1. Linear combinations and spans

Definition 1.1. A *linear combination* of vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ is a vector of the form

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

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

The set of all linear combinations of $\mathbf{a}_1, \ldots, \mathbf{a}_n$ is called the *span of* $\mathbf{a}_1, \ldots, \mathbf{a}_n$ and written $\langle \mathbf{a}_1, \ldots, \mathbf{a}_n \rangle$:

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

If V is a set of vectors in $\mathbb{R}^m$, we say that vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ *span* U if

$$\langle \mathbf{a}_1, \ldots, \mathbf{a}_n \rangle = U.$$

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

where

$$A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix},$$

we deduce:

Theorem 1.2. *Suppose $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}$. Then the span of the column vector of A is the set of matrix-vector products $A\mathbf{x}$ for $\mathbf{x} \in \mathbb{R}^n$:*

$$\langle \mathbf{a}_1, \ldots, \mathbf{a}_n \rangle = \{A\mathbf{x} : \mathbf{x} \in \mathbb{R}^n\}.$$

2. Bases and basis matrices of $\mathbb{R}^m$

Let

Definition 2.1. We say that vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ form a *basis of* $\mathbb{R}^m$ if every vector $\mathbf{v} \in \mathbb{R}^m$ can be written uniquely as a linear combination of the columns of A, i.e., for every vector $\mathbf{v} \in \mathbb{R}^m$, there are unique scalars, $x_1, \ldots, x_n$,

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

If the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ form a basis of $\mathbb{R}^m$, we say that the matrix

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

is a *basis matrix of* $\mathbb{R}^m$.

The following result follows from the fundamental correspondence between linear combinations and matrix-vector products (Theorem 1.2):

Theorem-Definition 2.1. *A is a basis matrix of $\mathbb{R}^m$ if and only if, for every $\mathbf{v} \in \mathbb{R}^m$, the equation*

$$A\mathbf{x} = \mathbf{v} \tag{1}$$

has a unique solution. This solution is called the coordinate vector of $\mathbf{v}$ with respect to A *and is written* $[\mathbf{v}]_A$.

Theorem 2.2. *Suppose $A \in \mathbb{R}^{m\times m}$ is invertible. Then A is a basis matrix of $\mathbb{R}^n$.*

Proof. If A is invertible, then $\mathbf{x} := A^{-1}\mathbf{v}$ is the unique solution to $A\mathbf{x} = \mathbf{v}$. □

Below, we will show that every basis matrix of $\mathbb{R}^m$ is (square and) invertible.

3. Linear independence

Let $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$. The condition under which vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ form a basis of $\mathbb{R}^m$ is the combination of two weaker subconditions:

(1) Every vector $\mathbf{v} \in \mathbb{R}^m$ can be written in *at least one way* as a linear combination of the $\mathbf{a}_j$.
(2) Every vector $\mathbf{v} \in \mathbb{R}^m$ can be written in *at most one way* as a linear combination of the $\mathbf{a}_j$.

We introduce some terminology, useful in exploring these subconditions.

Definition 3.1. The set of all linear combinations of $\mathbf{a}_1, \ldots, \mathbf{a}_n$ is called the *span of* $\mathbf{a}_1, \ldots, \mathbf{a}_n$ and written $\langle \mathbf{a}_1, \ldots, \mathbf{a}_n \rangle$:

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

If V is a set of vectors in $\mathbb{R}^m$, we say that vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ *span* U if

$$\langle \mathbf{a}_1, \ldots, \mathbf{a}_n \rangle = U.$$

Condition (1), above, says that the vectors $\mathbf{a}$ span $\mathbb{R}^m$.

Definition 3.2. The vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are *linearly independent* if the identity

$$x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{0}$$

holds only when the scalars x_j are all zero.

Theorem 3.3. *Condition (2) holds if and only if the vectors* $\mathbf{a}_1, \ldots, \mathbf{a}_n$ *are linearly independent.*

Proof. Suppose condition (2) holds. We need to show that the $\mathbf{a}_1, \ldots, \mathbf{a}_{nj}$ are linearly independent. To that end, consider the equation

$$(*) \qquad x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{0}.$$

It always has at least one solution, namely $x_j = 0$, for all j. By condition (2), the vector $\mathbf{0}$ can be written in at most one way as a linear combination of the $\mathbf{a}_j$. In other words, $(*)$ has at most one solution. Thus, $(*)$ has exactly one solution: $x_j = 0$, for all j. Therefore, the $\mathbf{a}_j$ are linearly independent.

Conversely, suppose the $\mathbf{a}_j$ are linearly independent. We need to show that condition (2) holds. To that end, let $\mathbf{v} \in \mathbb{R}^m$. To prove that $\mathbf{v}$ can be written in at most one way as a linear combination of the $\mathbf{a}_j$, suppose

$$(\dagger) \qquad x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{v}.$$

and

$$(\ddagger) \qquad y_1\mathbf{a}_1 + \cdots + y_n\mathbf{a}_n = \mathbf{v}.$$

Subtracting $(\ddagger)$ from $(\dagger)$, we get

$$(x_1 - y_1)\mathbf{a}_1 + \cdots + (x_n - y_n)\mathbf{a}_n = \mathbf{0}.$$

As the $\mathbf{a}_j$ are linearly indepenent, by hypothesis, we must have

$$x_j - y_j = 0, \quad \text{for all } j.$$

Thus, $x_j = y_j$ for all j, and the two ostensibly different expressions $(\dagger)$ and $(\ddagger)$ for $\mathbf{v}$ as linear combinations of the $\mathbf{a}_j$ are, in fact, the same. Condition (2) follows.

Corollary 3.4. *A is a basis matrix of $\mathbb{R}^m$ if and only if its columns vectors are linearly independent and they span $\mathbb{R}^m$.*

$\square$

4. Reduced row echelon form

In the next few sections, we will prove some important results concerning the numerology of systems of linear equations. They imply, in particular, that basis matrices are *square*. The main tool in the proofs of these results is the following (hopefully) familiar notion:

Definition 4.1. 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.

The next theorem paraphrases the fact that a matrix can be brought to reduced row echelon form via a sequence of elementary row operations, and that these operations can be effected by matrix multiplication.

Theorem 4.2. *Let $A \in \mathbb{R}^{m\times n}$. Then there is an invertible matrix $\gamma \in \mathbb{R}^{m\times m}$ such that γA is in reduced row echelon form.*

Theorem 4.3. *Let A be its matrix and let $R = \operatorname{rref} A$.*

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

5. Linear independent sets are small

Theorem 5.1. *Let $A \in \mathbb{R}^{m\times n}$. Suppose $m < n$. Then there is a* nonzero *vector $\mathbf{x} \in \mathbb{R}^n$ such that the equation $A\mathbf{x} = \mathbf{0}$.*

Proof. Invoking Theorem 5.1, let $\gamma \in \mathbb{R}^{m\times m}$ be an invertible matrix such that $R := \gamma A$ is in reduced row echelon form. Since R has m rows, at most m of its n columns can contain leading ones. Since $m < n$, by hypothesis, at least one column of R does not contain a leading one; suppose the leftmost such column is the j-th. Then

$$R = \begin{bmatrix} 1 & 0 & \cdots & 0 & r_{1,j} & * & * & \cdots & * \\ 0 & 1 & \cdots & 0 & r_{2,j} & * & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & r_{j-1,j} & * & * & \cdots & * \\ 0 & 0 & \cdots & 0 & 0 & * & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & * & * & \cdots & * \end{bmatrix}.$$

Let

$$\mathbf{x}_1 := \begin{bmatrix} -r_{1,j} \\ -r_{2,j} \\ \vdots \\ -r_{j-1,j} \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \neq \mathbf{0}.$$

You can verify (exercise!) that $R\mathbf{x} = \mathbf{0}$. Multiplying both sides of this identity by γ^{-1}, we get

$$A\mathbf{x} = \gamma^{-1}R\mathbf{x} = \gamma^{-1}\mathbf{0} = \mathbf{0}.$$

Thus, $\mathbf{x}_j$ is a nonzero solution of $A\mathbf{x} = \mathbf{0}$. □

Remark 5.2. The solution $\mathbf{x}_1$ arises from setting $t_1 = 1$, $t_j = 0$ for $j \neq 1$, in the parametric solution of $A\mathbf{x} = \mathbf{0}$.

Corollary 5.3. *Let $A \in \mathbb{R}^{m\times n}$. If the column vectors of A are linearly independent, then $n \leq m$.*

Proof. The statement of the preceding theorem is equivalent to the statement that if $A\mathbf{x} = \mathbf{v}$ has a solution for all $\mathbf{v} \in \mathbb{R}^m$, then $m \leq n$. Saying that $A\mathbf{x} = \mathbf{v}$ has a solution for all $\mathbf{v}$ is another way of saying that the column vectors of A span $\mathbb{R}^m$. □

6. Spanning sets are big

Theorem 6.1. *Let $A \in \mathbb{R}^{m\times n}$. Suppose $n < m$. Then there is a vector $\mathbf{v} \in \mathbb{R}^m$ such that the equation $A\mathbf{x} = \mathbf{v}$ has no solution.*

Proof. Invoking Theorem 5.1, let $\gamma \in \mathbb{R}^{m\times m}$ be an invertible matrix such that $R := \gamma A$ is in reduced row echelon form. Since R has n columns and each column contains at most a single leading one, at most n of its rows m can contain leading ones. As $n < m$, by hypothesis, at least one row of R does not contain a leading one and, as R is in reduced row echelon form, must therefore be a zero row.

Since R is in reduced row echelon form and has a zero row, its bottom row must be a zero row. Therefore,

$$R\mathbf{x} = \begin{bmatrix} * \\ * \\ \vdots \\ * \\ 0 \end{bmatrix},$$

for all $\mathbf{x} \in \mathbb{R}^n$. It follows that the equation $R\mathbf{x} = \mathbf{i}_m$ has no solution. Set

$$\mathbf{v} := \gamma^{-1}\mathbf{i}_m.$$

As γ is invertible and $\gamma A = R$, the equations

$$A\mathbf{x} = \mathbf{v} \quad \text{and} \qquad R\mathbf{x} = \mathbf{i}_m$$

have the same solution set. Since $R\mathbf{x} = \mathbf{i}_m$ has no solution, neither does $A\mathbf{x} = \mathbf{v}$. □

Corollary 6.2. *Let $A \in \mathbb{R}^{m\times n}$. If the column vectors of A span $\mathbb{R}^m$, then $m \leq n$.*

Proof. The statement of the preceding theorem is equivalent to the statement that if $A\mathbf{x} = \mathbf{v}$ has a solution for all $\mathbf{v} \in \mathbb{R}^m$, then $m \leq n$. Saying that $A\mathbf{x} - = \mathbf{v}$ has a solution for all $\mathbf{v}$ is another way of saying that the column vectors of A span $\mathbb{R}^m$. □

7. Basis matrices are square

Theorem 7.1. *Let $A \in \mathbb{R}^{m\times n}$ be a basis matrix of $\mathbb{R}^m$. Then $m = n$.*

Proof. Combine Corollaries 5.3 and 6.2. □

8. Big linearly independent sets span

Theorem 8.1. *Suppose $A \in \mathbb{R}^{m\times n}$ has linearly independent columns. Then*

$$\operatorname{rref} A = \begin{bmatrix} I_n \\ \mathbf{0} \end{bmatrix}.$$

Theorem 8.2. *Suppose R is a matrix in rref. Then R has a column with no leading one in it if and only if $R\mathbf{x} = \mathbf{0}$ has a nontrivial solution.*

Suppose R is a matrix in rref. Then R has a zero row if and only if there is a $\mathbf{v}$ such that $A\mathbf{x} = \mathbf{v}$ has no solution.

9. Small spanning sets are linearly independent

10. Basis matrices are invertible

In this section, we prove that the A being a basis matrix is equivalent to A being invertible. This is plausible, thanks to Theorem **??**.