

12: Singular value decomposition

“many problems of linear algebra can be better understood if we first ask the question: what if we take the SVD?”

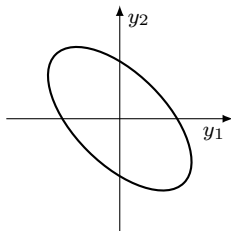
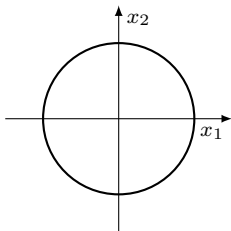
– Trefethen & Bau p.25

- SVD definition
- relationship to eigenvalue decomposition
- SVD properties
- low-rank approximation
- least squares

Geometric observation

the image of the unit sphere under any $m \times n$ matrix is a **hyperellipse**

$$x \mapsto Ax = y$$



Hyperellipses

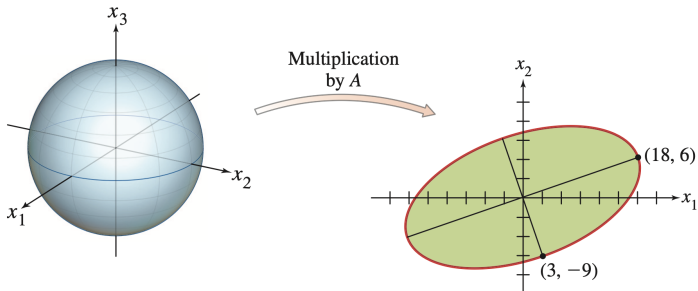
- hyper: m -dimensional generalization of an ellipse
- stretches vectors on the unit sphere S by σ_i in orthogonal direction $u_i \in \mathbb{R}^m$

$$\mathbf{image}(AS) = \{Ax \mid x \in S\}$$

Rectangular matrix

consider the matrix $A \in \mathbb{R}^{2 \times 3}$ and its action $x \mapsto Ax$ on the unit sphere

$$A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$$



the matrix A is a mapping from \mathbb{R}^3 to \mathbb{R}^2

[Image from **Linear Algebra and its Applications**, Lay, Lay, and McDonal]

Singular value decomposition (SVD)

any matrix $A \in \mathbb{C}^{m \times n}$ (m and n arbitrary) can be factored as

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

where

- $U \in \mathbb{C}^{m \times m}$ is **unitary**
- $V \in \mathbb{C}^{n \times n}$ is **unitary**
- $\Sigma \in \mathbb{R}^{m \times n}$ is **diagonal***
- diagonal elements of Σ , denoted σ_j are non-negative
- by convention, we order the **singular values**

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0, \quad p = \min\{m, n\}$$

- this is often referred to as the **full SVD**

SVD for tall matrices

$$\begin{array}{c} m \\ A \end{array} \begin{array}{c} n \\ \end{array} = \begin{array}{c} m \times m \\ U \end{array} \begin{array}{c} m \times n \\ \Sigma \end{array} \begin{array}{c} n \times n \\ V \end{array}$$

SVD for wide matrices

$$\begin{array}{c} m \\ A \end{array} \begin{array}{c} n \\ \end{array} = \begin{array}{c} m \times m \\ U \end{array} \begin{array}{c} m \times n \\ \Sigma \end{array} \begin{array}{c} n \times n \\ V \end{array}$$

- Σ always has the same shape as A

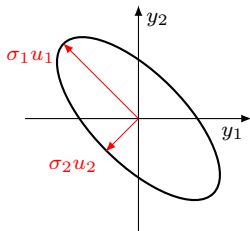
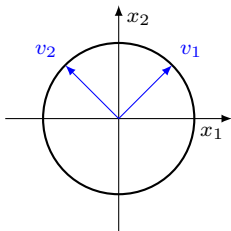
Singular vectors

when A is real, so are its singular vectors

$$A = U\Sigma V^T, \quad \text{with} \quad U^T U = U U^T = I \quad \text{and} \quad V^T V = V V^T = I$$

right multiplying both sides by V gives $AV = U\Sigma$

$$Av_i = \sigma_i u_i \quad \text{for} \quad i = 1, \dots, p$$



Sum of rank-1 matrices

without loss of generality assume $m > n$, consider $A = U\Sigma V^T$

$$\begin{aligned}
 & \begin{bmatrix} u_1 & \dots & u_n & \dots & u_m \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_n & & \\ \text{---} & 0 & \text{---} & & \\ \text{---} & 0 & \text{---} & & \end{bmatrix} \begin{bmatrix} \text{---} v_1^T \text{---} \\ \vdots \\ \text{---} v_n^T \text{---} \end{bmatrix} \\
 &= \begin{bmatrix} u_1 & \dots & u_n \end{bmatrix} \begin{bmatrix} \text{---} \sigma_1 v_1^T \text{---} \\ \vdots \\ \text{---} \sigma_n v_n^T \text{---} \end{bmatrix} \\
 &= \underbrace{\sigma_1 u_1 v_1^T}_{\text{rank 1}} + \underbrace{\sigma_2 u_2 v_2^T}_{\text{rank 1}} + \dots + \underbrace{\sigma_n u_n v_n^T}_{\text{rank 1}}
 \end{aligned}$$

- in general, replace n above with $p = \min\{m, n\}$
- rank is the number of linearly independent rows/cols in a matrix

Relationship to eigenvalue decomposition

recall (p.5-15) that for an $n \times n$ square matrix W with linearly independent eigenvectors

$$A = X\Lambda X^{-1}, \quad \text{where } \Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \quad \text{and} \quad X = \begin{bmatrix} | & & | \\ x_1 & \dots & x_n \\ | & & | \end{bmatrix}$$

Relationship to eigenvalue decomposition

recall (p.5-15) that for an $n \times n$ square matrix W with linearly independent eigenvectors

$$A = X\Lambda X^{-1}, \quad \text{where } \Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \quad \text{and } X = \begin{bmatrix} | & & | \\ x_1 & \dots & x_n \\ | & & | \end{bmatrix}$$

- consider the Gram matrix $A^T A$ with $A = U\Sigma V^T$

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

Relationship to eigenvalue decomposition

recall (p.5-15) that for an $n \times n$ square matrix W with linearly independent eigenvectors

$$A = X\Lambda X^{-1}, \quad \text{where } \Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \quad \text{and } X = \begin{bmatrix} | & & | \\ x_1 & \dots & x_n \\ | & & | \end{bmatrix}$$

- consider the Gram matrix $A^T A$ with $A = U\Sigma V^T$

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

- the matrix $\Sigma^T \Sigma$ has the form

$$\begin{bmatrix} \sigma_1^2 & & & \\ & \ddots & & \\ & & \sigma_p^2 & \\ & & & 0_k \end{bmatrix}, \quad \text{where } k = n - \min\{m, n\}$$

Relationship to eigenvalue decomposition

recall (p.5-15) that for an $n \times n$ square matrix W with linearly independent eigenvectors

$$A = X\Lambda X^{-1}, \quad \text{where } \Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \quad \text{and} \quad X = \begin{bmatrix} | & & | \\ x_1 & \dots & x_n \\ | & & | \end{bmatrix}$$

- consider the Gram matrix $A^T A$ with $A = U\Sigma V^T$

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

- the matrix $\Sigma^T \Sigma$ has the form

$$\begin{bmatrix} \sigma_1^2 & & & \\ & \ddots & & \\ & & \sigma_p^2 & \\ & & & 0_k \end{bmatrix}, \quad \text{where } k = n - \min\{m, n\}$$

- $A^T A$ is **symmetric** and V is **orthogonal**, so $A^T A = V(\Sigma^T \Sigma)V^T$ is an **eigen-decomposition**

Relationship to eigenvalue decomposition

recall (p.5-15) that for an $n \times n$ square matrix W with linearly independent eigenvectors

$$A = X\Lambda X^{-1}, \quad \text{where } \Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \quad \text{and} \quad X = \begin{bmatrix} | & & | \\ x_1 & \dots & x_n \\ | & & | \end{bmatrix}$$

- consider the Gram matrix $A^T A$ with $A = U\Sigma V^T$

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

- the matrix $\Sigma^T \Sigma$ has the form

$$\begin{bmatrix} \sigma_1^2 & & & \\ & \ddots & & \\ & & \sigma_p^2 & \\ & & & 0_k \end{bmatrix}, \quad \text{where } k = n - \min\{m, n\}$$

- $A^T A$ is **symmetric** and V is **orthogonal**, so $A^T A = V(\Sigma^T \Sigma) V^T$ is an **eigen-decomposition**
- V is the matrix of eigenvectors for $A^T A$

Summary

- $A^T A = V(\Sigma^T \Sigma)V^T$ is an eigen-decomposition
- v_i (i^{th} right singular vector) is an eigenvector of $A^T A$
- $\sigma_i = \sqrt{\lambda_i(A^T A)}$

following the same arguments as on p.12-8, we arrive at the following conclusion

- $AA^T = U(\Sigma \Sigma^T)U^T$ is an eigen-decomposition
- u_i (i^{th} left singular vector) is an eigenvector of AA^T
- $\sigma_i = \sqrt{\lambda_i(AA^T)}$

$A^T A$ and AA^T share the first p eigenvalues

$$\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2$$

Existence and uniqueness

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

- $U = [u_1, u_2, \dots, u_m]$: the column vectors u_i are the **left singular vectors**
- the n columns of V are the **right singular vectors**
- $\sigma_1, \dots, \sigma_{\min\{m,n\}}$ are the singular values

Existence and uniqueness

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

- $U = [u_1, u_2, \dots, u_m]$: the column vectors u_i are the **left singular vectors**
- the n columns of V are the **right singular vectors**
- $\sigma_1, \dots, \sigma_{\min\{m,n\}}$ are the singular values

Theorem

Every matrix $A \in \mathbb{C}^{m \times n}$ has a full singular value decomposition. The singular values $\{\sigma_j\}$ are uniquely determined, and, if $m = n$ and σ_j are distinct, the left and right singular vectors are uniquely determined up to complex signs.

- For the proof, see Trefethen and Bau p.29

Non-uniqueness example

- consider $A = U\Sigma V^*$ for A with distinct singular values
- consider the scalar $e^{i\theta}$ where $\theta \bmod 2\pi \neq 0$:

$$|e^{i\theta}| = 1$$

- define $\tilde{U} = e^{i\theta}U$ and $\tilde{V} = e^{i\theta}V$, then

$$A = U\Sigma V^* = \tilde{U}\Sigma\tilde{V}^* \quad \text{and} \quad U \neq \tilde{U}, \quad V \neq \tilde{V}$$

Summary

- singular values are unique
- A and A^T have the same singular values
- singular vectors are unique up to sign

Image of the unit ball – revisited

Rank

Rank of the product XY

- let $X \in \mathbb{R}^{m \times r}$ have linearly independent columns
- and $Y \in \mathbb{R}^{r \times n}$ have linearly independent rows
- then $\mathbf{rank}(XY) = r$

it will become clear how to prove this later

Rank

Rank of the product XY

- let $X \in \mathbb{R}^{m \times r}$ have linearly independent columns
- and $Y \in \mathbb{R}^{r \times n}$ have linearly independent rows
- then $\mathbf{rank}(XY) = r$

it will become clear how to prove this later

Theorem

Consider $A \in \mathbb{R}^{m \times n}$ with singular values

$$\sigma_1 \geq \sigma_2 \geq \dots \sigma_r > 0 = \sigma_{r+1} = \dots = \sigma_p.$$

The number of positive singular values is the rank of the matrix, i.e.,

$$\mathbf{rank}(A) = r.$$

Proof:

- partition U, Σ, V so that A is factored as

$$A = [U_1 \quad U_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}$$

where Σ_1 is an $r \times r$ diagonal matrix

- expand the right hands side to get $A = U_1 \Sigma_1 V_1^T$
- by definition, U_1 and V_1 have orthonormal columns so the rank of A is determined by Σ which has rank r

“Economy” SVD

- for large m or large n the matrices U and V will be large
- it is often better to work with a reduced SVD

$$A = \sum_{i=1}^r \sigma_i u_i v_i^T = U_r \Sigma_1 V_r^*$$

- if $m > n$ then $V_r = V$, if $m < n$ the $U_r = U$

Tall matrices

$$\begin{array}{c} m \\ \boxed{A} \end{array} \begin{array}{c} n \\ \end{array} = \begin{array}{c} m \times r \\ \boxed{U_r} \end{array} \begin{array}{cc} r \times r & r \times n \\ \boxed{\Sigma_1} & \boxed{V^*} \end{array}$$

- V is unitary
- U_r has orthonormal columns

Frobenius norm

for $A \in \mathbb{F}^{m \times n}$, the **Frobenius-norm** of A is

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2} = \sqrt{\sum_i^{\min\{m,n\}} \sigma_i^2(A)}$$

to see this, we use the fact that U and V have orthonormal columns, and so

$$\|A\|_F = \|U\Sigma V^*\|_F = \|\Sigma\|_F = \sqrt{\sum_i^{\min\{m,n\}} \sigma_i^2(A)}$$

the Frobenius norm is equivalent to the Euclidean-norm of the vectorization of A

$$\|A\|_F = \|\mathbf{vec}(A)\|_2$$

sometimes called the Schatten 2-norm

Spectral norm

for $A \in \mathbb{F}^{m \times n}$, the **spectral-norm** of A is

$$\|A\|_2 = \sigma_1(A) = \|U\Sigma V^*\|_2 = \|\Sigma\|_2$$

the spectral-norm is an example of an **induced** norm, an equivalent definition is

$$\|A\|_2 = \underset{\|x\|_2=1}{\text{maximize}} \frac{\|Ax\|_2}{\|x\|_2}$$

also known as the Schatten ∞ -norm

$$\|A\|_2 = \|\sigma\|_\infty$$

also also known as the operator norm

Low-rank approximation

we have seen that $A = U\Sigma V^T$ can be expressed as

$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \dots + \sigma_n u_n v_n^T$$

recall that the singular values are ordered by magnitude

$$\sigma_1 \geq \sigma_2 \geq \dots \sigma_r > \underbrace{\sigma_{r+1} = \dots = \sigma_p}_{=0}$$

if σ_i is small, then it doesn't contribute much to A

$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \dots + \sigma_{r-1} u_{r-1} v_{r-1}^T + \sigma_r u_r v_r^T$$

a low-rank approximation of A ignores the singular values in the tail

$$A_l = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \dots + \sigma_l u_l v_l^T \quad (l < r)$$

Optimal low-rank approximation

Theorem

For any matrix A with $\mathbf{rank}(A) = r$, define for any l such that $0 < l \leq r$,

$$A_l = \sum_{i=1}^l \sigma_i u_i v_i^*.$$

For any rank- l matrix X , A_l provides the best

- Frobenius-norm low- l approximation of A , specifically,

$$\|A - X\|_F \geq \|A - A_l\|_F = \sqrt{\sigma_{l+1}^2 + \dots + \sigma_r^2}.$$

- spectral-norm rank- l approximation of A , specifically,

$$\|A - X\|_2 \geq \|A - A_l\|_2 = \sigma_{l+1}.$$

Low-rank approximation

- 1350×1080 pixels (1,458,000 words)



Low-rank approximation

- rank-20 approximation, 48,600 words, 30x compression



Low-rank approximation

- rank-100 approximation, 243,000 words, 6x compression



Low-rank approximation

- rank-250 approximation, 607,500 words, 3x compression

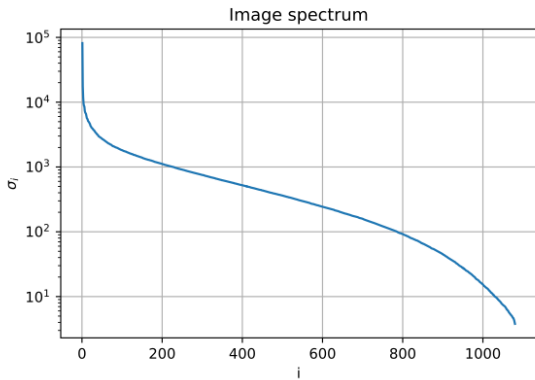


Low-rank approximation

- original

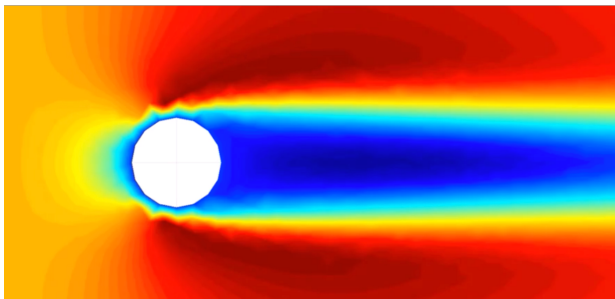


Low-rank approximation



Low-rank approximation

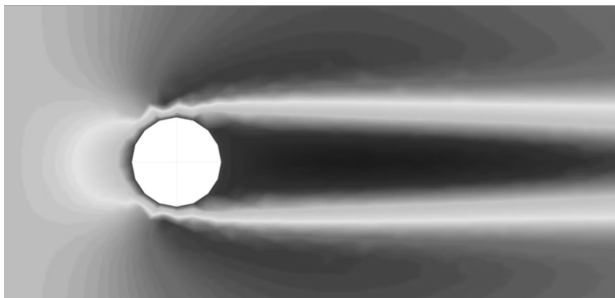
- fluid flow simulation



- $194 \times 954 \times 3$ RGB image

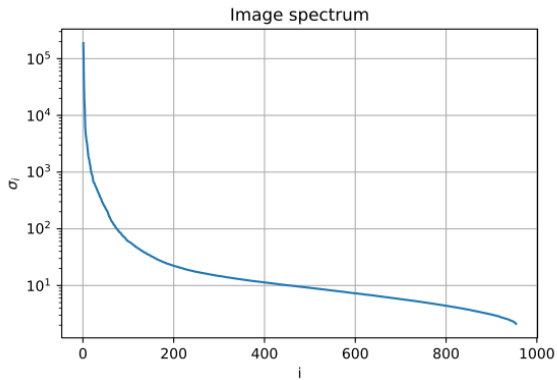
Low-rank approximation

- fluid flow simulation

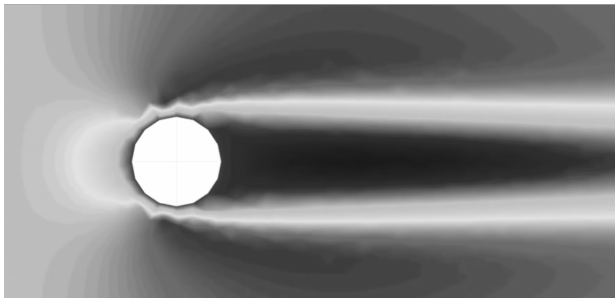


- 1994×954 grayscale image (1,902,276 words)

Low-rank approximation

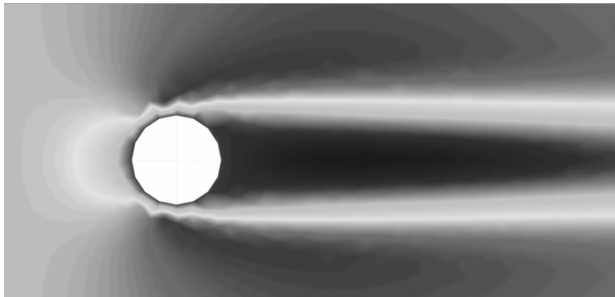


Low-rank approximation



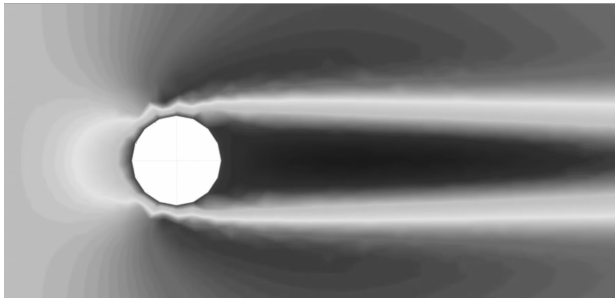
- rank-800 approximation
- 2,358,400 words (reduction?!)

Low-rank approximation



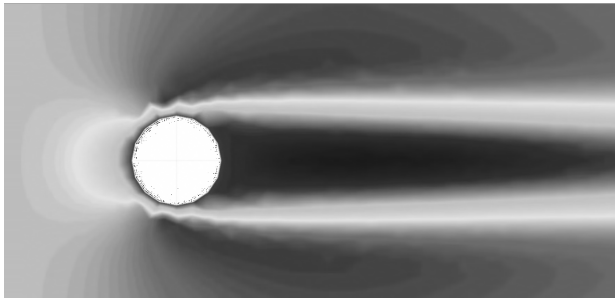
- rank-500 approximation
- 1,474,000 words 0.75x compression

Low-rank approximation



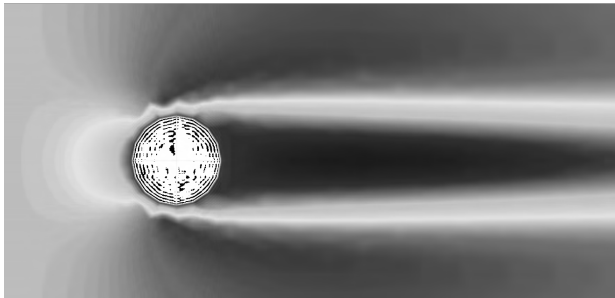
- rank-200 approximation
- 589,600 words

Low-rank approximation



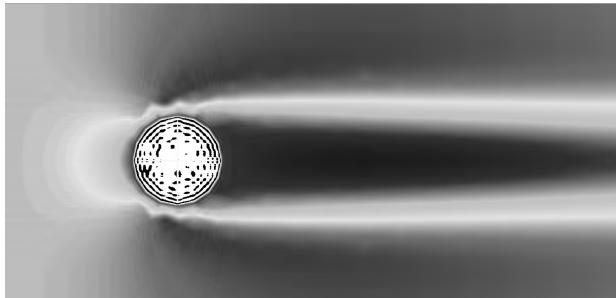
- rank-70 approximation
- 206,360 words

Low-rank approximation



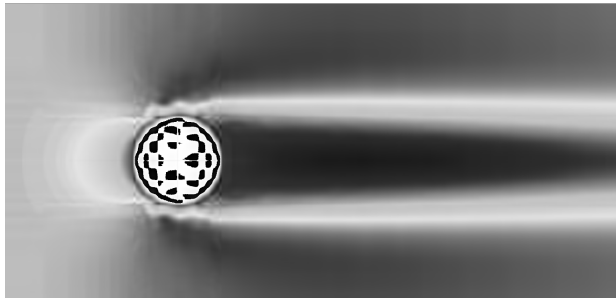
- rank-30 approximation
- 88,440 words

Low-rank approximation



- rank-20 approximation
- 58,960

Low-rank approximation



- rank-10 approximation
- 29,480 words

Matlab & Python

to compute an SVD in MATLAB

- `>> [U,S,V]=SVD(A)`
- `>> [U,S,V]=SVD(A,'econ')`

to compute an SVD in Python

```
>>> from scipy import linalg
```

- `>>> U, S, V = linalg.svd(A)`
- `>>> U, S, V = linalg.svd(A,fullmatrices=False)`

Least squares via SVD

recall the least squares solution $\hat{x} = (A^T A)^{-1} A^T b$ solves

$$\text{minimize } \|Ax - b\|_2^2$$

rewrite least squares solution using reduced SVD: $A = U_r \Sigma_1 V_r^T$

$$\begin{aligned}\hat{x} &= (A^T A)^{-1} A^T b = ((U \Sigma V^T)^T (U \Sigma V^T))^{-1} (U \Sigma V^T)^T b \\ &= (V \Sigma^T U^T U \Sigma V^T)^{-1} V \Sigma^T U^T b \\ &= (V \Sigma^2 V^T)^{-1} V \Sigma U^T b \\ &= V \Sigma^{-2} V^T V \Sigma U^T b \\ &= V \Sigma^{-1} U^T b.\end{aligned}$$

$$\boxed{\hat{x} = V \Sigma^{-1} U^T b}$$

Algorithm

- 1 Compute reduced SVD $A = U \Sigma V^T$
- 2 Matrix-vector product $z = U^T b$
- 3 Solve diagonal system $\Sigma y = z$
- 4 Form $\hat{x} = V y$

Further topics

- four subspaces associated with an $m \times n$ matrix
- Courant–Fischer theorem
- sensitivity of linear equations