Worksheet 18: Orthogonal Transformations (§5.3)

Definition: A linear transformation $\mathbb{R}^n \xrightarrow{T} \mathbb{R}^n$ is **orthogonal** if it preserves dot products—that is, if $\vec{x} \cdot \vec{y} = T(\vec{x}) \cdot T(\vec{y})$ for all vectors $\vec{x}$ and $\vec{y}$ in $\mathbb{R}^n$.

Theorem A: Let $\mathbb{R}^n \xrightarrow{T} \mathbb{R}^n$ be a linear transformation. Then T is orthogonal if and only if it preserves the length of every vector—that is, $||T(\vec{x})|| = ||\vec{x}||$ for all $\vec{x} \in \mathbb{R}^n$.

Problem 1. Which of the following maps are orthogonal transformations? Short, geometric justifications are preferred, where possible.

(a) The identity map $\mathbb{R}^3 \to \mathbb{R}^3$.

(b) Rotation counterclockwise through θ in $\mathbb{R}^2$.

(c) The reflection $\mathbb{R}^3 \to \mathbb{R}^3$ over a plane (though the origin).

(d) The projection $\mathbb{R}^3 \to \mathbb{R}^3$ onto a subspace V of dimension 2.

(e) Dilation $\mathbb{R}^3 \to \mathbb{R}^3$ by a factor of 3.

(f) Multiplication by $\begin{bmatrix} 3 & 1 \\ -2 & 5 \end{bmatrix}$.

Solution:

(a) Yes, dot product is obviously the same before or after doing nothing.

(b) Yes, rotation obviously preserves the LENGTH of every vector, so by the theorem, this means rotation is an orthogonal transformation.

(c) Yes, reflection obviously preserves the LENGTH of every vector, so by the theorem, this means reflection is an orthogonal transformation.

(d) No, projection will typically shorten the length of vectors.

(e) No, dilation by three obviously takes each vector to a vector of length 3 times as long.

(f) No. The vector $\vec{e}_1$ is sent to $\begin{bmatrix} 3 \\ -2 \end{bmatrix}$, which has length $\sqrt{13}$, not 1 like $\vec{e}_1$.

Problem 2. Let $\mathbb{R}^n \xrightarrow{T} \mathbb{R}^n$ be an orthogonal transformation.

(a) Prove that T is **injective.** [Hint: consider the kernel.]

(b) Prove that T is an isomorphism. [Hint: Note that the source and target here have the same dimension.]

(c) Prove that the matrix of T (in standard coordinates) has columns that are *orthonormal.*

(d) Prove the composition of orthogonal transformations is orthogonal. [Hint: Use the Theorem!]

Solution:

(a) Let $\vec{v} \in \ker T$. We want to show that $\vec{v} = 0$. We know that $T(\vec{v}) = 0$, so $||T(\vec{v})|| = 0$. Since T is orthogonal, it preserves lengths (by the theorem), so also $||\vec{v}|| = 0$. This means $\vec{v} = 0$ since no other vector has length 0. So the kernel of T is trivial and T is injective.

(b) We already know T is injective, so we just need to check it is surjective. By rank-nullity, $n = \dim \ker T + \text{rank} T$. So by (1), we have $\text{rank} T = n$, so T is surjective. Thus T is a bijective linear transformation, that is, an isomorphism.

(c) Let A be the standard matrix of T. The columns of A are $T(\vec{e}_1), \dots, T(\vec{e}_n)$ by our old friend the Key Theorem. We need to show these are orthonormal. Compute for each i: $T(\vec{e}_i) \cdot T(\vec{e}_i) = \vec{e}_i \cdot \vec{e}_i = 1$ (where we have used the fact that T preserves dot product). Also compute for each pair $i \neq j$: $T(\vec{e}_i) \cdot T(\vec{e}_j) = \vec{e}_i \cdot \vec{e}_j = 0$. This means that $T(\vec{e}_1), \dots, T(\vec{e}_n)$ are orthonormal.

(d) Yes! Assume T and S are orthonormal transformations with source and target $\mathbb{R}^n$. To show that $S \circ T$ is orthogonal, we can show it preserves LENGTHS, by the theorem. Take any vector $\vec{x} \in R^n$. Then $||\vec{x}|| = ||T(\vec{x})||$ since T is orthogonal (by the Theorem, or by the book definition). So also $||T(\vec{x})|| = ||S(T(\vec{x}))||$ since S is orthogonal. Putting these together we have $||\vec{x}|| = ||S(T(\vec{x}))||$. Since $\vec{x}$ was arbitrary, we see that $S \circ T$ preserves every length. QED.

Definition: An $n \times n$ matrix A is **orthogonal** if $A^\top A = I_n$—i.e., if its transpose is its inverse.

Problem 3. Which of the following matrices are orthogonal?

(i) $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ (ii) $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ (iii) $\begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}$ (iv) $\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$.

Solution: Yes, yes, no, no.

Problem 4. Suppose that $A = [\vec{v}_1 \ \vec{v}_2 \ \vec{v}_3]$ is a 3×3 matrix with columns $\vec{v}_1, \vec{v}_2, \vec{v}_3 \in \mathbb{R}^3$.

(a) Recalling that $A^\top = \begin{bmatrix} \vec{v}_1^\top \\ \vec{v}_2^\top \\ \vec{v}_3^\top \end{bmatrix}$, show that

$$A^\top A = \begin{bmatrix} \vec{v}_1 \cdot \vec{v}_1 & \vec{v}_1 \cdot \vec{v}_2 & \vec{v}_1 \cdot \vec{v}_3 \\ \vec{v}_2 \cdot \vec{v}_1 & \vec{v}_2 \cdot \vec{v}_2 & \vec{v}_2 \cdot \vec{v}_3 \\ \vec{v}_3 \cdot \vec{v}_1 & \vec{v}_3 \cdot \vec{v}_2 & \vec{v}_3 \cdot \vec{v}_3 \end{bmatrix}.$$

(b) Does the argument work for any size square matrix?

(c) Use (a)/(b) to prove a square matrix is orthogonal if and only if its columns are orthonormal.

(d) Is $B = \begin{bmatrix} 3/5 & 4/5 & 0 \\ -4/5 & 3/5 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ orthogonal? Find B^{-1}. [Be clever! There are easy ways and hard ways!]

Solution:

(a) Multiply $A^\top A = \begin{bmatrix} \vec{v}_1^\top \\ \vec{v}_2^\top \\ \vec{v}_3^\top \end{bmatrix} \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \vec{v}_3 \end{bmatrix} = \begin{bmatrix} \vec{v}_1^\top \vec{v}_1 & \vec{v}_1^\top \vec{v}_2 & \vec{v}_1^\top \vec{v}_3 \\ \vec{v}_2^\top \vec{v}_1 & \vec{v}_2^\top \vec{v}_2 & \vec{v}_2^\top \vec{v}_3 \\ \vec{v}_3^\top \vec{v}_1 & \vec{v}_3^\top \vec{v}_2 & \vec{v}_3^\top \vec{v}_3 \end{bmatrix}$, which produces the desired matrix since for any $n \times 1$ matrices (column vectors) $\vec{w}$ and $\vec{v}$, the matrix product $\vec{w}^\top \vec{v}$ is the same as the dot product $\vec{w} \cdot \vec{v}$.

(b) Yes! Let A have columns $\vec{v}_1, \ldots, \vec{v}_n$. Then the ij-th entry of $A^\top A$ is $\vec{v}_i^\top \vec{v}_j = \vec{v}_i \cdot \vec{v}_j$. So the columns are orthonormal if and only if $A^\top A$ is the identity matrix.

(c) For a square matrix $A = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix}$, the ij-th entry of $A^\top\ A$ is $\vec{v}_i \cdot \vec{v}_j$ by (b). So the columns are orthonormal if and only if $A^\top\ A = I_n$.

(d) Yes–it's easy to check its columns are orthonormal. The inverse is the transpose, which is $\begin{bmatrix} 3/5 & -4/5 & 0 \\ 4/5 & 3/5 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Problem 5. Show that if A and B are orthogonal $n \times n$ matrices, then the matrices $A^\top$, A^{-1}, and AB are also orthogonal.

Solution: Since A is orthgonal, $A^{-1} = A^\top$, so it's enough to show that $A^\top$ is orthogonal. We check: $(A^\top)^\top A^\top = A\ A^\top = I_n$ because a matrix always commutes with its inverse. For AB, we check: $(AB)^\top (AB) = (B^\top A^\top)(AB)$, and now using $A^{-1} = A^\top$, this collapses to $B^\top B$, which is I_n, since B is orthogonal.

Theorem B: *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation with standard matrix A. Then T is orthogonal if and only if A is orthogonal.*

Problem 6. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation.

(a) Prove Theorem B just above. [Hint: Use Problems 2 and 4.]

(b) If $\mathcal{B}$ is an arbitrary basis, is it still true that T is orthogonal if and if $[T]_\mathcal{B}$ is orthogonal? What about if $\mathcal{B}$ is an *orthonormal basis*?

Solution: For (a), we need to show two things:

1). If T is orthogonal, then A is orthogonal. For this is suffices to show A has orthonormal columns. This was shown in 2c.

2). If A has orthonormal columns, T is orthogonal. By the Theorem on page 1, it suffices to show that for any $\vec{x} \in \mathbb{R}^n$, $||T(\vec{x})|| = ||\vec{x}||$. Since the length is always a non-negative number, it suffices to show $||T(\vec{x})||^2 = ||\vec{x}||^2$. That is, it suffices to show $T(\vec{x}) \cdot T(\vec{x}) = \vec{x} \cdot \vec{x}$. For, this we take an arbitrary $\vec{x}$ and write it in the basis $\{\vec{e}_1, \ldots, \vec{e}_n\}$. Note that

$$\vec{x} \cdot \vec{x} = (x_1\vec{e}_1 + \cdots + x_n\vec{e}_n) \cdot (x_1\vec{e}_1 + \cdots + x_n\vec{e}_n) = \sum_{ij}(x_i\vec{e}_i) \cdot (x_j\vec{e}_j) = x_1^2 + \cdots + x_n^2.$$

Here the third equality is using some basic properties of dot product (like "foil"), and the third equality is using the fact that the $\vec{e}_i$ are orthonormal so that $(x_i\vec{e}_i) \cdot (x_j\vec{e}_j) = 0$ if $i \neq j$. On the other hand, we also have

$$\begin{aligned} T(\vec{x}) \cdot T(\vec{x}) =& T(x_1\vec{e}_1 + \cdots + x_n\vec{e}_n) \cdot T(x_1\vec{e}_1 + \cdots + x_n\vec{e}_n) \\ =& \sum_{ij}(x_iT(\vec{e}_i)) \cdot (x_jT(\vec{e}_j)) \\ =& \sum_{ij} x_ix_j\, T(\vec{e}_i) \cdot T(\vec{e}_j) = x_1^2 + \cdots + x_n^2 \end{aligned}$$

with the last equality coming from the fact that the $T(\vec{e}_i)$'s are the columns of A and hence orthonormal. QED.

For (b), the same proof works for any orthonormal basis—we really only used the fact that $\vec{e}_i \cdot \vec{e}_j$ is zero or one (if $i = j$), so that so is $T(\vec{e}_i) \cdot T(\vec{e}_j)$. Alternatively, we can deduce it from Theorem B as follows: Suppose $T : \mathbb{R}^n \to \mathbb{R}^n$ is an orthogonal transformation, so the standard matrix $[T]_{\mathcal{E}}$ of T is orthogonal by Theorem B. Let $\mathcal{U} = (\vec{u}_1, \ldots, \vec{u}_n)$ be an orthonormal basis of $\mathbb{R}^n$, and write $Q = [\vec{u}_1 \ \cdots \ \vec{u}_n]$, so that Q is an orthogonal matrix and thus so is the matrix $Q^{-1} = Q^\top$ by Problem 1. Then

$$[T]_{\mathcal{U}} \;=\; S_{\mathcal{E}\to\mathcal{U}}[T]_{\mathcal{E}}S_{\mathcal{U}\to\mathcal{E}} \;=\; Q^\top[T]_{\mathcal{E}}Q$$

is also orthogonal, since products of orthogonal matrices are orthogonal (again by Problem 5).

For an arbitrary basis $\mathcal{B}$ is false! For example, let T be rotation by $90°$ in $\mathbb{R}^2$. let $\mathcal{B} = (\vec{e}_1, \vec{e}_1 + \vec{e}_2)$. The $\mathcal{B}$-matrix of T is $\begin{bmatrix} 1 & -2 \\ -1 & 1 \end{bmatrix}$, which is clearly not orthogonal as its columns do not have dot product zero. For an arbitrary basis, this fails.

Problem 7. Prove Theorem A from page 1. [Hint: For the harder direction, consider $T(\vec{x} + \vec{y})$.]

Solution: First assume that T is orthogonal. Take arbitrary $\vec{x} \in \mathbb{R}^n$. We want to show $||\vec{x}|| = ||T(\vec{x})||$. By definition $||\vec{x}|| = (\vec{x} \cdot \vec{x})^{1/2}$ and $||T(\vec{x})|| = (T(\vec{x}) \cdot T(\vec{x}))^{1/2}$. Since T is orthogonal, we know that $\vec{x} \cdot \vec{x} = T(\vec{x}) \cdot T(\vec{x})$, so the result follows.

For the converse, assume T preserves lengths. Take arbitrary $\vec{x}, \vec{y} \in \mathbb{R}^n$. We need to show that $T(\vec{x}) \cdot T(\vec{y}) = \vec{x} \cdot \vec{y}$. By assumption $||T(\vec{x} + \vec{y})|| = ||(\vec{x} + \vec{y})||$. This is the same as $T(\vec{x}+\vec{y}) \cdot T(\vec{x}+\vec{y}) = (\vec{x}+\vec{y}) \cdot (\vec{x}+\vec{y})$. By linearity, this is the same as $(T(\vec{x})+T(\vec{y})) \cdot (T(\vec{x})+T(\vec{y})) = (\vec{x} + \vec{y}) \cdot (\vec{x} + \vec{y})$. Expanding out using the rules of dot product, we get

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

Now, we know that $T(\vec{x}) \cdot T(\vec{x}) = ||T(\vec{x})||^2 = ||\vec{x}||^2 = (\vec{x} \cdot \vec{x})$ and similarly for $\vec{y}$, so this cancels down to

$$2T(\vec{x}) \cdot T(\vec{y}) = 2\vec{x} \cdot \vec{y}$$

and so $T(\vec{x}) \cdot T(\vec{y}) = \vec{x} \cdot \vec{y}$, as needed.

Problem 8. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation, and let A be the standard matrix of T. Which of following are equivalent?

(a) T *preserves length*, i.e., $\|T(v)\| = \|v\|$ for all $v \in \mathbb{R}^n$.

(b) T *preserves distance*, i.e., $\|T(v) - T(w)\| = \|v - w\|$ for all $v, w \in \mathbb{R}^n$.

(c) T is an orthogonal transformation, i.e., T preserves the dot product.

(d) T maps any orthonormal basis of $\mathbb{R}^n$ to an orthonormal basis of $\mathbb{R}^n$.

(e) T maps the standard basis of $\mathbb{R}^n$ to an orthonormal basis of $\mathbb{R}^n$.

(f) The columns of A form an orthonormal basis of $\mathbb{R}^n$.

(g) $A^\top A = I_n$.

(h) $AA^\top = I_n$.

(i) A is an orthogonal matrix.

(j) The rows of A form an orthonormal basis of $\mathbb{R}^n$.

Solution: These are all equivalent, and we've bassically shown them above. Here are some direct proofs again.
$(a) \Leftrightarrow (b)$: If T is linear and preserves lengths, then for all $v, w \in \mathbb{R}^n$, $\|T(v) - T(w)\| = \|T(v - w)\| = \|v - w\|$. Conversely, if T is linear and preserves distances, then for all $v \in \mathbb{R}^n$, $\|T(v)\| = \|T(v) - 0\| = \|T(v) - T(0)\| = \|v - 0\| = \|v\|$.
$(a \wedge b) \Rightarrow (c)$: Let $v, w \in \mathbb{R}^n$. Expanding each side of $\|T(v - w)\| = \|v - w\|$ in terms of the dot product and using the facts that $\|T(v)\| = \|v\|$ and $\|T(w)\| = \|w\|$ to simplify, we find that $T(v) \cdot T(w) = v \cdot w$.
$(c) \Rightarrow (d)$: Assuming (c), if $u_i \cdot u_j = \delta_{ij}$ for each i, j, then also $T(u_i) \cdot T(u_j) = \delta_{ij}$ for each i, j.
$(d) \Rightarrow (e)$: Immediate, since the standard basis of $\mathbb{R}^n$ is orthonormal.
$(e) \Rightarrow (f)$: Follows from the fact that $A = [\,T(\vec{e}_1) \ \cdots \ T(\vec{e}_n)\,]$.
$(f) \Leftrightarrow (g)$: If we write $A = [\,\vec{u}_1 \ \cdots \ \vec{u}_n\,]$, the (i, j)-entry of $A^\top A$ is $\vec{u}_i \cdot \vec{u}_j$.
$(g) \Leftrightarrow (h)$: Follows from Theorem 2.4.8 in the text.
$(h) \Leftrightarrow (i)$: Definition of orthogonal.
$(h) \Leftrightarrow (j)$: If we write $A = \begin{bmatrix} — & \vec{u}_1^\top & — \\ & \vdots & \\ — & \vec{u}_n^\top & — \end{bmatrix}$, the (i, j)-entry of $AA^\top$ is $\vec{u}_i \cdot \vec{u}_j$.
$(g) \Rightarrow (c)$: If $A^\top A = I_n$, then $T(v) \cdot T(w) = Av \cdot Aw = v^\top A^\top A w = v^\top w = v \cdot w$.
$(c) \Rightarrow (a)$: If T preserves the dot product, then for all v, $\|T(v)\|^2 = T(v) \cdot T(v) = v \cdot v = \|v\|^2$.

Problem 9. Let $A \in \mathbb{R}^{n\times d}$ and $B \in \mathbb{R}^{d\times p}$. Prove that $(AB)^\top = B^\top A^\top$ using the ideas from Problem 4. [Note: You already proved this in the homework, most likely, a clumsier way!]

Solution: Let $\vec{\alpha}_i^\top$ denote the i-th row of A. So $A = \begin{bmatrix} \vec{\alpha}_1^\top \\ \vdots \\ \vec{\alpha}_n^\top \end{bmatrix} = \begin{bmatrix} \vec{\alpha}_1 & \cdots & \vec{\alpha}_d \end{bmatrix}^\top$. Let $\vec{\beta}_j$ denote the j-th column of B, so that $B = \begin{bmatrix} \vec{\beta}_1 & \cdots & \vec{\beta}_p \end{bmatrix}$. Then

$$AB = \begin{bmatrix} \vec{\alpha}_1^\top \\ \vdots \\ \vec{\alpha}_n^\top \end{bmatrix} \begin{bmatrix} \vec{\beta}_1 & \cdots & \vec{\beta}_p \end{bmatrix} = [\vec{\alpha}_i^\top \vec{\beta}_j] = [\vec{\alpha}_i \cdot \vec{\beta}_j],$$

so the ij-entry of $(AB)^\top$ is $\vec{\alpha}_j \cdot \vec{\beta}_i$. And

$$B^\top A^\top = \begin{bmatrix} \vec{\beta}_1^\top \\ \vdots \\ \vec{\beta}_p^\top \end{bmatrix} \begin{bmatrix} \vec{\alpha}_1 & \cdots & \vec{\alpha}_d \end{bmatrix} = [\vec{\beta}_i^\top \vec{\alpha}_j] = [\vec{\beta}_i \cdot \vec{\alpha}_j],$$

when is the same, since dot product is symmetric.