# Singular Value Decomsition

#### Hakbae Lee

The Department of Applied Statistics, Yonsei University

### **Matrices**

#### **Definition of Matrix**

Define an  $m \times n$  matrix **A** 

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \dots & a_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} = (a_{ij}) = \begin{pmatrix} \mathbf{R}_1 \\ \mathbf{R}_2 \\ \vdots \\ \mathbf{R}_m \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{C}_1 & \mathbf{C}_2 & \dots & \mathbf{C}_n \end{pmatrix}$$

where

$$\mathbf{R}_i = i$$
-th 1 × n row vector,  $i = 1, ..., m$ 

$$\mathbf{C}_i = j$$
-th  $m \times 1$  column vector,  $j = 1, \dots, n$ 

Each symmetric matrix  $\mathbf{A}(p \times p)$  can be written as

$$\mathbf{A} = \mathbf{\Gamma} \mathbf{\Lambda} \mathbf{\Gamma}^T = \sum_{j=1}^{\rho} \lambda_j \gamma_j \gamma_j^T$$

where

$$\boldsymbol{\Lambda} = \operatorname{diag}\{\lambda_1, \dots, \lambda_p\} = \left( \begin{array}{cccc} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \lambda_p \end{array} \right) : \boldsymbol{p} \times \boldsymbol{p}$$

where

$$\Gamma = (\gamma_1, \gamma_2, \dots, \gamma_p) : p \times p$$

where is an orthogonal matrix consisting of the eigenvectors  $\gamma_j$  of **A**.

Let **A** be  $p \times p$  symmetric matrix of rank r,  $(r \leq p)$ . Then there exists  $p \times p$  orthogonal matrix  $\Gamma$  so that  $\Gamma^T \Gamma = \mathbf{I}_p$  and

$$\mathbf{A} = \mathbf{\Gamma} \mathbf{\Lambda} \mathbf{\Gamma}^T = \mathbf{\Gamma} \begin{pmatrix} \mathbf{\Lambda}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{\Gamma}^T = \mathbf{\Gamma}_1 \mathbf{\Lambda}_1 \mathbf{\Gamma}_1^T$$

where letting  $\delta_i = i$ -th eigenvalue, i = 1, ..., r

$$\Gamma = (\Gamma_1, \Gamma_0) : \Gamma_1 : p \times r, \Gamma_0 : p \times (p-r)$$

$$\mathbf{\Lambda}_1 = \operatorname{diag}\{\lambda_1, \dots, \lambda_r\} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \lambda_r \end{pmatrix} : r \times r.$$

$$\begin{split} \Gamma_1^T \Gamma_1 &= \mathbf{I}_r, \ \Gamma_1^T \Gamma_0 = \mathbf{0}, \ \Gamma_1^T \Gamma_0 = \mathbf{0} \text{ and} \\ \mathbf{A}^2 &= \mathbf{A}^T \mathbf{A} = \mathbf{A} \mathbf{A}^T \\ &= (\Gamma_1 \boldsymbol{\Lambda}_1 \boldsymbol{\Gamma}_1^T)^T \boldsymbol{\Gamma}_1 \boldsymbol{\Lambda}_1 \boldsymbol{\Gamma}_1^T = \boldsymbol{\Gamma}_1 \boldsymbol{\Lambda}_1 \boldsymbol{\Gamma}_1^T \boldsymbol{\Gamma}_1 \boldsymbol{\Lambda}_1 \boldsymbol{\Gamma}_1^T \\ &= \boldsymbol{\Gamma}_1 \boldsymbol{\Lambda}_1^2 \boldsymbol{\Gamma}_1^T. \end{split}$$

Let  $\gamma_i$  be *i*-th  $p \times 1$  column vector of  $\Gamma$ . Then

$$\gamma_i^T \gamma_j = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j. \end{cases}$$

#### Thus

$$\mathbf{A} = \mathbf{\Gamma}_{1} \mathbf{\Lambda}_{1} \mathbf{\Gamma}_{1}^{T} = \sum_{i=1}^{r} \lambda_{i} \gamma_{i} \gamma_{i}^{T}$$

$$\mathbf{A}^{T} \mathbf{A} = \mathbf{\Gamma}_{1} \mathbf{\Lambda}_{1}^{2} \mathbf{\Gamma}_{1}^{T} = \sum_{i=1}^{r} \lambda_{i}^{2} \gamma_{i} \gamma_{i}^{T}$$

$$\mathbf{A} \mathbf{A}^{T} = \mathbf{\Gamma}_{1} \mathbf{\Lambda}_{1}^{2} \mathbf{\Gamma}_{1}^{T} = \sum_{i=1}^{r} \lambda_{i}^{2} \gamma_{i} \gamma_{i}^{T}$$

$$\gamma_{k}^{T} \mathbf{A} = \lambda_{k} \gamma_{k}^{T} \gamma_{k} \gamma_{k}^{T} = \lambda_{k} \gamma_{k}^{T}$$

$$\mathbf{A} \gamma_{k} = \lambda_{k} \gamma_{k} \gamma_{k}^{T} \gamma_{k} = \lambda_{k} \gamma_{k}.$$

#### Remark 1

Let  $\Gamma$  be an orthogonal matrix so that  $\Gamma^T\Gamma=I$ . Then

$$det(\Gamma) = |\Gamma| = 1.$$

#### Remark 2

Let **A** be  $p \times p$  symmetric matrix of full rank. Then, by the Spectral Decomposition,

$$\det(\mathbf{A}) = |\mathbf{A}| = |\mathbf{\Gamma}||\mathbf{\Lambda}||\mathbf{\Gamma}^T|$$

$$= |\mathbf{\Lambda}| = \begin{vmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \lambda_p \end{vmatrix} = \prod_{i=1}^p \lambda_i$$

#### Remark 3

Let **A** be  $p \times p$  symmetric matrix of full rank. Then, by the Spectral Decomposition,

$$\mathbf{A}^{\alpha} = (\mathbf{\Gamma} \mathbf{\Lambda} \mathbf{\Gamma}^{T})(\mathbf{\Gamma} \mathbf{\Lambda} \mathbf{\Gamma}^{T}) \dots (\mathbf{\Gamma} \mathbf{\Lambda} \mathbf{\Gamma}^{T}) = \mathbf{\Gamma} \mathbf{\Lambda}^{\alpha} \mathbf{\Gamma}^{T}$$
 for some  $\alpha \in \mathbb{R}$  In particular, a covariance matrix  $\mathbf{\Sigma}$  can be written by

$$\Sigma - \Gamma \Lambda \Gamma^T - \sum_{i}^{T} \lambda_{i} \gamma_{i} \gamma_{i}^{T}$$

$$\Sigma = \Gamma \Lambda \Gamma^T = \sum_{i=1}^r \lambda_i \gamma_i \gamma_i^T$$

then

$$\Sigma^{-1} = \Gamma \Lambda^{-1} \Gamma^{T} = \sum_{i=1}^{r} \lambda_{i}^{-1} \gamma_{i} \gamma_{i}^{T}$$

$$\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{\Gamma} \boldsymbol{\Lambda}^{-1/2} \boldsymbol{\Gamma}^T = \sum_{i=1}^r \lambda_i^{-1/2} \gamma_i \gamma_i^T$$

## Singular value Decomposition: General-version

 Any arbitrary matrix A(n × p) with rank r can be decomposed as

$$\mathbf{A} = \mathbf{\Gamma} \mathbf{\Lambda} \mathbf{\Delta}^T = \sum_{j=1}^r \lambda_j \gamma_j \delta_j^T$$

where  $\Gamma(n \times r)$  and  $\Delta(p \times r)$ .

- Both Γ and Δ are column orthogonal, i.e.,  $\Gamma^T \Gamma = \Delta^T \Delta = I_r$  and  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_r), \ \lambda_i > 0$ .
- The values  $\lambda_1, \dots, \lambda_r$  are the non-zero eigenvalues of the matrices  $\mathbf{A}\mathbf{A}^T$  and  $\mathbf{A}^T\mathbf{A}$ .
- Γ and Δ consist of the corresponding r eigenvectors of theses matrices.

# Singular Value Decomposition: General-version

Thus

$$\mathbf{A} = \mathbf{\Gamma} \mathbf{\Lambda} \mathbf{\Delta}^T = \sum_{j=1}^r \lambda_j \gamma_j \delta_j^T$$

$$\mathbf{A}^T \mathbf{A} = \mathbf{\Delta} \mathbf{\Lambda}^2 \mathbf{\Delta}^T = \sum_{j=1}^r \lambda_j^2 \delta_j \delta_j^T$$

$$\mathbf{A} \mathbf{A}^T = \mathbf{\Gamma} \mathbf{\Lambda}^2 \mathbf{\Gamma}^T = \sum_{j=1}^r \lambda_j^2 \gamma_j \gamma_j^T$$

$$\gamma_k^T \mathbf{A} = \gamma_k^T \mathbf{\Gamma} \mathbf{\Lambda} \mathbf{\Delta}^T = \lambda_k \delta_k^T$$

$$\mathbf{A} \delta_k = \mathbf{\Gamma} \mathbf{\Lambda} \mathbf{\Delta}^T \delta_k = \lambda_k \gamma_k.$$

# Singular Value Decomposition: General-version

#### G-inverse (Generalized inverse) matrix A-

$$\mathbf{A} = \mathbf{\Gamma} \mathbf{\Lambda} \mathbf{\Delta}^T = \sum_{j=1}^r \lambda_j \gamma_j \delta_j^T$$

Define

$$\mathbf{A}^{-} = \mathbf{\Delta} \mathbf{\Lambda}^{-1} \mathbf{\Gamma}^{T} = \sum_{j=1}^{r} \lambda_{j}^{-1} \delta_{j} \gamma_{j}^{T}$$

Then

$$\mathbf{A}\mathbf{A}^{-}\mathbf{A} = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Delta}^{T}\mathbf{\Delta}\mathbf{\Lambda}^{-1}\mathbf{\Gamma}^{T}\mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Delta}^{T} = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Delta}^{T} = \sum_{j=1}^{r} \lambda_{j}\gamma_{j}\delta_{j}^{T} = \mathbf{A}$$

### Singular value Decomposition: Another-version

 Any arbitrary matrix A(n × p) with rank r can be decomposed as

$$\mathbf{A} = \mathbf{\Gamma} \mathbf{\Lambda} \mathbf{\Delta}^T = \sum_{j=1}^r \lambda_j^{1/2} \gamma_j \delta_j^T$$

where  $\Gamma(n \times r)$  and  $\Delta(p \times r)$ .

- Both Γ and Δ are column orthogonal, i.e.,  $\Gamma^T \Gamma = \Delta^T \Delta = I_r$  and  $\Lambda = \operatorname{diag}\left(\lambda_1^{1/2}, \dots, \lambda_r^{1/2}\right), \ \lambda_j^{1/2} > 0.$
- The values  $\lambda_1, \dots, \lambda_r$  are the non-zero eigenvalues of the matrices  $\mathbf{A}\mathbf{A}^T$  and  $\mathbf{A}^T\mathbf{A}$ .
- Γ and Δ consist of the corresponding r eigenvectors of theses matrices.

### **Quadratic Forms**

• A quadratic form Q(x) is defined to be

$$Q(x) = x^T \mathbf{A} x = \sum_{i=1}^{p} \sum_{j=1}^{p} a_{ij} x_i x_j$$

for a symmetric matrix  $\mathbf{A}(p \times p)$  and a vector  $x \in \mathbb{R}^p$ 

$$Q(x) > 0$$
 for all  $x \neq 0$ : Positive definite  $Q(x) \geq 0$  for all  $x \neq 0$ : Positive semidefinite

- A is called positive definite(semidefinite) if the corresponding quadratic form Q(⋅) is positive definite(semidefinite).
- Notation:  $\mathbf{A} > 0 (\geq 0)$

### **Quadratic Forms**

### Proposition 1

If **A** is symmetric and  $Q(x) = x^T \mathbf{A} x$  is the corresponding quadratic form, then there exists a transformation  $y = \mathbf{\Gamma}^T x$  such that

$$Q(x) = x^T \mathbf{A} x = \sum_{i=1}^{\rho} \lambda_i y_i^2$$

where  $\lambda_i$ 's are the eigenvalues of **A**.

### Proposition 2

 $\mathbf{A} > 0$  if and only of all  $\lambda_i > 0$ ,  $i = 1, \dots, p$ 

### Corollary 1

If A > 0, then  $A^{-1}$  exists and |A| > 0.

### Quadratic Forms

### **Proposition 3**

- If **A** and **B** are symmetric and **B** > 0, then the maximum of  $\frac{x^T \mathbf{A} x}{x^T \mathbf{B} x}$  is given by the largest eigenvalues of  $\mathbf{B}^{-1} \mathbf{A}$ .
- More generally,

$$\max \frac{x^T \mathbf{A} x}{x^T \mathbf{B} x} = \lambda_1 \ge \lambda_2 \ge \dots \ge \lambda_p = \min \frac{x^T \mathbf{A} x}{x^T \mathbf{B} x}$$

where  $\lambda_1, \ldots, \lambda_p$  denote the eigenvalues of  $\mathbf{B}^{-1}\mathbf{A}$ .

- The vector which maximizes(minimizes)  $\frac{x^T \mathbf{A}x}{x^T \mathbf{B}x}$  is the eigenvector of  $\mathbf{B}^{-1} \mathbf{A}$  which corresponds to the largest(smallest) eigenvalue of  $\mathbf{B}^{-1} \mathbf{A}$ .
- If  $x^T \mathbf{B} x = 1$ , then

$$\max x^T \mathbf{A} x = \lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_p = \min x^T \mathbf{A} x$$

### **Partitioned Matrices**

#### Note

Let  $\mathbf{A}(n \times p)$  and  $\mathbf{B}(p \times n)$  be any two matrices and  $n \ge p$ . Then

$$egin{array}{c|c} -\lambda \mathbf{I}_n & -\mathbf{A} \\ \mathbf{B} & \mathbf{I}_p \end{array} = (-\lambda)^{n-p} |\mathbf{B}\mathbf{A} - \lambda \mathbf{I}_n| = |\mathbf{A}\mathbf{B} - \lambda \mathbf{I}_n|$$

#### Proposition 4

For  $\mathbf{A}(n \times p)$  and  $\mathbf{B}(p \times n)$ , the non-zero eigenvalues of  $\mathbf{AB}$  and  $\mathbf{BA}$  are the same and have the same multiplicity. If x is an eigenvector of  $\mathbf{AB}$  for an eigenvalues  $\lambda \neq 0$ , then  $y = \mathbf{B}x$  is an eigenvector of  $\mathbf{BA}$ .

#### Corollary 2

For  $\mathbf{A}(n \times p)$ ,  $\mathbf{B}(q \times n)$ ,  $a(p \times 1)$  and  $b(q \times 1)$ ,

$$\text{rank}(\textbf{A}ab\textbf{B}) \leq 1$$

The non-zero eigenvalue, if it exists, equals  $b^T \mathbf{B} \mathbf{A} a$  with eigenvector  $\mathbf{A} a$ 

## **Geometrical Aspects**

#### Note

- x<sub>1</sub>, x<sub>2</sub>,..., x<sub>k</sub> are mutually orthogonal if and only if x<sub>i</sub><sup>T</sup>x<sub>j</sub> = 0 for all i, j.
- In that case,  $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k)$  has rank k, and  $\mathbf{X}^T \mathbf{X}$  is a diagonal matrix with  $\mathbf{x}_i^T \mathbf{x}_i$  in the i-th diagonal position.
- Let's consider bivariate data  $(x_i, y_i)$ , i = 1, ..., n, and let  $\tilde{x}_i = x_i \bar{\mathbf{x}}$  and  $\tilde{y}_i = y_i \bar{\mathbf{y}}$ . Then the correlation between  $\mathbf{x}$  and  $\mathbf{y}$  is

$$\frac{\sum_{i=1}^{n}(x_i-\bar{\mathbf{x}}_i)(y_i-\bar{\mathbf{y}}_i)}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{\mathbf{x}}_i)^2\sum_{i=1}^{n}(y_i-\bar{\mathbf{x}}_i)^2}}=\frac{\tilde{\mathbf{x}}^T\tilde{\mathbf{y}}}{||\tilde{\mathbf{x}}||||\tilde{\mathbf{y}}||}=\cos(\theta)$$

where  $\theta$  is the angle between the deviation vectors  $\tilde{\mathbf{x}}$  and  $\tilde{\mathbf{y}}.$ 

### Geometrical Aspects

#### Rotations

For two dimensions, the clockwise rotation can be expressed:

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$
$$= \mathbf{\Gamma} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \mathbf{\Gamma} \mathbf{x}$$

the counter-clockwise rotation can be expressed

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$
$$= \mathbf{\Gamma}^T \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \mathbf{\Gamma}^T \mathbf{x}$$

#### **Definition**

Consider an  $n \times p$  matrix **X**.

$$\mathcal{C}(\mathbf{X}) = \{x \in \mathbb{R}^n | \exists a \in \mathbb{R}^p \text{ so that } \mathbf{X}a = x\} \subseteq \mathbb{R}^n$$

$$= \text{ the } \mathbf{column(range) space of } \mathbf{X}$$

$$\mathcal{N}(\mathbf{X}) = \{y \in \mathbb{R}^p | \mathbf{X}y = 0\} \subseteq \mathbb{R}^p$$

$$= \text{ the } \mathbf{null space of } \mathbf{X}$$

$$\mathcal{R}(\mathbf{X}) = \{z \in \mathbb{R}^p | \exists b \in \mathbb{R}^n \text{ so that } \mathbf{X}^T b = z\} \subseteq \mathbb{R}^p$$

$$= \text{ the } \mathbf{row space of } \mathbf{X}$$

$$= \mathcal{C}(\mathbf{X}^T) = \text{the } \mathbf{column space of } \mathbf{X}^T$$

Consider an  $n \times p$  matrix **X** with rank(**X**) = r

### Spaces by Singular Value Decomposition: General-version

$$\mathbf{X} = \mathbf{\Gamma} \mathbf{\Lambda} \mathbf{\Delta}^T = \sum_{j=1}^r \lambda_j \gamma_j \delta_j^T$$

$$\mathcal{C}(\mathbf{X}) = \{\gamma_1, \gamma_2, \dots, \gamma_r\}$$

$$\mathcal{N}(\mathbf{X}) = \{\delta_{r+1}, \delta_{r+2}, \dots, \delta_p\}$$

$$\mathcal{R}(\mathbf{X}) = \{\delta_1, \delta_2, \dots, \delta_r\}$$

### Note 1: Let X be $n \times p$ matrix. Then

$$\mathcal{N}(\mathbf{X}) = \mathcal{C}(\mathbf{X}^T)^{\perp} = \mathcal{R}(\mathbf{X})^{\perp}$$
  
 $\mathcal{N}(\mathbf{X})^{\perp} = \mathcal{C}(\mathbf{X}^T) = \mathcal{R}(\mathbf{X})$ 

#### Note 2: Let X be $n \times p$ matrix. Then

$$C(\mathbf{X}^T\mathbf{X}) = C(\mathbf{X}^T) = \mathcal{R}(\mathbf{X})$$

### Note 3: Let X be $n \times p$ matrix. Then

- $\dim(\mathcal{C}(\mathbf{X})) = \dim(\mathcal{R}(\mathbf{X})) = \operatorname{rank}(\mathbf{X}) = \operatorname{rank}(\mathbf{X}^T) = \operatorname{rank}(\mathbf{X}^T\mathbf{X}) = r \leq \min(n, p)$
- X<sup>T</sup>X has full rank (is nonsingular) if and only if X has full column rank (X has linearly independent columns).

#### Example: arbitrary $3 \times 4$ matrix A

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 5 & 1 \\ 2 & 2 & 10 & 6 \\ 3 & 4 & 15 & 1 \end{pmatrix} = \sum_{i=1}^{3} \lambda_{i} \gamma_{i} \delta_{i}^{T} = \lambda_{1} \gamma_{1} \delta_{1}^{T} + \lambda_{2} \gamma_{2} \delta_{2}^{T} + \lambda_{3} \gamma_{3} \delta_{3}^{T}$$

$$= 20.15 \begin{pmatrix} -0.274 \\ -0.568 \\ -0.776 \end{pmatrix} (-0.186, -0.238, -0.927, -0.221)$$

$$+4.40 \begin{pmatrix} -0.072 \\ 0.817 \\ -0.572 \end{pmatrix} (-0.035, -0.182, -0.177, 0.967)$$

$$+0.69 \begin{pmatrix} -0.959 \\ 0.101 \\ 0.265 \end{pmatrix} (0.053, -0.954, 0.265, -0.129)$$

### Example: arbitrary $3 \times 4$ matrix A

The eigenvalues and eigenvectors by Singular Value Decomposition are given by

$$\lambda_{1} = 20.15 \qquad \lambda_{2} = 4.40 \qquad \lambda_{3} = 0.69$$

$$\gamma_{1} = \begin{pmatrix} -0.274 \\ -0.568 \\ -0.776 \end{pmatrix} \qquad \gamma_{2} = \begin{pmatrix} -0.072 \\ 0.817 \\ -0.572 \end{pmatrix} \qquad \gamma_{3} = \begin{pmatrix} -0.959 \\ 0.101 \\ 0.265 \end{pmatrix}$$

$$\delta_{1} = \begin{pmatrix} -0.186 \\ -0.238 \\ -0.927 \\ -0.221 \end{pmatrix} \qquad \delta_{2} = \begin{pmatrix} -0.035 \\ -0.182 \\ -0.177 \\ 0.967 \end{pmatrix} \qquad \delta_{3} = \begin{pmatrix} 0.053 \\ -0.954 \\ 0.265 \\ -0.129 \end{pmatrix}$$