Factorization

CE40282-1: Linear Algebra
Hamid R. Rabiee and Maryam Ramezani
Sharif University of Technology

Review

- Eigenvalue and Eigenvector
- A positive definite matrix S has positive eigenvalues, positive pivots, positive determinants, and positive energy $v^T S v$ for every vector v. $S = A^T A$ is always positive definite if A has independent columns.
- Positive Definite Matrix
  - Five Tests
    - All eigenvalues are greater than 0
    - $x^T S x > 0$ for all x (other than zero-vector)
    - If S is positive definite $S = A^T A$ (A must have independent columns)
    - All upper left determinants must be > 0
    - Every pivot must be > 0

Positive Definite Matrix

- If S is positive definite $S = A^T A$ (A must have independent columns): $A^T A$ is positive definite iff the columns of A are linearly independent.
  - Proof?

Positive Definite Matrix

- All upper left determinants must be > 0

$$A = \begin{bmatrix} \boxed{2} & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

Positive Definite Matrix

- Every pivot must be > 0
  - Pivots are, in general, way easier to calculate than eigenvalues.
  - Just perform elimination and examine the diagonal terms.
  - Example: Is the following matrix positive definite matrix? $$\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$$
  - Note: number of positive (negative) pivots = number of positive (negative) eigenvalue

LU-factorization

- Review: Gaussian Elimination, row operations are used to change the coefficient matrix to an upper triangular matrix.
- LU Decomposition is very useful when we have large matrices n x n and if we use gauss-jordan or the other methods, we can get errors.

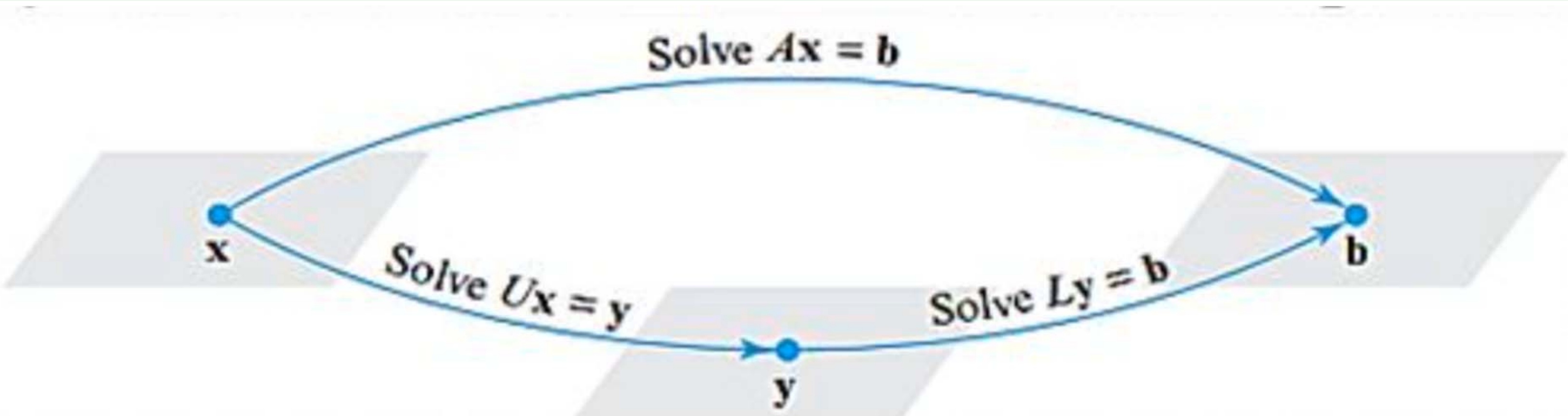
DEFINITION 1 A factorization of a square matrix A as

$$A = LU \tag{1}$$

where L is lower triangular and U is upper triangular, is called an ***LU-decomposition*** (or ***LU-factorization***) of A.

Method of LU Factorization

1) Rewrite the system $Ax = b$ as $LUx = b$
2) Define a new $n \times 1$ matrix y by $Ux = y$
3) Use $Ux = y$ to rewrite $LUx = b$ as $Ly = b$ and solve the system for y
4) Substitute y in $Ux = y$ and solve for x



Constructing LU Factorization

1) Reduce ***A*** to a REF form ***U*** by Gaussian elinmination without row exchanges, keeping track of the multipliers used to introduce the leading ***1s*** and multipliers used to introduce the zeros below the leading ***1s***
2) In each position along the main diagonal of ***L*** place the reciprocal of the multiplier that introduced the leading ***1*** in that position in ***U***
3) In each position below the main diagonal of ***L*** place negative of the multiplier used to introduce the zero in that position in ***U***
4) Form the decomposition ***A = LU***

Constructing LU Factorization

- Example

$$A = \begin{bmatrix} 6 & -2 & 0 \\ 9 & -1 & 1 \\ 3 & 7 & 5 \end{bmatrix}$$

$$A = \begin{bmatrix} 6 & -2 & 0 \\ 9 & -1 & 1 \\ 3 & 7 & 5 \end{bmatrix} \qquad \begin{bmatrix} \bullet & 0 & 0 \\ \bullet & \bullet & 0 \\ \bullet & \bullet & \bullet \end{bmatrix}$$

• denotes an unknown entry of L.

$$\begin{bmatrix} 1 & -\frac{1}{3} & 0 \\ 9 & -1 & 1 \\ 3 & 7 & 5 \end{bmatrix} \longleftarrow \text{multiplier} = \tfrac{1}{6} \qquad \begin{bmatrix} 6 & 0 & 0 \\ \bullet & \bullet & 0 \\ \bullet & \bullet & \bullet \end{bmatrix}$$

$$\begin{bmatrix} 1 & -\frac{1}{3} & 0 \\ 0 & 2 & 1 \\ 0 & 8 & 5 \end{bmatrix} \begin{matrix} \\ \longleftarrow \text{multiplier} = -9 \\ \longleftarrow \text{multiplier} = -3 \end{matrix} \qquad \begin{bmatrix} 6 & 0 & 0 \\ 9 & \bullet & 0 \\ 3 & \bullet & \bullet \end{bmatrix}$$

$$\begin{bmatrix} 1 & -\frac{1}{3} & 0 \\ 0 & 1 & \frac{1}{2} \\ 0 & 8 & 5 \end{bmatrix} \longleftarrow \text{multiplier} = \tfrac{1}{2} \qquad \begin{bmatrix} 6 & 0 & 0 \\ 9 & 2 & 0 \\ 3 & \bullet & \bullet \end{bmatrix}$$

$$\begin{bmatrix} 1 & -\frac{1}{3} & 0 \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix} \longleftarrow \text{multiplier} = -8 \qquad \begin{bmatrix} 6 & 0 & 0 \\ 9 & 2 & 0 \\ 3 & 8 & \bullet \end{bmatrix}$$

$$U = \begin{bmatrix} 1 & -\frac{1}{3} & 0 \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix} \longleftarrow \text{multiplier} = 1 \qquad L = \begin{bmatrix} 6 & 0 & 0 \\ 9 & 2 & 0 \\ 3 & 8 & 1 \end{bmatrix}$$

No actual operation is performed here since there is already a leading 1 in the third row.

Thus, we have constructed the LU-decomposition

$$A = LU = \begin{bmatrix} 6 & 0 & 0 \\ 9 & 2 & 0 \\ 3 & 8 & 1 \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{3} & 0 \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix}$$

Some Notes

- Sometimes it is impossible to write a matrix in the form “lower triangular”×“upper triangular”.
- An invertible matrix A has an LU decomposition provided that all upper left determinants are non-zero.

PLU Factorization

- if A is $n \times n$ and nonsingular, then it can be factored as

$$A = PLU$$

P is a permutation matrix, L is unit lower triangular, U is upper triangular

- not unique; there may be several possible choices for P, L, U
- interpretation: permute the rows of A and factor $P^T A$ as $P^T A = LU$
- also known as *Gaussian elimination with partial pivoting* (GEPP)

$$\begin{bmatrix} 0 & 5 & 5 \\ 2 & 9 & 0 \\ 6 & 8 & 8 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1/3 & 1 & 0 \\ 0 & 15/19 & 1 \end{bmatrix} \begin{bmatrix} 6 & 8 & 8 \\ 0 & 19/3 & -8/3 \\ 0 & 0 & 135/19 \end{bmatrix}$$

- we will skip the details of calculating P, L, U

Cholesky Factorization

every positive definite matrix $A \in \mathbf{R}^{n \times n}$ can be factored as

$$A = R^T R$$

where R is upper triangular with positive diagonal elements

- complexity of computing R is $(1/3)n^3$ flops
- R is called the *Cholesky factor* of A
- can be interpreted as “square root” of a positive definite matrix
- gives a practical method for testing positive definiteness

Cholesky factorization algorithm

$$\begin{bmatrix} A_{11} & A_{1,2:n} \\ A_{2:n,1} & A_{2:n,2:n} \end{bmatrix} = \begin{bmatrix} R_{11} & 0 \\ R^T_{1,2:n} & R^T_{2:n,2:n} \end{bmatrix} \begin{bmatrix} R_{11} & R_{1,2:n} \\ 0 & R_{2:n,2:n} \end{bmatrix}$$

$$= \begin{bmatrix} R^2_{11} & R_{11}R_{1,2:n} \\ R_{11}R^T_{1,2:n} & R^T_{1,2:n}R_{1,2:n} + R^T_{2:n,2:n}R_{2:n,2:n} \end{bmatrix}$$

1. compute first row of R:

$$R_{11} = \sqrt{A_{11}}, \qquad R_{1,2:n} = \frac{1}{R_{11}}A_{1,2:n} \qquad \boxed{A_{11} > 0}$$

if A is positive definite

2. compute $2,2$ block $R_{2:n,2:n}$ from

$$A_{2:n,2:n} - R^T_{1,2:n}R_{1,2:n} = R^T_{2:n,2:n}R_{2:n,2:n} \boxed{= A_{2:n,2:n} - \frac{1}{A_{11}}A_{2:n,1}A^T_{2:n,1}}$$

this is a Cholesky factorization of order $n-1$

Cholesky factorization algorithm

- Example

$$\begin{bmatrix} 25 & 15 & -5 \\ 15 & 18 & 0 \\ -5 & 0 & 11 \end{bmatrix} = \begin{bmatrix} R_{11} & 0 & 0 \\ R_{12} & R_{22} & 0 \\ R_{13} & R_{23} & R_{33} \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ 0 & R_{22} & R_{23} \\ 0 & 0 & R_{33} \end{bmatrix}$$

- first row of R

$$\begin{bmatrix} 25 & 15 & -5 \\ 15 & 18 & 0 \\ -5 & 0 & 11 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 0 \\ 3 & R_{22} & 0 \\ -1 & R_{23} & R_{33} \end{bmatrix} \begin{bmatrix} 5 & 3 & -1 \\ 0 & R_{22} & R_{23} \\ 0 & 0 & R_{33} \end{bmatrix}$$

- second row of R

$$\begin{bmatrix} 18 & 0 \\ 0 & 11 \end{bmatrix} - \begin{bmatrix} 3 \\ -1 \end{bmatrix} \begin{bmatrix} 3 & -1 \end{bmatrix} = \begin{bmatrix} R_{22} & 0 \\ R_{23} & R_{33} \end{bmatrix} \begin{bmatrix} R_{22} & R_{23} \\ 0 & R_{33} \end{bmatrix}$$

$$\begin{bmatrix} 9 & 3 \\ 3 & 10 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 1 & R_{33} \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 0 & R_{33} \end{bmatrix}$$

- third column of R: $10 - 1 = R_{33}^2$, *i.e.*, $R_{33} = 3$

Singular Value Decomposition (SVD)

- Handy mathematical technique that has application to many problems
- Given any $m \times n$ matrix A, algorithm to find matrices U, V, and $\sum$ such that

$$\boldsymbol{A = U \sum V^T}$$

U is $m \times n$ and orthonormal
$\sum$ is $n \times n$ and diagonal
V is $n \times n$ and orthonormal

Singular Value Decomposition

The SVD is a factorization of a $m \times n$ matrix into

$$\boldsymbol{A} = \boldsymbol{U}\,\boldsymbol{\Sigma}\,\boldsymbol{V}^{\boldsymbol{T}}$$

where $\boldsymbol{U}$ is a $m \times m$ orthogonal matrix, $\boldsymbol{V}^{\boldsymbol{T}}$ is a $n \times n$ orthogonal matrix and $\boldsymbol{\Sigma}$ is a $m \times n$ diagonal matrix.

For a square matrix ($\boldsymbol{m} = \boldsymbol{n}$):

$$\sigma_1 \geq \sigma_2 \geq \sigma_3 \ldots$$

$$\boldsymbol{A} = \begin{pmatrix} \vdots & \cdots & \vdots \\ \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n \\ \vdots & \cdots & \vdots \end{pmatrix} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix} \begin{pmatrix} \cdots & \mathbf{v}_1^T & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & \mathbf{v}_n^T & \cdots \end{pmatrix}$$

$$\boldsymbol{A} = \begin{pmatrix} \vdots & \cdots & \vdots \\ \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n \\ \vdots & \cdots & \vdots \end{pmatrix} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix} \begin{pmatrix} \vdots & \cdots & \vdots \\ \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n \\ \vdots & \cdots & \vdots \end{pmatrix}^T$$

SVD

- The $\sum_i$ are called the singular values of A
- If A is singular, some of the $\sum_i$ will be 0
- In general *rank*(A) = number of nonzero $\sum_i$
- SVD is mostly unique (up to permutation of singular values, or if some $\sum_i$ are equal)

Reduced SVD

What happens when $\boldsymbol{A}$ is not a square matrix?

1) $\boldsymbol{m} > \boldsymbol{n}$

$$\boldsymbol{A} = \boldsymbol{U}\,\boldsymbol{\Sigma}\,\boldsymbol{V}^{\boldsymbol{T}} = \underbrace{\left(\boxed{\begin{matrix} \vdots & \cdots & \vdots \\ \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n \\ \vdots & \cdots & \vdots \end{matrix}} \begin{matrix} \cdots & \vdots \\ \cdots & \boldsymbol{u}_m \\ \cdots & \vdots \end{matrix}\right)}_{m\times m} \underbrace{\begin{pmatrix} \boxed{\begin{matrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{matrix}} \\ \begin{matrix} & & 0 \\ & & \vdots \\ & & 0 \end{matrix} \end{pmatrix}}_{m\times n} \underbrace{\begin{pmatrix} \cdots & \mathbf{v}_1^T & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & \mathbf{v}_n^T & \cdots \end{pmatrix}}_{n\times n}$$

We can instead re-write the above as:

$$\boldsymbol{A} = \boldsymbol{U}_{\boldsymbol{R}}\,\boldsymbol{\Sigma}_{\boldsymbol{R}}\boldsymbol{V}^{\boldsymbol{T}}$$

Where $\boldsymbol{U}_{\boldsymbol{R}}$ is a $m\times n$ matrix and $\boldsymbol{\Sigma}_{\boldsymbol{R}}$ is a $n\times n$ matrix

Reduced SVD

2) $n > m$

$$A = U\,\Sigma\,V^T = \begin{pmatrix} \vdots & \cdots & \vdots \\ u_1 & \cdots & u_m \\ \vdots & \cdots & \vdots \end{pmatrix} \begin{pmatrix} \boxed{\begin{matrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_m \end{matrix}} & 0 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix} \begin{pmatrix} \boxed{\begin{matrix} \cdots & \mathbf{v}_1^T & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & \mathbf{v}_m^T & \cdots \end{matrix}} \\ \begin{matrix} \vdots & \vdots & \vdots \\ \cdots & \mathbf{v}_n^T & \cdots \end{matrix} \end{pmatrix}$$

$m \times m$ $\qquad$ $m \times n$ $\qquad$ $n \times n$

We can instead re-write the above as:

$$A = U\,\Sigma_R {V_R}^T$$

where V_R is a $n \times m$ matrix and Σ_R is a $m \times m$ matrix

In general:

$$A = U_R \Sigma_R {V_R}^T$$

U_R is a $m \times k$ matrix
Σ_R is a $k \times k$ matrix
V_R is a $n \times k$ matrix

$k = \min(m, n)$

Reduced SVD

Let's take a look at the product $\mathbf{\Sigma}^T\mathbf{\Sigma}$, where $\mathbf{\Sigma}$ has the singular values of a $\boldsymbol{A}$, a $m\times n$ matrix.

$$\mathbf{\Sigma}^T\mathbf{\Sigma} = \underset{n\times m}{\begin{pmatrix} \sigma_1 & & & 0 & & \\ & \ddots & & & \ddots & \\ & & \sigma_n & & & 0 \end{pmatrix}} \underset{m\times n}{\begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ & & 0 \\ & & \vdots \\ & & 0 \end{pmatrix}} = \underset{n\times n}{\boxed{\begin{pmatrix} {\sigma_1}^2 & & \\ & \ddots & \\ & & {\sigma_n}^2 \end{pmatrix}}}$$

$m > n$

$$\mathbf{\Sigma}^T\mathbf{\Sigma} = \underset{n\times m}{\begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_m \\ & & 0 \\ & & \vdots \\ & & 0 \end{pmatrix}} \underset{m\times n}{\begin{pmatrix} \sigma_1 & & & 0 & & \\ & \ddots & & & \ddots & \\ & & \sigma_m & & & 0 \end{pmatrix}} = \underset{n\times n}{\begin{pmatrix} \boxed{\begin{matrix} {\sigma_1}^2 & & \\ & \ddots & \\ & & {\sigma_m}^2 \end{matrix}} & \begin{matrix} 0 & & \\ & \ddots & \\ & & 0 \end{matrix} \\ \begin{matrix} 0 & & 0 \\ & \ddots & \\ & & 0 \end{matrix} & \begin{matrix} & \ddots & \\ & & 0 \end{matrix} \end{pmatrix}}$$

$n > m$

SVD

Assume $\boldsymbol{A}$ with the singular value decomposition $\boldsymbol{A} = \boldsymbol{U}\,\boldsymbol{\Sigma}\,\boldsymbol{V}^{\boldsymbol{T}}$. Let's take a look at the eigenpairs corresponding to $\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A}$:

$$\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A} = \left(\boldsymbol{U}\,\boldsymbol{\Sigma}\,\boldsymbol{V}^{\boldsymbol{T}}\right)^{\boldsymbol{T}}\left(\boldsymbol{U}\,\boldsymbol{\Sigma}\,\boldsymbol{V}^{\boldsymbol{T}}\right)$$

$$\left(\boldsymbol{V}^{\boldsymbol{T}}\right)^{\boldsymbol{T}}(\,\boldsymbol{\Sigma}\,)^{\boldsymbol{T}}\boldsymbol{U}^{\boldsymbol{T}}\left(\boldsymbol{U}\,\boldsymbol{\Sigma}\,\boldsymbol{V}^{\boldsymbol{T}}\right) = \boldsymbol{V}\boldsymbol{\Sigma}^{\boldsymbol{T}}\boldsymbol{U}^{\boldsymbol{T}}\,\boldsymbol{U}\,\boldsymbol{\Sigma}\,\boldsymbol{V}^{\boldsymbol{T}} = \boldsymbol{V}\,\boldsymbol{\Sigma}^{\boldsymbol{T}}\boldsymbol{\Sigma}\,\boldsymbol{V}^{\boldsymbol{T}}$$

Hence $\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A} = \boldsymbol{V}\,\boldsymbol{\Sigma}^{\boldsymbol{2}}\,\boldsymbol{V}^{\boldsymbol{T}}$

Recall that columns of $\boldsymbol{V}$ are all linear independent (orthogonal matrix), then from diagonalization ($\boldsymbol{B} = \boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^{-\boldsymbol{1}}$), we get:

- the columns of $\boldsymbol{V}$ are the eigenvectors of the matrix $\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A}$
- The diagonal entries of $\boldsymbol{\Sigma}^{\boldsymbol{2}}$ are the eigenvalues of $\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A}$

Let's call λ the eigenvalues of $\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A}$, then ${\sigma_i}^2 = \lambda_i$

SVD

- In a similar way,

$$\boldsymbol{AA^T} = \left(\boldsymbol{U\,\Sigma\,V^T}\right)\left(\boldsymbol{U\,\Sigma\,V^T}\right)^{\boldsymbol{T}}$$

$$\left(\boldsymbol{U\,\Sigma\,V^T}\right)\left(\boldsymbol{V^T}\right)^{\boldsymbol{T}}(\,\boldsymbol{\Sigma}\,)^{\boldsymbol{T}}\boldsymbol{U^T} = \boldsymbol{U\,\Sigma\,}{\color{red}\boldsymbol{V^TV}}\boldsymbol{\Sigma^T U^T} = \boldsymbol{U\Sigma\,\Sigma^T U^T}$$

Hence $\boldsymbol{AA^T} = \boldsymbol{U\,\Sigma^2\,U^T}$

Recall that columns of $\boldsymbol{U}$ are all linear independent (orthogonal matrices), then from diagonalization ($\boldsymbol{B} = \boldsymbol{XDX^{-1}}$), we get:

- The columns of $\boldsymbol{U}$ are the eigenvectors of the matrix $\boldsymbol{AA^T}$

How can we compute an SVD of a matrix A?

1. Evaluate the n eigenvectors $\mathbf{v}_i$ and eigenvalues λ_i of $\boldsymbol{A}^T\boldsymbol{A}$
2. Make a matrix $\boldsymbol{V}$ from the normalized vectors $\mathbf{v}_i$. The columns are called "right singular vectors".

$$\boldsymbol{V} = \begin{pmatrix} \vdots & \cdots & \vdots \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ \vdots & \cdots & \vdots \end{pmatrix}$$

3. Make a diagonal matrix from the square roots of the eigenvalues.

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix} \quad \sigma_i = \sqrt{\lambda_i} \quad \text{and} \quad \sigma_1 \geq \sigma_2 \geq \sigma_3 \ldots$$

4. Find $\boldsymbol{U}$: $\boldsymbol{A} = \boldsymbol{U}\,\boldsymbol{\Sigma}\,\boldsymbol{V}^T \Rightarrow \boldsymbol{U}\,\boldsymbol{\Sigma} = \boldsymbol{A}\,\boldsymbol{V} \Rightarrow \boldsymbol{U} = \boldsymbol{A}\,\boldsymbol{V}\,\boldsymbol{\Sigma}^{-1}$. The columns are called the "left singular vectors".