CSE 840: Computational Foundations of Artificial Intelligence January 27, 2025

Transposes, Change of Basis, Rank of a Matrix, Determinant

Instructor: Vishnu Boddeti *Scribe: Vaibhav Vennam, Roshni Bhowmik*

1 Transpose

Given a matrix $A = (a_{ij})_{ij} \in \mathbb{F}^{m\times n}$, the transpose matrix is given by $\left(A^T\right)_{kj} = A_{jk}$.

For example:

$$A = \begin{bmatrix} 1 & -9 & 3 \\ 1 & 2 & -2 \\ -2 & 1 & 1 \end{bmatrix} \implies A^T = \begin{bmatrix} 1 & 1 & -2 \\ -9 & 2 & 1 \\ 3 & -2 & 1 \end{bmatrix}$$

If $\mathbb{F} = \mathbb{C}$, then the conjugate transpose matrix is defined as $(A^*)_{ij} = \bar{A}_{ij}$, where
(Note: '*' subscript is used for conjugate transpose)

$$x = a + ib \implies \bar{x} = a - ib \; \forall a, b \in \mathbb{F}$$

This is useful in finding the adjoint of an operator.

Adjoint of an operator: It is a transformed version of the operator that preserves inner products. For a matrix, the adjoint is its **conjugate transpose** (flip the matrix and take complex conjugates). It helps in understanding symmetry, reversibility, and self-adjoint operators in math and physics.

2 Change of Basis

Consider the identity mapping $I : V \in \mathbb{R}^n \to V \in \mathbb{R}^n$, that is, the original vector space is getting mapped to the same space.

That being said, vector space V can be represented by multiple bases. If we represented the original space through the basis $\mathcal{A}$, and the target space with the basis $\mathcal{B}$, we can represent the identity mapping matrix as $M(I, \mathcal{A}, \mathcal{B})$ using the map I to **change** the bases.

If $\mathcal{A} = \{a_1, a_2, ..., a_n\}$ and $\mathcal{B} = \{b_1, b_2, ..., b_n\}$ are different bases of V, then we can represent the basis vectors in $\mathcal{A}$ using the vectors in $\mathcal{B}$ since $\mathcal{A} \subset V$:

$$a_1 = t_{11}b_1 + t_{21}b_2 + \cdots + t_{n1}b_n$$
$$a_2 = t_{12}b_1 + t_{22}b_2 + \cdots + t_{n2}b_n$$
$$a_n = t_{1n}b_1 + t_{2n}b_2 + \cdots + t_{nn}b_n$$

We can represent these vectors in a matrix form:

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}_{n\times 1} = \begin{bmatrix} t_{11} & t_{12} & \dots & t_{1n} \\ t_{21} & t_{22} & \dots & t_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ t_{n1} & t_{n2} & \dots & t_{nn} \end{bmatrix}_{n\times n} \times \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}_{n\times 1}$$

We can call the mapping matrix from $V \to V$ that changes the basis from $\mathcal{A}$ to $\mathcal{B}$ as:

$$M(T, \mathcal{A}, \mathcal{B}) = \begin{bmatrix} t_{11} & t_{12} & \dots & t_{1n} \\ t_{21} & t_{22} & \dots & t_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ t_{n1} & t_{n2} & \dots & t_{nn} \end{bmatrix}_{n\times n}$$

If the original basis and the target basis are the same, our mapping matrix would be an identity matrix:

$$M(I, \mathcal{A}, \mathcal{A}) = \mathbb{I}_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}_{n\times n}$$

since

$$\begin{aligned} a_1 &= 1.a_1 + 0.a_2 + \ldots + 0.a_n, \\ a_2 &= 0.a_1 + 1.a_2 + \ldots + 0.a_n, \\ &\vdots \\ a_n &= 0.a_1 + 0.a_2 + \ldots + 1.a_n. \end{aligned}$$

Each column represents the change of a basis vector $a_i \in \mathcal{A} \to b_i \in \mathcal{B}$

So for a change of basis from $\mathcal{A} \to \mathcal{A}$:

$$T_{\mathcal{A}} := [a_i]_{\mathcal{A}} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{n\times 1}$$

Similarly, for a change of basis from $\mathcal{A} \to \mathcal{B}$:

$$T_{\mathcal{A}} := [a_i]_{\mathcal{B}} = \begin{pmatrix} t_{11} \\ t_{21} \\ \vdots \\ t_{n1} \end{pmatrix}$$

Proposition 1 *Let $\mathcal{A}, \mathcal{B}$ be two basis of vector space V. Then the matrices $M(I, \mathcal{A}, \mathcal{B})$ and $M(I, \mathcal{B}, \mathcal{A})$ are invertible and each is the inverse of the other.*

$$T_{\mathcal{A}\to\mathcal{B}} = T^{-1}_{\mathcal{B}\to\mathcal{A}}$$

Proposition 2 *Let $\mathcal{A}$ and $\mathcal{A}$ be two bases of vector space V. Consider the transformation matrix*

$$A = M(I, \mathcal{A}, \mathcal{B}) \text{ and } A^{-1} = M(I, \mathcal{B}, \mathcal{A})$$

Let $T : V \to V$ be a linear map, and $X = M(T, \mathcal{A}, \mathcal{A})$. Then, $Y := A.X.A^{-1}$ represents T in the basis $\mathcal{B}$, i.e., $Y = M(T, \beta, \beta)$.

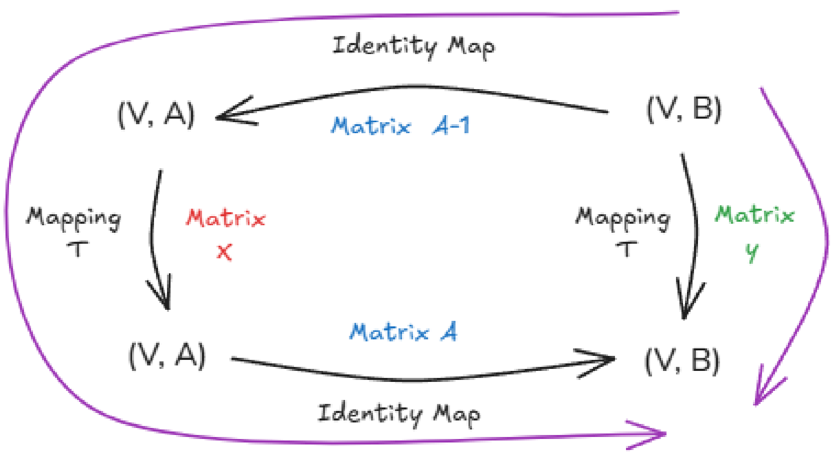

Figure 1: Two paths to Mapping (purple)

3 Rank of a Matrix

The rank of a matrix represents the maximum number of linearly independent rows or columns in the matrix. It provides important information about the matrix's structure and properties, such as:

1. **Linear Independence**: The rank indicates how many of the rows (or columns) are linearly independent. If a matrix has full rank, all of its rows or columns are independent.
2. **Solutions to Linear Systems**: The rank is related to the number of solutions a system of linear equations has. If the rank is equal to the number of variables (columns), the system has a unique solution (assuming it is consistent). If the rank is less than the number of variables, the system may have infinitely many solutions or none, depending on consistency.
3. **Invertibility**: A square matrix is invertible if and only if it has full rank, meaning its rank equals its number of rows or columns.

4. **Dimension of the Row and Column Spaces**: The rank of a matrix also tells you the dimensions of the row space (the span of the rows) and the column space (the span of the columns).

In summary, the rank reflects the degree of linear independence and the overall structural richness of the matrix.

Let $A \in \mathbb{F}^{m \times n}$. Then the **column rank** of A is dim(span(column vectors of A)) and the **row rank** is dim(span(row vectors of A)).

3.1 Properties

1. The row and column rank of a matrix are always the same. We can generalize by calling it the **rank** of the matrix.

2. This implies that the rank of a matrix A is the same as its transpose A^T.

$$rank(A) = rank(A^T)$$

3. For a mapping $T \in \mathcal{L}(V, W)$, there exists a mapping matrix $M(T, \mathcal{A}, \mathcal{B})$ whose rank is independent of the choice of the basis, and represented as $rank(M(T))$.

4. Let $T \in \mathcal{L}(V, W)$. Then the rank of the mapping matrix $M(T)$ is equal to the dimension of range of the mapping.

$$rank(M(T)) = dim(range(T))$$

4 Determinant of a Matrix

A linear mapping $d : \mathbb{F}^{n \times n} \to \mathbb{F}$ is called a determinant if:

- It is **multilinear**. Let A be a matrix with columns $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ represented as.

$$A = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \ldots & a_n \\ | & | & & | \end{bmatrix} = (a_1, a_2, \ldots, a_n)$$

  Consider the column $\mathbf{a}_i$, assume $\mathbf{a}_i = \mathbf{a}_i' + \mathbf{a}_i''$. Then

$$\det((a_1, a_2, \ldots, a_i, \ldots, a_n)) = \det((a_1, a_2, \ldots, a_i', \ldots, a_n)) + \det((a_1, a_2, \ldots, a_i'', \ldots, a_n))$$

$$\det((a_1, a_2, \ldots, \lambda.a_i, \ldots, a_n)) = \lambda.\det((a_1, a_2, \ldots, a_i, \ldots, a_n))$$

- It is **alternating**: If A has two identical columns, then $\det(A) = 0$
- It is **normed**: $\det(\mathbb{I}) = 1$

4.1 Properties

1. Similar to rank, determinant of a mapping matrix does not depend on choice of the basis.

2. The mapping d exists and is unique for all matrices.

3. **Additional Properties:**

- $\det(c.A) = c^n \det(A); A \in \mathbb{F}^{m\times n} \;\; \& \;\; c \in \mathbb{F}$
- $\det(A.B) = \det(A).\det(B)$
- $\det(A^T) = \det(A)$
- $\det(A^{-1}) = \frac{1}{\det(A)}$(if A is invertible)
- If A is invertible then $\det(A) \neq 0$
- $\det(A + B) \neq \det(A) + \det(B)$
- If A is a upper triangular matrix, then $\det(A)$ is the product of the diagonals of the matrix
- If you swap any two rows or columns of a matrix, the determinant of the matrix changes its sign

4.2 Calculating Determinant

Leibniz Formula:

The **Leibniz Formula** provides a way to compute the determinant of an $n\times n$ matrix. It is expressed as:

$$\det(A) = \sum_{\sigma \in S_n} \text{sign}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)}$$

Where:

- $A = [a_{i,j}]$ is an $n \times n$ matrix.
- S_n is the set of all permutations of $\{1, 2, 3, \ldots, n\}$.
- σ is a permutation in S_n.
- $\text{sign}(\sigma)$ is the sign of the permutation σ which is $+1$ for even permutations and -1 for odd permutations.
- $a_{i,\sigma(i)}$ represents the elements of the matrix corresponding to the permutation σ.

The Leibniz formula expresses the determinant as a sum over all permutations of the matrix indices, multiplying the appropriate matrix entries for each permutation and accounting for the sign of the permutation.

Laplace Formula:

The **Laplace Formula** (also known as cofactor expansion) provides a recursive method to compute the determinant of a matrix. It expresses the determinant of an $n \times n$ matrix in terms of the determinants of smaller $(n-1) \times (n-1)$ matrices.

For a square matrix $A = [a_{ij}]$, the Laplace expansion along the first row is given by:

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

Where:

- A_{1j} is the submatrix formed by removing the first row and the j-th column of A.
- a_{1j} are the elements of the first row of A.
- The factor $(-1)^{1+j}$ is the sign factor that alternates for each term.

More generally, for any row i or column j, the Laplace expansion is:

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij})$$

Where:

- A_{ij} is the submatrix formed by removing the i-th row and the j-th column of A.

The determinant of a matrix can be computed by expanding along any row or column, but typically the first row or column is chosen to simplify the calculations.

4.3 Special Cases

- n=1, $\det(a) = a$
- n=2, $\det\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right) = a_{11}a_{22} - a_{12}a_{21}$
- n=3, $\det\left(\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}\right) = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$

4.4 Geometric Intuition

Imagine a 2×2 matrix with columns a_1, a_2. We can form a parallelogram in a 2-dimensional plane using the column vectors a_1, a_2.

For a matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, the column vectors would be $a_1 = a\hat{i} + c\hat{j}, a_2 = b\hat{i} + d\hat{j}$. The resulting parallelogram would look like this:

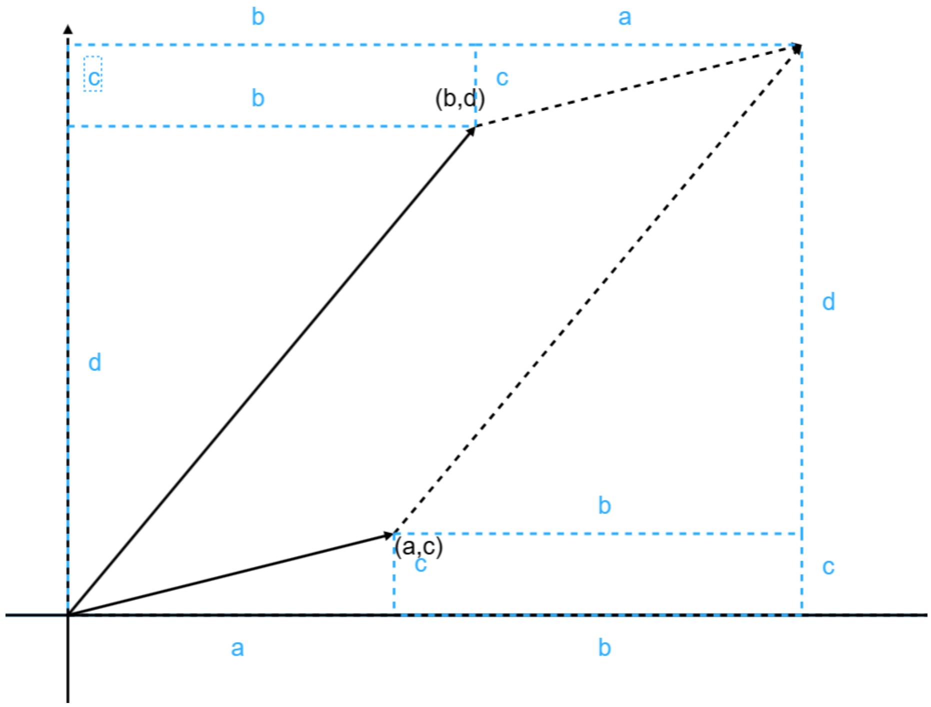

Figure 2: Parallelogram formed by column vectors

The area of the parallelogram can be calculated by subtracting the triangles and rectangles from the bigger rectangle.

$$\text{Area}_{//} = (a+b)(c+d) - \frac{1}{2}ac - \frac{1}{2}bd - bc - \frac{1}{2}ac - \frac{1}{2}bd - bc$$

$$\text{Area}_{//} = (ac + bc + ad + bd) - ac - bd - 2bc$$

$$\text{Area}_{//} = ad - bc = \det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right)$$

Essentially, the area of the parallelogram formed by the column vectors of A is its determinant. If we flipped the column vectors, the sign of the determinant would change since the new area would be $bc - ad$.

In a 3-dimensional plane, we can form a parallelotope with the column vectors of the matrix $A_{3\times 3}$. Using the same logic, we can assess that the volume of the parallelotope formed by the column vectors a_1, a_2, a_3 is the determinant of the matrix A. But this gets harder as we go into an n-dimensional plane where we are unable to assess the shape formed by the column vectors, or calculate the determinant of an $n \times n$ matrix. This forces us to intuitively think about how we can reduce the resulting shape into something less computationally intensive.

In case of our 2-D matrix, if we rotate and translate the column vectors to form the new coordinate axes X_{a_1} and Y_{a_2}, the resulting shape would just be a rectangle in this new axes. In case of a 3-D matrix, we would just turn the parallelotope into a rectangular prism (cuboid) and the volume would just be the product of the dimensions of the new cuboid ($V = lbh$).

In these translated axes, the lengths of the shape would also be squished and morphed to form, lets say, $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ as the new lengths where n is the number of dimensions of the shape formed by the column vectors of an $n \times n$ matrix. These can be called the eigenvalues of the matrix. Essentially, the original matrix A would be converted this way:

$$A_{\mathcal{R}^n} = \begin{bmatrix} a_{11} & a_{21} & \ldots & a_{n1} \\ a_{12} & a_{22} & \ldots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \ldots & a_{nn} \end{bmatrix} \rightarrow A_{\text{new axes}} = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix}$$

For an $n \times n$ matrix A with columns $(a_1, a_2, ..., a_n)$, the determinant gives the signed volume of the parallelotope formed by these vectors.

$$\det(A_{n\times n}) = \Pi_{i=1}^{n}(\lambda_i) = \lambda_1.\lambda_2 \ldots .\lambda_n.$$

4.5 Application to Integrals

The importance of determinants gathered from the insights we derived from the geometric intuition is more noticeable in multivariate calculus, especially evaluating integrals over transformed regions.

Proposition 3 *Let $\Omega \in \mathbb{R}$ be an open set, $\sigma : \Omega \rightarrow \mathbb{R}^n$ be differentiable, and $f : \sigma(\Omega) \rightarrow \mathbb{R}$. Then:*

$$\int_{\sigma(\Omega)} f(y)dy = \int_{\Omega} f(\sigma(x))|\det(\sigma'(x))|dx$$

Lets break this down: we have a differentiable mapping $\sigma : \Omega \rightarrow \mathbb{R}^n$ and a function f defined on $\sigma(\Omega)$. To accurately compute the integral of f over the transformed region $\sigma(\Omega) \in \mathbb{R}^n$, we can calculate the integral in the original space $\Omega \in R^n$ and relate it to the transformed space using the change of variables formula. This is done by multiplying the integrand $f(\sigma(x))$ where $x \in \Omega$ by the absolute value of the Jacobian determinant $|\det(\sigma'(x)|$. This adjustment works on the same principle of scaling the axes. In this case, we are adjusting for each volume element after transformation. Essentially, the Jacobian determinant quantifies the factor by which the mapping operator σ scales the volume near that point.