

# Notes on Matrices and Linear Algebra

Albert Jojo

February 6, 2025

## 1 Introduction

Linear Algebra is a branch of mathematics concerning vector spaces and linear mappings between them. Matrices are used to represent and solve systems of linear equations, and they have numerous applications in fields like physics, engineering, computer science, and economics.

## 2 Matrix Algebra in Big Data Analysis

### 2.1 Definitions of Matrices and Vectors

- A **matrix** is an arrangement of numbers in rectangular form.

A  $(j \times k)$  matrix  $A$  can be written as:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{j1} & a_{j2} & \cdots & a_{jk} \end{bmatrix}$$

- A **square matrix** has equal numbers of rows and columns. A  $(2 \times 2)$  matrix is:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

- A **vector** is a special matrix with either one row or one column:

– **Row vector** (order  $1 \times k$ ):

$$\sigma = [\sigma_1 \quad \sigma_2 \quad \cdots \quad \sigma_k]$$

– **Column vector** (order  $k \times 1$ ):

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$$

## 2.2 Addition, Subtraction, Multiplication

- Matrix addition:  $A + B = C$  (element-wise addition)
- Commutative property:  $A + B = B + A$
- Associative property:  $(A + B) + C = A + (B + C)$

## 2.3 Transposition

For a matrix  $Q$  of order  $(3 \times 2)$ :

$$Q = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \\ q_{31} & q_{32} \end{bmatrix}, \quad Q^T = \begin{bmatrix} q_{11} & q_{21} & q_{31} \\ q_{12} & q_{22} & q_{32} \end{bmatrix}$$

## 2.4 Symmetric Matrices

A matrix  $A$  is **symmetric** if:

$$A = A^T$$

This requires  $A$  to be square and  $a_{ij} = a_{ji}$  for all  $i, j$ .

## 3 Matrix Properties

A matrix  $A$  of size  $m \times n$  has several important properties and operations associated with it. Some key properties are:

- Matrix multiplication is generally **not** commutative:  $AB \neq BA$ .
- Associative property:  $(AB)C = A(BC)$ .
- Distributive property:  $A(B + C) = AB + AC$ .
- Transpose of a product:  $(AB)^T = B^T A^T$ .
- Power of a matrix:  $A^n = A \times A \times \cdots \times A$  (with  $n$  factors of  $A$ ).

## 4 Diagonal and Identity Matrices

A **diagonal matrix** is a square matrix in which all the entries outside the main diagonal are zero. A diagonal matrix is generally written as:

$$D = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix}$$

The **identity matrix**  $I$  is a square matrix where all the elements of the principal diagonal are 1, and all other elements are 0:

$$I = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

The identity matrix acts as a multiplicative identity for matrices, meaning that for any matrix  $A$ , we have:

$$A \times I = A \quad \text{and} \quad I \times A = A$$

## 5 Echelon Form

An  $m \times n$  matrix is in **row echelon form** if it satisfies the following conditions:

1. The first non-zero element in each row, called the **leading entry**, is 1.
2. Each leading entry is to the right of the leading entry in the row just above.
3. Rows with all zero elements are at the bottom of the matrix.

A matrix is in **reduced row echelon form (RREF)** if it satisfies the following additional condition:

- Each leading 1 is the only non-zero entry in its column.

For example, the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 0 \end{bmatrix}$$

is in row echelon form but not in reduced row echelon form. We can apply row operations to reduce it to the following RREF:

$$A_{RREF} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 4 \\ 0 & 0 & 0 \end{bmatrix}$$

## 6 Invertibility of Diagonal Matrices

A diagonal matrix  $D$  is invertible if and only if all of its diagonal elements are non-zero. The inverse  $D^{-1}$  of a diagonal matrix is also a diagonal matrix, where each diagonal element is the reciprocal of the corresponding diagonal element in  $D$ .

For example, consider the diagonal matrix:

$$D = \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix}$$

Then, the inverse  $D^{-1}$  is:

$$D^{-1} = \begin{bmatrix} \frac{1}{d_1} & 0 & 0 \\ 0 & \frac{1}{d_2} & 0 \\ 0 & 0 & \frac{1}{d_3} \end{bmatrix}$$

This inverse exists only if  $d_1, d_2, d_3 \neq 0$ .

## 7 Matrix Multiplication

Matrix multiplication involves the dot product of rows of the first matrix with columns of the second matrix. The product of two matrices  $A = [a_{ij}]$  of size  $m \times n$  and  $B = [b_{ij}]$  of size  $n \times p$  is a matrix  $C = AB$  of size  $m \times p$ , where each element  $c_{ij}$  is computed as:

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}$$

Example: Let

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$$

Then, the product  $AB$  is:

$$AB = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} = \begin{bmatrix} 1 \cdot 5 + 2 \cdot 7 & 1 \cdot 6 + 2 \cdot 8 \\ 3 \cdot 5 + 4 \cdot 7 & 3 \cdot 6 + 4 \cdot 8 \end{bmatrix} = \begin{bmatrix} 19 & 22 \\ 43 & 50 \end{bmatrix}$$

## 8 More types of Matrices

Here are some other important types of matrices:

- **Upper-Triangular Matrix:** An upper triangular matrix is a square matrix in which all entries below the main diagonal are zero. That is, for a matrix  $A = [a_{ij}]$ ,  $a_{ij} = 0$  for all  $i > j$ . An example of an upper-triangular matrix is:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$$

- **Lower-Triangular Matrix:** A lower triangular matrix is a square matrix in which all entries above the main diagonal are zero. That is, for a matrix  $A = [a_{ij}]$ ,  $a_{ij} = 0$  for all  $i < j$ . An example of a lower-triangular matrix is:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 3 & 0 \\ 4 & 5 & 6 \end{bmatrix}$$

- **Symmetric Matrix:** A symmetric matrix is a square matrix that is equal to its transpose. That is,  $A = A^T$ . For example:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{bmatrix}$$

- **Skew-Symmetric Matrix:** A skew-symmetric matrix is a square matrix that is the negative of its transpose. That is,  $A = -A^T$ . For example:

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

- **Orthogonal Matrix:** A square matrix  $A$  is orthogonal if its rows and columns are orthonormal vectors. This means  $A^T A = A A^T = I$ , where  $I$  is the identity matrix. For example:

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

- **Zero Matrix:** A matrix in which every element is zero. For example:

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

## 9 Proofs

### 9.1 Invertibility

(1) Suppose  $A$  is an  $n \times n$  upper-triangular matrix, and the  $k$ -th entry on the main diagonal of  $A$  is zero, i.e.,

$$a_{kk} = 0.$$

**Solution:**

Consider the submatrix  $B$ , which is formed by taking the last  $n - k + 1$  rows of  $A$ . This submatrix  $B$  is of size  $(n - k + 1) \times n$ .

The key observation is that the  $k$ -th column of  $B$  consists entirely of zeros. Therefore, the rank of the submatrix  $B$  is at most  $n - k$ , because there are at most  $n - k$  linearly independent columns in  $B$ . In other words, we have:

$$\text{rank}(B) \leq n - k.$$

Since the rank of  $B$  is strictly less than  $n$ , it follows that  $B$  is not of full column rank. This implies that the rows of  $B$  are not linearly independent.

By the rank-nullity theorem, the rank of a matrix is equal to the number of linearly independent rows (or columns). Since  $A$  has a zero entry on the diagonal, the matrix  $A$  has fewer than  $n$  linearly independent rows and columns.

Therefore, the rank of  $A$  is less than  $n$ , which means  $A$  is **not full rank**, and thus:

$A$  is not invertible.

(2) Suppose  $A$  is not invertible.

This means that there exists a linear dependence among the rows of  $A$ . Specifically, let the  $k$ -th row of  $A$  be a linear combination of the other rows:

$$A_k = \sigma_1 A_1 + \sigma_2 A_2 + \cdots + \sigma_n A_n.$$

For all rows above  $A_k$ , the coefficients  $\sigma_1, \sigma_2, \dots, \sigma_{k-1}$  must be zero, because the rows above  $A_k$  are linearly independent of it. Therefore, we can express  $A_k$  as:

$$A_k = \sigma_{k+1} A_{k+1} + \cdots + \sigma_n A_n.$$

This implies that the entry  $a_{kk}$ , which corresponds to the diagonal element of  $A_k$ , must be zero because all the corresponding elements in columns  $k+1, \dots, n$  of the submatrix  $A_{k+1}, \dots, A_n$  are zero.

Thus, we conclude that  $a_{kk} = 0$ , which completes the proof.

## 10 Identity Matrix

- The identity matrix is a square, consisting of 1's along the diagonal and zeros elsewhere.

$$I_{n \times n} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

## 10.1 Properties

1.  $AI = IA = A$

**Example:**

Consider the matrix  $A$ :

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

Multiplying  $A$  by the identity matrix  $I$  (of size  $3 \times 3$ ) on either side gives:

$$AI = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = A$$

Similarly, for  $IA$ :

$$IA = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = A$$

**Zero Matrix:** A matrix that only consists of all zeros.

## 11 Determinant and Eigenvectors

- Determinants are defined only for square matrices.
- Example: For a  $2 \times 2$  matrix  $A$ , the determinant is calculated as:

$$\det(A) = |A| = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

- The determinant is a scalar value that encodes important information about the matrix, such as whether it is invertible (if the determinant is non-zero) or its volume scaling factor in geometric terms.

## 12 Cofactor and Determinant Calculation

- In general, we need to first define the cofactor  $Q_{r,s}$  for each element of matrix  $A$ . The cofactor of  $a_{r,s}$  is given by:

$$Q_{r,s} = (-1)^{r+s} |A_{r,s}|$$

where  $A_{r,s}$  is the matrix formed by deleting row  $r$  and column  $s$  from the original matrix  $A$ .

**Example:**

Let the matrix  $A$  be:

$$A = \begin{bmatrix} 1 & 3 & 2 \\ 4 & 5 & 6 \\ 8 & 7 & 9 \end{bmatrix}$$

To find the determinant of matrix  $A$  using cofactor expansion, we expand along the first row. The determinant is given by the formula:

$$\det(A) = \sum_{s=1}^n (-1)^{1+s} a_{1s} |A_{1s}|$$

Where  $A_{1s}$  denotes the submatrix formed by deleting the first row and the  $s$ -th column. We now expand along the first row, i.e., for  $s = 1, 2, 3$ :

$$\det(A) = (-1)^{1+1} a_{11} |A_{11}| + (-1)^{1+2} a_{12} |A_{12}| + (-1)^{1+3} a_{13} |A_{13}|$$

Substitute the values from matrix  $A$ :

$$\det(A) = 1 \cdot \begin{vmatrix} 5 & 6 \\ 7 & 9 \end{vmatrix} - 3 \cdot \begin{vmatrix} 4 & 6 \\ 8 & 9 \end{vmatrix} + 2 \cdot \begin{vmatrix} 4 & 5 \\ 8 & 7 \end{vmatrix}$$

Now, calculate the 2x2 determinants:

$$\begin{vmatrix} 5 & 6 \\ 7 & 9 \end{vmatrix} = (5 \cdot 9) - (6 \cdot 7) = 45 - 42 = 3$$

$$\begin{vmatrix} 4 & 6 \\ 8 & 9 \end{vmatrix} = (4 \cdot 9) - (6 \cdot 8) = 36 - 48 = -12$$

$$\begin{vmatrix} 4 & 5 \\ 8 & 7 \end{vmatrix} = (4 \cdot 7) - (5 \cdot 8) = 28 - 40 = -12$$

Substitute these values back into the determinant expansion:

$$\det(A) = 1 \cdot 3 - 3 \cdot (-12) + 2 \cdot (-12)$$

Simplify the expression:

$$\det(A) = 3 + 36 - 24 = 15$$

Thus, the determinant of matrix  $A$  is:

$$\det(A) = 15$$



## 12.1 Properties

1.  $\det(I) = 1$
2. If exchanging two rows of a matrix, we only need to reverse the sign of its determinant
3. If we multiply one row of a matrix by  $t$ , the determinant is also multiplied by a factor of  $t$ .
4. The determinant always behaves like a linear function on the rows of the matrix.
5. If rows of a matrix are equal, its determinant is zero.
6. If  $i \neq j$ , subtracting  $t$  times row  $i$  from row  $j$  doesn't change the determinant

$$\begin{aligned} \begin{bmatrix} a & b \\ c - ta & d - tb \end{bmatrix} &= \begin{bmatrix} a & b \\ c & d \end{bmatrix} - \begin{bmatrix} ta & tb \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \\ &= \begin{bmatrix} a & b \\ c & d \end{bmatrix} - t \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \end{aligned}$$

7. If  $A$  has a row that is all zeros, then  $|A| = 0$ .
8. The determinant of a triangular matrix is the product of the diagonal entries  $d_1, d_2, \dots, d_n$ .
9.  $|A| = 0$  when  $A$  is singular.
10.  $|AB| = |A| \cdot |B| \Rightarrow |A_1 A_2 \dots A_n| = |A_1| |A_2| \dots |A_n|$
11.  $|A^T| = |A|$

## 12.2 Tridiagonal Matrix

A is one for which the only non-zero entries lie on or adjacent to the diagonal:

$$A_{4 \times 4} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

$$|A_n| = |A_{n-1}| - |A_{n-2}|$$

$$|A_1| = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = 0$$

## 13 Definitions of Eigenvalues and Eigenvectors

### 13.1 Application

- Stability analysis
- Vibration analysis
- Atomic orbitals
- Facial recognition

Consider an  $n \times n$  matrix  $A$  and a non-zero vector  $\mathbf{v}$  of length  $n$ . If multiplying  $A$  with  $\mathbf{v}$  simply scales  $\mathbf{v}$  by  $\lambda$ , then:

$$A\mathbf{v} = \lambda\mathbf{v} \quad \Rightarrow \quad (A - \lambda I)\mathbf{v} = 0$$

Example: Find all the eigenvalues and eigenvectors of

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

The characteristic equation is given by:

$$|A - \lambda I| = 0$$

Now, let's compute  $A - \lambda I$ :

$$A - \lambda I = \begin{bmatrix} -\lambda & 0 & 1 \\ 0 & 1 - \lambda & 0 \\ 1 & 0 & -\lambda \end{bmatrix}$$

To find the determinant of  $A - \lambda I$ , we compute:

$$|A - \lambda I| = \det \begin{bmatrix} -\lambda & 0 & 1 \\ 0 & 1 - \lambda & 0 \\ 1 & 0 & -\lambda \end{bmatrix}$$

We can expand this determinant:

$$\begin{aligned} |A - \lambda I| &= (-\lambda) \begin{vmatrix} 1 - \lambda & 0 \\ 0 & -\lambda \end{vmatrix} - 0 + 1 \begin{vmatrix} 0 & 1 - \lambda \\ 1 & 0 \end{vmatrix} \\ &= (-\lambda) [(-\lambda)(1 - \lambda) - (0)(0)] + 1 [(0)(0) - (1 - \lambda)(1)] \\ &= -\lambda [\lambda(\lambda - 1)] + (-(1 - \lambda)) \\ &= -\lambda^2(\lambda - 1) + (1 - \lambda) \end{aligned}$$

Simplifying:

$$|A - \lambda I| = -\lambda^3 + \lambda^2 + \lambda - 1$$

The characteristic equation is then:

$$-\lambda^3 + \lambda^2 + \lambda - 1 = 0$$

Factoring the cubic equation:

$$(\lambda - 1)(-\lambda^2 - 1) = 0$$

So, the eigenvalues are:

$$\lambda_1 = 1, \quad \lambda_2 = -1 \quad (\text{two orthogonal eigenvalues})$$

Eigenvectors For  $\lambda_1 = 1$ , substitute into  $(A - \lambda I)\mathbf{v} = 0$ :

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

The corresponding eigenvector  $\mathbf{v}_1$  can be found by solving  $(A - I)\mathbf{v} = 0$ . A solution is:

$$\mathbf{v}_1 = \begin{bmatrix} -t \\ m \\ t \end{bmatrix}$$

where  $t$  and  $m$  are free parameters.

For  $\lambda_2 = -1$ , substitute into  $(A - (-1)I)\mathbf{v} = 0$ :

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

The corresponding eigenvector  $\mathbf{v}_2$  can be found by solving  $(A + I)\mathbf{v} = 0$ . A solution is:

$$\mathbf{v}_2 = \begin{bmatrix} n \\ 0 \\ -n \end{bmatrix}$$

where  $n$  is a free parameter.

## 13.2 Properties of Eigenvalues

### 13.3 Algebraic Multiplicity

Sometimes the eigenvalues may have fewer than  $n$  distinct roots/solutions. The characteristic polynomial can be factored as:

$$(\lambda - \lambda_1)^{t_1}(\lambda - \lambda_2)^{t_2} \dots (\lambda - \lambda_k)^{t_k}$$

where  $\lambda_1, \dots, \lambda_k$  are the distinct roots (eigenvalues), and  $t_1, \dots, t_k$  are positive integers such that:

$$\sum_{i=1}^k t_i = n$$

Here,  $t_i$  is the **\*\*algebraic multiplicity\*\*** of the eigenvalue  $\lambda_i$ .

February 6, 2025

### 13.4 Geometric Multiplicity

The **\*\*geometric multiplicity\*\*** of an eigenvalue is at most its algebraic multiplicity. It is defined as the dimension of the eigenspace associated with that eigenvalue.

**\*\*Definition\*\***: The geometric multiplicity of an eigenvalue  $\lambda_i$  is the dimension of the null space of  $(A - \lambda_i I)$ , i.e., the number of linearly independent eigenvectors corresponding to  $\lambda_i$ . Lemma : The geometric multiplicity of an eigenvalue is at most its algebraic multiplicity

ex:

We are given the matrix:

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

The characteristic equation is:

$$|A - \lambda I| = \begin{vmatrix} -1 - \lambda & 1 & 0 \\ -4 & 3 - \lambda & 0 \\ 1 & 0 & 2 - \lambda \end{vmatrix} = 0$$

To compute the determinant, we simplify:

$$|A - \lambda I| = -(1 + \lambda)(3 - \lambda)(2 - \lambda) - 4(\lambda - 2) = 0$$

The equation becomes:

$$(\lambda^2 + 2\lambda - 1)^2(\lambda - 2) = 0$$

Thus, the eigenvalues are:

$$\lambda_1 = 1, \quad \lambda_2 = 2$$

with multiplicities:

$$t_1 = 2, \quad t_2 = 2$$

**Eigenvector Calculation** for  $\lambda_1 = 1$ : Let the eigenvector corresponding to  $\lambda_1 = 1$  be  $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ .

We solve the equation:

$$(A - \lambda_1 I)x = 0$$

Substitute  $\lambda_1 = 1$ :

$$A - I = \begin{bmatrix} -2 & 1 & 0 \\ -4 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

Now, we compute the system of equations:

$$\begin{bmatrix} -2 & 1 & 0 \\ -4 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -2x_1 + x_2 \\ -4x_1 + 2x_2 \\ x_1 + x_3 \end{bmatrix} = 0$$

This gives the following system of equations:

$$-2x_1 + x_2 = 0 \quad \Rightarrow \quad x_2 = 2x_1$$

$$-4x_1 + 2x_2 = 0 \quad \Rightarrow \quad x_2 = 2x_1 \quad (\text{consistent with the previous equation})$$

$$x_1 + x_3 = 0 \quad \Rightarrow \quad x_3 = -x_1$$

Thus, the eigenvector corresponding to  $\lambda_1 = 1$  is:

$$x = \begin{bmatrix} x_1 \\ 2x_1 \\ -x_1 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$$

Therefore, the set of solutions is the eigenspace corresponding to  $\lambda_1$ , given by the set of vectors:

$$x = x_1 \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$$

This set has dimension 1. Hence, the **geometric multiplicity** is 1.

### 13.5 Geometric Multiplicity of Eigenvalues

- It is the dimension of the linear space of its associated eigenvectors. Let  $A$  be a  $k \times k$  matrix,  $\lambda_k$  be one of the eigenvalues of  $A$ , and denote its associated eigenspace by  $E_k$ . Then the dimension of  $E_k$  is the geometric multiplicity of its eigenvalue  $\lambda_k$ .

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

The characteristic equation is:

$$|A - \lambda I| = 0$$

which gives:

$$A - \lambda I = \begin{bmatrix} 2 - \lambda & 0 \\ 0 & 2 - \lambda \end{bmatrix}$$

The determinant is:

$$|A - \lambda I| = (2 - \lambda)^2 = 0$$

Thus, we have the eigenvalue:

$$\lambda_1 = 2$$

Now, solving for the eigenspace:

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

The equation  $(A - 2I)\mathbf{x} = 0$  becomes:

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

This equation is trivially satisfied for any values of  $x_1$  and  $x_2$ , meaning the eigenspace is spanned by:

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Thus, the eigenspace corresponding to  $\lambda_1 = 2$  is  $\mathbb{R}^2$ , and its dimension (geometric multiplicity) is 2.

### 13.6 Proof

Suppose that the geometric multiplicity of  $\lambda_k$  is equal to  $g$ , so that there are  $g$  linearly independent eigenvectors  $x_1, \dots, x_g$  associated with  $\lambda_k$ .

Let us randomly select  $k - g$  vectors  $x_{g+1}, \dots, x_k$ , all having dimension  $k \times 1$ , and such that the  $k$  column vectors  $x_1, x_2, \dots, x_k$  are linearly independent.

Define the  $k \times k$  matrix:

$$x = [x_1, x_2, \dots, x_k]$$

For any  $g$ , denote by  $b_g$  the vector that solves:

$$xb_g = Ax_g$$

where  $x_g = [x_1, x_2, \dots, x_g]$ .

• **Step 1: Eigenvalue Property of  $x_1, x_2, \dots, x_g$ :**

We know that the first  $g$  eigenvectors  $x_1, x_2, \dots, x_g$  correspond to the eigenvalue  $\lambda_k$ . By the definition of an eigenvector:

$$Ax_i = \lambda_k x_i \quad \text{for each } i = 1, 2, \dots, g$$

Thus, the action of  $A$  on the first  $g$  columns of  $x$  is simply scaling each eigenvector by  $\lambda_k$ . In matrix form:

$$A \begin{bmatrix} x_1 & x_2 & \dots & x_g \end{bmatrix} = \begin{bmatrix} \lambda_k x_1 & \lambda_k x_2 & \dots & \lambda_k x_g \end{bmatrix}$$

Or equivalently:

$$Ax_g = \lambda_k x_g$$

where  $x_g = [x_1, x_2, \dots, x_g]$  is the matrix formed by the first  $g$  eigenvectors.

• **Step 2: Define the Vector  $b_g$ :**

We want to find  $b_g$ , a  $k \times 1$  vector, such that:

$$xb_g = Ax_g$$

Let  $b_g$  be written as:

$$b_g = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

where  $b_1$  is a  $g \times 1$  vector corresponding to the first  $g$  eigenvectors, and  $b_2$  is a  $(k - g) \times 1$  vector corresponding to the remaining  $k - g$  eigenvectors.

Thus, multiplying  $x$  by  $b_g$  gives:

$$xb_g = [x_1, x_2, \dots, x_k] \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = [x_1 b_1 + x_{g+1} b_2, \dots, x_g b_1 + x_k b_2]$$

For the equation  $xb_g = Ax_g$  to hold, the vector on the right-hand side must match the action of  $A$  on  $x_g$ . We know:

$$Ax_g = \begin{bmatrix} \lambda_k x_1 & \lambda_k x_2 & \dots & \lambda_k x_g \end{bmatrix} = \lambda_k [x_1, x_2, \dots, x_g]$$

Thus, we require:

$$[x_1, x_2, \dots, x_k] \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \lambda_k [x_1, x_2, \dots, x_g]$$

• **Step 3: Solve for  $b_g$ :**

We now equate the two expressions:

$$[x_1, x_2, \dots, x_k] \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \lambda_k [x_1, x_2, \dots, x_g]$$

This equation can be split into two parts: - For the first  $g$  columns:

$$x_1 b_1 + x_{g+1} b_2 = \lambda_k x_1$$

and similarly for the other columns  $x_2, \dots, x_g$ .

- For the last  $k-g$  columns, we assume these additional vectors  $x_{g+1}, \dots, x_k$  are chosen such that they span the remaining space, and hence they do not contribute to the solution for the first  $g$  eigenvectors.

Thus, solving for  $b_1$  and  $b_2$ , we have:

$$b_1 = \lambda_k^{-1} \quad \text{and} \quad b_2 = 0$$

• **Conclusion:**

The vector  $b_g$  that satisfies  $x b_g = A x_g$  is the vector that corresponds to the eigenvalue equation  $Ax = \lambda_k x$ , with the appropriate scaling. The additional vectors  $x_{g+1}, \dots, x_k$  are chosen such that the entire set of vectors  $x_1, \dots, x_k$  remains linearly independent. Therefore, we conclude:

$$x b_g = A x_g$$

as required, and the vector  $b_g$  provides the solution to the equation.

## 14 Homework 2(f)

Conclusion: Thus, there is no effective eigenvalue, since  $g < t$ .

February 6, 2025

## 15 Inverse and Singularity

**Definition:** Suppose  $A$  is a square matrix. We look for an inverse matrix  $A^{-1}$  of the same size, such that  $A^{-1}A = I$ . Thus,  $A^{-1}Ax = x$ . Multiplying  $Ax = b$  by  $A^{-1}$  gives  $A^{-1}Ax = A^{-1}b$ . If the determinant of  $A$  is non-zero, then  $A^{-1}$  exists. In this case,  $A$  is **invertible**.

### 15.1 Example: Calculating $A^{-1}$ via Gauss-Jordan Elimination

Given the matrix:

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

we augment  $A$  with the identity matrix  $I$  to form  $[A|I]$ :

$$[A|I] = \left[ \begin{array}{ccc|ccc} 2 & -1 & 0 & 1 & 0 & 0 \\ -1 & 2 & -1 & 0 & 1 & 0 \\ 0 & -1 & 2 & 0 & 0 & 1 \end{array} \right]$$



**Step 1:** Normalize the first pivot and eliminate below.

$$R_1 \leftarrow \frac{1}{2}R_1 \Rightarrow \left[ \begin{array}{ccc|ccc} 1 & -\frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ -1 & 2 & -1 & 0 & 1 & 0 \\ 0 & -1 & 2 & 0 & 0 & 1 \end{array} \right]$$

$$R_2 \leftarrow R_2 + R_1 \Rightarrow \left[ \begin{array}{ccc|ccc} 1 & -\frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{3}{2} & -1 & \frac{1}{2} & 1 & 0 \\ 0 & -1 & 2 & 0 & 0 & 1 \end{array} \right]$$

**Step 2:** Normalize the second pivot and eliminate above/below.

$$R_2 \leftarrow \frac{2}{3}R_2 \Rightarrow \left[ \begin{array}{ccc|ccc} 1 & -\frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 1 & -\frac{2}{3} & \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & -1 & 2 & 0 & 0 & 1 \end{array} \right]$$

$$R_3 \leftarrow R_3 + R_2 \Rightarrow \left[ \begin{array}{ccc|ccc} 1 & -\frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 1 & -\frac{2}{3} & \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & 0 & \frac{4}{3} & \frac{1}{3} & \frac{2}{3} & 1 \end{array} \right]$$

**Step 3:** Normalize the third pivot and eliminate above.

$$R_3 \leftarrow \frac{3}{4}R_3 \Rightarrow \left[ \begin{array}{ccc|ccc} 1 & -\frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 1 & -\frac{2}{3} & \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & 0 & 1 & \frac{1}{4} & \frac{1}{2} & \frac{3}{4} \end{array} \right]$$

$$R_2 \leftarrow R_2 + \frac{2}{3}R_3 \Rightarrow \left[ \begin{array}{ccc|ccc} 1 & -\frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 & \frac{1}{2} & 1 & \frac{1}{2} \\ 0 & 0 & 1 & \frac{1}{4} & \frac{1}{2} & \frac{3}{4} \end{array} \right]$$

$$R_1 \leftarrow R_1 + \frac{1}{2}R_2 \Rightarrow \left[ \begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{3}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 1 & 0 & \frac{1}{2} & 1 & \frac{1}{2} \\ 0 & 0 & 1 & \frac{1}{4} & \frac{1}{2} & \frac{3}{4} \end{array} \right]$$

The inverse of  $A$  is:

$$A^{-1} = \begin{bmatrix} \frac{3}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{3}{4} \end{bmatrix}$$

## 15.2 Calculate $A^{-1}$ using determinant

$$A^{-1} = \frac{1}{|A|} \text{adj}(A) = Q_A^T$$

**Example**

$$A = \begin{bmatrix} 1 & 3 & 2 \\ 4 & 5 & 6 \\ 8 & 7 & 9 \end{bmatrix}, \quad Q_A = \begin{bmatrix} 3 & 12 & -12 \\ -13 & -7 & 17 \\ 8 & 2 & -7 \end{bmatrix}, \quad Q_A^T = \begin{bmatrix} 3 & -13 & 8 \\ 12 & -7 & 2 \\ -12 & 17 & -7 \end{bmatrix}$$

$$|A| = 15 \neq 0$$

$$A^{-1} = \frac{1}{|A|} Q_A^T = \frac{1}{15} \begin{bmatrix} 3 & -13 & 8 \\ 12 & -7 & 2 \\ -12 & 17 & -7 \end{bmatrix} = \begin{bmatrix} \frac{1}{5} & \frac{-13}{15} & \frac{8}{15} \\ \frac{4}{5} & \frac{-7}{15} & \frac{2}{15} \\ \frac{-4}{5} & \frac{17}{15} & \frac{-7}{15} \end{bmatrix}$$

## Lemma

Let  $A$  be a square matrix. If  $A$  is invertible, then every equation  $Ax = b$  has a unique solution.

**Proof:** Let  $B$  be the inverse of  $A$ . Suppose  $v$  is a solution of  $Ax = b$ . Then:

$$Av = b.$$

Multiply both sides by  $B$ :

$$B(Av) = Bb \Rightarrow (BA)v = Bb \Rightarrow Iv = Bb \Rightarrow v = Bb.$$

This shows that  $v = A^{-1}b$  is the **only** solution. Thus, uniqueness is guaranteed.

## Inverse Properties

1.  $(A^{-1})^{-1} = A$
2.  $(A^T)^{-1} = (A^{-1})^T$
3.  $\det(A^{-1}) = \frac{1}{\det(A)} = |A|^{-1}$
4.  $(AB)^{-1} = B^{-1}A^{-1}$
5. For invertible matrices  $A_1, A_2, \dots, A_n$ :  $(A_1A_2 \cdots A_n)^{-1} = A_n^{-1} \cdots A_2^{-1}A_1^{-1}$
6.  $(A^n)^{-1} = (A^{-1})^n$  for  $n \in \mathbb{N}$
7. For scalar  $k \neq 0$ :  $(kA)^{-1} = \frac{1}{k}A^{-1}$
8. If  $D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$  with  $\lambda_i \neq 0$ , then  $D^{-1} = \begin{bmatrix} \frac{1}{\lambda_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\lambda_n} \end{bmatrix}$
9.  $(A + B)^{-1} \neq A^{-1} + B^{-1}$  (in general)
10. For a non-singular matrix  $A$  of order  $n$ :  $\text{adj}(\text{adj}(A)) = |A|^{n-2}A$
11. If  $AB=0$ , then at least one of  $A$  or  $B$  must be singular

## 16 Singularity

1.  $|A| = 0 \implies A$  is a singular matrix (no inverse exists).
2. For a matrix  $A_{k \times k} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_k]$ , where  $\mathbf{a}_i \in \mathbb{R}^{k \times 1}$ , the columns are **linearly dependent** if there exist scalars  $\alpha_j$  (not all zero) such that:

$$\mathbf{a}_i = \sum_{j \neq i} \alpha_j \mathbf{a}_j$$

3. The **rank** of a matrix is the number of linearly independent columns/rows. If all columns/rows are independent, the matrix is **full rank**.

**Example:**

$$A = \begin{bmatrix} 2 & -4 \\ 3 & -6 \end{bmatrix}, \quad \mathbf{a}_1 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} -4 \\ -6 \end{bmatrix}$$

Since  $\mathbf{a}_2 = -2\mathbf{a}_1$ , the columns are linearly dependent. Thus:

$$\text{rank}(A) = 1$$

### 16.1 Block Matrix

A submatrix of a matrix  $M$  is formed by deleting selected rows and/or columns from  $M$ .

**Example:** Let  $M = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$ . By deleting the first row and third column,

we get the submatrix:

$$M_{\text{sub}} = \begin{bmatrix} 4 & 5 \\ 7 & 8 \end{bmatrix}$$

### 16.2 Partitioned Matrix

A partitioned (or block) matrix divides a matrix into smaller submatrices (blocks), often to simplify operations.

**Example:** Partition the matrix  $M = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$  into blocks:

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} = \left[ \begin{array}{cc|c} 1 & 2 & 3 \\ 4 & 5 & 6 \\ \hline 7 & 8 & 9 \end{array} \right]$$

where:

$$M_{11} = \begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix}, \quad M_{12} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}, \quad M_{21} = \begin{bmatrix} 7 & 8 \end{bmatrix}, \quad M_{22} = \begin{bmatrix} 9 \end{bmatrix}$$

Partitioning allows operations like block matrix multiplication. For instance,

$$M_{11} \cdot M_{12} = \begin{bmatrix} 1 \cdot 3 + 2 \cdot 6 \\ 4 \cdot 3 + 5 \cdot 6 \end{bmatrix} = \begin{bmatrix} 15 \\ 42 \end{bmatrix}.$$

## 16.3 Singularity Properties

### 1. Sherman-Morrison Formula

Let  $A$  be a  $k \times k$  nonsingular matrix, and let  $\mathbf{u}, \mathbf{v}$  be  $k \times 1$  column vectors. If  $1 + \mathbf{v}^T A^{-1} \mathbf{u} \neq 0$ , then:

$$(A + \mathbf{u}\mathbf{v}^T)^{-1} = A^{-1} - \frac{A^{-1}\mathbf{u}\mathbf{v}^T A^{-1}}{1 + \mathbf{v}^T A^{-1} \mathbf{u}}$$

**Example:** Let  $A = I_2$  (identity matrix),  $\mathbf{u} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ ,  $\mathbf{v} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ . Then:

$$A + \mathbf{u}\mathbf{v}^T = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \quad A^{-1} = I_2$$

Applying the formula:

$$(A + \mathbf{u}\mathbf{v}^T)^{-1} = I_2 - \frac{I_2 \mathbf{u}\mathbf{v}^T I_2}{1 + \mathbf{v}^T I_2 \mathbf{u}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \frac{\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}}{2} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix}$$

*Application:* Efficient matrix updates in least-squares problems and neural network training.

February 6, 2025

## 17 Systems of Equations

A system of  $K$  linear equations in  $L$  unknowns is a set of equations.

$$A_{11}x_1 + A_{12}x_2 + \dots + A_{1L}x_L = b_1, A_{21}x_1 + A_{22}x_2 + \dots + A_{2L}x_L = b_2, A_{31}x_1 + A_{32}x_2 + \dots + A_{3L}x_L = b_3, \dots, A_{K1}x_1 + A_{K2}x_2 + \dots + A_{KL}x_L = b_K$$

1.  $A\mathbf{x} = \mathbf{b}$  —> It has a solution if and only if  $\mathbf{b}$  belongs to the span of the columns of  $A$ . If  $A\mathbf{x} = \mathbf{b}$  has a solution, then the solution is unique if and only if the columns of  $A$  are linearly independent.

$$\begin{bmatrix} x_1 + 2x_2 + 0x_3 = 5 \\ 2x_1 + 1x_2 + 3x_3 = 4 \end{bmatrix}$$

### 17.1 Augmented Matrix

**Augmented Matrix to solve Linear Eqns** It is the result of the joining the columns of two or more matrices having the same number of rows.

Let  $A$  be a  $K \times L$  matrix,  $B$  a  $K \times M$  matrix, then the augmented matrix  $(A|B)$  is the  $K \times (L + M)$  matrix obtained by appending the columns of  $B$  to the right of  $A$ .

## 17.2 Homogeneous System vs. Non-Homogeneous System

A system of linear equations is called:

- **Homogeneous** if all constant terms are zero:  $A\mathbf{x} = \mathbf{0}$
- **Non-Homogeneous** if at least one constant term is non-zero:  $A\mathbf{x} = \mathbf{b}$  where  $\mathbf{b} \neq \mathbf{0}$

### Homogeneous System Example

$$\begin{cases} x_1 + 2x_2 - x_3 = 0 \\ 3x_1 - x_2 + 4x_3 = 0 \\ 2x_1 + x_2 + x_3 = 0 \end{cases}$$

Matrix form with augmented column (all zeros):

$$\left[ \begin{array}{ccc|c} 1 & 2 & -1 & 0 \\ 3 & -1 & 4 & 0 \\ 2 & 1 & 1 & 0 \end{array} \right]$$

**Property:** Always has at least the trivial solution  $\mathbf{x} = \mathbf{0}$ .

### Non-Homogeneous System Example

$$\begin{cases} x_1 + 2x_2 - x_3 = 5 \\ 3x_1 - x_2 + 4x_3 = 2 \\ 2x_1 + x_2 + x_3 = 7 \end{cases}$$

Matrix form with non-zero augmented column:

$$\left[ \begin{array}{ccc|c} 1 & 2 & -1 & 5 \\ 3 & -1 & 4 & 2 \\ 2 & 1 & 1 & 7 \end{array} \right]$$

**Property:** Solutions (if they exist) can be expressed as:  $\mathbf{x} = \mathbf{x}_p + \mathbf{x}_h$ , where  $\mathbf{x}_p$  is a particular solution and  $\mathbf{x}_h$  is a solution to the homogeneous system.

**Other Properties** Through elementary row operations, any non-homogeneous system  $A\mathbf{x} = \mathbf{b}$  can be transformed into an equivalent system  $R\mathbf{x} = \mathbf{b}_R$ , where:

- $R$  is a matrix in **Row Echelon Form (REF)**
- The system is **inconsistent** (no solution) if  $R$  contains a zero row  $R_i$  with  $\mathbf{b}_R[i] \neq 0$
- The system has **free variables** if  $R$  has non-pivotal columns

### Example of Inconsistency:

$$\left[ \begin{array}{ccc|c} 1 & 3 & 2 & 4 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 3 \end{array} \right] \leftarrow \text{Contradiction}$$

This system has no solution because  $0x_1 + 0x_2 + 0x_3 = 3$  is impossible.

## 17.3 Partitioned System

For a  $k \times l$  REF matrix  $R$  with  $r$  basic columns:

- **Basic variables:** Correspond to pivot columns (first  $r$  columns)
- **Free variables:** Correspond to non-pivot columns (last  $l - r$  columns)

### Matrix Partitioning

Split  $R$  and  $\mathbf{x}$  into:

$$R = [B \quad N], \quad \mathbf{x} = \begin{bmatrix} \mathbf{x}_b \\ \mathbf{x}_n \end{bmatrix}$$

- $B$ :  $k \times r$  matrix of basic columns
- $N$ :  $k \times (l - r)$  matrix of non-basic columns
- $\mathbf{x}_b$ :  $r \times 1$  vector of basic variables
- $\mathbf{x}_n$ :  $(l - r) \times 1$  vector of free variables

### Rewriting the System

The system  $R\mathbf{x} = \mathbf{b}_R$  becomes:

$$B\mathbf{x}_b + N\mathbf{x}_n = \mathbf{b}_R$$

Leading to the general solution:

$$\mathbf{x}_b = B^{-1}(\mathbf{b}_R - N\mathbf{x}_n)$$

(Note:  $B$  is always invertible for consistent systems)

**Example:** Given system in REF:

$$\left[ \begin{array}{cc|c} 1 & 2 & 5 \\ 0 & 1 & 3 \end{array} \right]$$

- Basic variables:  $x_1, x_2$
- Free variables: None (full rank)
- Unique solution:  $x_1 = -1, x_2 = 3$

**Example with Free Variable:**

$$\left[ \begin{array}{ccc|c} 1 & 3 & 2 & 4 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 \end{array} \right]$$

• Partitioning:  $B = \begin{bmatrix} 1 & 3 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, N = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}$

- Solutions:  $x_3$  is free. Let  $x_3 = t$ :

$$\begin{cases} x_2 = 1 + t \\ x_1 = 4 - 3(1 + t) - 2t = 1 - 5t \end{cases}$$

General solution:  $\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -5 \\ 1 \\ 1 \end{bmatrix}$

February 6, 2025

## 18 Singular Value Decomposition

### Four Fundamental Subspaces

The four fundamental subspaces are the ranges (column spaces) and kernels (null spaces) of a matrix  $A$  and its transpose  $A^\top$ .

Consider the following spaces and linear map:

- $S = \mathbb{R}^3$ : the space of all  $3 \times 1$  column vectors.
- $T = \mathbb{R}^2$ : the space of all  $2 \times 1$  column vectors.
- $A$ : a  $2 \times 3$  matrix defining the linear map  $f : S \rightarrow T$  via  $f(\mathbf{s}) = A\mathbf{s}$ .

Given the matrix

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

the linear map transforms  $3 \times 1$  vectors in  $S$  into  $2 \times 1$  vectors in  $T$ .

**Example:** Let  $\mathbf{s} = \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} \in S$ . Then,

$$A\mathbf{s} = \begin{bmatrix} 2 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} (2 \cdot 3) + (0 \cdot 0) + (1 \cdot 1) \\ (1 \cdot 3) + (0 \cdot 0) + (1 \cdot 1) \end{bmatrix} = \begin{bmatrix} 7 \\ 4 \end{bmatrix} \in T.$$

## 18.1 Column Space

The column space (range) of  $A$ , denoted  $\mathcal{R}(A)$ , is defined as:

$$\mathcal{R}(A) = \text{range}(f) = \{A\mathbf{s} : \mathbf{s} \in S\}$$

This consists of all possible linear combinations of  $A$ 's column vectors.

Given the matrix:

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

any vector  $A\mathbf{s}$  can be expressed as:

$$A\mathbf{s} = \begin{bmatrix} 2 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} S_1 \\ S_2 \\ S_3 \end{bmatrix} = S_1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + S_2 \begin{bmatrix} 0 \\ 0 \end{bmatrix} + S_3 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = S_1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + S_3 \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

The column space is therefore spanned by the first and third columns:

$$\text{span} \left( \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right).$$

## 18.2 Null Space

The null space (kernel) of  $A$ , denoted  $\mathcal{N}(A)$ , is defined as:

$$\mathcal{N}(A) = \text{null}(f) = \{\mathbf{s} \in S : A\mathbf{s} = \mathbf{0}\}.$$

**Example:** To find the null space of

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

we solve  $A\mathbf{s} = \mathbf{0}$ :

$$\begin{bmatrix} 2 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} S_1 \\ S_2 \\ S_3 \end{bmatrix} = S_1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + S_2 \begin{bmatrix} 0 \\ 0 \end{bmatrix} + S_3 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

This reduces to the system:

$$\begin{cases} 2S_1 + S_3 = 0 \\ S_1 + S_3 = 0 \end{cases} \implies S_1 = -S_3 \text{ and } S_3 = -S_1.$$



Letting  $S_2$  be a free variable ( $S_2 = t$ ) and  $S_1 = -S_3$ , we get:

$$\mathbf{s} = t \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad t \in \mathbb{R}.$$

Thus, the null space is:

$$\text{span} \left( \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right).$$

### 18.3 Row Space

A linear map  $g : T \rightarrow S$  defined by:

$$g(\mathbf{t}) = A^\top \mathbf{t}$$

It is the range of the linear map defined by  $A^\top$ :

$$\mathcal{R}(A^\top) = \text{range}(g) = \{A^\top \mathbf{t} : \mathbf{t} \in T\},$$

which is called the row space of  $A$ .

**Example:**

$$A_{2 \times 3} = \begin{bmatrix} 2 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}, \quad A_{3 \times 2}^\top = \begin{bmatrix} 2 & 1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix}.$$

For any  $\mathbf{t} = \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} \in T$ :

$$A^\top \mathbf{t} = \begin{bmatrix} 2 & 1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} = t_1 \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} + t_2 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

The row space of  $A$  is therefore spanned by:

$$\mathcal{R}(A^\top) = \text{span} \left( \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right).$$

### 18.4 Left-Hand Null Space

It is the kernel of the linear map defined by  $A^\top$ :

$$\mathcal{N}(A^\top) = \text{null}(g) = \{\mathbf{t} \in T : A^\top \mathbf{t} = \mathbf{0}\}.$$

**Example:**

$$A^\top \mathbf{t} = \begin{bmatrix} 2 & 1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} = t_1 \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} + t_2 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Solving the system:

$$\begin{cases} 2t_1 + t_2 = 0 \\ t_1 + t_2 = 0 \end{cases} \implies t_1 = 0, t_2 = 0.$$

Thus, the left-hand null space is trivial:

$$\mathcal{N}(A^\top) = \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\}.$$

## 18.5 Definition of SVD

Let  $M$  be an  $m \times n$  matrix with rank  $r$ . There exists a matrix factorization called the **\*\*reduced singular value decomposition (SVD)\*\*** of  $M$ :

$$M = U \cdot \Sigma \cdot V^\top,$$

where:

1.  $U$  is an  $m \times r$  column-orthonormal matrix:
  - Each column is a unit vector.
  - Columns are mutually orthogonal (dot product of any two columns is 0).
2.  $\Sigma$  is an  $r \times r$  diagonal matrix. The diagonal entries  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$  are called the singular values of  $M$ .
3.  $V^\top$  is an  $r \times n$  row-orthonormal matrix:
  - Each row is a unit vector.
  - Rows are mutually orthogonal.

## 18.6 Propositions

1. **Proposition 1:** For any matrix  $M$ , the matrices  $M^\top M$  and  $MM^\top$  have non-negative eigenvalues.
2. **Proposition 2:** For any matrix  $M$ ,  $\text{rank}(M) = \text{rank}(M^\top M) = \text{rank}(MM^\top)$ .

## 18.7 Calculation

1. The columns of  $V$  (rows of  $V^\top$ ) are the eigenvectors corresponding to the positive eigenvalues of  $M^\top M$ , and the squares of the singular values in  $\Sigma$  are these eigenvalues.

### 18.7.1 Proof

: Let the reduced Singular Value Decomposition (SVD) of  $M$  be:

$$M = U\Sigma V^\top,$$

where:

- $U \in \mathbb{R}^{m \times m}$  and  $V \in \mathbb{R}^{n \times n}$  are orthonormal matrices,
- $\Sigma \in \mathbb{R}^{m \times n}$  is a diagonal matrix with non-negative entries (the singular values of  $M$ ).

Now, compute  $M^\top M$ :

$$M^\top M = (U\Sigma V^\top)^\top (U\Sigma V^\top) = V\Sigma^\top U^\top U\Sigma V^\top.$$

Since  $U$  is orthonormal,  $U^\top U = I$ , so this simplifies to:

$$M^\top M = V\Sigma^2 V^\top.$$

This is the eigen-decomposition of  $M^\top M$ , where:

- $V$  contains the eigenvectors of  $M^\top M$ ,
- $\Sigma^2$  contains the eigenvalues of  $M^\top M$ , which are the squares of the singular values of  $M$ .

Thus, the non-zero eigenvalues of  $M^\top M$  are the squares of the singular values  $\sigma_i$  in  $\Sigma$ , and the columns of  $V$  are the corresponding eigenvectors.

## 18.8 Lemma: Eigenvalue Relationship

Suppose  $(v, \lambda)$  is the eigenstructure of  $MM^\top$  with  $\lambda \neq 0$ . Then:

$$MM^\top v = \lambda v.$$

Multiplying both sides by  $M^\top$ , we get:

$$M^\top M M M^\top v = \lambda M^\top v.$$

Let  $x = M^\top v$ , then:

$$M^\top M x = \lambda x.$$

This shows that  $\lambda$  is also an eigenvalue of  $M^\top M$ . Hence, the eigenvalues of  $MM^\top$  and  $M^\top M$  are the same.

## 18.9 Full Singular Value Decomposition (SVD)

The full SVD of an  $m \times n$  matrix  $M$  is defined as:

$$M = U\Sigma V^\top,$$

where:

- $U \in \mathbb{R}^{m \times m}$  is an orthogonal matrix containing all the eigenvectors of  $MM^\top$ , including those corresponding to the zero eigenvalues,
- $\Sigma \in \mathbb{R}^{m \times n}$  is a diagonal matrix, where the diagonal entries  $\Sigma_{ii}$  are the singular values of  $M$ ,
- $V^\top \in \mathbb{R}^{n \times n}$  is an orthogonal matrix containing all the eigenvectors of  $M^\top M$ , including those corresponding to zero eigenvalues.

The full SVD gives a complete set of eigenvectors for both  $MM^\top$  and  $M^\top M$ , including the eigenvectors associated with zero eigenvalues.

## 18.10 Reduced vs Full SVD

The full SVD can be truncated to give the reduced SVD, which is a more compact representation when dealing with low-rank matrices.

### 18.10.1 Full SVD

For a matrix  $M \in \mathbb{R}^{m \times n}$ , the full SVD is:

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

where: -  $U \in \mathbb{R}^{m \times m}$  contains the eigenvectors of  $MM^\top$ , -  $\Sigma \in \mathbb{R}^{m \times n}$  is a rectangular diagonal matrix containing the singular values of  $M$ , -  $V^\top \in \mathbb{R}^{n \times n}$  contains the eigenvectors of  $M^\top M$ .

### 18.10.2 Reduced SVD

The reduced SVD retains only the non-zero singular values and their corresponding eigenvectors. It is given by:

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

where: -  $U_r \in \mathbb{R}^{m \times r}$ , with  $r = \text{rank}(M)$ , contains the eigenvectors of  $MM^\top$  corresponding to non-zero eigenvalues, -  $\Sigma_r \in \mathbb{R}^{r \times r}$  is a square diagonal matrix containing only the non-zero singular values, -  $V_r^\top \in \mathbb{R}^{r \times n}$  contains the eigenvectors of  $M^\top M$  corresponding to non-zero eigenvalues.

### 18.10.3 Key Differences Between Full and Reduced SVD

- In the full SVD,  $U$  and  $V^\top$  are square matrices ( $m \times m$  and  $n \times n$ , respectively), and  $\Sigma$  is an  $m \times n$  rectangular matrix.
- In the reduced SVD,  $U_r$  is  $m \times r$ ,  $\Sigma_r$  is  $r \times r$ , and  $V_r^\top$  is  $r \times n$ , where  $r = \text{rank}(M)$ .
- The reduced SVD is more efficient in storage and computation, particularly when the matrix  $M$  is low-rank.

### 18.10.4 How the Reduced SVD Relates to the Full SVD

The reduced SVD is essentially a truncation of the full SVD. Specifically: -  $U_r$  is the first  $r$  columns of  $U$  from the full SVD, -  $V_r^\top$  is the first  $r$  rows of  $V^\top$  from the full SVD, -  $\Sigma_r$  contains only the non-zero singular values, corresponding to the non-zero diagonal entries in  $\Sigma$  from the full SVD.

### 18.10.5 Example

Let  $M \in \mathbb{R}^{4 \times 3}$  be a matrix with rank  $r = 2$ . The full SVD of  $M$  would look like:

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

where: -  $U \in \mathbb{R}^{4 \times 4}$ , -  $\Sigma \in \mathbb{R}^{4 \times 3}$ , -  $V^\top \in \mathbb{R}^{3 \times 3}$ .

The reduced SVD would be:

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

where: -  $U_r \in \mathbb{R}^{4 \times 2}$ , -  $\Sigma_r \in \mathbb{R}^{2 \times 2}$ , -  $V_r^\top \in \mathbb{R}^{2 \times 3}$ .

### 18.10.6 Visualizing Full SVD vs Reduced SVD

**Full SVD:**

$$M = \underbrace{\begin{bmatrix} u_1 & u_2 & \dots & u_m \end{bmatrix}}_U \cdot \underbrace{\begin{bmatrix} \Sigma_1 & 0 & \dots & 0 \\ 0 & \Sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Sigma_n \end{bmatrix}}_\Sigma \cdot \underbrace{\begin{bmatrix} v_1^\top \\ v_2^\top \\ \vdots \\ v_n^\top \end{bmatrix}}_{V^\top}$$

**Reduced SVD:**

$$M = \underbrace{\begin{bmatrix} u_1 & u_2 & \dots & u_r \end{bmatrix}}_{U_r} \cdot \underbrace{\begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix}}_{\Sigma_r} \cdot \underbrace{\begin{bmatrix} v_1^\top \\ v_2^\top \end{bmatrix}}_{V_r^\top}$$

Where  $r = \text{rank}(M)$ .

## 18.11 Conclusion

The full SVD provides a complete representation of the matrix, including both non-zero and zero singular values. The reduced SVD is a more compact representation that captures only the non-zero singular values, making it more efficient for low-rank matrices or dimensionality reduction tasks.

## Spectral Decomposition

### 18.12 Spectral Theorem of Decomposition

Let  $A \in \mathbb{R}^{n \times n}$  be real and symmetric. Then

1. The eigenvalues of  $A$  are real
2.  $A$  is diagonalizable
3. there is an orthonormal basis of  $\mathbb{R}^n$  consisting of eigenvectors of  $A$ .

Thus,  $A$  may be orthogonally diagonalized:  $A = V \Lambda V^T$  where  $V \in \mathbb{R}^{n \times n}$  is an orthonormal matrix of eigenvectors of  $A$ , and  $\Lambda \in \mathbb{R}^{n \times n}$  is a real diagonal matrix of eigenvalues.

$$A = V \Lambda V^T = \sum_{i=1}^n \lambda_i v_i v_i^T$$

1. If for the symmetric matrix  $A$ ,  $\lambda_i > 0, i = 1, \dots, n$  then it is called positive definite and writes as  $A > 0$ . If  $\lambda_i \geq 0, i = 1, \dots, n$ , then  $A$  is called non-negative definite and write as  $A \geq 0$ .

## Spectral Decomposition Calculation

Let  $A$  be a square  $n \times n$  matrix with eigenvalues  $\lambda_i$  and corresponding eigenvectors  $v_i$ . The spectral decomposition of  $A$  can be written as:

$$A = [v_1, v_2, \dots, v_n] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} v_1^T & v_2^T & \cdots & v_n^T \end{bmatrix}$$

or equivalently:

$$A = \sum_{i=1}^n \lambda_i v_i v_i^T$$

**Example:**

Consider the matrix  $A$ :

$$A = \begin{bmatrix} -3 & 5 \\ 4 & -2 \end{bmatrix}$$

**Step 1: Find the Eigenvalues**

The eigenvalues are found by solving the characteristic equation  $\det(A - \lambda I) = 0$ :

$$\det(A - \lambda I) = \begin{vmatrix} -3 - \lambda & 5 \\ 4 & -2 - \lambda \end{vmatrix} = 0$$

Expanding the determinant:

$$(-3 - \lambda)(-2 - \lambda) - 20 = 0$$

Simplifying:

$$\lambda^2 + 5\lambda - 6 - 20 = 0 \quad \Rightarrow \quad \lambda^2 + 5\lambda - 26 = 0$$

Solving for  $\lambda$ , we find:

$$\lambda_1 = -7, \quad \lambda_2 = 2$$

**Step 2: Find the Eigenvectors**

To find the eigenvectors corresponding to  $\lambda_1 = -7$ , we solve  $(A - \lambda_1 I)v_1 = 0$ :

$$A - (-7)I = \begin{bmatrix} 4 & 5 \\ 4 & 5 \end{bmatrix}$$

This leads to the eigenvector:

$$v_1 = \begin{bmatrix} \frac{-5}{\sqrt{41}} \\ \frac{-4}{\sqrt{41}} \end{bmatrix}$$

For  $\lambda_2 = 2$ , we solve  $(A - 2I)v_2 = 0$ :

$$A - 2I = \begin{bmatrix} -5 & 5 \\ 4 & -4 \end{bmatrix}$$

This leads to the eigenvector:

$$v_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

### Step 3: Construct the Matrix $V$ and $\Lambda$

The matrix  $V$  of eigenvectors is:

$$V = \begin{bmatrix} \frac{-5}{\sqrt{41}} & \frac{1}{\sqrt{2}} \\ \frac{-4}{\sqrt{41}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

The diagonal matrix  $\Lambda$  of eigenvalues is:

$$\Lambda = \begin{bmatrix} -7 & 0 \\ 0 & 2 \end{bmatrix}$$

Thus, the spectral decomposition of  $A$  is:

$$A = V\Lambda V^T$$

## Definitions, Properties, and Derivatives of Matrix Traces

### Introduction

#### 1. Vector Norms:

- For any vector  $x \in \mathbb{R}^n$ , the  $l_2$ -norm of  $x$  is given by:

$$\|x\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \sqrt{\sum_{i=1}^n x_i^2}$$

where  $x_i$  is the  $i$ -th entry of  $x$  for all  $i \in [n]$ .

#### 2. Inner Product:

- For vectors  $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$  and  $y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ , the angle  $\theta$  between them can be computed as:

$$\cos \theta = \frac{x_1 y_1 + x_2 y_2}{\sqrt{x_1^2 + x_2^2} \sqrt{y_1^2 + y_2^2}} = \frac{\langle x, y \rangle}{\|x\|_2 \|y\|_2}$$

where  $\langle x, y \rangle = x^T y = x_1 y_1 + x_2 y_2$  is the inner product of  $x$  and  $y$ .

#### 3. Matrix Trace:

- For any square matrix  $X \in \mathbb{R}^{n \times n}$ , the trace  $\text{tr}(X)$  is the sum of the diagonal entries:

$$\text{tr}(X) = \sum_{i=1}^n x_{ii}$$

where  $x_{ii}$  is the  $(i, i)$ -th entry of  $X$ . Additionally,  $\text{tr}(X) = \text{tr}(X^T)$ , meaning the trace is invariant under transposition.



### 18.13 Properties of Matrix Trace

- $\text{tr}(x + y) = \text{tr}(x) + \text{tr}(y)$
- $\text{tr}(\alpha x + \beta y) = \alpha \text{tr}(x) + \beta \text{tr}(y)$
- $\text{tr}(xy) = \text{tr}(yx)$
- $\text{tr}(x^T x) = \|x\|_F^2$

## 19 Notes For homework

Proof: should be general.. Verify: Can use specific examples Learn the proofs on Singular Value Decomposition.