Math 2270-004 Week 13 notes
We will not necessarily finish the material from a given day's notes on that day. We may also add or subtract some material as the week progresses, but these notes represent an in-depth outline of what we plan to cover. These notes cover material in 6.4-6.8

Mon Apr 9
- 6.4   Gram Schmidt and $A = QR$ decomposition. Orthogonal matrices

Announcements:

(Recall)

Warm-up Exercise: Can you still use Gram-Schmidt to convert $\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 4 \end{bmatrix} \right\}$ into an orthonormal basis for $\mathbb{R}^2$?

hint: $\vec{u}_1 = \dfrac{\vec{v}_1}{\|\vec{v}_1\|}$ ✓

$\vec{z}_2 = \vec{v}_2 - (\vec{v}_2 \cdot \vec{u}_1)\vec{u}_1$

$\vec{u}_2 = \dfrac{\vec{z}_2}{\|\vec{z}_2\|}$

(page 3)

$$\left\{ \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}, \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \right\}$$

(if it was up to you, reorder $\left\{ \begin{bmatrix} 0 \\ 4 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$
$\downarrow$
$\left\{ \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$)

$$\vec{z}_2 = \begin{bmatrix} 0 \\ 4 \end{bmatrix} - \left(\begin{bmatrix} 0 \\ 4 \end{bmatrix} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right) \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 0 \\ 4 \end{bmatrix} - \frac{1}{2}(4)\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 4 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \underline{\begin{bmatrix} -2 \\ 2 \end{bmatrix}} = 2\begin{bmatrix} -1 \\ 1 \end{bmatrix},$$

$$\vec{u}_2 = \frac{\vec{z}_2}{\|\vec{z}_2\|} = \frac{1}{\sqrt{8}}\begin{bmatrix} -2 \\ 2 \end{bmatrix} = \frac{1}{2\sqrt{2}}\begin{bmatrix} -2 \\ 2 \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

better way is to normalize $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$: $\frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \end{bmatrix}$

(there is only one unit vector in direction of any $\vec{v}$, normalize any appropriate positive scalar mult of $\vec{v}$ to get it. This will help on HW)

We begin on Monday with a continuation of the discussion of Gram-Schmidt orthogonalization from 6.4. Keeping track of the G.S. process carefully yields the $A = QR$ matrix product decomposition theorem, where Q is an "orthogonal matrix" consisting of an orthonormal basis for the span of the columns of A and R is an upper triangular matrix with positive entries along the diagonal. This decomposition is one way to understand why matrix determinants correspond to $\pm$ Volumes, in $\mathbb{R}^n$, among other uses.

T Section 6.5, *Least square solutions* is about finding approximate solutions to inconsistent matrix equations, and relies on many of the ideas we've been studying in Chapter 6 up to this point.

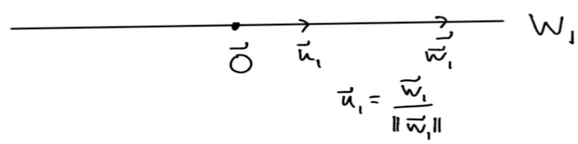
W Section 6.6, *Applications to linear models,* is an application of the least squares method to e.g. linear regression in statistics.

F Finally, sections 6.7 and 6.8 generalize our orthogonality discussions that began with the dot product, to *inner products* in other vector spaces such as function spaces. These ideas lie at the heart of physics applications that use Fourier series, and more recent applications such as image and audio compression.

Recall the Gram-Schmidt process from Friday:

Start with a basis $B = \{\underline{\mathbf{w}}_1, \underline{\mathbf{w}}_2, \ldots \underline{\mathbf{w}}_p\}$ for a subspace W of $\mathbb{R}^n$. How can you convert it into an orthonormal basis? Here's how! The inductive process is called Gram-Schmidt orthogonalization.

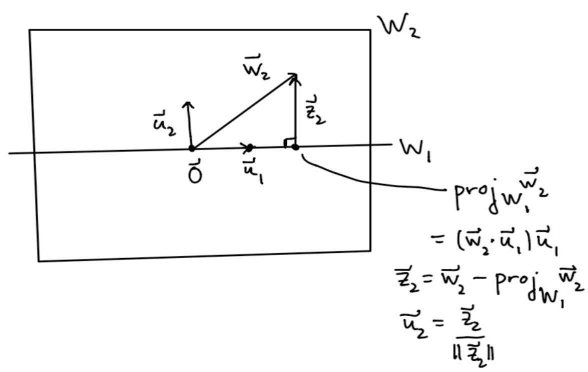
- Let $W_1 = span\{\underline{\mathbf{w}}_1\}$. Define $\underline{\mathbf{u}}_1 = \dfrac{\underline{\mathbf{w}}_1}{\|\underline{\mathbf{w}}_1\|}$. Then $\{\underline{\mathbf{u}}_1\}$ is an orthonormal basis for W_1.



Let $W_2 = span\{\underline{\mathbf{w}}_1, \underline{\mathbf{w}}_2\} = span\{\underline{\mathbf{u}}_1, \underline{\mathbf{w}}_2\}$.

Let $\underline{\mathbf{z}}_2 = \underline{\mathbf{w}}_2 - proj_{W_1}\underline{\mathbf{w}}_2 = \underline{\mathbf{w}}_2 - (\underline{\mathbf{w}}_2 \cdot \underline{\mathbf{u}}_1)\underline{\mathbf{u}}_1$ so $\underline{\mathbf{z}}_2 \perp \underline{\mathbf{u}}_1$.

Define $\underline{\mathbf{u}}_2 = \dfrac{\underline{\mathbf{z}}_2}{\|\underline{\mathbf{z}}_2\|}$. So $\{\underline{\mathbf{u}}_1, \underline{\mathbf{u}}_2\}$ is an orthonormal basis for W_2.



Inductively,

Let $W_j = span\{\underline{\mathbf{w}}_1, \underline{\mathbf{w}}_2, \ldots \underline{\mathbf{w}}_j\} = span\{\underline{\mathbf{u}}_1, \underline{\mathbf{u}}_2, \ldots \underline{\mathbf{u}}_{j-1}, \underline{\mathbf{w}}_j\}$.

Let $\underline{\mathbf{z}}_j = \underline{\mathbf{w}}_j - proj_{W_{j-1}}\underline{\mathbf{w}}_j = \underline{\mathbf{w}}_j - (\underline{\mathbf{w}}_j \cdot \underline{\mathbf{u}}_1)\underline{\mathbf{u}}_1 - (\underline{\mathbf{w}}_j \cdot \underline{\mathbf{u}}_2)\underline{\mathbf{u}}_2 - \ldots (\underline{\mathbf{w}}_j \cdot \underline{\mathbf{u}}_{j-1})\underline{\mathbf{u}}_{j-1}$.

...so $\underline{\mathbf{z}}_j \perp span\{\underline{\mathbf{u}}_1, \underline{\mathbf{u}}_2, \ldots \underline{\mathbf{u}}_{j-1}\}$.

Define $\underline{\mathbf{u}}_j = \dfrac{\underline{\mathbf{z}}_j}{\|\underline{\mathbf{z}}_j\|}$. Then $\{\underline{\mathbf{u}}_1, \underline{\mathbf{u}}_2, \ldots \underline{\mathbf{u}}_j\}$ is an orthonormal basis for W_j.

Continue up to $j = p$.

We're denoting the original basis for W by $B = \{\underline{\boldsymbol{w}}_1, \underline{\boldsymbol{w}}_2, \ldots \underline{\boldsymbol{w}}_p\}$. Denote the orthonormal basis we've constructed with Gram-Schmidt by $O = \{\underline{\boldsymbol{u}}_1, \underline{\boldsymbol{u}}_2, \ldots \underline{\boldsymbol{u}}_p\}$. Because O is orthonormal it's easy to express these two bases in terms of each other. Notice

$$W_j = span\ \{\underline{\boldsymbol{w}}_1, \underline{\boldsymbol{w}}_2, \ldots \underline{\boldsymbol{w}}_j\} = \underbrace{span\ \{\underline{\boldsymbol{u}}_1, \underline{\boldsymbol{u}}_2, \ldots \underline{\boldsymbol{u}}_j\}}_{\text{orthormal set}} \qquad \text{for each } 1 \le j \le p.$$

$\vec{x} \in W_j$

$\vec{x} = (\vec{x}\cdot\vec{u}_1)\vec{u}_1 + (\vec{x}\cdot\vec{u}_2)\vec{u}_2 + \cdots (\vec{x}\cdot\vec{u}_j)u_j$

So,

1 $$\underline{\boldsymbol{w}}_1 = (\underline{\boldsymbol{w}}_1 \cdot \underline{\boldsymbol{u}}_1)\underline{\boldsymbol{u}}_1$$

2 $$\underline{\boldsymbol{w}}_2 = (\underline{\boldsymbol{w}}_2 \cdot \underline{\boldsymbol{u}}_1)\underline{\boldsymbol{u}}_1 + (\underline{\boldsymbol{w}}_2 \cdot \underline{\boldsymbol{u}}_2)\underline{\boldsymbol{u}}_2$$

$$\vdots$$

$$\underline{\boldsymbol{w}}_j = (\underline{\boldsymbol{w}}_j \cdot \underline{\boldsymbol{u}}_1)\underline{\boldsymbol{u}}_1 + (\underline{\boldsymbol{w}}_j \cdot \underline{\boldsymbol{u}}_2)\underline{\boldsymbol{u}}_2 + \ldots\ (\underline{\boldsymbol{w}}_j \cdot \underline{\boldsymbol{u}}_j)\underline{\boldsymbol{u}}_j$$

$$\vdots$$

$$\underline{\boldsymbol{w}}_p = \sum_{l=1}^{p} (\underline{\boldsymbol{w}}_l \cdot \underline{\boldsymbol{u}}_l)\underline{\boldsymbol{u}}_l\ .$$

Notice that the coefficients of the last terms in the sums above, namely $(\underline{\boldsymbol{w}}_j \cdot \underline{\boldsymbol{u}}_j)$ can be computed as

keep in mind. $$(\underline{\boldsymbol{w}}_j \cdot \underline{\boldsymbol{u}}_j) = \underline{\boldsymbol{z}}_j \cdot \frac{\underline{\boldsymbol{z}}_j}{\|\underline{\boldsymbol{z}}_j\|} = \|\underline{\boldsymbol{z}}_j\|$$

$$\vec{u}_j \cdot \left[\, \vec{z}_j = \vec{w}_j - \text{proj}_{W_{j-1}} \vec{w}_j \,\right]$$

In matrix form (column by column) we have

$$* \quad \underbrace{\begin{array}{ccc} 1 & 2 & p \\ \end{array}\begin{bmatrix} \vec{w}_1 & \vec{w}_2 & \cdots & \vec{w}_p \end{bmatrix}}_{\text{"A"}} = \underbrace{\begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_p \end{bmatrix}}_{\text{"Q"}} \underbrace{\begin{array}{c} \begin{array}{ccccc} 1 & 2 & & & p \end{array} \\ \begin{bmatrix} \vec{w}_1\cdot\vec{u}_1 & \vec{w}_2\cdot\vec{u}_1 & \vec{w}_3\cdot\vec{u}_1 & \cdots & \vec{w}_p\cdot\vec{u}_1 \\ 0 & \vec{w}_2\cdot\vec{u}_2 & \vec{w}_3\cdot\vec{u}_2 & & \vec{w}_p\cdot\vec{u}_2 \\ 0 & 0 & \vec{w}_3\cdot\vec{u}_3 & \ddots & \vdots \\ \vdots & \vdots & 0 & & \\ 0 & 0 & 0 & & \vec{w}_p\cdot\vec{u}_p \end{bmatrix} \end{array}}_{\text{"R"}}$$

"A" columns are original basis for W

$A_{n\times p}$

"Q" columns are orthonormal

$Q_{n\times p}$

"R" upper Δ'ular, with positive. diagonal entries

$R_{p\times p}$

$\vec{w}_j\cdot\vec{u}_j = \|\vec{z}_j\|$

R also known as $P_{\cancel{B\leftarrow O}}$ $O \leftarrow B$ ☺

Thus any matrix with linearly independent columns may be written in factored form as above, ($W = Col\, A$),

$$A_{n\times p} = Q_{n\times p} R_{p\times p}.$$

This factorization contains geometric information and can simplify the computational work needed to solve matrix equations $A\,\underline{\boldsymbol{x}} = \underline{\boldsymbol{b}}$.

From previous page...

$$* \qquad A_{n \times p} = Q_{n \times p} R_{p \times p}$$

<u>shortcut</u> (or what to do if you forgot the formulas for the entries of R) If you just know Q you can recover R by multiplying both sides of the $*$ equation on the previous page by the transpose Q^T of the Q matrix:

$$Q^T \qquad A \quad = \quad Q^T \qquad Q \qquad R$$

$$\begin{bmatrix} - - \vec{u}_1^{\,T} - - \\ - - \vec{u}_2^{\,T} - - \\ \vdots \\ - - \vec{u}_p^{\,T} - - \end{bmatrix} \begin{bmatrix} \vdots & \vdots & & \vdots \\ \vec{w}_1 & \vec{w}_2 & \cdots & \vec{w}_p \\ \vdots & \vdots & & \vdots \end{bmatrix} = \underbrace{\begin{bmatrix} - - \vec{u}_1^{\,T} - - \\ - - \vec{u}_2^{\,T} - - \\ \vdots \\ \vec{u}_p^{\,T} \end{bmatrix} \begin{bmatrix} \vdots & \vdots & & \vdots \\ \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_p \\ \vdots & \vdots & & \vdots \end{bmatrix}} R = IR = R!$$

$$\begin{bmatrix} u_i^T \end{bmatrix} \begin{bmatrix} \vec{u}_j \end{bmatrix} = \vec{u}_i \cdot \vec{u}_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

$$Q^T Q = \begin{bmatrix} 1 & 0 & \\ 0 & 1 & \\ 0 & 0 & \\ \vdots & 0 & \\ 0 & \vdots & \end{bmatrix}$$

$$A = Q\,R$$
$$Q^T A = Q^T Q\, R = I\, R = R.$$

<u>Example)</u> From last Friday,

$$\vec{w}_1 \qquad \vec{w}_2 \qquad\qquad\qquad \vec{u}_1 \qquad\qquad \vec{u}_2$$

$$B = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 4 \end{bmatrix} \right\}, \quad O = \left\{ \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \right\}.$$

$$\vec{w}_1 \cdot \vec{u}_1 = \frac{2}{\sqrt{2}} = \sqrt{2}$$
$$\vec{w}_2 \cdot \vec{u}_1 = \frac{4}{\sqrt{2}} = 2\sqrt{2}$$
$$\vec{w}_2 \cdot \vec{u}_2 = \frac{4}{\sqrt{2}} = 2\sqrt{2}$$

$$B \quad = \quad Q \qquad\qquad R$$

$$\begin{bmatrix} 1 & 0 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \boldsymbol{w}_1 \cdot \boldsymbol{u}_1 & \boldsymbol{w}_2 \cdot \boldsymbol{u}_1 \\ 0 & \boldsymbol{w}_2 \cdot \boldsymbol{u}_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{2} & 2\sqrt{2} \\ 0 & 2\sqrt{2} \end{bmatrix} = Q\,R.$$

<u>Exercise 1</u>) Verify that R could have been recovered via the formula

$$Q^T A = R$$

"A" = B

$$A = QR$$
$$Q^T A = Q^T Q R = IR.$$

$$\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} \sqrt{2} & 2\sqrt{2} \\ 0 & 2\sqrt{2} \end{bmatrix} \checkmark$$

From previous page ...

$$\begin{bmatrix} 1 & 0 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \underbrace{\begin{bmatrix} \sqrt{2} & 2\sqrt{2} \\ 0 & 2\sqrt{2} \end{bmatrix}}_{R}.$$

<u>Exercise 2)</u> Verify that the $A = Q\,R$ factorization in this example may be further factored as

$$\begin{bmatrix} 1 & 0 \\ 1 & 4 \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}}_{Rot_{\pi/4}} \underbrace{\begin{bmatrix} \sqrt{2} & 0 \\ 0 & 2\sqrt{2} \end{bmatrix}}_{\text{diag. stretch matrix}} \underbrace{\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}}_{\text{area preserving shear}}.$$

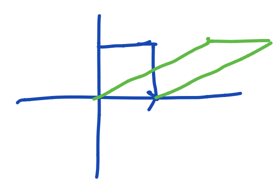
diag. stretch matrix
(x_1- dir by $\sqrt{2}$
x_2- dir by $2\sqrt{2}$)

area preserving shear

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

$$2 = \frac{2\sqrt{2}}{\sqrt{2}}$$

$$T(\vec{x}) = A\vec{x}$$



- So, the transformation $T(\underline{\boldsymbol{x}}) = A\,\underline{\boldsymbol{x}}$ is a composition of (1) an area-preserving shear, followed by (2) a diagonal scaling that increases area by a factor of $\sqrt{2}\cdot 2\sqrt{2} = 4$, followed by a rotation of $\frac{\pi}{4}$, which does not effect area. Since determinants of products matrices are the products of determinants (we checked this back when we studied determinants), and area expansion factors of compositions are also the products of the area expansion factors, the generalization of this example explains why the determinant of A (or its absolute value in general) coincides with the area expansion factor, in the 2×2 case. You show in your homework that the only possible Q matrices in the 2×2 case are rotations as above, or reflections across lines through the origin. In the latter case, the determinant of Q is -1, and the determinant of A is negative.

$$Q_{2\times 2} \quad \underbrace{\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}}_{Rot_\theta,\ \det = +1} \quad \text{or} \quad \underbrace{\begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}}_{\det = -1} \quad \text{HW}$$

($Refl_L$ L is at angle $\frac{\theta}{2}$)

<u>Example</u> from last Friday.

$$B = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 4 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} \right\} \qquad O = \left\{ \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}, \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

<u>Exercise 3a</u> Find the $A = Q\,R$ factorization based on the data above, for

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

$$\text{solution } A = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{2} & 2\sqrt{2} & -\frac{1}{\sqrt{2}} \\ 0 & 2\sqrt{2} & -\frac{3}{\sqrt{2}} \\ 0 & 0 & 3 \end{bmatrix}$$

<u>Exercise 3b</u> Further factor R into a diagonal matrix times a volume-preserving shear and interpret the transformation $T(\underline{\mathbf{x}}) = A\,\underline{\mathbf{x}}$ as a composition of (1) a volume preserving shear, followed by (2) a coordinate scaling that increases volume by a factor of 12, followed by a rotation about the x_3 axis in $\mathbb{R}^3$, which preserves volume. The generalization of this example explains why the determinant of A (or its absolute value in general) is the volume expansion factor for the transformation $T(\underline{\mathbf{x}}) = A\,\underline{\mathbf{x}}$.

Definition A square $n \times n$ matrix Q is called *orthogonal* if its columns are ortho-normal. (You can read more about orthogonal matrices at e.g. Wikipedia.)

Theorem. Let Q be an orthogonal matrix. Then

a) $Q^{-1} = Q^T$.

b) The rows of Q are also ortho-normal.

c) the transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ given by

$$T(\underline{\boldsymbol{x}}) = Q\,\underline{\boldsymbol{x}}$$

preserves dot products and magnitudes, (so also volumes, since cubes generated by perpendicular vectors will be transformed into equal-volume cubes). In other words, for all $\underline{\boldsymbol{x}}, \underline{\boldsymbol{y}} \in \mathbb{R}^n$,

$$T(\underline{\boldsymbol{x}}) \cdot T(\underline{\boldsymbol{y}}) = \underline{\boldsymbol{x}} \cdot \underline{\boldsymbol{y}}$$

$$\|T(\underline{\boldsymbol{x}})\| = \|\underline{\boldsymbol{x}}\|.$$

d) The only matrix transformations $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ that preserve dot products are orthogonal transformations. (These transformations are often referred to as *isometries*.)