My Book Title

Author Name

October 11, 2025

Contents

Chapter 1

The Four Fundamental Subspaces

The goal of this section

The goal of this section is to understand the *four fundamental subspaces* associated with a finite dimensional matrix A over the real numbers $\mathbb{R}$. We begin by reviewing matrix multiplication, then define our four subspaces and prove our main theorem. Next we introduce the dual space and adjoint operators. We define the annihilator of a subspace and give an alternative proof of the main theorem. Finally we prove the rank-nullity theorem and also prove that the rank of the transpose of a matrix is equal to the rank of the original matrix.

Matrix Multiplication

Let us recall briefly the formula for the multiplication of a matrix by a vector. If $Bv = w$ then $w_k = \sum_j B_{kj} v_j$. This can be interpreted in two different ways. Firstly as the dot product of each of the rows of B with the vector v.

$$\begin{bmatrix} - & r_1 & - \\ - & r_2 & - \\ - & r_3 & - \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} r_1.v \\ r_2.v \\ r_3.v \end{bmatrix}$$

Secondly as a linear combination of the columns.

$$\begin{bmatrix} | & | & | \\ c_1 & c_2 & c_3 \\ | & | & | \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = v_1 \begin{bmatrix} | \\ c_1 \\ | \end{bmatrix} + v_2 \begin{bmatrix} | \\ c_2 \\ | \end{bmatrix} + v_3 \begin{bmatrix} | \\ c_3 \\ | \end{bmatrix}$$

It is worth pausing here for a moment to make sure you are clear on this. Here is an example:

Example

Let

$$B = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}, \qquad \mathbf{v} = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}.$$

Denote the rows of B by

$$r_1 = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}, \quad r_2 = \begin{bmatrix} 4 & 5 & 6 \end{bmatrix}, \quad r_3 = \begin{bmatrix} 7 & 8 & 9 \end{bmatrix},$$

and the columns by

$$c_1 = \begin{bmatrix} 1 \\ 4 \\ 7 \end{bmatrix}, \quad c_2 = \begin{bmatrix} 2 \\ 5 \\ 8 \end{bmatrix}, \quad c_3 = \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix}.$$

1. Row View: Dot Products

Each entry of $\mathbf{w} = B\mathbf{v}$ is the dot product of a row of B with $\mathbf{v}$:

$$\begin{aligned} w_1 &= r_1 \cdot \mathbf{v} = (1)(2) + (2)(-1) + (3)(1) = 2 - 2 + 3 = 3, \\ w_2 &= r_2 \cdot \mathbf{v} = (4)(2) + (5)(-1) + (6)(1) = 8 - 5 + 6 = 9, \\ w_3 &= r_3 \cdot \mathbf{v} = (7)(2) + (8)(-1) + (9)(1) = 14 - 8 + 9 = 15. \end{aligned}$$

Thus,

$$\begin{bmatrix} - & r_1 & - \\ - & r_2 & - \\ - & r_3 & - \end{bmatrix} \mathbf{v} = \begin{bmatrix} r_1 \cdot \mathbf{v} \\ r_2 \cdot \mathbf{v} \\ r_3 \cdot \mathbf{v} \end{bmatrix} = \begin{bmatrix} 3 \\ 9 \\ 15 \end{bmatrix}.$$

2. Column View: Linear Combination

The product $B\mathbf{v}$ is a linear combination of the columns of B, weighted by the entries of $\mathbf{v}$:

$$B\mathbf{v} = v_1 c_1 + v_2 c_2 + v_3 c_3 = 2c_1 + (-1)c_2 + 1c_3.$$

Compute:

$$2c_1 = \begin{bmatrix} 2 \\ 8 \\ 14 \end{bmatrix}, \quad -1c_2 = \begin{bmatrix} -2 \\ -5 \\ -8 \end{bmatrix}, \quad 1c_3 = \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix}.$$

Add them:

$$\begin{bmatrix} 2 \\ 8 \\ 14 \end{bmatrix} + \begin{bmatrix} -2 \\ -5 \\ -8 \end{bmatrix} + \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix} = \begin{bmatrix} 2-2+3 \\ 8-5+6 \\ 14-8+9 \end{bmatrix} = \begin{bmatrix} 3 \\ 9 \\ 15 \end{bmatrix}.$$

So,

$$\begin{bmatrix} | & | & | \\ c_1 & c_2 & c_3 \\ | & | & | \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} = 2 \begin{bmatrix} | \\ c_1 \\ | \end{bmatrix} - 1 \begin{bmatrix} | \\ c_2 \\ | \end{bmatrix} + 1 \begin{bmatrix} | \\ c_3 \\ | \end{bmatrix} = \begin{bmatrix} 3 \\ 9 \\ 15 \end{bmatrix}.$$

Conclusion

Both interpretations yield the same result:

$$B\mathbf{v} = \begin{bmatrix} 3 \\ 9 \\ 15 \end{bmatrix}.$$

The Transpose

Suppose that A is an m by n matrix. This means that A has m rows and n columns and is a map from $\mathbb{R}^n$ to $\mathbb{R}^m$.

The *transpose* map A^T is defined by: $A^T_{ij} = A_{ji}$. The transpose A^T is a map from $\mathbb{R}^m$ to $\mathbb{R}^n$. It has n rows and m columns.

$$A = \begin{array}{cc} n & \\ \begin{bmatrix} * & * & * \\ * & * & * \\ * & * & * \\ * & * & * \\ * & * & * \end{bmatrix} & m \end{array}$$

$$A^T = \begin{array}{cc} m & \\ \begin{bmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{bmatrix} & n \end{array}$$

The *image* of A is the subspace of $\mathbb{R}^m$ consisting of vectors of the form Av where $v \in \mathbb{R}^n$. This is also called the *column space* of A and denoted by $\mathrm{Col}(A)$.

The *kernel* of A is the subspace of $\mathbb{R}^n$ consisting of vectors v such that $Av = 0$. This is also called the *null space* of A and denoted by $\mathrm{Nul}(A)$.

To re-iterate:

$$\mathrm{Col}(A) \subseteq \mathbb{R}^m \qquad \mathrm{Nul}(A) \subseteq \mathbb{R}^n$$

It is important to note that these two subspaces to *n*ot lie in the same ambiant space.

By similar reasoning the *image* of A^T is the subspace of $\mathbb{R}^n$ consisting of vectors of the form A^Tv where $v \in \mathbb{R}^m$. This is also called the *column space* of A^T and denoted by $\mathrm{Col}(A^T)$. This subspace is sometimes also called the *row space* of A.

The *kernel* of A^T is the subspace of $\mathbb{R}^m$ consisting of vecors v such that $A^Tv = 0$. This is also called the *null space* of A^T and denoted by $\mathrm{Nul}(A^T)$.

We have:

$$\mathrm{Col}(A^T) \subseteq \mathbb{R}^n \qquad \mathrm{Nul}(A^T) \subseteq \mathbb{R}^m$$

Both $\mathrm{Col}(A^T)$ and $\mathrm{Nul}(A)$ lie in $\mathbb{R}^n$. And both $\mathrm{Col}(A)$ and $\mathrm{Nul}(A^T)$ lie in R^m

The main result of this section is the following:

Theorem 1. *The four fundamental subspaces satisfy the following orthogonal decompositions:*

$$\mathbb{R}^n = \mathrm{Col}(A^T) \oplus \mathrm{Nul}(A), \qquad \mathbb{R}^m = \mathrm{Col}(A) \oplus \mathrm{Nul}(A^\top)$$

where $\oplus$ *denotes an orthogonal direct sum.*

Proof. We must show that:

$$\boxed{\mathrm{Nul}(A) = \mathrm{Col}(A^T)^\perp} \quad \text{and} \quad \boxed{\mathrm{Nul}(A^\top) = \mathrm{Col}(A)^\perp}$$

Note that by symmetry we only really need to prove one of these identities. We will prove the one on the left.

Note firstly that the columns of A^T are simply the rows of A. So what we will actually prove is that:

$$\mathrm{Nul}(A) = \mathrm{Row}(A)^\perp$$

We shall show that a vector $\mathbf{x} \in \mathbb{R}^n$ satisfies $A\mathbf{x} = \mathbf{0}$ if and only if $\mathbf{x}$ is orthogonal to every vector in the row space of A.

Let the rows of A be $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_m \in \mathbb{R}^n$. Then:

$$A\mathbf{x} = \begin{bmatrix} \mathbf{r}_1 \cdot \mathbf{x} \\ \mathbf{r}_2 \cdot \mathbf{x} \\ \vdots \\ \mathbf{r}_m \cdot \mathbf{x} \end{bmatrix}$$

Thus,

$$A\mathbf{x} = \mathbf{0} \iff \mathbf{r}_i \cdot \mathbf{x} = 0 \quad \text{for all } i = 1, \ldots, m$$

Any vector $\mathbf{v} \in \mathrm{Row}(A)$ is a linear combination:

$$\mathbf{v} = c_1\mathbf{r}_1 + \cdots + c_m\mathbf{r}_m$$

Then:

$$\mathbf{v} \cdot \mathbf{x} = \sum_{i=1}^{m} c_i(\mathbf{r}_i \cdot \mathbf{x}) = 0$$

So $\mathbf{x} \perp \mathbf{v}$, hence $\mathbf{x} \in \text{Row}(A)^{\perp}$.

Conversely, if $\mathbf{x} \in \text{Row}(A)^{\perp}$, then in particular $\mathbf{x} \perp \mathbf{r}_i$ for each row $\mathbf{r}_i$, so $A\mathbf{x} = \mathbf{0}$, meaning $\mathbf{x} \in \text{Nul}(A)$.

Therefore:

$$\boxed{\text{Nul}(A) = \text{Row}(A)^{\perp}}$$

$\square$

Theorem (Rank–Nullity Theorem)

Let $T : V \to W$ be a linear transformation between finite-dimensional vector spaces over a field $\mathbb{R}$. Then

$$\dim(V) = \dim(\ker T) + \dim(\operatorname{Im} T).$$

In particular, for any $m \times n$ matrix A over $\mathbb{R}$, viewing A as a linear map $\mathbb{R}^n \to \mathbb{R}^m$, we have

$$n = \text{nullity}(A) + \text{rank}(A).$$

Proof

Let $K = \ker T \subseteq V$. Since V is finite-dimensional, so is K. Let

$$\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$$

be a basis for K, so $k = \dim(\ker T)$.

Extend this to a basis for all of V:

$$\{\mathbf{u}_1, \dots, \mathbf{u}_k, \mathbf{v}_1, \dots, \mathbf{v}_r\},$$

so that $\dim(V) = k + r$.

We claim that the set

$$\{T(\mathbf{v}_1), \dots, T(\mathbf{v}_r)\}$$

is a basis for $\operatorname{Im} T$.

1. Spanning

Let $\mathbf{w} \in \operatorname{Im} T$. Then $\mathbf{w} = T(\mathbf{x})$ for some $\mathbf{x} \in V$. Write

$$\mathbf{x} = \sum_{i=1}^{k} a_i \mathbf{u}_i + \sum_{j=1}^{r} b_j \mathbf{v}_j.$$

Applying T, and using $T(\mathbf{u}_i) = \mathbf{0}$ (since $\mathbf{u}_i \in \ker T$), we get

$$T(\mathbf{x}) = \sum_{j=1}^{r} b_j T(\mathbf{v}_j).$$

Thus, $\{T(\mathbf{v}_1), \ldots, T(\mathbf{v}_r)\}$ spans $\operatorname{Im} T$.

2. Linear Independence

Suppose

$$\sum_{j=1}^{r} c_j T(\mathbf{v}_j) = \mathbf{0}.$$

Then

$$T\left(\sum_{j=1}^{r} c_j \mathbf{v}_j\right) = \mathbf{0},$$

so $\sum_{j=1}^{r} c_j \mathbf{v}_j \in \ker T = \operatorname{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$.

But the full set $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is linearly independent. Therefore, the only linear combination of the $\mathbf{v}_j$'s that lies in $\operatorname{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is the trivial one. Hence, $c_j = 0$ for all j, and the set $\{T(\mathbf{v}_1), \ldots, T(\mathbf{v}_r)\}$ is linearly independent.

3. Conclusion

We have shown that $\dim(\operatorname{Im} T) = r$. Since $\dim(V) = k + r$, it follows that

$$\dim(V) = \dim(\ker T) + \dim(\operatorname{Im} T).$$

This completes the proof.

Introduction

A fundamental fact in linear algebra is that the rank of a matrix A is equal to the rank of its transpose $A^\top$. In other words, the maximum number of linearly independent **columns** of A is the same as the maximum number of linearly independent **rows** of A.

This might seem surprising at first—after all, rows and columns live in different spaces! But in $\mathbb{R}^n$ (or $\mathbb{C}^n$), the presence of an inner product (the dot product) creates a deep symmetry between rows and columns.

We will prove this result using two key ideas:

1. The **Rank–Nullity Theorem**,
2. The **orthogonal relationship** between the null space and the row space.

This proof works for real matrices (and complex matrices with minor adjustments), but it relies on the geometry of the dot product.

Step 1: What the Rank–Nullity Theorem Tells Us

Let A be an $m \times n$ real matrix. We can think of A as a linear transformation that maps vectors from $\mathbb{R}^n$ to $\mathbb{R}^m$.

The Rank–Nullity Theorem says:

(dimension of domain) = (dimension of null space)+(dimension of column space).

In symbols:

$$n = \dim(\mathrm{Nul}(A)) + \dim(\mathrm{Col}(A)). \tag{1}$$

Here:

- $\mathrm{Nul}(A) = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{0}\}$ is the set of vectors that A sends to zero,
- $\mathrm{Col}(A)$ is the span of the columns of A, and its dimension is the **column rank** of A.

Step 2: The Geometric Link Between Rows and Null Space

Now consider the **row space** of A, denoted $\mathrm{Row}(A)$. This is the subspace of $\mathbb{R}^n$ spanned by the rows of A.

Here's the key geometric insight:

A vector $\mathbf{x}$ is in the null space of A if and only if it is perpendicular to every row of A.

Why? Because the equation $A\mathbf{x} = \mathbf{0}$ means that the dot product of each row of A with $\mathbf{x}$ is zero.

In other words:

$$\mathrm{Nul}(A) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \perp \mathbf{r} \text{ for every row } \mathbf{r} \text{ of } A\}.$$

This is precisely the definition of the **orthogonal complement** of the row space. So we have:

$$\mathrm{Nul}(A) = \mathrm{Row}(A)^{\perp}. \tag{2}$$

Step 3: Use Dimensions of Orthogonal Complements

In $\mathbb{R}^n$, if S is any subspace, then:

$$\dim(S) + \dim(S^{\perp}) = n.$$

Apply this to $S = \mathrm{Row}(A)$. Using (2), we get:

$$\dim(\mathrm{Row}(A)) + \dim(\mathrm{Nul}(A)) = n. \tag{3}$$

But $\dim(\mathrm{Row}(A))$ is exactly the **row rank** of A.

Step 4: Compare with Rank–Nullity

Now look back at equation (1) from Rank–Nullity:

$$n = \dim(\mathrm{Nul}(A)) + \dim(\mathrm{Col}(A)).$$

And equation (3) from orthogonality:

$$n = \dim(\mathrm{Nul}(A)) + \dim(\mathrm{Row}(A)).$$

Both right-hand sides equal n, and both contain the term $\dim(\mathrm{Nul}(A))$. Therefore, the remaining terms must be equal:

$$\dim(\mathrm{Col}(A)) = \dim(\mathrm{Row}(A)).$$

In other words:

$$\text{column rank of } A = \text{row rank of } A.$$

Since the row rank of A is the same as the column rank of $A^\top$, we conclude:

$$\mathrm{rank}(A) = \mathrm{rank}(A^\top).$$

Chapter 2

The Dual Space

The Dual Space

Hopefully the reader has seen the idea of the *dual space* before, otherwise this section will probably be a little confusing.

Let V be a finite dimensional vector space over the real numbers $\mathbb{R}$ A **linear functional** is a function $f : V \to \mathbb{R}$ such that for all vectors $\mathbf{u}, \mathbf{v} \in V$ and scalars $\alpha, \beta \in \mathbb{R}$,

$$f(\alpha\mathbf{u} + \beta\mathbf{v}) = \alpha f(\mathbf{u}) + \beta f(\mathbf{v}).$$

The **dual space** of V, denoted V^*, is the set of all **linear functionals** on V. That is,

$$V^* = \{f : V \to \mathbb{R} \mid f \text{ is linear}\}.$$

The dual space V^* is itself a vector space over $\mathbb{R}$, with vector addition and scalar multiplication defined *pointwise*:

$$(f + g)(\mathbf{v}) = f(\mathbf{v}) + g(\mathbf{v}),$$
$$(\alpha f)(\mathbf{v}) = \alpha\, f(\mathbf{v}),$$

for all $f, g \in V^*$, $\alpha \in \mathbb{R}$, and $\mathbf{v} \in V$.

Key Point

One of the most important properties of the dual space of V is that the standard dot product allows us to associate every vector $\mathbf{v} \in V$ with a linear functional $f_{\mathbf{v}} \in V^*$ via the rule

$$\boxed{f_{\mathbf{v}}(\mathbf{y}) = \mathbf{v} \cdot \mathbf{y} = \mathbf{v}^\top \mathbf{y}, \quad \text{for all } \mathbf{y} \in V}.$$

Theorem 2 (Finite-Dimensional Riesz Representation Theorem)**.** *Let $V = \mathbb{R}^n$ with the standard inner product. Then the map*

$$\psi : V \to V^*, \quad \psi(\mathbf{v})(\mathbf{y}) = \mathbf{v}^\top \mathbf{y},$$

is a vector space isomorphism.

Proof. First, linearity: for all $\alpha, \beta \in \mathbb{R}$ and $\mathbf{v}, \mathbf{w}, \mathbf{y} \in V$,

$$f_{\alpha\mathbf{v}+\beta\mathbf{w}}(\mathbf{y}) = (\alpha\mathbf{v} + \beta\mathbf{w})^\top \mathbf{y} = \alpha\mathbf{v}^\top \mathbf{y} + \beta\mathbf{w}^\top \mathbf{y} = \alpha f_{\mathbf{v}}(\mathbf{y}) + \beta f_{\mathbf{w}}(\mathbf{y}).$$

Next, injectivity: if $f_{\mathbf{v}} = 0$, then $\mathbf{v}^\top \mathbf{y} = 0$ for all $\mathbf{y}$. In particular, $\mathbf{v}^\top \mathbf{v} = \|\mathbf{v}\|^2 = 0$, so $\mathbf{v} = \mathbf{0}$.

Finally, surjectivity: given $g \in V^*$, define $v_i = g(\mathbf{e}_i)$ for the standard basis $\{\mathbf{e}_i\}$. Set $\mathbf{v} = \sum_i v_i \mathbf{e}_i$. Then for each basis vector,

$$f_{\mathbf{v}}(\mathbf{e}_i) = \mathbf{v}^\top \mathbf{e}_i = v_i = g(\mathbf{e}_i).$$

Since $f_{\mathbf{v}}$ and g agree on a basis, they are equal.

Thus ψ is linear, injective, and surjective, hence an isomorphism. ☐

Corollary 1. $\dim V^* = \dim V$ *for finite-dimensional V*

As a consequence of the above, suppose that $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is a basis for V, then there is a uniquely associated **dual basis** $\{\mathbf{e}^1, \ldots, \mathbf{e}^n\} \subseteq V^*$ defined by

$$\mathbf{e}^i(\mathbf{e}_j) = \delta^i_j = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Every linear functional $f \in V^*$ can be uniquely expressed as

$$f = \sum_{i=1}^{n} f(\mathbf{e}_i)\, \mathbf{e}^i.$$

Some intuition

It can sometimes help to think of vectors in V as column vectors and vectors in V^* as row vectors. The dot product is now the application of a linear functional to a vector. The transpose is the isomoprhism between V and V^*.

The Adjoint

The reason why we have gone to all the trouble of introducing the dual space is that the transpose of a matrix is the natural matrix representation of the **adjoint** of a linear map.

The Adjoint of a Linear Map

Let $T : V \to W$ be a linear map between finite-dimensional real vector spaces. The **dual map** $T^* : W^* \to V^*$ is defined by

$$(T^* f)(\mathbf{v}) = f(T\mathbf{v}) \quad \text{for all } f \in W^*,\ \mathbf{v} \in V.$$

In words: to evaluate $T^* f$ at a vector $\mathbf{v}$, first apply T to $\mathbf{v}$, then apply the functional f to the result.

Theorem 3. *The matrix of the dual map T^* is the transpose of the matrix of T.*

Proof. Choose bases:

- $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ for V,
- $\mathcal{C} = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ for W,
- $\mathcal{B}^* = \{\mathbf{v}^1, \ldots, \mathbf{v}^n\}$ and $\mathcal{C}^* = \{\mathbf{w}^1, \ldots, \mathbf{w}^m\}$ for the dual bases.

Suppose the matrix of T with respect to $\mathcal{B}, \mathcal{C}$ is $A = (a_{ij})$, then the image of v_1 is the first column of A:

$$T(\mathbf{v}_1) = \sum_{i=1}^{m} A_{i1}\, \mathbf{w}_i$$

More generally, the image of v_j is the jth column of A:

$$T(\mathbf{v}_j) = \sum_{i=1}^{m} a_{ij}\mathbf{w}_i.$$

We compute the matrix of T^* with respect to $\mathcal{C}^*, \mathcal{B}^*$. For any $k \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, n\}$,

$$(T^*\mathbf{w}^k)(\mathbf{v}_j) = \mathbf{w}^k(T\mathbf{v}_j) = \mathbf{w}^k\left(\sum_{i=1}^{m} a_{ij}\mathbf{w}_i\right) = a_{kj}.$$

On the other hand, if the matrix of T^* is $B = (b_{\ell k})$, then

$$T^*(\mathbf{w}^1) = \sum_{\ell=1}^{n} b_{\ell 1}\,\mathbf{v}^\ell$$

More generally:

$$T^*(\mathbf{w}^k) = \sum_{\ell=1}^{n} b_{\ell k}\mathbf{v}^\ell,$$

so

$$(T^*\mathbf{w}^k)(\mathbf{v}_j) = b_{jk}.$$

Comparing both expressions gives $b_{jk} = a_{kj}$, so $B = A^\top$.

That is the matrix of the dual map T^* is the transpose of the matrix of T. $\square$

The Four Fundamental Subspaces

Recall from Chapter 1 that the key orthogonal decompositions are:

$$\mathbb{R}^n = \mathrm{Row}(A) \oplus \mathrm{Nul}(A), \quad \mathbb{R}^m = \mathrm{Col}(A) \oplus \mathrm{Nul}(A^\top),$$

or equivalently:

$$\mathrm{Nul}(A) = \mathrm{Row}(A)^\perp, \quad \mathrm{Nul}(A^\top) = \mathrm{Col}(A)^\perp.$$

The Annihilator

Let V be a finite-dimensional vector space, and let $W \subseteq V$ be a subspace. The **annihilator** of W is:

$$W^\circ = \{f \in V^* : f(\mathbf{w}) = 0 \text{ for all } \mathbf{w} \in W\},$$

where V^* is the **dual space** (the space of linear functionals on V).

Some notation

If one is not careful then it is easy to confuse the vector space V with the vector spacd V^*. For this reason we will use the following notation. Suppose that $A \subseteq V$ and $B \subseteq V^* R$ then we write:

$$A \cong B$$

When we mean to say that:

$$A = \psi^{-1}[B]$$

where ψ is the isomorphism defined in Theorem 2. We come now to the most important result in this Chapter.

Theorem 4.

$$\mathrm{Nul}(A^\top) \cong (\mathrm{Col}(A))^\circ, \qquad \mathrm{Nul}(A) \cong (\mathrm{Row}(A))^\circ,$$

where the annihilator is taken inside the appropriate dual space, and the isomorphism is the one induced by the dot product.

Proof. We only need to prove one of these results, and the second one follows by symmetry.

Let A be an $m \times n$ real matrix. We will prove that, under the natural identification of vectors in $\mathbb{R}^m$ with linear functionals on $\mathbb{R}^m$ provided by the dot product, the null space of $A^\top$ corresponds, under isomorphism, to the annihilator of the column space of A. That is,

$$\mathrm{Nul}(A^\top) \cong (\mathrm{Col}(A))^\circ.$$

We proceed step by step, explaining the meaning of each concept and how they relate.

The **column space** of A is the subspace of $\mathbb{R}^m$ defined by

$$\mathrm{Col}(A) = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\} \subseteq \mathbb{R}^m.$$

Its **annihilator** is the set of all linear functionals on $\mathbb{R}^m$ that vanish on every vector in $\mathrm{Col}(A)$:

$$(\mathrm{Col}(A))^\circ = \{f \in (\mathbb{R}^m)^* \mid f(\mathbf{y}) = 0 \text{ for all } \mathbf{y} \in \mathrm{Col}(A)\}.$$

Because every $\mathbf{y} \in \mathrm{Col}(A)$ can be written as $\mathbf{y} = A\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^n$, we can rephrase this condition as:

$$f \in (\mathrm{Col}(A))^\circ \quad \Longleftrightarrow \quad f(A\mathbf{x}) = 0 \text{ for all } \mathbf{x} \in \mathbb{R}^n.$$

Now consider a vector $\mathbf{v} \in \mathbb{R}^m$, and let $f_\mathbf{v}$ be the corresponding functional: $f_\mathbf{v}(\mathbf{y}) = \mathbf{v}^\top \mathbf{y}$. We ask: *When does $f_\mathbf{v}$ belong to $(\mathrm{Col}(A))^\circ$?*

By the characterization above, this happens precisely when

$$f_\mathbf{v}(A\mathbf{x}) = 0 \quad \text{for all } \mathbf{x} \in \mathbb{R}^n.$$

Substituting the definition of $f_\mathbf{v}$, this becomes:

$$\mathbf{v}^\top (A\mathbf{x}) = 0 \quad \text{for all } \mathbf{x} \in \mathbb{R}^n. \tag{2.1}$$

Step 4: Rewrite the condition using properties of the transpose

Using the fact that $(XY)^\top = Y^\top X^\top$, we may rewrite this as:

$$\mathbf{v}^\top A\mathbf{x} = (A^\top \mathbf{v})^\top \mathbf{x}.$$

Thus, condition (2.1) is equivalent to:

$$(A^\top \mathbf{v})^\top \mathbf{x} = 0 \quad \text{for all } \mathbf{x} \in \mathbb{R}^n.$$

In order for this to be true for all x the linear functional $f_{A^T\mathbf{v}}$ must be zero, which iimplies that $A^Tv = 0$ which implies that $v \in \mathrm{Nul}(A^T)$.

We have shown the following chain of equivalences:

$$\begin{aligned}
\mathbf{v} \in \mathrm{Nul}(A^\top) &\iff A^\top \mathbf{v} = \mathbf{0} \\
&\iff (A^\top \mathbf{v})^\top \mathbf{x} = 0 \text{ for all } \mathbf{x} \in \mathbb{R}^n \\
&\iff \mathbf{v}^\top A\mathbf{x} = 0 \text{ for all } \mathbf{x} \in \mathbb{R}^n \\
&\iff f_{\mathbf{v}}(A\mathbf{x}) = 0 \text{ for all } \mathbf{x} \in \mathbb{R}^n \\
&\iff f_{\mathbf{v}} \in (\mathrm{Col}(A))^\circ.
\end{aligned}$$

Therefore, under the isomorphism $\mathbf{v} \leftrightarrow f_{\mathbf{v}}$, the subspace $\mathrm{Nul}(A^\top) \subseteq \mathbb{R}^m$ corresponds, under isomorphism, to the subspace $(\mathrm{Col}(A))^\circ \subseteq (\mathbb{R}^m)^*$.

□

Geometric Interpretation

This result has a clean geometric meaning: a vector $\mathbf{v} \in \mathbb{R}^m$ is orthogonal (with respect to the dot product) to every vector in the column space of A if and only if $A^\top \mathbf{v} = \mathbf{0}$. But "orthogonal to the column space" is precisely what it means for the functional $f_{\mathbf{v}}(\mathbf{y}) = \mathbf{v} \cdot \mathbf{y}$ to vanish on $\mathrm{Col}(A)$ — i.e., to be in the annihilator. Thus, in $\mathbb{R}^m$, the annihilator $(\mathrm{Col}(A))^\circ$ is naturally identified with the orthogonal complement $\mathrm{Col}(A)^\perp$, and we recover the familiar fundamental theorem of linear algebra:

$$\mathrm{Nul}(A^\top) = \mathrm{Col}(A)^\perp.$$

Theorem 5. $\mathrm{Nul}(A) \cong (\mathrm{Row}(A))^\circ$

Proof. This follows immediately by symmetry. □

Conclusion. Via the isomorphism ψ between $\mathbb{R}^k$ $(\mathbb{R}^k)^*$, we have the following identities:

$$\boxed{\mathrm{Nul}(A^\top) \cong (\mathrm{Col}(A))^\circ}, \qquad \boxed{\mathrm{Nul}(A) \cong (\mathrm{Row}(A))^\circ}.$$

The Annihilator reveals that the four subspaces reflect a fundamental duality. The matrix $A : \mathbb{R}^n \to \mathbb{R}^m$ induces a **dual map** $A^* : (\mathbb{R}^m)^* \to (\mathbb{R}^n)^*$, and we have:

$$\ker(A^*) = (\operatorname{Im} A)^\circ, \qquad \operatorname{Im}(A^*) = (\ker A)^\circ.$$

When we identify $(\mathbb{R}^k)^* \cong \mathbb{R}^k$ via the isomorphism ψ, these become:

$$\operatorname{Nul}(A^\top) = \operatorname{Col}(A)^\perp, \qquad \operatorname{Row}(A) = \operatorname{Nul}(A)^\perp.$$

Hopefully this chapter has been illuminating.

Chapter 3

Orthogonal Projections

A projection is a linear operator P satisfying $P^2 = P$. One can think of P as a transformation that **fixes every vector in its column space**—that is, if $\mathbf{v} \in \mathrm{Col}(P)$, then $P\mathbf{v} = \mathbf{v}$.

Given any vector $\mathbf{x}$, we can decompose it as

$$\mathbf{x} = P\mathbf{x} + (\mathbf{x} - P\mathbf{x}),$$

where:

- $P\mathbf{x}$ is the **projected part**, lying in $\mathrm{Col}(P)$;
- $\mathbf{x} - P\mathbf{x}$ is the **residual** (or error), which lies in the **null space** $\mathrm{Nul}(P)$, because
$$P(\mathbf{x} - P\mathbf{x}) = P\mathbf{x} - P^2\mathbf{x} = P\mathbf{x} - P\mathbf{x} = \mathbf{0}.$$

Thus, the entire space splits as a **direct sum**:

$$\mathbb{R}^n = \mathrm{Col}(P) \oplus \mathrm{Nul}(P).$$

Every vector can be written uniquely as a sum of a vector in the column space and a vector in the null space.

An **orthogonal projection** can be described in two seemingly different ways:

1. **Geometrically**: a projection P is orthogonal if its column space and null space are perpendicular, i.e.,
$$\mathrm{Col}(P) \perp \mathrm{Nul}(P).$$

2. **Algebraically**: a projection P is orthogonal if it is symmetric, i.e.,

$$P^\top = P.$$

The purpose of this note is to **prove the equivalence of these two definitions**. That is, for any matrix P satisfying $P^2 = P$, we will show:

$$\mathrm{Col}(P) \perp \mathrm{Nul}(P) \quad \Longleftrightarrow \quad P^\top = P.$$

Theorem 6 (The Orthogonal Projection Theorem)**.** *Let P be a linear operator on $\mathbb{R}^n$ such that $P^2 = P$ (i.e., P is a projection). Then the following are equivalent:*

1. *P is an **orthogonal projection**, meaning* $\mathrm{Col}(P) \perp \mathrm{Nul}(P)$*;*
2. *P is **symmetric**, i.e., $P^\top = P$.*

In other words, for a projection, symmetry is equivalent to orthogonality of the column and null spaces.

Proof. We prove both directions.

(1) Symmetry $\implies$ Orthogonality. Assume $P^2 = P$ and $P^\top = P$. By the Fundamental Theorem of Linear Algebra,

$$\mathrm{Col}(P)^\perp = \mathrm{Nul}(P^\top).$$

Since P is symmetric, $P^\top = P$, so $\mathrm{Nul}(P^\top) = \mathrm{Nul}(P)$. Hence,

$$\mathrm{Col}(P)^\perp = \mathrm{Nul}(P),$$

which means $\mathrm{Col}(P) \perp \mathrm{Nul}(P)$. Thus, P is an orthogonal projection.

(2) Orthogonality $\implies$ Symmetry. Assume $P^2 = P$ and $\mathrm{Col}(P) \perp \mathrm{Nul}(P)$. Then $\mathrm{Nul}(P) = \mathrm{Col}(P)^\perp$. Again by the Fundamental Theorem,

$$\mathrm{Col}(P)^\perp = \mathrm{Nul}(P^\top),$$

so we obtain

$$\mathrm{Nul}(P) = \mathrm{Nul}(P^\top).$$

To show $P = P^\top$, it suffices to verify that for all vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,

$$\mathbf{x}^\top P \mathbf{y} = \mathbf{x}^\top P^\top \mathbf{y}.$$

Because P is an orthogonal projection, every vector decomposes orthogonally as

$$\mathbf{z} = P\mathbf{z} + (\mathbf{z} - P\mathbf{z}), \quad \text{with } P\mathbf{z} \in \mathrm{Col}(P),\ \mathbf{z} - P\mathbf{z} \in \mathrm{Nul}(P).$$

Now consider arbitrary $\mathbf{x}, \mathbf{y}$. Since $\mathbf{x} - P\mathbf{x} \in \text{Nul}(P)$ and $P\mathbf{y} \in \text{Col}(P)$, orthogonality gives

$$(\mathbf{x} - P\mathbf{x})^\top P\mathbf{y} = 0 \quad \Longrightarrow \quad \mathbf{x}^\top P\mathbf{y} = (P\mathbf{x})^\top P\mathbf{y}. \tag{*}$$

Similarly, $P\mathbf{x} \in \text{Col}(P)$ and $\mathbf{y} - P\mathbf{y} \in \text{Nul}(P)$ are orthogonal, so

$$(P\mathbf{x})^\top(\mathbf{y} - P\mathbf{y}) = 0 \quad \Longrightarrow \quad (P\mathbf{x})^\top\mathbf{y} = (P\mathbf{x})^\top P\mathbf{y}. \tag{**}$$

From (*) and (**), we conclude

$$\mathbf{x}^\top P\mathbf{y} = (P\mathbf{x})^\top\mathbf{y} = \mathbf{x}^\top P^\top\mathbf{y}.$$

Since this holds for all $\mathbf{x}, \mathbf{y}$, it follows that $P = P^\top$.

Thus, the two conditions are equivalent. □

Finding particular projection matrices

Now that we understand what an orthogonal projection is, we would like to be able to construct explicit matrices which project onto a given subspace. Let us begin with the one dimensional case.

Goal

Let $\mathbf{u}_1 \in \mathbb{R}^n$ be a unit vector (i.e., $\|\mathbf{u}_1\| = 1$). We want to find the matrix P that orthogonally projects any vector $\mathbf{x} \in \mathbb{R}^n$ onto the line spanned by $\mathbf{u}_1$.

Geometric Definition of Orthogonal Projection

The orthogonal projection $\hat{\mathbf{x}} = P\mathbf{x}$ must satisfy:

1. $\hat{\mathbf{x}}$ lies on the line: $\hat{\mathbf{x}} = c\mathbf{u}_1$ for some scalar c,
2. The error $\mathbf{x} - \hat{\mathbf{x}}$ is perpendicular to the line: $(\mathbf{x} - \hat{\mathbf{x}})^\top\mathbf{u}_1 = 0$.

Deriving the Projection Formula

From condition (2):

$$(\mathbf{x} - c\mathbf{u}_1)^\top \mathbf{u}_1 = 0 \quad \Rightarrow \quad \mathbf{x}^\top \mathbf{u}_1 - c\,\mathbf{u}_1^\top \mathbf{u}_1 = 0.$$

Since $\mathbf{u}_1$ is a unit vector, $\mathbf{u}_1^\top \mathbf{u}_1 = 1$, so

$$c = \mathbf{u}_1^\top \mathbf{x}.$$

Thus, the projection is

$$\hat{\mathbf{x}} = (\mathbf{u}_1^\top \mathbf{x})\,\mathbf{u}_1.$$

Expressing as a Matrix Multiplication

We now seek a matrix P such that $\hat{\mathbf{x}} = P\mathbf{x}$ for all $\mathbf{x}$. Rewrite the expression using the fact that $(\mathbf{u}_1^\top \mathbf{x})$ is a scalar together with the associativity of matrix multiplication:

$$\hat{\mathbf{x}} = (\mathbf{u}_1^\top \mathbf{x})\,\mathbf{u}_1 = \mathbf{u}_1(\mathbf{u}_1^\top \mathbf{x}) = (\mathbf{u}_1\mathbf{u}_1^\top)\mathbf{x}.$$

Since this holds for every $\mathbf{x}$, we identify

$$P = \mathbf{u}_1\mathbf{u}_1^\top.$$

Verifying Projection Properties

- **Idempotent**:

  $$P^2 = (\mathbf{u}_1\mathbf{u}_1^\top)(\mathbf{u}_1\mathbf{u}_1^\top) = \mathbf{u}_1(\mathbf{u}_1^\top\mathbf{u}_1)\mathbf{u}_1^\top = \mathbf{u}_1(1)\mathbf{u}_1^\top = P.$$

- **Symmetric**:

  $$P^\top = (\mathbf{u}_1\mathbf{u}_1^\top)^\top = \mathbf{u}_1\mathbf{u}_1^\top = P.$$

  Thus, P is an orthogonal projection matrix.

Geometric Interpretation

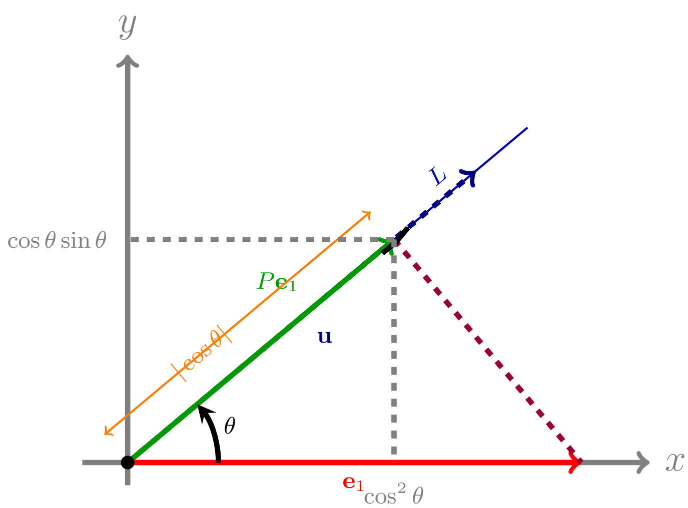
The matrix $\mathbf{u}_1\mathbf{u}_1^\top$ is the **outer product** of $\mathbf{u}_1$ with itself. - The inner product $\mathbf{u}_1^\top\mathbf{x}$ gives the scalar coordinate along $\mathbf{u}_1$, - The outer product $\mathbf{u}_1\mathbf{u}_1^\top$ converts this into a linear transformation that projects any vector onto the line.

Example: 2D Projection

Let $\mathbf{u}_1 = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$. Then

$$P = \mathbf{u}_1 \mathbf{u}_1^\top = \begin{bmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{bmatrix},$$

This is the standard orthogonal projection matrix onto a line at angle θ. To see this let L be the line through the origin in $\mathbb{R}^2$ that makes an angle θ with the positive x-axis.



We consider the **orthogonal projection** of the vector $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ onto L. We will show geometrically that the projected point has coordinates

$$(\cos^2\theta,\ \cos\theta\sin\theta).$$

Step 1: The length of the projection is $\cos\theta$

The direction of the line L is given by the **unit vector**

$$\mathbf{u} = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}.$$

The orthogonal projection of any vector $\mathbf{x}$ onto L lies along $\mathbf{u}$, at a distance equal to the scalar component of $\mathbf{x}$ in the direction of $\mathbf{u}$.

For $\mathbf{x} = \mathbf{e}_1 = (1, 0)$, this scalar component is the dot product:

$$\mathbf{u}^\top \mathbf{e}_1 = \cos\theta \cdot 1 + \sin\theta \cdot 0 = \cos\theta.$$

Thus, the projected point P lies on L, at a distance $\cos\theta$ from the origin.

Step 2: Coordinates of a point at distance $\cos\theta$ along L

Any point on L at distance r from the origin has coordinates

$$(r\cos\theta,\ r\sin\theta),$$

because you travel r units in the direction $(\cos\theta, \sin\theta)$.

Here, $r = \cos\theta$, so the coordinates of the projection are:

$$\big(\cos\theta \cdot \cos\theta,\ \cos\theta \cdot \sin\theta\big) = \big(\cos^2\theta,\ \cos\theta\sin\theta\big).$$

Hence, the orthogonal projection of $\mathbf{e}_1$ onto the line at angle θ is

$$P\mathbf{e}_1 = \begin{bmatrix} \cos^2\theta \\ \cos\theta\sin\theta \end{bmatrix}.$$

A similar argument holds for the projection of e_2.

Conclusion

The matrix $\mathbf{u}_1\mathbf{u}_1^\top$ is the unique orthogonal projection matrix onto the line spanned by the unit vector $\mathbf{u}_1$. It arises naturally from the geometric definition of orthogonal projection and satisfies all the required algebraic properties.

Higher Dimensions

Theorem

Let $\mathcal{S} \subseteq \mathbb{R}^n$ be a k-dimensional subspace, and let $U \in \mathbb{R}^{n\times k}$ be a matrix whose columns $\mathbf{u}_1, \ldots, \mathbf{u}_k$ form an orthonormal basis for $\mathcal{S}$ (i.e., $U^\top U = I_k$). Then the matrix

$$P = UU^\top$$

is the unique orthogonal projection matrix onto $\mathcal{S}$.

Proof

We verify that P satisfies the defining properties of an orthogonal projection onto $\mathcal{S}$.

1. $P\mathbf{x} \in \mathcal{S}$ for all $\mathbf{x} \in \mathbb{R}^n$

For any $\mathbf{x} \in \mathbb{R}^n$,

$$P\mathbf{x} = UU^\top \mathbf{x} = U(U^\top \mathbf{x}).$$

Let $\mathbf{z} = U^\top \mathbf{x} \in \mathbb{R}^k$. Then

$$P\mathbf{x} = U\mathbf{z} = z_1\mathbf{u}_1 + z_2\mathbf{u}_2 + \cdots + z_k\mathbf{u}_k,$$

which is a linear combination of the basis vectors of $\mathcal{S}$. Hence, $P\mathbf{x} \in \mathcal{S}$.

2. The error $\mathbf{x} - P\mathbf{x}$ is orthogonal to $\mathcal{S}$

We show $\mathbf{x} - P\mathbf{x} \perp \mathbf{v}$ for every $\mathbf{v} \in \mathcal{S}$. Since $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ spans $\mathcal{S}$, it suffices to show orthogonality to each basis vector $\mathbf{u}_i$.

Compute the inner product with $\mathbf{u}_i$:

$$\mathbf{u}_i^\top(\mathbf{x} - P\mathbf{x}) = \mathbf{u}_i^\top \mathbf{x} - \mathbf{u}_i^\top UU^\top \mathbf{x}.$$

Note that $\mathbf{u}_i^\top U$ is the i-th row of $U^\top U = I_k$, so $\mathbf{u}_i^\top U = \mathbf{e}_i^\top$, where $\mathbf{e}_i$ is the i-th standard basis vector in $\mathbb{R}^k$. Thus,

$$\mathbf{u}_i^\top UU^\top \mathbf{x} = \mathbf{e}_i^\top U^\top \mathbf{x} = (U^\top \mathbf{x})_i = \mathbf{u}_i^\top \mathbf{x}.$$

Therefore,

$$\mathbf{u}_i^\top(\mathbf{x} - P\mathbf{x}) = \mathbf{u}_i^\top \mathbf{x} - \mathbf{u}_i^\top \mathbf{x} = 0.$$

Since this holds for all $i = 1, \ldots, k$, the error $\mathbf{x} - P\mathbf{x}$ is orthogonal to every vector in $\mathcal{S}$.

3. P is symmetric and idempotent

- **Symmetric**: $P^\top = (UU^\top)^\top = UU^\top = P$.
- **Idempotent**: $P^2 = (UU^\top)(UU^\top) = U(U^\top U)U^\top = UI_kU^\top = UU^\top = P$, where we used the orthonormality condition $U^\top U = I_k$.

4. Uniqueness

Suppose Q is another matrix such that $Q\mathbf{x} \in \mathcal{S}$ and $\mathbf{x} - Q\mathbf{x} \perp \mathcal{S}$ for all $\mathbf{x}$. Since $Q\mathbf{x} \in \mathcal{S}$, we can write $Q\mathbf{x} = U\mathbf{w}$ for some $\mathbf{w} \in \mathbb{R}^k$. Orthogonality implies $U^\top(\mathbf{x} - Q\mathbf{x}) = \mathbf{0}$, so

$$U^\top\mathbf{x} - U^\top U\mathbf{w} = \mathbf{0} \quad \Rightarrow \quad U^\top\mathbf{x} - I_k\mathbf{w} = \mathbf{0} \quad \Rightarrow \quad \mathbf{w} = U^\top\mathbf{x}.$$

Thus, $Q\mathbf{x} = UU^\top\mathbf{x} = P\mathbf{x}$ for all $\mathbf{x}$, so $Q = P$.

Conclusion

The matrix $P = UU^\top$ is the unique orthogonal projection matrix onto the subspace $\mathcal{S} \subseteq \mathbb{R}^n$, for any dimension $k = \dim(\mathcal{S}) \geq 1$.