# Lecture 20 - Matrix Transpose, Quadratic Forms, and Orthogonal Matrices

November 14, 2022

**Goals:** Be able to compute the transpose of a matrix and use it to invert orthogonal matrices.

Often times, it is convenient to swap the rows and columns of a matrix:

**Definition:** Given an $m \times n$ matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix},$$

its **transpose**, denoted $A^\top$, is defined to be the $n \times m$ matrix obtained by turning the rows into columns and columns into rows. The $i$th *row* of $A^\top$ is the $i$th *column* of $A$:

$$A^\top = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}.$$

Equivalently, the $ij$-entry of $A^\top$ = $ji$-entry of $A$.

**Facts:** $(A^\top)^\top = A$, and $(A+B)^\top = A^\top + B^\top$. respects matrix addition

**Example 1:** For the matrices $A$ and $B$ below, calculate $A^\top$, $B^\top$, $(AB)^\top$, and $B^\top A^\top$.

$$\underset{2\times3}{A = \begin{bmatrix} 1 & 2 & 3 \\ -2 & 0 & 1 \end{bmatrix}}, \qquad \underset{3\times1}{B = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}}$$

$$\underset{3\times2}{A^T} = \begin{bmatrix} 1 & -2 \\ 2 & 0 \\ 3 & 1 \end{bmatrix} \qquad \underset{2\times1}{AB} = \begin{bmatrix} 3 \\ -2 \end{bmatrix} \qquad B^TA^T = \underset{1\times3}{\begin{bmatrix} 1 & 1 & 0 \end{bmatrix}} \underset{3\times2}{\begin{bmatrix} 1 & -2 \\ 2 & 0 \\ 3 & 1 \end{bmatrix}}$$

$$\underset{1\times3}{B^T} = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \qquad \underset{1\times2}{(AB)^T} = \begin{bmatrix} 3 & -2 \end{bmatrix} \qquad = \underset{1\times2}{\begin{bmatrix} 3 & -2 \end{bmatrix}}$$

$$(AB)^T = B^TA^T$$

**Example 2:** For $A$ below, compute $A^{-1}$, $(A^{-1})^\top$, and $(A^\top)^{-1}$.

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}, \quad \det(A) = ad-bc = 2-0 = 2$$

$$A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$$= \frac{1}{2}\begin{bmatrix} 2 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1/2 \\ 0 & 1/2 \end{bmatrix}$$

$$(A^{-1})^T = \begin{bmatrix} 1 & 0 \\ -1/2 & 1/2 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 2 & 0 \\ -1 & 1 \end{bmatrix}$$

$$A^T = \begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix}, \quad \det(A^T) = 2-0 = 2$$

$$(A^T)^{-1} = \frac{1}{2}\begin{bmatrix} 2 & 0 \\ -1 & 1 \end{bmatrix}$$

- $(A^{-1})^T = (A^T)^{-1}$
- for $2\times2$, $\det(A) = \det(A^T)$

**Example 3:** Compute $v^T w$ [redacted] for $\mathbf{v} = \begin{bmatrix}1\\1\\1\end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix}2\\1\\3\end{bmatrix}$.

$$v^T w = \begin{bmatrix}1 & 1 & 1\end{bmatrix}\begin{bmatrix}2\\1\\3\end{bmatrix} = \begin{bmatrix}6\end{bmatrix} \qquad v \cdot w = 6$$

$v^T w = v \cdot w$ for n-vectors

The ideas we saw from the first three examples are actually true in general and are summarized as follows:

**More properties of matrix algebra:**

- (T1) It reverses the order of matrix multiplication $(AB)^\top = B^\top A^\top$.
- (T2) If an $n \times n$ matrix $A$ is invertible, then so is $A^\top$, with $(A^\top)^{-1} = (A^{-1})^\top$.
- (T3) If $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ are viewed as $n \times 1$ matrices, then

$$\mathbf{v}^\top \mathbf{w} = \mathbf{v} \cdot \mathbf{w},$$

  where $\mathbf{v} \cdot \mathbf{w}$ is viewed as a $1 \times 1$ matrix.

When dealing with dot products, we have seen that we can move scalars from one vector to another or just factor it out completely. Here is what happens if we try it with matrices:

$\Rightarrow \vec{x} \cdot (c\,\vec{y}) = c\,(\vec{x} \cdot \vec{y})$

**Theorem 20.1.4**: For any $m \times n$ matrix $A$ and vectors $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$,

$$\overbrace{(\underset{m\times n}{A}\underset{n\times1}{\mathbf{x}})}^{m\times1} \cdot \underset{m\times1}{\mathbf{y}} = \underset{n\times1}{\mathbf{x}} \cdot \overbrace{(\underset{n\times m}{A^\top}\underset{m\times1}{\mathbf{y}})}^{n\times1}.$$

In other words: a matrix moves across a dot product via its transpose.

**Example 4:** Suppose a factory is to produce 200 units of the Ferrari "296 GTS" and 120 units of the Ferrari "812 GTS". Suppose each car is made of three raw materials - steel, plastic, and aluminum; the materials cost $p_1$, $p_2$, and $p_3$ dollars / kg respectively. If the "296 GTS" needs 750kg of steel, 120kg of plastics, and 180kg of aluminum to produce, and the "812 GTS" needs 1200kg of steel, 170kg of plastics, and 240kg of aluminum, find an expression for the total cost of the factory to produce the cars.

$$\vec{q} = \begin{bmatrix}200\\120\end{bmatrix}\begin{array}{l}\text{units of "296"}\\ \text{units of "812"}\end{array}, \qquad \vec{p} = \begin{bmatrix}p_1\\p_2\\p_3\end{bmatrix}\begin{array}{l}\text{price of steel /kg}\\ \text{price of plastic /kg}\\ \text{price of alu / kg}\end{array}$$

$$\text{Let } A = \begin{array}{cccl} \text{amount of steel} & \text{plastic} & \text{alu} & \\ 750 & 120 & 180 & \text{\# of materials for "296"} \\ 1200 & 170 & 240 & \text{\# of materials for "812"} \end{array}$$

Total cost = (kg needed × price /kg)

$$\text{kg needed} = A^T \vec{q} = \begin{bmatrix}750 & 1200\\120 & 170\\180 & 240\end{bmatrix}\begin{bmatrix}200\\120\end{bmatrix} = \begin{bmatrix}750(200) + 1200(120)\\120(200) + 170(120)\\180(200) + 240(120)\end{bmatrix}\begin{array}{l}\text{total kg of steel}\\ \text{total kg of plastic}\\ \text{total kg of alu}\end{array}$$

$$\text{Total cost} = \vec{p} \cdot (A^T \vec{q})$$
$$= p_1(750 \cdot 200 + 1200 \cdot 120)$$
$$+ p_2(120 \cdot 200 + 170 \cdot 120)$$
$$+ p_3(180 \cdot 200 + 240 \cdot 120)$$

**Method 2**

$$\text{Total cost} = (\underbrace{\text{price /car}}_{A\vec{p}} \times \underbrace{\text{\# of cars}}_{\vec{q}})$$

$$= \begin{bmatrix}750p_1 + 120p_2 + 180p_3\\1200p_1 + 170p_2 + 240p_3\end{bmatrix} \cdot \begin{bmatrix}200\\120\end{bmatrix} = (A\vec{p}) \cdot \vec{q}$$

= total cost

**Definition:** A matrix is called **symmetric** if $A^\top = A$ (in other words, the $ij$-entry and the $ji$-entry coincide for all $i, j$). A symmetric matrix is always square.

**Example 5:** Let $f(x, y) = x^2 + 2xy$. Compute $D(\nabla f)(x, y)$ and its transpose.

$$\nabla f = \begin{bmatrix} 2x+2y \\ 2x \end{bmatrix}, \quad D(\nabla f)(x,y) = \begin{bmatrix} \leftarrow \nabla(2x+2y) \rightarrow \\ \leftarrow \nabla(2x) \rightarrow \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 2 & 0 \end{bmatrix}$$

$$\left(D(\nabla f)\right)^T(x,y) = \begin{bmatrix} 2 & 2 \\ 2 & 0 \end{bmatrix}^T = \begin{bmatrix} 2 & 2 \\ 2 & 0 \end{bmatrix}$$ so $D(\nabla f)$ is symmetric

proof

$$\nabla f = \begin{bmatrix} f_x \\ f_y \end{bmatrix}, \quad D(\nabla f) = \begin{bmatrix} - \nabla(f_y) - \\ - \nabla(f_y) - \end{bmatrix} = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}$$

we know $f_{xy} = f_{yx}$, so diagonals are same & its symmetric

$\longrightarrow$ Hessian matrix

**Example 6:** If $A$ is symmetric and invertible, is it true that $A^{-1}$ must be symmetric?

$(A^{-1})^T = (A^T)^{-1}$, since $A$ is symmetric, $A = A^T$,

so $(A^T)^{-1} = A^{-1}$, so $A^{-1} = (A^{-1})^T$, so $A^{-1}$ is symmetric

**Theorem 20.3.8:** For any $m \times n$ matrix $M$, the $n \times n$ matrix $M^\top M$ and the $m \times m$ matrix $MM^\top$ are symmetric (these are called **Gram** matrices).

transpose reverses order

$(M^TM)^T = M^T M^{TT} = M^T M \therefore M^TM = (M^TM)^T$, so $M^TM$ is symmetric

$(MM^T)^T = M^{TT}M^T = MM^T \therefore MM^T = (MM^T)^T$, so $MM^T$ is symmetric

One use of symmetric matrices is as follows:

Let $A$ be a symmetric $n \times n$ matrix. For any vector $\mathbf{v} \in \mathbb{R}^n$, the product $\mathbf{v}^\top A\mathbf{v}$ is the $1 \times 1$ matrix with entry $\mathbf{v} \cdot (A\mathbf{v})$. This type of construction shows up in a lot of places, so we give it the notation $q_A(\mathbf{v}) = \mathbf{v}^\top \cdot (A\mathbf{v})$. We call $q_A$ a **quadratic form**. For $n = 2, 3$, it is worth "memorizing" rather than computing it each time:

$$\begin{bmatrix} x \\ y \end{bmatrix}^\top \overset{2\times2}{\begin{bmatrix} a & u \\ u & b \end{bmatrix}} \overset{2\times1}{\begin{bmatrix} x \\ y \end{bmatrix}} = ax^2 + by^2 + 2uxy, \qquad 2\times2$$

Diagonal entries $c_{ii}$ is coefficient for $x_i^2$, and $c_{ij} = c_{ji}$ multiplies $2x_ix_j$.

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}^\top \begin{bmatrix} a & u & v \\ u & b & w \\ v & w & c \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = ax^2 + by^2 + cz^2 + 2uxy + 2vxz + 2wyz. \qquad 3\times3$$

The general pattern in $q_A(\mathbf{x})$ for an $n \times n$ symmetric matrix $A = (c_{ij})$ is that the diagonal entry $c_{ii}$ multiplies against $x_i^2$ and the off diagonal entry $c_{ij} = c_{ji}$ multiplies against $x_ix_j$ up to a factor of 2. Given a quadratic form $Q$, we can find a **unique** *symmetric* matrix $A$ for which $Q = q_A$.

Quadratic forms will show up later in Chapter 26. For now, we will use them to discuss a special class of matrices whose geometric meaning is visualized in terms of "rigid motions" (length preserving) of $\mathbb{R}^n$.

if you rotate, length doesn't change

Let $R$ be the matrix of a rotation of $\mathbb{R}^3$ around the origin. Note that $\mathbf{v}$ and $R\mathbf{v}$ must have the same length, i.e. $||R\mathbf{v}|| = ||\mathbf{v}||$, which implies $||R\mathbf{v}||^2 = ||\mathbf{v}||^2$, which then means $\mathbf{v}^\top\mathbf{v} = (R\mathbf{v})^\top(R\mathbf{v})$. We obtain:

$$\mathbf{v}^\top\mathbf{v} = \underbrace{(R\mathbf{v})^\top(R\mathbf{v}) = \mathbf{v}^\top(\overset{A\vec{v}}{R^\top R\mathbf{v}})}_{} = q_{R^\top R}(\mathbf{v})$$

$v \cdot v = (Rv) \cdot (Rv)$

$q_A(\vec{v}) = v \cdot (A\vec{v})$

$(Rv) \cdot (Rv)$
$= v \cdot (R^T R v)$ theorem 20.1.4

for any $\mathbf{v} \in \mathbb{R}^3$. This means $q_{R^\top R}(\mathbf{v}) = ||\mathbf{v}||^2 = q_{I_3}(\mathbf{v})$, so we *must* have that $R^\top R = I_3$, i.e. the inverse of $R$ is

$q_{I_3}(\vec{v}) = \vec{v} \cdot (I_3\vec{v}) = \vec{v} \cdot \vec{v} = v^Tv$

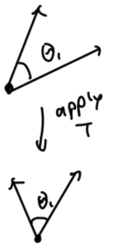
$\Rightarrow I_3 = R^TR \Rightarrow R^T = R^{-1}$

its transpose $R^\top$. Moreover, since the $ij$-entry of $R^\top R$ is the dot product of the $i$th and $j$th columns of $R$, this means that the columns are orthonormal.

We extend these ideas to a general $m \times n$ matrix $A$.

**Theorem 20.4.1:** The following conditions are equivalent for a general $m \times n$ matrix $A$:

- The linear transformation $T_A : \mathbb{R}^n \to \mathbb{R}^m$ is length preserving. → $\|T\vec{v}\| = \|\vec{v}\|$
- $A^\top A = I_n$.
- (only when $m = n$) $AA^\top = I_n$. columns are orthogonal
- ☆ The $n$ columns of $A$ are an orthonormal collection of $m$-vectors
- (only when $m = n$) The rows of $A$ are an orthonormal collection.



$\theta_1 = \theta_2$ angle preserving

If these hold, then $T_A$ is also angle preserving, so when $m = n$ it is a "rigid motion" of $\mathbb{R}^n$ fixing $\mathbf{0}$. → Rotations / reflections

**Definition:** Any square matrix satisfying the equivalent conditions above is called an **orthogonal matrix**.

**Proposition 20.4.4:** If $A$ is orthogonal, then $A^{-1} = A^\top$.

**Example 7:** Solve the following system of equations:

$$\begin{aligned} \frac{1}{\sqrt{3}}x + \frac{2}{\sqrt{6}}z &= 1 \\ \frac{1}{\sqrt{3}}x + \frac{1}{\sqrt{2}}y - \frac{1}{\sqrt{6}}z &= 0 \\ -\frac{1}{\sqrt{3}}x + \frac{1}{\sqrt{2}}y + \frac{1}{\sqrt{6}}z &= -1 \end{aligned}$$

$$A = \begin{array}{c} \begin{array}{ccc} c_1 & c_2 & c_3 \end{array} \\ \begin{bmatrix} 1/\sqrt{3} & 0 & 2/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{2} & -1/\sqrt{6} \\ -1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \end{bmatrix} \end{array}$$

☆ $\|c_1\| = \sqrt{1/3 + 1/3 + 1/3} = 1$

$\|c_2\| = 1$

$\|c_3\| = 1$

$c_1 \cdot c_2 = 0$

$c_2 \cdot c_3 = 0$

∴ columns of A are orthonormal

∴ A is an orthogonal matrix

→ Since A is orthogonal,

$A^{-1} = A^T$

$\Rightarrow \vec{x} = A^{-1}\vec{b}$

$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = A^T \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$

For an orthogonal matrix $O$, $O^{-1} = O^T$

Proof:

Let $O = \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{bmatrix}$, $O^T = \begin{bmatrix} — & v_1 & — \\ — & v_2 & — \\ & \vdots & \\ — & v_n & — \end{bmatrix}$

$v_1 \perp$ to all $v_i$

$\|v_i\| = 1 \rightarrow v_i \cdot v_i = 1$

$$O^T O = \begin{bmatrix} — & v_1 & — \\ — & v_2 & — \\ & \vdots & \\ — & v_n & — \end{bmatrix} \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{bmatrix} \longrightarrow \begin{array}{cc} & \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{bmatrix} \\ \begin{bmatrix} — & v_1 & — \\ — & v_2 & — \\ & \vdots & \\ — & v_n & — \end{bmatrix} & \begin{bmatrix} v_1 \cdot v_1 & v_1 \cdot v_2 & \cdots & v_1 \cdot v_n \\ v_2 \cdot v_1 & v_2 \cdot v_2 & \cdots & v_2 \cdot v_n \\ \vdots & \vdots & \ddots & \\ v_n \cdot v_1 & v_n \cdot v_2 & & v_n \cdot v_n \end{bmatrix} \end{array} \leftarrow \text{result}$$

$$\downarrow$$

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & & 1 \end{bmatrix} = I_n \quad \checkmark$$