
CHAPTER 4: ORTHOGONALITY

Introduction to Linear Algebra

Franklin Chen

Gilbert Strang

Juanita High School

26 February 2025

Contents

1	Orthogonality of the Four Subspaces	3
2	Projections	6
2.1	Projection onto a Line	7
2.2	Projection onto a Subspace	8
2.2.1	Alternate Calculation for a projection	8
2.3	random other things	8
3	Least Squares Approximations	9
4	Orthogonal Bases and Gram-Schmidt	9

1 Orthogonality of the Four Subspaces

Definition 1.1 (Orthogonal Subspaces)

Subspaces V and W of a vector space are orthogonal if every vector in V is perpendicular/orthogonal to every vector in W .

$$\vec{v} \cdot \vec{w} = (\vec{v})^T \times \vec{w} = 0 \quad \forall \vec{v} \in V, \vec{w} \in W$$

Example 1.2 (Floors and Walls)

Suppose you live in a rectangular room. The floor of your room (extended to infinity) is a subspace F . The line where two walls meet is a subspace W (one-dimensional). F and W are orthogonal; every vector along W is perpendicular to every vector along F . The origin $(0, 0, 0)$ is in the corner.

Example 1.3 (Zero)

If a vector is in two mutually orthogonal subspaces, then $\vec{v} \cdot \vec{v} = 0$. The only \vec{v} that has this property is $\vec{0}$. $\vec{0}$ is the only vector that will be in two mutually orthogonal subspaces.

The crucial examples for Linear Algebra come from the fundamental subspaces.

Theorem 1.4

The nullspace of A is orthogonal to the row space of A .

Proof. Suppose \vec{x} is in the nullspace of A . Then, $A\vec{x} = \vec{0}$. By the definitions of matrix-vector multiplication:

$$A\vec{x} = \begin{bmatrix} \text{row 1} \\ \text{row 2} \\ \vdots \\ \text{row } m \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} \text{row 1} \cdot \vec{x} \\ \text{row 2} \cdot \vec{x} \\ \vdots \\ \text{row } m \cdot \vec{x} \end{bmatrix} = \vec{0}$$

Because every cross product between the rows of A and \vec{x} equal zero, every row of A is perpendicular to \vec{x} . This implies that the rows of A - the row space- and any vector in the nullspace of A are mutually perpendicular. Thus, the nullspace of A is orthogonal to the row space of A . \square

Theorem 1.5

The left nullspace of A (the nullspace of the row space) is orthogonal to the column space of A .

Proof. (Very similar proof to Theorem 1.4.) Suppose \vec{y} is in the left nullspace. Then, $A^T \vec{y} = \vec{0}$. Taking the transpose, $(A^T \vec{y})^T = (\vec{y})^T A = (\vec{0})^T$. Expanding this out:

$$\begin{aligned} (\vec{y})^T A &= \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} \begin{bmatrix} \text{col } 1 & \text{col } 2 & \cdots & \text{col } n \end{bmatrix} \\ &= \begin{bmatrix} \text{col } 1 \cdot (\vec{y})^T & \text{col } 2 \cdot (\vec{y})^T & \cdots & \text{col } n \cdot (\vec{y})^T \end{bmatrix} = (\vec{0})^T \end{aligned}$$

Every cross product between the cols of A and $(\vec{y})^T$ equals zero, implying that the cols of A and $(\vec{y})^T$ are mutually perpendicular. Thus, the left nullspace of A is orthogonal to the column space of A . \square

Note that the fundamental subspaces are more than orthogonal: they are **orthogonal compliments** (which matter for some reason?)

Definition 1.6 (Orthogonal Complement)

The orthogonal complement of a subspace V contains **all** vectors orthogonal to V . The orthogonal complement is denoted V^\perp ("v perp").

By definition, the nullspace is the orthogonal complement of the row space, and similar with the left nullspace and the column space. **Keep in mind what this implies:** every \vec{x} orthogonal to the row space automatically satisfies $A\vec{x} = \vec{0}$, and every \vec{y} orthogonal to the null space is automatically part of the row space. $N(A)^\perp = R(A^T)$ and $R(A^T)^\perp = N(A)$! Similar applies for the left nullspace and the column space.

This is all crystalised in the **Second Fundamental Theorem of Linear Algebra**:

Theorem 1.7 (Second Fundamental Theorem of Linear Algebra)

The nullspace is the orthogonal complement of the row space in \mathbb{R}^n . The left nullspace is the orthogonal complement of the col space in \mathbb{R}^m .

From what I understand, this implies that the nullspace plus the row space is equal to \mathbb{R}^n . Thus, any input $\vec{x} \in \mathbb{R}^n$ can be broken down into the sum of a *row space component* \vec{x}_r and a *null space component* \vec{x}_n . So, for any $\vec{b} \in R(A)$:

1. By the definition of the column space, there exists a vector $\vec{x} \in \mathbb{R}^n$ such that $A\vec{x} = \vec{b}$.
2. Shown above, $\vec{x} = \vec{x}_r + \vec{x}_n$.
3. $A\vec{x} = A(\vec{x}_r + \vec{x}_n) = A\vec{x}_r + \vec{0} = A\vec{x}_r = \vec{b}$!

Furthermore, Theorem 1.8 can be shown:

Theorem 1.8 (Row Space and Col Space one-to-one mapping)

For every vector \vec{y} in the column space, there exists *one and only one* vector \vec{x} in the row space such that $A\vec{x} = \vec{y}$.

Proof. Suppose there exists \vec{x}' in the row space such that $A\vec{x} = A\vec{x}'$. The difference $\vec{x} - \vec{x}'$ can be observed to be in the nullspace (maps to 0). By the additive property of subspaces, $\vec{x} - \vec{x}'$ is also part of the row space. As zero is the only vector that can be in two mutually orthogonal subspaces, this difference must equal zero. As such, $\vec{x} = \vec{x}'$. \square

If we "throw away" the nullspaces, we can see that *there is always(?) an invertible matrix "hiding" within any matrix* because of this one-to-one mapping.

Example 1.9 (Hidden Invertible Matrix)

$$A = \begin{bmatrix} 3 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \text{ contains } \begin{bmatrix} 3 & 0 \\ 0 & 5 \end{bmatrix}$$

The hidden invertible matrix will always have size $r \times r$: both the row space and col space are of dimension r .

The following is quite useful, and is provided without proof:

Theorem 1.10 (Linear Independence and Span)

Any n linearly independent vectors in \mathbb{R}^n must span \mathbb{R}^n . Any n vectors that span \mathbb{R}^n must be linearly independent.

2 Projections

Definition 2.1 (Projection)

The **projection** of some vector \vec{b} onto a subspace A is the vector \vec{p} such that $\vec{p} \in \text{span}(A)$ and $|\vec{p} - \vec{b}|$ is minimized.

Example 2.2 (Projections as Components)

Suppose $\vec{b} = (2, 3, 4)$. The projection of \vec{b} onto the z-axis is the "shadow of \vec{b} on the z-axis": $\vec{b}_z = (0, 0, 4)$. Similarly, the projection of \vec{b} onto the xy-plane is the "shadow of \vec{b} on the xy-plane": $\vec{b}_{xy} = (2, 3, 0)$.

Such projections have matrices associated with them. Generally, the projections of a vector onto a line will have a rank one matrix associated, and a projection onto a plane a rank two matrix associated. Projection matrices should generally satisfy $\vec{p} = P\vec{b}$. For this example:

$$P_z = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$P_{xy} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Generally, we will be projecting vectors onto the column space of some arbitrary matrix A . In this case, these matrices are nice because we're projecting onto the core axes and planes. Keep in mind, often, this projection will be much nastier.

$$A_z = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$A_{xy} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

2.1 Projection onto a Line

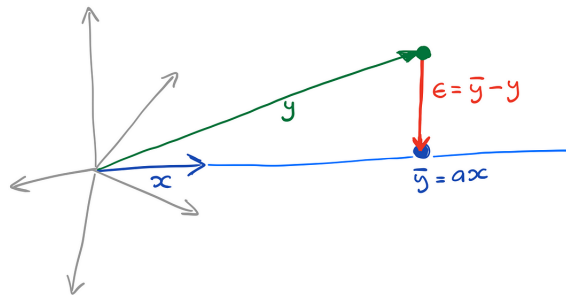


Figure 1: The projection of \vec{y} onto \vec{x} , denoted \bar{y} . Ideally, we want to minimize $\epsilon = \bar{y} - \vec{y}$.

Suppose we are given two vectors \vec{x} and \vec{y} in \mathbb{R}^m , and we want to find the projection \bar{y} on the line \vec{x} such that \vec{y} and its projection are as close as possible. **The key is orthogonality:** the line connecting \vec{y} and \bar{y} (ϵ) is *perpendicular* to \vec{x} .

The projection is some multiple of \vec{x} : $\bar{y} = a\vec{x}$. $\epsilon = \bar{y} - \vec{y} = a\vec{x} - \vec{y}$ is perpendicular to \vec{x} , so:

$$\begin{aligned}\vec{x} \cdot (a\vec{x} - \vec{y}) &= 0 \\ &= a\vec{x} \cdot \vec{x} - \vec{y} \cdot \vec{x} \\ \implies a &= \frac{\vec{x} \cdot \vec{y}}{\vec{x} \cdot \vec{x}} = \frac{(\vec{x})^T \vec{y}}{(\vec{x})^T \vec{x}} \\ \implies \bar{y} &= a\vec{x} = \frac{(\vec{x})^T \vec{y}}{(\vec{x})^T \vec{x}} \vec{x}\end{aligned}$$

This provides an explicit formula for the projection \bar{y} , without dealing with any nasty trig functions. It is provided without proof that this equation for projection lines up with the traditional trigonometric way of finding projections.

The general **projection matrix** is also visible in this equation. If $\bar{y} = a\vec{x} = P\vec{y}$, then, by rearranging the fraction (bring up \vec{x} and put down \vec{y}):

$$P = \frac{\vec{x}(\vec{x})^T}{(\vec{x})^T \vec{x}}$$

P is a column times row, so it is a $m \times m$ matrix of *rank one* (as we are projecting onto a one-dimensional subspace: the line through \vec{x}).

Various remarks about P :

- Because P is defined using the *line* through \vec{x} , only the **direction** of \vec{x} matters: changing the magnitude of \vec{x} changes nothing.
- $P^n = P$: projecting multiple times onto the same line does nothing.
- (?) The sum of the diagonal of P equals 1.

- $I - P$ is a projection onto the line through ϵ : the perpendicular subspace.

2.2 Projection onto a Subspace

Now, instead of projecting a vector \vec{y} onto another vector \vec{x} , we want to project onto a **subspace**. We want to find the "closest" vector to a target vector \vec{b} within the subspace. In other words: **find the combination of basis vectors $\vec{x}_1\vec{a}_1 + \dots + \vec{x}_n\vec{a}_n$ that is closest to the target vector.**

When $n = 1$, this is projection onto a line. However, in general, we are projecting onto an linearly independent matrix A : we are looking for the particular $\vec{p} = A\vec{x}$, such that \vec{p} is closest to \vec{b} . We solve this problem in three steps: finding \vec{x} , finding the projection $\vec{p} = A\vec{x}$, and finding the matrix P .

Similar to the $n = 1$ case, **the key is in orthogonality**: the error vector $\vec{b} - A\vec{x}$ is *perpendicular* to the column space of A . ($\vec{b} - A\vec{x}$ is in the left nullspace of A). Mathematically:

$$\begin{aligned} (\vec{a}_1)^T(\vec{b} - A\vec{x}) &= 0 \\ &\vdots \\ (\vec{a}_n)^T(\vec{b} - A\vec{x}) &= 0 \end{aligned}$$

The coefficients $(\vec{a}_n)^T$ can be extracted out as A^T . Thus, the n equations are exactly $A^T(\vec{b} - A\vec{x}) = \vec{0}$. Or, in another form, $A^T A\vec{x} = A^T \vec{b}$.

This equation allows us to derive \vec{x} as $\vec{x} = (A^T A)^{-1} A^T \vec{b}$. From this, the projection of \vec{b} onto the subspace is $\vec{p} = A\vec{x} = A(A^T A)^{-1} A^T \vec{b}$. It can be observed that $P = A(A^T A)^{-1} A^T$. It is proved without proof that this alligns with the formula derived for $n = 1$.

2.2.1 Alternate Calculation for a projection

From above, $A^T(\vec{b} - A\vec{x}) = \vec{0}$. This implies that $\vec{b} - A\vec{x}$ is in the left nullspace. Denoting the left nullspace matrix A_\perp , $\vec{b} - A\vec{x} = A_\perp \vec{y} \rightarrow \vec{b} = A\vec{x} + A_\perp \vec{y}$. We can combine this into a segmented dot product:

$$\begin{bmatrix} A & A_\perp \end{bmatrix} \begin{bmatrix} \vec{x} \\ \vec{y} \end{bmatrix} = \vec{b}$$

We can then use RREF to find \vec{x} , and calculate the projection as $A\vec{x}$. This method also provides \vec{y} , which captures the part of \vec{b} not captured by the projection: $A_\perp \vec{y}$. (this can also be observed in the formula for \vec{b} as the orthogonal component!)

So, to find the projection \vec{p} from a vector \vec{p} on a subspace A :

1. Use RREF to find the basis of the nullspace of A^T . Use these as the columns for A_\perp .
2. Use RREF to solve the segmented matrix equation for \vec{x} and \vec{y} .
3. Calculate $\vec{p} = A\vec{x}$. If needed, calculate the error as $A_\perp \vec{y}$.

2.3 random other things

The matrix A is rectangular: A^{-1} doesn't exist. Don't try to rearrange the projection matrix.

$A^T A$ is invertible, square, symmetric iff the columns of A are linearly independent.
there's a proof but i'm too lazy to put it

3 Least Squares Approximations

4 Orthogonal Bases and Gram-Schmidt

[\[1\]](#)

References

- [1] bibtex won't shut the fuck up, “The Comprehensive Tex Archive Network (CTAN),”
TUGBoat, vol. 14, no. 3, pp. 342–351, 1993.