

Overview

Last time we introduced the notion of an orthonormal basis for a subspace. We also saw that if a square matrix U has orthonormal columns, then U is invertible and $U^{-1} = U^T$. Such a matrix is called an *orthogonal* matrix.

At the beginning of the course we developed a formula for computing the projection of one vector onto another in \mathbb{R}^2 or \mathbb{R}^3 . Today we'll generalise this notion to higher dimensions.

From Lay, §6.3

Review

Recall from Stewart that if $\mathbf{u} \neq \mathbf{0}$ and \mathbf{y} are vectors in \mathbb{R}^n , then

$\text{proj}_{\mathbf{u}}\mathbf{y} = \frac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}$ is the orthogonal projection of \mathbf{y} onto \mathbf{u} .

(Lay uses the notation “ $\hat{\mathbf{y}}$ ” for this projection, where \mathbf{u} is understood.)

Review

Recall from Stewart that if $\mathbf{u} \neq \mathbf{0}$ and \mathbf{y} are vectors in \mathbb{R}^n , then

$\text{proj}_{\mathbf{u}}\mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u}$ is the orthogonal projection of \mathbf{y} onto \mathbf{u} .

(Lay uses the notation “ $\hat{\mathbf{y}}$ ” for this projection, where \mathbf{u} is understood.)

How would you describe the vector $\text{proj}_{\mathbf{u}}\mathbf{y}$ in words?

Review

Recall from Stewart that if $\mathbf{u} \neq \mathbf{0}$ and \mathbf{y} are vectors in \mathbb{R}^n , then

$\text{proj}_{\mathbf{u}}\mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u}$ is the orthogonal projection of \mathbf{y} onto \mathbf{u} .

(Lay uses the notation “ $\hat{\mathbf{y}}$ ” for this projection, where \mathbf{u} is understood.)

How would you describe the vector $\text{proj}_{\mathbf{u}}\mathbf{y}$ in words?

One possible answer:

\mathbf{y} can be written as the sum of a vector parallel to \mathbf{u} and a vector orthogonal to \mathbf{u} ; $\text{proj}_{\mathbf{u}}\mathbf{y}$ is the summand parallel to \mathbf{u} .

Or alternatively,

\mathbf{y} can be written as the sum of a vector in the line spanned by \mathbf{u} and a vector orthogonal to \mathbf{u} ; $\text{proj}_{\mathbf{u}}\mathbf{y}$ is the summand in $\text{Span}\{\mathbf{u}\}$.

Review

Recall from Stewart that if $\mathbf{u} \neq \mathbf{0}$ and \mathbf{y} are vectors in \mathbb{R}^n , then

$\text{proj}_{\mathbf{u}}\mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u}$ is the orthogonal projection of \mathbf{y} onto \mathbf{u} .

(Lay uses the notation “ $\hat{\mathbf{y}}$ ” for this projection, where \mathbf{u} is understood.)

How would you describe the vector $\text{proj}_{\mathbf{u}}\mathbf{y}$ in words?

One possible answer:

\mathbf{y} can be written as the sum of a vector parallel to \mathbf{u} and a vector orthogonal to \mathbf{u} ; $\text{proj}_{\mathbf{u}}\mathbf{y}$ is the summand parallel to \mathbf{u} .

Or alternatively,

\mathbf{y} can be written as the sum of a vector in the line spanned by \mathbf{u} and a vector orthogonal to \mathbf{u} ; $\text{proj}_{\mathbf{u}}\mathbf{y}$ is the summand in $\text{Span}\{\mathbf{u}\}$.

We'd like to generalise this, replacing $\text{Span}\{\mathbf{u}\}$ by an arbitrary subspace:

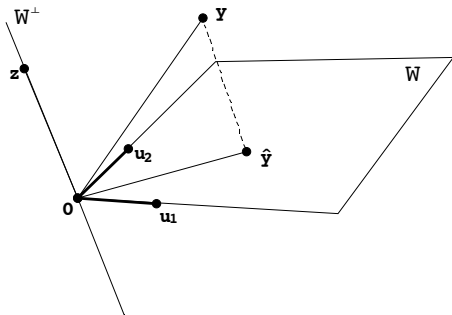
Given \mathbf{y} and a subspace W in \mathbb{R}^n , we'd like to write \mathbf{y} as a sum of a vector in W and a vector in W^\perp .

Example 1

Suppose that $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal basis for \mathbb{R}^3 and let $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$. Write \mathbf{y} as the sum of a vector $\hat{\mathbf{y}}$ in W and a vector \mathbf{z} in W^\perp .

Example 1

Suppose that $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal basis for \mathbb{R}^3 and let $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$. Write \mathbf{y} as the sum of a vector $\hat{\mathbf{y}}$ in W and a vector \mathbf{z} in W^\perp .



Recall that for any orthogonal basis, we have

$$\mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2} \mathbf{u}_2 + \frac{\mathbf{y} \cdot \mathbf{u}_3}{\mathbf{u}_3 \cdot \mathbf{u}_3} \mathbf{u}_3.$$

Recall that for any orthogonal basis, we have

$$\mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2} \mathbf{u}_2 + \frac{\mathbf{y} \cdot \mathbf{u}_3}{\mathbf{u}_3 \cdot \mathbf{u}_3} \mathbf{u}_3.$$

It follows that

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2} \mathbf{u}_2$$

and

$$\mathbf{z} = \frac{\mathbf{y} \cdot \mathbf{u}_3}{\mathbf{u}_3 \cdot \mathbf{u}_3} \mathbf{u}_3.$$

Since \mathbf{u}_3 is orthogonal to \mathbf{u}_1 and \mathbf{u}_2 , its scalar multiples are orthogonal to $\text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$. Therefore $\mathbf{z} \in W^\perp$

All this can be generalised to any vector \mathbf{y} and subspace W of \mathbb{R}^n , as we will see next.

The Orthogonal Decomposition Theorem

Theorem

Let W be a subspace in \mathbb{R}^n . Then each $\mathbf{y} \in \mathbb{R}^n$ can be written uniquely in the form

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z} \quad (1)$$

where $\hat{\mathbf{y}} \in W$ and $\mathbf{z} \in W^\perp$.

If $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is any orthogonal basis of W , then

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \dots + \frac{\mathbf{y} \cdot \mathbf{u}_p}{\mathbf{u}_p \cdot \mathbf{u}_p} \mathbf{u}_p \quad (2)$$

The vector $\hat{\mathbf{y}}$ is called the **orthogonal projection of \mathbf{y} onto W** .

Note that it follows from this theorem that to calculate the decomposition $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$, it is enough to know one orthogonal basis for W explicitly. Any orthogonal basis will do, and all orthogonal bases will give the same decomposition $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$.

Example 2

Given

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \mathbf{u}_3 = \begin{bmatrix} 0 \\ -1 \\ 1 \\ -1 \end{bmatrix}$$

let W be the subspace of \mathbb{R}^4 spanned by $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$.

Write $\mathbf{y} = \begin{bmatrix} 2 \\ -3 \\ 4 \\ 1 \end{bmatrix}$ as the sum of a vector in W and a vector orthogonal to W .

The orthogonal projection of \mathbf{y} onto W is given by

$$\begin{aligned}\hat{\mathbf{y}} &= \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2} \mathbf{u}_2 + \frac{\mathbf{y} \cdot \mathbf{u}_3}{\mathbf{u}_3 \cdot \mathbf{u}_3} \mathbf{u}_3 \\ &= \frac{-2}{3} \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \end{bmatrix} + \frac{7}{3} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} + \frac{6}{3} \begin{bmatrix} 0 \\ -1 \\ 1 \\ -1 \end{bmatrix} \\ &= \frac{1}{3} \begin{bmatrix} 5 \\ -8 \\ 13 \\ 3 \end{bmatrix}\end{aligned}$$

The orthogonal projection of \mathbf{y} onto W is given by

$$\begin{aligned}\hat{\mathbf{y}} &= \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2} \mathbf{u}_2 + \frac{\mathbf{y} \cdot \mathbf{u}_3}{\mathbf{u}_3 \cdot \mathbf{u}_3} \mathbf{u}_3 \\ &= \frac{-2}{3} \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \end{bmatrix} + \frac{7}{3} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} + \frac{6}{3} \begin{bmatrix} 0 \\ -1 \\ 1 \\ -1 \end{bmatrix} \\ &= \frac{1}{3} \begin{bmatrix} 5 \\ -8 \\ 13 \\ 3 \end{bmatrix}\end{aligned}$$

Also

$$\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} 2 \\ -3 \\ 4 \\ 1 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} 5 \\ -8 \\ 13 \\ 3 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 \\ -1 \\ -1 \\ 0 \end{bmatrix}$$

Thus the desired decomposition of \mathbf{y} is

$$\begin{aligned}\mathbf{y} &= \hat{\mathbf{y}} + \mathbf{z} \\ \begin{bmatrix} 2 \\ -3 \\ 4 \\ 1 \end{bmatrix} &= \frac{1}{3} \begin{bmatrix} 5 \\ -8 \\ 13 \\ 3 \end{bmatrix} + \frac{1}{3} \begin{bmatrix} 1 \\ -1 \\ -1 \\ 0 \end{bmatrix}.\end{aligned}$$

The Orthogonal Decomposition Theorem ensures that $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$ is in W^\perp . However, verifying this is a good check against computational mistakes.

Thus the desired decomposition of \mathbf{y} is

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$$
$$\begin{bmatrix} 2 \\ -3 \\ 4 \\ 1 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 5 \\ -8 \\ 13 \\ 3 \end{bmatrix} + \frac{1}{3} \begin{bmatrix} 1 \\ -1 \\ -1 \\ 0 \end{bmatrix}.$$

The Orthogonal Decomposition Theorem ensures that $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$ is in W^\perp . However, verifying this is a good check against computational mistakes.

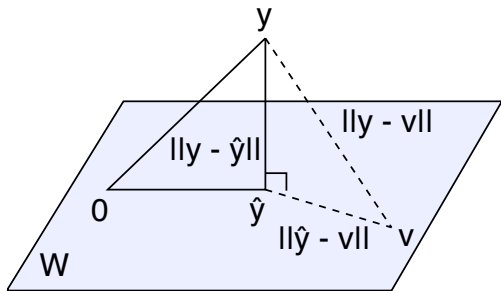
This problem was made easier by the fact that $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal basis for W . If you were given an arbitrary basis for W instead of an orthogonal basis, what would you do?

Theorem (The Best Approximation Theorem)

Let W be a subspace of \mathbb{R}^n , \mathbf{y} any vector in \mathbb{R}^n , and $\hat{\mathbf{y}}$ the orthogonal projection of \mathbf{y} onto W . Then $\hat{\mathbf{y}}$ is the closest vector in W to \mathbf{y} , in the sense that

$$\|\mathbf{y} - \hat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{v}\| \quad (3)$$

for all \mathbf{v} in W , $\mathbf{v} \neq \hat{\mathbf{y}}$.



Proof

Let \mathbf{v} be any vector in W , $\mathbf{v} \neq \hat{\mathbf{y}}$. Then $\hat{\mathbf{y}} - \mathbf{v} \in W$. By the Orthogonal Decomposition Theorem, $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to W . In particular $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to $\hat{\mathbf{y}} - \mathbf{v}$. Since

$$\mathbf{y} - \mathbf{v} = (\mathbf{y} - \hat{\mathbf{y}}) + (\hat{\mathbf{y}} - \mathbf{v})$$

the Pythagorean Theorem gives

$$\|\mathbf{y} - \mathbf{v}\|^2 = \|\mathbf{y} - \hat{\mathbf{y}}\|^2 + \|\hat{\mathbf{y}} - \mathbf{v}\|^2.$$

Hence $\|\mathbf{y} - \mathbf{v}\|^2 > \|\mathbf{y} - \hat{\mathbf{y}}\|^2$. □

We can now define the distance from a vector \mathbf{y} to a subspace W of \mathbb{R}^n .

Definition

Let W be a subspace of \mathbb{R}^n and let \mathbf{y} be a vector in \mathbb{R}^n . The *distance* from \mathbf{y} to W is

$$\|\mathbf{y} - \hat{\mathbf{y}}\|$$

where $\hat{\mathbf{y}}$ is the orthogonal projection of \mathbf{y} onto W .

Example 3

Consider the vectors

$$\mathbf{y} = \begin{bmatrix} 3 \\ -1 \\ 1 \\ 13 \end{bmatrix}, \mathbf{u}_1 = \begin{bmatrix} 1 \\ -2 \\ -1 \\ 2 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} -4 \\ 1 \\ 0 \\ 3 \end{bmatrix}.$$

Find the closest vector to \mathbf{y} in $W = \text{Span} \{ \mathbf{u}_1, \mathbf{u}_2 \}$.

$$\begin{aligned} \hat{\mathbf{y}} &= \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2} \mathbf{u}_2 \\ &= \frac{30}{10} \begin{bmatrix} 1 \\ -2 \\ -1 \\ 2 \end{bmatrix} + \frac{26}{26} \begin{bmatrix} -4 \\ 1 \\ 0 \\ 3 \end{bmatrix} = \begin{bmatrix} -1 \\ -5 \\ -3 \\ 9 \end{bmatrix}. \end{aligned}$$

Example 3

Consider the vectors

$$\mathbf{y} = \begin{bmatrix} 3 \\ -1 \\ 1 \\ 13 \end{bmatrix}, \mathbf{u}_1 = \begin{bmatrix} 1 \\ -2 \\ -1 \\ 2 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} -4 \\ 1 \\ 0 \\ 3 \end{bmatrix}.$$

Find the closest vector to \mathbf{y} in $W = \text{Span} \{ \mathbf{u}_1, \mathbf{u}_2 \}$.

$$\begin{aligned} \hat{\mathbf{y}} &= \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2} \mathbf{u}_2 \\ &= \frac{30}{10} \begin{bmatrix} 1 \\ -2 \\ -1 \\ 2 \end{bmatrix} + \frac{26}{26} \begin{bmatrix} -4 \\ 1 \\ 0 \\ 3 \end{bmatrix} = \begin{bmatrix} -1 \\ -5 \\ -3 \\ 9 \end{bmatrix}. \end{aligned}$$

Therefore the distance from \mathbf{y} to W is $\| \begin{bmatrix} 3 \\ -1 \\ 1 \\ 13 \end{bmatrix} - \begin{bmatrix} -1 \\ -5 \\ -3 \\ 9 \end{bmatrix} \| = \| \begin{bmatrix} 4 \\ 4 \\ 4 \\ 4 \end{bmatrix} \| = 8$

Theorem

If $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p\}$ is an orthonormal basis for a subspace W of \mathbb{R}^n , then for all \mathbf{y} in \mathbb{R}^n we have

$$\text{proj}_W \mathbf{y} = (\mathbf{y} \cdot \mathbf{u}_1) \mathbf{u}_1 + (\mathbf{y} \cdot \mathbf{u}_2) \mathbf{u}_2 + \cdots + (\mathbf{y} \cdot \mathbf{u}_p) \mathbf{u}_p.$$

This theorem is an easy consequence of the usual projection formula:

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \cdots + \frac{\mathbf{y} \cdot \mathbf{u}_p}{\mathbf{u}_p \cdot \mathbf{u}_p} \mathbf{u}_p.$$

When each \mathbf{u}_j is a unit vector, the denominators are all equal to 1.

Theorem

If $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p\}$ is an orthonormal basis for a subspace W of \mathbb{R}^n , then for all \mathbf{y} in \mathbb{R}^n we have

$$\text{proj}_W \mathbf{y} = (\mathbf{y} \cdot \mathbf{u}_1) \mathbf{u}_1 + (\mathbf{y} \cdot \mathbf{u}_2) \mathbf{u}_2 + \cdots + (\mathbf{y} \cdot \mathbf{u}_p) \mathbf{u}_p.$$

This theorem is an easy consequence of the usual projection formula:

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \cdots + \frac{\mathbf{y} \cdot \mathbf{u}_p}{\mathbf{u}_p \cdot \mathbf{u}_p} \mathbf{u}_p.$$

When each \mathbf{u}_i is a unit vector, the denominators are all equal to 1.

Theorem

If $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p\}$ is an orthonormal basis for W and $U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_p \end{bmatrix}$, then for all \mathbf{y} in \mathbb{R}^n we have

$$\text{proj}_W \mathbf{y} = UU^T \mathbf{y}. \quad (4)$$

The proof is a matrix calculation; see the posted slides for details.

Note that if U is a $n \times p$ matrix with orthonormal columns, then we have $U^T U = I_p$ (see Lay, Theorem 6 in Chapter 6). Thus we have

$$U^T U \mathbf{x} = I_p \mathbf{x} = \mathbf{x} \quad \text{for every } \mathbf{x} \text{ in } \mathbb{R}^p$$

$$U U^T \mathbf{y} = \text{proj}_W \mathbf{y} \quad \text{for every } \mathbf{y} \text{ in } \mathbb{R}^n, \text{ where } W = \text{Col } U.$$

Note: Pay attention to the sizes of the matrices involved here. Since U is $n \times p$ we have that U^T is $p \times n$. Thus $U^T U$ is a $p \times p$ matrix, while $U U^T$ is an $n \times n$ matrix.

The previous theorem shows that the function which sends \mathbf{x} to its orthogonal projection onto W is a linear transformation. The kernel of this transformation is ...

The previous theorem shows that the function which sends \mathbf{x} to its orthogonal projection onto W is a linear transformation. The kernel of this transformation is ...

...the set of all vectors orthogonal to W , i.e., W^\perp .

The range is W itself.

The previous theorem shows that the function which sends \mathbf{x} to its orthogonal projection onto W is a linear transformation. The kernel of this transformation is ...

...the set of all vectors orthogonal to W , i.e., W^\perp .

The range is W itself.

The theorem also gives us a convenient way to find the closest vector to \mathbf{x} in W : find an orthonormal basis for W and let U be the matrix whose columns are these basis vectors. Then multiply \mathbf{x} by UU^T .

Examples

Example 4

Let $W = \text{Span} \left\{ \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix} \right\}$ and let $\mathbf{x} = \begin{bmatrix} 4 \\ 8 \\ 1 \end{bmatrix}$. What is the closest vector to \mathbf{x} in W ?

$$\text{Set } \mathbf{u}_1 = \begin{bmatrix} 2/3 \\ 1/3 \\ 2/3 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} -2/3 \\ 2/3 \\ 1/3 \end{bmatrix},$$

$$U = \begin{bmatrix} 2/3 & -2/3 \\ 1/3 & 2/3 \\ 2/3 & 1/3 \end{bmatrix}.$$

We check that $U^T U = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, so U has orthonormal columns.

The closest vector is

$$\text{proj}_W \mathbf{x} = UU^T \mathbf{x} = \frac{1}{9} \begin{bmatrix} 8 & -2 & 2 \\ -2 & 5 & 4 \\ 2 & 4 & 5 \end{bmatrix} \begin{bmatrix} 4 \\ 8 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix}.$$

We check that $U^T U = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, so U has orthonormal columns.

The closest vector is

$$\text{proj}_W \mathbf{x} = U U^T \mathbf{x} = \frac{1}{9} \begin{bmatrix} 8 & -2 & 2 \\ -2 & 5 & 4 \\ 2 & 4 & 5 \end{bmatrix} \begin{bmatrix} 4 \\ 8 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix}.$$

We can also compute distance from \mathbf{x} to W :

$$\|\mathbf{x} - \text{proj}_W \mathbf{x}\| = \left\| \begin{bmatrix} 4 \\ 8 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix} \right\| = \left\| \begin{bmatrix} 2 \\ 4 \\ -4 \end{bmatrix} \right\| = 6.$$

Because this example is about vectors in \mathbb{R}^3 , so we could also use cross products:

$$\begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} \times \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 2 & 1 & 2 \\ -2 & 2 & 1 \end{vmatrix} = -3\mathbf{i} - 6\mathbf{j} + 6\mathbf{k} = \mathbf{n}$$

gives a vector orthogonal to W , so the distance is the length of the projection of \mathbf{x} onto \mathbf{n} :

$$\begin{bmatrix} 4 \\ 8 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} -1/3 \\ -2/3 \\ 2/3 \end{bmatrix} = -6,$$

and the closest vector is

$$\begin{bmatrix} 4 \\ 8 \\ 1 \end{bmatrix} + 6 \begin{bmatrix} -1/3 \\ -2/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix}.$$

This example showed that the standard matrix for projection to

$$W = \text{Span} \left\{ \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix} \right\} \text{ is } \frac{1}{9} \begin{bmatrix} 8 & -2 & 2 \\ -2 & 5 & 4 \\ 2 & 4 & 5 \end{bmatrix}.$$

If we instead work with $\mathcal{B} = \left\{ \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ -2 \\ 2 \end{bmatrix} \right\}$ coordinates, what is the orthogonal projection matrix?

This example showed that the standard matrix for projection to

$$W = \text{Span} \left\{ \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix} \right\} \text{ is } \frac{1}{9} \begin{bmatrix} 8 & -2 & 2 \\ -2 & 5 & 4 \\ 2 & 4 & 5 \end{bmatrix}.$$

If we instead work with $\mathcal{B} = \left\{ \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ -2 \\ 2 \end{bmatrix} \right\}$ coordinates, what is the orthogonal projection matrix?

Observe that the three basis vectors were chosen very carefully: \mathbf{b}_1 and \mathbf{b}_2 span W , and \mathbf{b}_3 is orthogonal to W . Thus each of the basis vectors is an eigenvector for the linear transformation. (Why?)

The linear transformation is represented by a diagonal matrix when it's

written in terms of an eigenbasis. Thus we get the matrix $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

This example showed that the standard matrix for projection to

$$W = \text{Span} \left\{ \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix} \right\} \text{ is } \frac{1}{9} \begin{bmatrix} 8 & -2 & 2 \\ -2 & 5 & 4 \\ 2 & 4 & 5 \end{bmatrix}.$$

If we instead work with $\mathcal{B} = \left\{ \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ -2 \\ 2 \end{bmatrix} \right\}$ coordinates, what is the orthogonal projection matrix?

Observe that the three basis vectors were chosen very carefully: \mathbf{b}_1 and \mathbf{b}_2 span W , and \mathbf{b}_3 is orthogonal to W . Thus each of the basis vectors is an eigenvector for the linear transformation. (Why?)

The linear transformation is represented by a diagonal matrix when it's

written in terms of an eigenbasis. Thus we get the matrix $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

What does this tell you about orthogonal projection matrices in general?

Example 5

$\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}$ are orthogonal and span a subspace W of \mathbb{R}^4 . Find a vector orthogonal to W .

Normalize the columns and set

$$U = \begin{bmatrix} 1/\sqrt{2} & 1/2 \\ 0 & 1/2 \\ 1/\sqrt{2} & -1/2 \\ 0 & -1/2 \end{bmatrix}.$$

Then the standard matrix for the orthogonal projection is has matrix

$$UU^T = \frac{1}{4} \begin{bmatrix} 3 & 1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 3 & 1 \\ -1 & -1 & 1 & 1 \end{bmatrix}.$$

Thus, choosing a vector $\mathbf{v} = \begin{bmatrix} 3 \\ 2 \\ 0 \\ 1 \end{bmatrix}$ not in W , the closest vector to \mathbf{v} in W is given by

$$UU^T \begin{bmatrix} 3 \\ 2 \\ 0 \\ 1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 5 \\ 2 \\ 1 \\ -2 \end{bmatrix}.$$

In particular, $\mathbf{v} - UU^T\mathbf{v} = \begin{bmatrix} 3 \\ 2 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 5 \\ 2 \\ 1 \\ -2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 \\ 2 \\ -1 \\ 4 \end{bmatrix}$ lies in W^\perp .

Thus $\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 2 \\ -1 \\ 4 \end{bmatrix}$ are orthogonal in \mathbb{R}^4 , and span a subspace W_1 of dimension 3.

But now we can repeat the process with W_1 ! This time take

$$U = \begin{bmatrix} 1/\sqrt{2} & 1/2 & 1/\sqrt{22} \\ 0 & 1/2 & 2/\sqrt{22} \\ 1/\sqrt{2} & -1/2 & -1/\sqrt{22} \\ 0 & -1/2 & 4/\sqrt{22} \end{bmatrix},$$

$$UU^T = \frac{1}{44} \begin{bmatrix} 35 & 15 & 9 & -3 \\ 15 & 19 & -15 & 5 \\ 9 & -15 & 35 & 3 \\ -3 & 5 & 3 & 43 \end{bmatrix}.$$

Taking $\mathbf{x} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$, $(I_4 - UU^T)\mathbf{x} = 1/44 \begin{bmatrix} 3 \\ -5 \\ -3 \\ 1 \end{bmatrix}$ and then

$\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 2 \\ -1 \\ 4 \end{bmatrix}$, $\begin{bmatrix} 3 \\ -5 \\ -3 \\ 1 \end{bmatrix}$ is an orthogonal basis for \mathbb{R}^4 . □