Linear Algebra Review

1 Vector Spaces

Definition 1.1. *A set $\mathcal{V}$ is called a* vector space *if $\forall x \in \mathcal{V}$, $\forall y \in \mathcal{V}$, $\forall \lambda \in \mathbb{R}$, (i) $x + y \in \mathcal{V}$ and (ii) $\lambda x \in \mathcal{V}$ (i.e., if $\mathcal{V}$ is closed under addition and scalar multiplication).*

Examples:

1. $\mathcal{V} = \mathbb{R}^n$: the n-dimensional Euclidean space
2. $\mathcal{V} = \mathbb{R}^{m \times n}$: the space of $m \times n$ real matrices
3. $\mathcal{V} = \{\mathbf{0}\}$: the vector of all zeros in $\mathbb{R}^n$

Notes: Every vector space should contain a special "zero" element (simply pick $\lambda = 0$). We will represent the number zero by 0 and the vector of all zeros in $\mathbb{R}^n$ by $\mathbf{0}$. The dimension of the vector will always be clear from the context.

2 Subspaces

Definition 2.1. *Let $\mathcal{V}$ be a vector space and let $\mathcal{S} \subseteq \mathcal{V}$ be a nonempty subset. $\mathcal{S}$ is called a* subspace *if $\forall x \in \mathcal{S}$, $\forall y \in \mathcal{S}$, $\forall \lambda \in \mathbb{R}$, (i) $x + y \in \mathcal{S}$ and (ii) $\lambda x \in \mathcal{S}$ (i.e., if $\mathcal{S}$ is closed under addition and scalar multiplication).*

Examples:

1. $\mathcal{S} = \{\mathbf{0}\} \subset \mathbb{R}^n$
2. $\mathcal{S} = \mathbb{R}^n$
3. $\mathcal{S} = \{x \in \mathbb{R}^2 : x_1 + x_2 = 0\}$ (verify the definition)

Notes:

1. Every vector space is a subspace.
2. Subspaces are defined with respect to a given vector space.
3. Every subspace should contain the zero element.

3 Span

Definition 3.1. *Let $\{p^1, \ldots, p^k\} \subseteq \mathbb{R}^n$. Then, span $\left(\{p^1, \ldots, p^k\}\right)$ is the set of all vectors that can be written as linear combinations of $\{p^1, \ldots, p^k\}$, i.e.,*

$$span\left(\{p^1, \ldots, p^k\}\right) = \left\{ y \in \mathbb{R}^n : y = \sum_{i=1}^{k} \lambda_i p^i \text{ for some } \lambda_1, \ldots, \lambda_k \in \mathbb{R} \right\}.$$

Proposition 3.1. *Let $\{p^1, \ldots, p^k\} \subseteq \mathbb{R}^n$. Then, span $\left(\{p^1, \ldots, p^k\}\right)$ is a subspace.*

Proof. Left as an exercise. □

4 Linear Independence

Definition 4.1. *A set of vectors $\{p^1, \ldots, p^k\} \subseteq \mathbb{R}^n$ is said to be* linearly independent *if no nontrivial linear combination of the vectors $\{p^1, \ldots, p^k\}$ yields the zero vector, i.e.,*

$$\sum_{i=1}^{k} \lambda_i p^i = \mathbf{0} \Longleftrightarrow \lambda_i = 0, \quad i = 1, \ldots, k.$$

Otherwise, they are called linearly dependent*.*

Note: A linearly independent set of vectors **cannot** contain the zero vector $\mathbf{0} \in \mathbb{R}^n$ (why not?).

Proposition 4.1. *Let $\mathcal{T} = \{p^1, \ldots, p^k\}$. If the vectors $\{p^1, \ldots, p^k\}$ are linearly dependent, then there exists $j \in \{1, \ldots, k\}$ such that $p^j \in span\left(\{p^1, \ldots, p^{j-1}\}\right)$ (i.e., one of the vectors in $\mathcal{T}$ can be written as a linear combination of the previous vectors in T).*

Proof. Suppose that the vectors $\{p^1, \ldots, p^k\}$ are linearly dependent. Then, there exist $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ such that $\sum_{i=1}^{k} \lambda_i p^i = \mathbf{0}$ and at least one $\lambda_i \neq 0$, where $i \in \{1, \ldots, k\}$. Let $j \in \{1, \ldots, k\}$ be the largest index such that $\lambda_j \neq 0$. Then,

$$\lambda_j p^j = \sum_{i=1}^{j-1} \lambda_i p^i.$$

Since $\lambda_j \neq 0$, dividing both sides by λ_j, we obtain $p^j \in \text{span}\left(\{p^1, \ldots, p^{j-1}\}\right)$. □

Proposition 4.2. *Let $\mathcal{T} = \{p^1, \ldots, p^k\}$. If $\mathcal{T}$ is a linearly independent set of vectors, then any subset of vectors $\mathcal{U} \subseteq \mathcal{T}$ is also linearly independent. If $\mathcal{T}$ is a linearly dependent set of vectors, then any set of vectors $\mathcal{W}$ with $\mathcal{T} \subseteq \mathcal{W}$ is also linearly dependent.*

Proof. Left as an exercise. □

5 Basis and Dimension

Definition 5.1. *Let $\mathcal{S} \subseteq \mathbb{R}^n$ be a subspace. A set of vectors $\{p^1, \ldots, p^r\}$ is said to be a* basis of $\mathcal{S}$ *if (i) span $\left(\{p^1, \ldots, p^r\}\right) = \mathcal{S}$ and (ii) the vectors $\{p^1, \ldots, p^r\}$ are linearly independent.*

Example: For $i = 1, \ldots, n$, let $e^i \in \mathbb{R}^n$ denote the ith unit vector, i.e., the ith component of e^i is one and all the remaining components are zero. Then, it is easy to verify that $\{e^1, \ldots, e^n\}$ is a basis of $\mathbb{R}^n$. In fact, this basis is called the *standard basis* of $\mathbb{R}^n$. For $n = 3$,

$$e^1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad e^2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad e^3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Proposition 5.1. *Let $\mathcal{S} \subseteq \mathbb{R}^n$ be a subspace. Suppose that $\{p^1, \ldots, p^r\}$ is a basis of $\mathcal{S}$. Let $\mathcal{T} = \{q^1, \ldots, q^s\} \subseteq \mathcal{S}$ be a set of linearly independent vectors. Then, $s \leq r$.*

Proof. We will prove this assertion by contradiction. Suppose, for a contradiction, that $r < s$. Since span $\left(\{p^1, \ldots, p^r\}\right) = \mathcal{S}$ and $q^1 \in \mathcal{S}$, q^1 can be written as a linear combination of $\{p^1, \ldots, p^r\}$. Therefore, the set $\{q^1, p^1, \ldots, p^r\}$ is linearly dependent. By Proposition 4.1, there exists a vector that can be written as a linear combination of the previous vectors in this set. This vector cannot be q^1. It has to be some vector p^j, where $j \in \{1, \ldots, r\}$. Therefore, let $\mathcal{U}_1 = \{q^1, p^1, \ldots, p^r\} \backslash \{p^j\}$. We claim that span$(\mathcal{U}_1) = S$. Indeed, any vector in $\mathcal{S}$ can be written as a linear combination of $\{p^1, \ldots, p^r\}$ and p^j can be written as a linear combination of $\{q^1, p^1, \ldots, p^{j-1}\}$, which establishes the claim.

In a similar manner, consider $\{q^2\} \cup \mathcal{U}_1$. By a similar reasoning, this set is linearly dependent. By Proposition 4.1, there exists a vector that can be written as a linear combination of the previous vectors in this set. This vector cannot be q^1 or q^2 since they are linearly independent by Proposition 4.2. It has to be some vector p^k. Let $\mathcal{U}_2 = (\{q^2\} \cup \mathcal{U}_1) \backslash \{p^k\}$. By a similar argument, $\text{span}(\mathcal{U}_2) = \mathcal{S}$.

We can continue this argument by adding one element from $\mathcal{T}$ and removing one element from $\{p^1, \ldots, p^r\}$. Note that the number of vectors is always equal to r. Therefore, after repeating this procedure r times, we obtain $\text{span}\left(\{q^1, \ldots, q^r\}\right) = \mathcal{S}$. Since $r < s$, we have $q^{r+1} \in \mathcal{S}$, which implies $q^{r+1} \in \text{span}\left(\{q^1, \ldots, q^r\}\right)$. Therefore, $\{q^1, \ldots, q^{r+1}\}$ is linearly dependent, which contradicts our assumption that $\mathcal{T} = \{q^1, \ldots, q^s\}$ is linearly independent. Therefore, we have $s \leq r$. □

Corollary 5.1. *Every basis of a subspace contains the same number of vectors.*

Definition 5.2. *Let $\mathcal{S} \subseteq \mathbb{R}^n$ be a subspace. The* dimension *of $\mathcal{S}$ is the number of vectors in any basis of $\mathcal{S}$ and is denoted by $\dim(\mathcal{S})$.*

Examples:

1. $\dim(\mathbb{R}^n) = n$ since $\{e^1, \ldots, e^n\}$ is a basis of $\mathbb{R}^n$.
2. $\dim(\{\mathbf{0}\}) = 0$.
3. Let $\mathcal{S} = \{x \in \mathbb{R}^2 : x_1 + x_2 = 0\}$. Then, $\dim(\mathcal{S}) = 1$. To see this, let

$$\hat{x} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

   Then, you can show that $\text{span}(\{\hat{x}\}) \subseteq \mathcal{S}$. Conversely, for any $x \in \mathcal{S}$, $x_1 + x_2 = 0$, which implies $x_2 = -x_1$, i.e., $x = \lambda \hat{x}$, where $\lambda = x_1$. Therefore, $x \in \text{span}(\{\hat{x}\})$. Combining these two inclusions, we obtain $\mathcal{S} = \text{span}(\{\hat{x}\})$. Since $\{\hat{x}\}$ is linearly independent, it is a basis of $\mathcal{S}$. Therefore, $\dim(\mathcal{S}) = 1$.

Notes:

1. For a subspace $\mathcal{S} \subseteq \mathbb{R}^n$, $\dim(\mathcal{S})$ is the largest number of linearly independent vectors in $\mathcal{S}$.
2. If $\mathcal{S} = \text{span}\left(\{p^1, \ldots, p^r\}\right)$, then $\dim(\mathcal{S}) \leq r$. Furthermore, $\dim(\mathcal{S}) = r$ if and only if $\{p^1, \ldots, p^r\}$ is linearly independent.

6 Column Spaces and Null Spaces

Definition 6.1. *A matrix $A \in \mathbb{R}^{m \times n}$ is given by*

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}.$$

Notes:

1. A matrix $A \in \mathbb{R}^{m \times n}$ can be expressed in terms of its rows or its columns:

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} = \begin{bmatrix} (r^1)^T \\ \vdots \\ (r^m)^T \end{bmatrix} = \begin{bmatrix} c^1 \ldots c^n \end{bmatrix},$$

where

$$r^i = \begin{bmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{in} \end{bmatrix} \in \mathbb{R}^n, \quad i = 1, \ldots, m; \quad c^j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix} \in \mathbb{R}^m, \quad j = 1, \ldots, n.$$

2. Let $x \in \mathbb{R}^n$. Then,

$$\begin{aligned} Ax &= \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\ &= \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \end{bmatrix} \\ &= \begin{bmatrix} (r^1)^T x \\ \vdots \\ (r^m)^T x \end{bmatrix} \\ &= x_1 c^1 + x_2 c^2 + \ldots + x_n c^n. \end{aligned}$$

Therefore, the matrix-vector multiplication Ax can be viewed as taking a linear combination of the columns $c^1, \ldots, c^n$ of A using multipliers $x_1, \ldots, x_n$.

Definition 6.2. *Let $A \in \mathbb{R}^{m \times n}$. The* column space (range space) *of A is given by*

$$\mathcal{R}(A) = \{y \in \mathbb{R}^m : y = Ax \quad \textit{for some } x \in \mathbb{R}^n\}$$

(i.e., it is the set of all vectors that can be written as a linear combination of the columns of A).

Notes:

1. Let $A \in \mathbb{R}^{m \times n}$. Then, $\mathcal{R}(A) = \text{span}\left(\{c^1, \ldots, c^n\}\right)$, where $c^1, \ldots, c^n$ are the columns of A. Therefore $\mathcal{R}(A)$ is a subspace in $\mathbb{R}^m$.
2. Let $A \in \mathbb{R}^{m \times n}$ and let $b \in \mathbb{R}^m$. The system of linear equations $Ax = b$ has a solution in $x \in \mathbb{R}^n$ if and only if $b \in \mathcal{R}(A)$.
3. Let $A \in \mathbb{R}^{m \times n}$. The system of linear equations $Ax = b$ has a solution in $x \in \mathbb{R}^n$ for every $b \in \mathbb{R}^m$ if and only if $\mathcal{R}(A) = \mathbb{R}^m$.

Definition 6.3. *Let $A \in \mathbb{R}^{m \times n}$. The* column rank *of A is given by $\dim(\mathcal{R}(A))$, i.e., it is the largest number of linearly independent columns of A.*

Note: Let $A \in \mathbb{R}^{m \times n}$. Since $\mathcal{R}(A) = \text{span}\left(\{c^1, \ldots, c^n\}\right)$, it follows that the column rank of A satisfies $\dim(\mathcal{R}(A)) \leq n$. If $\dim(\mathcal{R}(A)) = n$, then $\{c^1, \ldots, c^n\}$ is linearly independent and we say A has full column rank.

Definition 6.4. *Let $A \in \mathbb{R}^{m \times n}$. The transpose of A is given by*

$$A^T = \begin{bmatrix} a_{11} & a_{21} & \ldots & a_{m1} \\ a_{12} & a_{22} & \ldots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \ldots & a_{mn} \end{bmatrix} = \begin{bmatrix} r^1 \ldots r^m \end{bmatrix} = \begin{bmatrix} (c^1)^T \\ \vdots \\ (c^n)^T \end{bmatrix}.$$

Note that $A^T \in \mathbb{R}^{n \times m}$.

Notes:

1. Let $A \in \mathbb{R}^{m\times n}$ and $x \in \mathbb{R}^n$. Then $(Ax)^T = x^T A^T$ (please verify).
2. Let $A \in \mathbb{R}^{m\times n}$. Then, $(A^T)^T = A$ (please verify).

Definition 6.5. *Let $A \in \mathbb{R}^{m\times n}$. The* row rank *of A is given by the column rank of A^T, i.e., it is equal to $dim(\mathcal{R}(A^T)) = dim\left(span\left(\{r^1, \ldots, r^m\}\right)\right)$.*

Note: Let $A \in \mathbb{R}^{m\times n}$. Note that $\mathcal{R}(A) \subseteq \mathbb{R}^m$ whereas $\mathcal{R}(A^T) \subseteq \mathbb{R}^n$.

Proposition 6.1. *Let $A \in \mathbb{R}^{m\times n}$. The row rank of A is equal to the column rank of A.*

Proof. Let k denote the row rank of A, i.e., $k = \dim(\mathcal{R}(A^T))$. Suppose that $\{p^1, \ldots, p^k\} \subset \mathbb{R}^n$ is a basis of $\mathcal{R}(A^T)$. Then, each row r^i can be written as a linear combination of these vectors, i.e.,

$$r^i = \sum_{\ell=1}^{k} \lambda_{i,\ell} p^\ell, \quad i = 1, \ldots, m.$$

Note that

$$c^j = Ae^j = \begin{bmatrix} (r^1)^T e^j \\ (r^2)^T e^j \\ \vdots \\ (r^m)^T e^j \end{bmatrix}, \quad j = 1, \ldots, n,$$

where $e^j \in \mathbb{R}^n$ denotes the jth unit vector, $j = 1, \ldots, n$. Therefore,

$$c^j = \begin{bmatrix} \left(\sum_{\ell=1}^{k} \lambda_{1,\ell} p^\ell\right)^T e^j \\ \left(\sum_{\ell=1}^{k} \lambda_{2,\ell} p^\ell\right)^T e^j \\ \vdots \\ \left(\sum_{\ell=1}^{k} \lambda_{m,\ell} p^\ell\right)^T e^j \end{bmatrix} = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \ldots & \lambda_{1k} \\ \lambda_{21} & \lambda_{22} & \ldots & \lambda_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{m1} & \lambda_{m2} & \ldots & \lambda_{mk} \end{bmatrix} \begin{bmatrix} (p^1)^T e^j \\ (p^2)^T e^j \\ \vdots \\ (p^k)^T e^j \end{bmatrix} = Mw^j, \quad j = 1, \ldots, n.$$

It follows that $c^j \in \mathcal{R}(M)$ for each $j = 1, \ldots, n$. Therefore, $\mathcal{R}(A) = \text{span}\left(\{c^1, \ldots, c^n\}\right) \subseteq \mathcal{R}(M)$ since every linear combination of the vectors $\{c^1, \ldots, c^n\}$ can also be written as a linear combination of the columns of M. Since M has k columns, $\dim(\mathcal{R}(M)) \leq k$, which implies that $\dim(\mathcal{R}(A)) \leq k = \dim(\mathcal{R}(A^T))$.

The same argument can be repeated starting with a basis for $\mathcal{R}(A)$ and one obtains the reverse inequality $k = \dim(\mathcal{R}(A^T)) \leq \dim(\mathcal{R}(A))$. Combining the two inequalities, we obtain the desired result. □

Definition 6.6. *Let $A \in \mathbb{R}^{m\times n}$. Then, the* rank *of A, denoted by rank(A), is given by the column (or, equivalently row) rank of A.*

Note: Let $A \in \mathbb{R}^{m\times n}$. Then, $\text{rank}(A) \leq \min\{m, n\}$. If $\text{rank}(A) = m$, then we say that A has full row rank. If $\text{rank}(A) = n$, then we say that A has full column rank.

Definition 6.7. *Let $A \in \mathbb{R}^{m\times n}$. The null space of A is given by*

$$\mathcal{N}(A) = \{z \in \mathbb{R}^n : Az = \mathbf{0}\}.$$

Notes:

1. Let $A \in \mathbb{R}^{m\times n}$. You can verify that $\mathcal{N}(A)$ is a subspace.
2. Note that $\mathcal{N}(A) \neq \emptyset$ since $\mathbf{0} \in \mathcal{N}(A)$.
3. $\mathcal{N}(A) = \{\mathbf{0}\}$ if and only if $\text{rank}(A) = n$ (i.e., A has full column rank).
4. Note that $\mathcal{N}(A) \subseteq \mathbb{R}^n$ whereas $\mathcal{R}(A) \subseteq \mathbb{R}^m$.

7 Orthogonal Complements and Fundamental Theorem of Linear Algebra

Definition 7.1. *Let $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$. We say that x and y are* orthogonal *if*

$$x^T y = \sum_{i=1}^{n} x_i y_i = 0.$$

Definition 7.2. *Let $\mathcal{S} \subseteq \mathbb{R}^n$ be a subspace. The* orthogonal complement *of $\mathcal{S}$ is given by*

$$\mathcal{S}^\perp = \{y \in \mathbb{R}^n : x^T y = 0 \text{ for all } x \in \mathcal{S}\},$$

i.e., it is the set of all vectors which are orthogonal to every vector in $\mathcal{S}$.

Notes:

1. Let $\mathcal{S} \subseteq \mathbb{R}^n$ be a subspace. It is easy to verify that $\mathcal{S}^\perp$ is a subspace.
2. Let $\mathcal{S} \subseteq \mathbb{R}^n$ be a subspace. Then, $\mathcal{S} \cap \mathcal{S}^\perp = \{\mathbf{0}\}$. To see this, let $x \in \mathbb{R}^n$ be a vector such that $x \in \mathcal{S} \cap \mathcal{S}^\perp$. Then, since $x \in \mathcal{S}$ and $x \in \mathcal{S}^\perp$, we have $x^T x = 0 = \sum_{i=1}^n x_i^2$, which is true if and only if $x = \mathbf{0} \in \mathbb{R}^n$.
3. Let $\mathcal{S} \subseteq \mathbb{R}^n$ be a subspace. Then, $(\mathcal{S}^\perp)^\perp = \mathcal{S}$ (proof left as an exercise)
4. $\{\mathbf{0}\}^\perp = \mathbb{R}^n$
5. $(\mathbb{R}^n)^\perp = \{\mathbf{0}\}$

Example: Let $\mathcal{S} = \{x \in \mathbb{R}^2 : x_1 + x_2 = 0\}$. Then, $\mathcal{S}^\perp = \{x \in \mathbb{R}^2 : x_1 - x_2 = 0\}$. (Check the definition and verify by drawing the two lines.)

Proposition 7.1. *Let $\mathcal{S} \subseteq \mathbb{R}^n$ be a subspace. $\mathcal{S}^\perp = \{\mathbf{0}\}$ if and only if $\mathcal{S} = \mathbb{R}^n$.*

Proof. $\Rightarrow$: Let $\mathcal{S} \subseteq \mathbb{R}^n$ be a subspace such that $\mathcal{S}^\perp = \{0\}$. Let $\{p^1, \ldots, p^r\} \subset \mathbb{R}^n$ be a basis of $\mathcal{S}$ and let $P = [p^1 \cdots p^r] \in \mathbb{R}^{n \times r}$. We claim that $\mathcal{N}(P^T) = \{\mathbf{0}\}$. Suppose, for a contradiction, that $\mathcal{N}(P^T) \neq \{\mathbf{0}\}$. Then, there exists $p \in \mathbb{R}^n$ such that $p \neq \mathbf{0}$ and $P^T p = \mathbf{0}$. Therefore, $(p^i)^T p = 0$ for each $i = 1, \ldots, r$. We claim that p is orthogonal to every vector in $\mathcal{S}$. Indeed, for any $x \in \mathcal{S}$, there exist real numbers $\lambda_1, \ldots, \lambda_r$ such that $x = \sum_{i=1}^r \lambda_i p^i$. Therefore, $p^T x = \sum_{i=1}^r \lambda_i (p^i)^T p = 0$. Since this is true for each $x \in \mathcal{S}$, we have $p \in \mathcal{S}^\perp$, which contradicts our assumption that $\mathcal{S}^\perp = \{0\}$. It follows that $\mathcal{N}(P^T) = \{0\}$, which is true if and only if P^T has full column rank. Therefore, $\text{rank}(P^T) = n$. It follows that the row rank of P^T is equal to n. However, the row rank of P^T is the same as the column rank of P. Therefore, $\text{rank}(P) = r = n$.

We now claim that $\mathcal{S} = \text{span}\{p^1, \ldots, p^n\} = \mathcal{R}(P) = \mathbb{R}^n$. Suppose, for a contradiction, that $\mathcal{S} \neq \mathbb{R}^n$. Then, there exists $x \in \mathbb{R}^n$ such that $x \notin \mathcal{S}$. Then, we claim that the set $\{x, p^1, \ldots, p^n\}$ is linearly independent. To see this, if the vectors were linearly dependent, there would be real numbers $\lambda_0, \lambda_1, \ldots, \lambda_n$, which are not all equal to zero, such that

$$\lambda_0 x + \sum_{i=1}^{n} \lambda_i p^i = \mathbf{0}.$$

Note that $\lambda_0 \neq 0$ since otherwise $\{p^1, \ldots, p^n\}$ would be linearly dependent. Then, dividing both sides of this equation by λ_0, we see that

$$x = \sum_{i=1}^{n} \left(-\frac{\lambda_i}{\lambda_0}\right) p^i,$$

which implies that $x \in \mathcal{S}$, which is a contradiction. Therefore, the set $\{x, p^1, \ldots, p^n\}$ is linearly independent. However, this set contains $n+1$ linearly independent vectors in $\mathbb{R}^n$ and since $\{e^1, \ldots, e^n\}$ is a basis for $\mathbb{R}^n$, we get a contradiction by Proposition 5.1. It follows that $\mathcal{S} = \text{span}\{p^1, \ldots, p^n\} = \mathcal{R}(P) = \mathbb{R}^n$.

$\Leftarrow$: Let $\mathcal{S} = \mathbb{R}^n$. Then, $\{e^1, \ldots, e^n\} \subset \mathcal{S}$. Therefore, for any $x \in \mathcal{S}^\perp$, we have $(e^i)^T x = x_i = 0$ for each $i = 1, \ldots, n$. It follows that $x = \mathbf{0}$, i.e., $\mathcal{S}^\perp = \{0\}$. □

Note: The proof of Proposition 7.1 reveals that any set of n linearly independent vectors in $\mathbb{R}^n$ constitutes a basis of $\mathbb{R}^n$.

Proposition 7.2. *Let $\mathcal{S} \subseteq \mathbb{R}^n$ be a subspace. Then, any vector $z \in \mathbb{R}^n$ can be written uniquely as $z = x + y$, where $x \in \mathcal{S}$ and $y \in \mathcal{S}^\perp$, i.e.,*

$$\mathbb{R}^n = \mathcal{S} + \mathcal{S}^\perp = \{x + y : x \in \mathcal{S}, y \in \mathcal{S}^\perp\}.$$

Proof. Let $\mathcal{V} = \mathcal{S} + \mathcal{S}^\perp$. It is easy to verify that $\mathcal{V}$ is a subspace in $\mathbb{R}^n$. Therefore, $\mathcal{V} \subseteq \mathbb{R}^n$. Suppose, for a contradiction, that $\mathcal{V} \neq \mathbb{R}^n$. Then, $\mathcal{V}^\perp \neq \{\mathbf{0}\}$ by Proposition 7.1, i.e., there exists $z \in \mathbb{R}^n$ such that $z \neq \mathbf{0}$ and z is orthogonal to every vector in $\mathcal{V}$. Let $x \in \mathcal{S}$ be any arbitrary vector. Then $x = x + \mathbf{0} \in \mathcal{V}$ since $\mathbf{0} \in S^\perp$. It follows that $z^T x = 0$, which implies that $z \in \mathcal{S}^\perp$. Since $z \in \mathcal{S}^\perp$, we have $z = \mathbf{0} + z \in \mathcal{V}$ since $\mathbf{0} \in \mathcal{S}$. Therefore, $z^T z = 0$, which implies that $z = \mathbf{0}$, a contradiction. It follows that $\mathcal{V} = \mathbb{R}^n$.

Let $z \in \mathbb{R}^n$ be an arbitrary vector. Therefore, there exist $x \in \mathcal{S}$ and $y \in \mathcal{S}^\perp$ such that $z = x + y$. To see that this decomposition is unique, suppose, for a contradiction, that there exist $x' \in \mathcal{S}$ and $y' \in \mathcal{S}^\perp$ such that $x \neq x'$, $y \neq y'$, and $z = x + y = x' + y'$. Then, $(x - x') + (y - y') = \mathbf{0}$. Note that $x - x' = x + (-1)x' \in \mathcal{S}$ and $y - y' = y + (-1)y' \in \mathcal{S}^\perp$ since $\mathcal{S}$ and $\mathcal{S}^\perp$ are both subspaces. Therefore, $(x - x')^T(y - y') = (x - x')^T(x' - x) = 0$, which implies that $x - x' = \mathbf{0}$, or equivalently, $x = x'$, which is a contradiction. Therefore, z can be uniquely written as $z = x + y$, where $x \in \mathcal{S}$ and $y \in \mathcal{S}^\perp$. □

Proposition 7.3. *Let $\mathcal{S} \subseteq \mathbb{R}^n$ be a subspace. Then,*

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

Proof. Let $\{p^1, \ldots, p^r\} \subset \mathbb{R}^n$ and $\{q^1, \ldots, q^s\} \subset \mathbb{R}^n$ be a basis for $\mathcal{S}$ and $\mathcal{S}^\perp$, respectively. Therefore, $\dim(\mathcal{S}) = r$ and $\dim(\mathcal{S}^\perp) = s$. We claim that $\text{span}\left(\{p^1, \ldots, p^r, q^1, \ldots, q^s\}\right) = \mathbb{R}^n$. Let $z \in \mathbb{R}^n$ be an arbitrary vector. By Proposition 7.2, there exists a unique $x \in \mathcal{S}$ and a unique $y \in \mathcal{S}^\perp$ such that $z = x + y$. Since $x \in \text{span}\left(\{p^1, \ldots, p^r\}\right)$ and $y \in \text{span}\left(\{q^1, \ldots, q^s\}\right)$, our claim follows. Therefore, the set $\{p^1, \ldots, p^r, q^1, \ldots, q^s\}$ contains n linearly independent vectors, which implies that $n \leq r + s$.

Next, we claim that the set $\{p^1, \ldots, p^r, q^1, \ldots, q^s\}$ is linearly independent. Suppose, for a contradiction, that this is not true. Then, there exist real numbers $\alpha_i,\ i = 1, \ldots, r$ and $\beta_k,\ k = 1, \ldots, s$ such that they are not all equal to zero and

$$\mathbf{0} = \sum_{i=1}^{r} \alpha_i p^i + \sum_{k=1}^{s} \beta_k q^k.$$

Let $z \in \mathbb{R}^n$ be an arbitrary vector. Then, by Proposition 7.2, there exists a unique $x \in \mathcal{S}$ and a unique $y \in \mathcal{S}^\perp$ such that $z = x + y$. Let $x' = x + \sum_{i=1}^{r} \alpha_i p^i$ and $y' = y + \sum_{k=1}^{s} \beta_k q^k$. Note that $x' \in \mathcal{S}$, $y' \in \mathcal{S}^\perp$, and $z = x + y = x' + y'$ using the previous equation. By Proposition 7.2, we have $x = x'$ and $y = y'$, which implies that $\sum_{i=1}^{r} \alpha_i p^i = \sum_{k=1}^{s} \beta_k q^k = \mathbf{0}$. Since $\{p^1, \ldots, p^r\}$ and $\{q^1, \ldots, q^s\}$ are both linearly independent, we have $\alpha_i = 0$ for each $i = 1, \ldots, r$ and $\beta_k = 0$ for each $k = 1, \ldots, s$, which is a contradiction. Therefore, the set $\{p^1, \ldots, p^r, q^1, \ldots, q^s\}$ is linearly independent. □

Proposition 7.4 (Fundamental Theorem of Linear Algebra)**.** *Let $A \in \mathbb{R}^{m \times n}$. Then,*

$$\mathcal{R}(A)^\perp = \mathcal{N}(A^T).$$

Proof. We will show that $\mathcal{R}(A)^\perp \subseteq \mathcal{N}(A^T)$ and $\mathcal{N}(A^T) \subseteq \mathcal{R}(A)^\perp$, respectively.

$\mathcal{R}(A)^\perp \subseteq \mathcal{N}(A^T)$: Let $w \in \mathcal{R}(A)^\perp$ be an arbitrary vector. Then, for each $e^1, \ldots, e^n \in \mathbb{R}^n$, we have $Ae^i \in \mathcal{R}(A),\ i = 1, \ldots, n$. Therefore, $w^T(Ae^i) = (e^i)^T A^T w = 0$, for each $i = 1, \ldots, n$. This implies that the ith component of $A^T w$ is equal to zero for each $i = 1, \ldots, n$. Therefore, $A^T w = 0$, which implies that $w \in \mathcal{N}(A^T)$.

$\mathcal{N}(A^T) \subseteq \mathcal{R}(A)^\perp$: Let $w \in \mathcal{N}(A^T)$. Then, $A^T w = 0$. For any $y \in \mathcal{R}(A)$, there exists $u \in \mathbb{R}^m$ such that $y = Au$. Therefore, $w^T y = w^T(Au) = u^T(A^T w) = 0$, which implies that $w \in \mathcal{R}(A)^\perp$. □

Corollary 7.1. *Let $A \in \mathbb{R}^{m \times n}$. Then,*

1. $dim(\mathcal{R}(A)) + dim(\mathcal{N}(A^T)) = rank(A) + dim(\mathcal{N}(A^T)) = n.$
2. $dim(\mathcal{R}(A^T)) + dim(\mathcal{N}(A)) = rank(A) + dim(\mathcal{N}(A)) = n.$

Notes:

1. Let $A \in \mathbb{R}^{m\times n}$. Every $z \in \mathbb{R}^m$ can be uniquely decomposed as $z = x + y$, where $x \in \mathcal{R}(A)$ and $y \in \mathcal{N}(A^T)$.
2. Let $A \in \mathbb{R}^{m\times n}$. Every $z \in \mathbb{R}^n$ can be uniquely decomposed as $z = x + y$, where $x \in \mathcal{R}(A^T)$ and $y \in \mathcal{N}(A)$.

7.1 Systems of Equations

Let $A \in \mathbb{R}^{m\times n}$ and let $b \in \mathbb{R}^m$. Consider the system of linear equations:

$$Ax = b,$$

where $x \in \mathbb{R}^n$.

1. The system $Ax = b$ has at least one solution if and only if $b \in \mathcal{R}(A)$.
2. Suppose that $\hat{x} \in \mathbb{R}^n$ is a solution of the system $Ax = b$, i.e. $A\hat{x} = b$. Then, for every $d \in \mathcal{N}(A)$, $\hat{x} + d \in \mathbb{R}^n$ is also a solution of the system since $A(\hat{x} + d) = b + Ad = b$.
3. The system $Ax = b$ has exactly one solution if and only if $b \in \mathcal{R}(A)$ and $\mathcal{N}(A) = \{\mathbf{0}\}$ (i.e., A has full column rank).
4. The system $Ax = b$ has no solution (i.e., is inconsistent) if and only if $b \notin \mathcal{R}(A)$. This is true if and only if $\text{rank}([A\,|\,b]) > \text{rank}(A)$, where $[A\,|\,b] \in \mathbb{R}^{m\times(n+1)}$ is the matrix obtained by appending b to A.
5. The system $Ax = b$ has a solution for every $b \in \mathbb{R}^m$ if and only if $\text{rank}(A) = m$ (i.e., if A has full row rank). In this case, $\min\{m, n + 1\} \geq m \geq \text{rank}([A\,|\,b]) \geq \text{rank}(A) = m$, which implies that $\text{rank}([A\,|\,b]) = \text{rank}(A)$.
6. If the system $Ax = b$ has at least one solution and $\text{rank}(A) < n$, then it has infinitely many solutions since $\dim(\mathcal{N}(A)) = n - \text{rank}(A) > 0$, which implies that $\mathcal{N}(A) \neq \{\mathbf{0}\}$.
7. In summary, for a given system $Ax = b$, exactly one of the following three outcomes are possible: (i) There is no solution; (ii) there is a unique solution; (iii) there is an infinite number of solutions.

7.2 Representation of Subspaces

Let $\mathcal{S} \subseteq \mathbb{R}^n$ be a subspace and let $\dim(\mathcal{S}) = r$. Then, every basis of $\mathcal{S}$ contains exactly r vectors. Let $\{p^1, \ldots, p^r\} \subset \mathbb{R}^n$ be a basis of $\mathcal{S}$ and let

$$P = \begin{bmatrix} p^1 & p^2 & \cdots & p^r \end{bmatrix} \in \mathbb{R}^{n\times r}.$$

Then, $\mathcal{S} = \text{span}\left(\{p^1, \ldots, p^r\}\right) = \mathcal{R}(P)$.

Alternatively, $\mathcal{S} = (\mathcal{S}^\perp)^\perp$. By Proposition 7.4, $\mathcal{R}(P)^\perp = \mathcal{N}(P^T)$. Therefore,

$$\mathcal{S} = \mathcal{N}(P^T)^\perp = \{y \in \mathbb{R}^n : x^T y = 0 \text{ for all } x \in \mathcal{N}(P^T)\}.$$

Note that $\dim(\mathcal{N}(P^T)) = n - \dim(\mathcal{R}(P)) = n - r$. Let $\{q^1, \ldots, q^{n-r}\} \subset \mathbb{R}^n$ be a basis for $\mathcal{N}(P^T)$. Let $y \in \mathcal{S}$ be an arbitrary vector. Since $q^i \in \mathcal{N}(P^T)$ for each $i = 1, \ldots, n - r$, we have $(q^i)^T y = 0$ for each $i = 1, \ldots, n - r$. Therefore, if

$$Q = \begin{bmatrix} (q^1)^T \\ (q^2)^T \\ \vdots \\ (q^{n-r})^T \end{bmatrix} \in \mathbb{R}^{(n-r)\times n},$$

then $y \in S$ implies $Qy = \mathbf{0}$. Therefore, $\mathcal{S} \subseteq \mathcal{N}(Q)$.

Conversely, if $y \in \mathcal{N}(Q)$, then $Qy = \mathbf{0}$, which implies $(q^i)^T y = 0$ for each $i = 1, \ldots, n-r$. Let $x \in \mathcal{N}(P^T)$ be an arbitrary vector. Then, there exist real numbers $\lambda_1, \ldots, \lambda_{n-r}$ such that $x = \sum_{i=1}^{n-r} \lambda_i q^i$. Therefore, $x^T y = \sum_{i=1}^{n-r} \lambda_i (q^i)^T y = 0$, which implies that $y \in \mathcal{S}$. Therefore, $\mathcal{N}(Q) \subseteq \mathcal{S}$. It follows that

$$\mathcal{S} = \mathcal{R}(P) = \mathcal{N}(Q),$$

i.e., every subspace can be represented as the range space of a matrix, or alternatively, as the null space of a related matrix. Note that the relation between $\{p^1, \ldots, p^r\}$ and $\{q^1, \ldots, q^{n-r}\}$ is as follows. Since $p^j \in \mathcal{S}$ for each $j = 1, \ldots, r$, $q^i \in \mathcal{N}(P^T)$ for each $i = 1, \ldots, n-r$, and $\mathcal{S} = \mathcal{N}(P^T)^\perp$, we have

$$(q^i)^T p^j = 0, \quad i = 1, \ldots, n-r; \; j = 1, \ldots, r.$$

Intuitively, this derivation shows that each subspace can either be represented as the span of its basis or it can be represented as the set of vectors which are orthogonal to every vector in its orthogonal complement.
Example: Let

$$p^1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \in \mathbb{R}^3, \quad p^2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad P = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{3\times 2},$$

and let $\mathcal{S} = \text{span}(\{p^1, p^2\}) = \mathcal{R}(A)$. Let

$$q^1 = \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix} \in \mathbb{R}^3, \quad Q = \begin{bmatrix} 1 & -1 & -1 \end{bmatrix} \in \mathbb{R}^{1\times 3}.$$

Note that $(q^1)^T p^1 = (q^1)^T p^2 = 0$. Therefore, $\mathcal{S} = \mathcal{N}(Q) = \{y \in \mathbb{R}^3 : Qy = \mathbf{0}\}$.

8 Affine Subspaces

Definition 8.1. *Let $\mathcal{S} \subseteq \mathbb{R}^n$ be a subspace and let $\hat{x} \in \mathbb{R}^n$. Then, an* affine subspace *is given by*

$$\mathcal{S}_0 = \mathcal{S} + \{\hat{x}\} = \{w + \hat{x} : w \in \mathcal{S}\}.$$

The dimension of an affine subspace is given by the dimension of the underlying subspace, i.e., $dim(\mathcal{S}_0) = dim(\mathcal{S})$.

Notes:

1. If $\hat{x} = \mathbf{0} \in \mathbb{R}^n$, then $\mathcal{S}_0 = \mathcal{S}$. Therefore, every subspace is also an affine subspace. However, the converse is not necessarily true since an affine subspace need not contain $\mathbf{0} \in \mathbb{R}^n$.
2. An affine subspace is given by translating (i.e., shifting) a subspace by the vector $\hat{x}$.

Example : Let $\mathcal{S}_0 = \{x \in \mathbb{R}^2 : x_1 + x_2 = 1\}$. Then,

$$\mathcal{S}_0 = \{x \in \mathbb{R}^2 : x_1 + x_2 = 0\} + \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\} = \mathcal{S} + \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\},$$

where $\mathcal{S} = \{x \in \mathbb{R}^2 : x_1 + x_2 = 0\}$, which is a subspace. Therefore, $\mathcal{S}_0$ is an affine subspace. Furthermore, $\dim(\mathcal{S}_0) = \dim(\mathcal{S}) = 1$.

8.1 Representation of Affine Subspaces

Let $\mathcal{S}_0 \subseteq \mathbb{R}^n$ be an affine subspace given by

$$\mathcal{S}_0 = \mathcal{S} + \{\hat{x}\},$$

where $\mathcal{S} \subseteq \mathbb{R}^n$ is a subspace and let $\hat{x} \in \mathbb{R}^n$. Let $\dim(\mathcal{S}) = r$ and let $\{p^1, \ldots, p^r\} \subset \mathbb{R}^n$ be a basis of $\mathcal{S}$. Define

$$P = \begin{bmatrix} p^1 & p^2 & \cdots & p^r \end{bmatrix} \in \mathbb{R}^{n \times r}.$$

Then, $\mathcal{S}_0 = \mathcal{R}(P) + \{\hat{x}\} = \{x \in \mathbb{R}^n : x = \hat{x} + Pu \quad \text{for some} \quad u \in \mathbb{R}^r\}$.

Alternatively, let $\{q^1, \ldots, q^{n-r}\} \subset \mathbb{R}^n$ be a basis for $\mathcal{N}(P^T)$ and let

$$Q = \begin{bmatrix} (q^1)^T \\ (q^2)^T \\ \vdots \\ (q^{n-r})^T \end{bmatrix} \in \mathbb{R}^{(n-r) \times n}.$$

Then,

$$\mathcal{S}_0 = \mathcal{N}(Q) + \{\hat{x}\}.$$

Therefore, $x \in \mathcal{S}_0$ if and only there exists $y \in \mathcal{N}(Q)$ such that $x = y + \hat{x}$. If $x \in \mathcal{S}_0$, then $Qx = Qy + Q\hat{x} = Q\hat{x} = \hat{q}$. Therefore, if $x \in \mathcal{S}_0$, then x is a solution of the system $Qx = \hat{q}$, where $\hat{q} = Q\hat{x}$. Conversely, consider the system $Qx = \hat{q}$. Clearly, $x = \hat{x}$ is a solution of this system. Therefore, for any $d \in \mathcal{N}(Q)$, $x = \hat{x} + d$ is also a solution of this system since $Q(\hat{x} + d) = Q\hat{x} + \mathbf{0} = \hat{q}$. It follows that every solution of the system $Qx = \hat{q}$ can be written as $x = d + \hat{x}$, which implies that $x \in \mathcal{N}(Q) + \{\hat{x}\}$, or equivalently, $x \in \mathcal{S}_0$. Combining these two inclusions, we have

$$\mathcal{S}_0 = \{x \in \mathbb{R}^n : Qx = \hat{q}\}.$$

Note that $\dim(\mathcal{S}_0) = \dim(\mathcal{S}) = r$. Furthermore, $\text{rank}(Q) = n - r$ since the rows of Q are linearly independent (i.e., Q has full row rank). Therefore,

$$\dim(\mathcal{S}_0) = n - \text{rank}(Q) = n - (n - r) = r.$$

8.2 Relation with Systems of Equations

Let $A \in \mathbb{R}^{m \times n}$ and let $b \in \mathbb{R}^m$. Consider the system of linear equations:

$$Ax = b,$$

where $x \in \mathbb{R}^n$. Let

$$\mathcal{S}_0 = \{x \in \mathbb{R}^n : Ax = b\},$$

i.e., $\mathcal{S}_0$ is the set of all solutions of the system $Ax = b$. By the previous discussion, $\mathcal{S}_0$ is an affine subspace. Furthermore,

$$\dim(\mathcal{S}_0) = n - \text{rank}(A).$$

9 Square Matrices

Definition 9.1. *Let $A \in \mathbb{R}^{n \times n}$ be a square matrix. Then, A is* invertible *(or* nonsingular*) if there exists a square matrix $B \in \mathbb{R}^{n \times n}$ such that $AB = BA = I$, where I is the identity matrix given by*

$$I = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

If A is not invertible, it is called singular.

Proposition 9.1. *Let $A \in \mathbb{R}^{n\times n}$ be a nonsingular matrix. Then, the inverse of A is unique.*

Proof. Suppose, for a contradiction, that $B_1 \in \mathbb{R}^{n\times n}$ and $B_2 \in \mathbb{R}^{n\times n}$ are both inverses of A such that $B_1 \neq B_2$. Then, $AB_1 = B_1A = I$ and $AB_2 = B_2A = I$. Since $AB_1 = I$, multiplying both sides by B_2 from the left, we obtain $B_2AB_1 = B_2$. Since $B_2A = I$, we obtain $B_1 = B_2$, a contradiction. Therefore, the inverse of A is unique. □

Definition 9.2. *Let $A \in \mathbb{R}^{n\times n}$ be a nonsingular matrix. The unique inverse of A is denoted by $A^{-1} \in \mathbb{R}^{n\times n}$.*

Proposition 9.2. *Let $A \in \mathbb{R}^{n\times n}$ be a square matrix. A is nonsingular if and only if rank$(A) = n$ (i.e., A has both full row and full column rank).*

Proof. $\Rightarrow$: Let $A \in \mathbb{R}^{n\times n}$ be a nonsingular matrix. We claim that the system $Ax = b$ has a solution for each $b \in \mathbb{R}^m$. Indeed, $A^{-1}Ax = A^{-1}b$, which implies that $x = A^{-1}b$ is a solution of this system. It follows that A has full row rank, i.e., rank$(A) = n$.

$\Leftarrow$: Let $A \in \mathbb{R}^{n\times n}$ be a square matrix. Suppose that rank$(A) = n$. Consider the system

$$Ax = e^j, \quad j = 1, \ldots, n,$$

where $e^j \in \mathbb{R}^n$ denotes the jth unit vector, $j = 1, \ldots, n$. Since rank$(A) = n$, A has full row and column rank. Therefore, each of these n systems has a unique solution. Let $x^j \in \mathbb{R}^n$ denote the unique solution of the system $Ax = e^j,\ j = 1, \ldots, n$. Let

$$B = \begin{bmatrix} x^1 & \cdots & x^n \end{bmatrix}.$$

Then,

$$AB = \begin{bmatrix} Ax^1 & \cdots & Ax^n \end{bmatrix} = \begin{bmatrix} e^1 & \cdots & e^n \end{bmatrix} = I.$$

Now, we claim that rank$(B) = n$. Consider the following system:

$$\sum_{j=1}^{n} \lambda_j x^j = 0.$$

Multiplying both sides from the left by the matrix A, we obtain $\sum_{j=1}^n \lambda_j Ax^j = \sum_{j=1}^n \lambda_j e^j = 0$, which implies that $\lambda_1 = \ldots = \lambda_n = 0$. Therefore, $\{x^1, \ldots, x^n\}$ are linearly independent. It follows that B has full column rank, i.e., rank$(B) = n$. Since B is a square matrix, it also has full row rank. Following a similar argument, there exists a matrix $C \in \mathbb{R}^{n\times n}$ such that $BC = I$. Therefore, $ABC = A$, which implies that $C = A$ since $AB = I$. Therefore, $BC = BA = I$. Since $AB = BA = I$, it follows that $B = A^{-1}$. Therefore, A is nonsingular. □

Notes:

1. The argument in the proof of Proposition 9.2 shows that if A and B are two $n \times n$ matrices such that $AB = I$, then we also have $BA = I$. Therefore, $A^{-1} = B$ and $B^{-1} = A$.

2. Let $A \in \mathbb{R}^{n\times n}$ be a nonsingular matrix. Then, for every $b \in \mathbb{R}^m$, the system $Ax = b$ has a unique solution given by $x = A^{-1}b$.