UPPSALA
UNIVERSITET

Department of Information Technology

Scientific Computing for Data Analysis

Davoud Mirzaei

Block 3

Numerical Linear Algebra- II

Lecture 8: Singular Value Decomposition (SVD)

Agenda

- Eigenvalues and eigenvectors
- Singular value decomposition (SVD)
- SVD in Python

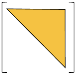
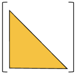
From Linear Algebra (Eigenvalues and Eigenvectors)

- Given a square matrix A of size $n \times n$, we say $\boldsymbol{v} \neq 0$ is an eigenvector corresponding to eigenvalue λ for A if

$$A\boldsymbol{v} = \lambda \boldsymbol{v}$$

- Meaning: when A acts on the vector $\boldsymbol{v}$ it multiplies $\boldsymbol{v}$ by a constant λ (the eigenvalue) - matrix A only shrinks/stretches the vector $\boldsymbol{v}$.
- We can solve

$$\det(A - \lambda I) = 0$$

  to obtain eigenvalues, but never used in computer computations, numerical methods always used
- A is singular $\Longleftrightarrow$ at least one $\lambda_j = 0$
- If  or  then eigenvalues on main diagonal

Eigendecomposition

- If A is real and symmetric then (spectral theorem)

$$A = \underbrace{\begin{bmatrix} | & & | \\ \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n \\ | & & | \end{bmatrix}}_{V} \underbrace{\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}}_{D} \underbrace{\begin{bmatrix} -- & \boldsymbol{v}_1^T & -- \\ & \vdots & \\ -- & \boldsymbol{v}_n^T & -- \end{bmatrix}}_{V^T}$$

  $\boldsymbol{v}_j$ orthonormal eigenvectors, λ_j always real.

- If A is non-symmetric but has n independent eigenvectors $\boldsymbol{v}_j$ then it can be diagonalized as

$$A = \underbrace{\begin{bmatrix} | & & | \\ \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n \\ | & & | \end{bmatrix}}_{V} \underbrace{\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}}_{D} \underbrace{\begin{bmatrix} | & & | \\ \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n \\ | & & | \end{bmatrix}^{-1}}_{V^{-1}}$$

- In the most general case (dependent eigenvector) D is replaced by J, the Jordan form: $A = VJV^{-1}$

Singular Value Decomposition (SVD)

- Eigenvalues only defined for square matrices (same number of rows as columns)
- SVD - a generalization of eigenvalues/eigenvectors, works for all matrices
- SVD is used in many data reduction techniques
- Basis for PCA - to find pattern of correlations
- Use SVD to solve least squares problem (regression) also for singular matrices
- Can be seen as a data driven generalization of the Fourier transform (based on sine and cosine functions to approximate functions)
- SVD allow us to tailor a coordinate system based on the data we have
- ...

SVD: definition

Let $m \geq n$. Every $m \times n$ matrix A has a SVD of the form

$$A = U\Sigma V^T$$

where $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthogonal matrices and $\Sigma \in \mathbb{R}^{m\times n}$ is a diagonal matrix that carries the singular values $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_n \geqslant 0$ on its diagonal (decreasing order).

$$\underbrace{\begin{pmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{pmatrix}}_{A} = \underbrace{\begin{pmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \end{pmatrix}}_{U} \underbrace{\begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}}_{\Sigma} \underbrace{\begin{pmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{pmatrix}}_{V^T}$$

Reduced SVD

The last $m-n$ rows of Σ are zeros so the last $m-n$ columns of U have no contribution to the product (but are still important!):

$$\underbrace{\begin{pmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{pmatrix}}_{A} = \underbrace{\begin{pmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \end{pmatrix}}_{[U_1 \quad U_2]} \underbrace{\begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \\ \hline 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}}_{\begin{bmatrix} \Sigma_1 \\ 0 \end{bmatrix}} \underbrace{\begin{pmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{pmatrix}}_{V^T}$$

The reduced SVD is:

$$A = U_1 \Sigma_1 V^T$$

where U_1 is $m \times (m-n)$ and Σ is $n \times n$ and diagonal. V is $n \times n$ as before.

An example:

$$A = \begin{bmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{18}} & \frac{2}{3} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{18}} & -\frac{2}{3} \\ 0 & \frac{4}{\sqrt{18}} & -\frac{1}{3} \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 3 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

The reduced SVD for this example is:

$$A = \begin{bmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{18}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{18}} \\ 0 & \frac{4}{\sqrt{18}} \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

How matrices U and V and singular values σ_k can be computed? A simple approach comes in the next page ...

SVD (connection to eigenvalues/eigenvectors)

Compute A^TA and AA^T via SVD:

$$A^TA = (U\Sigma V^T)^T(U\Sigma V^T) = V\Sigma^T \underbrace{U^TU}_{I} \Sigma V^T = V \underbrace{\Sigma^T\Sigma}_{D} V^T = VDV^T$$

where D is a $n \times n$ diagonal matrix

$$D = \Sigma^T\Sigma = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_n^2 \end{bmatrix}$$

- Since V is orthogonal, VDV^T is the **eigendecomposition** of symmetrix matrix A^TA.
- This means that $\sigma_1^2, \dots, \sigma_n^2$ are eigenvalues and columns of V are corresponding eigenvectors of A^TA.

Next page ...

SVD (connection to eigenvalues/eigenvectors)

- The same computation for AA^T shows that

$$AA^T = U\Sigma\Sigma^T U^T = UDU^T$$

where D is a $m \times m$ diagonal matrix

$$D = \Sigma\Sigma^T = \begin{bmatrix} \sigma_1^2 & & & & & 0 \\ & \ddots & & & & \\ & & \sigma_n^2 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ 0 & & & & & 0 \end{bmatrix}$$

- The columns of U are eigenvectors of AA^T corresponding to the same eigenvalues $\sigma_1^2, \ldots, \sigma_n^2$ plus $m-n$ zero eigenvalues $\sigma_{n+1}^2 = \cdots = \sigma_m^2 = 0$.
- Conclusion: Singular values of A are square roots of eigenvalues of A^TA. Columns of V are eigenvectors of A^TA. Columns of U are eigenvectors of AA^T
- A different and computationally more stable algorithm is implemented in computers.

Example: To compute the SVD of

$$A = \begin{bmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{bmatrix}$$

we form

$$A^T A = \begin{bmatrix} 17 & 8 \\ 8 & 17 \end{bmatrix}, \quad \det(A^T A - \lambda I) = (17-\lambda)^2 - 64 = 0 \Longrightarrow \lambda_1 = 25,\ \lambda_2 = 9$$

which gives $\sigma_1 = \sqrt{\lambda_1} = \sqrt{25} = 5$ and $\sigma_2 = \sqrt{\lambda_2} = \sqrt{9} = 3$. We can also show that eigenvectors of $A^T A$ are

$$\boldsymbol{v}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \ \boldsymbol{v}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \Longrightarrow V = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Next page ...

Columns of U are eigenvectors of AA^T:

$$AA^T = \begin{bmatrix} 13 & 12 & 2 \\ 12 & 13 & -2 \\ 23 & -2 & 8 \end{bmatrix}$$

No need to compute eigenvalues as we know they are $\lambda_1 = 25$, $\lambda_2 = 9$ and $\lambda_3 = 0$ (why?)
But eigenvectors of AA^T are are different from those of A^TA! They are

$$\boldsymbol{u}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \; \boldsymbol{u}_2 = \frac{1}{\sqrt{18}} \begin{bmatrix} 1 \\ -1 \\ 4 \end{bmatrix}, \; \boldsymbol{u}_3 = \frac{1}{3} \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}$$

$$\Longrightarrow U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{18}} & \frac{2}{3} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{18}} & -\frac{2}{3} \\ 0 & \frac{4}{\sqrt{18}} & -\frac{1}{3} \end{bmatrix}$$

SVD in python

The commonly used algorithm for SVD is the **Golub-Kahan-Reinsch** algorithm. It is implemented in `numpy.linalg` (and also `scipy.linalg` modules. The `numpy` syntax is:

```
numpy.linalg.svd(A)
```

which is equivalent to:

```
numpy.linalg.svd(A, full_matrices=True, compute_uv=True, hermitian=False)
```

- `full_matrices=False`: the output will be the reduced SVD
- `compute_uv=False`: unitary matrices U and V are not computed and the output is vector of singular values only
- `hermitian=True`: A is assumed to be Hermitian (symmetric if real-valued), enabling a more efficient method for finding singular values
- The command `numpy.linalg.svd(A,0)` also gives the reduced SVD.

An example:

```python
import numpy as np
A = np.array([[1,2,3],[2,1,4],[1,1,-1],[2,5,-3],[2,-4,-1]])

U,S,Vt = np.linalg.svd(A)
print('Full SVD:','\n U =\n',U,'\n V =\n',Vt.T,'\n S =\n',S)

U,S,Vt = np.linalg.svd(A, full_matrices = False)
print('Reduced SVD:','\n U =\n',U,'\n V =\n',Vt.T,'\n S =\n',S)
```

The outputs of `svd` function in Python are U, V^T instead of V, and vector $S = [\sigma_1, \ldots, \sigma_n]$ instead of a diagonal matrix (even in the full version).

Outputs: next page ...

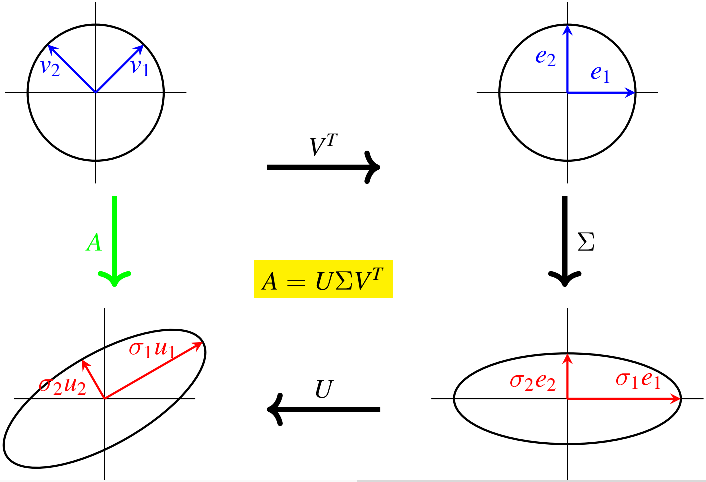
(Numbers edited to 3 decimal places)

```
Full SVD:
U =
[[-0.253  0.548 -0.077 -0.57   0.552]
 [-0.123  0.715 -0.384  0.427 -0.378]
 [-0.185 -0.127 -0.248 -0.659 -0.674]
 [-0.806 -0.368 -0.351  0.236  0.189]
 [ 0.486 -0.192 -0.814 -0.048  0.25 ]]
V =
[[-0.19   0.122 -0.974]
 [-0.974  0.102  0.202]
 [ 0.124  0.987  0.099]]
S =
 [6.973 6.003 3.513]

Reduced SVD:
U =
[[-0.253  0.548 -0.077]
 [-0.123  0.715 -0.384]
 [-0.185 -0.127 -0.248]
 [-0.806 -0.368 -0.351]
 [ 0.486 -0.192 -0.814]]
V =
[[-0.19   0.122 -0.974]
 [-0.974  0.102  0.202]
 [ 0.124  0.987  0.099]]
S =
 [6.973 6.003 3.513]
```

A geometric interpretation of SVD

- Let $S = \{\boldsymbol{x} \in \mathbb{R}^n : \|x\|_2 = 1\}$, an sphere in $\mathbb{R}^n$ (a circle in $\mathbb{R}^2$)
- If A is a $m \times n$ matrix then AS is an hyperellipsoid in $\mathbb{R}^m$ (A shrinks/stretches and rotates S)
- A geometric interpretation via SVD:



Some useful properties of SVD

If the $m \times n$ matrix A has SVD $A = U\Sigma V^T$ then:

1. $\|A\|_2 = \sigma_1$,
2. $\|A\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_n^2}$,
3. $A^{-1} = V\Sigma^{-1}U^T$ when $m = n$ and A is nonsingular ($\sigma_n \neq 0$)

$$\Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma_1} & & \\ & \ddots & \\ & & \frac{1}{\sigma_n} \end{bmatrix}$$

4. $\|A^{-1}\|_2 = \frac{1}{\sigma_n}$ when $m = n$ and A is nonsingular,
5. $\mathrm{cond}_2(A) = \frac{\sigma_1}{\sigma_n}$ if A is nonsingular,
6. $\mathrm{rank}(A) =$ number of nonzeros singular values.

- The proof of all above properties follows from the facts that V and U are orthogonal and multiplication by an orthogonal matrix does not change the 2-norm, the Frobenius norm, and the rank of a matrix. (left as exercise!)

Computing pseudoinverse via SVD

- If A has rank $r < n$ then the last $n - r$ singular values are zeros: $\sigma_r \neq 0$, $\sigma_{r+1} = \cdots = \sigma_n = 0$ (In this case A is called rank-deficient)

$$\Sigma = \left[\begin{array}{ccc|ccc} \sigma_1 & & & & & \\ & \ddots & & & 0 & \\ & & \sigma_r & & & \\ \hline & & & 0 & & \\ & 0 & & & \ddots & \\ & & & & & 0 \\ \hdashline & 0 & & & 0 & \end{array}\right] = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix}$$

  Then $A = U_1 \Sigma_1 V_1^T$ where U_1 and V_1 are the first r columns of U and V, respectively

- The **pseudoinverse** of a general matrix A then is defined by

$$A^+ = V_1 \Sigma_1^{-1} U_1^T$$

Example: The SVD factors of a matrix A are given by:

$$U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \Sigma = \begin{bmatrix} 2\sqrt{3} & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} V = \begin{bmatrix} \frac{\sqrt{6}}{3} & 0 & 0 & \frac{-1}{\sqrt{3}} \\ 0 & 0 & 1 & 0 \\ \frac{1}{\sqrt{6}} & \frac{-1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{3}} \end{bmatrix}$$

What is the pseudoinverse of A?

$$A^+ = V_1\Sigma_1^{-1}U_1^T = \begin{bmatrix} \frac{\sqrt{6}}{3} & 0 \\ 0 & 0 \\ \frac{1}{\sqrt{6}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{2\sqrt{3}} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{\sqrt{2}} & 0 \\ \frac{-1}{\sqrt{2}} & 0 & 0 & \frac{1}{\sqrt{2}} & 0 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{6} & 0 & 0 & \frac{1}{6} & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Analysis: A matrix has SVD $A = U\Sigma V^T$ with

$$\Sigma = \begin{bmatrix} 14.6 & 0 & 0 & 0 \\ 0 & 8.4 & 0 & 0 \\ 0 & 0 & 1.3 & 0 \\ 0 & 0 & 0 & 0.03 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

- What is rank of A?
- What are $\|A\|_2$ and $\|A\|_F$?
- What is $\mathrm{cond}_2(A)$?
- What are eigenvalues of A^TA? what are eigenvalues of AA^T?
- what are rank of A^TA and AA^T?

The second part of a question for higher grades:

B) (4 points) For which value(s) of $b \in \mathbb{R}, -3 < b < 3,$ is for the following matrix

$$\begin{pmatrix} 1 & b \\ b & 9 \end{pmatrix}$$

the singular value $\sigma_2 \geq 1$? **Justify** your response. No python allowed for this problem.

(If you write your solution on a piece of paper, make a note here about that.)

Fill in your answer here