1. The equation $A\mathbf{x} = \mathbf{v}$

Let $A \in \mathbb{R}^{m\times n}$. To describe the nature of the solution set of the equation $A\mathbf{x} = \mathbf{v}$, for $\mathbf{v} \in \mathbb{R}^m$, is governed by two sets:

Definition 1.1.

(1) The set
$$C(A) := \{A\mathbf{x} : \mathbf{x} \in \mathbb{R}^n\}$$
is called the *column space of* A.

(2) The set
$$N(A) = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{0}\}$$
is called the *nullspace of* A

Remark 1.2. As $A\mathbf{0} = \mathbf{0}$, we have $\mathbf{0} \in N(A)$ for any matrix A.

The set $C(A)$ governs the solubility of $A\mathbf{x} = \mathbf{v}$, while $N(A)$ determines the size of the corresponding set. More precisely:

Theorem 1.3.

(1) *The column space* $C(A)$ *consists of vectors* $\mathbf{v} \in \mathbb{R}^m$ *such that* $A\mathbf{x} = \mathbf{v}$ *has* at least one *solution.*

(2) $A\mathbf{x} = \mathbf{v}$ *has a solution for all* $\mathbf{v} \in \mathbb{R}^m$ *if and only if* $C(A) = \mathbb{R}^m$.

(3) *Let* $\mathbf{v} \in C(A)$, *so that* $A\mathbf{x} = \mathbf{v}$ *has at least one solution,* $\mathbf{x}_0$, *say. Then*
$$\{\mathbf{x} : A\mathbf{x} = \mathbf{v}\} = \{\mathbf{x}_0 + \mathbf{y} : \mathbf{y} \in N(A)\}. \tag{$*$}$$

(4) *Let* $\mathbf{v} \in C(A)$. *Then* $A\mathbf{x} = \mathbf{v}$ *has a unique solution if and only if* $N(A) = \{\mathbf{0}\}$.

Proof.

(1) Statement (1) follows, essentially, from the definition of $C(A)$. If $\mathbf{v} \in C(A)$, then $\mathbf{v} = A\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^n$, so $\mathbf{x}$ is a solution of $A\mathbf{x} = \mathbf{v}$. Conversely, if $\mathbf{v} \notin C(A)$, then there is no vector $\mathbf{x} \in \mathbb{R}^m$ such that $A\mathbf{x} = \mathbf{v}$ i.e., $A\mathbf{x} = \mathbf{v}$ has no solution.

(2) follows directly from (1).

(3) Let $\mathbf{v} \in C(A)$. We must prove $(*)$. Let $\mathbf{x} \in \{\mathbf{x} : A\mathbf{x} = \mathbf{v}\}$. Then
$$A(\mathbf{x} - \mathbf{x}_0) = A\mathbf{x} - A\mathbf{x}_0 = \mathbf{v} - \mathbf{v} = \mathbf{0}.$$
Thus, $y := \mathbf{x} - \mathbf{x}_0 \in N(A)$ and $\mathbf{x} = \mathbf{x}_0 + \mathbf{y}$. Therefore, $\mathbf{x} \in \{\mathbf{x}_0 + \mathbf{y} : \mathbf{y} \in N(A)\}$. Since $\mathbf{x} \in \{\mathbf{x} : A\mathbf{x} = \mathbf{v}\}$ was arbitrary,
$$\{\mathbf{x} : A\mathbf{x} = \mathbf{v}\} \subseteq \{\mathbf{x}_0 + \mathbf{y} : \mathbf{y} \in N(A)\}.$$

Conversely, suppose $\mathbf{x} \in \{\mathbf{x}_0 + \mathbf{y} : \mathbf{y} \in N(A)\}$, i.e., $\mathbf{x} = \mathbf{x}_0 + \mathbf{y}$, for some $\mathbf{y} \in N(A)$. Now $A\mathbf{x}_0 = \mathbf{0}$ by hypothesis and $A\mathbf{y} = \mathbf{0}$ by definition of $N(A)$. Therefore,

$$A\mathbf{x} = A(\mathbf{x}_0 + y) = A\mathbf{x}_0 + A\mathbf{y} = \mathbf{0} + \mathbf{0} = \mathbf{0}.$$

Thus, $\mathbf{x} \in N(A)$. Since $\mathbf{x} \in \{\mathbf{x}_0 + \mathbf{y} : \mathbf{y} \in N(A)\}$ was arbitrary,

$$\{\mathbf{x}_0 + \mathbf{y} : \mathbf{y} \in N(A)\} \subseteq \{\mathbf{x} : A\mathbf{x} = \mathbf{v}\}.$$

Identity $(*)$ follows.

(4) Let $\mathbf{v} \in C(A)$ and let $\mathbf{x}_0 \in \mathbb{R}^n$ be some solution of $A\mathbf{x} = \mathbf{v}$. By (3), the set of all solutions of $A\mathbf{x} = \mathbf{v}$ is

$$\{\mathbf{x}_0 + \mathbf{y} : \mathbf{y} \in N(A)\}.$$

If $N(A) = \{\mathbf{0}\}$, then

$$\{\mathbf{x}_0 + \mathbf{y} : \mathbf{y} \in N(A)\} = \{\mathbf{x}_0 + \mathbf{0}\} = \{\mathbf{x}_0\},$$

in which case $\mathbf{x}_0$ is the unique solution of $A\mathbf{x} = \mathbf{v}$.

If, conversely, $N(A) \neq \{\mathbf{0}\}$, then $N(A)$ must contain a nonzero vector, y. ($N(A)$ cannot be empty by Remark 1.2.) Set $\mathbf{x}_1 := \mathbf{x}_0 + \mathbf{y}$. Then $\mathbf{x}_1 \neq \mathbf{x}_0$ as $\mathbf{y} \neq \mathbf{0}$ and $A\mathbf{x}_1 = \mathbf{0}$ as $\mathbf{y} \in N(A)$. Thus, $A\mathbf{x} = \mathbf{v}$ has at least two solutions. □

To compute the sets $C(A)$ and $N(A)$, we use the *reduced row echelon form* of A.

2. Reduced row echelon form

Definition 2.1. A matrix R is in *reduced row echelon form* if:

(1) All zero rows of R are at the bottom.
(2) Every nonzero row has a leading one.
(3) A leading one is the right of those in the rows above it.
(4) A leading one is the only nonzero entry in its column.

Definition 2.2. Let $A, B \in \mathbb{R}^{m\times n}$. We say that A *and* B *are row equivalent* or that A *is row equivalent to* B if there is an *invertible* matrix $\gamma \in \mathbb{R}^{m\times m}$ such that $\gamma A = B$.

Remarks 2.3.

(1) Since $\gamma A = B$ if and only if $\gamma^{-1}B = A$, the notion of row equivalence is symmetric in A and B: A is row equivalent to B if and only if B is row equivalent to A.
(2) A and B are row equivalent if and only if A can be transformed into B via a sequence of elementary row operations. The theory of elementary matrices connects this formulation of row equivalence with that of Definition 2.2.

Theorem 2.4. *A matrix A is row equivalent to a unique matrix in reduced row echelon form.*

3. Linear combinations, span, and the column space

Let

$$A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \cdots \mathbf{a}_n \end{bmatrix} \in \mathbb{R}^{m \times n}.$$

The theory of the equation $A\mathbf{x} = \mathbf{v}$ — an algebraic theory, so far — can be understood in terms of the geometry of the column vectors $\mathbf{a}_1, \dots, \mathbf{a}_n$ of A and the vector $\mathbf{v}$. This geometric perspective is well worth developing: It lets us to apply algebraic methods to geometry. Even more valuable, perhaps, it gives us a framework for thinking geometrically (visually) about algebra. There is no free lunch, however: In developing a geometric point of view on linear algebra, we incur some overhead, mainly in the form of new terminology.

Definition 3.1. A *linear combination of the vectors* $\mathbf{a}_1, \dots, \mathbf{a}_n$ is an expression of the form

$$x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n,$$

where $t_j \in \mathbb{R}$.

Example-Definition 3.2. The zero vector can be expressed as a linear comination of any sequence of vectors $\mathbf{a}_1, \dots, \mathbf{a}_n$:

$$\mathbf{0} = 0\mathbf{a}_1 + \cdots + 0\mathbf{a}_n.$$

The expression on the right hand side is called the *trivial linear combination* of $\mathbf{a}_1, \dots, \mathbf{a}_n$.

Example-Definition 3.3. Let $\mathbf{i}_j \in \mathbb{R}^m$ be the j-th column of the identity matrix $I \in \mathbb{R}^{m \times m}$:

$$\mathbf{i}_j := \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{matrix} \\ \\ \\ \leftarrow j\text{-th row.} \\ \\ \\ \\ \end{matrix}$$

It's called the *j-th standard basis vector of* $\mathbb{R}^m$. Any vector $\mathbf{v} \in \mathbb{R}^m$ can be expressed as a linear combination of these:

$$\begin{aligned}
\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{bmatrix} &= \begin{bmatrix} v_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ v_2 \\ \vdots \\ 0 \end{bmatrix} + \cdots + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ v_m \end{bmatrix} \\
&= v_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + v_2 \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \cdots + v_m \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \\
&= v_1\mathbf{i}_1 + v_2\mathbf{i}_2 + \cdots + v_m\mathbf{i}_m
\end{aligned}$$

Here's a trivial, yet crucial, observation: By the definition of matrix multiplication,

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

We record (a paraphrase of) this simple fact, prominently, for ease of reference.

Key fact 3.4. Linear combinations are just matrix-vector products.

Definitions 3.5. The *span* of the vectors $\mathbf{a}_1, \dots, \mathbf{a}_n$, written $\langle \mathbf{a}_1, \dots, \mathbf{a}_n \rangle$, is the set of their linear combinations:

$$\langle \mathbf{a}_1, \dots, \mathbf{a}_n \rangle := \{x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n : x_1, \dots, x_n \in \mathbb{R}\}.$$

Example 3.6. Let $\mathbf{i}_1, \dots, \mathbf{i}_m$ be the standard basis of $\mathbb{R}^m$, as in Example-Definition 3.3. As every vector $\mathbf{v} \in \mathbb{R}^m$ can be expressed as a linear comnination of $\mathbf{i}_1, \dots, \mathbf{i}_m$, we have

$$\langle \mathbf{i}_1, \dots, \mathbf{i}_m \rangle = \mathbb{R}^m.$$

Example 3.7. By Key fact 4.3 and Definition 1.1, we have:

$$C(A) = \langle \mathbf{a}_1, \dots, \mathbf{a}_n \rangle. \tag{1}$$

Theorem 3.8. *Let A be its matrix and let $R = \operatorname{rref} A$.*

(1) *The equation $A\mathbf{x} = \mathbf{v}$ has a solution for all vectors $\mathbf{v}$ if and only if every row of R has a leading one.*
(2) *The equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution, $\mathbf{x} = \mathbf{0}$, if and only if every column of R has a leading one.*

Theorem 3.8 suggests taking a closer look at the following two sets of vectors associated to a matrix $A \in \mathbb{R}^{m\times n}$:

Definitions 3.9.

(1) The set of all vectors $\mathbf{v} \in \mathbb{R}^m$ such $A\mathbf{x} = \mathbf{v}$ has a solution is called the *column space of A* and written $C(A)$:
$$C(A) = \{A\mathbf{x} : \mathbf{x} \in \mathbb{R}^n\}.$$
(2) The set of all vectors $\mathbf{x} \in \mathbb{R}^n$ such that $A\mathbf{x} = \mathbf{0}$ is called the *nullspace of A* and written $N(A)$:
$$N(A) = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{0}\}.$$

Corollary 3.10. *Let $A \in \mathbb{R}^{m\times n}$.*

(1) *If $A\mathbf{x} = \mathbf{v}$ has a solution for all $\mathbf{v} \in \mathbb{R}^m$, i.e., if*
$$C(A) = \mathbb{R}^m,$$
then $m \le n$.
(2) *If $A\mathbf{x} = \mathbf{0}$ has only the trivial solution, i.e., if*
$$N(A) = \{\mathbf{0}\},$$
then $n \le m$.

Corollary 3.11. *The following are equivalent for a square matrix $A \in \mathbb{R}^{m\times m}$.*

(1) *$A\mathbf{x} = \mathbf{v}$ has a solution for all vectors $\mathbf{v}$.*
(2) *$C(A) = \mathbb{R}^m$*
(3) *$A\mathbf{x} = \mathbf{0}$ has only the trivial solution.*
(4) *$N(A) = \{\mathbf{0}\}$*
(5) *$\operatorname{rref} A = I$*

(6) *A is invertible.*

Later, we'll give an intrinsic definition of the following essential notion:

Provisional definition 3.12. A *subspace of* $\mathbb{R}^m$ is a subset of the form $C(A)$, where $A \in \mathbb{R}^{m\times n}$, or $N(B)$, where $B \in \mathbb{R}^{\ell\times m}$.

Row equivalence

Uniqueness of rref

Pivot columns

4. Linear combinations

Definition 4.1. A *linear combination* of vectors $\mathbf{a}_1, \dots, \mathbf{a}_n$ is an expression of the form

$$x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n,$$

where $t_j \in \mathbb{R}$.

Example-Definition 4.2. The zero vector can be written as a linear comination of any sequence of vectors $\mathbf{a}_1, \dots, \mathbf{a}_n$:

$$\mathbf{0} = 0\mathbf{a}_1 + \cdots + 0\mathbf{a}_n.$$

The expression on the right hand side is called the *trivial linear combination* of $\mathbf{a}_1, \dots, \mathbf{a}_n$.

Here's a trivial, yet crucial, observation. Let

$$A = \begin{bmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_n\end{bmatrix} \in \mathbb{R}^{m\times n}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}.$$

Then, by the definition of matrix multiplication,

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

We record (a paraphrase of) this simple fact, prominently, for ease of reference.

Key fact 4.3. Linear combinations are just matrix-vector products.

5. Span

Definitions 5.1. The *span* of a sequence $\mathbf{a}_1, \dots, \mathbf{a}_n$ of vectors in $\mathbb{R}^m$, written $\langle \mathbf{a}_1, \dots, \mathbf{a}_n \rangle$, is the set of all linear combinations of these vectors:

$$\langle \mathbf{a}_1, \dots, \mathbf{a}_n \rangle := \{x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n : x_1, \dots, x_n \in \mathbb{R}\}.$$

We say $\mathbf{a}_1, \dots, \mathbf{a}_n$ *span* $\mathbb{R}^m$ if every vector $\mathbf{v} \in \mathbb{R}^m$ can be written as such a linear combination, i.e., if

$$\langle \mathbf{a}_1, \dots, \mathbf{a}_n \rangle = \mathbb{R}^m.$$

If

$$A = \begin{bmatrix}\mathbf{a}_1 & \dots & \mathbf{a}_n\end{bmatrix} \in \mathbb{R}^{m\times n},$$

we abbreviate to $\langle \mathbf{a}_1, \dots, \mathbf{a}_n \rangle$ to $C(A)$, and refer to this set as the *column space* of A:

$$C(A) := \langle \mathbf{a}_1, \dots, \mathbf{a}_n \rangle.$$

Note that, by Key Fact 4.3,

$$C(A) = \{A\mathbf{x} : \mathbf{x} \in \mathbb{R}^n\}.$$

Theorem 5.2. *The following are equivalent for a matrix $A \in \mathbb{R}^{m \times n}$.*

(1) $C(A) = \mathbb{R}^m$
(2) *The equation $A\mathbf{x} = \mathbf{v}$ has a solution for all $\mathbf{v}$.*
(3) rref A *has a leading one in every row.*

Moreover, if any (hence, all) of these conditions hold, then $m \leq n$.

6. Linear independence

Definition 6.1. Vectors $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^m$ are *linearly independent* if the only linear combination of $\mathbf{a}_1, \dots, \mathbf{a}_n$ that equals $\mathbf{0}$ is the trivial one (see Example-Definition 4.2).