

Chapter 4: Orthogonality

Val Anthony Balagon

January 2019

Abstract

This chapter focuses on the orthogonality of the four subspaces, projections, and least squares approximations.

1 Orthogonality of the Four Subspaces

Two vectors are orthogonal when their dot product is zero $\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^T \mathbf{w} = 0$. This chapter will revolve around orthogonal subspaces, orthogonal bases, and orthogonal matrices.

Definition 1.1. Orthogonal vectors have the following properties:

- i. $\mathbf{v}^T \mathbf{w} = 0$
- ii. $\|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 = \|\mathbf{v} + \mathbf{w}\|^2 \rightarrow \mathbf{v}^T \mathbf{v} + \mathbf{w}^T \mathbf{w} = (\mathbf{v} + \mathbf{w})^T (\mathbf{v} + \mathbf{w})$

Remark. *The zero vector is orthogonal to any vector.*

Remark. *The subspaces have orthogonal properties.*

1. **The row space $C(A^T)$ is perpendicular to the nullspace $N(A)$.** Every row of A is perpendicular to the solution of $A\mathbf{x} = \mathbf{0}$.
2. **The column space $C(A)$ is perpendicular to the left nullspaces $N(A^T)$.** When \mathbf{b} is outside of the column space when we're trying to solve for $A\mathbf{x} = \mathbf{b}$, then this nullspace of A^T comes into its own. It contains the error $\mathbf{e} = \mathbf{b} - A\mathbf{x}$ in the least-squares solution.

Definition 1.2. Two subspaces \mathbf{V} and \mathbf{W} of a vector space are orthogonal if every vector \mathbf{v} in \mathbf{V} is perpendicular to every vector \mathbf{w} in \mathbf{W} .

$$\mathbf{v}^T \mathbf{w} = 0 \text{ for all } \mathbf{v} \text{ in } \mathbf{V} \text{ and all } \mathbf{w} \text{ in } \mathbf{W}.$$

Theorem 1. Every vector \mathbf{x} in the nullspace is perpendicular to every row of A , because $A\mathbf{x} = \mathbf{0}$. The nullspace $N(A)$ and the row space $C(A^T)$ are orthogonal subspaces of \mathbb{R}^n .

$$A\mathbf{x} = \begin{bmatrix} \text{row 1} \\ \vdots \\ \text{row } m \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$C_1(\text{row}_1^T) = 0$$

$$C_2(\text{row}_2^T) = 0$$

$$\vdots$$

$$C_m(\text{row}_m^T) = 0$$

(row 1) $\cdot \mathbf{x}$ is zero and (row m) $\cdot \mathbf{x}$ is also zero. Every row has a zero dot product with \mathbf{x} . Then \mathbf{x} is perpendicular to every combination of the rows. **The whole row space $C(A^T)$ is orthogonal to $N(A)$.**

Proof. The vectors in the row space are combinations of $A^T \mathbf{y}$ of the rows. We take the dot product of $A^T \mathbf{y}$ with any \mathbf{x} in the nullspace.

$$\mathbf{x} \cdot (A^T \mathbf{y}) = \mathbf{x}^T (A^T \mathbf{y}) = (A\mathbf{x})^T \mathbf{y} = \mathbf{0}^T \mathbf{y} = 0$$

□

Example 1. The rows of A are perpendicular to $\mathbf{x} = (1, 1, -1)$ in the nullspace:

$$A\mathbf{x} = \begin{bmatrix} 1 & 3 & 4 \\ 5 & 2 & 7 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 + 3 - 4 \\ 5 + 2 - 7 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

In this example, the column space is all of \mathbb{R}^2 . The nullspace of A^T is the zero vector. The column space of A and the nullspace of A^T are always orthogonal subspaces.

Theorem 2. Every vector \mathbf{y} in the nullspace of A^T is perpendicular to every column of A . The left nullspace $N(A^T)$ and the column space $C(A)$ are orthogonal in \mathbb{R}^m .

Proof. The nullspace of A^T is orthogonal to the row space of A^T , which is the column space of A .

$$A^T \mathbf{y} = \begin{bmatrix} (\text{column 1})^T \\ \vdots \\ (\text{column } n)^T \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

□

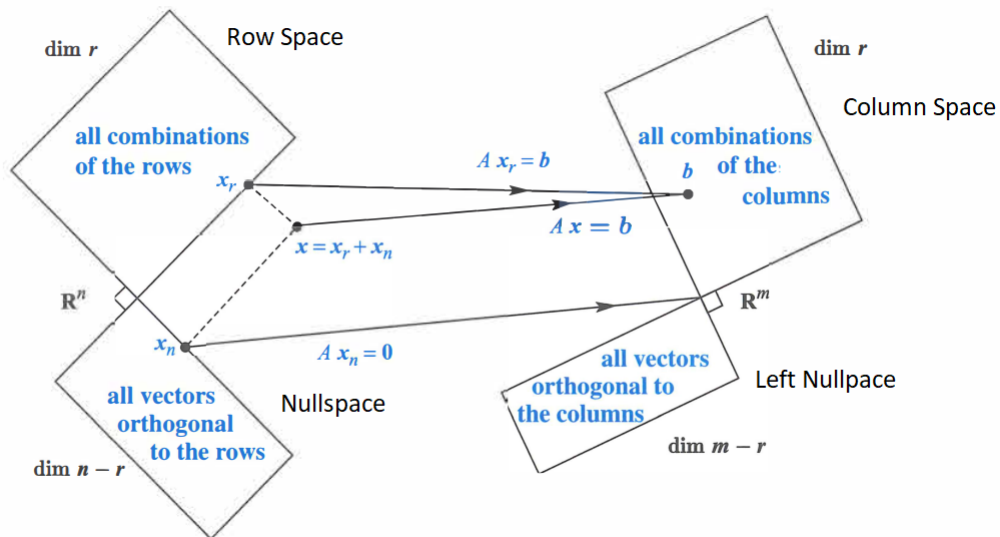


Figure 1: The Four Subspaces. There are two pairs of orthogonal subspaces.

Theorem 3. If a vector \mathbf{v} is orthogonal to itself, then \mathbf{v} is the zero vector.

Theorem 4. Fundamental Theorem of Linear Algebra, Part 2:

$N(A)$ is the orthogonal complement of the row space $C(A^T)$ in \mathbb{R}^n .

$N(A^T)$ is the orthogonal complement of the column space $C(A)$ in \mathbb{R}^m .

Things to note from Figure 1

1. When A multiplies to $\mathbf{x} = \mathbf{x}_r + \mathbf{x}_n$, it goes to \mathbf{b} which is in the column space.
2. When A multiplies to \mathbf{x}_r , it goes to \mathbf{b} which is also in the column space.
3. When A multiplies to \mathbf{x}_n , the nullspace component goes to $\mathbf{0}$.

1.1 Combining Bases from Subspaces

Theorem 5. Any independent vectors in \mathbb{R}^n must span \mathbb{R}^n . So they are a basis.

Any n vectors that span \mathbb{R}^n must be independent. So they are a basis

Theorem 6. If the n columns of A are independent, they span \mathbb{R}^n . So $A\mathbf{x} = \mathbf{b}$ is solvable. If the n columns span \mathbb{R}^n , they are independent. So $A\mathbf{x} = \mathbf{b}$ has only one solution.

2 Projections

Let's say we are given an arbitrary vector \mathbf{b} . When \mathbf{b} is projected onto a line, its projection \mathbf{p} is the part of \mathbf{b} along that line. If \mathbf{b} is projected onto a plane, \mathbf{p} is a part in that plane. The projection \mathbf{p} is $P\mathbf{b}$. The projection matrix P multiplies \mathbf{b} to give \mathbf{p} .

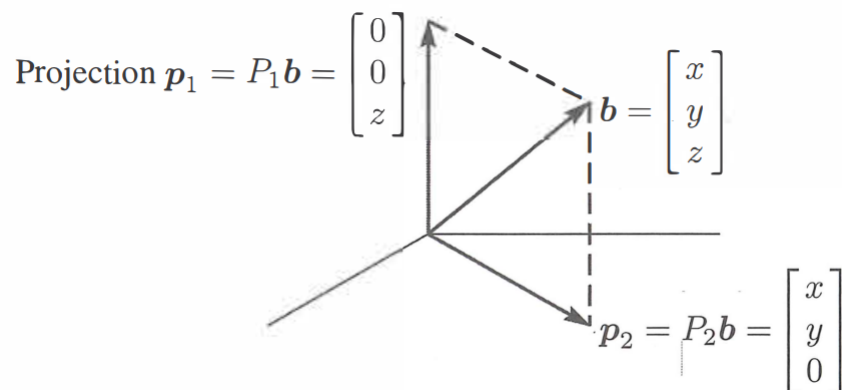


Figure 2: The projections \mathbf{p}_1 and \mathbf{p}_2 onto the z axis and the xy plane.

Say $\mathbf{b} = [2 \ 3 \ 4]^T$, its projections are $\mathbf{p}_1 = [0 \ 0 \ 4]^T$ and $\mathbf{p}_2 = [2 \ 3 \ 0]^T$ in Figure 2. \mathbf{p}_1 and \mathbf{p}_2 are the projections of \mathbf{b} onto the z axis and xy axis, respectively. The projection matrices P_1 and P_2 are 3×3 matrices. Projection onto a line comes from a rank one matrix, while a projection onto a planes comes from a rank two matrix.

$$\begin{aligned} \text{Onto the } z \text{ axis: } P_1 &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} & \text{Onto the } xy \text{ plane: } P_2 &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ \mathbf{p}_1 = P_1 \mathbf{b} &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 4 \end{bmatrix} & \mathbf{p}_2 = P_2 \mathbf{b} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix} \end{aligned}$$

The projections \mathbf{p}_1 and \mathbf{p}_2 are perpendicular. The xy plane and the z axis are orthogonal subspaces. The line and plane are also orthogonal complements. Their dimensions add to $1 + 2 = 3$. Every vector \mathbf{b} in the whole space is the sum of its parts in the two subspaces. The projections \mathbf{p}_1 and \mathbf{p}_2 are exactly those two parts of \mathbf{b} :

$$\text{The vectors give } \mathbf{p}_1 + \mathbf{p}_2 = \mathbf{b} \quad \text{The matrices give } P_1 + P_2 = I$$

In general, the objective is to find \mathbf{p} in each subspace, and the projection matrix P that produces that part $\mathbf{p} = P\mathbf{b}$. Every subspace of \mathbb{R}^m has its own $m \times m$ projection matrix. To compute P , we need a good description of the subspace that it projects onto. The best description of a subspace is a basis. We put the basis vectors into the columns of A . Now we are projecting onto the column space of A . The problem now is to project any \mathbf{b} onto the column space of any $m \times n$ matrix.

2.1 2-D Case

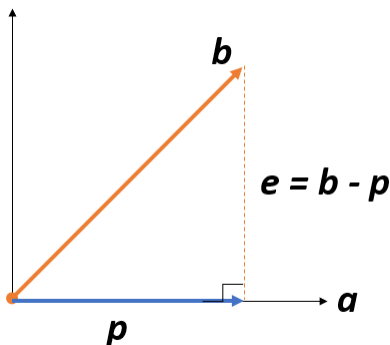


Figure 3: The projection of \mathbf{b} onto \mathbf{a}

A line goes through the origin in the direction \mathbf{a} . Along that line, we want the point \mathbf{p} closest to \mathbf{b} . The key to projection is orthogonality: The line from \mathbf{b} to \mathbf{p} is perpendicular to the vector \mathbf{a} . This is the dotted line marked $\mathbf{e} = \mathbf{b} - \mathbf{p}$ for the error on the left side of Figure 3. We now compute \mathbf{p} by algebra.

The projection \mathbf{p} will be a multiple of \mathbf{a} . We call it $\mathbf{p} = \hat{\mathbf{x}}\mathbf{a}$. Computing this number $\hat{\mathbf{x}}$ will give the vector \mathbf{p} . Then from the formula \mathbf{p} , we will read off the projection matrix P . These three steps will lead to all projection matrices: **find $\hat{\mathbf{x}}$, then find the vector \mathbf{p} , then find the matrix P .**

Projecting \mathbf{b} onto \mathbf{a} with error $\mathbf{e} = \mathbf{b} - \hat{\mathbf{x}}\mathbf{a}$:

$$\mathbf{a} \cdot (\mathbf{b} - \hat{\mathbf{x}}\mathbf{a}) = 0 \quad \text{or} \quad \mathbf{a} \cdot \mathbf{b} - \hat{\mathbf{x}}\mathbf{a} \cdot \mathbf{a} = 0$$

$$\hat{\mathbf{x}} = \frac{\mathbf{a} \cdot \mathbf{b}}{\mathbf{a} \cdot \mathbf{a}} = \boxed{\frac{\mathbf{a}^T \mathbf{b}}{\mathbf{a}^T \mathbf{a}}}$$

Theorem 7. The projection of \mathbf{b} onto the line through \mathbf{a} is the vector $\mathbf{p} = \hat{\mathbf{x}}\mathbf{a} = \frac{\mathbf{a}^T \mathbf{b}}{\mathbf{a}^T \mathbf{a}} \mathbf{a}$

Special Case 1: If $\mathbf{b} = \mathbf{a}$, then $\hat{\mathbf{x}} = 1$. The projection of \mathbf{a} onto \mathbf{a} is itself. $P\mathbf{a} = \mathbf{a}$.

Special Case 2: If \mathbf{b} is perpendicular to \mathbf{a} , then $\mathbf{a}^T \mathbf{b} = 0$. The projection is $\mathbf{p} = 0$.

Example 2. Project $\mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ onto $\mathbf{a} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ to find $\mathbf{p} = \hat{\mathbf{x}}\mathbf{a}$.

$$\mathbf{p} = \hat{\mathbf{x}}\mathbf{a} = \frac{\mathbf{a}^T \mathbf{b}}{\mathbf{a}^T \mathbf{a}} \mathbf{a} = \frac{\begin{bmatrix} 1 & 2 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}{\begin{bmatrix} 1 & 2 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}} \mathbf{a} = \boxed{\frac{5}{9} \mathbf{a}}$$

The error vector between \mathbf{b} and \mathbf{p} is $\mathbf{e} = \mathbf{b} - \mathbf{p}$. Those vectors \mathbf{p} and \mathbf{e} will add to $\mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

$$\mathbf{p} = \left(\frac{5}{9}, \frac{10}{9}, \frac{10}{9} \right) \quad \text{and} \quad \mathbf{e} = \mathbf{b} - \mathbf{p} = \left(\frac{4}{9}, -\frac{1}{9}, -\frac{1}{9} \right)$$

To get the magnitude of \mathbf{p} :

$$\|\mathbf{p}\| = \frac{\|\mathbf{a}\| \|\mathbf{b}\| \cos \theta}{\|\mathbf{a}\|^2} \|\mathbf{a}\| = \|\mathbf{b}\| \cos \theta$$

Now to find the projection matrix,

$$\mathbf{p} = \mathbf{a} \hat{x} = \mathbf{a} \frac{\mathbf{a}^T \mathbf{b}}{\mathbf{a}^T \mathbf{a}} = P \mathbf{b} \text{ then } P = \frac{\mathbf{a} \mathbf{a}^T}{\mathbf{a}^T \mathbf{a}}$$

$$P = \begin{bmatrix} \frac{1}{9} & \frac{2}{9} & \frac{2}{9} \\ \frac{2}{9} & \frac{4}{9} & \frac{4}{9} \\ \frac{2}{9} & \frac{4}{9} & \frac{4}{9} \end{bmatrix}$$

P is a column times a row! Then column is \mathbf{a} , the row is \mathbf{a}^T . Then divide by the number $\mathbf{a}^T \mathbf{a}$. The projection matrix P is an $m \times m$ matrix, but its rank is 1. We are projecting onto a 1-D subspace, a line through \mathbf{a} . *That line is the column space of P .* Next, we try to project a second time.

$$P^2 = \begin{bmatrix} \frac{1}{9} & \frac{2}{9} & \frac{2}{9} \\ \frac{2}{9} & \frac{4}{9} & \frac{4}{9} \\ \frac{2}{9} & \frac{4}{9} & \frac{4}{9} \end{bmatrix} \begin{bmatrix} \frac{1}{9} & \frac{2}{9} & \frac{2}{9} \\ \frac{2}{9} & \frac{4}{9} & \frac{4}{9} \\ \frac{2}{9} & \frac{4}{9} & \frac{4}{9} \end{bmatrix} = \begin{bmatrix} \frac{1}{9} & \frac{2}{9} & \frac{2}{9} \\ \frac{2}{9} & \frac{4}{9} & \frac{4}{9} \\ \frac{2}{9} & \frac{4}{9} & \frac{4}{9} \end{bmatrix} = P$$

$P^2 = P$

Notice that projecting a second time does not change anything. Notice also that P is symmetric.

To summarize:

1. $C(P)$ is a line through \mathbf{a}
2. $\text{rank}(P) = 1$
3. $P^T = P$, P is symmetric
4. $P^2 = P$, projecting a second time does not change anything

2.2 Projection Onto A Subspace

We now move to n vectors a_1, \dots, a_n in \mathbb{R}^m . We assume that these \mathbf{a} 's are linearly independent. The problem now becomes: how do we find the combination $\mathbf{p} = \hat{x}_1 a_1 + \dots + \hat{x}_n a_n$ closest to a given vector \mathbf{b} ? We are projecting \mathbf{b} in \mathbb{R}^m onto the subspace spanned by the \mathbf{a} 's.

We compute projections onto n -dimensional subspaces in three steps as before: **Find a vector $\hat{\mathbf{x}}$, find the projection $\mathbf{p} = A\hat{\mathbf{x}}$, find the projection matrix P .**

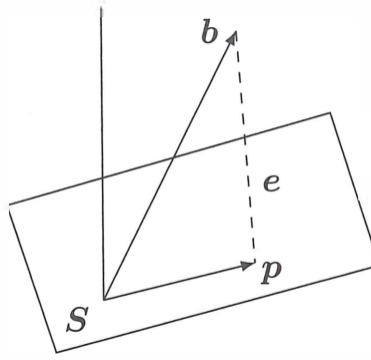


Figure 4: The projection \mathbf{p} of \mathbf{b} onto $\mathbf{S} = \text{column space of } A$

The error vector $\mathbf{b} - A\hat{\mathbf{x}}$ is perpendicular to the subspace. In n -dimensions, there will be n equations for $\hat{\mathbf{x}}$. The error makes a right angle with all the vectors a_1, \dots, a_n in the base. The n right angles give the n equations for $\hat{\mathbf{x}}$.

$$\begin{bmatrix} - & a_1^T & - \\ & \vdots & \\ - & a_n^T & - \end{bmatrix} \quad (1)$$

The combination $\mathbf{p} = \hat{x}_1 \mathbf{a}_1 + \dots + \hat{x}_n \mathbf{a}_n = A\hat{\mathbf{x}}$ that is closest to \mathbf{b} comes from $\hat{\mathbf{x}}$:

$$\boxed{A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}} \quad \text{or} \quad \boxed{A^T A \hat{\mathbf{x}} = A^T \mathbf{b}} \quad (2)$$

This symmetric $A^T A$ is $n \times n$. It is invertible if \mathbf{a} 's are independent. The solution $\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}$. The projection of \mathbf{b} onto the subspace is \mathbf{p} :

$$\boxed{\mathbf{p} = A\hat{\mathbf{x}} = A(A^T A)^{-1} A^T \mathbf{b}} \quad (3)$$

To find the projection matrix P :

$$\boxed{P = A(A^T A)^{-1} A^T} \quad (4)$$

The equations above are equivalent to their $n = 1$ counterparts. We use the inverse $(A^T A)^{-1}$ instead of $\frac{1}{a^T a}$. Matrix inversion is a guarantee because the columns of A are linearly independent.

Here are interesting properties:

1. Our subspace is the column space of A
2. The error vector $\mathbf{b} - A\hat{\mathbf{x}}$ is perpendicular to that column space
3. Therefore $\mathbf{b} - A\hat{\mathbf{x}}$ is in the nullspace of A^T . This means that $A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$

Example 3. If $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$ find $\hat{\mathbf{x}}$ and \mathbf{p} .

Computing for $A^T A$ and $A^T \mathbf{b}$

$$A^T A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}$$

$$A^T \mathbf{b} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$$

Computing for $\hat{\mathbf{x}}$

$$\begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix} \rightarrow \hat{\mathbf{x}} = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}$$

Finding \mathbf{p} :

$$\mathbf{p} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 5 \\ -3 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix}$$

Finding the error:

$$\mathbf{e} = \mathbf{b} - \mathbf{p} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

Finding P :

$$(A^T A)^{-1} = \frac{1}{6} \begin{bmatrix} 5 & -3 \\ -3 & 3 \end{bmatrix}$$

$$P = A(A^T A)^{-1}A^T = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} * \frac{1}{6} \begin{bmatrix} 5 & -3 \\ -3 & 3 \end{bmatrix} * \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix}$$

Projecting it twice,

$$P^2 = \frac{1}{6} \begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix} * \frac{1}{6} \begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix}$$

$$P^2 = P$$

If A is a rectangular matrix, then we won't be able to do $(A^T A)^{-1} = A^{-1}(A^T)^{-1}$. This is because A does not have an inverse.

Theorem 8. $A^T A$ is invertible if and only if A has linearly independent columns.

Theorem 9. When A has independent columns, $A^T A$ is square and invertible.

Proof.

Take the transpose of $A^T A$:

$$(A^T A)^T = A^T (A^T)^T = A^T A$$

$$(A^T A)^T = A^T A$$

Hence, $A^T A$ is symmetric. Since A has independent columns, this means that A^T is also invertible. Hence, $A^T A$ is invertible. \square

3 Least Squares Approximations

4 Orthonormal Bases and Gram-Schmidt

5 Problems

Problem 1.1. asd

Solution. soln

\square