ORTHOGONAL MATRICES AND THE TRANSPOSE

1. Pythagorean Theorem and Cauchy Inequality

We wish to generalize certain geometric facts from $\mathbb{R}^2$ to $\mathbb{R}^n$.

Theorem 1.1. *(Pythagorean Theorem) Given two vectors $\vec{x}, \vec{y} \in \mathbb{R}^n$ we have*

$$||\vec{x}+\vec{y}||^2 = ||\vec{x}||^2 + ||\vec{y}||^2 \iff \vec{x}\cdot\vec{y} = 0.$$

Proof. One computes

$$||\vec{x}+\vec{y}||^2 = (\vec{x}+\vec{y})\cdot(\vec{x}+\vec{y}) = ||\vec{x}||^2 + 2\vec{x}\cdot\vec{y} + ||\vec{y}||^2.$$

Hence, $||\vec{x}+\vec{y}||^2 = ||\vec{x}||^2 + ||\vec{y}||^2 \iff \vec{x}\cdot\vec{y} = 0$. $\square$

A consequence, of this is that for any subspace, $V \subset \mathbb{R}^n$, $||\text{proj}_V(\vec{x})|| \leq ||\vec{x}||$. To see this, note that $\vec{x} = \text{proj}_V(\vec{x}) + \vec{x}^\perp$ and two terms are orthogonal. Hence, by the Pythagorean theorem

$$||\vec{x}||^2 = ||\text{proj}_V(\vec{x})||^2 + ||\vec{x}^\perp||^2 \geq ||\text{proj}_V(\vec{x})||^2$$

Theorem 1.2. *(Cauchy-Schwarz Inequality) If $\vec{x}, \vec{y} \in \mathbb{R}^n$, then $|\vec{x}\cdot\vec{y}| \leq ||\vec{x}||||\vec{y}||$.*

Proof. Suppose $\vec{y} \neq 0$ (its trivial otherwise). Let $\vec{u} = \frac{1}{||\vec{y}||}\vec{y}$. If $V = \text{span}(\vec{y}) = \text{span}(\vec{u})$, then

$$\text{proj}_V(\vec{x}) = ||(\vec{x}\cdot\vec{u})\vec{u}|| = |\vec{x}\cdot\vec{u}| = \frac{|\vec{x}\cdot\vec{y}|}{||\vec{y}||}.$$

Hence,

$$||\vec{x}|| \geq ||\text{proj}_V(\vec{x})|| = \frac{|\vec{x}\cdot\vec{y}|}{||\vec{y}||}$$

multiplying through completes proof. $\square$

Recall, for non-zero $\vec{x}, \vec{y} \in \mathbb{R}^2$ we observed that

$$\vec{x}\cdot\vec{y} = ||\vec{x}||||\vec{y}||\cos\theta$$

where θ was the angle between $\vec{x}$ and $\vec{y}$. The Cauchy-Schwarz inequality implies we can define the angle, θ, between non-zero $\vec{x}, \vec{y} \in \mathbb{R}^n$ in general to be

$$\theta = \arccos \frac{\vec{x}\cdot\vec{y}}{||\vec{x}||||\vec{y}||}.$$

2. Orthogonal Transformations and Matrices

Linear transformations that preserve length are of particular interest. A linear transform $T : \mathbb{R}^n \to \mathbb{R}^n$ is *orthogonal* if for all $\vec{x} \in \mathbb{R}^n$

$$||T(\vec{x})|| = ||\vec{x}||.$$

Likewise, a matrix $U \in \mathbb{R}^{n\times n}$ is *orthogonal* if $U = [T]$ for T an orthogonal transformation. That is, for all $\vec{x}$,

$$||U\vec{x}|| = ||\vec{x}||.$$

EXAMPLE: $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$, rotation counter-clockwise by θ, is orthogonal.

NON-EXAMPLE: If $V \neq \mathbb{R}^n$, then $\text{proj}_V : \mathbb{R}^n \to \mathbb{R}^n$ is not orthogonal. Indeed, $\vec{w} \notin V$ satisfies $||\text{proj}_V(\vec{w})|| < ||\vec{w}||$.

3. Properties of Orthogonal transformations

Orthogonal transformations are so called as they preserve orthogonality:

Theorem 3.1. *If $T : \mathbb{R}^n \to \mathbb{R}^n$ is orthogonal and $\vec{v} \cdot \vec{w} = 0$, then $T(\vec{v}) \cdot T(\vec{w}) = 0$.*

The matrices kI_n preserve orthogonality, but are only orthogonal when $|k| = 1$.

Proof. We have

$$||T(\vec{v}) + T(\vec{w})||^2 = ||T(\vec{v} + \vec{w})||^2 = ||\vec{v} + \vec{w}||^2 = ||\vec{v}||^2 + ||\vec{w}||^2 = ||T(\vec{v})||^2 + ||T(\vec{w})||^2$$

Hence, the claim follows from the Pythagorean theorem. □

We have

(1) Alinear tranform $T : \mathbb{R}^n \to \mathbb{R}^n$ is orthogonal if and only if $T(\vec{e}_1), \dots, T(\vec{e}_n)$ is an orthonormal basis of $\mathbb{R}^n$
(2) Similar, $U \in \mathbb{R}^{n\times n}$ is orthogonal if and only if the columns of U form an orthonormal basis of $\mathbb{R}^n$.

To see the first claim, note that if T is orthogonal, then by definition $T(\vec{e}_i)$ is unit and the previous result implies $T(\vec{e}_i) \cdot T(\vec{e}_j) = 0$ for $i \neq j$ (as $\vec{e}_i \cdot \vec{e}_j = 0$). Hence, $T(\vec{e}_1), \dots, T(\vec{e}_n)$ is an orthonormal basis. Conversely, suppose $T(\vec{e}_1), \dots, T(\vec{e}_n)$ forms an orthonormal basis. Consider $\vec{x} = x_1\vec{e}_1 + \dots + x_n\vec{e}_n$ We compute

$$||T(\vec{x})||^2 = ||x_1T(\vec{e}_1) + \dots + x_nT(\vec{e}_n)||^2 = ||x_1T(\vec{e}_1)||^2 + \dots + ||x_nT(\vec{e}_n)||^2 = ||\vec{x}||^2.$$

This completes the proof of Claim (1). The second claim is immediate.

Theorem 3.2. *If $A, B \in \mathbb{R}^{n\times n}$ are orthogonal, then so is AB. Moreover, A is invertible and A^{-1} is also orthogonal.*

Proof. As A and B are orthogonal, we have for any $\vec{x} \in \mathbb{R}^n$

$$||AB\vec{x}|| = ||A(B\vec{x})|| = ||B\vec{x}|| = ||\vec{x}||.$$

This proves the first claim. For the second claim, note that if $A\vec{z} = \vec{0}$, then

$$||\vec{z}|| = ||A\vec{z}|| = ||\vec{0}|| = 0$$

hence, $\vec{x} = \vec{0}$. In particular, $\ker(A) = \left\{\vec{0}\right\}$ so A is invertible. Furthermore,

$$||A^{-1}\vec{x}|| = ||A(A^{-1}\vec{x})|| = ||\vec{x}||$$

so A^{-1} is orthonormal. □

4. Transpose

Consider an $m \times n$ matrix A. The *transpose*, $A^\top$, of A is the $n \times m$ matrix whose entry in the ith row and jth column is the entry of A in the jth row and ith column. Geometrically, $A^\top$ is obtained from A by reflecting across the diagonal of A.We say A is *symmetric* if $A^\top = A$ and A is *skew-symmetric* if $A^\top = -A$.

EXAMPLE:

$$\begin{bmatrix} 2 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}^\top = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

We compute for $\vec{v}, \vec{w} \in \mathbb{R}^n$ that

$$\vec{v} \cdot \vec{w} = \vec{v}^\top \vec{w}$$

EXAMPLE:

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}^\top \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 1 \end{bmatrix} = 1(-1) + 2(1) = 1.$$

Theorem 4.1. *$A \in \mathbb{R}^{n \times n}$ is orthogonal if and only if $A^\top A = I_n$, i.e., $A^{-1} = A^\top$.*

Proof. Let

$$A = \begin{bmatrix} \vec{v}_1 & | & \cdots & | & \vec{v}_n \end{bmatrix}.$$

Hence,

$$A^\top = \begin{bmatrix} \vec{v}_1^\top \\ \vdots \\ \vec{v}_n^\top \end{bmatrix}$$

and so

$$A^\top A = \begin{bmatrix} \vec{v}_1 \cdot \vec{v}_2 & \cdots & \vec{v}_1 \cdot \vec{v}_n \\ \vdots & \ddots & \vdots \\ \vec{v}_n \cdot \vec{v}_i & \cdots & \vec{v}_n \cdot \vec{v}_n \end{bmatrix}.$$

Clearly, $A^\top A = I_n$ if and only if the columns of A are orthonormal. □

5. Properties of Transpose

The transpose has has several natural algebraic properties

(1) $(A+B)^\top = A^\top + B^\top$ for $A, B \in \mathbb{R}^{n \times m}$.
(2) $(kA)^\top = kA^\top$ for $A \in \mathbb{R}^{n \times m}$, $k \in \mathbb{R}$.
(3) $(AB)^\top = B^\top A^\top$ for $A \in \mathbb{R}^{n \times p}, B \in \mathbb{R}^{p \times m}$.
(4) $(A\vec{x}) \cdot \vec{y} = \vec{x} \cdot A^\top \vec{y}$ for $A \in \mathbb{R}^{n \times m}, \vec{x} \in \mathbb{R}^m, \vec{y} \in \mathbb{R}^n$.
(5) $(A^\top)^{-1} = (A^{-1})^\top$ for $A \in \mathbb{R}^{n \times n}$ invertible.

Properties (1) and (2) are straightforward to check. To see property (3), write A and B in terms of of their rows and columns

$$A = \begin{bmatrix} \vec{w}_1^\top \\ \vdots \\ \vec{w}_n^\top \end{bmatrix} \text{ and } B = \begin{bmatrix} \vec{v}_1 & | & \cdots & | & \vec{v}_m \end{bmatrix}$$

Where $\vec{v}_i, \vec{w}_j \in \mathbb{R}^p$. Likewise,

$$A^\top = \begin{bmatrix} \vec{w}_1 & | & \cdots & | & \vec{w}_n \end{bmatrix} \text{ and } B^\top = \begin{bmatrix} \vec{v}_1^\top \\ \vdots \\ \vec{v}_m^\top \end{bmatrix}$$

$$AB = \begin{bmatrix} \vec{w}_1^\top \vec{v}_1 & \cdots & \vec{w}_1^\top \vec{v}_m \\ \vdots & \ddots & \vdots \\ \vec{w}_n^\top \vec{v}_1 & \cdots & \vec{w}_n^\top \vec{v}_m \end{bmatrix} = \begin{bmatrix} \vec{v}_1 \cdot \vec{w}_1 & \cdots & \vec{v}_m \cdot \vec{w}_1 \\ \vdots & \ddots & \vdots \\ \vec{v}_1 \cdot \vec{w}_n & \cdots & \vec{v}_m \cdot \vec{w}_n \end{bmatrix}$$

$$B^\top A^\top = \begin{bmatrix} \vec{v}_1^\top \vec{w}_1 & \cdots & \vec{v}_1^\top \vec{w}_n \\ \vdots & \ddots & \vdots \\ \vec{v}_m^\top \vec{w}_1 & \cdots & \vec{v}_m^\top \vec{w}_n \end{bmatrix} = \begin{bmatrix} \vec{v}_1 \cdot \vec{w}_1 & \cdots & \vec{v}_1 \cdot \vec{w}_n \\ \vdots & \ddots & \vdots \\ \vec{v}_m \cdot \vec{w}_1 & \cdots & \vec{v}_m \cdot \vec{w}_n \end{bmatrix}$$

One concludes that $(AB)^\top = B^\top A^\top$. To see Property (4) use

$$(A\vec{x})\cdot\vec{y} = (A\vec{x})^\top\vec{y} = (\vec{x}^\top A^\top)\vec{y} = \vec{x}^\top(A^\top\vec{y}) = \vec{x}\cdot(A^\top\vec{y}).$$

Finally, Property (5) can be seen by noting that, as A is invertible,

$$I_n = AA^{-1} \Rightarrow I_n = I_n^\top = (AA^{-1})^\top = (A^{-1})^\top A^\top$$

Hence, $(A^\top)^{-1} = (A^{-1})^\top$.

Theorem 5.1. *Consider a $A\in\mathbb{R}^{n\times n}$. The following are equivalent*

1. *A is orthogonal matrix*
2. *The transformation $T(\vec{x}) = A\vec{x}$ is orthogonal (i.e. preserves length)*
3. *The columns of A form a orthonormal basis of $\mathbb{R}^n$*
4. *$A^\top A = I_n$*
5. *$A^{-1} = A^\top$*
6. *A preserves the dot product, i.e. $A\vec{x}\cdot A\vec{y} = \vec{x}\cdot\vec{y}$*

Proof. We've already seen why (1)-(4) are equivalent. (4) $\iff$ (5) is immediate. Finally,

$$A\vec{x}\cdot A\vec{y} = \vec{x}(A^\top A)\vec{y}$$

So (4)$\Rightarrow$ (6). Going the other direction,

$$||A\vec{x}||^2 = A\vec{x}\cdot A\vec{x}$$

so (6)$\Rightarrow$ (1). □

6. Matrix of an orthogonal projection

Recall, if $\vec{u}_1$ is unit and $V = \text{span}(\vec{u}_1)$, then

$$\text{proj}_V(\vec{x}) = (\vec{x}\cdot\vec{u}_1)\vec{u}_1 = (\vec{u}^\top\vec{x})\vec{u}_1 = \vec{u}_1(\vec{u}_1^\top\vec{x}) = (\vec{u}_1\vec{u}_1^\top)\vec{x}.$$

Hence, in this case

$$[\text{proj}_V] = \vec{u}_1\vec{u}_1^\top \in \mathbb{R}^{n\times n}.$$

More generally, if $\vec{u}_1,\dots,\vec{u}_m$ is an orthonormal basis of $V\subset\mathbb{R}^n$, then we have

$$\text{proj}_V(\vec{x}) = (\vec{x}\cdot\vec{u}_1)\vec{u}_1 + \cdots + (\vec{x}\cdot\vec{u}_m)\vec{u}_m = \begin{bmatrix}\vec{u}_1 & | & \cdots & | & \vec{u}_m\end{bmatrix}\begin{bmatrix}\vec{u}_1^\top \\ \vdots \\ \vec{u}_m^\top\end{bmatrix}\vec{x}$$

That is if

$$Q = \begin{bmatrix}\vec{u}_1 & | & \cdots & | & \vec{u}_m\end{bmatrix},$$

then $\text{proj}_V(\vec{x}) = QQ^\top\vec{x}$ Hence,

$$[\text{proj}_V] = QQ^\top.$$

EXAMPLE: Let

$$\vec{u}_1 = \frac{1}{2}\begin{bmatrix}1\\1\\1\\-1\end{bmatrix}, \vec{u}_2 = \frac{1}{2}\begin{bmatrix}-1\\1\\1\\1\end{bmatrix}$$

be an orthonormal basis of $V \subset \mathbb{R}^4$. Clearly,

$$P = [\mathrm{proj}_V] = \frac{1}{4}\begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 & -1 \\ -1 & 1 & 1 & 1 \end{bmatrix} = \frac{1}{4}\begin{bmatrix} 2 & 0 & 0 & -2 \\ 0 & 2 & 2 & 0 \\ 0 & 2 & 2 & 0 \\ -2 & 0 & 0 & 2 \end{bmatrix}.$$