

# Column Space, Null Space and Rank-Nullity

Last modified: 2026-10-27 20:53:54 (PDT)

## Column space and null space

**Definition 0.1** (Column space, range, image). The **column space** [1] of an  $m \times n$  matrix  $\mathbf{A}$  is

$$\mathcal{C}(\mathbf{A}) \stackrel{\text{def}}{=} \{\mathbf{A}\tilde{x} : \tilde{x} \in \mathbb{R}^n\},$$

the set of all vectors  $\mathbf{A}\tilde{x}$  in  $\mathbb{R}^m$ . It is also called the **range** or **image** of  $\mathbf{A}$ .

**Definition 0.2** (Row space). The **row space** of an  $m \times n$  matrix  $\mathbf{A}$  is  $\mathcal{C}(\mathbf{A}^\top)$ , a set of vectors in  $\mathbb{R}^n$ .

**Example 0.1** (The column space of a rank-one matrix). Let

$$\mathbf{A} = \begin{bmatrix} 1 & -2 & -2 \\ 3 & -6 & -6 \end{bmatrix}.$$

For any  $\tilde{x} \in \mathbb{R}^3$ ,

$$\begin{aligned} \mathbf{A}\tilde{x} &= (x_1 - 2x_2 - 2x_3, 3(x_1 - 2x_2 - 2x_3)) \\ &= (x_1 - 2x_2 - 2x_3)(1, 3) \end{aligned}$$

([Definition 13 in Matrices](#)). The number  $x_1 - 2x_2 - 2x_3$  can be any real number (take  $x_2 = 0$  and  $x_3 = 0$ ), so  $\mathcal{C}(\mathbf{A}) = \{c(1, 3) : c \in \mathbb{R}\}$ , a line in  $\mathbb{R}^2$ . The vector  $(1, 0)$  is not in  $\mathcal{C}(\mathbf{A})$ :  $c(1, 3)$  has second entry  $3c$ , which is 0 only if  $c = 0$ , and then the first entry is 0, not 1.

**Theorem 0.1** (The column space is the span of the columns). *If  $\mathbf{A}$  is an  $m \times n$  matrix with columns  $\tilde{a}_1, \dots, \tilde{a}_n$ , then*

$$\mathcal{C}(\mathbf{A}) = \text{span}\{\tilde{a}_1, \dots, \tilde{a}_n\},$$

so  $\mathcal{C}(\mathbf{A})$  is a subspace of  $\mathbb{R}^m$ , and the row space  $\mathcal{C}(\mathbf{A}^\top)$  is a subspace of  $\mathbb{R}^n$ .

### **i** Proof

*Proof.* By [Theorem 3 in Subspaces and Rank](#),  $\mathbf{A}\tilde{x} = x_1\tilde{a}_1 + \dots + x_n\tilde{a}_n$ , and as  $\tilde{x}$  ranges over  $\mathbb{R}^n$ , the coefficients  $x_1, \dots, x_n$  range over all lists of  $n$  numbers. So the set of all  $\mathbf{A}\tilde{x}$  ([Definition 0.1](#)) is the set of all linear combinations of the columns, which is their span ([Definition 5 in Subspaces and Rank](#)). A span is a subspace ([Theorem 2 in Subspaces and Rank](#)). Applying the same argument to  $\mathbf{A}^\top$ , whose columns are the rows of  $\mathbf{A}$ , shows the row space is a subspace of  $\mathbb{R}^n$ .  $\square$

**Example 0.2** (Reading off the column space and row space from the columns and rows). For  $\mathbf{A}$  in [Example 0.1](#), the columns are  $(1, 3)$ ,  $(-2, -6) = -2(1, 3)$  and  $(-2, -6) = -2(1, 3)$ , so by [Theorem 0.1](#)  $\mathcal{C}(\mathbf{A}) = \text{span}\{(1, 3)\}$ , the same line found in [Example 0.1](#). The rows are  $(1, -2, -2)$  and  $(3, -6, -6) = 3(1, -2, -2)$ , so the row space is  $\text{span}\{(1, -2, -2)\}$ , a line in  $\mathbb{R}^3$ .

**Definition 0.3** (Null space, kernel). The **null space** of an  $m \times n$  matrix  $\mathbf{A}$  is

$$\mathcal{N}(\mathbf{A}) \stackrel{\text{def}}{=} \{\tilde{x} \in \mathbb{R}^n : \mathbf{A}\tilde{x} = \tilde{0}_m\},$$

the set of [solutions](#) of the linear system  $\mathbf{A}\tilde{x} = \tilde{0}_m$ . It is also called the **kernel** of  $\mathbf{A}$ .

**Definition 0.4** (Left null space). The **left null space** of an  $m \times n$  matrix  $\mathbf{A}$  is  $\mathcal{N}(\mathbf{A}^\top)$ , a set of vectors in  $\mathbb{R}^m$ .

**Example 0.3** (The null space and left null space of a rank-one matrix). For  $\mathbf{A}$  in [Example 0.1](#),  $\mathbf{A}\tilde{x} = (x_1 - 2x_2 - 2x_3, 3(x_1 - 2x_2 - 2x_3))$ , so  $\mathbf{A}\tilde{x} = \tilde{0}_2$  exactly when  $x_1 = 2x_2 + 2x_3$ . Writing

$$\begin{aligned} \tilde{x} &= (2x_2 + 2x_3, x_2, x_3) \\ &= x_2(2, 1, 0) + x_3(2, 0, 1) \end{aligned}$$

shows  $\mathcal{N}(\mathbf{A}) = \text{span}\{(2, 1, 0), (2, 0, 1)\}$ , a plane in  $\mathbb{R}^3$ .

For the left null space,  $\mathbf{A}^\top y = (y_1 + 3y_2, -2(y_1 + 3y_2), -2(y_1 + 3y_2))$ , which is  $\tilde{0}_3$  exactly when  $y_1 = -3y_2$ , so  $\mathcal{N}(\mathbf{A}^\top) = \text{span}\{(-3, 1)\}$ , a line in  $\mathbb{R}^2$ .

**Theorem 0.2** (A null space is a subspace). *For any  $m \times n$  matrix  $\mathbf{A}$ ,  $\mathcal{N}(\mathbf{A})$  (Definition 0.3) is a subspace of  $\mathbb{R}^n$  (Definition 4 in Subspaces and Rank).*

**i** Proof

*Proof.*  $\mathbf{A}\mathbf{0}_n = \mathbf{0}_m$ , so  $\mathbf{0}_n \in \mathcal{N}(\mathbf{A})$ , and the null space is not empty. If  $\mathbf{u}, \mathbf{w} \in \mathcal{N}(\mathbf{A})$  and  $c \in \mathbb{R}$ , then

$$\begin{aligned} \mathbf{A}(\mathbf{u} + \mathbf{w}) &= \mathbf{A}\mathbf{u} + \mathbf{A}\mathbf{w} && \text{(Theorem 6 in Matrices)} \\ &= \mathbf{0}_m + \mathbf{0}_m && (\mathbf{u} \text{ and } \mathbf{w} \text{ are in the null space)} \\ &= \mathbf{0}_m && (\text{adding } \mathbf{0}_m \text{ changes nothing}) \end{aligned}$$

and, for each entry  $i = 1, \dots, m$ ,

$$\begin{aligned} (\mathbf{A}(c\mathbf{u}))_i &= a_{i1}(cu_1) + \dots + a_{in}(cu_n) && \text{(Definition 13 in Matrices)} \\ &= c(a_{i1}u_1 + \dots + a_{in}u_n) && (\text{factor } c \text{ out of each term}) \\ &= c(\mathbf{A}\mathbf{u})_i && \text{(Definition 13 in Matrices)} \\ &= c \cdot 0 && (\mathbf{u} \text{ is in the null space}) \\ &= 0, && (\text{arithmetic}) \end{aligned}$$

so  $\mathbf{A}(c\mathbf{u}) = \mathbf{0}_m$ . Therefore  $\mathbf{u} + \mathbf{w}$  and  $c\mathbf{u}$  are in  $\mathcal{N}(\mathbf{A})$ . □

**Example 0.4** (The solutions of  $\mathbf{A}\mathbf{x} = \mathbf{b}$  with  $\mathbf{b} \neq \mathbf{0}$  are not a subspace). In Example 0.3,  $(2, 1, 0)$  and  $(2, 0, 1)$  are in  $\mathcal{N}(\mathbf{A})$ , and so is their sum  $(4, 1, 1)$ , since  $4 - 2 \cdot 1 - 2 \cdot 1 = 0$ . By contrast, the solutions of  $\mathbf{A}\mathbf{x} = (1, 3)$  for the same  $\mathbf{A}$  are the vectors with  $x_1 - 2x_2 - 2x_3 = 1$ . That set does not contain  $\mathbf{0}_3$ , so by Theorem 1 in Subspaces and Rank it is not a subspace.

**Theorem 0.3** (Independent columns exactly when the null space is trivial). *For an  $m \times n$  matrix  $\mathbf{A}$ , the following are equivalent:*

1. The columns of  $\mathbf{A}$  are linearly independent (Definition 1 in Subspaces and Rank).
2.  $\mathbf{A}\mathbf{x} = \mathbf{0}_m$  only for  $\mathbf{x} = \mathbf{0}_n$ .
3.  $\mathcal{N}(\mathbf{A}) = \{\mathbf{0}_n\}$  (Definition 0.3).

So  $\mathbf{A}$  has full column rank (Definition 3 in Subspaces and Rank) exactly when its null space contains only the zero vector.

**i** Proof

*Proof.* Write  $\underline{a}_1, \dots, \underline{a}_n$  for the columns of  $\mathbf{A}$ . For  $\underline{x} = (x_1, \dots, x_n)$ , the matrix-vector product is the combination of the columns with weights  $x_1, \dots, x_n$ :

$$\mathbf{A}\underline{x} = x_1\underline{a}_1 + \dots + x_n\underline{a}_n.$$

So  $\mathbf{A}\underline{x} = \underline{0}_m$  says exactly that  $x_1\underline{a}_1 + \dots + x_n\underline{a}_n = \underline{0}_m$ . By [Definition 1 in Subspaces and Rank](#), the columns are linearly independent exactly when the only weights with this property are  $x_1 = \dots = x_n = 0$ . That is statement 2, so statements 1 and 2 are equivalent. Statement 2 says that the only vector in  $\mathcal{N}(\mathbf{A})$  is  $\underline{0}_n$ , which is statement 3. The last sentence is the definition of full column rank ([Definition 3 in Subspaces and Rank](#)).  $\square$

**Example 0.5** (Checking independence through the null space). For  $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$ ,

$$\mathbf{A}\underline{x} = \begin{bmatrix} x_1 + x_2 \\ x_1 + 2x_2 \\ x_1 + 3x_2 \end{bmatrix}.$$

This is  $\underline{0}_3$  only if  $x_2 = 0$  (subtract the first entry from the second), and then  $x_1 = 0$ . So  $\mathcal{N}(\mathbf{A}) = \{\underline{0}_2\}$  and the two columns are linearly independent.

For  $\mathbf{B} = \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 2 \end{bmatrix}$ ,  $\mathbf{B}(2, -1) = \underline{0}_3$ , so  $\mathcal{N}(\mathbf{B})$  contains a nonzero vector and the columns of  $\mathbf{B}$  are linearly dependent.

**Theorem 0.4** (The row space and null space share only the zero vector). *For any  $m \times n$  matrix  $\mathbf{A}$ ,*

$$\mathcal{C}(\mathbf{A}^\top) \cap \mathcal{N}(\mathbf{A}) = \{\underline{0}_n\}.$$

**i** Proof

*Proof.* Both sets are subspaces, by Theorem 0.1 and Theorem 0.2, so both contain  $\underline{0}_n$  ([Theorem 1 in Subspaces and Rank](#)).

Now take any  $\underline{x}$  in both sets. Being in the row space means  $\underline{x} = \mathbf{A}^\top \underline{u}$  for some  $\underline{u} \in \mathbb{R}^m$  ([Definition 0.1](#)), and being in the null space means  $\mathbf{A}\underline{x} = \underline{0}_m$  ([Definition 0.3](#)). Then

$$\begin{aligned}
\tilde{x}^\top \tilde{x} &= (\mathbf{A}^\top \tilde{u})^\top \tilde{x} && \text{(substitute } \tilde{x} = \mathbf{A}^\top \tilde{u} \text{ in the first factor)} \\
&= (\tilde{u}^\top (\mathbf{A}^\top)^\top) \tilde{x} && \text{(Theorem 17 in Matrices)} \\
&= (\tilde{u}^\top \mathbf{A}) \tilde{x} && \text{(transposing twice returns the original matrix)} \\
&= \tilde{u}^\top (\mathbf{A} \tilde{x}) && \text{(Theorem 5 in Matrices)} \\
&= \tilde{u}^\top \tilde{0}_m && \text{(substitute } \mathbf{A} \tilde{x} = \tilde{0}_m \text{)} \\
&= 0. && \text{(every term of the inner product is 0)}
\end{aligned}$$

But  $\tilde{x}^\top \tilde{x} = x_1^2 + \dots + x_n^2$ , which is 0 only when every  $x_i = 0$ . So  $\tilde{x} = \tilde{0}_n$ .  $\square$

**Example 0.6** (The row space and null space in Example 0.3). For  $\mathbf{A}$  in Example 0.1, the row space is  $\text{span}\{(1, -2, -2)\}$  and the null space is  $\text{span}\{(2, 1, 0), (2, 0, 1)\}$  (Example 0.3). A vector  $c(1, -2, -2)$  of the row space is in the null space only if

$$\begin{aligned}
\mathbf{A}(c(1, -2, -2)) &= c(1 - 2 \cdot (-2) - 2 \cdot (-2), 3(1 - 2 \cdot (-2) - 2 \cdot (-2))) && \text{(Example 0.3's formula for } \mathbf{A}\tilde{x}\text{)} \\
&= c(9, 27) && \text{(arithmetic)}
\end{aligned}$$

is  $\tilde{0}_2$ , that is, only if  $c = 0$ . So the two subspaces share only  $\tilde{0}_3$ .

**Definition 0.5** (Cross-product matrix). Let  $\mathbf{A}$  be an  $m \times n$  matrix with columns  $\tilde{a}_1, \dots, \tilde{a}_n$  and  $\mathbf{B}$  an  $m \times q$  matrix with columns  $\tilde{b}_1, \dots, \tilde{b}_q$ . The **cross-product matrix** of  $\mathbf{A}$  and  $\mathbf{B}$  is the  $n \times q$  matrix  $\mathbf{A}^\top \mathbf{B}$ , whose entry  $(i, j)$  is the dot product  $\tilde{a}_i \cdot \tilde{b}_j$ .

This is what R's `crossprod(A, B)` computes, and `crossprod(A)` is  $\mathbf{A}^\top \mathbf{A}$ . In NumPy it has no name of its own: write `A.T @ B`.

For two column vectors it is the  $1 \times 1$  matrix holding their **dot product**. It is unrelated to the cross product  $\tilde{u} \times \tilde{v}$  of two vectors in  $\mathbb{R}^3$ , and differs from the Frobenius inner product  $\text{tr}(\mathbf{A}^\top \mathbf{B})$ , which sums the diagonal of this matrix into a single number.

**Example 0.7** (A cross-product matrix). Let  $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}$  and  $\mathbf{B} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ . Then  $\mathbf{A}^\top \mathbf{B}$  is  $2 \times 1$ , and

$$\begin{aligned}
\mathbf{A}^\top \mathbf{B} &= \begin{bmatrix} \underline{a}_1 \cdot \underline{b}_1 \\ \underline{a}_2 \cdot \underline{b}_1 \end{bmatrix} && \text{(Definition 0.5)} \\
&= \begin{bmatrix} 1 + 2 + 0 \\ 0 + 2 + 3 \end{bmatrix} && \text{(multiply and add)} \\
&= \begin{bmatrix} 3 \\ 5 \end{bmatrix}. && \text{(add)}
\end{aligned}$$

**Definition 0.6** (Gram matrix). The **Gram matrix** [2] of an  $m \times n$  matrix  $\mathbf{A}$  is the  $n \times n$  matrix  $\mathbf{A}^\top \mathbf{A}$ , the **cross-product matrix** of  $\mathbf{A}$  with itself.

**Example 0.8** (A Gram matrix and its entries). Let  $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}$ , with columns  $\underline{a}_1 = (1, 1, 0)$  and  $\underline{a}_2 = (0, 1, 1)$ . Then

$$\begin{aligned}
\mathbf{A}^\top \mathbf{A} &= \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} && \text{(Definition 3 in Matrices)} \\
&= \begin{bmatrix} 1 \cdot 1 + 1 \cdot 1 + 0 \cdot 0 & 1 \cdot 0 + 1 \cdot 1 + 0 \cdot 1 \\ 0 \cdot 1 + 1 \cdot 1 + 1 \cdot 0 & 0 \cdot 0 + 1 \cdot 1 + 1 \cdot 1 \end{bmatrix} && \text{(Definition 7 in Matrices)} \\
&= \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}. && \text{(multiply and add)}
\end{aligned}$$

Row  $i$  of  $\mathbf{A}^\top$  is column  $i$  of  $\mathbf{A}$ , so entry  $(i, j)$  of the Gram matrix is the dot product  $\underline{a}_i \cdot \underline{a}_j$ : for instance,

$$\begin{aligned}
\underline{a}_1 \cdot \underline{a}_2 &= 0 + 1 + 0 \\
&= 1
\end{aligned}$$

is entry  $(1, 2)$ .

**Theorem 0.5** ( $\mathbf{A}^\top \mathbf{A}$  has the same null space as  $\mathbf{A}$ ). For any  $m \times n$  matrix  $\mathbf{A}$ ,

$$\mathcal{N}(\mathbf{A}^\top \mathbf{A}) = \mathcal{N}(\mathbf{A}).$$

**i** Proof

*Proof.*  $\mathcal{N}(\mathbf{A}) \subseteq \mathcal{N}(\mathbf{A}^\top \mathbf{A})$ . If  $\mathbf{A}\underline{x} = \underline{0}_m$ , then

$$\begin{aligned}\mathbf{A}^\top \mathbf{A} \tilde{x} &= \mathbf{A}^\top \tilde{0}_m \\ &= \tilde{0}_n.\end{aligned}$$

$\mathcal{N}(\mathbf{A}^\top \mathbf{A}) \subseteq \mathcal{N}(\mathbf{A})$ . If  $\mathbf{A}^\top \mathbf{A} \tilde{x} = \tilde{0}_n$ , then

$$\begin{aligned}0 &= \tilde{x}^\top \tilde{0}_n && \text{(every term of the inner product is 0)} \\ &= \tilde{x}^\top (\mathbf{A}^\top \mathbf{A} \tilde{x}) && \text{(substitute } \tilde{0}_n = \mathbf{A}^\top \mathbf{A} \tilde{x}) \\ &= (\tilde{x}^\top \mathbf{A}^\top) (\mathbf{A} \tilde{x}) && \text{(Theorem 5 in Matrices)} \\ &= (\mathbf{A} \tilde{x})^\top (\mathbf{A} \tilde{x}) && \text{(Theorem 17 in Matrices)} \\ &= (\mathbf{A} \tilde{x}) \cdot (\mathbf{A} \tilde{x}) && \text{(Example 6 in Matrices)} \\ &= \|\mathbf{A} \tilde{x}\|^2. && \text{(square both sides of Equation 2 in Vectors)}\end{aligned}$$

A vector whose Euclidean norm is 0 has every entry 0, so  $\mathbf{A} \tilde{x} = \tilde{0}_m$ . □

**Example 0.9** (The null space of a Gram matrix). For  $\mathbf{A}$  in Example 0.1,

$$\begin{aligned}\mathbf{A}^\top \mathbf{A} &= \begin{bmatrix} 1 & 3 \\ -2 & -6 \\ -2 & -6 \end{bmatrix} \begin{bmatrix} 1 & -2 & -2 \\ 3 & -6 & -6 \end{bmatrix} \\ &= \begin{bmatrix} 10 & -20 & -20 \\ -20 & 40 & 40 \\ -20 & 40 & 40 \end{bmatrix}.\end{aligned}$$

Every row is a multiple of  $(1, -2, -2)$ , so  $\mathbf{A}^\top \mathbf{A} \tilde{x} = \tilde{0}_3$  exactly when  $x_1 - 2x_2 - 2x_3 = 0$ , the same condition as for  $\mathcal{N}(\mathbf{A})$  in Example 0.3.

## Rank-nullity and rank factorization

### Source

This section is adapted from Zhou [3], used under the MIT License (see the license text in [Section 2 in Subspaces and Rank](#)). Theorem 0.10 and Corollary 0.2 restore the source's statements that a product has rank at most the rank of either factor, and that multiplying by a matrix of full column rank or by a matrix whose rank is its number of rows does not change the rank. These parts of the source are not part of this section:

- the names “full row rank”, “full rank” and “singular”
- the statement that  $\mathcal{C}(\mathbf{C}) \subseteq \mathcal{C}(\mathbf{A})$  exactly when  $\mathbf{C} = \mathbf{A}\mathbf{B}$  for some  $\mathbf{B}$ , and its row space version, and the containment  $\mathcal{N}(\mathbf{B}) \subseteq \mathcal{N}(\mathbf{A}\mathbf{B})$
- the dimension of a subspace as the largest number of linearly independent vectors

in it, or the smallest number of vectors that span it

- the properties of a rank factorization  $\mathbf{A} = \mathbf{C}\mathbf{R}$  ( $\text{rank}(\mathbf{C}) = \text{rank}(\mathbf{R}) = r$ ,  $\mathcal{C}(\mathbf{A}) = \mathcal{C}(\mathbf{C})$ ,  $\mathcal{N}(\mathbf{A}) = \mathcal{N}(\mathbf{R})$ ), the fact that it is not unique, and the second proof of  $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}^\top)$  that it gives
- calling  $\mathbf{A}\mathbf{A}^\top$  a Gram matrix, and the equation  $\mathcal{N}(\mathbf{A}\mathbf{A}^\top) = \mathcal{N}(\mathbf{A}^\top)$  (Theorem 0.5 applied to  $\mathbf{A}^\top$  gives it)
- the statement that  $\mathcal{C}(\mathbf{A})$  and  $\mathcal{N}(\mathbf{A}^\top)$  share only the zero vector (it follows from Theorem 10 in Direct Sums and Orthogonal Complements)

**Theorem 0.6** (The rank is the dimension of the column space). *For any  $m \times n$  matrix  $\mathbf{A}$ ,*

$$\text{rank}(\mathbf{A}) = \dim(\mathcal{C}(\mathbf{A})).$$

### Proof

*Proof.* Let  $r = \text{rank}(\mathbf{A})$  (Definition 2 in Subspaces and Rank), and choose  $r$  linearly independent columns of  $\mathbf{A}$ ; by Definition 2 in Subspaces and Rank, no larger set of columns is linearly independent. We show the chosen columns are a basis of  $\mathcal{C}(\mathbf{A})$  (Definition 6 in Subspaces and Rank).

**Every column is in the span of the chosen columns.** A chosen column is in that span trivially. If some unchosen column  $\tilde{a}_j$  were not in the span, then adding it to the chosen columns would give  $r + 1$  linearly independent columns (Theorem 8 in Subspaces and Rank), more than  $r$ . So every column of  $\mathbf{A}$  is in the span of the chosen columns.

**The chosen columns span  $\mathcal{C}(\mathbf{A})$ .** The span of the chosen columns is a subspace (Theorem 2 in Subspaces and Rank), so it is closed under addition and scalar multiplication, and it contains every linear combination of the columns of  $\mathbf{A}$ . By Theorem 0.1, those combinations make up  $\mathcal{C}(\mathbf{A})$ , so  $\mathcal{C}(\mathbf{A})$  is contained in the span of the chosen columns. The chosen columns are themselves columns of  $\mathbf{A}$ , so their span is contained in  $\mathcal{C}(\mathbf{A})$ , and the two sets are equal.

The chosen columns are linearly independent and span  $\mathcal{C}(\mathbf{A})$ , so they are a basis of it, and  $\dim(\mathcal{C}(\mathbf{A})) = r$  (Definition 9 in Subspaces and Rank). If  $r = 0$ , every column is  $\tilde{0}_m$ ,  $\mathcal{C}(\mathbf{A}) = \{\tilde{0}_m\}$ , and both sides are 0.  $\square$

**Example 0.10** (The rank of the matrix in Example 0.1). For  $\mathbf{A}$  in Example 0.1,  $\mathcal{C}(\mathbf{A}) = \text{span}\{(1, 3)\}$  (Example 0.2), and  $(1, 3)$  is linearly independent because it is nonzero, so it is a basis and  $\mathcal{C}(\mathbf{A})$  has dimension 1. So  $\text{rank}(\mathbf{A}) = 1$ , by Theorem 0.6. Directly: the first column  $(1, 3)$  is linearly independent on its own, and any two columns are dependent, because each column is a multiple of  $(1, 3)$ .

**Definition 0.7** (Nullity). The **nullity** [4] of a matrix  $\mathbf{A}$  is the dimension of its null space (Definition 0.3):

$$\text{nullity}(\mathbf{A}) \stackrel{\text{def}}{=} \dim(\mathcal{N}(\mathbf{A})).$$

**Example 0.11** (The nullity of the matrix in Example 0.1). For  $\mathbf{A}$  in Example 0.1,  $\mathcal{N}(\mathbf{A}) = \text{span}\{(2, 1, 0), (2, 0, 1)\}$  (Example 0.3). Those two vectors are linearly independent:  $c_1(2, 1, 0) + c_2(2, 0, 1) = (2c_1 + 2c_2, c_1, c_2)$ , which is  $\mathbf{0}_3$  only if

$$\begin{aligned} c_1 &= c_2 \\ &= 0. \end{aligned}$$

So they are a basis of  $\mathcal{N}(\mathbf{A})$ , and  $\text{nullity}(\mathbf{A}) = 2$ . The  $2 \times 2$  identity matrix has nullity 0, because  $\mathbf{I}\tilde{x} = \tilde{x}$  is  $\mathbf{0}_2$  only when  $\tilde{x} = \mathbf{0}_2$ .

**Theorem 0.7** (Rank-nullity theorem). For any  $m \times n$  matrix  $\mathbf{A}$ ,

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

### **i** Proof

*Proof.* Let  $\nu = \text{nullity}(\mathbf{A})$ . Choose a basis  $\tilde{x}_1, \dots, \tilde{x}_\nu$  of  $\mathcal{N}(\mathbf{A})$ , and extend it to a basis  $\tilde{x}_1, \dots, \tilde{x}_\nu, \tilde{y}_1, \dots, \tilde{y}_{n-\nu}$  of  $\mathbb{R}^n$  (Theorem 9 in Subspaces and Rank); it has  $n$  vectors, because every basis of  $\mathbb{R}^n$  has  $\dim(\mathbb{R}^n) = n$  vectors (Theorem 7 in Subspaces and Rank, Example 16 in Subspaces and Rank). We show that  $\mathbf{A}\tilde{y}_1, \dots, \mathbf{A}\tilde{y}_{n-\nu}$  are a basis of  $\mathcal{C}(\mathbf{A})$ , using Theorem 4 in Subspaces and Rank to apply  $\mathbf{A}$  to linear combinations.

**They are linearly independent.** Suppose  $\sum_{i=1}^{n-\nu} v_i \mathbf{A}\tilde{y}_i = \mathbf{0}_m$ . Then  $\mathbf{A}(\sum_{i=1}^{n-\nu} v_i \tilde{y}_i) = \mathbf{0}_m$  (Theorem 4 in Subspaces and Rank), so  $\sum_{i=1}^{n-\nu} v_i \tilde{y}_i \in \mathcal{N}(\mathbf{A})$ , and it equals  $\sum_{j=1}^\nu u_j \tilde{x}_j$  for some numbers  $u_j$ , since the  $\tilde{x}_j$  span  $\mathcal{N}(\mathbf{A})$ . Moving everything to one side,

$$\sum_{j=1}^\nu (-u_j) \tilde{x}_j + \sum_{i=1}^{n-\nu} v_i \tilde{y}_i = \mathbf{0}_n,$$

and because the  $\tilde{x}$ 's and  $\tilde{y}$ 's together form a basis of  $\mathbb{R}^n$ , which is linearly independent, every coefficient is 0; in particular every  $v_i = 0$ .

**They span  $\mathcal{C}(\mathbf{A})$ .** Take any  $\tilde{w} \in \mathcal{C}(\mathbf{A})$ , so  $\tilde{w} = \mathbf{A}\tilde{z}$  for some  $\tilde{z} \in \mathbb{R}^n$  (Definition 0.1). Write  $\tilde{z} = \sum_j a_j \tilde{x}_j + \sum_i b_i \tilde{y}_i$  in the basis of  $\mathbb{R}^n$ . Then

$$\begin{aligned}
w_{\sim} &= \mathbf{A} \left( \sum_{j=1}^{\nu} a_j x_j + \sum_{i=1}^{n-\nu} b_i y_i \right) && \text{(substitute } z) \\
&= \sum_{j=1}^{\nu} a_j \mathbf{A} x_j + \sum_{i=1}^{n-\nu} b_i \mathbf{A} y_i && \text{(Theorem 4 in Subspaces and Rank)} \\
&= \sum_{j=1}^{\nu} a_j \mathbf{0}_m + \sum_{i=1}^{n-\nu} b_i \mathbf{A} y_i && \text{(each } x_j \in \mathcal{N}(\mathbf{A})) \\
&= \sum_{i=1}^{n-\nu} b_i \mathbf{A} y_i, && \text{(drop the zero terms)}
\end{aligned}$$

a linear combination of  $\mathbf{A} y_1, \dots, \mathbf{A} y_{n-\nu}$ . These vectors are in  $\mathcal{C}(\mathbf{A})$ , so their span is exactly  $\mathcal{C}(\mathbf{A})$ .

So  $\dim(\mathcal{C}(\mathbf{A})) = n - \nu$ , and  $\text{rank}(\mathbf{A}) = n - \nu$  by Theorem 0.6.  $\square$

**Example 0.12** (Checking rank-nullity on the matrix in Example 0.1). The matrix  $\mathbf{A}$  in Example 0.1 has  $n = 3$  columns, rank 1 (Example 0.10) and nullity 2 (Example 0.11), and  $1 + 2 = 3$ . The  $2 \times 2$  identity matrix has rank 2 and nullity 0 (Example 0.11), and  $2 + 0 = 2$ .

**Theorem 0.8** (Multiplying on the right cannot increase the rank). *If  $\mathbf{A}$  is  $m \times n$  and  $\mathbf{B}$  is  $n \times k$ , then*

$$\text{rank}(\mathbf{AB}) \leq \text{rank}(\mathbf{A}).$$

### **i** Proof

*Proof.*  $\mathcal{C}(\mathbf{AB}) \subseteq \mathcal{C}(\mathbf{A})$ . Any vector in  $\mathcal{C}(\mathbf{AB})$  is  $(\mathbf{AB}) x_{\sim}$  for some  $x \in \mathbb{R}^k$  (Definition 0.1), and

$$(\mathbf{AB}) x_{\sim} = \mathbf{A} (\mathbf{B} x_{\sim}) \quad \text{(Theorem 5 in Matrices)}$$

is  $\mathbf{A}$  times the vector  $\mathbf{B} x_{\sim} \in \mathbb{R}^n$ , so it is in  $\mathcal{C}(\mathbf{A})$ .

**Compare dimensions.** Both column spaces are subspaces (Theorem 0.1), so

$$\begin{aligned}
\text{rank}(\mathbf{AB}) &= \dim(\mathcal{C}(\mathbf{AB})) && \text{(Theorem 0.6)} \\
&\leq \dim(\mathcal{C}(\mathbf{A})) && \text{(Theorem 10 in Subspaces and Rank, part 2)} \\
&= \text{rank}(\mathbf{A}). && \text{(Theorem 0.6)}
\end{aligned}$$

$\square$

**Example 0.13** (A product can lose rank). Let

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad \mathbf{AB} = \begin{bmatrix} 1 \cdot 0 + 0 \cdot 1 & 1 \cdot 0 + 0 \cdot 0 \\ 0 \cdot 0 + 0 \cdot 1 & 0 \cdot 0 + 0 \cdot 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

$\mathbf{A}$  has rank 1, but  $\mathbf{AB}$  has rank 0, so the inequality in Theorem 0.8 can be strict. With the  $2 \times 2$  identity in place of  $\mathbf{B}$ ,  $\mathbf{AI} = \mathbf{A}$  keeps rank 1, and the inequality is an equality.

**Theorem 0.9** (Fundamental theorem of ranks). *For any  $m \times n$  matrix  $\mathbf{A}$ ,*

$$\begin{aligned} \text{rank}(\mathbf{A}) &= \text{rank}(\mathbf{A}^\top) \\ &= \text{rank}(\mathbf{A}^\top \mathbf{A}) \\ &= \text{rank}(\mathbf{AA}^\top). \end{aligned}$$

*In particular, the largest number of linearly independent rows of  $\mathbf{A}$  equals the largest number of linearly independent columns.*

#### **i** Proof

*Proof.*  $\text{rank}(\mathbf{A}^\top \mathbf{A}) = \text{rank}(\mathbf{A})$ . Both matrices have  $n$  columns, and  $\mathcal{N}(\mathbf{A}^\top \mathbf{A}) = \mathcal{N}(\mathbf{A})$  (Theorem 0.5), so they have the same nullity (Definition 0.7). Then

$$\begin{aligned} \text{rank}(\mathbf{A}^\top \mathbf{A}) &= n - \text{nullity}(\mathbf{A}^\top \mathbf{A}) \quad (\text{Theorem 0.7}) \\ &= n - \text{nullity}(\mathbf{A}) \quad (\text{equal null spaces}) \\ &= \text{rank}(\mathbf{A}). \quad (\text{Theorem 0.7}) \end{aligned}$$

$\text{rank}(\mathbf{AA}^\top) = \text{rank}(\mathbf{A}^\top)$ . Apply the previous step to the  $n \times m$  matrix  $\mathbf{A}^\top$ , using  $(\mathbf{A}^\top)^\top = \mathbf{A}$ .

**Chain the inequalities.**

$$\begin{aligned} \text{rank}(\mathbf{A}) &= \text{rank}(\mathbf{A}^\top \mathbf{A}) \quad (\text{first step}) \\ &\leq \text{rank}(\mathbf{A}^\top) \quad (\text{Theorem 0.8, with } \mathbf{A}^\top \text{ on the left}) \\ &= \text{rank}(\mathbf{AA}^\top) \quad (\text{second step}) \\ &\leq \text{rank}(\mathbf{A}). \quad (\text{Theorem 0.8, with } \mathbf{A} \text{ on the left}) \end{aligned}$$

The chain starts and ends at  $\text{rank}(\mathbf{A})$ , so every quantity in it equals  $\text{rank}(\mathbf{A})$ . The columns of  $\mathbf{A}^\top$  are the rows of  $\mathbf{A}$ , which gives the statement about rows.  $\square$

**Example 0.14** (Four equal ranks). For  $\mathbf{A}$  in Example 0.1,  $\text{rank}(\mathbf{A}) = 1$  (Example 0.10).

Its transpose  $\mathbf{A}^\top = \begin{bmatrix} 1 & 3 \\ -2 & -6 \\ -2 & -6 \end{bmatrix}$  has second column 3 times its first, so its rank is 1.  $\mathbf{A}^\top \mathbf{A}$  in Example 0.9 has every row a multiple of  $(1, -2, -2)$ , so its rank is 1 too, and

$$\begin{aligned} \mathbf{A}\mathbf{A}^\top &= \begin{bmatrix} 1 \cdot 1 + (-2)(-2) + (-2)(-2) & 1 \cdot 3 + (-2)(-6) + (-2)(-6) \\ 3 \cdot 1 + (-6)(-2) + (-6)(-2) & 3 \cdot 3 + (-6)(-6) + (-6)(-6) \end{bmatrix} \\ &= \begin{bmatrix} 9 & 27 \\ 27 & 81 \end{bmatrix}, \end{aligned}$$

whose second column is 3 times its first, so its rank is 1.

**Corollary 0.1** (The rank is at most the smaller dimension). *For any  $m \times n$  matrix  $\mathbf{A}$ ,  $\text{rank}(\mathbf{A}) \leq \min\{m, n\}$ .*

#### **i** Proof

*Proof.*  $\mathbf{A}$  has  $n$  columns, so at most  $n$  of them are linearly independent, and  $\text{rank}(\mathbf{A}) \leq n$  (Definition 2 in Subspaces and Rank). Likewise  $\mathbf{A}^\top$  has  $m$  columns, so  $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}^\top) \leq m$  (Theorem 0.9).  $\square$

**Example 0.15** (A  $2 \times 3$  matrix has rank at most 2). By Corollary 0.1, every  $2 \times 3$  matrix has rank at most  $\min\{2, 3\} = 2$ , even though it has 3 columns. The matrix  $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$  of Remark 1 in Subspaces and Rank reaches that bound: its first two columns  $(1, 0)$  and  $(0, 1)$  are linearly independent, so its rank is 2.

**Theorem 0.10** (Multiplying on the left cannot increase the rank). *If  $\mathbf{A}$  is  $m \times n$  and  $\mathbf{B}$  is  $n \times k$ , then*

$$\text{rank}(\mathbf{AB}) \leq \text{rank}(\mathbf{B}).$$

#### **i** Proof

*Proof.*

$$\begin{aligned} \text{rank}(\mathbf{AB}) &= \text{rank}((\mathbf{AB})^\top) \quad (\text{Theorem 0.9}) \\ &= \text{rank}(\mathbf{B}^\top \mathbf{A}^\top) \quad (\text{Theorem 17 in Matrices}) \\ &\leq \text{rank}(\mathbf{B}^\top) \quad (\text{Theorem 0.8, with } \mathbf{B}^\top \text{ on the left}) \\ &= \text{rank}(\mathbf{B}). \quad (\text{Theorem 0.9}) \end{aligned}$$

□

**Example 0.16** (A product of rank 1 with a factor of rank 2). Let

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \quad \mathbf{AB} = \begin{bmatrix} 1 \cdot 1 + 1 \cdot 1 & 1 \cdot 0 + 1 \cdot 0 \\ 0 \cdot 1 + 1 \cdot 1 & 0 \cdot 0 + 1 \cdot 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 1 & 0 \end{bmatrix}.$$

$\mathbf{A}$  has rank 2, because its columns  $(1, 0)$  and  $(1, 1)$  are linearly independent.  $\mathbf{B}$  has rank 1, because its second column is  $\mathbf{0}_2$  and its first is not, and  $\mathbf{AB}$  has rank 1 for the same reason. So  $\text{rank}(\mathbf{AB}) = 1 \leq \text{rank}(\mathbf{B}) = 1$  (Theorem 0.10), and the bound of Theorem 0.8, 2, is not the tighter one.

**Corollary 0.2** (Multiplying by a matrix of full rank does not change the rank). Let  $\mathbf{A}$  be  $m \times n$  and  $\mathbf{B}$  be  $n \times k$ .

1. If  $\mathbf{A}$  has full column rank (Definition 3 in Subspaces and Rank), then  $\text{rank}(\mathbf{AB}) = \text{rank}(\mathbf{B})$ .
2. If  $\text{rank}(\mathbf{B}) = n$ , then  $\text{rank}(\mathbf{AB}) = \text{rank}(\mathbf{A})$ .

### **i** Proof

*Proof. Part 1.* Because  $\mathbf{A}$  has full column rank,  $\mathbf{A}\tilde{y} = \mathbf{0}_m$  only for  $\tilde{y} = \mathbf{0}_n$  (Theorem 0.3). So for  $\tilde{x} \in \mathbb{R}^k$ ,

$$\mathbf{AB}\tilde{x} = \mathbf{0}_m \iff \mathbf{B}\tilde{x} = \mathbf{0}_n \quad (\text{Theorem 0.3, with } \tilde{y} = \mathbf{B}\tilde{x})$$

so  $\mathcal{N}(\mathbf{AB}) = \mathcal{N}(\mathbf{B})$  (Definition 0.3), and the two matrices have the same nullity (Definition 0.7). Both have  $k$  columns, so by Theorem 0.7 they have the same rank.

**Part 2.**  $\mathcal{C}(\mathbf{B})$  is a subspace of  $\mathbb{R}^n$  (Theorem 0.1) of dimension  $n$  (Theorem 0.6), so  $\mathcal{C}(\mathbf{B}) = \mathbb{R}^n$  (Theorem 11 in Subspaces and Rank). Every vector of  $\mathcal{C}(\mathbf{AB})$  is  $\mathbf{A}(\mathbf{B}\tilde{x})$  for some  $\tilde{x}$  (Definition 0.1), and  $\mathbf{B}\tilde{x}$  takes every vector of  $\mathbb{R}^n$ , so  $\mathcal{C}(\mathbf{AB}) = \mathcal{C}(\mathbf{A})$  and the two matrices have the same rank (Theorem 0.6). □

**Example 0.17** (Ranks of products with a full-rank factor). Let  $\mathbf{B} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$ , which has rank 1.

- Part 1:  $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$  has two linearly independent columns, so it has full column

rank, and  $\mathbf{AB} = \begin{bmatrix} 2 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}$  has rank 1 = rank( $\mathbf{B}$ ).

- Part 2: let  $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ , which has rank 1, and let  $\mathbf{B} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ , which has rank 2.

Then  $\mathbf{AB} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$  has rank 1 = rank( $\mathbf{A}$ ).

**Theorem 0.11** (A centered data matrix has rank at most  $n-1$ ). *Let  $\mathbf{X}$  be an  $n \times p$  matrix with rows  $\tilde{x}_1^\top, \dots, \tilde{x}_n^\top$ , let  $\bar{\tilde{x}} = \frac{1}{n} \sum_{i=1}^n \tilde{x}_i$  be the mean of the rows, and let  $\mathbf{X}_c = \mathbf{X} - \tilde{1} \bar{\tilde{x}}^\top$  be the matrix whose row  $i$  is  $(\tilde{x}_i - \bar{\tilde{x}})^\top$ , where  $\tilde{1}$  is the ones vector of length  $n$  (Definition 16 in Vectors). Then  $\text{rank}(\mathbf{X}_c) \leq n-1$ .*

### **i** Proof

*Proof.* The entries of each column of  $\mathbf{X}_c$  sum to zero:

$$\begin{aligned} \tilde{1}^\top \mathbf{X}_c &= \tilde{1}^\top \mathbf{X} - (\tilde{1}^\top \tilde{1}) \bar{\tilde{x}}^\top \quad (\text{Theorem 6 in Matrices}) \\ &= \sum_{i=1}^n \tilde{x}_i^\top - n \bar{\tilde{x}}^\top \quad (\tilde{1}^\top \tilde{1} = n) \\ &= n \bar{\tilde{x}}^\top - n \bar{\tilde{x}}^\top \quad (\text{definition of } \bar{\tilde{x}}) \\ &= \mathbf{0}_p^\top. \end{aligned}$$

Transposing,  $\mathbf{X}_c^\top \tilde{1} = \mathbf{0}_p$ . The ones vector is nonzero, so it is in the null space of the  $p \times n$  matrix  $\mathbf{X}_c^\top$ , and  $\text{nullity}(\mathbf{X}_c^\top) \geq 1$ . By Theorem 0.7, applied to the  $n$  columns of  $\mathbf{X}_c^\top$ ,

$$\begin{aligned} \text{rank}(\mathbf{X}_c) &= \text{rank}(\mathbf{X}_c^\top) \quad (\text{Theorem 0.9}) \\ &= n - \text{nullity}(\mathbf{X}_c^\top) \quad (\text{Theorem 0.7}) \\ &\leq n - 1. \quad (\text{nullity}(\mathbf{X}_c^\top) \geq 1) \end{aligned}$$

□

**Example 0.18** (Two points in three dimensions).  $\mathbf{X} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix}$  has  $\bar{\tilde{x}} = (2, 2, 2)$ , so

$$\mathbf{X}_c = \begin{bmatrix} -1 & 0 & 1 \\ 1 & 0 & -1 \end{bmatrix}.$$

The second row is  $-1$  times the first, so  $\text{rank}(\mathbf{X}_c) = 1 = n - 1$ , and the column sums are zero:  $\tilde{1}^\top \mathbf{X}_c = \mathbf{0}_3^\top$ .

**Corollary 0.3** (A sample covariance matrix from  $n \leq p$  points is singular). *With  $\mathbf{X}_c$  as in Theorem 0.11, the sample covariance matrix*

$$\mathbf{S} = \frac{1}{n-1} \mathbf{X}_c^\top \mathbf{X}_c$$

*is  $p \times p$  with  $\text{rank}(\mathbf{S}) \leq n-1$ . So if  $n \leq p$ , then  $\mathbf{S}$  is not invertible.*

#### **i** Proof

*Proof.*

$$\begin{aligned} \text{rank}(\mathbf{S}) &= \text{rank}(\mathbf{X}_c^\top \mathbf{X}_c) \quad (\text{multiplying by } \frac{1}{n-1} \neq 0 \text{ does not change the rank}) \\ &= \text{rank}(\mathbf{X}_c) \quad (\text{Theorem 0.9}) \\ &\leq n-1. \quad (\text{Theorem 0.11}) \end{aligned}$$

If  $n \leq p$ , then  $\text{rank}(\mathbf{S}) \leq n-1 < p$ , so  $\mathbf{S}$  does not have full column rank, and it is not invertible by [Theorem 1 in Diagonalization and Cholesky](#).  $\square$

**Example 0.19** (As many points as dimensions). The  $n = 3$  points  $(1, 0, 0)$ ,  $(0, 1, 0)$  and  $(0, 0, 1)$  in  $p = 3$  dimensions have  $\tilde{x} = \frac{1}{3}(1, 1, 1)$ , so

$$\mathbf{X}_c = \mathbf{I}_3 - \frac{1}{3} \tilde{x} \tilde{x}^\top, \quad \mathbf{S} = \frac{1}{2} \mathbf{X}_c = \frac{1}{6} \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix},$$

using  $\mathbf{X}_c^\top \mathbf{X}_c = \mathbf{X}_c$  (the centering matrix is a symmetric projection). Then  $\mathbf{S} \tilde{x} = 0_3$  and  $\text{rank}(\mathbf{S}) = 2 = n-1 < 3$ . With one more point than dimensions ( $n = p+1$ ), the rank bound allows  $\mathbf{S}$  to have full rank  $p$ .

**Definition 0.8** (Affine subspace). An **affine subspace** [5] of  $\mathbb{R}^p$  is a set of the form

$$\{\tilde{x}_0 + \tilde{s} : \tilde{s} \in \mathcal{S}\}$$

for some vector  $\tilde{x}_0 \in \mathbb{R}^p$  and some subspace  $\mathcal{S}$  of  $\mathbb{R}^p$  ([Definition 4 in Subspaces and Rank](#)): the **image** of  $\mathcal{S}$  under the translation by  $\tilde{x}_0$  ([Definition 18 in Matrices](#)).

**Example 0.20** (A line that misses the origin is an affine subspace). The line  $\mathcal{T} = \{(x, 1) : x \in \mathbb{R}\}$  of [Example 4 in Subspaces and Rank](#) is not a subspace, but it is an affine subspace, with  $\tilde{x}_0 = (0, 1)$  and  $\mathcal{S} = \text{span}\{(1, 0)\}$  ([Definition 5 in Subspaces and Rank](#)):

$$\begin{aligned}(0, 1) + c(1, 0) &= (0 + c, 1 + 0) \quad (\text{Definition 6 in Matrices, Definition 7 in Vectors}) \\ &= (c, 1), \quad (\text{add})\end{aligned}$$

and as  $c$  ranges over  $\mathbb{R}$ ,  $(c, 1)$  ranges over all of  $\mathcal{T}$ . Every subspace  $\mathcal{S}$  is also an affine subspace, with  $\tilde{x}_0 = \tilde{0}$ .

**Theorem 0.12** (A hyperplane is a translated subspace of dimension  $p - 1$ ). *Let  $\tilde{w} \in \mathbb{R}^p$  be nonzero, let  $b \in \mathbb{R}$ , let  $\mathcal{H} = \{\tilde{x} \in \mathbb{R}^p : \tilde{w}^\top \tilde{x} + b = 0\}$  be the hyperplane with normal vector  $\tilde{w}$  and offset  $b$  (Definition 22 in Matrices), and let  $\mathcal{H}_0 = \{\tilde{x} \in \mathbb{R}^p : \tilde{w}^\top \tilde{x} = 0\}$  be the hyperplane with the same normal vector and offset 0.*

1.  $\mathcal{H}_0$  is a subspace of  $\mathbb{R}^p$  (Definition 4 in Subspaces and Rank) of dimension  $p - 1$  (Definition 9 in Subspaces and Rank).
2.  $\mathcal{H}$  is not empty, and for any  $\tilde{x}_0 \in \mathcal{H}$ ,  $\mathcal{H} = \{\tilde{x}_0 + \tilde{s} : \tilde{s} \in \mathcal{H}_0\}$ , so  $\mathcal{H}$  is an affine subspace (Definition 0.8).
3. For any  $\tilde{x}, \tilde{y} \in \mathcal{H}$ , the normal vector is orthogonal to their difference:  $\tilde{w} \perp (\tilde{x} - \tilde{y})$  (Definition 19 in Vectors).

### **i** Proof

*Proof.* Here  $\tilde{w}^\top$  is a  $1 \times p$  matrix, and  $\tilde{w}^\top \tilde{x} = \tilde{w} \cdot \tilde{x}$  (Example 6 in Matrices).

**Part 1.**  $\mathcal{H}_0$  is the null space  $\mathcal{N}(\tilde{w}^\top)$  (Definition 0.3), so it is a subspace (Theorem 0.2). The  $p$  columns of  $\tilde{w}^\top$  are the numbers  $w_1, \dots, w_p$ . Some  $w_j \neq 0$ , because  $\tilde{w} \neq \tilde{0}$ , and that column on its own is linearly independent ( $c w_j = 0$  forces  $c = 0$ ), so  $\text{rank}(\tilde{w}^\top) \geq 1$  (Definition 2 in Subspaces and Rank); and  $\text{rank}(\tilde{w}^\top) \leq \min\{1, p\} = 1$  (Corollary 0.1). So

$$\begin{aligned}\dim(\mathcal{H}_0) &= \text{nullity}(\tilde{w}^\top) \quad (\text{Definition 0.7}) \\ &= p - \text{rank}(\tilde{w}^\top) \quad (\text{Theorem 0.7, for the } p \text{ columns of } \tilde{w}^\top) \\ &= p - 1. \quad (\text{the rank is 1})\end{aligned}$$

**Part 2.**  $\tilde{w} \cdot \tilde{w} = w_1^2 + \dots + w_p^2 > 0$ , because some  $w_j \neq 0$ . So  $\tilde{x}^* \stackrel{\text{def}}{=} -\frac{b}{\tilde{w} \cdot \tilde{w}} \tilde{w}$  is defined, and

$$\begin{aligned}\tilde{w}^\top \tilde{x}^* + b &= -\frac{b}{\tilde{w} \cdot \tilde{w}} (\tilde{w}^\top \tilde{w}) + b \quad (\text{homogeneity of } \tilde{x} \mapsto \tilde{w}^\top \tilde{x}, \text{ Theorem 10 in Matrices}) \\ &= -b + b \quad (\tilde{w}^\top \tilde{w} = \tilde{w} \cdot \tilde{w}) \\ &= 0, \quad (\text{arithmetic})\end{aligned}$$

so  $\tilde{x}^* \in \mathcal{H}$ , and  $\mathcal{H}$  is not empty. Now take any  $\tilde{x}_0 \in \mathcal{H}$ , so  $\tilde{w}^\top \tilde{x}_0 = -b$ . For any  $\tilde{x} \in \mathbb{R}^p$ ,

$$\begin{aligned}
\tilde{w}^\top(\tilde{x} - \tilde{x}_0) &= \tilde{w}^\top \tilde{x} - \tilde{w}^\top \tilde{x}_0 && \text{(Theorem 4 in Subspaces and Rank, with coefficients 1 and } -1) \\
&= \tilde{w}^\top \tilde{x} - (-b) && (\tilde{x}_0 \in \mathcal{H}) \\
&= \tilde{w}^\top \tilde{x} + b, && \text{(arithmetic)}
\end{aligned}$$

so  $\tilde{x} \in \mathcal{H}$  exactly when  $\tilde{x} - \tilde{x}_0 \in \mathcal{H}_0$ , that is, exactly when  $\tilde{x} = \tilde{x}_0 + \tilde{s}$  with  $\tilde{s} = \tilde{x} - \tilde{x}_0 \in \mathcal{H}_0$ .  $\mathcal{H}_0$  is a subspace (part 1), so  $\mathcal{H}$  has the form in Definition 0.8.

**Part 3.** For  $\tilde{x}, \tilde{y} \in \mathcal{H}$ , both  $\tilde{w}^\top \tilde{x}$  and  $\tilde{w}^\top \tilde{y}$  equal  $-b$ , so

$$\begin{aligned}
\tilde{w} \cdot (\tilde{x} - \tilde{y}) &= \tilde{w}^\top \tilde{x} - \tilde{w}^\top \tilde{y} && \text{(Theorem 4 in Subspaces and Rank, with coefficients 1 and } -1) \\
&= (-b) - (-b) && (\tilde{x}, \tilde{y} \in \mathcal{H}) \\
&= 0. && \text{(arithmetic)}
\end{aligned}$$

□

**Example 0.21** (The plane  $x_1 + x_2 + x_3 = 1$  in  $\mathbb{R}^3$ ). Take  $\tilde{w} = (1, 1, 1)$  and  $b = -1$ , so  $\mathcal{H} = \{\tilde{x} : x_1 + x_2 + x_3 - 1 = 0\}$ , the plane that Example 5 in Subspaces and Rank showed is not a subspace. By part 1,  $\mathcal{H}_0 = \{\tilde{x} : x_1 + x_2 + x_3 = 0\}$  is a subspace of dimension  $3 - 1 = 2$ .

- $\tilde{x}_0 = (1, 0, 0)$  is in  $\mathcal{H}$ , since  $1 + 0 + 0 - 1 = 0$ , and so is  $\tilde{y} = (0, 1, 0)$ .
- Their difference  $\tilde{y} - \tilde{x}_0 = (-1, 1, 0)$  has entries summing to  $-1 + 1 + 0 = 0$ , so it is in  $\mathcal{H}_0$ , and  $\tilde{y} = \tilde{x}_0 + (-1, 1, 0)$ , as part 2 says.

•

$$\begin{aligned}
\tilde{w} \cdot (-1, 1, 0) &= -1 + 1 + 0 \\
&= 0,
\end{aligned}$$

as part 3 says.

- The point built in the proof is

$$\begin{aligned}
\tilde{x}^* &= -\frac{-1}{3} (1, 1, 1) \\
&= \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right),
\end{aligned}$$

and  $\frac{1}{3} + \frac{1}{3} + \frac{1}{3} - 1 = 0$ .

**Theorem 0.13** (Where a hyperplane passes, and its shape in  $\mathbb{R}^2$  and  $\mathbb{R}^3$ ). Let  $\mathcal{H} = \{\underline{x} \in \mathbb{R}^p : \underline{w} \cdot \underline{x} + b = 0\}$  be a hyperplane ([Definition 22 in Matrices](#)).

1.  $\mathcal{H}$  contains the origin  $\underline{0}$  exactly when  $b = 0$ .
2. If  $p = 2$ , then  $\mathcal{H}$  is a line:  $\mathcal{H} = \{\underline{x}_0 + t \underline{v} : t \in \mathbb{R}\}$  for some  $\underline{x}_0 \in \mathbb{R}^2$  and some nonzero  $\underline{v} \in \mathbb{R}^2$ .
3. If  $p = 3$ , then  $\mathcal{H}$  is a plane:  $\mathcal{H} = \{\underline{x}_0 + s \underline{v}_1 + t \underline{v}_2 : s, t \in \mathbb{R}\}$  for some  $\underline{x}_0 \in \mathbb{R}^3$  and some linearly independent  $\underline{v}_1, \underline{v}_2 \in \mathbb{R}^3$ .

### **i** Proof

*Proof.* **Part 1.** Since  $\underline{w} \cdot \underline{0} = 0$ ,

$$\begin{aligned} \underline{w} \cdot \underline{0} + b &= 0 + b \quad (\text{the dot product with } \underline{0} \text{ is } 0) \\ &= b, \quad (\text{Theorem 4 in Convexity, Infimum and Sums}) \end{aligned}$$

so  $\underline{0} \in \mathcal{H}$  exactly when  $b = 0$ .

**Parts 2 and 3.** By Theorem 0.12,  $\mathcal{H} = \{\underline{x}_0 + \underline{s} : \underline{s} \in \mathcal{H}_0\}$  for any  $\underline{x}_0 \in \mathcal{H}$ , where  $\mathcal{H}_0$  is a subspace of  $\mathbb{R}^p$  of dimension  $p - 1$ .

For  $p = 2$ ,  $\mathcal{H}_0$  has dimension 1, so it has a basis ([Definition 6 in Subspaces and Rank](#)) with one vector  $\underline{v}$ . That vector is nonzero, and  $\mathcal{H}_0 = \{t \underline{v} : t \in \mathbb{R}\}$ .

For  $p = 3$ ,  $\mathcal{H}_0$  has dimension 2, so it has a basis with two vectors  $\underline{v}_1$  and  $\underline{v}_2$ . They are linearly independent, and  $\mathcal{H}_0 = \{s \underline{v}_1 + t \underline{v}_2 : s, t \in \mathbb{R}\}$ .  $\square$

**Example 0.22** (A line in  $\mathbb{R}^2$  and a plane in  $\mathbb{R}^3$ , checked on a grid of points). In  $\mathbb{R}^2$ , take  $\underline{w} = (1, 1)$  and  $b = -1$ , the line of [Example 20 in Matrices](#). The point  $\underline{x}_0 = -\frac{b}{\underline{w} \cdot \underline{w}} \underline{w}$  is on it, and the vector  $\underline{v} = (-w_2, w_1)$  is nonzero with  $\underline{w} \cdot \underline{v} = 0$ , so it spans the subspace  $\mathcal{H}_0$  of Theorem 0.12. In  $\mathbb{R}^3$ , take  $\underline{w} = (1, 1, 1)$  and  $b = -1$ , the plane of [Example 0.21](#), with the independent vectors  $\underline{v}_1 = (1, -1, 0)$  and  $\underline{v}_2 = (0, 1, -1)$  in  $\mathcal{H}_0$ . The code evaluates  $\underline{w} \cdot \underline{x} + b$  at the origin and at points  $\underline{x}_0 + t \underline{v}$  and  $\underline{x}_0 + s \underline{v}_1 + t \underline{v}_2$ .

```

level <- function(w, b, x) sum(w * x) + b

w2 <- c(1, 1)
b2 <- -1
point_line <- -b2 / sum(w2 * w2) * w2
v <- c(-w2[2], w2[1])
line_levels <- sapply(-2:2, function(t) level(w2, b2, point_line + t * v))

w3 <- c(1, 1, 1)
b3 <- -1
point_plane <- -b3 / sum(w3 * w3) * w3
v1 <- c(1, -1, 0)
v2 <- c(0, 1, -1)
grid <- expand.grid(s = -2:2, t = -2:2)
plane_levels <- mapply(
  function(s, t) level(w3, b3, point_plane + s * v1 + t * v2),
  grid$s, grid$t
)

c(
  origin_in_R2 = level(w2, b2, c(0, 0)),
  origin_in_R3 = level(w3, b3, c(0, 0, 0)),
  origin_with_b0 = level(w3, 0, c(0, 0, 0))
)
#>      origin_in_R2      origin_in_R3 origin_with_b0
#>             -1             -1             0

```

At the origin the left side equals the offset, so the origin is on the line only if  $b$  is 0: it gives -1 for the line and -1 for the plane, and 0 once the offset is changed to 0. All 5 points  $\tilde{x}_0 + t \tilde{v}$  on the line and all 25 points  $\tilde{x}_0 + s \tilde{v}_1 + t \tilde{v}_2$  on the plane satisfy the equation up to rounding error: 30 of the 30 points have  $|\tilde{w} \cdot \tilde{x} + b| < 10^{-10}$ .

**Definition 0.9** (Open half-space). Let  $\mathcal{H} = \{\tilde{x} \in \mathbb{R}^p : \tilde{w} \cdot \tilde{x} + b = 0\}$  be a hyperplane ([Definition 22 in Matrices](#)). The two **open half-spaces** [6] of  $\mathcal{H}$  are

$$\mathcal{H}^+ \stackrel{\text{def}}{=} \{\tilde{x} \in \mathbb{R}^p : \tilde{w} \cdot \tilde{x} + b > 0\} \quad \text{and} \quad \mathcal{H}^- \stackrel{\text{def}}{=} \{\tilde{x} \in \mathbb{R}^p : \tilde{w} \cdot \tilde{x} + b < 0\}. \quad (1)$$

**Definition 0.10** (Closed half-space). The **closed half-spaces** of the hyperplane  $\mathcal{H}$  of [Definition 0.9](#) also include  $\mathcal{H}$  itself:  $\mathcal{H}^+ \cup \mathcal{H}$  and  $\mathcal{H}^- \cup \mathcal{H}$ .

**i Remark**

*Remark.* Replacing  $(\underline{w}, b)$  by  $(-\underline{w}, -b)$  describes the same hyperplane but swaps the names  $\mathcal{H}^+$  and  $\mathcal{H}^-$ . The sign of  $\underline{w} \cdot \underline{x} + b$  tells you which side of the hyperplane  $\underline{x}$  is on. A linear classifier uses this sign to assign one of two labels.

**Example 0.23** (Half-spaces of a line in the plane). Let  $\mathcal{H}$  be the line in  $\mathbb{R}^2$  with  $\underline{w} = (1, 2)$  and  $b = -2$  (Definition 22 in Matrices), so  $\mathcal{H} = \{\underline{x} : x_1 + 2x_2 - 2 = 0\}$ . Evaluate  $\underline{w} \cdot \underline{x} + b$  at three points (Definition 0.9):

- At  $(0, 0)$  it equals  $0 + 0 - 2 = -2 < 0$ , so  $(0, 0) \in \mathcal{H}^-$ .
- At  $(2, 0)$  it equals  $2 + 0 - 2 = 0$ , so  $(2, 0) \in \mathcal{H}$ .
- At  $(2, 1)$  it equals  $2 + 2 - 2 = 2 > 0$ , so  $(2, 1) \in \mathcal{H}^+$ .

Moving from  $(2, 0)$  in the direction  $\underline{w} = (1, 2)$  gives  $(3, 2)$ , where the value is  $3 + 4 - 2 = 5 > 0$ . So the direction  $\underline{w}$  leads into  $\mathcal{H}^+$ .

**i Remark** (The two sides of a hyperplane)

*Remark 0.1* (The two sides of a hyperplane). Every point of  $\mathbb{R}^p$  is in exactly one of  $\mathcal{H}^+$ ,  $\mathcal{H}$  and  $\mathcal{H}^-$ , because the number  $\underline{w} \cdot \underline{x} + b$  is positive, zero or negative. The vector  $\underline{w}$  points into  $\mathcal{H}^+$ : the number  $\underline{w} \cdot \underline{x} + b$  goes up when  $\underline{x}$  moves in the direction  $\underline{w}$ .

**Theorem 0.14** (Differences of points in a hyperplane and in a half-space). Let  $\underline{w} \in \mathbb{R}^p$  be nonzero, let  $b \in \mathbb{R}$ , and let  $\mathcal{H}$ ,  $\mathcal{H}^+$  be the hyperplane and open half-space of Definition 0.9. For a set  $S \subseteq \mathbb{R}^p$ , let  $D(S) \stackrel{\text{def}}{=} \{\underline{y} - \underline{x} : \underline{x}, \underline{y} \in S\}$  be the set of differences of its points.

1.  $D(\mathcal{H}) = \{\underline{s} : \underline{w} \cdot \underline{s} = 0\}$ , a subspace of dimension  $p - 1$ .
2.  $D(\mathcal{H}^+) = \mathbb{R}^p$ , a subspace of dimension  $p$ .

The same holds for  $\mathcal{H}^-$ .

**i Proof**

*Proof. Part 1.* Let  $\underline{x}, \underline{y} \in \mathcal{H}$ . By Theorem 0.12 part 3,  $\underline{w} \cdot (\underline{y} - \underline{x}) = 0$ , so  $\underline{y} - \underline{x} \in \mathcal{H}_0 = \{\underline{s} : \underline{w} \cdot \underline{s} = 0\}$ . Conversely, take any  $\underline{x}_0 \in \mathcal{H}$  (Theorem 0.12 part 2 gives such a point). For any  $\underline{s} \in \mathcal{H}_0$ , the point  $\underline{x}_0 + \underline{s}$  is in  $\mathcal{H}$  by the same part, and  $(\underline{x}_0 + \underline{s}) - \underline{x}_0 = \underline{s}$ . So  $D(\mathcal{H}) = \mathcal{H}_0$ , which has dimension  $p - 1$  by Theorem 0.12 part 1.

**Part 2.** Every difference is in  $\mathbb{R}^p$ , so  $D(\mathcal{H}^+) \subseteq \mathbb{R}^p$ . For the reverse inclusion, take any  $\underline{x}_0 \in \mathcal{H}$ . Take any  $\underline{v} \in \mathbb{R}^p$ . For a number  $t > 0$  to be chosen, let  $\underline{x}_t \stackrel{\text{def}}{=} \underline{x}_0 + t\underline{w}$ . Then

$$\begin{aligned}\tilde{w} \cdot \tilde{x}_t + b &= (\tilde{w} \cdot \tilde{x}_0 + b) + t \tilde{w} \cdot \tilde{w} \quad (\text{linearity of } \tilde{x} \mapsto \tilde{w} \cdot \tilde{x}) \\ &= t \tilde{w} \cdot \tilde{w}, \quad (\tilde{x}_0 \in \mathcal{H})\end{aligned}$$

which is positive, so  $\tilde{x}_t \in \mathcal{H}^+$ . Next,

$$\begin{aligned}\tilde{w} \cdot (\tilde{x}_t + \tilde{v}) + b &= (\tilde{w} \cdot \tilde{x}_t + b) + \tilde{w} \cdot \tilde{v} \quad (\text{linearity of } \tilde{x} \mapsto \tilde{w} \cdot \tilde{x}) \\ &= t \tilde{w} \cdot \tilde{w} + \tilde{w} \cdot \tilde{v}. \quad (\text{the display above})\end{aligned}$$

Because  $\tilde{w} \cdot \tilde{w} > 0$ , this is positive once  $t > -\tilde{w} \cdot \tilde{v} / \tilde{w} \cdot \tilde{w}$ . Choose such a  $t > 0$ . Then  $\tilde{x}_t$  and  $\tilde{x}_t + \tilde{v}$  are both in  $\mathcal{H}^+$ , and their difference is  $\tilde{v}$ . So  $\tilde{v} \in D(\mathcal{H}^+)$ . The whole space  $\mathbb{R}^p$  is a subspace of dimension  $p$  ([Definition 9 in Subspaces and Rank](#)).

The proof for  $\mathcal{H}^-$  is the same with  $-\tilde{w}$  and  $-b$  in place of  $\tilde{w}$  and  $b$ .  $\square$

**Example 0.24** (Two iris species on either side of a line). The `iris` data set in R has the petal length and petal width, in centimeters, of 150 irises. Take the 100 flowers of the species `versicolor` and `virginica`, and let  $\tilde{x} = (x_1, x_2)$  be the petal length and petal width of one flower. A logistic regression of the species on  $\tilde{x}$  gives a coefficient vector  $\tilde{w}$  and an intercept  $b$ . The line  $\tilde{w} \cdot \tilde{x} + b = 0$  is a hyperplane in  $\mathbb{R}^2$  ([Definition 22 in Matrices](#)), and the two species mostly fall in its two open half-spaces ([Definition 0.9](#)). [Figure 1](#) shows the line and the flowers.

```
two_species <- dplyr::filter(iris, Species != "setosa") |> droplevels()
fit <- glm(
  Species ~ Petal.Length + Petal.Width,
  family = binomial,
  data = two_species
)
w <- coef(fit)[c("Petal.Length", "Petal.Width")]
b <- coef(fit)[["(Intercept)"]]
two_species$score <- as.vector(
  as.matrix(two_species[c("Petal.Length", "Petal.Width")]) %*% w + b
)
two_species$side <- ifelse(two_species$score > 0, "positive", "negative")
ggplot2::ggplot(two_species) +
  ggplot2::aes(x = Petal.Length, y = Petal.Width, colour = Species) +
  ggplot2::geom_point() +
  ggplot2::geom_abline(intercept = -b / w[[2]], slope = -w[[1]] / w[[2]]) +
  ggplot2::labs(x = "Petal length (cm)", y = "Petal width (cm)")
```

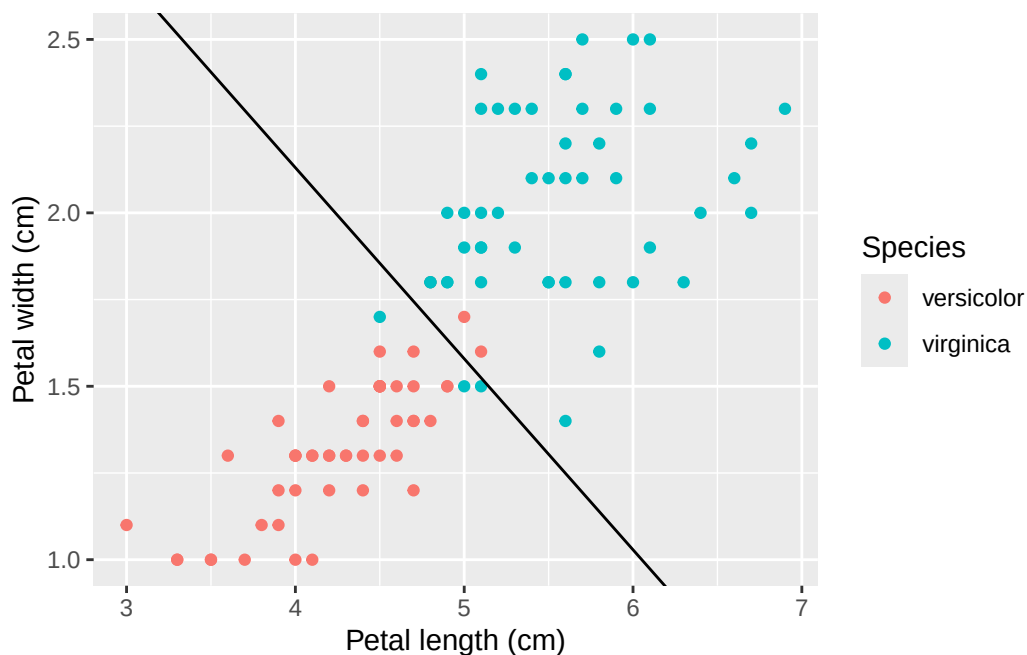


Figure 1: Petal length and width of versicolor and virginica irises, with the line where a logistic regression is indifferent between the two species

```
table(two_species$Species, two_species$side)
#>
#>           negative positive
#> versicolor      47       3
#> virginica       3      47
```

The vector  $w$  points toward the virginica side, so the positive half-space  $\mathcal{H}^+$  holds 50 of the 100 flowers, and 47 of them are virginica. The line is a set of dimension  $p - 1 = 1$ , and each side is a region of the plane of dimension  $p = 2$  (Theorem 0.14).

**Exercise 0.1** (Which side of the plane?). Let  $\mathcal{H}$  be the plane in  $\mathbb{R}^3$  with  $\tilde{w} = (1, 1, 1)$  and  $b = -1$  (Example 0.21).

1. For each point, say whether it is in  $\mathcal{H}^+$ ,  $\mathcal{H}$  or  $\mathcal{H}^-$  (Definition 0.9):  $(0, 0, 0)$ ,  $(1, 0, 0)$  and  $(1, 1, 1)$ .
2. Write the plane and its half-spaces using  $-\tilde{w}$  and  $-b$ . Which points from part 1 change sides?
3. State the dimension of  $D(\mathcal{H})$  and of  $D(\mathcal{H}^+)$  (Theorem 0.14).

## Solution

### Solution 0.1.

1. Compute  $\tilde{w} \cdot \tilde{x} + b = x_1 + x_2 + x_3 - 1$  at each point.
  - At  $(0, 0, 0)$  it equals  $0 + 0 + 0 - 1 = -1 < 0$ , so the point is in  $\mathcal{H}^-$ .
  - At  $(1, 0, 0)$  it equals  $1 + 0 + 0 - 1 = 0$ , so the point is in  $\mathcal{H}$ .
  - At  $(1, 1, 1)$  it equals  $1 + 1 + 1 - 1 = 2 > 0$ , so the point is in  $\mathcal{H}^+$ .
2. With  $-\tilde{w} = (-1, -1, -1)$  and  $-b = 1$ , the equation is  $-x_1 - x_2 - x_3 + 1 = 0$ . Multiplying both sides by  $-1$  gives the original equation, so it is the same plane. The new expression is the negative of the old one at every point, so the new positive half-space is the old  $\mathcal{H}^-$  and the new negative half-space is the old  $\mathcal{H}^+$ . The origin  $(0, 0, 0)$  is now in the positive half-space, and  $(1, 1, 1)$  is now in the negative half-space. The point  $(1, 0, 0)$  stays on the plane. So the origin and  $(1, 1, 1)$  both change names, and the point on the plane does not.
3. Here  $p = 3$ .  $D(\mathcal{H})$  is the subspace  $\{\tilde{s} : s_1 + s_2 + s_3 = 0\}$ , which has dimension  $p - 1 = 2$ .  $D(\mathcal{H}^+) = \mathbb{R}^3$  has dimension  $p = 3$ .

**Definition 0.11** (Rank factorization). Let  $\mathbf{A}$  be an  $m \times n$  matrix with  $\text{rank}(\mathbf{A}) = r \geq 1$ . A **rank factorization** [7] of  $\mathbf{A}$  is a product

$$\underset{m \times n}{\mathbf{A}} = \underset{m \times r}{\mathbf{C}} \underset{r \times n}{\mathbf{R}}.$$

**Example 0.25** (A rank factorization of a rank-one matrix). The matrix  $\mathbf{A}$  in Example 0.1 has rank 1 (Example 0.10), and

$$\begin{aligned} \begin{bmatrix} 1 \\ 3 \end{bmatrix} \begin{bmatrix} 1 & -2 & -2 \end{bmatrix} &= \begin{bmatrix} 1 \cdot 1 & 1 \cdot (-2) & 1 \cdot (-2) \\ 3 \cdot 1 & 3 \cdot (-2) & 3 \cdot (-2) \end{bmatrix} \\ &= \begin{bmatrix} 1 & -2 & -2 \\ 3 & -6 & -6 \end{bmatrix}, \end{aligned}$$

so  $\mathbf{C} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$  ( $2 \times 1$ ) and  $\mathbf{R} = \begin{bmatrix} 1 & -2 & -2 \end{bmatrix}$  ( $1 \times 3$ ) are a rank factorization. It is not unique:  $\mathbf{C} = \begin{bmatrix} 2 \\ 6 \end{bmatrix}$  and  $\mathbf{R} = \begin{bmatrix} \frac{1}{2} & -1 & -1 \end{bmatrix}$  give the same product. A product  $\mathbf{C}\mathbf{R}$  with  $\mathbf{C}$  of size  $2 \times 2$  is not a rank factorization of this  $\mathbf{A}$ , because the inner dimension must equal  $\text{rank}(\mathbf{A}) = 1$ .

**Theorem 0.15** (Every nonzero matrix has a rank factorization). *Every  $m \times n$  matrix  $\mathbf{A}$  with  $\text{rank}(\mathbf{A}) = r \geq 1$  has a rank factorization (Definition 0.11). One is given by taking the columns of  $\mathbf{C}$  to be any basis of  $\mathcal{C}(\mathbf{A})$ .*

### Proof

*Proof.*  $\mathcal{C}(\mathbf{A})$  has dimension  $r$  (Theorem 0.6), so it has a basis  $\underline{c}_1, \dots, \underline{c}_r$  (Theorem 9 in Subspaces and Rank); let  $\mathbf{C}$  be the  $m \times r$  matrix with these columns. Each column  $\underline{a}_j$  of  $\mathbf{A}$  is in  $\mathcal{C}(\mathbf{A})$ , so  $\underline{a}_j = r_{1j}\underline{c}_1 + \dots + r_{rj}\underline{c}_r$  for some numbers  $r_{ij}$ , because the basis spans  $\mathcal{C}(\mathbf{A})$ . Let  $\mathbf{R}$  be the  $r \times n$  matrix with entries  $r_{ij}$ , whose column  $j$  is  $\underline{r}_j = (r_{1j}, \dots, r_{rj})$ . Then column  $j$  of  $\mathbf{CR}$  is

$$\begin{aligned} \mathbf{C}\underline{r}_j &= r_{1j}\underline{c}_1 + \dots + r_{rj}\underline{c}_r && \text{(Theorem 3 in Subspaces and Rank)} \\ &= \underline{a}_j, && \text{(choice of the } r_{ij}) \end{aligned}$$

so  $\mathbf{CR} = \mathbf{A}$ . □

**Example 0.26** (Building a rank factorization from a basis). For  $\mathbf{A}$  in Example 0.1,  $(1, 3)$  spans  $\mathcal{C}(\mathbf{A})$  (Example 0.2), and it is linearly independent because it is nonzero ( $c(1, 3) = \underline{0}_2$  forces  $c = 0$ ), so it is a basis of  $\mathcal{C}(\mathbf{A})$ . The columns of  $\mathbf{A}$  are  $(1, 3) = 1 \cdot (1, 3)$ ,  $(-2, -6) = -2 \cdot (1, 3)$  and  $(-2, -6) = -2 \cdot (1, 3)$ , so the construction in Theorem 0.15 gives  $\mathbf{C} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$  and  $\mathbf{R} = \begin{bmatrix} 1 & -2 & -2 \end{bmatrix}$ , the factorization in Example 0.25.

- [1] Wikipedia contributors, “Row and column spaces — Wikipedia, the free encyclopedia.” 2026. Available: [https://en.wikipedia.org/wiki/Row\\_and\\_column\\_spaces](https://en.wikipedia.org/wiki/Row_and_column_spaces)
- [2] Wikipedia contributors, “Gram matrix — Wikipedia, the free encyclopedia.” 2026. Available: [https://en.wikipedia.org/wiki/Gram\\_matrix](https://en.wikipedia.org/wiki/Gram_matrix)
- [3] H. Zhou, “Rank and nullity.” Lecture notes for Biostat 216, Mathematical Methods for Biostatistics, University of California, Los Angeles, 2024. Accessed: Oct. 05, 2026. [Online]. Available: <https://ucla-biostat-216.github.io/2024fall/slides/05-rank/05-rank.html>
- [4] Wikipedia contributors, “Rank–nullity theorem — Wikipedia, the free encyclopedia.” 2026. Available: [https://en.wikipedia.org/wiki/Rank%E2%80%93nullity\\_theorem](https://en.wikipedia.org/wiki/Rank%E2%80%93nullity_theorem)
- [5] Wikipedia contributors, “Affine space — Wikipedia, the free encyclopedia.” 2026. Available: [https://en.wikipedia.org/wiki/Affine\\_space](https://en.wikipedia.org/wiki/Affine_space)
- [6] Wikipedia contributors, “Half-space (geometry) — Wikipedia, the free encyclopedia.” 2026. Available: [https://en.wikipedia.org/wiki/Half-space\\_\(geometry\)](https://en.wikipedia.org/wiki/Half-space_(geometry))
- [7] Wikipedia contributors, “Rank factorization — Wikipedia, the free encyclopedia.” 2026. Available: [https://en.wikipedia.org/wiki/Rank\\_factorization](https://en.wikipedia.org/wiki/Rank_factorization)