma} A \mathbf{u}$  we have  $A \mathbf{v} = \sigma \mathbf{u}$  and  $\mathbf{v}$  is an eigenvector of  $A^H A$  for the eigenvalue  $\sigma^2$ .



The Courant-Fisher Theorem for eigenvalues allows the formulation of a similar result for singular values.

## Theorem

*Let  $A$  be a matrix,  $A \in \mathbb{C}^{m \times n}$ . If  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_k \geq \dots$  is the non-increasing sequence of singular values of  $A$ , then*

$$\sigma_k = \min_{\dim(S)=n-k+1} \max\{\|A\mathbf{x}\|_2 \mid \mathbf{x} \in S \text{ and } \|\mathbf{x}\|_2 = 1\}$$

$$\sigma_k = \max_{\dim(T)=k} \min\{\|A\mathbf{x}\|_2 \mid \mathbf{x} \in T \text{ and } \|\mathbf{x}\|_2 = 1\},$$

*where  $S$  and  $T$  range over subspaces of  $\mathbb{C}^n$ .*



# Proof

We give the argument only for the second equality of the theorem; the first can be shown in a similar manner.

We saw that  $\sigma_k$  equals the  $k^{\text{th}}$  largest absolute value of the eigenvalue  $|\lambda_k|$  of the matrix  $A^H A$ . By Courant-Fisher Theorem, we have

$$\begin{aligned}\lambda_k &= \max_{\dim(T)=k} \min_{\mathbf{x}} \{ \mathbf{x}^H A^H A \mathbf{x} \mid \mathbf{x} \in T \text{ and } \|\mathbf{x}\|_2 = 1 \} \\ &= \max_{\dim(T)=k} \min_{\mathbf{x}} \{ \|\mathbf{A} \mathbf{x}\|_2^2 \mid \mathbf{x} \in T \text{ and } \|\mathbf{x}\|_2 = 1 \},\end{aligned}$$

which implies the second equality of the theorem.



## Theorem

Let  $A$  be a matrix,  $A \in \mathbb{C}^{m \times n}$ . If  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_k \geq \dots$  is the non-increasing sequence of singular values of  $A$ , then

$$\begin{aligned}\sigma_k &= \min_{\mathbf{w}_1, \dots, \mathbf{w}_{k-1}} \max \{ \|A\mathbf{x}\|_2 \mid \mathbf{x} \perp \mathbf{w}_1, \dots, \mathbf{x} \perp \mathbf{w}_{k-1} \text{ and } \|\mathbf{x}\|_2 = 1 \} \\ &= \max_{\mathbf{w}_1, \dots, \mathbf{w}_{n-k}} \min \{ \|A\mathbf{x}\|_2 \mid \mathbf{x} \perp \mathbf{w}_1, \dots, \mathbf{x} \perp \mathbf{w}_{n-k} \text{ and } \|\mathbf{x}\|_2 = 1 \}.\end{aligned}$$



## Corollary

*The smallest singular value of a matrix  $A \in \mathbb{C}^{m \times n}$  equals*

$$\min\{\|Ax\|_2 \mid x \in \mathbb{C}^n \text{ and } \|x\|_2 = 1\}.$$

*The largest singular value of a matrix  $A \in \mathbb{C}^{m \times n}$  equals*

$$\max\{\|Ax\|_2 \mid x \in \mathbb{C}^n \text{ and } \|x\|_2 = 1\}.$$



The SVD theorem can also be proven by induction on  $q = \min\{m, n\}$ . In the base case,  $q = 1$ , we have  $A \in \mathbb{C}^{1 \times 1}$ , or  $A \in \mathbb{C}^{m \times 1}$ , or  $A \in \mathbb{C}^{1 \times n}$ . Suppose, for example, that  $A = \mathbf{a} \in \mathbb{C}^{m \times 1}$ , where

$$\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix}$$

and let  $a = \|\mathbf{a}\|_2$ . We seek  $U \in \mathbb{C}^{m \times m}$ ,  $V = (v) \in \mathbb{C}^{1 \times 1}$  such that

$$\mathbf{a} = UDv,$$

where  $D \in \mathbb{C}^{m \times 1}$  is

$$D = \begin{pmatrix} a \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{C}^{m \times 1}.$$



The role of the matrix  $U$  is played by any unitary matrix which has the first column equal to

$$\begin{pmatrix} \frac{a_1}{a} \\ \frac{a_2}{a} \\ \vdots \\ \frac{a_n}{a} \end{pmatrix},$$

and we can adopt  $v = 1$ . The remaining base subcases can be treated in a similar manner.



Suppose now that the statement holds when at least one of the numbers  $m$  and  $n$  is less than  $q$  and let us prove the assertion when at least one of  $m$  and  $n$  is less than  $q + 1$ .

Let  $\mathbf{u}_1$  be a unit eigenvector of  $AA^H$  that corresponds to the eigenvalue  $\sigma_1^2$ . We have:

$$AA^H \mathbf{u}_1 = \sigma_1^2 \mathbf{u}_1.$$

Define  $\mathbf{v}_1 = \frac{1}{\sigma_1} A^H \mathbf{u}_1$ . Note that:

$$\begin{aligned} \|\mathbf{v}_1\|^2 &= \mathbf{v}_1^H \mathbf{v}_1 = \frac{1}{\sigma_1^2} \mathbf{u}_1^H A A^H \mathbf{u}_1 \\ &= \mathbf{u}_1^H \mathbf{u}_1 = \|\mathbf{u}_1\|^2 = 1. \end{aligned}$$



Also, we have:

$$A\mathbf{v}_1 = \frac{1}{\sigma_1}AA^H\mathbf{u}_1 = \sigma_1\mathbf{u}_1,$$

which shows that  $(\mathbf{v}_1, \mathbf{u}_1)$  is a pair of singular vectors corresponding to the singular value  $\sigma_1$ .

We have also

$$\mathbf{u}_1^HA^H\mathbf{v}_1 = \frac{1}{\sigma_1}\mathbf{u}_1^HAA^H\mathbf{u}_1 = \sigma_1.$$



Define  $U = (\mathbf{u}_1 \ U_1)$  and  $V = (\mathbf{v}_1 \ V_1)$  as unitary matrices having  $\mathbf{u}_1$  and  $\mathbf{v}_1$  as their first columns, respectively. Then,

$$\begin{aligned} U^H A^H V &= \begin{pmatrix} \mathbf{u}_1^H \\ U_1^H \end{pmatrix} A^H (\mathbf{v}_1 \ V_1) \\ &= \begin{pmatrix} \mathbf{u}_1^H A^H \\ U_1^H A^H \end{pmatrix} (\mathbf{v}_1 \ V_1) \\ &= \begin{pmatrix} \mathbf{u}_1^H A^H \mathbf{v}_1 & \mathbf{u}_1^H A^H V_1 \\ U_1^H A^H \mathbf{v}_1 & U_1^H A^H V_1 \end{pmatrix} \end{aligned}$$



Since  $U$  is a unitary matrix every column of  $U_1$  is orthogonal to  $\mathbf{u}_1$ . Therefore,

$$U_1^H A \mathbf{v}_1 = \frac{1}{\sigma_1} U_1^H A A^H \mathbf{u}_1 = \sigma_1 U_1^H \mathbf{u}_1 = \mathbf{0},$$

and, similarly,

$$\mathbf{u}_1^H A^H V_1 = \sigma_1 \mathbf{v}_1^H V_1 = \mathbf{0}',$$

because  $\mathbf{v}_1$  is orthogonal on all columns of  $V_1$ . Thus,

$$U^H A V = \begin{pmatrix} \sigma_1 & \mathbf{0}' \\ \mathbf{0} & U_1^H A V_1 \end{pmatrix}.$$

The matrix  $U_1^H A V_1$  has fewer rows and columns than  $U^H A V$ , so we can apply the inductive hypothesis to  $B = U_1^H A V_1$ . Therefore, by the inductive hypothesis,  $B$  can be written as  $B = X D Y^H$ , where  $X$  and  $Y$  are unitary matrices and  $D$  is a diagonal matrix.



This allows us to write

$$U^H A V = \begin{pmatrix} \sigma_1 & \mathbf{0}' \\ \mathbf{0} & X D Y^H \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & X \end{pmatrix} \begin{pmatrix} \sigma_1 & \mathbf{0}' \\ \mathbf{0} & D \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Y^H \end{pmatrix}.$$

Since the matrices

$$\begin{pmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & X \end{pmatrix} \text{ and } \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Y^H \end{pmatrix}$$

are unitary, we obtain the desired conclusion.



If  $A \in \mathbb{C}^{n \times n}$  is an invertible matrix and  $\sigma$  is a singular value of  $A$ , then  $\frac{1}{\sigma}$  is a singular value of the matrix  $A^{-1}$ .



# Reminder

Let

$$\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \in \mathbb{C}^n - \{\mathbf{0}\}.$$

Recall that the matrix  $\mathbf{a}\mathbf{a}^H$  has the same non-zero eigenvalues as the matrix  $\mathbf{a}^H\mathbf{a}$ . Since

$$\mathbf{a}^H\mathbf{a} = \bar{a}_1 a_1 + \cdots + \bar{a}_n a_n \in \mathbb{C}$$

is a scalar, its unique eigenvalue is  $|a_1|^2 + \cdots + |a_n|^2 = \|\mathbf{a}\|^2$ , hence the matrix  $\mathbf{a}\mathbf{a}^H$  has a unique eigenvalue is  $\|\mathbf{a}\|^2$ .



## Example

Let

$$\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

be a non-zero vector in  $\mathbb{C}^n$ , which can also be regarded as a matrix in  $\mathbb{C}^{n \times 1}$ . The square of a singular value of  $A$  is an eigenvalue of the matrix

$$\mathbf{a}^H \mathbf{a} = \begin{pmatrix} \bar{a}_1 a_1 & \cdots & \bar{a}_n a_1 \\ \bar{a}_1 a_2 & \cdots & \bar{a}_n a_2 \\ \vdots & \cdots & \vdots \\ \bar{a}_1 a_n & \cdots & \bar{a}_n a_n \end{pmatrix}$$

and we have seen that the unique non-zero eigenvalue of this matrix is  $\|\mathbf{a}\|_2^2$ . Thus, the unique singular value of  $\mathbf{a}$  is  $\|\mathbf{a}\|_2$ .

## Example

Let  $A \in \mathbb{R}^{3 \times 2}$  be the matrix

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

The matrices  $AA^H$  and  $A^H A$  are given by:

$$AA^H = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{pmatrix} \text{ and } A^H A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

The eigenvalues of  $A^H A$  are the roots of the polynomial  $\lambda^2 - 4\lambda + 3$ , and therefore, they are  $\lambda_1 = 3$  and  $\lambda_2 = 1$ . The eigenvalues of  $AA^H$  are 3, 1 and 0.



## Example

Unit eigenvectors of  $A^H A$  that correspond to 3 and 1 are

$$\mathbf{v}_1 = \alpha_1 \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix} \text{ and } \mathbf{v}_2 = \alpha_2 \begin{pmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{pmatrix},$$

respectively, where  $\alpha_i \in \{-1, 1\}$  for  $i = 1, 2$ .



## Example

Unit eigenvectors of  $AA^H$  that correspond to 3, 1 and 0 are:

$$\mathbf{u}_1 = \beta_1 \begin{pmatrix} \frac{\sqrt{6}}{6} \\ \frac{\sqrt{6}}{3} \\ \frac{\sqrt{6}}{6} \end{pmatrix}, \mathbf{u}_2 = \beta_2 \begin{pmatrix} \frac{\sqrt{2}}{2} \\ 0 \\ -\frac{\sqrt{2}}{2} \end{pmatrix}, \mathbf{u}_3 = \beta_3 \begin{pmatrix} \frac{\sqrt{3}}{3} \\ -\frac{\sqrt{3}}{3} \\ \frac{\sqrt{3}}{3} \end{pmatrix},$$

respectively, where  $\beta_i \in \{-1, 1\}$  for  $i = 1, 2, 3$ .

## Example

The choice of the columns of the matrices  $U$  and  $V$  must be done such that for a pair of eigenvectors  $(u, v)$  that correspond to a singular values  $\sigma$  we have  $\mathbf{v} = \frac{1}{\sigma} A^H \mathbf{u}$  or, equivalently,  $\mathbf{u} = \frac{1}{\sigma} A \mathbf{v}$ . For instance, if we choose  $\alpha_1 = \alpha_2 = 1$ , then

$$\mathbf{v}_1 = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{pmatrix},$$

$\mathbf{u}_1 = \frac{1}{\sqrt{3}} A \mathbf{v}_1$  and  $\mathbf{u}_2 = A \mathbf{v}_2$ , that is,

$$\mathbf{u}_1 = \begin{pmatrix} \frac{\sqrt{6}}{6} \\ \frac{\sqrt{6}}{3} \\ \frac{\sqrt{6}}{6} \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} -\frac{\sqrt{2}}{2} \\ 0 \\ \frac{\sqrt{2}}{2} \end{pmatrix},$$

which means that  $\beta_1 = 1$  and  $\beta_2 = -1$ ; the value of  $\beta_3$  that corresponds to the eigenvalue of 0 can be chosen arbitrarily.

## Example

Thus, an SVD of  $A$  is:

$$A = \begin{pmatrix} \frac{\sqrt{6}}{6} & -\frac{\sqrt{2}}{2} & \frac{\sqrt{3}}{3} \\ \frac{\sqrt{6}}{3} & 0 & -\frac{\sqrt{3}}{3} \\ \frac{\sqrt{6}}{6} & \frac{\sqrt{2}}{2} & \frac{\sqrt{3}}{3} \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{pmatrix}.$$

A variant of the SVD Decomposition Theorem is given next.

## Corollary

**(The Thin SVD Decomposition Corollary)** Let  $A \in \mathbb{C}^{m \times n}$  be a matrix having non-zero singular values  $\sigma_1, \sigma_2, \dots, \sigma_r$ , where  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$  and  $r \leq \min\{m, n\}$ . Then,  $A$  can be factored as  $A = UDV^H$ , where  $U \in \mathbb{C}^{m \times r}$  and  $V \in \mathbb{C}^{n \times r}$  are matrices having orthonormal sets of columns and  $D$  is the diagonal matrix

$$D = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{pmatrix}.$$

## Proof.

The statement is an immediate consequence of the theorem. □

The decomposition described above is known as a *thin SVD decomposition* of the matrix  $A$ .

### Example

The thin SVD decomposition of the matrix  $A$

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

is

$$A = \begin{pmatrix} \frac{\sqrt{6}}{6} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{6}}{3} & 0 \\ \frac{\sqrt{6}}{6} & \frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{pmatrix}.$$



Since  $U$  and  $V$  in the thin SVD have orthonormal columns it is easy to see that

$$U^H U = V^H V = I_p. \quad (2)$$



## Lemma

Let  $D \in \mathbb{R}^{n \times n}$  be a diagonal matrix, where  $D = \text{diag}(\sigma_1, \dots, \sigma_r)$  and  $\sigma_1 \geq \dots \geq \sigma_r$ . Then, we have  $\|D\|_2 = \sigma_1$ , and  $\|D\|_F = \sqrt{\sum_{i=1}^r \sigma_i^2}$ .

## Proof.

By the definition of  $\|D\|_2$  we have:

$$\|D\|_2 = \max\{\|D\mathbf{x}\|_2 \mid \|\mathbf{x}\| = 1\} = \max\left\{\sqrt{\sum_{i=1}^r \sigma_i^2 |x_i|^2} \mid \sum_{i=1}^n |x_i|^2 = 1\right\}.$$

Since  $\sum_{i=1}^n |x_i|^2 = 1$ , we have:  $\sum_{i=1}^r \sigma_i^2 |x_i|^2 \leq \sigma_1^2 (\sum_{i=1}^r |x_i|^2) \leq \sigma_1^2$ .  
It follows that

$$\max\left\{\sqrt{\sum_{i=1}^r \sigma_i^2 |x_i|^2} \mid \sum_{i=1}^n |x_i|^2 = 1\right\} = \sigma_1.$$

The second part is immediate.

## Theorem

Let  $A \in \mathbb{C}^{m \times n}$  be a matrix whose singular values are  $\sigma_1 \geq \dots \geq \sigma_r$ . Then  $\|A\|_2 = \sigma_1$ , and  $\|A\|_F = \sqrt{\sum_{i=1}^r \sigma_i^2}$ .

## Proof.

Suppose that the SVD of  $A$  is  $A = UDV^H$ , where  $U$  and  $V$  are unitary matrices. Then, by previous results, we have:

$$\begin{aligned}\|A\|_2 &= \|UDV^H\|_2 = \|D\|_2 = \sigma_1, \\ \|A\|_F &= \|UDV^H\|_F = \|D\|_F = \sqrt{\sum_{i=1}^r \sigma_i^2}.\end{aligned}$$



## Corollary

If  $A \in \mathbb{C}^{m \times n}$  is a matrix, then  $\|A\|_2 \leq \|A\|_F \leq \sqrt{n} \|A\|_2$ .

## Proof.

Suppose that  $\sigma_1(A)$  is the largest of the singular values of  $A$ . Then, since  $\|A\|_F = \sqrt{\sum_{i=1}^r \sigma_i^2}$ , we have

$$\sigma_1(A) \leq \|A\|_F \leq \sqrt{n \max_i \sigma_i(A)^2} = \sigma_1(A) \sqrt{n},$$

which is desired double inequality. □



## Theorem

Let  $A \in \mathbb{C}^{n \times n}$  be an invertible matrix. If the singular values of  $A$  are  $\sigma_1 \geq \dots \geq \sigma_n > 0$ , then

$$\text{cond}(A) = \frac{\sigma_1}{\sigma_n}.$$

## Proof.

We have shown that  $\|A\|_2 = \sigma_1$ . Since the singular values of  $A^{-1}$  are

$$\frac{1}{\sigma_n} \geq \dots \geq \frac{1}{\sigma_1}$$

it follows that  $\|A^{-1}\|_2 = \frac{1}{\sigma_n}$ . The desired equality follows immediately.  $\square$



## Corollary

Let  $A \in \mathbb{C}^{n \times n}$  be an invertible matrix. We have  $\text{cond}(A^H A) = (\text{cond}(A))^2$ .

## Proof.

Let  $\sigma$  be a singular value of  $A$  and let  $\mathbf{u}, \mathbf{v}$  be two left and right singular vectors corresponding to  $\sigma$ , respectively. We have

$$A\mathbf{v} = \sigma\mathbf{u} \text{ and } A^H\mathbf{u} = \sigma\mathbf{v}.$$

This implies  $A^H A\mathbf{v} = \sigma A^H\mathbf{u} = \sigma^2\mathbf{v}$ , which shows that the singular values of the matrix  $A^H A$  are the squares of the singular values of  $A$ , which produces the desired conclusion. □



Let  $A = UDV^H$  be an SVD of  $A$ . If we write  $U$  and  $V$  using their columns as  $U = (\mathbf{u}_1 \cdots \mathbf{u}_m)$  and  $V = (\mathbf{v}_1 \cdots \mathbf{v}_n)$ , then  $A$  can be written as:

$$\begin{aligned}
 A &= UDV^H \\
 &= (\mathbf{u}_1 \cdots \mathbf{u}_n) \begin{pmatrix} \sigma_1 & 0 & \cdots & \cdots & 0 \\ 0 & \sigma_2 & \cdots & \cdots & 0 \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ 0 & 0 & \cdots & \sigma_r & 0 \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{v}_1^H \\ \vdots \\ \mathbf{v}_m^H \end{pmatrix} \\
 &= (\mathbf{u}_1 \cdots \mathbf{u}_m) \begin{pmatrix} \sigma_1 \mathbf{v}_1^H \\ \vdots \\ \sigma_r \mathbf{v}_p^H \end{pmatrix} \\
 &= \sigma_1 \mathbf{u}_1 \mathbf{v}_1^H + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_p^H.
 \end{aligned}$$



Since  $\mathbf{u}_i \in \mathbb{C}^m$  and  $\mathbf{v}_i \in \mathbb{C}^n$ , each of the matrices  $\mathbf{u}_i \mathbf{v}_i^H$  is a  $m \times n$  matrix of rank 1. Thus, the SVD yields an expression of  $A$  as a sum of  $r$  matrices of rank 1, where  $r$  is the number of non-zero singular values of  $A$ .



## Theorem

The rank-1 matrices of the form  $\mathbf{u}_i \mathbf{v}_i^H$ , where  $1 \leq i \leq r$  are pairwise orthogonal. Moreover,  $\|\mathbf{u}_i \mathbf{v}_i^H\|_F = 1$  for  $1 \leq i \leq r$ .

## Proof.

For  $i \neq j$  and  $1 \leq i, j \leq r$  we have:

$$\text{trace}(\mathbf{u}_i \mathbf{v}_i^H (\mathbf{u}_j \mathbf{v}_j^H)^H) = \text{trace}(\mathbf{u}_i \mathbf{v}_i^H \mathbf{v}_j \mathbf{u}_j^H) = 0,$$

because the vectors  $\mathbf{v}_i$  and  $\mathbf{v}_j$  are orthogonal. Thus,  $(\mathbf{u}_i \mathbf{v}_i^H, \mathbf{u}_j \mathbf{v}_j^H) = 0$ . We have

$$\begin{aligned} \|\mathbf{u}_i \mathbf{v}_i^H\|_F^2 &= \text{trace}((\mathbf{u}_i \mathbf{v}_i^H)^H \mathbf{u}_i \mathbf{v}_i^H) \\ &= \text{trace}(\mathbf{v}_i \mathbf{u}_i^H \mathbf{u}_i \mathbf{v}_i^H) = 1, \end{aligned}$$

because the matrices  $U$  and  $V$  are unitary. □

## Theorem

*Let  $A \in \mathbb{C}^{m \times n}$  be a matrix that has the singular value decomposition  $A = UDV^H$ . If  $\text{rank}(A) = r$ , then the first  $r$  columns of  $U$  form an orthonormal basis for  $\text{range}(A)$ , and the last  $n - r$  columns of  $V$  constitute an orthonormal basis for  $\text{null}(A)$ .*



# Proof

Since both  $U$  and  $V$  are unitary matrices, it is clear that  $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ , the set of the first  $r$  columns of  $U$ , and  $\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$ , the set of the last  $n - r$  columns of  $V$ , are linearly independent sets. Thus, we only need to show that  $\langle \mathbf{u}_1, \dots, \mathbf{u}_r \rangle = \text{range}(A)$  and  $\langle \mathbf{v}_{r+1}, \dots, \mathbf{v}_n \rangle = \text{null}(A)$ .

We have

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^H + \dots + \sigma_r \mathbf{u}_r \mathbf{v}_r^H.$$

If  $\mathbf{t} \in \text{range}(A)$ , then  $\mathbf{t} = A\mathbf{s}$  for some  $\mathbf{s} \in \mathbb{C}^n$ . Therefore,  $\mathbf{t} = \sigma_1 \mathbf{u}_1 (\mathbf{v}_1^H \mathbf{s}) + \dots + \sigma_r \mathbf{u}_r (\mathbf{v}_r^H \mathbf{s})$ , and, since the every product  $\mathbf{v}_j^H \mathbf{s}$  is a scalar for  $1 \leq j \leq r$ , it follows that  $\mathbf{t} \in \langle \mathbf{u}_1, \dots, \mathbf{u}_r \rangle$ , so  $\text{range}(A) \subseteq \langle \mathbf{u}_1, \dots, \mathbf{u}_r \rangle$ .



## Proof cont'd

To prove the reverse inclusion note that

$$A \left( \frac{1}{\sigma_i} \mathbf{v}_i \right) = \mathbf{u}_i,$$

for  $1 \leq i \leq r$ , due to the orthogonality of the columns of  $V$ . Thus,  $\langle \mathbf{u}_1, \dots, \mathbf{u}_r \rangle = \text{range}(A)$ .

Thus,  $A\mathbf{v}_j = \mathbf{0}$  for  $r+1 \leq j \leq n$ , so  $\langle \mathbf{v}_{r+1}, \dots, \mathbf{v}_n \rangle \subseteq \text{null}(A)$ . Conversely, suppose that  $A\mathbf{r} = \mathbf{0}$ . Since the columns of  $V$  form a basis of  $\mathbb{C}^n$  we have  $\mathbf{r} = a_1\mathbf{v}_1 + \dots + a_n\mathbf{v}_n$ , so  $A\mathbf{r} = a_1A\mathbf{v}_1 + \dots + a_r\mathbf{v}_r = \mathbf{0}$ . The linear independence of  $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$  implies  $a_1 = \dots = a_r = 0$ , so  $\mathbf{r} = a_{r+1}\mathbf{v}_{r+1} + \dots + a_n\mathbf{v}_n$ , which shows that  $\text{null}(A) \subseteq \langle \mathbf{v}_{r+1}, \dots, \mathbf{v}_n \rangle$ . Thus,  $\text{null}(A) = \langle \mathbf{v}_{r+1}, \dots, \mathbf{v}_n \rangle$ .



## Corollary

*Let  $A \in \mathbb{C}^{m \times n}$  be a matrix that has the singular value decomposition  $A = UDV^H$ . If  $\text{rank}(A) = r$ , then the first  $r$  transposed columns of  $V$  form an orthonormal basis for the subspace of  $\mathbb{R}^n$  generated by the rows of  $A$ .*

## Proof.

This statement follows immediately from the previous theorem applied to  $A^H$ . □



The singular value decomposition of a matrix can be computed in MATLAB using the function `svd`. To illustrate its use in the simplest form consider the matrix

$$A = \begin{pmatrix} \frac{1}{3} & \frac{4}{3} & \frac{5}{3} \\ \frac{11}{5} & \frac{4}{5} & 3 \\ \frac{1}{7} & \frac{4}{7} & \frac{5}{7} \\ \frac{1}{9} & \frac{4}{9} & \frac{5}{9} \end{pmatrix},$$

defined in MATLAB by

```
A = [1/3 4/3 5/3 ; 11/5 4/5 3 ; 1/7 4/7 5/7 ; 1/9 4/9 5/9]
```

A call the `svd(A)` function yields a vector containing the singular values of  $A$ , as in

```
>> svd(A)
ans =
    4.3674
    1.2034
    0.0000
```



It is interesting to note that  $\text{rank}(A) = 2$  since the last column of  $A$  is the sum of the first two columns. Thus, we would expect to see two non-zero singular values.

To compute  $\|A\|_2$ , which equals the largest singular value of  $A$  we can use  $\max(\text{svd}(A))$ .



Another variant of the `svd` function,  $[U, S, V] = \text{svd}(A)$  yields a diagonal matrix  $S$ , of the same format as  $A$  and with nonnegative diagonal elements in decreasing order, and unitary matrices  $U$  and  $V$  so that  $A = USV^H$ . For the matrix  $A$  shown above we obtain:



## Example

```
>> [U,S,V] = svd(A)
```

```
U =
```

-0.4487	0.7557	0.4769	0.0161
-0.8599	-0.5105	0.0000	-0.0000
-0.1923	0.3239	-0.6726	-0.6370
-0.1496	0.2519	-0.5658	0.7707

```
S =
```

4.3674	0	0
0	1.2034	0
0	0	0.0000
0	0	0

```
V =
```

-0.4775	-0.6623	0.5774
-0.3349	0.7447	0.5774
-0.8123	0.0823	-0.5774

The “economical form” of the `svd` function is

`[U,S,V] = svd(A,'econ')`

If  $A \in \mathbb{R}^{m \times n}$  and  $m > n$ , only the first  $n$  columns of  $U$  are computed and  $S \in \mathbb{R}^{n \times n}$ . If  $m < n$  only the first  $m$  columns of  $V$  are computed.

### Example

Starting from the matrix

$$A = \begin{pmatrix} 18 & 8 & 20 \\ -4 & 20 & 1 \\ 25 & 8 & 27 \\ 9 & 4 & 10 \end{pmatrix} \in \mathbb{R}^{4 \times 3}$$

a call the economical variant of the `svd` function yields



## Example

```
>> [U,D,V] = svd(A,'econ')
```

```
U =
```

-0.5717	-0.0211	0.8095
-0.0721	-0.9933	-0.0669
-0.7656	0.1133	-0.4685
-0.2859	-0.0105	-0.3474

```
S =
```

49.0923	0	0
0	20.2471	0
0	0	0.0000

```
V =
```

-0.6461	0.3127	0.6963
-0.2706	-0.9468	0.1741
-0.7137	0.0760	-0.6963

The function `svapprox` given below computes the successive approximations  $B(k) = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^H$  of a matrix  $A \in \mathbb{R}^{m \times n}$  having the SVD  $A = UDV^H$  and produces a three-dimensional array  $C \in \mathbb{R}^{m \times n \times r}$ , where  $r$  is the numerical rank of  $A$  and  $C(:, :, k) = B(k)$  for  $1 \leq k \leq r$ .



```

function [C] = svapprox(A)
%SVAPPROX computes the successive approximations
%      of A using the singular component decomposition.
%      The number of approximations equals the
%      numerical rank of A.

% determine the format of A and its numerical rank
[m, n] = size(A);
r = rank(A,10^(-5));
% compute the SVD of A
[U,D,V] = svd(A);
C = zeros(m,n,r);
C(:, :, 1) = D(1,1) * U(:,1) * (V(:,1))';
for k=2:r
    C(:, :, k) = D(k,k) * U(:,k) * (V(:,k))' + C(:, :, k-1);
end;

```



In the next figure we have an image of the digit 4 created from a pgm file that contains the representation of this digit.

The numerical rank of the matrix  $A$  introduced in the example mentioned above is 8. Therefore, the array  $C$  computed by  $C = \text{svapprox}(A)$  consists of 8 matrices. To represent these matrices in the pgm format we cast the components of  $C$  to integers of the type `uint8` using  $D = \min(16, \text{uint8}(C))$ . Thus,  $D(:, :, j)$  contains the rounded  $j^{\text{th}}$  approximation of  $A$ .



The images for the first four approximations are represented next:



(a)



(b)



(c)



(d)

Successive Approximations of  $A$ .

Note that the digit four is easily recognizable beginning with the second approximation.

