GIAN Course on Solving Linear Systems and Computing Generalized Inverses Using Recurrent Neural Networks June 09-19, 2025, IIT Indore, (The Least Squares Problem and SVD)

Sk. Safique Ahmad

June 13, 2025

Second Step of the Algorithm

- The second step operates on the submatrix obtained by ignoring the first row and column.
- Otherwise, it is identical to the first step:
  - Compute and apply Householder reflector.
  - Identify pivot and possibly swap columns.
- When columns are interchanged, the **full columns** are swapped, not just the parts in the submatrix.
- This is equivalent to performing the interchange **before** the QR process starts.

Case: Matrix Has Full Rank

- If the matrix has full rank, the algorithm terminates after m steps.
- The result is a decomposition:

$$A\Pi = QR$$

- Where:
  - Π is a column permutation matrix,
  - $Q \in \mathbb{R}^{n\times n}$ is orthogonal,
  - $R \in \mathbb{R}^{n\times m}$ is upper triangular and nonsingular.

Case: Matrix Does Not Have Full Rank

- If A does not have full rank, at some step we will encounter $\tau_i = 0$.
- This occurs when all entries in the remaining submatrix are zero.
- Suppose this occurs after r steps.
- Let $Q_i \in \mathbb{R}^{n \times n}$ denote the reflector used at step i.

Structure of R and Reflectors

- Let $R_H \in \mathbb{R}^{r \times r}$ be the upper triangular part constructed from the first r steps.
- Then:
$$R = \begin{bmatrix} R_H & * \\ 0 & 0 \end{bmatrix}$$
- The diagonal entries of R_H are $-\tau_1, -\tau_2, \ldots, -\tau_r$, all nonzero.
- Clearly, $\text{rank}(R) = r$.

Final Form of the Decomposition

- Let:

$$Q = Q_1 Q_2 \cdots Q_r$$

- Then:

$$Q^T = Q_r Q_{r-1} \cdots Q_1$$

- Therefore:

$$Q^T A = R \quad \text{and} \quad A = QR$$

- Since $\text{rank}(A) = \text{rank}(R) = r$, we conclude:

$$\text{rank}(A) = r$$

Theorem 3.3.11

Theorem: Let $A \in \mathbb{R}^{n\times m}$ with rank $r > 0$. Then there exist matrices:

- $\Pi \in \mathbb{R}^{m\times m}$: a permutation matrix,
- $Q \in \mathbb{R}^{n\times n}$: orthogonal,
- $R \in \mathbb{R}^{n\times m}$: upper triangular,

such that:

$$A\Pi = QR$$

where:

$$R = \begin{bmatrix} R_H & * \\ 0 & 0 \end{bmatrix}, \quad R_H \in \mathbb{R}^{r\times r} \text{ is nonsingular.}$$

Least Squares Problem Setup

- Given $A \in \mathbb{R}^{n\times m}$, $b \in \mathbb{R}^n$, we seek $x \in \mathbb{R}^m$ that minimizes:

$$\|Ax - b\|_2$$

- If A has **full column rank**, the solution is unique.
- If A is **rank-deficient** (i.e., $\text{rank}(A) = r < m$), the problem has **infinitely many solutions**.

QR Decomposition with Column Pivoting

- Apply QR with column pivoting: $A\Pi = QR$
- $Q \in \mathbb{R}^{n\times n}$: orthogonal
- $R = \begin{bmatrix} R_1 & R_2 \\ 0 & 0 \end{bmatrix}$, where:
  - $R_1 \in \mathbb{R}^{r\times r}$ is upper triangular and nonsingular,
  - $r = \text{rank}(A) < m$
- $\Pi \in \mathbb{R}^{m\times m}$: permutation matrix

Reduced Least Squares System

- Let $y = \Pi^T x$, then:

$$QRy \approx b \Rightarrow Ry \approx Q^T b$$

- Partition $y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$, where:

$$R = \begin{bmatrix} R_1 & R_2 \\ 0 & 0 \end{bmatrix}, \quad y_1 \in \mathbb{R}^r, \quad y_2 \in \mathbb{R}^{m-r}$$

- Solve:

$$R_1 y_1 = (Q^T b)_{1:r}$$

- y_2 is **free** (arbitrary) $\Rightarrow$ infinite solutions.

General Solution Form

- General solution to the least squares problem:

$$x = \Pi \begin{bmatrix} R_1^{-1}(Q^T b)_{1:r} \\ \text{free vector } y_2 \end{bmatrix}$$

- The solution set forms an affine subspace:

$$x = x_{\text{particular}} + \text{null}(A)$$

- The set of solutions is infinite due to $\dim(\text{null}(A)) = m - r > 0$

Minimum Norm Solution

- Among infinite solutions, one may choose the one with minimum $\|x\|_2$
- This is called the **minimum norm least squares solution**:

$$x_{\min} = A^{\dagger} b$$

  where $A^{\dagger}$ is the Moore–Penrose pseudoinverse.
- In QR terms:

$$x_{\min} = \Pi \begin{bmatrix} R_1^{-1}(Q^T b)_{1:r} \\ 0 \end{bmatrix}$$

MATLAB Code: Infinite Least Squares Solutions

```
% Rank-deficient matrix A and vector b
A = [1 2 3 4; 2 4 6 8; 3 6 9 12];  % rank 2
b = [1; 2; 3];

% QR decomposition with column pivoting
[Q, R, P] = qr(A, 'vector');

% Determine rank numerically
tol = max(size(A)) * eps(norm(R, 'fro'));
r = sum(abs(diag(R)) > tol);

% Solve R1 * y1 = Q^T * b (first r components)
R1 = R(1:r, 1:r);
Qt_b = Q' * b;
b1 = Qt_b(1:r);
y1 = R1 \ b1;
```

```
x = zeros(size(A,2), 1); x(P) = y;
disp('Minimum-norm solution:'); disp(x);
```

Classical Gram-Schmidt Algorithm

- Given linearly independent vectors $v_1, v_2, \ldots, v_m \in \mathbb{R}^n$
- Produces orthonormal vectors $q_1, q_2, \ldots, q_m$ such that:

$$\text{span}\{q_1, \ldots, q_i\} = \text{span}\{v_1, \ldots, v_i\}, \quad \text{for } i = 1, \ldots, m$$

- Algorithm:

$$q_1 = \frac{v_1}{\|v_1\|}$$

$$\text{for } k = 2 \text{ to } m$$

$$\quad \text{for } j = 1 \text{ to } k-1$$

$$\qquad r_{jk} = q_j^\top v_k$$

$$\qquad v_k = v_k - r_{jk} q_j$$

$$\quad \text{end}$$

$$\quad r_{kk} = \|v_k\|, \quad q_k = \frac{v_k}{r_{kk}}$$

end

Gram-Schmidt as QR Decomposition

- Let $A = [v_1, v_2, \ldots, v_m] \in \mathbb{R}^{n\times m}$
- The Gram-Schmidt process gives:

$$A = QR$$

  where:
  - $Q = [q_1, q_2, \ldots, q_m]$ is orthonormal ($Q^\top Q = I$)
  - R is upper triangular with entries $r_{jk} = q_j^\top v_k$
- Each v_k can be written as:

$$v_k = \sum_{j=1}^{k} r_{jk} q_j$$

- In matrix form:

$$A = QR \quad \text{(Gram-Schmidt gives QR)}$$

MATLAB Code for Classical Gram-Schmidt

```
function [Q, R] = classical_gs(A)
    [n, m] = size(A);
    Q = zeros(n, m);
    R = zeros(m, m);

    for k = 1:m
        v = A(:,k);
        for j = 1:k-1
            R(j,k) = Q(:,j)' * A(:,k);
            v = v - R(j,k) * Q(:,j);
        end
        R(k,k) = norm(v);
        Q(:,k) = v / R(k,k);
    end
end
```

Summary

- Classical Gram-Schmidt orthogonalizes vectors sequentially.
- It is numerically unstable for nearly linearly dependent vectors.
- The resulting decomposition $A = QR$ links the process directly to matrix factorization.
- Modified Gram-Schmidt improves numerical stability.

Modified Gram-Schmidt Overview

- Modified Gram-Schmidt (MGS) is a numerically more stable variant of the classical method.
- Orthogonalizes column by column using updated vectors.
- Better handles near-linear dependence in columns.
- Produces $A = QR$ where:
  - Q: orthonormal columns
  - R: upper triangular matrix

MATLAB Code: Modified Gram-Schmidt

```
function [Q, R] = modified_gs(A)
    [n, m] = size(A);
    Q = zeros(n, m);
    R = zeros(m, m);
    V = A;

    for i = 1:m
        R(i,i) = norm(V(:,i));
        Q(:,i) = V(:,i) / R(i,i);
        for j = i+1:m
            R(i,j) = Q(:,i)' * V(:,j);
            V(:,j) = V(:,j) - R(i,j) * Q(:,i);
        end
    end
end
```

Example Matrix

- Let

$$A = \begin{bmatrix} 1 & 1 \\ 10^{-10} & 0 \\ 0 & 10^{-10} \end{bmatrix}$$

- Columns of A are nearly linearly dependent
- Classical Gram-Schmidt fails due to loss of orthogonality
- Modified Gram-Schmidt maintains orthogonality

Numerical Stability Comparison

- Compute $Q^\top Q$:
  - Classical GS: $Q^\top Q \neq I$ (orthogonality lost)
  - Modified GS: $Q^\top Q \approx I$
- Stability matters in ill-conditioned problems
- Use ‘norm(Q’*Q - eye(size(Q,2)))‘ in MATLAB to test

Remark

- Modified Gram-Schmidt is more stable than the classical version.
- Especially important when vectors are nearly linearly dependent.
- For numerical work, prefer MGS or Householder QR over classical GS.

Given Vectors

- $\mathbf{v}_1 = \begin{bmatrix} 3 \\ -3 \\ 3 \\ -3 \end{bmatrix}$
- $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$
- Define $S = \text{span}\{\mathbf{v}_1, \mathbf{v}_2\} \subset \mathbb{R}^4$

Step (a): Gram-Schmidt Process

$$\mathbf{u}_1 = \mathbf{v}_1$$

$$r_{11} = \|\mathbf{u}_1\| = \sqrt{3^2 + (-3)^2 + 3^2 + (-3)^2} = \sqrt{36} = 6$$

$$\mathbf{q}_1 = \frac{\mathbf{u}_1}{r_{11}} = \frac{1}{6}\begin{bmatrix} 3 \\ -3 \\ 3 \\ -3 \end{bmatrix} = \begin{bmatrix} 0.5 \\ -0.5 \\ 0.5 \\ -0.5 \end{bmatrix}$$

Continue Gram-Schmidt

$$r_{12} = \mathbf{q}_1^T \mathbf{v}_2 = [0.5, -0.5, 0.5, -0.5] \cdot [1, 2, 3, 4]^T$$
$$= 0.5(1) + (-0.5)(2) + 0.5(3) + (-0.5)(4) = -1$$

$$\mathbf{u}_2 = \mathbf{v}_2 - r_{12}\mathbf{q}_1 = \mathbf{v}_2 + \mathbf{q}_1 = \begin{bmatrix} 1.5 \\ 1.5 \\ 3.5 \\ 3.5 \end{bmatrix}$$

$$r_{22} = \|\mathbf{u}_2\| = \sqrt{1.5^2 + 1.5^2 + 3.5^2 + 3.5^2} = \sqrt{32} = 4\sqrt{2}$$

$$\mathbf{q}_2 = \frac{\mathbf{u}_2}{r_{22}} = \frac{1}{4\sqrt{2}} \begin{bmatrix} 1.5 \\ 1.5 \\ 3.5 \\ 3.5 \end{bmatrix}$$

Step (b): Construct Q and R

$$Q = \begin{bmatrix} 0.5 & \frac{1.5}{4\sqrt{2}} \\ -0.5 & \frac{1.5}{4\sqrt{2}} \\ 0.5 & \frac{3.5}{4\sqrt{2}} \\ -0.5 & \frac{3.5}{4\sqrt{2}} \end{bmatrix}, \quad R = \begin{bmatrix} 6 & -1 \\ 0 & 4\sqrt{2} \end{bmatrix}$$

Then, $V = QR$

QR Decomposition Method Comparison

- Let $V \in \mathbb{R}^{30\times 20}$ with entries:

$$V(i,j) = \left(\frac{j}{20}\right)^{i-1}, \quad i = 1,\ldots,30,\; j = 1,\ldots,20$$

- Such matrices are called **Vandermonde matrices**.
- Highly ill-conditioned:

$$\kappa_2(V) \approx 3 \times 10^{13}$$

- Indicates columns are nearly linearly dependent.

Numerical Experiment

- Goal: Compare orthogonality of Q from QR decomposition methods.
- Metric: $\|I - Q^\top Q\|_2$
- IEEE double-precision unit roundoff: $u \approx 10^{-16}$
- Expect error for stable methods: $\approx \kappa(V) \cdot u \approx 3 \times 10^{-3}$

QR Method Comparison on Vandermonde Matrix

- Comparison of orthogonality error $\|I - Q^\top Q\|_2$:

Method	$\|I - Q^\top Q\|_2$
Classical Gram-Schmidt	12.4
Modified Gram-Schmidt	$\approx 3 \times 10^{-4}$
Householder QR (Reflectors)	$\approx 1.9 \times 10^{-15}$

Table: *

QR decomposition results for highly ill-conditioned Vandermonde matrix

Remark

- **Classical Gram-Schmidt:** fails in preserving orthogonality for ill-conditioned matrices.
- **Modified Gram-Schmidt:** more stable but still sensitive to ill-conditioning.
- **Householder QR:** highly stable, preferred in practice.
- Recommendation: **Use Householder QR** or SVD for high-accuracy applications.

Singular Value Decomposition (SVD), Moore Penrose Inverse

Bases and Matrices in the SVD

The Singular Value Decomposition is a highlight of linear algebra. A is any $m \times n$ matrix, square or rectangular. Its rank is r. We will diagonalize this A, but not by $X^{-1}AX$.

The eigenvectors in X have three big problems:

- They are usually not orthogonal,
- there are not always enough eigenvectors, and $Ax = \lambda x$ requires A to be a square matrix.
- The singular vectors of A solve all those problems in a perfect way.

Theorem 4.1.1 (SVD Theorem)

Let $A \in \mathbb{R}^{n \times m}$ be a nonzero matrix of rank r. Then:
Singular Value Decomposition (SVD) There exist orthogonal matrices:

$$U \in \mathbb{R}^{n \times n}, \quad V \in \mathbb{R}^{m \times m}$$

and a "diagonal" matrix:

$$\Sigma \in \mathbb{R}^{n \times m}$$

such that:

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

Structure of the SVD

- $U = [\mathbf{u}_1 \ \mathbf{u}_2 \ \ldots \ \mathbf{u}_n]$ with $U^\top U = I_n$
- $V = [\mathbf{v}_1 \ \mathbf{v}_2 \ \ldots \ \mathbf{v}_m]$ with $V^\top V = I_m$
- Σ has the form:

$$\Sigma = \begin{bmatrix} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \ddots & & \\ & & & \sigma_r & \\ & & & & \mathbf{0} \end{bmatrix}$$

where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ are the singular values of A.

Geometric Interpretation

- A maps the orthonormal basis vectors of $\mathbb{R}^m$ (columns of V) to scaled orthogonal vectors in $\mathbb{R}^n$ (columns of U).
- Each σ_i represents the stretching factor along the direction $\mathbf{v}_i$.
- The rank r of A equals the number of nonzero singular values.

Remark

- Every real matrix $A \in \mathbb{R}^{n \times m}$ has a Singular Value Decomposition.
- $A = U\Sigma V^\top$, where:
  - $U \in \mathbb{R}^{n \times n}$ and $V \in \mathbb{R}^{m \times m}$ are orthogonal,
  - $\Sigma \in \mathbb{R}^{n \times m}$ is diagonal with singular values.
- The SVD is a fundamental tool in numerical linear algebra, data compression, and PCA.

Theorem (Geometric SVD Theorem)

Let $A \in \mathbb{R}^{n \times m}$ be a nonzero matrix of rank r. Then:

- There exists an orthonormal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ of $\mathbb{R}^m$
- And an orthonormal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ of $\mathbb{R}^n$
- And singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$

Such that:

$$A\mathbf{v}_i = \sigma_i \mathbf{u}_i \quad \text{for } i = 1, \ldots, r$$

$$A\mathbf{v}_i = \mathbf{0} \quad \text{for } i = r+1, \ldots, m$$

$$A^T\mathbf{u}_i = \sigma_i \mathbf{v}_i \quad \text{for } i = 1, \ldots, r$$

$$A^T\mathbf{u}_i = \mathbf{0} \quad \text{for } i = r+1, \ldots, n$$

Geometric Interpretation

- A maps the unit vectors $\mathbf{v}_i$ in $\mathbb{R}^m$ to scaled orthogonal vectors $\sigma_i \mathbf{u}_i$ in $\mathbb{R}^n$.
- The first r directions are scaled by $\sigma_i > 0$, and the rest are mapped to 0.
- The image of the unit sphere in $\mathbb{R}^m$ under A is a hyperellipse in $\mathbb{R}^n$.

Exercise 4.1.5: From Algebraic SVD to Geometric SVD

Let $A = U\Sigma V^\top$ be the SVD of A, where:

- Columns of $V = [\mathbf{v}_1, \ldots, \mathbf{v}_m]$
- Columns of $U = [\mathbf{u}_1, \ldots, \mathbf{u}_n]$

Then:

$$A\mathbf{v}_i = U\Sigma V^\top \mathbf{v}_i = \sigma_i \mathbf{u}_i \quad (i = 1, \ldots, r)$$

- The matrix multiplication $AV = U\Sigma$ implies that each $\mathbf{v}_i$ is mapped to $\sigma_i \mathbf{u}_i$.
- When $\sigma_i = 0$, $A\mathbf{v}_i = \mathbf{0}$.

Thus, the geometric form follows directly from the standard SVD expression.

SVD and the Four Fundamental Subspaces

The SVD provides orthonormal bases for the fundamental subspaces of $A \in \mathbb{R}^{n\times m}$.

- $\mathcal{R}(A) = \text{Col}(A) = \text{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_r\} \subseteq \mathbb{R}^n$
- $\mathcal{N}(A) = \text{Null}(A) = \text{span}\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_m\} \subseteq \mathbb{R}^m$
- $\mathcal{R}(A^\top) = \text{Row}(A) = \text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_r\} \subseteq \mathbb{R}^m$
- $\mathcal{N}(A^\top) = \text{Left Null Space} = \text{span}\{\mathbf{u}_{r+1}, \ldots, \mathbf{u}_n\} \subseteq \mathbb{R}^n$

These follow directly from the SVD:

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

Corollary: Rank-Nullity Relation

Corollary 4.1.9 Let $A \in \mathbb{R}^{n\times m}$. Then:

$$\dim(\mathcal{R}(A)) + \dim(\mathcal{N}(A)) = m$$

- That is, the sum of the dimensions of the column space and the null space equals the number of columns.
- This result, also known as the **Rank-Nullity Theorem**, follows from the orthogonality and completeness of the columns of $V \in \mathbb{R}^{m\times m}$.
- Similarly, $\dim(\mathcal{R}(A^\top)) + \dim(\mathcal{N}(A^\top)) = n$.

(a) Structure of Rank-1 Matrix

Proof: Let $A \in \mathbb{R}^{n\times m}$ have rank 1.

- Then all columns of A lie in $\text{Range}(A) = \text{span}(\mathbf{u}_1)$.
- Choose $\|\mathbf{u}_1\| = 1$, then $A = \mathbf{u}_1\mathbf{w}^T$ for some $\mathbf{w} \in \mathbb{R}^m$.
- Define $\mathbf{v}_1 = \frac{\mathbf{w}}{\|\mathbf{w}\|}$, $\sigma_1 = \|\mathbf{w}\|$, so:

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T$$

(b) Orthonormal Extension

- Extend $\mathbf{u}_1$ to an orthonormal basis of $\mathbb{R}^n$: $U = [\mathbf{u}_1 \ \cdots] \in \mathbb{R}^{n\times n}$
- Similarly, extend $\mathbf{v}_1$ to $V = [\mathbf{v}_1 \ \cdots] \in \mathbb{R}^{m\times m}$
- Define:

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \cdots \\ 0 & 0 & \cdots \\ \vdots & \vdots & \end{bmatrix}_{n\times m} \Rightarrow A = U\Sigma V^T$$

- This gives the SVD of a rank-1 matrix.

(c) Leading Singular Value and Vector

Let $A \in \mathbb{R}^{n \times m}$, $\text{rank}(A) = r > 1$

- Let $\mathbf{v}_1$ maximize $\|A\mathbf{v}\|_2$ over unit vectors.
- Then $\mathbf{u}_1 = \frac{A\mathbf{v}_1}{\|A\mathbf{v}_1\|}$, and define:

$$\sigma_1 = \|A\mathbf{v}_1\| = \|A\|_2$$

- Let $U = [\mathbf{u}_1 \ \cdots]$, $V = [\mathbf{v}_1 \ \cdots]$, define:

$$B = U^T A V = \begin{bmatrix} \sigma_1 & \mathbf{z}^T \\ \mathbf{0} & A_1 \end{bmatrix} \Rightarrow A = UBV^T$$

(d) Showing $\mathbf{z} = \mathbf{0}$

- Suppose $B = \begin{bmatrix} \sigma_1 & \mathbf{z}^T \\ 0 & A_1 \end{bmatrix}$
- Take $\mathbf{x} = \begin{bmatrix} \cos\theta \\ \sin\theta\mathbf{w} \end{bmatrix}$, $\|\mathbf{w}\| = 1$
- Then $\|B\mathbf{x}\|^2 = \sigma_1^2 \cos^2\theta + \|\mathbf{z}^T\mathbf{w}\|^2 \sin^2\theta + \|A_1\mathbf{w}\|^2 \sin^2\theta$
- Since σ_1 is the largest singular value, optimization implies $\mathbf{z} = 0$

(e) Completing the SVD Inductively

- Since $\mathbf{z} = 0$, $B = \begin{bmatrix} \sigma_1 & 0 \\ 0 & A_1 \end{bmatrix}$
- $\text{rank}(A_1) = r - 1$. By induction, SVD of $A_1 = U_1 \Sigma_1 V_1^T$
- Embed into full SVD of A as:
$$A = U \begin{bmatrix} \sigma_1 & 0 \\ 0 & \Sigma_1 \end{bmatrix} V^T$$
- This gives an SVD for general $A \in \mathbb{R}^{n \times m}$ of rank r.

Theorem 4.1.10: Condensed SVD

Condensed Singular Value Decomposition (SVD)
Let $A \in \mathbb{R}^{n\times m}$ be a nonzero matrix of rank r. Then:
Condensed SVD Form There exist:

- $U_r \in \mathbb{R}^{n\times r}$ with orthonormal columns ($U_r^\top U_r = I_r$)
- $V_r \in \mathbb{R}^{m\times r}$ with orthonormal columns ($V_r^\top V_r = I_r$)
- A diagonal matrix $\Sigma_r \in \mathbb{R}^{r\times r}$ with positive entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$

such that:

$$A = U_r \Sigma_r V_r^\top$$

Exercise: Proving the Condensed SVD

- Start from the full SVD:
$$A = U\Sigma V^\top$$
where $U \in \mathbb{R}^{n\times n}$, $\Sigma \in \mathbb{R}^{n\times m}$, $V \in \mathbb{R}^{m\times m}$
- Partition as:
$$U = [U_r \;\; \bar{U}], \quad \Sigma = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix}, \quad V = [V_r \;\; \bar{V}]$$
- Then:
$$A = U\Sigma V^\top = U_r \Sigma_r V_r^\top$$
by removing the zero blocks.

Example: Compute the SVD of $A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}$

The rank is $r = 2$, so A has two positive singular values σ_1 and σ_2. We will find:
- $\sigma_1 > \lambda_{\text{max}} = 5$ - $\sigma_2 < \lambda_{\text{min}} = 3$
Begin by computing $A^T A$ and AA^T:

$$A^T A = \begin{bmatrix} 3 & 4 \\ 0 & 5 \end{bmatrix}^T \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} = \begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix}, \quad AA^T = \begin{bmatrix} 9 & 12 \\ 12 & 41 \end{bmatrix}$$

Both matrices have the same trace (50) and determinant (225). Their eigenvalues are:

$$\sigma_1^2 = 45, \quad \sigma_2^2 = 5 \Rightarrow \sigma_1 = \sqrt{45}, \quad \sigma_2 = \sqrt{5}$$

Then $\sigma_1 \sigma_2 = 15$, which is the determinant of A.
Now, we find the eigenvectors of $A^T A$:
For $\sigma_1^2 = 45$:

$$\begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 45 \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Rightarrow v_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

For $\sigma_2^2 = 5$:

$$\begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 5 \begin{bmatrix} 1 \\ -1 \end{bmatrix} \Rightarrow v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Then we compute $u_i = \frac{Av_i}{\sigma_i}$ to get the columns of U.
This gives the full SVD: $A = U\Sigma V^T$.
The right singular vectors are:

$$v_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Now compute:

$$Av_1 = A \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 3 \\ 4 \end{bmatrix} + \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 5 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 3 \\ 9 \end{bmatrix} = \sqrt{45} \cdot \frac{1}{\sqrt{10}} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \sigma_1 u_1$$

$$Av_2 = A \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 3 \\ 4 \end{bmatrix} - \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 5 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 3 \\ -1 \end{bmatrix} = \sqrt{5} \cdot \frac{1}{\sqrt{10}} \begin{bmatrix} 3 \\ -1 \end{bmatrix} = \sigma_2 u_2$$

This gives:

$$u_1 = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \quad u_2 = \frac{1}{\sqrt{10}} \begin{bmatrix} 3 \\ -1 \end{bmatrix}$$

The singular value decomposition is:

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

Where:

$$U = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 & 3 \\ 3 & -1 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sqrt{45} & 0 \\ 0 & \sqrt{5} \end{bmatrix}, \quad V = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{7}$$

U and V contain orthonormal bases for the column space and row space of A. These bases diagonalize A:

$$AV = U\Sigma \Rightarrow U^T AV = \Sigma$$

A as a Sum of Rank-One Matrices

$$\sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T = \sqrt{45} \cdot \frac{1}{\sqrt{10}} \begin{bmatrix} 1 \\ 3 \end{bmatrix} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \end{bmatrix} + \sqrt{5} \cdot \frac{1}{\sqrt{10}} \begin{bmatrix} 3 \\ -1 \end{bmatrix} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \end{bmatrix}$$

$$= \frac{\sqrt{45}}{\sqrt{20}} \begin{bmatrix} 1 \\ 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} + \frac{\sqrt{5}}{\sqrt{20}} \begin{bmatrix} 3 \\ -1 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} = A$$

Consider:

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Observations:

- All eigenvalues of A are 0.
- Only one eigenvector: $(1, 0, 0, 0)^T$.
- Singular values: $\sigma = 3, 2, 1, 0$
- Singular vectors are columns of the identity matrix.

This example shows how the SVD provides much more structural insight than the eigen-decomposition, especially for non-symmetric or defective matrices.

SVD Setup

Let $A \in \mathbb{R}^{n\times m}$ with rank r, and let:

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

where:

- $U \in \mathbb{R}^{n\times n}$, $V \in \mathbb{R}^{m\times m}$ are orthogonal,
- $\Sigma = \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{n\times m}$
- $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$

Transforming the Problem

We want to solve the least squares problem:

$$\min_x \|Ax - b\|_2$$

Using the orthogonality of U and V, let:

$$c = U^T b, \quad y = V^T x$$

Then:

$$\|Ax - b\|_2 = \|U\Sigma V^T x - b\|_2 = \|\Sigma y - c\|_2$$

Minimizing the Residual

The residual becomes:

$$\|\Sigma y - c\|_2^2 = \sum_{i=1}^{r}(\sigma_i y_i - c_i)^2 + \sum_{i=r+1}^{n} c_i^2$$

This is minimized when:

$$y_i = \frac{c_i}{\sigma_i}, \quad i = 1, \ldots, r$$

$$y_{r+1}, \ldots, y_m \text{ arbitrary (do not affect residual)}$$

Minimum Norm Solution

To find the solution x with minimal $\|x\|_2$, we must minimize $\|y\|_2$.
This is achieved when:

$$y_{r+1} = \cdots = y_m = 0$$

Hence, the minimum-norm least-squares solution is:

$$x = Vy = \sum_{i=1}^{r} \frac{c_i}{\sigma_i} \mathbf{v}_i$$

or equivalently:

$$x = A^+ b$$

where $A^+ = V\Sigma^+ U^T$ is the Moore-Penrose pseudoinverse.

Moore-Penrose Pseudoinverse

Let $A \in \mathbb{R}^{n \times m}$ be a matrix of rank r, with SVD:

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

Then the Moore-Penrose pseudoinverse $A^\dagger$ is given by:

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

where $\Sigma^\dagger \in \mathbb{R}^{m \times n}$ is formed by:

$$\Sigma^\dagger = \begin{bmatrix} 1/\sigma_1 & & & \\ & \ddots & & \\ & & 1/\sigma_r & \\ & & & \mathbf{0} \end{bmatrix}$$

with $\sigma_1, \dots, \sigma_r > 0$ the nonzero singular values of A.

Exercise– Matrix Form of Pseudoinverse

Given the full SVD of A:

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

Then:

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

Because U is orthogonal:

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

This representation is exact and satisfies all four Moore-Penrose conditions.

Condensed Form of the SVD

Let $U_r \in \mathbb{R}^{n\times r}$, $V_r \in \mathbb{R}^{m\times r}$ denote the first r columns of U and V, and $\Sigma_r \in \mathbb{R}^{r\times r}$ the diagonal matrix of nonzero singular values.
Then:

$$A = U_r \Sigma_r V_r^T$$

$$A^\dagger = V_r \Sigma_r^{-1} U_r^T$$

This form is efficient and commonly used in numerical computation.

Why Use the Pseudoinverse?

- Solves the least-squares problem:

$$\min_x \|Ax - b\|_2 \Rightarrow x = A^\dagger b$$

- Works even when A is not full-rank.
- Provides the minimum-norm solution when there are infinitely many.
- The pseudoinverse is essential in data fitting, control theory, and machine learning.

Exercise: Moore-Penrose Characterization

Theorem: Let $A \in \mathbb{R}^{n\times m}$, and let $B \in \mathbb{R}^{m\times n}$. Then $B = A^{\dagger}$ if and only if:

$$\begin{aligned} &(1) \quad ABA = A \\ &(2) \quad BAB = B \\ &(3) \quad (AB)^T = AB \\ &(4) \quad (BA)^T = BA \end{aligned}$$

Proof Outline:

- If $B = A^{\dagger}$ from SVD, all four properties hold.
- Conversely, any matrix B satisfying these four conditions must be $A^{\dagger}$.

Setup: SVD of A

Let $A \in \mathbb{R}^{n\times m}$ with full column rank m, and let

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

be the singular value decomposition, where:

- $U \in \mathbb{R}^{n\times n}$, $U^T U = I_n$
- $V \in \mathbb{R}^{m\times m}$, $V^T V = I_m$
- $\Sigma = \begin{bmatrix} \text{diag}(\sigma_1, \ldots, \sigma_m) \\ \mathbf{0} \end{bmatrix} \in \mathbb{R}^{n\times m}$

SVDs of Related Matrices

Matrix	**SVD Expression**	**Singular Values**
A^TA	$V\Sigma^T\Sigma V^T$	σ_i^2
$(A^TA)^{-1}$	$V(\Sigma^T\Sigma)^{-1}V^T$	$1/\sigma_i^2$
$(A^TA)^{-1}A^T$	$V\Sigma^{-1}U^T$	$1/\sigma_i$
$A(A^TA)^{-1}$	$U\Sigma^{-1}V^T$	$1/\sigma_i$

Observations

- A^TA and $(A^TA)^{-1}$ are symmetric and positive definite.
- $(A^TA)^{-1}A^T = A^\dagger$: the Moore-Penrose pseudoinverse of A.
- $A(A^TA)^{-1}$ is the pseudoinverse of A^T.
- All SVDs use the same orthogonal matrices U and V, but scale differently.

Diagonal Structure of $A^T A$ and AA^T

We always start with $A^T A$ and AA^T. They are diagonal (with easy v's and u's):

$$A^T A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 9 \end{bmatrix}, \quad AA^T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 9 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The eigenvalues of $A^T A$ (and AA^T) are $\sigma^2 = 9, 4, 1$ (nonzero), corresponding to singular values $\sigma_1 = 3$, $\sigma_2 = 2$, $\sigma_3 = 1$.

Their corresponding orthonormal eigenvectors (in order of decreasing singular values) are:

$$U = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad V = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The first columns u_1 and v_1 have 1's in positions 3 and 4. Then the matrix $u_1 \sigma_1 v_1^T$ picks out the largest number in A, which is $A_{3,4} = 3$.
Thus the SVD of A is:

$$A = U \Sigma V^T = 3 u_1 v_1^T + 2 u_2 v_2^T + 1 u_3 v_3^T$$

Effect of Removing a Zero Row

Suppose we remove the last row of A (which is entirely zeros). Then A becomes a 3×4 matrix and AA^T becomes 3×3. Its fourth row and column disappear.
However, the eigenvalues of A^TA and AA^T remain the same: $\lambda = 1, 4, 9$, so the singular values are still $\sigma = 3, 2, 1$. We just remove the last row of Σ, and the last row and column of U:

$$A_{3\times4} = U_{3\times3}\Sigma_{3\times4}V_{4\times4}^T$$

The SVD naturally accommodates rectangular matrices.

Stability of Singular Values vs. Eigenvalue Instability

The 4×4 matrix A provides a powerful illustration of the instability of eigenvalues. Suppose the $(4, 1)$ entry of A is changed slightly—from 0 to $\frac{1}{60{,}000}$.
Consider the matrix:

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ \frac{1}{60{,}000} & 0 & 0 & 0 \end{bmatrix}$$

This small change in the $(4, 1)$ entry (only $1/60{,}000$) creates a much larger effect in the eigenvalues of A. Originally, with a zero in the $(4, 1)$ position, the eigenvalues of A were all zero:

$$\lambda = 0, 0, 0, 0$$

After the change, the eigenvalues move to four points on a circle in the complex plane centered at the origin, with radius $\frac{1}{10}$:

At the other extreme, when $A^T A = AA^T$ (i.e., A is a **normal matrix**), the eigenvectors of A are orthogonal, and the eigenvalues are completely stable.

Singular Values Are Stable.

By contrast, the singular values of A remain stable under small perturbations. In this case, the new singular values are:

$$\sigma = 3, 2, 1, \frac{1}{60{,}000}$$

The singular vectors (U and V) remain essentially unchanged. The fourth piece of the SVD is:

$$\sigma_4 u_4 v_4^T = \frac{1}{60{,}000} u_4 v_4^T$$

—mostly zeros, except for the new small entry.

Thank You !