

# Linear Systems

## Singular Value Decomposition

Sivakumar Balasubramanian

Department of Bioengineering  
Christian Medical College, Bagayam  
Vellore 632002

## References

- ▶ G Strang, Linear Algebra: Chapters .

## Matrices are basis dependent

- ▶ For linear transformations represented as matrices depend on the choice of basis. For example, if  $\mathbf{A} : \mathbb{R}^n \rightarrow \mathbb{R}^n$  represents a linear transformation in the standard basis, then the same transformation in a basis  $V$  is given by,

$$\mathbf{V}^{-1}\mathbf{A}\mathbf{V} : \text{Similarity transformation}$$

- ▶ In fact, for specific a choice of basis, it is possible to have the simplest possible representation for  $\mathbf{A} \rightarrow$  *Eigen decomposition*.

When a matrix  $\mathbf{A}$  has  $n$  eigenpairs  $\{(\lambda_i, \mathbf{x}_i)\}_{i=1}^n$ , with linearly independent eigenvectors, we have

$$\mathbf{A} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}^{-1}$$

- ▶ What about rectangular matrices  $\mathbf{A} \in \mathbb{R}^{m \times n}$ ? Can we talk about “similar” matrices in this case?

## Matrix equivalence

- ▶ Consider a linear transformation  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , such that  $\mathbf{y} = T(\mathbf{x})$ , where  $\mathbf{x} \in \mathbb{R}^n$  and  $\mathbf{y} \in \mathbb{R}^m$ .  $T$  can be represented as a matrix  $\mathbf{A}$ , such that  $\mathbf{y} = \mathbf{A}\mathbf{x}$ .
- ▶ Exact entries of  $\mathbf{A}$  will depend on the choice of basis for both the input and the output spaces. Let us assume that the matrix  $\mathbf{A}$  is the representation when the standard basis is used for the input and output spaces.
- ▶ If a different set of basis are chosen for the input and output spaces, namely  $V = \{\mathbf{v}_i\}_{i=1}^n$  ( $\mathbf{v}_i \in \mathbb{R}^n$ ) and  $W = \{\mathbf{w}_i\}_{i=1}^m$  ( $\mathbf{w}_i \in \mathbb{R}^m$ ). Then the corresponding matrix representation for the linear transformation  $T$  is,

$$\mathbf{A}_{VW} = \mathbf{W}^{-1}\mathbf{A}\mathbf{V}$$

where, the  $\mathbf{V} = [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \dots \quad \mathbf{v}_n]$  and  $\mathbf{W} = [\mathbf{w}_1 \quad \mathbf{w}_2 \quad \dots \quad \mathbf{w}_m]$ .

- ▶  $\mathbf{A}$  and  $\mathbf{A}_{VW}$  are called *equivalent matrices*.

## Singular Value Decomposition: Diagonalizing any matrix

- ▶ Eigen-decomposition provided a way to do this for a square matrix with full rank.  $\mathbf{A} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}^{-1}$ . When  $\mathbf{A}$  is symmetric,  $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$ .
- ▶ For rectangular and rank-deficient matrices, we can do this using *singular value decomposition*.
- ▶ Consider a matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$  with  $\text{rank}(\mathbf{A}) = r$ .

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_m] \begin{bmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \dots \quad \mathbf{v}_n]^T$$

where,  $\mathbf{U} \in \mathbb{R}^{m \times m}$ ,  $\mathbf{U}\mathbf{U}^T = \mathbf{I}$ ;  $\mathbf{V} \in \mathbb{R}^{n \times n}$ ,  $\mathbf{V}\mathbf{V}^T = \mathbf{I}$ ; and  $\mathbf{D} = \text{diag}(\sigma_1 \dots \sigma_r)$ .

- ▶ Columns  $\mathbf{U}$  are eigenvectors of  $\mathbf{A}^T\mathbf{A}$ , forming an orthonormal basis for  $\mathbb{R}^m$ .
- ▶ Columns  $\mathbf{V}$  are eigenvectors of  $\mathbf{A}\mathbf{A}^T$ , forming an orthonormal basis for  $\mathbb{R}^n$ .
- ▶  $\sigma_i^2 = \lambda_i$ , where  $\lambda_i$ s are the eigenvalues of  $\mathbf{A}^T\mathbf{A}$  and  $\mathbf{A}\mathbf{A}^T$ .

# Singular Value Decomposition: Diagonalizing any matrix

- For  $\mathbf{A}$ ,

$$\begin{aligned} C(\mathbf{A}) &= \text{span}\{\hat{\mathbf{u}}_1 \dots \hat{\mathbf{u}}_r\} & N(\mathbf{A}^T) &= \text{span}\{\hat{\mathbf{u}}_{r+1} \dots \hat{\mathbf{u}}_m\} \\ C(\mathbf{A}^T) &= \text{span}\{\hat{\mathbf{v}}_1 \dots \hat{\mathbf{v}}_r\} & N(\mathbf{A}) &= \text{span}\{\hat{\mathbf{v}}_{r+1} \dots \hat{\mathbf{v}}_n\} \end{aligned}$$

where, the  $\hat{\mathbf{u}}_i$ s and the  $\hat{\mathbf{v}}_i$ s are any orthonormal basis for  $\mathbb{R}^m$  and  $\mathbb{R}^n$ , respectively.

$$\hat{\mathbf{U}}_{cs} = [\hat{\mathbf{u}}_1 \dots \hat{\mathbf{u}}_r], \quad \hat{\mathbf{U}}_{lns} = [\hat{\mathbf{u}}_{r+1} \dots \hat{\mathbf{u}}_m], \quad \hat{\mathbf{V}}_{rs} = [\hat{\mathbf{v}}_1 \dots \hat{\mathbf{v}}_r], \quad \hat{\mathbf{V}}_{ns} = [\hat{\mathbf{v}}_{r+1} \dots \hat{\mathbf{v}}_n]$$

- Now,  $\mathbf{A}$  can be written as,

$$\mathbf{A} = [\hat{\mathbf{U}}_{cs} \quad \hat{\mathbf{U}}_{lns}] \begin{bmatrix} \mathbf{R} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{V}}_{rs}^T \\ \hat{\mathbf{V}}_{ns}^T \end{bmatrix}$$

where,  $\mathbf{R} \in \mathbb{R}^{r \times r}$ .

It can be shown that two orthogonal matrices  $\mathbf{P}$  and  $\mathbf{Q}$  can be chosen, such that

$$\mathbf{A} = [\hat{\mathbf{U}}_{cs} \quad \hat{\mathbf{U}}_{lns}] \mathbf{P} \begin{bmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{Q}^T \begin{bmatrix} \hat{\mathbf{V}}_{rs}^T \\ \hat{\mathbf{V}}_{ns}^T \end{bmatrix} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T$$

# Singular Value Decomposition: Diagonalizing any matrix

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_m] \begin{bmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \dots \mathbf{v}_n]^T$$

- ▶ Orthonormal basis for  $C(\mathbf{A}) \rightarrow \{\mathbf{u}_1 \dots \mathbf{u}_r\}$ .
- ▶ Orthonormal basis for  $N(\mathbf{A}^T) \rightarrow \{\mathbf{u}_{r+1} \dots \mathbf{u}_m\}$ .
- ▶ Orthonormal basis for  $C(\mathbf{A}^T) \rightarrow \{\mathbf{v}_1 \dots \mathbf{v}_r\}$ .
- ▶ Orthonormal basis for  $N(\mathbf{A}) \rightarrow \{\mathbf{v}_{r+1} \dots \mathbf{v}_n\}$ .

$$\mathbf{D} = \begin{bmatrix} \sigma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_r \end{bmatrix}, \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0.$$

$$\text{▶ Reduced SVD: } \mathbf{A} = [\mathbf{u}_1 \dots \mathbf{u}_r] \begin{bmatrix} \sigma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_r \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_r^T \end{bmatrix}$$

# Geometry of SVD

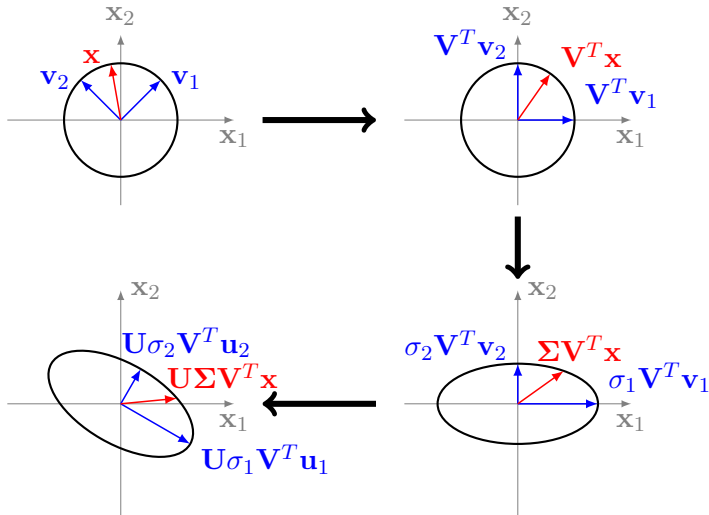
$$\mathbf{y} = \mathbf{A}\mathbf{x}, \text{ where } \mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \\ \mathbf{A} \in \mathbb{R}^{n \times n} \text{ and } \text{rank}(\mathbf{A}) = n.$$

$$\begin{aligned} 1 &= \|\mathbf{x}\|^2 = \|\mathbf{A}^{-1}\mathbf{y}\|^2 \\ &= \mathbf{y}^T \mathbf{A}^{-T} \mathbf{A}^{-1} \mathbf{y} \\ &= \mathbf{y}^T (\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^T)^T \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^T \mathbf{y} \\ &= \mathbf{y}^T \mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{V}^T \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^T \mathbf{y} \\ &= \mathbf{w}^T \mathbf{\Sigma}^{-2} \mathbf{w} \end{aligned}$$

where,  $\mathbf{w} = \mathbf{U}^T \mathbf{y}$ ; and

$$\mathbf{\Sigma}^{-2} = \begin{bmatrix} \mathbf{D}^{-2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

$$x_1^2 + \dots + x_n^2 = \frac{w_1^2}{\sigma_1^2} + \frac{w_2^2}{\sigma_2^2} + \dots + \frac{w_n^2}{\sigma_n^2} = 1$$





## Singular Value Decomposition: Diagonalizing any matrix

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_m] \begin{bmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \dots \quad \mathbf{v}_n]^T$$

SVD allows us to obtain low rank approximation of the given matrix  $\mathbf{A}$ , which has lots of applications in signal processing and data analysis.

$$\mathbf{A} = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \dots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T, \quad \text{rank}(\mathbf{A}) = r$$

where,  $\mathbf{u}_i \mathbf{v}_i^T$  are rank one matrices.

We can obtain a matrix of rank  $k < r$  by setting  $\sigma_i = 0, \forall k < i \leq r$ .

$$\mathbf{A}_k = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \dots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T$$

SVD gives the best possible low rank approximations in terms of the distance between  $\mathbf{A}$  and  $\mathbf{A}_k$ .

$$\min_{\text{rank}(\mathbf{B})=k} \|\mathbf{A} - \mathbf{B}\|_2 = \|\mathbf{A} - \mathbf{A}_k\|_2 = \sigma_{k+1}$$

$$\min_{\text{rank}(\mathbf{B})=k} \|\mathbf{A} - \mathbf{B}\|_F = \|\mathbf{A} - \mathbf{A}_k\|_F = \left( \sum_{i=k+1}^r \sigma_i^2 \right)^{1/2}$$

## Singular Value Decomposition: Diagonalizing any matrix

- ▶ Geometrically, low rank approximations corresponds a  $r$ -dimensional hyperellipsoid approximated by a lower dimensional hyperellipsoid, obtained by flattening the hyperellipsoid along its smallest principal axis.
- ▶ **Principal component analysis:**
  - ▶ Multi-dimensional data often have structure in the form of correlations between the individual variables. Such data can be approximated by a lower dimensional representation.

