Lecture Notes for the Course

MAA111E Linear Algebra for Science Students

V. Ravichandran
School of Mathematical Sciences
Universiti Sains Malaysia

Contents

CHAPTER 1

Euclidean Space

Let n be a positive integer. A sequence of n real numbers $(a_1, a_2, a_3, \dots, a_n)$ is called an **ordered n-tuple**. The set of all ordered n-tuples, denoted by $\mathbb{R}^n$, is called the **Euclidean n-space**. We use the terms **ordered pair** for 2-tuple and **ordered triple** for 3-tuple. We call the elements of $\mathbb{R}^n$ as points in $\mathbb{R}^n$ or as vectors in $\mathbb{R}^n$.

DEFINITION 1.1. Let $\mathbf{u} = (u_1, u_2, u_3, \dots, u_n)$, $\mathbf{v} = (v_1, v_2, v_3, \dots, v_n)$ be two vectors in $\mathbb{R}^n$. The two vectors $\mathbf{u}$ and $\mathbf{v}$ are equal if

$$u_1 = v_1, \quad u_2 = v_2, \dots, \quad u_n = v_n.$$

The sum $\mathbf{u} + \mathbf{v}$ and the difference $\mathbf{u} - \mathbf{u}$ are defined by

$$\mathbf{u} + \mathbf{v} := (u_1 + v_1, u_2 + v_2, \dots, u_n + v_n)$$
$$\mathbf{u} - \mathbf{v} := (u_1 - v_1, u_2 - v_2, \dots, u_n - v_n)$$

and the scalar multiple $\lambda\mathbf{u}$ (where λ is a real number) is defined by

$$\lambda\mathbf{u} := (\lambda u_1, \lambda u_2, \dots, \lambda u_n).$$

The zero vector, denoted by $\mathbf{0}$, is defined by

$$\mathbf{0} = (0, 0, \dots, 0).$$

When $\lambda = -1$, the scalar multiple $-\mathbf{u}$ given by

$$-\mathbf{u} = (-u_1, -u_2, -u_3, \dots, -u_n)$$

is called the negative of $\mathbf{u}$ or the additive inverse of $\mathbf{u}$. Note that $\mathbf{u} - \mathbf{v} = \mathbf{u} + (-\mathbf{v})$ and $\mathbf{u} - \mathbf{u} = \mathbf{0}$.

THEOREM 1.1. *Let* $\mathbf{u}$, $\mathbf{v}$ *and* $\mathbf{w}$ *be vectors in* $\mathbb{R}^n$ *and* λ *and* μ *be scalars. Then we have the following*

(1) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
(2) $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
(3) $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}$
(4) $\mathbf{u} + (-\mathbf{u}) = (-\mathbf{u}) + \mathbf{u} = \mathbf{0}$
(5) $\lambda(\mu\mathbf{u}) = (\lambda\mu)\mathbf{u}$
(6) $\lambda(\mathbf{u} + \mathbf{v}) = \lambda\mathbf{u} + \lambda\mathbf{v}$
(7) $(\lambda + \mu)\mathbf{u} = \lambda\mathbf{u} + \mu\mathbf{u}$
(8) $1\mathbf{u} = \mathbf{u}$

(1) is the commutative law for addition; (2) is the associative law for addition; $\mathbf{0}$ is the identity element for addition; $-\mathbf{u}$ is additive inverse of $\mathbf{u}$.

PROOF. We prove only the commutative law for addition. Let

$$\mathbf{u} = (u_1, u_2, u_3, \dots, u_n), \quad \mathbf{v} = (v_1, v_2, v_3, \dots, v_n).$$

Then

$$\begin{aligned}\mathbf{u}+\mathbf{v} &= (u_1, u_2, u_3, \ldots, u_n) + (v_1, v_2, v_3, \ldots, v_n)\\ &= (u_1+v_1, u_2+v_2, \ldots, u_n+v_n)\\ &= (v_1+u_1, v_2+u_2, \ldots, v_n+u_n)\\ &= \mathbf{v}+\mathbf{u}.\end{aligned}$$

□

EXAMPLE 1.1. *Let* $\mathbf{u} = (1, 3, -1)$, $\mathbf{v} = (4, -2, 5)$ *be two vectors in* $\mathbb{R}^3$. *Then*

$$\begin{aligned}\mathbf{u}+\mathbf{v} &= (1, 3, -1) + (4, -2, 5)\\ &= (1+4, 3+(-2), -1+5) = (5, 1, 4),\\ \mathbf{v}+\mathbf{u} &= (4, -2, 5) + (1, 3, -1) = (5, 1, 4),\\ 3\mathbf{u} &= 3(1, 3, -1) = (3, 9, -3),\\ 3\mathbf{v} &= 3(4, -2, 5) = (12, -6, 15),\\ 3\mathbf{u}+3\mathbf{v} &= (3, 9, -3) + (12, -6, 15) = (15, 3, 12),\\ 3(\mathbf{u}+\mathbf{v}) &= 3(5, 1, 4) = (15, 3, 12).\end{aligned}$$

Clearly $\mathbf{u}+\mathbf{v} = \mathbf{v}+\mathbf{u}$, $3(\mathbf{u}+\mathbf{v}) = 3\mathbf{u}+3\mathbf{v}$. *Also*

$$\mathbf{u}-\mathbf{v} = (1, 3, -1) - (4, -2, 5) = (-3, 5, -6).$$

EXAMPLE 1.2. *If* $\mathbf{x}+\mathbf{u} = \mathbf{v}$, *then by adding* $-\mathbf{u}$ *both sides of the equation, we get*

$$(\mathbf{x}+\mathbf{u}) + (-\mathbf{u}) = \mathbf{v} + (-\mathbf{u}).$$

By using associative law, we get

$$\mathbf{x} + (\mathbf{u} + (-\mathbf{u})) = \mathbf{v} - \mathbf{u}.$$

Thus we have

$$\mathbf{x} + \mathbf{0} = \mathbf{v} - \mathbf{u}$$

or

$$\mathbf{x} = \mathbf{v} - \mathbf{u}.$$

The generalization of dot product of vectors on $\mathbb{R}^3$ to $\mathbb{R}^n$ is given in the following:

DEFINITION 1.2. The Euclidean inner product of two vectors $\mathbf{u} = (u_1, u_2, u_3, \ldots, u_n)$, $\mathbf{v} = (v_1, v_2, v_3, \ldots, v_n)$ in $\mathbb{R}^n$, denoted by $\mathbf{u}\cdot\mathbf{v}$, is defined by

$$\mathbf{u}\cdot\mathbf{v} = u_1v_1 + u_2v_2 + \cdots + u_nv_n = \sum_{i=1}^{n} u_iv_i.$$

If we write the vectors $\mathbf{u}$, $\mathbf{v}$ as a column vectors

$$\mathbf{u} = \begin{pmatrix} u_1\\ u_2\\ \vdots\\ u_n\end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} v_1\\ v_2\\ \vdots\\ v_n\end{pmatrix},$$

then

$$\mathbf{u}^T\mathbf{v} = \begin{pmatrix} u_1 & u_2 & \cdots & u_n \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$
$$= [u_1v_1 + u_2v_2 + \cdots + u_nv_n]$$
$$= [\mathbf{u} \cdot \mathbf{v}] = \mathbf{u} \cdot \mathbf{v}.$$

Also it follows that $\mathbf{v}^T\mathbf{u} = \mathbf{u} \cdot \mathbf{v}$. Thus we have

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T\mathbf{v} = \mathbf{v}^T\mathbf{u}.$$

THEOREM 1.2. *If* $\mathbf{u}$, $\mathbf{v}$ *and* $\mathbf{w}$ *are vectors in* $\mathbb{R}^n$ *and* k *a scalar, then*

(1) $\mathbf{u}.\mathbf{v} = \mathbf{v} \cdot \mathbf{u}$
(2) $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$
(3) $(\lambda\mathbf{u}) \cdot \mathbf{v} = \lambda(\mathbf{u} \cdot \mathbf{v})$
(4) $\mathbf{u} \cdot \mathbf{u} \geq 0$ *and* $\mathbf{u} \cdot \mathbf{u} = 0$ *if and only if* $\mathbf{u} = \mathbf{0}$.

PROOF. Let $\mathbf{u} = (u_1, u_2, u_3, \ldots, u_n)$,
$\mathbf{v} = (v_1, v_2, v_3, \ldots, v_n)$ and $\mathbf{w} = (w_1, w_2, w_3, \ldots, w_n)$.

(1) Since $u_iv_i = v_iu_i$, we have

$$\mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^{n} u_iv_i = \sum_{i=1}^{n} v_iu_i = \mathbf{v} \cdot \mathbf{u}$$

(2) Since

$$\mathbf{u} + \mathbf{v} = (u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n),$$

we have

$$(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \sum_{i=1}^{n}(u_i + v_i)w_i$$
$$= \sum_{i=1}^{n}(u_iw_i + v_iw_i)$$
$$= \sum_{i=1}^{n} u_iw_i + \sum_{i=1}^{n} v_iw_i$$
$$= \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$$

(3) Since $\lambda\mathbf{u} = (\lambda u_1, \lambda u_2, \lambda u_3, \ldots, \lambda u_n)$, we have

$$(\lambda\mathbf{u}) \cdot \mathbf{v} = \sum_{i=1}^{n}(\lambda u_i)v_i = \lambda\sum_{i=1}^{n} u_iv_i = \lambda(\mathbf{u} \cdot \mathbf{v}).$$

(4) Clearly $\mathbf{u} \cdot \mathbf{u} = \sum_{i=1}^{n} u_i^2 \geq 0$. If $\mathbf{u} \cdot \mathbf{u} = 0$, then $u_i = 0$ for every i and therefore $\mathbf{u} = \mathbf{0}$. If $\mathbf{u} = \mathbf{0}$, then $\mathbf{u}.\mathbf{u} = 0$.

□

EXAMPLE 1.3. *Using the above theorem, we can compute the Euclidean inner product in the same way we compute ordinary product. For example,*

$$\begin{aligned}(\mathbf{u}+\mathbf{v}).(\mathbf{u}-\mathbf{v}) &= \mathbf{u}.(\mathbf{u}-\mathbf{v})+\mathbf{v}.(\mathbf{u}-\mathbf{v})\\ &= \mathbf{u}\cdot\mathbf{u}-\mathbf{u}\cdot\mathbf{v}+\mathbf{v}\cdot\mathbf{u}-\mathbf{v}\cdot\mathbf{v}\\ &= \mathbf{u}\cdot\mathbf{u}-\mathbf{v}\cdot\mathbf{v}.\end{aligned}$$

DEFINITION 1.3. The Euclidean norm or Euclidean length of a vector $\mathbf{u}=(u_1,u_2,u_3,\dots,u_n)$ in $\mathbb{R}^n$, denoted by $\|\mathbf{u}\|$, is defined by

$$\|\mathbf{u}\| := \sqrt{\mathbf{u}\cdot\mathbf{u}} = \sqrt{u_1^2+u_2^2+\cdots+u_n^2} = \sqrt{\sum_{i=1}^{n} u_i^2}.$$

The distance between two vectors $\mathbf{u}$, $\mathbf{v}$ in $\mathbb{R}^n$, denoted by $d(\mathbf{u},\mathbf{v})$, is defined by

$$\begin{aligned}d(\mathbf{u},\mathbf{v}) &:= \|\mathbf{u}-\mathbf{v}\|\\ &= \sqrt{(u_1-v_1)^2+(u_2-v_2)^2+\cdots+(u_n-v_n)^2}\\ &= \sqrt{\sum_{i=1}^{n}(u_i-v_i)^2}\end{aligned}$$

EXAMPLE 1.4. *For two vectors* $\mathbf{u}$, $\mathbf{v}$ *in* $\mathbb{R}^n$, *we have*

$$\begin{aligned}\|\mathbf{u}+\mathbf{v}\|^2 &= (\mathbf{u}+\mathbf{v})\cdot(\mathbf{u}+\mathbf{v})\\ &= \mathbf{u}\cdot(\mathbf{u}+\mathbf{v})+\mathbf{v}\cdot(\mathbf{u}+\mathbf{v})\\ &= \mathbf{u}\cdot\mathbf{u}+\mathbf{u}\cdot\mathbf{v}+\mathbf{v}\cdot\mathbf{u}+\mathbf{v}\cdot\mathbf{v}\\ &= \mathbf{u}\cdot\mathbf{u}+2(\mathbf{u}\cdot\mathbf{v})+\mathbf{v}\cdot\mathbf{v}\\ &= \|\mathbf{u}\|^2+2(\mathbf{u}\cdot\mathbf{v})+\|\mathbf{v}\|^2\end{aligned}$$

and

$$\begin{aligned}\|\mathbf{u}-\mathbf{v}\|^2 &= (\mathbf{u}-\mathbf{v}).(\mathbf{u}-\mathbf{v})\\ &= \mathbf{u}\cdot(\mathbf{u}-\mathbf{v})-\mathbf{v}\cdot(\mathbf{u}-\mathbf{v})\\ &= \mathbf{u}\cdot\mathbf{u}-\mathbf{u}\cdot\mathbf{v}-\mathbf{v}\cdot\mathbf{u}+\mathbf{v}\cdot\mathbf{v}\\ &= \mathbf{u}\cdot\mathbf{u}-2(\mathbf{u}\cdot\mathbf{v})+\mathbf{v}\cdot\mathbf{v}\\ &= \|\mathbf{u}\|^2-2(\mathbf{u}\cdot\mathbf{v})+\|\mathbf{v}\|^2.\end{aligned}$$

Adding the two equalities, we get

$$\|\mathbf{u}+\mathbf{v}\|^2+\|\mathbf{u}-\mathbf{v}\|^2 = 2[\|\mathbf{u}\|^2+\|\mathbf{v}\|^2].$$

Also we get

$$\|\mathbf{u}+\mathbf{v}\|^2-\|\mathbf{u}-\mathbf{v}\|^2 = 4\mathbf{u}\cdot\mathbf{v}$$

and therefore the Euclidean inner product can be expressed in term of the norms as

$$\mathbf{u}\cdot\mathbf{v} = \frac{1}{4}[\|\mathbf{u}+\mathbf{v}\|^2-\|\mathbf{u}-\mathbf{v}\|^2].$$

EXAMPLE 1.5. *Let* $\mathbf{u}=(1,2,-1,3)$ *and* $\mathbf{v}=(2,0,3,1)$ *be two vectors in* $\mathbb{R}^4$. *Then*

$$\begin{aligned}\|\mathbf{u}\| &= \sqrt{1^2+2^2+(-1)^2+3^2} = \sqrt{15},\\ \|\mathbf{v}\| &= \sqrt{2^2+0^2+3^2+1^2} = \sqrt{14},\\ \mathbf{u}\cdot\mathbf{v} &= 1\times 2+2\times 0+(-1)\times 3+3\times 1 = 2.\end{aligned}$$

Note that

$$|(\mathbf{u}\cdot\mathbf{v})| = 2 \le \sqrt{15}\sqrt{14} = \|\mathbf{u}\|\,\|\mathbf{v}\|\cdot$$

THEOREM 1.3. *If* $\mathbf{u}$ *is vector in* $\mathbb{R}^n$ *and* k *a scalar, then*

(1) $\|\mathbf{u}\| \ge 0$ *and* $\|\mathbf{u}\| = 0$ *if and only if* $\mathbf{u} = \mathbf{0}$
(2) $\|\lambda\mathbf{u}\| = |\lambda|\,\|\mathbf{u}\|$.

PROOF. We prove only that $\|\lambda\mathbf{u}\| = |\lambda|\,\|\mathbf{u}\|$. Since $\lambda\mathbf{u} = (\lambda u_1, \lambda u_2, \cdots, \lambda u_n)$, we have

$$\begin{aligned}\|\lambda\mathbf{u}\| &= \sqrt{\sum_{i=1}^{n}(\lambda u_i)^2}\\ &= \sqrt{\lambda^2\sum_{i=1}^{n}u_i^2}\\ &= |\lambda|\sqrt{\sum_{i=1}^{n}(u_i)^2}\\ &= |\lambda|\,\|\mathbf{u}\|.\end{aligned}$$

□

EXAMPLE 1.6. *By taking* $\lambda = -1$ *in* $\|\lambda\mathbf{u}\| = |\lambda|\,\|\mathbf{u}\|$, *we obtain the following: For any* $\mathbf{u} \in \mathbb{R}^n$,

$$\|-\mathbf{u}\| = \|\mathbf{u}\|.$$

THEOREM 1.4 (Cauchy-Schwarz inequality). *Let* $\mathbf{u} = (u_1, u_2, u_3, \ldots, u_n)$, $\mathbf{v} = (v_1, v_2, v_3, \ldots, v_n)$ *be two vectors in* $\mathbb{R}^n$. *Then*

$$|(\mathbf{u}\cdot\mathbf{v})| \le \|\mathbf{u}\|\,\|\mathbf{v}\|$$

or equivalently

$$\left|\sum_{i=1}^{n} u_i v_i\right| \le \sqrt{\sum_{i=1}^{n} u_i^2}\sqrt{\sum_{i=1}^{n} v_i^2}.$$

PROOF. Let $a = \|\mathbf{v}\|^2$, $b = \mathbf{u}\cdot\mathbf{v}$ and $c = \|\mathbf{u}\|^2$. Note that $a \ge 0$. If $a = 0$, then $\mathbf{v} = 0$ and in this case, $\mathbf{u}\cdot\mathbf{v} = 0$ and $\|\mathbf{v}\| = 0$ and therefore the inequality is satisfied. So assume that $a > 0$. Now

$$\begin{aligned}0 &\le \|\mathbf{u} + \lambda\mathbf{v}\|^2\\ &= \|\mathbf{u}\|^2 + 2\lambda(\mathbf{u}.\mathbf{v}) + \lambda^2\|\mathbf{v}\|^2\\ &= a\lambda^2 + 2b\lambda + c\end{aligned}$$

for all λ. By taking $\lambda = -b/a$, we see that

$$a\frac{b^2}{a^2} - 2b\frac{b}{a} + c \ge 0$$

or

$$-\frac{b^2}{a} + c \ge 0.$$

Since $a > 0$, we obtain $b^2 \le ac$ or equivalently

$$(\mathbf{u}\cdot\mathbf{v})^2 \le \|\mathbf{u}\|^2\|\mathbf{v}\|^2.$$

By taking square roots both side, we get the desired result. □

EXAMPLE 1.7. *When $n = 2$, Cauchy-Schwarz inequality becomes*

$$|u_1v_1 + u_2v_2|^2 \leq (u_1^2 + u_2^2)(v_1^2 + v_2^2).$$

By taking $u_1 = a$, $u_2 = b$, $v_1 = \cos\theta$ and $v_2 = \sin\theta$, we get

$$\begin{aligned} |a\cos\theta + b\sin\theta| &\leq \sqrt{(a^2+b^2)(\sin^2\theta + \cos^2\theta)} \\ &= \sqrt{a^2+b^2} \end{aligned}$$

or equivalently $(a\cos\theta + b\sin\theta)^2 \leq a^2 + b^2$.

EXAMPLE 1.8. *By taking $v_1 = v_2 = \cdots = v_n = 1$ in Cauchy-Schwarz inequality, we have*

$$(u_1 + u_2 + \cdots + u_n)^2 \leq n(u_1^2 + u_2^2 + \ldots + u_n^2)\cdot$$

THEOREM 1.5 (Triangle inequality). *If $\mathbf{u}$ and $\mathbf{v}$ are two vectors in $\mathbb{R}^n$, then*

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|\cdot$$

PROOF. By using Cauchy-Schwartz inequality, we get

$$\begin{aligned} \|\mathbf{u}+\mathbf{v}\|^2 &= \|\mathbf{u}\|^2 + 2(\mathbf{u}\cdot\mathbf{v}) + \|\mathbf{v}\|^2 \\ &\leq \|\mathbf{u}\|^2 + 2\|\mathbf{u}\|\|\mathbf{v}\| + \|\mathbf{v}\|^2 \\ &= (\|\mathbf{u}\| + \|\mathbf{v}\|)^2. \end{aligned}$$

By taking square roots, the result follows. □

EXAMPLE 1.9. *Let $\mathbf{u}$ and $\mathbf{v}$ be two vectors in $\mathbb{R}^n$. Then*

$$|\ \|\mathbf{u}\| - \|\mathbf{v}\|\ | \leq \|\mathbf{u} - \mathbf{v}\|\cdot$$

PROOF. By using the triangle inequality, we get

$$\begin{aligned} \|\mathbf{u}\| &= \|(\mathbf{u}-\mathbf{v}) + \mathbf{v}\| \\ &\leq \|\mathbf{u}-\mathbf{v}\| + \|\mathbf{v}\| \end{aligned}$$

or

$$\|\mathbf{u}\| - \|\mathbf{v}\| \leq \|\mathbf{u}-\mathbf{v}\|\cdot$$

Similarly

$$\begin{aligned} \|\mathbf{v}\| &= \|(\mathbf{v}-\mathbf{u}) + \mathbf{u}\| \\ &\leq \|\mathbf{v}-\mathbf{u}\| + \|\mathbf{u}\| \\ &= \|-(\mathbf{u}-\mathbf{v})\| + \|\mathbf{u}\| \\ &= \|\mathbf{u}-\mathbf{v}\| + \|\mathbf{u}\| \end{aligned}$$

or

$$\|\mathbf{v}\| - \|\mathbf{u}\| \leq \|\mathbf{u}-\mathbf{v}\|$$

or

$$-(\|\mathbf{u}\| - \|\mathbf{v}\|) \leq \|\mathbf{u}-\mathbf{v}\|\cdot$$

Thus we have $|\ \|\mathbf{u}\| - \|\mathbf{v}\|\ | \leq \|\mathbf{u}-\mathbf{v}\|$. □

DEFINITION 1.4. Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are orthogonal if $\mathbf{u}.\mathbf{v} = 0$.

EXAMPLE 1.10. *Consider* $\mathbf{u} = (1, 1, 1, 1, 1, 1)$ *and* $\mathbf{v} = (1, -1, 1, -1, 1, -1)$ *in* $\mathbb{R}^6$. *Then*

$$\mathbf{u} \cdot \mathbf{v} = 1 - 1 + 1 - 1 + 1 - 1 = 0$$

and therefore $\mathbf{u}$ *and* $\mathbf{v}$ *are orthogonal in* $\mathbb{R}^6$. *Note that* $\|\mathbf{u}\| = \sqrt{6} = \|\mathbf{v}\|$,

$$\begin{aligned} \|\mathbf{u} + \mathbf{v}\| &= \|(2, 0, 2, 0, 2, 0)\| \\ &= \sqrt{4 + 0 + 4 + 0 + 4 + 0} = \sqrt{12} \end{aligned}$$

and therefore

$$\|\mathbf{u} + \mathbf{v}\|^2 = 12 = 6 + 6 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2.$$

THEOREM 1.6 (Pythagorean Theorem). *Two vectors* $\mathbf{u}$ *and* $\mathbf{v}$ *in* $\mathbb{R}^n$ *are orthogonal if and only if*

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2.$$

PROOF. Since $\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + 2(\mathbf{u} \cdot \mathbf{v}) + \|\mathbf{v}\|^2$, we have $\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$ if and only if $\mathbf{u} \cdot \mathbf{v} = 0$ or equivalently if and only if $\mathbf{u}$ and $\mathbf{v}$ are orthogonal. □

THEOREM 1.7. *Two vectors* $\mathbf{u}$ *and* $\mathbf{v}$ *in* $\mathbb{R}^n$ *are orthogonal if and only if*

$$\|\mathbf{u} + \mathbf{v}\| = \|\mathbf{u} - \mathbf{v}\|.$$

PROOF. Since

$$\mathbf{u} \cdot \mathbf{v} = \frac{1}{4}[\|\mathbf{u} + \mathbf{v}\|^2 - \|\mathbf{u} - \mathbf{v}\|^2],$$

$\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are orthogonal if and only if $\|\mathbf{u} + \mathbf{v}\|^2 - \|\mathbf{u} - \mathbf{v}\|^2 = 0$ or equivalently if and if $\|\mathbf{u} + \mathbf{v}\| = \|\mathbf{u} - \mathbf{v}\|$. □

1. Linear Transformations

Let A and B be two sets. Then a function $f : A \to B$ is a rule that associate with each element $a \in A$ with exactly one element $b \in B$; the element b is called the image of a and is written as $b = f(a)$. The set A is called the domain of f and B is called the codomain of f. The set $\{f(a) : a \in A\}$, the set of all images of points in A is called the range of f. Two functions f and g are equal if they have same domain and $f(a) = g(a)$ for all a in the domain.

EXAMPLE 1.11. *The following are examples of functions.*

(1) $f : \mathbb{R} \to \mathbb{R}$ *where* $f(x) = x^2 + x$,
(2) $f : \mathbb{R}^3 \to \mathbb{R}$ *where* $f(x, y, z) = x^2 + xy + z^2$,
(3) $f : \mathbb{R}^n \to \mathbb{R}$ *where* $f(x_1, x_2, \ldots, x_n) = x_1 + x_2^2 + x_3^3 + \ldots + x_n^n$,
(4) $f : \mathbb{R} \to \mathbb{R}^2$ *where* $f(x) = (x^2, x + 1)$,
(5) $f : \mathbb{R}^2 \to \mathbb{R}^2$ *where* $f(x, y) = (x^2, x + y^2)$,
(6) $f : \mathbb{R}^m \to \mathbb{R}^n$ *where* $f(x_1, x_2, \ldots, x_m) = (x_1, x_1 + x_2, \ldots, x_{n-1} + x_n)$ $(n \leq m)$.

A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is called a map or transformation and we say that f maps $\mathbb{R}^n$ to $\mathbb{R}^m$ or f is a mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$. A mapping $f : \mathbb{R}^n \to \mathbb{R}^n$ is called an operator on

$\mathbb{R}^n$. For $i = 1, 2, \dots, m$, let $f_i : \mathbb{R}^n \to \mathbb{R}$ be functions. For each $(x_1, x_2, x_3, \cdots, x_n) \in \mathbb{R}^n$, let

$$\begin{aligned} w_1 &= f_1(x_1, x_2, x_3, \dots, x_n) \\ w_2 &= f_2(x_1, x_2, x_3, \dots, x_n) \\ \dots & \vdots \\ w_m &= f_m(x_1, x_2, x_3, \dots, x_n). \end{aligned}$$

Thus for each $(x_1, x_2, x_3, \cdots, x_n) \in \mathbb{R}^n$ we can associate $(w_1, w_2, w_3, \cdots, w_m) \in \mathbb{R}^m$, and therefore these equations defines a function $f : \mathbb{R}^n \to \mathbb{R}^m$ where

$$f(x_1, x_2, x_3, \cdots, x_n) = (w_1, w_2, w_3, \cdots, w_m).$$

Example 1.12. *Consider the functions*

$$\begin{aligned} w_1 &= f_1(x_1, x_2, x_3) = x_1 + x_1x_2 + x_3 \\ w_2 &= f_2(x_1, x_2, x_3) = x_2 + x_3. \end{aligned}$$

Then we can define a function $f : \mathbb{R}^3 \to \mathbb{R}^2$ *by*

$$f(x_1, x_2, x_3) = (w_1, w_2) = (x_1 + x_1x_2 + x_3, x_2 + x_3).$$

If the functions f_i defining f are all linear equations, then $f : \mathbb{R}^n \to \mathbb{R}^m$ is called a linear transformation. When $n = m$, we call a linear transformation as a linear operator. Thus a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is given by

$$T(x_1, x_2, x_3, \cdots, x_n) = (w_1, w_2, w_3, \cdots, w_m)$$

where

$$\begin{aligned} w_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ w_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots & \qquad\qquad \vdots \\ w_m &= a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n. \end{aligned}$$

These equations can be written as

$$\begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_m \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_m \end{pmatrix}.$$

If we write

$$\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_m \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_m \end{pmatrix},$$

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{pmatrix},$$

then we have $\mathbf{w} = A\mathbf{x}$. The matrix A is called the standard matrix of the linear transformation T and T is called multiplication by A. In this case, we write $\mathbf{w} = T(\mathbf{x}) = A\mathbf{x}$.

EXAMPLE 1.13. *The transform $T : \mathbb{R}^4 \to \mathbb{R}^3$ defined by*

$$\begin{aligned} w_1 &= 2x_1 + x_2 + 4x_3 - 2x_4, \\ w_2 &= 3x_1 + x_2 - x_3 + 2x_4, \\ w_3 &= x_1 + x_2 + x_3 + 3x_4. \end{aligned}$$

is a linear transformation from $\mathbb{R}^4$ to $\mathbb{R}^3$. This linear transformation can be written as

$$\begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 4 & -2 \\ 3 & 1 & -1 & 2 \\ 1 & 1 & 1 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}.$$

Thus the standard matrix of this linear transformation is given by

$$A = \begin{pmatrix} 2 & 1 & 4 & -2 \\ 3 & 1 & -1 & 2 \\ 1 & 1 & 1 & 3 \end{pmatrix}.$$

When $(x_1, x_2, x_3, x_4) = (1, 2, -1, 0)$, we have

$$\begin{aligned} w_1 &= 2 + 2 - 4 - 0 = 0, \\ w_2 &= 3 + 2 + 1 + 0 = 6, \\ w_3 &= 1 + 2 - 1 + 0 = 2. \end{aligned}$$

This can also be computed using the standard matrix of the linear transformation T:

$$\begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 4 & -2 \\ 3 & 1 & -1 & 2 \\ 1 & 1 & 1 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 6 \\ 2 \end{pmatrix}.$$

THEOREM 1.8. *A transform $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation if and only if*

$$T(\lambda\mathbf{u} + \mu\mathbf{v}) = \lambda T(\mathbf{u}) + \mu T(\mathbf{v})$$

for all vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and for every scalars λ, μ.

PROOF. Assume that T is a linear transformation and A its standard matrix. Then

$$\begin{aligned} T(\lambda\mathbf{u} + \mu\mathbf{v}) &= A(\lambda\mathbf{u} + \mu\mathbf{v}) \\ &= A(\lambda\mathbf{u}) + A(\mu\mathbf{v}) \\ &= \lambda A\mathbf{u} + \mu A\mathbf{v} \\ &= \lambda T(\mathbf{u}) + \mu T(\mathbf{v}). \end{aligned}$$

Conversely assume that the transformation T has the property that

$$T(\lambda\mathbf{u} + \mu\mathbf{v}) = \lambda T(\mathbf{u}) + \mu T(\mathbf{v})$$

for all vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and for every scalars λ, μ. It follows easily that

$$\begin{aligned} &T(\lambda_1\mathbf{u}_1 + \lambda_2\mathbf{u}_2 + \cdots + \lambda_k\mathbf{u}_k) \\ &= \lambda_1 T(\mathbf{u}_1) + \lambda_2 T(\mathbf{u}_2) + \cdots + \lambda_k T(\mathbf{u}_k) \end{aligned}$$

for any vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k \in \mathbb{R}^n$ and any scalars $\lambda_1, \ldots, \lambda_n$. We complete the proof by showing that T is a multiplication by a matrix A. Let $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ be vectors given by

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \cdots, \quad \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

Let A be the matrix whose columns are $T(\mathbf{e}_1)$, $T(\mathbf{e}_2)$, $\ldots$, $T(\mathbf{e}_n)$. Then

$$\begin{aligned} A\mathbf{x} &= x_1 T(\mathbf{e}_1) + x_2 T(\mathbf{e}_2) + \cdots + x_n T(\mathbf{e}_n) \\ &= T(x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n) \\ &= T(\mathbf{x}). \end{aligned}$$

Thus T is multiplication by A and therefore it is linear. □

Definition 1.5. A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is one-to-one if

$$\mathbf{x} \neq \mathbf{y} \Rightarrow T(\mathbf{x}) \neq T(\mathbf{y})$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ or equivalently

$$T(\mathbf{x}) = T(\mathbf{y}) \Rightarrow \mathbf{x} = \mathbf{y}$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

Example 1.14. *Let $T(\mathbf{x}) = A\mathbf{x}$ be a linear transformation. Then*

$$T(\mathbf{x}) = T(\mathbf{y}) \Rightarrow A\mathbf{x} = A\mathbf{y} \Rightarrow \mathbf{x} = \mathbf{y}$$

provided A is invertible. Thus a linear is one-to-one if its matrix is invertible.

CHAPTER 2

Real Vector Spaces

1. Real Vector Spaces

Definition 2.1. A nonempty set V, whose elements are called vectors, together with two operation called addition and scalar multiplication denoted by $+$ and . respectively is called a vector space if the following axioms hold for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and for all scalars λ and μ:

(1) If $\mathbf{u}, \mathbf{v} \in V$, then $\mathbf{u}+\mathbf{v} \in V$.
(2) $\mathbf{u}+\mathbf{v} = \mathbf{v}+\mathbf{u}$
(3) $(\mathbf{u}+\mathbf{v})+\mathbf{w} = \mathbf{u}+(\mathbf{v}+\mathbf{w})$
(4) There is an element $\mathbf{0} \in V$ such that for any $\mathbf{u} \in V$, $\mathbf{u}+\mathbf{0} = \mathbf{0}+\mathbf{u} = \mathbf{u}$.
(5) For each $\mathbf{u} \in V$, there is an element $-\mathbf{u} \in V$ such that $\mathbf{u}+(-\mathbf{u}) = (-\mathbf{u})+\mathbf{u} = \mathbf{0}$. This element is called the (additive) inverse of $\mathbf{u}$.
(6) If $\mathbf{u} \in V$ and λ a scalar, then $\lambda\mathbf{u} \in V$.
(7) $\lambda(\mathbf{u}+\mathbf{v}) = \lambda\mathbf{u}+\lambda\mathbf{v}$
(8) $(\lambda+\mu)\mathbf{u} = \lambda\mathbf{u}+\mu\mathbf{u}$
(9) $\lambda(\mu\mathbf{u}) = (\lambda\mu)\mathbf{u}$
(10) $1u = u$

Example 2.1. *The set $V = \mathbb{R}^n$ with the addition and scalar multiplication is a vector space.*

Example 2.2. *The set M_2 of all square matrices of order 2 with usual matrix addition, matrix scalar multiplication is a vector space. This can be seen as follows.*

Let

$$\mathbf{u} = \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix}.$$

Then

(1) *The addition of $\mathbf{u}$ and $\mathbf{v}$ is given by*

$$\begin{aligned} \mathbf{u}+\mathbf{v} &= \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix} + \begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix} \\ &= \begin{pmatrix} u_{11}+v_{11} & u_{12}+v_{12} \\ u_{21}+v_{21} & u_{22}+v_{22} \end{pmatrix}. \end{aligned}$$

Clearly $\mathbf{u}+\mathbf{v} \in M_2$.

(2) *We have earlier proved the commutative law for matrix addition. We repeat the proof. By the definition of addition, we have*

$$\begin{aligned}\mathbf{u}+\mathbf{v} &= \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix} + \begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix} \\ &= \begin{pmatrix} u_{11}+v_{11} & u_{12}+v_{12} \\ u_{21}+v_{21} & u_{22}+v_{22} \end{pmatrix} \\ &= \begin{pmatrix} v_{11}+u_{11} & v_{12}+u_{12} \\ v_{21}+u_{21} & v_{22}+u_{22} \end{pmatrix} \\ &= \mathbf{v}+\mathbf{u}.\end{aligned}$$

(3) *Associative law for matrix addition is already proved.*

(4) *The zero matrix*

$$\mathbf{0} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

satisfies $\mathbf{u}+\mathbf{0}=\mathbf{0}+\mathbf{u}=\mathbf{u}$ *for all* $\mathbf{u} \in M_2$.

(5) *For*

$$\mathbf{u} = \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix} \in M_2,$$

the matrix

$$-\mathbf{u} = \begin{pmatrix} -u_{11} & -u_{12} \\ -u_{21} & -u_{22} \end{pmatrix} \in M_2$$

satisfies

$$\mathbf{u}+(-\mathbf{u}) = (-\mathbf{u})+\mathbf{u} = \mathbf{0}.$$

(6) *By the definition of matrix scalar multiplication, we have*

$$\lambda\mathbf{u} = \lambda \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix} = \begin{pmatrix} \lambda u_{11} & \lambda u_{12} \\ \lambda u_{21} & \lambda u_{22} \end{pmatrix} \in M_2.$$

(7) *By the definition of matrix addition and matrix scalar multiplication, we have*

$$\begin{aligned}\lambda(\mathbf{u}+\mathbf{v}) &= \lambda \left[\begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix} + \begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix} \right] \\ &= \lambda \begin{pmatrix} u_{11}+v_{11} & u_{12}+v_{12} \\ u_{21}+v_{21} & u_{22}+v_{22} \end{pmatrix} \\ &= \begin{pmatrix} \lambda(u_{11}+v_{11}) & \lambda(u_{12}+v_{12}) \\ \lambda(u_{21}+v_{21}) & \lambda(u_{22}+v_{22}) \end{pmatrix} \\ &= \begin{pmatrix} \lambda u_{11}+\lambda v_{11} & \lambda u_{12}+\lambda v_{12} \\ \lambda u_{21}+\lambda v_{21} & \lambda u_{22}+\lambda v_{22} \end{pmatrix} \\ &= \begin{pmatrix} \lambda u_{11} & \lambda u_{12} \\ \lambda u_{21} & \lambda u_{22} \end{pmatrix} + \begin{pmatrix} \lambda v_{11} & \lambda v_{12} \\ \lambda v_{21} & \lambda v_{22} \end{pmatrix} \\ &= \lambda \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix} + \lambda \begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix} \\ &= \lambda\mathbf{u}+\lambda\mathbf{v}.\end{aligned}$$

(8) *The proof of* $(\lambda+\mu)\mathbf{u} = \lambda\mathbf{u}+\mu\mathbf{u}$ *is similar.*

(9) *It is clear that*

$$\begin{aligned}\lambda(\mu\mathbf{u}) &= \lambda\begin{pmatrix}\mu u_{11} & \mu u_{12}\\ \mu u_{21} & \mu u_{22}\end{pmatrix}\\ &= \begin{pmatrix}\lambda\mu u_{11} & \lambda\mu u_{12}\\ \lambda\mu u_{21} & \lambda\mu u_{22}\end{pmatrix}\\ &= (\lambda\mu)\begin{pmatrix}u_{11} & u_{12}\\ u_{21} & u_{22}\end{pmatrix}\\ &= (\lambda\mu)\mathbf{u}.\end{aligned}$$

(10) *Clearly*

$$1\mathbf{u} = 1\begin{pmatrix}u_{11} & u_{12}\\ u_{21} & u_{22}\end{pmatrix} = \begin{pmatrix}1u_{11} & 1u_{12}\\ 1u_{21} & 1u_{22}\end{pmatrix} = \mathbf{u}.$$

Thus M_2 is a vector space.

EXAMPLE 2.3. *The set $M_{m\times n}$ of all matrices of order $m \times n$ with usual matrix addition, matrix scalar multiplication is a vector space.*

EXAMPLE 2.4. *The set V of all points in any plane through the origin is a vector space with usual addition and scalar multiplication. Any plane through the origin has the equation $ax + by + cz = 0$. At least one of the constant a, b, c, is non-zero. Let $c \neq 0$. Then the equation can be written as $z = (-a/c)x + (-b/c)y = Ax + By$. Thus the set V is given by*

$$V = \{(u_1, u_2, Au_1 + Bu_2) | u_1, u_2 \in \mathbb{R}\}.$$

(1) *If $\mathbf{u}, \mathbf{v} \in V$, then*

$$\mathbf{u} = (u_1, u_2, Au_1 + Bu_2), \quad \mathbf{v} = (v_1, v_2, Av_1 + Bv_2)$$

and therefore

$$\begin{aligned}&\mathbf{u} + \mathbf{v}\\ &= (u_1, u_2, Au_1 + Bu_2) + (v_1, v_2, Av_1 + Bv_2)\\ &= (u_1 + v_1, u_2 + v_2, A(u_1 + v_1) + B(u_2 + v_2))\\ &\in V.\end{aligned}$$

(2) *We omit the proofs of $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$, $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$.*
(3) *The element $\mathbf{0} = (0, 0, 0) \in V$ satisfies, any $\mathbf{u} \in V$, $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}$.*
(4) *For each $\mathbf{u} = (u_1, u_2, Au_1 + Bu_2) \in V$, the element $-\mathbf{u} = (-u_1, -u_2, -Au_1 - Bu_2) \in V$ satisfies $\mathbf{u} + (-\mathbf{u}) = (-\mathbf{u}) + \mathbf{u} = \mathbf{0}$.*
(5) *If $\mathbf{u}, \mathbf{v} \in V$ and λ a scalar, then it is easy to prove that $\lambda\mathbf{u} \in V$. $\lambda(\mathbf{u} + \mathbf{v}) = \lambda\mathbf{u} + \lambda\mathbf{v}$, $(\lambda + \mu)\mathbf{u} = \lambda\mathbf{u} + \mu\mathbf{u}$, $\lambda(\mu\mathbf{u}) = (\lambda\mu)\mathbf{u}$, $1u = u$.*

Therefore V is a vector space.

EXAMPLE 2.5. *Let V be the set of all functions $f : [a, b] \to \mathbb{R}$. If $f, g \in V$ and λ be any scalar, then let $f + g$, λf be the functions defined by*

$$(f + g)(x) = f(x) + g(x), \quad (\lambda f)(x) = \lambda f(x)$$

for all $x \in [a, b]$. Under these operations, V is a vector space.

Similarly the set $\mathbb{C}[a, b]$ of all continuous functions $f : [a, b] \to \mathbb{R}$ is a vector space under the above defined operations.

Example 2.6. *Let V be the set with a single element $\mathbf{0}$. Let addition and scalar multiplication be defined by*

$$\mathbf{0}+\mathbf{0}=\mathbf{0}, \quad \lambda\mathbf{0}=\mathbf{0}.$$

Then V is a vector space and it is called a zero vector space.

Theorem 2.1 (Cancellation law). *Let V be a vector space. For any vector $\mathbf{u},\mathbf{v},\mathbf{w}\in V$, we have*

$$\mathbf{u}+\mathbf{w}=\mathbf{v}+\mathbf{w}\Rightarrow\mathbf{u}=\mathbf{v}.$$

Proof. Since $\mathbf{w}\in V$, there is an element $-\mathbf{w}\in V$ such that $\mathbf{w}+(-\mathbf{w})=\mathbf{0}=(-\mathbf{w})+\mathbf{w}$. Now by adding $-\mathbf{w}$ to the equation $\mathbf{u}+\mathbf{w}=\mathbf{v}+\mathbf{w}$, we get

$$(\mathbf{u}+\mathbf{w})+(-\mathbf{w})=(\mathbf{v}+\mathbf{w})+(-\mathbf{w})$$

or by using associative law,

$$\mathbf{u}+(\mathbf{w}+(-\mathbf{w}))=\mathbf{v}+(\mathbf{w}+(-\mathbf{w}))$$

or

$$\mathbf{u}+\mathbf{0}=\mathbf{v}+\mathbf{0}$$

or

$$\mathbf{u}=\mathbf{v}.$$

□

Theorem 2.2. *Let V be a vector space. For any vector $\mathbf{u}\in V$ and scalar λ, we have*

(1) $0\mathbf{u}=\mathbf{0}$;
(2) $\lambda\mathbf{0}=\mathbf{0}$;
(3) $(-1)\mathbf{u}=-\mathbf{u}$;
(4) *If* $\lambda\mathbf{u}=\mathbf{0}$, *then* $\lambda=0$ *or* $\mathbf{u}=\mathbf{0}$.

Proof. (1) Since

$$0\mathbf{u}+0\mathbf{u}=(0+0)\mathbf{u}=0\mathbf{u}=\mathbf{0}+0\mathbf{u},$$

by cancellation law, we have $0\mathbf{u}=\mathbf{0}$.

(2) Since $\mathbf{0}+\mathbf{0}=\mathbf{0}$, we have

$$\lambda\mathbf{0}+\lambda\mathbf{0}=\lambda(\mathbf{0}+\mathbf{0})=\lambda\mathbf{0}=\mathbf{0}+\lambda\mathbf{0}.$$

By cancellation law, we have $\lambda\mathbf{0}=\mathbf{0}$.

(3) Since

$$\mathbf{u}+(-\mathbf{u})=\mathbf{0}$$

and

$$\mathbf{u}+(-1)\mathbf{u}=1\mathbf{u}+(-1)\mathbf{u}=(1+(-1))\mathbf{u}=0\mathbf{u}=\mathbf{0},$$

we have

$$\mathbf{u}+(-1)\mathbf{u}=\mathbf{u}+(-\mathbf{u})=\mathbf{0}.$$

By using cancellation law, we have

$$(-1)\mathbf{u}=-\mathbf{u}.$$

(4) If $\lambda=0$, then there is nothing to prove. Therefore assume that $\lambda\neq 0$. Since $\lambda\mathbf{u}=\mathbf{0}$, we have

$$\mathbf{u}=1\mathbf{u}=(\frac{1}{\lambda}\lambda)\mathbf{u}=\frac{1}{\lambda}(\lambda\mathbf{u})=\frac{1}{\lambda}\mathbf{0}=\mathbf{0}.$$

□

2. Subspaces

DEFINITION 2.2. A subset W of a vector space V is called a subspace if W is itself a vector space under the addition and multiplication defined on V.

THEOREM 2.3. *Let W be a non-empty set subset of a vector space V. Then W is a subspace of V if and only if*

(1) $\mathbf{u}+\mathbf{v} \in W$ *for all* $\mathbf{u}, \mathbf{v} \in W$,
(2) $\lambda\mathbf{u} \in W$ *for all* $\mathbf{u} \in W$ *and for all scalar* λ.

PROOF. Let W be a subspace of V. Then all vector space axioms hold in W and therefore $\mathbf{u}+\mathbf{v}, \lambda\mathbf{u} \in W$ for all $\mathbf{u}, \mathbf{v} \in W$ and for any scalar λ.

Conversely, let the conditions (1) and (2) holds. Then we need to show that W is subspace of V. In other words, we have to show that W is a vector space on its own. In view of (1), the first axiom of vector space holds. The commutative and associative axioms hold in V and therefore they holds in W. By taking $\lambda = 0$ in (2), it follows that $\mathbf{0} \in W$. Similarly by taking $\lambda = -1$ in (1), we see that $-\mathbf{u} \in W$ whenever $\mathbf{u} \in W$. The other axioms hold in V and therefore they hold in W too. This completes the proof. □

THEOREM 2.4. *Let W be a non-empty set subset of a vector space V. Then W is a subspace of V if and only if $\lambda\mathbf{u}+\mu\mathbf{v} \in W$ for all $\mathbf{u}, \mathbf{v} \in W$ and for all scalar λ, μ.*

PROOF. Let W be a subspace of V. Then all vector space axioms hold in W. If $\mathbf{u}, \mathbf{v} \in W$ and λ, μ are scalars, then $\lambda\mathbf{u}, \mu\mathbf{v} \in W$ and therefore $\lambda\mathbf{u}+\mu\mathbf{v} \in W$.

Conversely, if $\lambda\mathbf{u}+\mu\mathbf{v} \in W$ for all $\mathbf{u}, \mathbf{v} \in W$ and for all scalar λ, μ, then it by taking $\lambda = \mu = 1$ and $\mu = 0$ respectively, we get

(1) $\mathbf{u}+\mathbf{v} \in W$ for all $\mathbf{u}, \mathbf{v} \in W$,
(2) $\lambda\mathbf{u} \in W$ for all $\mathbf{u} \in W$ and for all scalar k.

By the previous theorem, it follows that W is a subspace of V. □

EXAMPLE 2.7. *Let $V = \mathbb{R}^3$ be the vector space under usual addition and scalar multiplication and $W = \{(x, y, 0) : x, y \in \mathbb{R}\}$. Let $\mathbf{u} = (x_1, y_1, 0)$ and $\mathbf{v} = (x_2 y_2, 0)$ be two vectors in W and λ, μ be scalars. Then*

$$\begin{aligned} \lambda\mathbf{u}+\mu\mathbf{v} &= \lambda(x_1, y_1, 0) + \mu(x_2, y_2, 0) \\ &= (\lambda x_1 + \mu x_2, \lambda y_1 + \mu y_2, 0) \in W. \end{aligned}$$

Thus W is a subspace of V.

EXAMPLE 2.8. *Let $V = \mathbb{R}^3$ be the vector space under usual addition and scalar multiplication and*

$$W = \{(x, y, ax + by) : x, y \in \mathbb{R}\}.$$

Let $\mathbf{u} = (x_1, y_1, ax_1 + by_1)$ and $\mathbf{v} = (x_2 y_2, ax_2 + by_2)$ be two vectors in W and λ, μ be scalars. Then

$$\begin{aligned} \lambda\mathbf{u}+\mu\mathbf{v} &= \lambda(x_1, y_1, ax_1 + by_1) + \mu(x_2, y_2, ax_2 + by_2) \\ &= (\lambda x_1 + \mu x_2, \lambda y_1 + \mu y_2, a(\lambda x_1 + \mu x_2) + b(\lambda y_1 + \mu y_2)) \in W. \end{aligned}$$

Thus W is a subspace of V.

Since the points $(x, y, z) \in W$ *satisfy*

$$ax + by = z,$$

we have proved that the set of points in a straight line passing through origin is a subspace of $\mathbb{R}^3$.

EXAMPLE 2.9. *Consider the vector space* $V = M_m$ *of all square matrices of order* m. *Let* W *be the subset consisting of symmetric matrices of order* m. *Since the sum of two symmetric matrices is symmetric and scalar multiple of symmetric matrix is symmetric,* W *is a subspace of* V.

Similarly the subsets consisting of lower triangular matrices, upper triangular matrices and diagonal matrices are all subspaces of V.

The subset consisting of invertible matrices of order m *is not a vector space since the sum of two invertible matrices need not be invertible.*

EXAMPLE 2.10. *Consider the vector space* V *of all functions* $f : [a, b] \to \mathbb{R}$ *under the addition and scalar multiplication of functions. Since the sum of two continuous is continuous and scalar multiple of a continuous function is also continuous, the subset* $C[a, b]$ *consisting of continuous functions is a subspace of* V.

THEOREM 2.5. *The set of all of solution of a homogeneous linear system* $A\mathbf{x} = \mathbf{0}$ *(with* m *equations and* n *unknowns) is a subspace of* $\mathbb{R}^n$.

PROOF. Let $\mathbf{x}, \mathbf{x}'$ be two solution of the linear system $A\mathbf{x} = \mathbf{0}$. Then $A\mathbf{x} = \mathbf{0}$ and $A\mathbf{x}' = \mathbf{0}$. For any scalar λ and μ, we have

$$A(\lambda\mathbf{x} + \mu\mathbf{x}') = \lambda A\mathbf{x} + \mu A\mathbf{x}' = \lambda\mathbf{0} + \mu\mathbf{0} = \mathbf{0}.$$

This shows that $\lambda\mathbf{x} + \mu\mathbf{x}'$ is also a solution of $A\mathbf{x} = \mathbf{0}$. Thus the set of all solutions of the homogeneous linear system $A\mathbf{x} = \mathbf{0}$ is a subspace of $\mathbb{R}^n$. □

EXAMPLE 2.11. *Consider solutions of the homogeneous linear system*

$$x + y + z = 0, \quad y - z = 0$$

given by

$$x = -2t, \quad y = t, \quad z = t.$$

The set of all solutions $W = \{(-2t, t, t) : t \in \mathbb{R}\}$ *is a subspace of* $\mathbb{R}^3$. *It can be verified directly. Let* $\mathbf{u} = (-2t, t, t)$ *and* $\mathbf{v} = (-2s, s, s)$ *and* λ *and* μ *be scalars. Then*

$$\lambda\mathbf{u} + \mu\mathbf{v} = (-2(\lambda t + \mu s), \lambda t + \mu s, \lambda t + \mu s)$$

is again in W.

DEFINITION 2.3. A vector $\mathbf{v}$ is a linear combination of the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ if there are scalars $\lambda_1, \lambda_2, \ldots, \lambda_k$ such that

$$\mathbf{v} = \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \ldots + \lambda_k\mathbf{v}_k.$$

EXAMPLE 2.12. *Consider the vector space* $\mathbb{R}^n$ *with usual addition and scalar multiplication. Consider the following vectors*

$$\begin{aligned}
\mathbf{e}_1 &= (1, 0, 0, \cdots, 0)\\
\mathbf{e}_2 &= (0, 1, 0, \cdots, 0)\\
\mathbf{e}_3 &= (0, 0, 1, \cdots, 0)\\
&\vdots\\
\mathbf{e}_n &= (0, 0, 0, \cdots, 1).
\end{aligned}$$

Any vector $\mathbf{v} = (v_1, v_2, \cdots, v_n) \in \mathbb{R}^n$ *can be written as*

$$\mathbf{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + v_3\mathbf{e}_3 + \cdots + v_n\mathbf{e}_n.$$

EXAMPLE 2.13. *Consider the vector space* $\mathbb{R}^3$. *Let* $\mathbf{e}_1 = (1,0,0)$, $\mathbf{e}_2 = (1,1,0)$ *and* $\mathbf{e}_3 = (1,1,1)$. *Let* $\mathbf{v} = (v_1, v_2, v_3) \in \mathbb{R}^3$ *be any vector. Since*

$$\lambda_1\mathbf{e}_1 + \lambda_2\mathbf{e}_2 + \lambda_3\mathbf{e}_3 = (\lambda_1 + \lambda_2 + \lambda_3, \lambda_2 + \lambda_3, \lambda_3),$$

$$\mathbf{v} = \lambda_1\mathbf{e}_1 + \lambda_2\mathbf{e}_2 + \lambda_3\mathbf{e}_3$$

if

$$\lambda_1 + \lambda_2 + \lambda_3 = v_1, \lambda_2 + \lambda_3 = v_2, \lambda_3 = v_3$$

or equivalently if

$$\lambda_1 = v_1 - v_2, \quad \lambda_2 = v_2 - v_3, \lambda_3 = v_3.$$

Thus

$$\mathbf{v} = (v_1 - v_2)\mathbf{e}_1 + (v_2 - v_3)\mathbf{e}_2 + v_3\mathbf{e}_3.$$

EXAMPLE 2.14. *Consider the vector space* $\mathbb{R}^3$. *Let* $\mathbf{v}_1 = (3,-1,2)$, $\mathbf{v}_2 = (-1,2,-1)$. *Consider the vectors* $\mathbf{u} = (7,-4,5)$ *and* $\mathbf{w} = (1,3,1)$.

If $\mathbf{u}$ *is a linear combination of* $\mathbf{v}_1$ *and* $\mathbf{v}_2$, *then there are scalars* λ_1 *and* λ_2 *such that*

$$\mathbf{u} = \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2.$$

Since

$$\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 = (3\lambda_1 - \lambda_2, -\lambda_1 + 2\lambda_2, 2\lambda_1 - \lambda_2)$$

we must have

$$3\lambda_1 - \lambda_2 = 7, \quad -\lambda_1 + 2\lambda_2 = -4, \quad 2\lambda_1 - \lambda_2 = 5.$$

The solution of the linear system is given by $\lambda_1 = 2$ *and* $\lambda_2 = -1$. *Thus*

$$\mathbf{u} = 2\mathbf{v}_1 - 2\mathbf{v}_2.$$

The linear system resulting from

$$\mathbf{w} = \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2$$

is

$$3\lambda_1 - \lambda_2 = 1, \quad -\lambda_1 + 2\lambda_2 = 3, \quad 2\lambda_1 - \lambda_2 = 1;$$

this system is inconsistent and therefore $\mathbf{w}$ *is not a linear combination of* $\mathbf{v}_1$ *and* $\mathbf{v}_2$.

THEOREM 2.6. *Let* V *be a vector space. The set* W *of all linear combination of* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k \in V$ *is a subspace of* V. *Any subspace* W' *containing* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ *contains the subspace* W.

PROOF. Since $\mathbf{v}_i = 0\mathbf{v}_1 + \cdots + 0\mathbf{v}_{i-1} + 1\mathbf{v}_i + 0\mathbf{v}_{i+1} + \cdots + 0\mathbf{v}_k$, $\mathbf{v}_i \in W$. Thus W is non-empty. Let $\mathbf{u}, \mathbf{v} \in W$. Then there are scalars λ_i, μ_i $(i = 1, 2, \ldots, k)$ such that

$$\mathbf{u} = \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \ldots + \lambda_k\mathbf{v}_k$$
$$\mathbf{v} = \mu_1\mathbf{v}_1 + \mu_2\mathbf{v}_2 + \ldots + \mu_k\mathbf{v}_k.$$

Thus we have

$$\begin{aligned}\lambda\mathbf{u} + \mu\mathbf{v} &= \lambda(\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \ldots + \lambda_k\mathbf{v}_k) \\ &\quad + \mu(\mu_1\mathbf{v}_1 + \mu_2\mathbf{v}_2 + \ldots + \mu_k\mathbf{v}_k) \\ &= (\lambda\lambda_1 + \mu\mu_1)\mathbf{v}_1 + (\lambda\lambda_2 + \mu\mu_2)\mathbf{v}_2 + \ldots \\ &\quad + (\lambda\lambda_k + \mu\mu_k)\mathbf{v}_k.\end{aligned}$$

Thus $\lambda\mathbf{u}+\mu\mathbf{v}$ is a linear combination of $\mathbf{v}_1$, $\mathbf{v}_2$, …, $\mathbf{v}_k \in V$. Therefore $\lambda\mathbf{u}+\mu\mathbf{v} \in W$ and W is therefore a subspace of V.

Since W' is a subspace of V, every linear combination of elements in W' is again in W'. Since $\mathbf{v}_1$, $\mathbf{v}_2$, …, $\mathbf{v}_k \in W'$, their linear combination is again in W'. Thus W is contained in W'. □

3. Span

DEFINITION 2.4. Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ be a subset of a vector space V. Then the subspace W consisting of all linear combination of elements in S is called the space spanned by $\mathbf{v}_1$, $\mathbf{v}_2$, …, $\mathbf{v}_k$ or simply span of S and is denoted by $\operatorname{span}(S)$ or $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$. If W is the span of S, we say that S spans W.

EXAMPLE 2.15. *Let V be a vector space. Let $\mathbf{0} \neq \mathbf{v} \in V$ and $S = \{\mathbf{v}\}$. Then*

$$\operatorname{span}(S) = \{\lambda\mathbf{v} : \lambda \in \mathbb{R}\}.$$

Let $\mathbf{v}_1, \mathbf{v}_2 \in V$ and $\mathbf{v}_1 \neq \lambda\mathbf{v}_2$ for any λ. If $S = \{\mathbf{v}_1, \mathbf{v}_2\}$, then

$$\operatorname{span} S = \{\lambda\mathbf{v}_1 + \mu\mathbf{v}_2 : \lambda, \mu \in \mathbb{R}\}.$$

EXAMPLE 2.16. *Consider the vector space $\mathbb{R}^n$. Consider the set S consisting of the following vectors*

$$\begin{aligned}
\mathbf{e}_1 &= (1, 0, 0, \cdots, 0)\\
\mathbf{e}_2 &= (0, 1, 0, \cdots, 0)\\
\mathbf{e}_3 &= (0, 0, 1, \cdots, 0)\\
&\vdots\\
\mathbf{e}_n &= (0, 0, 0, \cdots, 1).
\end{aligned}$$

For any vector $\mathbf{v} = (v_1, v_2, \cdots, v_n) \in \mathbb{R}^n$, we have

$$\mathbf{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + v_3\mathbf{e}_3 + \cdots + v_n\mathbf{e}_n.$$

Thus $\operatorname{span}(S) = \mathbb{R}^n$.

EXAMPLE 2.17. *Consider $\mathbf{v}_1 = (1, 0, 0)$, $\mathbf{v}_2 = (0, 1, 1)$ and $\mathbf{v}_3 = (1, 0, 1)$ in $\mathbb{R}^3$. Then*

$$\begin{aligned}
&\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \lambda_3\mathbf{v}_3\\
&= (\lambda_1 + \lambda_3, \lambda_2, \lambda_2 + \lambda_3)\\
&= (v_1, v_2, v_3)
\end{aligned}$$

where $v_1 = \lambda_1 + \lambda_3$, $v_2 = \lambda_2$, $v_3 = \lambda_3 + \lambda_2$ or $\lambda_1 = v_1 + v_2 - v_3$, $\lambda_2 = v_2$, $\lambda_3 = v_3 - v_2$. Thus any vector in $\mathbb{R}^3$ is a a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$ and $\mathbf{v}_3$. Thus

$$\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} = \mathbb{R}^3.$$

EXAMPLE 2.18. *Consider $\mathbf{v}_1 = (1, 0, 0)$, $\mathbf{v}_2 = (0, 1, 0)$ in $\mathbb{R}^3$. Then*

$$\begin{aligned}
&\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2\\
&= (\lambda_1, \lambda_2, 0).
\end{aligned}$$

Therefore

$$\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\} = \{(x, y, 0) : x, y \in \mathbb{R}\};$$

in this case, $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\} \neq \mathbb{R}^3$.

EXAMPLE 2.19. *The set* $\{1, x, x^2, \dots, x^n\}$ *spans the vector space of all polynomials of degree less than or equal to* n.

4. Linear dependence and independence

DEFINITION 2.5. Let V be a vector space. Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ be a subset of V. If there are scalars $\lambda_1, \lambda_2, \dots, \lambda_k$, not all zero, such that

$$\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \dots + \lambda_k\mathbf{v}_k = 0,$$

then the set S is *linearly dependent*. If S is not linearly dependent, then it is linearly independent. Thus S is linearly independent if

$$\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \dots + \lambda_k\mathbf{v}_k = 0,$$

then

$$\lambda_1 = \lambda_2 = \dots = \lambda_k = 0.$$

EXAMPLE 2.20. *A set with single vector* $\mathbf{v}$ *is linearly dependent if and only if there is a scalar* $\lambda \neq 0$ *such that* $\lambda\mathbf{v} = \mathbf{0}$. *This hold if and only if* $\mathbf{v} = \mathbf{0}$. *Thus a single element set* $\{\mathbf{v}\}$ *is linearly dependent if and only if* $\mathbf{v} = \mathbf{0}$.

The set $\{\mathbf{v}\}$ *is linearly independent if and only if* $\mathbf{v} \neq \mathbf{0}$.

EXAMPLE 2.21. *Consider a two element subset* $\{\mathbf{v}_1, \mathbf{v}_2\}$ *of a vector space* V. *The set is linearly dependent if there are scalars* λ_1 *and* λ_2, *not both zero, such that*

$$\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 = \mathbf{0}.$$

Since λ_1 *and* λ_2 *are not both zero, at least one of them, say* λ_1, *is nonzero. Then*

$$\mathbf{v}_1 = -\frac{\lambda_2}{\lambda_1}\mathbf{v}_2.$$

Thus $\mathbf{v}_1$ *is a scalar multiple of* $\mathbf{v}_2$.

If $\mathbf{v}_1$ *is a scalar multiple of* $\mathbf{v}_2$, *then* $\mathbf{v}_1 = \lambda\mathbf{v}_2$ *or* $\mathbf{v}_1 - \lambda\mathbf{v}_2 = \mathbf{0}$. *Thus* $\{\mathbf{v}_1, \mathbf{v}_2\}$ *is linearly dependent.*

Thus, the set $\{\mathbf{v}_1, \mathbf{v}_2\}$ *is linearly dependent if and only if one vector is scalar multiple of other vector. The set* $\{\mathbf{v}_1, \mathbf{v}_2\}$ *is linearly independent if and only if neither vector is scalar multiple of other vector.*

The set $\{(1,0,0), (0,1,0)\}$ *is linearly independent while* $\{(1,1,2), (2,2,4)\}$ *is linearly dependent.*

EXAMPLE 2.22. *Consider the vector space* $\mathbb{R}^n$. *Consider the set* S *consisting of the following vectors*

$$\begin{aligned}
\mathbf{e}_1 &= (1, 0, 0, \cdots, 0) \\
\mathbf{e}_2 &= (0, 1, 0, \cdots, 0) \\
\mathbf{e}_3 &= (0, 0, 1, \cdots, 0) \\
&\vdots \\
\mathbf{e}_n &= (0, 0, 0, \cdots, 1).
\end{aligned}$$

Then

$$\lambda_1\mathbf{e}_1 + \lambda_2\mathbf{e}_2 + \dots + \lambda_n\mathbf{e}_n = \mathbf{0}$$

becomes

$$(\lambda_1, \lambda_2, \ldots, \lambda_n) = (0, 0, \ldots, 0)$$

or

$$\lambda_1 = \lambda_2 = \cdots = \lambda_n = 0.$$

Thus S is a linearly independent set.

EXAMPLE 2.23. *Consider $\mathbf{v}_1 = (1, 0, 0)$, $\mathbf{v}_2 = (0, 1, 1)$ and $\mathbf{v}_3 = (1, 0, 1)$ in $\mathbb{R}^3$. Since*

$$\begin{aligned}&\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \lambda_3\mathbf{v}_3\\ &= (\lambda_1 + \lambda_3, \lambda_2, \lambda_2 + \lambda_3),\end{aligned}$$

the equation

$$\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \lambda_3\mathbf{v}_3 = \mathbf{0}$$

becomes

$$\lambda_1 + \lambda_3 = 0, \quad \lambda_2 = 0, \quad \lambda_2 + \lambda_3 = 0.$$

Solving these equations, we get

$$\lambda_1 = \lambda_2 = \lambda_3 = 0.$$

Thus $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a linearly independent set.

EXAMPLE 2.24. *The set $\{1, x, x^2, \ldots, x^n\}$ is linearly independent in the vector space of all polynomials of degree less than or equal to n. Let $\lambda_0, \lambda_1, \lambda_2, \ldots, \lambda_n$ be scalars such that*

$$\lambda_0 1 + \lambda_1 x + \lambda_2 x^2 + \cdots + \lambda_n x^n = 0.$$

Since this is true for all x, we get $\lambda_0 = 0$ by taking $x = 0$. Now by differentiating the above equation, we get

$$\lambda_1 + 2\lambda_2 x + \cdots + n\lambda_n x^{n-1} = 0.$$

By taking $x = 0$ in this equation, we get $\lambda_1 = 0$. By proceeding this way, we see that

$$\lambda_0 = \lambda_1 = \lambda_2 = \ldots = \lambda_n = 0$$

and therefore $\{1, x, x^2, \ldots, x^n\}$ is a linearly independent set.

EXAMPLE 2.25. *Consider $\mathbf{v}_1 = (1, 0, 0)$, $\mathbf{v}_2 = (0, 1, 1)$ and $\mathbf{v}_3 = (1, 1, 1)$ in $\mathbb{R}^3$. Since*

$$\begin{aligned}&\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \lambda_3\mathbf{v}_3\\ &= (\lambda_1 + \lambda_3, \lambda_2 + \lambda_3, \lambda_2 + \lambda_3),\end{aligned}$$

the equation

$$\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \lambda_3\mathbf{v}_3 = \mathbf{0}$$

becomes

$$\lambda_1 + \lambda_3 = 0, \quad \lambda_2 + \lambda_3 = 0, \quad \lambda_2 + \lambda_3 = 0.$$

Solving these equations, we get

$$\lambda_1 = \lambda_2 = t, \quad \lambda_3 = -t.$$

In particular, we can take $\lambda_1 = \lambda_2 = 1$ and $\lambda_3 = -1$. In this case, we have

$$\mathbf{v}_1 + \mathbf{v}_2 - \mathbf{v}_3 = \mathbf{0}.$$

Thus $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a linearly dependent set.

EXAMPLE 2.26. *If S is a subset of a vector space V and $\mathbf{0} \in S$, then S is linearly dependent. For, let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{k-1}, \mathbf{0}$ be the vectors in S. By taking $\lambda_1 = \lambda_2 = \cdots = \lambda_{k-1} = 0$ and $\lambda_k = 1$, we see that*

$$\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \ldots + \lambda_{k-1}\mathbf{v}_{k-1} + \lambda_k\mathbf{0} = \mathbf{0}.$$

Since not all λ_i's are zero (in this case, $\lambda_k \neq 0$), we see that S is linearly dependent.

THEOREM 2.7. *A subset $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ with two or more vector of a vector space V is linearly dependent if and only if at least one of the vector in S is a linear combination of other vectors in S.*

PROOF. Let S be linearly dependent. Then there are scalars $\lambda_1, \lambda_2, \dots, \lambda_k$, not all zero, such that

$$\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \dots + \lambda_k\mathbf{v}_k = \mathbf{0}.$$

Let $\lambda_i \neq 0$. Then

$$\lambda_i\mathbf{v}_i = -\lambda_1\mathbf{v}_1 - \lambda_2\mathbf{v}_2 - \dots - \lambda_{i-1}\mathbf{v}_{i-1}$$

$$-\lambda_{i+1}\mathbf{v}_{i+1} - \dots - \lambda_k\mathbf{v}_k$$

or

$$\mathbf{v}_i = \frac{-\lambda_1}{\lambda_1}\mathbf{v}_1 - \frac{\lambda_2}{\lambda_1}\mathbf{v}_2 - \dots - \frac{\lambda_{i-1}}{\lambda_1}\mathbf{v}_{i-1}$$

$$-\frac{\lambda_{i+1}}{\lambda_1}\mathbf{v}_{i+1} - \dots - \frac{\lambda_k}{\lambda_1}\mathbf{v}_k.$$

This shows that $\mathbf{v}_i$ is linear combination of other vectors in S.

Conversely, let one of the vector, say $\mathbf{v}_1$, is linear combination of other vectors in S. Then there are scalars $\lambda_2, \lambda_3, \dots, \lambda_k$ such that

$$\mathbf{v}_1 = \lambda_2\mathbf{v}_2 + \dots + \lambda_k\mathbf{v}_k.$$

Thus

$$-\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \dots + \lambda_k\mathbf{v}_k = \mathbf{0}.$$

Since not all constants in the linear combination is zero, it follows that S is linearly dependent. ☐

EXAMPLE 2.27. *Consider $\mathbf{v}_1 = (1, 0, 0)$, $\mathbf{v}_2 = (1, 0, 1)$ and $\mathbf{v}_3 = (3, 0, 1)$ in $\mathbb{R}^3$. Since*

$$\begin{aligned} 2\mathbf{v}_1 + \mathbf{v}_2 &= 2(1, 0, 0) + (1, 0, 1) \\ &= (2, 0, 0) + (1, 0, 1) \\ &= (3, 0, 1) \\ &= \mathbf{v}_3, \end{aligned}$$

the set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly dependent in $\mathbb{R}^3$.

THEOREM 2.8. *Any subset S of $\mathbb{R}^n$ having more than n vectors is linearly dependent.*

PROOF. Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ be a subset of $\mathbb{R}^n$ and $k > n$. Consider the equation

$$\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \dots + \lambda_k\mathbf{v}_k = \mathbf{0}.$$

We need to show that there is a solution where not all λ_i's are zero. Each side of the above equation is a vector in $\mathbb{R}^n$. By equating the corresponding coordinates, we get n homogeneous equations in k unknowns. Since a homogeneous linear system with more unknowns than the number of equations has nontrivial solution, we have non-zero solutions. Thus the set S is linearly dependent. ☐

5. Basis and dimension

DEFINITION 2.6. A subset S of a vector space V is called a **basis** for V if (a) S is linearly independent and (b) S spans V.

EXAMPLE 2.28. *Consider the vector space $\mathbb{R}^n$. Consider the set S consisting of the following vectors*

$$\begin{aligned}\mathbf{e}_1 &= (1, 0, 0, \cdots, 0)\\ \mathbf{e}_2 &= (0, 1, 0, \cdots, 0)\\ \mathbf{e}_3 &= (0, 0, 1, \cdots, 0)\\ &\vdots\\ \mathbf{e}_n &= (0, 0, 0, \cdots, 1).\end{aligned}$$

This S is a linearly independent set and also spans $\mathbb{R}^n$. Thus it is a basis for $\mathbb{R}^n$. It is called the ***standard basis*** *for $\mathbb{R}^n$.*

EXAMPLE 2.29. *Consider $\mathbf{v}_1 = (1, 0, 0)$, $\mathbf{v}_2 = (0, 1, 1)$ and $\mathbf{v}_3 = (1, 0, 1)$ in $\mathbb{R}^3$. We have shown that the set $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ spans $\mathbb{R}^3$ and linearly independent. Thus S is basis for $\mathbb{R}^3$.*

EXAMPLE 2.30. *The set $S = \{1, x, x^2, \ldots, x^n\}$ is linearly independent in the vector space P_n of all polynomials of degree less than or equal to n. Also it spans P_n and therefore S is basis for P_n.*

THEOREM 2.9. *Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ be a basis for a vector space V. Then every element $\mathbf{v} \in V$ can be expressed as a unique linear combination of elements in S:*

$$\mathbf{v} = \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_k\mathbf{v}_k.$$

PROOF. Since S is a basis for V, $V = \text{span}(S)$. Thus every vector in V is a linear combination of elements in S. We need to show that $\mathbf{v}$ can be expressed uniquely as linear combination of elements in S. Or equivalently, if $\mathbf{v}$ is expressed as linear combination in two ways

$$\mathbf{v} = \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_k\mathbf{v}_k$$

and

$$\mathbf{v} = \mu_1\mathbf{v}_1 + \mu_2\mathbf{v}_2 + \cdots + \mu_k\mathbf{v}_k,$$

then we need to show that

$$\lambda_1 = \mu_1, \ldots, \lambda_k = \mu_k.$$

By subtracting the two expression for $\mathbf{v}$, we now get

$$(\lambda_1 - \mu_1)\mathbf{v}_1 + \cdots + (\lambda_k - \mu_k)\mathbf{v}_k = \mathbf{0}.$$

Since S is a basis for V, the set S is linearly independent. Hence

$$\lambda_1 - \mu_1 = 0, \ldots, \lambda_k - \mu_k = 0$$

or

$$\lambda_1 = \mu_1, \ldots, \lambda_k = \mu_k.$$

□

Definition 2.7. Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ be a basis for a vector space V. Then every element $\mathbf{v} \in V$ can be expressed as a unique linear combination of elements in S:

$$\mathbf{v} = \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_k\mathbf{v}_k.$$

The scalars $\lambda_1, \lambda_2, \dots, \lambda_k$ are called the coordinates of the vector $\mathbf{v}$ with respect to the basis S. The vector $(\mathbf{v})_S := (\lambda_1, \lambda_1, \dots, \lambda_k) \in \mathbb{R}^k$ is the coordinate vector of $\mathbf{v}$ relative to S.

Example 2.31. *The coordinate vector of $\mathbf{v} = (v_1, v_2, v_3)$ relative to the standard basis of $\mathbb{R}^3$ is $(\mathbf{v})_S = (v_1, v_2, v_3)$ which is same as the vector $\mathbf{v}$.*

Consider the set $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ where $\mathbf{v}_1 = (1, 0, 0)$, $\mathbf{v}_2 = (0, 1, 1)$ and $\mathbf{v}_3 = (1, 0, 1)$ are in $\mathbb{R}^3$. Thus S is basis for $\mathbb{R}^3$. For any vector $\mathbf{v} = (v_1, v_2, v_3)$, we have

$$\mathbf{v} = (v_1 + v_2 - v_3)\mathbf{v}_1 + v_2\mathbf{v}_2 + (v_3 - v_2)\mathbf{v}_3.$$

Therefore the coordinate vector of $\mathbf{v}$ relative to the basis S is given by

$$(\mathbf{v})_S = (v_1 + v_2 - v_3, v_2, v_3 - v_2).$$

Example 2.32. *Consider the vector space M_2 of all matrices of order 2. Consider also the set*

$$S = \{E_1, E_2, E_3, E_4\}$$

where

$$E_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, E_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},$$
$$E_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, E_4 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Then any matrix

$$\mathbf{v} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

can be written as a linear combination of E_1, E_2, E_3 and E_4:

$$\mathbf{v} = aE_1 + bE_2 + cE_3 + dE_4.$$

Hence $\operatorname{span}(S) = M_2$. *Also if there are scalars $\lambda_1, \dots, \lambda_4$ such that*

$$\lambda_1 E_1 + \lambda_2 E_2 + \lambda_3 E_3 + \lambda_4 E_4 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

then we have

$$\begin{pmatrix} \lambda_1 & \lambda_2 \\ \lambda_3 & \lambda_4 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

and therefore

$$\lambda_1 = \ldots = \lambda_4 = 0.$$

Thus S is linearly independent and therefore S is a basis for M_2. For the matrix

$$\mathbf{v} = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

we have

$$(\mathbf{v})_S = (a, b, c, d).$$

Definition 2.8. A vector space V is finite-dimensional if it has a basis with finite number of elements. The vector space $\{\mathbf{0}\}$ is also considered as finite-dimensional (even though there is no basis for this vector space). Otherwise it is called infinite-dimensional.

The number of elements in a basis for a finite-dimensional vector space V is called the dimension of the vector space and is denoted by $\dim(V)$.

The above definition of dimension make sense only when all the basis of a finite-dimensional vector space have same number of elements. In fact, this is true and will be proved shortly.

THEOREM 2.10. *Let V be finite-dimensional vector space. Let S be a basis for V with n vectors. Any set with more than n elements is linearly dependent.*

PROOF. Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis for V and $S' = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$ be any set and $m > n$. We will show that S' is linearly dependent. To do this we must find scalars $\lambda_1, \lambda_2, \ldots, \lambda_m$ not all zero such that

$$\lambda_1\mathbf{w}_1 + \lambda_2\mathbf{w}_2 + \cdots + \lambda_m\mathbf{w}_m = \mathbf{0}. \tag{1}$$

Since S is a basis for V, all the vectors in S' are linear combination of vectors in S. Thus we have scalars a_{ij} such that

$$\begin{array}{ccccccccc} \mathbf{w}_1 & = & a_{11}\mathbf{v}_1 & + & a_{21}\mathbf{v}_2 & + & \cdots & + & a_{n1}\mathbf{v}_n \\ \mathbf{w}_2 & = & a_{12}\mathbf{v}_1 & + & a_{22}\mathbf{v}_2 & + & \cdots & + & a_{n2}\mathbf{v}_n \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ \mathbf{w}_m & = & a_{1m}\mathbf{v}_1 & + & a_{2m}\mathbf{v}_2 & + & \cdots & + & a_{nm}\mathbf{v}_n. \end{array}$$

Since $\mathbf{v}_i$'s are linearly independent and

$$\begin{aligned} \mathbf{0} &= \lambda_1\mathbf{w}_1 + \lambda_2\mathbf{w}_2 + \cdots + \lambda_m\mathbf{w}_m \\ &= (\lambda_1 a_{11} + \lambda_2 a_{12} + \cdots + \lambda_m a_{1m})\mathbf{v}_1 \\ &\quad + (\lambda_1 a_{21} + \lambda_2 a_{22} + \cdots + \lambda_m a_{2m})\mathbf{v}_2 \\ &\quad + \cdots \\ &\quad + (\lambda_1 a_{n1} + \lambda_2 a_{n2} + \cdots + \lambda_m a_{nm})\mathbf{v}_n, \end{aligned}$$

we have

$$\begin{aligned} \lambda_1 a_{11} + \lambda_2 a_{12} + \cdots + \lambda_m a_{1m} &= 0 \\ \lambda_1 a_{21} + \lambda_2 a_{22} + \cdots + \lambda_m a_{2m} &= 0 \\ &\vdots \\ \lambda_1 a_{n1} + \lambda_2 a_{n2} + \cdots + \lambda_m a_{nm} &= 0. \end{aligned}$$

Since there are n equations in m variables and $m > n$, there are nontrivial solution to the above linear system. Thus there are scalars $\lambda_1, \lambda_2, \ldots, \lambda_m$ not all zero such that (**??**) holds. This proves that S' is linearly dependent. □

THEOREM 2.11. *Let V be finite-dimensional vector space. Let S be a basis for V with n vectors. Any set with less than n elements does not span V.*

PROOF. Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis for V and $S' = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$ be any set and $m < n$. We will show that S' does not span V. Assume, on the contrary, that S' spans V. Then we show that this implies that S is linearly dependent, a contraction.

Since S' spans V, all the vectors in S are linear combination of vectors in S'. Thus we have scalars a_{ij} such that

$$\begin{array}{ccccccccc} \mathbf{v}_1 & = & a_{11}\mathbf{w}_1 & + & a_{21}\mathbf{w}_2 & + & \cdots & + & a_{m1}\mathbf{w}_m \\ \mathbf{v}_2 & = & a_{12}\mathbf{w}_1 & + & a_{22}\mathbf{w}_2 & + & \cdots & + & a_{m2}\mathbf{w}_m \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ \mathbf{v}_n & = & a_{1n}\mathbf{w}_1 & + & a_{2n}\mathbf{w}_2 & + & \cdots & + & a_{mn}\mathbf{w}_m. \end{array}$$

Note that

$$\begin{aligned}\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_n\mathbf{v}_n \\ =&(\lambda_1 a_{11} + \lambda_2 a_{12} + \cdots + \lambda_n a_{1n})\mathbf{w}_1 \\ &+ (\lambda_1 a_{21} + \lambda_2 a_{22} + \cdots + \lambda_n a_{2n})\mathbf{w}_2 \\ &+ \cdots \\ &+ (\lambda_1 a_{m1} + \lambda_2 a_{m2} + \cdots + \lambda_n a_{mn})\mathbf{w}_m,\end{aligned}$$

Consider the equation

$$\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_n\mathbf{v}_n = \mathbf{0}. \tag{2}$$

Also consider the system of equations

$$\begin{aligned}\lambda_1 a_{11} + \lambda_2 a_{12} + \cdots + \lambda_n a_{1n} &= 0 \\ \lambda_1 a_{21} + \lambda_2 a_{22} + \cdots + \lambda_n a_{2n} &= 0 \\ \cdots \\ \lambda_1 a_{m1} + \lambda_2 a_{m2} + \cdots + \lambda_n a_{mn} &= 0\end{aligned}$$

Since there are m equations in n variables and $n > m$, there are nontrivial solution to the above linear system. Thus there are scalars $\lambda_1, \lambda_2, \ldots, \lambda_n$ not all zero such that (**??**) holds. This proves that S is linearly dependent, a contraction to the assumption that S is a basis for V. Thus it follows that S' does not span V. □

THEOREM 2.12. *Any two bases for a finite-dimensional vector space V have same number of elements.*

PROOF. Let S be a basis for V having n elements. Let S' be another basis for V with n' elements. Since S is a basis for V, any set with more than n elements is linearly dependent. But S' is linearly independent, so we must have $n' \leq n$.

Since S is a basis for V, any set with less than n elements does not span V. But S' spans V and therefore we must have $n' \geq n$. Thus we have $n = n'$. □

EXAMPLE 2.33. *Since $\mathbb{R}^n$ has a basis with n elements,* $\dim(\mathbb{R}^n) = n$*. Also the vector space P_n of all polynomial of degree n has a basis with $n+1$ elements and therefore* $\dim(P_n) = n+1$*. The vector space M_2 of all square matrices of order 2 has a basis with four elements and therefore* $\dim(M_2) = 4$*. In general,* $\dim(M_{m\times n}) = mn$*.*

THEOREM 2.13. *Let V be a vector space and S be a nonempty subset of V. If S is linearly independent subset of V and $v \notin$ span(S), then the set $S' = S \cup \{v\}$ is also linearly independent.*

PROOF. Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$. We need to show that $S' = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k, \mathbf{v}\}$ is linearly independent. In other words, we need to show that

$$\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_k\mathbf{v}_k + \lambda_{k+1}\mathbf{v} = \mathbf{0}$$

implies

$$\lambda_1 = \lambda_2 = \cdots = \lambda_k = \lambda_{k+1} = 0.$$

The scalar $\lambda_{k+1} = 0$. For, if this scalar is nonzero, then

$$\mathbf{v} = -\frac{\lambda_1}{\lambda_{k+1}}\mathbf{v}_1 - \frac{\lambda_2}{\lambda_{k+1}}\mathbf{v}_2 + \cdots - \frac{\lambda_k}{\lambda_{k+1}}\mathbf{v}_k$$

and therefore $\mathbf{v} \in \operatorname{span}(S)$, contrary to the assumption that $\mathbf{v} \notin \operatorname{span}(S)$.

Since $\lambda_{k+1} = 0$, we have

$$\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \cdots + \lambda_k \mathbf{v}_k = \mathbf{0}$$

and by linear independence of S, it follows that

$$\lambda_1 = \lambda_2 = \cdots = \lambda_k = 0.$$

This proves that S' is linearly independent. □

THEOREM 2.14. *Let V be a vector space and S be a nonempty subset of V. If $\mathbf{v} \in S$ and $\mathbf{v} \in \operatorname{span}(S - \{\mathbf{v}\})$, then*

$$\operatorname{span}(S) = \operatorname{span}(S - \{\mathbf{v}\}).$$

PROOF. Let $S = \{\mathbf{v}, \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$. Since $\mathbf{v} \in \operatorname{span}(S - \{\mathbf{v}\})$, $\mathbf{v}$ is a linear combination of vectors in $S - \{\mathbf{v}\} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$; that is,

$$\mathbf{v} = \lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \cdots + \lambda_k \mathbf{v}_k.$$

To prove the theorem, we need to show that every vector that is expressible as linear combination of $\mathbf{v}, \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ is expressible as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$. Consider a vector $\mathbf{u} \in \operatorname{span} S$. Then

$$\begin{aligned} \mathbf{u} &= \mu_0 \mathbf{v} + \mu_1 \mathbf{v}_1 + \mu_2 \mathbf{v}_2 + \cdots + \mu_k \mathbf{v}_k \\ &= \mu_0(\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \cdots + \lambda_k \mathbf{v}_k) \\ &\qquad + \mu_1 \mathbf{v}_1 + \mu_2 \mathbf{v}_2 + \cdots + \mu_k \mathbf{v}_k \\ &= (\mu_0 \lambda_1 + \mu_1)\mathbf{v}_1 + \ldots + (\mu_0 \lambda_k + \mu_k)\mathbf{v}_k. \end{aligned}$$

Thus $\mathbf{u}$ is a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$. □

THEOREM 2.15. *Let V be a finite-dimensional vector space and S be a finite subset of V.*

(a) *If S spans V, but not a basis for V, a basis for V can obtained from S by removing appropriate vectors.*
(b) *If S is linearly independent, but not a basis for V, then S can enlarged to a basis for V by adding appropriate vectors into S.*

PROOF OF (a). If S spans V but not a basis for V, then S is linearly dependent and therefore there is a vector $\mathbf{v} \in S$ which linear combination of other vectors in S. The set S' obtained by removing $\mathbf{v}$ from S still spans V. If this set S' is linearly independent, then S' is a basis for V. Otherwise there is a vector $\mathbf{v}'$ in S' which linear combination of other vectors in S'. The set S'' obtained by removing $\mathbf{v}'$ from S' still spans V. If this set S'' is linearly independent, then S'' is a basis for V. This procedure is continued until we get a subset of S that is linearly independent and spans V. □

PROOF OF (b). If S is linearly independent set in V but not a basis for V, then S does not span V. The set S can be extended to another set S' by adding appropriate vector $\mathbf{v}$ so that S' is linearly independent. If S' spans V, then S' is a basis for V. If S' does not span V, then the set S' can be extended to another set S'' by adding appropriate vector $\mathbf{v}'$ so that S'' is linearly independent. If S'' spans V, then S'' is a basis for V. This procedure is continued until we get a linearly independent set that spans V. □

THEOREM 2.16. *Let V be an n-dimensional vector space and S be a subset of V with exactly n vectors. Then S is a basis for V if either S spans V or S is linearly independent.*

PROOF. Let V be vectors space of dimension n. Let S spans V. Then to show that S is a basis, we need to show that S is linearly independent. If S is not linearly independent, then some vector $\mathbf{v}$ is linear combination of other vectors in S. The set S' obtained from S by removing the vector $\mathbf{v}$ still spans V, which is not possible since S' has only $n-1$ elements and V is of dimension n. Thus S should be linearly independent and therefore S is a basis for V. (A set with less than n elements in a n-dimensional vector space V cannot span V.)

Let S be linearly independent. We will show that S spans V so that S becomes a basis for V. If S does not span, then we can find a vector $\mathbf{v}$ such that the set S' obtained adding $\mathbf{v}$ is still linearly independent, which is not possible since S' has $n+1$ elements and V of dimension n. (A set with more than n elements in a n-dimensional vector space V cannot be linearly independent in V.) □

EXAMPLE 2.34. *Consider the set S with three vectors $\mathbf{v}_1 = (1,2,3)$, $\mathbf{v}_2 = (2,5,3)$ and $\mathbf{v}_3 = (1,0,8)$. Let λ_1, λ_2 and λ_3 be scalars such that $\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \lambda_3\mathbf{v}_3 = \mathbf{0}$. Therefore we have*

$$\begin{array}{rcrcrl} \lambda_1 & + & 2\lambda_2 & + & \lambda_3 & = 0, \\ 2\lambda_1 & + & 5\lambda_2 & & & = 0, \\ 3\lambda_1 & + & 3\lambda_2 & + & 8\lambda_3 & = 0. \end{array}$$

Since the matrix

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 5 & 0 \\ 3 & 3 & 8 \end{pmatrix}$$

has nonzero determinant ($\det(A) = -1$), it is invertible and therefore there is only a trivial solution to the above equations. Thus the set S is linearly independent. Therefore S is a basis for $\mathbb{R}^3$.

EXAMPLE 2.35. *Consider the set S with three vectors $\mathbf{v}_1 = (1,0,0,0)$, $\mathbf{v}_2 = (a,1,0,0)$, $\mathbf{v}_3 = (b,c,1,0)$ and $\mathbf{v}_4 = (d,e,f,1)$. Let λ_1, λ_2, λ_3 and λ_4 be scalars such that $\lambda_1\mathbf{v}_1 + \lambda_2 + \lambda_3\mathbf{v}_3 + \lambda_4\mathbf{v}_4 = \mathbf{0}$. Therefore we have*

$$\begin{array}{rcrcrcrl} \lambda_1 & + & a\lambda_2 & + & b\lambda_3 & + & d\lambda_4 & = 0, \\ & & \lambda_2 & + & c\lambda_3 & + & e\lambda_4 & = 0, \\ & & & & \lambda_3 & + & f\lambda_4 & = 0, \\ & & & & & & \lambda_4 & = 0. \end{array}$$

Since the matrix

$$A = \begin{pmatrix} 1 & a & b & d \\ 0 & 1 & c & e \\ 0 & 0 & 1 & f \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

has nonzero determinant ($\det(A) = 1$), it is invertible and therefore there is only a trivial solution to the above equations. Thus the set S is linearly independent. Therefore S is a basis for $\mathbb{R}^4$.

EXAMPLE 2.36. *Consider the set S with three vectors $\mathbf{v}_1 = (1,1,1)$, $\mathbf{v}_2 = (1,1,0)$ and $\mathbf{v}_3 = (1,0,0)$. Since*

$$\mathbf{v}_3 = (1,0,0), \mathbf{v}_2 - \mathbf{v}_3 = (0,1,0), \mathbf{v}_1 - \mathbf{v}_2 = (0,0,1),$$

any vector $\mathbf{v} = (\lambda_1, \lambda_2, \lambda_3) \in \mathbb{R}^3$ can be written as

$$\begin{aligned} \mathbf{v} &= \lambda_1\mathbf{v}_3 + \lambda_2(\mathbf{v}_2 - \mathbf{v}_3) + \lambda_3(\mathbf{v}_1 - \mathbf{v}_2) \\ &= \lambda_3\mathbf{v}_1 + (\lambda_2 - \lambda_3)\mathbf{v}_2 + (\lambda_1 - \lambda_2)\mathbf{v}_3. \end{aligned}$$

Thus S spans $\mathbb{R}^3$ and therefore it is a basis for $\mathbb{R}^3$.

6. Row Space, Column Space and Null Space

DEFINITION 2.9. Let $A = [a_{ij}]$ be an $m \times n$ matrix. The vectors

$$\begin{aligned}\mathbf{r}_1 &= (a_{11}\ a_{12}\ a_{13}\ \cdots\ a_{1n}),\\ \mathbf{r}_2 &= (a_{21}\ a_{22}\ a_{23}\ \cdots\ a_{2n}),\\ &\vdots\\ \mathbf{r}_m &= (a_{m1}\ a_{m2}\ a_{m3}\ \cdots\ a_{mn}),\end{aligned}$$

obtained from the rows of A, are called the **row vectors** of A. The vectors

$$\mathbf{c}_1 = \begin{pmatrix} a_{11}\\ a_{21}\\ a_{31}\\ \vdots\\ a_{m1}\end{pmatrix}, \quad \mathbf{c}_2 = \begin{pmatrix} a_{12}\\ a_{22}\\ a_{32}\\ \vdots\\ a_{m2}\end{pmatrix}, \cdots, \mathbf{c}_n = \begin{pmatrix} a_{1n}\\ a_{2n}\\ a_{3n}\\ \vdots\\ a_{mn}\end{pmatrix},$$

obtained from the columns of A, are called the **column vectors** of A. The vector space spanned by the row vectors of A is called the **row space** of A and the vector space spanned by the column vectors is the **column space** of A. The vector space of all solutions of the linear system $A\mathbf{x} = \mathbf{0}$ is the **null space** of A. Note that row space and null spaces are subspaces of $\mathbb{R}^n$ while the column space is a subspace of $\mathbb{R}^m$.

EXAMPLE 2.37. *Consider the matrix*

$$A = \begin{pmatrix} 1 & 0 & 1\\ 1 & 1 & 1\end{pmatrix}.$$

The vectors $\mathbf{r}_1 = (1\ 0\ 1)$ *and* $\mathbf{r}_2 = (1\ 1\ 1)$ *are the row vectors of* A *while* $\mathbf{c}_1 = (1\ 1)^T$ *and* $\mathbf{c}_2 = (0\ 1)^T$ *and* $\mathbf{c}_3 = \mathbf{c}_1$ *are the column vectors of* A. *The subspace*

$$R = \{\lambda \mathbf{r}_1 + \mu \mathbf{r}_2 : \lambda, \mu \in \mathbb{R}\}$$

is the row space of A. *The subspace*

$$C = \{\lambda \mathbf{c}_1 + \mu \mathbf{c}_2 : \lambda, \mu \in \mathbb{R}\}$$

is the column space of A.

The solutions of the system $A\mathbf{x} = \mathbf{0}$ *are given by the equation*

$$x + z = 0, \quad x + y + z = 0$$

or equivalently by

$$x + z = 0, \quad y = 0.$$

By solving this, we get $x = t, y = 0, z = -t$. *Thus the set of all solutions of* $A\mathbf{x} = \mathbf{0}$ *is* $N = \{(t\ 0\ -t) : t \in \mathbb{R}\}$. *This is clearly a vector subspace of* $\mathbb{R}^3$ *and is the null space of* A.

THEOREM 2.17. *A linear system* $A\mathbf{x} = \mathbf{b}$ *is consistent if and only if* $\mathbf{b}$ *is in the column space of* A.

PROOF. Let $A = [a_{ij}]$ be a matrix of order $m \times n$ and $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n$ be column vectors of A. Then for any $\mathbf{x} = (x_1\ x_2\ x_3 \ldots x_n)^T$, we have

$$\begin{aligned} A\mathbf{x} &= (\mathbf{c}_1\ \mathbf{c}_2 \ldots \mathbf{c}_n)(x_1\ x_2\ \ldots x_n)^T\\ &= x_1\mathbf{c}_1 + x_2\mathbf{c}_2 + \cdots + x_n\mathbf{c}_n.\end{aligned}$$

Since the column space of A is just the set of all linear combinations of the column vectors, we see that $A\mathbf{x} = \mathbf{b}$ is consistent if and only if

$$\mathbf{b} = x_1\mathbf{c}_1 + x_2\mathbf{c}_2 + \cdots + x_n\mathbf{c}_n.$$

Thus $A\mathbf{x} = \mathbf{b}$ if and only if $\mathbf{b}$ is the column space of A. □

THEOREM 2.18. *A linear system $A\mathbf{x} = \mathbf{0}$ is has only trivial solution if and only if the column vectors of A are linearly independent.*

PROOF. Let $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n$ be column vectors of A so that

$$A\mathbf{x} = x_1\mathbf{c}_1 + x_2\mathbf{c}_2 + \cdots + x_n\mathbf{c}_n.$$

Let $A\mathbf{x} = 0$ has only trivial solutions. Consider

$$x_1\mathbf{c}_1 + x_2\mathbf{c}_2 + \cdots + x_n\mathbf{c}_n = \mathbf{0}.$$

Since this is same as $A\mathbf{x} = \mathbf{0}$, it follows that $\mathbf{x} = \mathbf{0}$ or $x_1 = x_2 = \cdots = x_n = 0$. Thus the column vectors are linearly independent.

Conversely assume that the column vectors of A are linearly independent. Consider $A\mathbf{x} = \mathbf{0}$. This is same as

$$x_1\mathbf{c}_1 + x_2\mathbf{c}_2 + \cdots + x_n\mathbf{c}_n = \mathbf{0}.$$

By linear independence of the column vectors, we get $x_1 = x_2 = \cdots = x_n = 0$ or $\mathbf{x} = 0$. Thus $A\mathbf{x} = \mathbf{0}$ has only a trivial solution. □

EXAMPLE 2.38. *Consider the linear system $A\mathbf{x} = \mathbf{b}$ where*

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad \textit{and} \quad \mathbf{b} = \begin{pmatrix} 4 \\ -1 \\ -1 \end{pmatrix}.$$

The augmented matrix for the system is

$$\begin{pmatrix} 1 & 0 & 1 & 4 \\ 1 & 1 & 0 & -1 \\ 0 & 1 & 1 & -1 \end{pmatrix}.$$

By subtracting the first row from the second, second row from the third and then dividing the third row by 2, we get the following matrix

$$\begin{pmatrix} 1 & 0 & 1 & 4 \\ 0 & 1 & -1 & -5 \\ 0 & 0 & 1 & 2 \end{pmatrix}.$$

Thus the linear system becomes $x + z = 4$, $y - z = -5$ and $z = 2$. Thus we have

$$x = 2, y = -3, z = 2.$$

Note that the vector $\mathbf{b}$ is in the column space of A, for

$$\begin{pmatrix} 4 \\ -1 \\ -1 \end{pmatrix} = 2\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} - 3\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + 2\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

The vector

$$\mathbf{b} = \begin{pmatrix}1\\1\\0\end{pmatrix} + \begin{pmatrix}0\\1\\1\end{pmatrix} = \begin{pmatrix}1\\2\\1\end{pmatrix}$$

is in the column space of A and therefore the system $A\mathbf{x} = \mathbf{b}$ is consistent and the solution is given by $x = 1, y = 1, z = 0$.

THEOREM 2.19. *Let $\mathbf{x}_0$ be any solution of consistent linear system $A\mathbf{x} = \mathbf{b}$ and $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ be basis vectors of the null space of A. Every solution $\mathbf{x}$ of $A\mathbf{x} = \mathbf{b}$ can be written as*

$$\mathbf{x} = \mathbf{x}_0 + \lambda_1\mathbf{x}_1 + \cdots + \lambda_k\mathbf{x}_k, \tag{3}$$

*where $\lambda_1, \ldots, \lambda_k$ are scalars. Conversely the vector $\mathbf{x}$ given by (**??**) is a solution of $A\mathbf{x} = \mathbf{b}$ every choice of scalars $\lambda_1, \ldots, \lambda_k$.*

PROOF. Since $\mathbf{x}_0$ is a solution of $A\mathbf{x} = \mathbf{b}$, we have $A\mathbf{x}_0 = \mathbf{b}$. If $\mathbf{x}$ is any solution of $A\mathbf{x} = \mathbf{b}$, then $A(\mathbf{x} - \mathbf{x}_0) = A\mathbf{x} - A\mathbf{x}_0 = \mathbf{b} - \mathbf{b} = \mathbf{0}$ and therefore $\mathbf{x} - \mathbf{x}_0$ is solution of $A\mathbf{x} = \mathbf{0}$; in other words, $\mathbf{x} - \mathbf{x}_0$ is in the null space of A. Since $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ are the basis vectors of the null space of A, there are scalars $\lambda_1, \ldots, \lambda_k$ such that

$$\mathbf{x} - \mathbf{x}_0 = \lambda_1\mathbf{x}_1 + \cdots + \lambda_k\mathbf{x}_k$$

or

$$\mathbf{x} = \mathbf{x}_0 + \lambda_1\mathbf{x}_1 + \cdots + \lambda_k\mathbf{x}_k.$$

This proves (**??**).

To prove the converse, let us compute $A\mathbf{x}$. Since $\mathbf{x}_0$ be any solution of $A\mathbf{x} = \mathbf{b}$, we have $A\mathbf{x}_0 = \mathbf{b}$. Since $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ are basis vectors of the null space of A, they are solutions of $A\mathbf{x} = \mathbf{0}$ and therefore $A\mathbf{x}_r = \mathbf{0}$ for every $r = 1, 2, \ldots, k$. Now

$$\begin{aligned} A\mathbf{x} &= A(\mathbf{x}_0 + \lambda_1\mathbf{x}_1 + \cdots + \lambda_k\mathbf{x}_k) \\ &= A\mathbf{x}_0 + \lambda_1 A\mathbf{x}_1 + \cdots + \lambda_k A\mathbf{x}_k \\ &= \mathbf{b} + \mathbf{0} + \mathbf{0} + \cdots + \mathbf{0} \\ &= \mathbf{b}. \end{aligned}$$

Thus $\mathbf{x}$ given by (**??**) is a solution of $A\mathbf{x} = \mathbf{b}$ for every choice of scalars $\lambda_1, \ldots, \lambda_k$. □

Any solution $\mathbf{x} = \mathbf{x}_0$ of the linear system $A\mathbf{x} = \mathbf{b}$ is called a **particular solution** of $A\mathbf{x} = \mathbf{b}$. The solution $\mathbf{x}$ given by (**??**) is the **general solution** of $A\mathbf{x} = \mathbf{b}$. The solution $\mathbf{x} = \lambda_1\mathbf{x}_1 + \cdots + \lambda_k\mathbf{x}_k$ is the **general solution** of $A\mathbf{x} = \mathbf{0}$.

EXAMPLE 2.39. *Consider the linear system*

$$\begin{pmatrix}1 & 0 & 0\\0 & 1 & 1\\1 & 1 & 1\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{pmatrix}1\\2\\3\end{pmatrix}.$$

Clearly $\mathbf{x}_0 = (1\ 1\ 1)^T$ is a solution of the given linear system. The solution space of the homogeneous linear system

$$\begin{pmatrix}1 & 0 & 0\\0 & 1 & 1\\1 & 1 & 1\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix}$$

is $\{(0,t,-t)^T : t \in \mathbb{R}\}$ *and its basis vector is* $\mathbf{x}_1 = (0,1,-1)^T$. *Thus the general solution of the given nonhomogeneous linear system is*

$$\mathbf{x} = \mathbf{x}_0 + t\mathbf{x}_1 = \begin{pmatrix}1\\1\\1\end{pmatrix} + t\begin{pmatrix}0\\1\\-1\end{pmatrix} = \begin{pmatrix}1\\1+t\\1-t\end{pmatrix}.$$

THEOREM 2.20. *Let B be a matrix obtained from the matrix A by a sequence of elementary row operations. Then the null space of the matrix A and the null space of B are same.*

PROOF. Since the solutions of the linear system $A\mathbf{x} = \mathbf{0}$ is same as the solutions of the linear system $B\mathbf{x} = \mathbf{0}$, the result follows. □

THEOREM 2.21. *Let B be a matrix obtained from the matrix A by a sequence of elementary row operations. Then the row space of a matrix A and the row space of B are same.*

PROOF. It is enough to prove that the row operations does not change the row space of a given matrix. There are three kinds of row operations. First, consider the row operation of interchanging two rows of A to get A'. Clearly the rows of A and A' are same and therefore their row spaces are same.

Now let A' be obtained from A by multiplying ith row of A by a nonzero scalar λ or by adding λ times jth row of A to the ith row of A. Let $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_k$ and $\mathbf{r}'_1, \mathbf{r}'_2, \dots, \mathbf{r}'_k$ be the row vectors of A and A' respectively. In this case, we have

$$\begin{aligned}
\mathbf{r}'_1 &= \mathbf{r}_1,\\
&\vdots\\
\mathbf{r}'_{i-1} &= \mathbf{r}_{i-1},\\
\mathbf{r}'_i &= \lambda\mathbf{r}_i \text{ or } \mathbf{r}_i + \lambda\mathbf{r}_j,\\
\mathbf{r}'_{i+1} &= \mathbf{r}_{i+1}\\
&\vdots\\
\mathbf{r}'_k &= \mathbf{r}_k.
\end{aligned}$$

Any vector which is a linear combination of $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_k$ is a linear combination of $\mathbf{r}'_1$, $\mathbf{r}'_2$, …, $\mathbf{r}'_k$ and conversely. Since the row space is the set of all linear combination of row vectors, it follows that the row space of A and A' are same. □

THEOREM 2.22. *Let B be a matrix obtained from the matrix A by a sequence of elementary row operations. Then a subset of column vectors of A is linearly independent if and only if the corresponding column vectors of B are linearly independent. The same is true for basis also.*

PROOF. The result is consequence of two results: The null space is not changed by row operations and a homogeneous linear system has only trivial solutions if and only the column vectors are linearly independent. □

THEOREM 2.23. *Let R be the reduced row echelon form of a matrix A. Then the row vectors of R with leading 1's forms a basis for the row space of A and the column vectors of A corresponding to column vectors of R with leading 1's forms a basis for the column space of A.*

Corollary 2.1. *The dimension of row space and column space are equal.*

Definition 2.10. The common value of the dimensions of row space and column space of a matrix A is the **rank** of A. The dimension of the null space of A is the **nullity** of A.

It follows that the rank of A is the number of leading 1's in the reduced row echelon form of the matrix A.

Theorem 2.24. *The matrix A and its transpose A^T have the same rank.*

Proof. Since the row space of A is same as the column space of A^T, we have: rank of A = dimension of row space of A= the dimension of column space of A^T = rank of A^T. □

Example 2.40. *Consider the matrix*

$$A = \begin{pmatrix} 1 & 2 & 0 & 0 & 2 & -1 \\ 2 & 4 & 1 & 0 & 7 & -1 \\ 1 & 2 & 0 & 1 & 4 & 0 \\ 1 & 2 & 1 & 1 & 7 & 1 \end{pmatrix}.$$

The reduced row echelon form of the matrix is given by

$$R = \begin{pmatrix} 1 & 2 & 0 & 0 & 2 & -1 \\ 0 & 0 & 1 & 0 & 3 & 1 \\ 0 & 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The row vectors

$$\mathbf{r}_1 = \begin{pmatrix} 1 & 2 & 0 & 0 & 2 & -1 \end{pmatrix}$$
$$\mathbf{r}_2 = \begin{pmatrix} 0 & 0 & 1 & 0 & 3 & 1 \end{pmatrix}$$
$$\mathbf{r}_3 = \begin{pmatrix} 0 & 0 & 0 & 1 & 2 & 1 \end{pmatrix}$$

forms a basis for the row space of A. Note that first, third and fourth columns corresponds to leading 1's. Therefore the column vectors

$$\mathbf{c}_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 1 \end{pmatrix} \quad \mathbf{c}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} \quad \mathbf{c}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}$$

forms a basis for the column space of A.

Note that the basis for the column space is just the column vectors of A while for the row space they are not the rows of A. We can find the basis for the row space consisting of rows of A by reducing A^T to reduced row echelon form.

The reduced row echelon form of A^T is

$$\begin{pmatrix} 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The rows corresponding to the first, second and third columns constitutes the basis for the row space of A; that is, the vectors

$$\mathbf{r}_1' = \begin{pmatrix}1 & 2 & 0 & 0 & 2 & -1\end{pmatrix}$$
$$\mathbf{r}_2' = \begin{pmatrix}2 & 4 & 1 & 0 & 7 & -1\end{pmatrix}$$
$$\mathbf{r}_3' = \begin{pmatrix}1 & 2 & 0 & 1 & 4 & 0\end{pmatrix}$$

forms a basis for the row space of A. In this example, the rank of A is 3.

The solutions of the linear system $A\mathbf{x} = \mathbf{0}$ is obtained by solving the equations (obtained from the reduced row echelon form of A):

$$\begin{aligned} x_1 + 2x_2 + 2x_5 - x_6 &= 0 \\ x_3 + 3x_5 + x_6 &= 0 \\ x_4 + 2x_5 + x_6 &= 0. \end{aligned}$$

By taking the unknowns that does not corresponds to leading 1's as parameters:

$$x_2 = r, \quad x_5 = s, \quad x_6 = t,$$

we get

$$x_1 = -t - 2s - 2r, \quad x_3 = -t - 3s, \quad x_4 = -t - 2s.$$

Thus the solution vector is given by

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} = \begin{pmatrix} t - 2s - 2r \\ r \\ -t - 3s \\ -t - 2s \\ s \\ t \end{pmatrix} = r \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + s \begin{pmatrix} -2 \\ 0 \\ -3 \\ -2 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} 1 \\ 0 \\ -1 \\ -1 \\ 0 \\ 1 \end{pmatrix}.$$

Since the vectors

$$\begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} -2 \\ 0 \\ -3 \\ -2 \\ 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ 0 \\ -1 \\ -1 \\ 0 \\ 1 \end{pmatrix}$$

are linearly independent, these vectors forms a basis for the null space of A. Thus the null space has dimension 3 (which is same as the number of variables that does not corresponds to leading 1's.)

We state the following theorem without proof.

THEOREM 2.25. *If A is an $m \times n$ matrix, then the rank of A is the number of leading ones in the reduced row echelon form of A or equivalently the number of leading variables in the solution of $A\mathbf{x} = \mathbf{0}$ and the nullity of A is the number of variables that does not correspond to leading 1's or equivalently the number parameters in the general solution of $A\mathbf{x} = \mathbf{0}$.*

THEOREM 2.26 (Dimension theorem). *If A is any matrix of order $m \times n$, then*

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

PROOF. From the previous theorem, it is clear that the sum of the rank of A and nullity of A is the number of variables in the equation $A\mathbf{x} = \mathbf{0}$. Since this is n, the proof is completed. □

Theorem 2.27 (Consistency Theorem). *For any linear system $A\mathbf{x} = \mathbf{b}$, the following are equivalent:*

(1) *$A\mathbf{x} = \mathbf{b}$ is consistent.*
(2) *$\mathbf{b}$ is in the column space of A.*
(3) *The rank of A and the rank of the augmented matrix $[A|\mathbf{b}]$ are equal.*

Proof. We have earlier proved that $A\mathbf{x} = \mathbf{b}$ is consistent if and only $\mathbf{b}$ is in the column space of A. We will now prove that $\mathbf{b}$ is in the column space of A if and only if A and $[A|\mathbf{b}]$ have the same rank.

Let $\mathbf{b}$ be in the column space of A. We will show that the column space of A and $[A|\mathbf{b}]$ are same. Let r be the rank of A. Then there is a basis S for the column space of A consisting of r column vectors of A. Thus span(S) is the column space of A. Any vector in the column space of $[A|\mathbf{b}]$ is a linear combination of vectors in S and $\mathbf{b}$. Since $\mathbf{b}$ is in the column space of A, any vector in the column space of $[A|\mathbf{b}]$ is a linear combination of vectors in S and conversely. Thus

$$\text{span}(S) = \text{span}(S \cup \{\mathbf{b}\}).$$

Thus the column space of A and $[A|\mathbf{b}]$ are same and hence the rank of A and $[A|\mathbf{b}]$ are equal.

Conversely, let the rank of A and the rank of the augmented matrix $[A|\mathbf{b}]$ are equal. Let r be the rank of A. Then there are r column vectors of A that forms a basis for the column space of A. These vectors are also basis for the column space of $[A|\mathbf{b}]$ and therefore $\mathbf{b}$ belongs to the column space of A. ☐

Example 2.41. *Find λ & μ so that the system of linear equations*

$$\begin{aligned} x + y + z &= 6 \\ x + 2y + 3z &= 10 \\ x + 2y + \lambda z &= \mu \end{aligned}$$

has no solution, unique solution, or infinite number of solutions. The given system is

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 2 & \lambda \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 6 \\ 10 \\ \mu \end{pmatrix}.$$

Augmented matrix

$$\begin{aligned} [A|\mathbf{b}] &= \begin{pmatrix} 1 & 1 & 1 & 6 \\ 1 & 2 & 3 & 10 \\ 1 & 2 & \lambda & \mu \end{pmatrix} \\ &\sim \begin{pmatrix} 1 & 1 & 1 & 6 \\ 0 & 1 & 2 & 4 \\ 0 & 1 & \lambda - 1 & \mu - 6 \end{pmatrix} \begin{matrix} R_2 = R_2 - R_1 \\ R_3 = R_3 - R_1 \end{matrix} \\ &\sim \begin{pmatrix} 1 & 1 & 1 & 6 \\ 0 & 1 & 2 & 4 \\ 0 & 0 & \lambda - 3 & \mu - 10 \end{pmatrix} \; R_3 = R_3 - R_2 \end{aligned}$$

Note that rank of A is 2 if $\lambda = 3$ and 3 if $\lambda \neq 3$ while the rank of $[A|\mathbf{b}]$ is 2 if $\lambda = 3$ and $\mu = 10$ and 3 otherwise.

The system has no solution when rank of $[A|\mathbf{b}]$ *is not equal to rank of* A *i.e when* $\lambda = 3$ *and* $\mu \neq 10$.

The system has a solution if the rank of A *is equal to rank of* $[A|\mathbf{b}]$, *ie when* $\lambda = 3$ *and* $\mu = 10$ *or when* $\lambda \neq 3$. *The solution is unique if* $\lambda \neq 3$.

CHAPTER 3

Eigenvalues and Eigenvectors

1. Eigenvalues and Eigenvectors

Let $A = [a_{ij}]$ be a square matrix of order n. The matrix $A - \lambda I$ is called the characteristic matrix of A. The equation $|A - \lambda I| = 0$ is called the **characteristic equation** of A. The determinant $|A - \lambda I|$ is always a polynomial in λ and it is called the **characteristic polynomial of** A. The roots of the characteristic equation are called **characteristic roots** or **eigenvalues** of A. Note that λ is the eigenvalue of A if $A - \lambda I$ is singular. Let λ be an eigenvalue of A. Then a vector $\mathbf{x} \neq 0$ is called an **eigenvector** corresponding to λ if $A\mathbf{x} = \lambda\mathbf{x}$. If $\mathbf{x}$ is an eigenvector corresponding to λ of the matrix A, then $A\mathbf{x} = \lambda\mathbf{x}$ and since, for any $k \neq 0$, $A(k\mathbf{x}) = kA\mathbf{x} = \lambda k\mathbf{x}$ it follows that $k\mathbf{x}$ is also an eigenvector.

Example 3.1. *Find the eigenvalues and eigenvectors of the matrix*

$$\begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 3 \end{pmatrix}.$$

The characteristic equation, $|A - \lambda I| = 0$, *is*

$$\begin{vmatrix} 3-\lambda & 1 & 1 \\ 1 & 3-\lambda & -1 \\ 1 & -1 & 3-\lambda \end{vmatrix} = 0$$

which becomes

$$\lambda^3 - 9\lambda^2 + 24\lambda - 16 = 0.$$

The roots of the above equation (or the eigenvalues of the matrix) are 1, 4, 4.

The eigenvector corresponding to the eigenvalue λ *is given by the matrix equation*

$$\begin{pmatrix} 3-\lambda & 1 & 1 \\ 1 & 3-\lambda & -1 \\ 1 & -1 & 3-\lambda \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0.$$

Case (i) $(\lambda = 1)$ *In this case, we have to solve*

$$\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & -1 \\ 1 & -1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0$$

or

$$2x_1 + x_2 + x_3 = 0,\ x_1 + 2x_2 - x_3 = 0,\ x_1 - x_2 + 2x_3 = 0.$$

Subtracting the first two equation, we get the third equation. Hence using the first two equation, we get

$$\frac{x_1}{3} = \frac{x_2}{-3} = \frac{x_3}{-3} = \frac{k}{3}.$$

Therefore $(k, -k, -k)$, $k \neq 0$ *is an eigenvector corresponding to* $\lambda = 1$.

Case (ii) ($\lambda = 4$) In this case, we have to solve

$$\begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0$$

or

$$-x_1 + x_2 + x_3 = 0, \; x_1 - x_2 - x_3 = 0, \; x_1 - x_2 - x_3 = 0.$$

Note that all the three equations are same. Hence we use the first equation alone. By putting $x_3 = 0$, we have $x_1 = x_2 = k$, where k is a constant. Similarly by putting $x_2 = 0$, we have $x_1 = x_3 = k$. Hence for nonzero k, $(k, k, 0)$ and $(k, 0, k)$ are the eigenvectors corresponding to $\lambda = 4$. Note that for the eigenvalue $\lambda = 4$ we have two linearly independent eigenvectors.

Example 3.2. *Find the eigenvalues and eigenvectors of*

$$\begin{pmatrix} 2 & -2 & 2 \\ 1 & 1 & 1 \\ 1 & 3 & -1 \end{pmatrix}.$$

The characteristic equation of the given matrix is

$$\begin{vmatrix} 2-\lambda & -2 & 2 \\ 1 & 1-\lambda & 1 \\ 1 & 3 & -1-\lambda \end{vmatrix} = 0.$$

This reduces to $(2 - \lambda)(\lambda^2 - 4) = 0$ which gives $\lambda = -2, 2, 2$. The eigenvectors corresponding to λ is given by

$$\begin{pmatrix} 2-\lambda & -2 & 2 \\ 1 & 1-\lambda & 1 \\ 1 & 3 & -1-\lambda \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0.$$

Case (i) ($\lambda = -2$) The corresponding eigenvector is given by the equations

$$4x_1 - 2x_2 + 2x_3 = 0, \; x_1 + 3x_2 + x_3 = 0, \; x_1 + 3x_2 + x_3 = 0.$$

Note that the second and third equations are same. Hence using the first and second, we have

$$\frac{x_1}{-4} = \frac{x_2}{-1} = \frac{x_3}{7}$$

and hence the corresponding eigenvector is $(-4k, -k, 7k)$ where $k \neq 0$.

Case (ii) ($\lambda = 2$) The corresponding eigenvector is given by the equations

$$-2x_2 + 2x_3 = 0, \; x_1 - x_2 + x_3 = 0, \; x_1 + 3x_2 - 3x_3 = 0.$$

Note the difference of the second and third equations gives an equation proportional to the first equation. Hence using the second and third, we have

$$\frac{x_1}{0} = \frac{x_2}{4} = \frac{x_3}{4} = \frac{k}{4}$$

and hence the corresponding eigenvector is $(0, k, k)$ where $k \neq 0$.

Example 3.3. *Find the eigenvalues and eigenvectors of*

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

The characteristic equation of the given matrix is

$$|A-\lambda I| = \begin{vmatrix} 1-\lambda & 0 & 1 \\ 0 & 1-\lambda & 1 \\ 1 & 1 & -\lambda \end{vmatrix} = 0$$

or

$$\lambda^3 - 2\lambda^2 - \lambda + 2 = 0.$$

The eigenvalues are $\lambda = -1, 1, 2$. *The corresponding eigenvectors are*

$$\mathbf{x}_1 = \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}, \quad \mathbf{x}_2 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{x}_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

LEMMA 3.1. *Let A be a square matrix and $\mathbf{x} \neq 0$, λ be a number. The vector $\mathbf{x}$ is an eigenvector corresponding to λ if and only if $A\mathbf{x} = \lambda\mathbf{x}$.*

PROOF. Clearly if $\mathbf{x}$ is an eigenvector corresponding to λ, then $A\mathbf{x} = \lambda\mathbf{x}$. Conversely, if $A\mathbf{x} = \lambda\mathbf{x}$, then we have to show that λ is an eigenvalue of A or equivalently that $A-\lambda I$ is singular. If $A-\lambda I$ is nonsingular, $(A-\lambda I)\mathbf{x} = 0$ has unique solution $\mathbf{x} = 0$ which violates $\mathbf{x} \neq 0$. Therefore $A - \lambda I$ is singular or equivalently λ is an eigenvalue of A. □

THEOREM 3.1. *The eigenvectors corresponding to distinct eigenvalues are linearly independent.*

PROOF. Let $\mathbf{x}_1$, $\mathbf{x}_2$, …, $\mathbf{x}_n$ be the eigenvectors corresponding to the distinct eigenvalues λ_1, λ_2, …, λ_n of a square matrix A. Assume that the eigenvectors are linearly dependent. Then there are scalars α_1, α_2, …, α_n such that

$$\alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \cdots + \alpha_n\mathbf{x}_n = 0.$$

Among all such relations let

$$\alpha_{i_1}\mathbf{x}_{i_1} + \alpha_{i_2}\mathbf{x}_{i_2} + \cdots + \alpha_{i_r}\mathbf{x}_{i_r} = 0 \tag{4}$$

be the one which is shortest and $\alpha_{i_j} \neq 0$ for all j. Then multiplying (**??**) by A we have

$$\alpha_{i_1}A\mathbf{x}_{i_1} + \alpha_{i_2}A\mathbf{x}_{i_2} + \cdots + \alpha_{i_r}A\mathbf{x}_{i_r} = 0$$

or

$$\alpha_{i_1}\lambda_{i_1}\mathbf{x}_{i_1} + \alpha_{i_2}\lambda_{i_2}\mathbf{x}_{i_2} + \cdots + \alpha_{i_r}\lambda_{ir}\mathbf{x}_{i_r} = 0.$$

Multiplying (**??**) by λ_{i_1} and subtracting it from the above equation, we have

$$(\lambda_{i_2} - \lambda_{i_1})\alpha_{i_2}\mathbf{x}_{i_2} + (\lambda_{i_3} - \lambda_{i_1})\alpha_{i_3} + \cdots (\lambda_{i_r} - \lambda_{i_1})\alpha_{i_r} = 0$$

which is shorter than the shortest. Therefore the vectors can not be linearly dependent; they are linearly independent. □

THEOREM 3.2. *Let λ_1, λ_2, …, λ_n be eigenvalues of a square matrix A of order n. Then we have the following:*

(i) *λ_1^m, λ_2^m, …, λ_n^m are eigenvalues of A^m for any positive integer m. In general, if $p(x)$ is a polynomial, then $p(\lambda_i)$, $i = 1, 2, \ldots, n$ are the eigenvalues of $p(A)$.*
(ii) *A is nonsingular if and only if no eigenvalue is zero. Equivalently A is singular if and only if at least one eigenvalue of A is zero.*
(iii) *If A is nonsingular, $1/\lambda_1$, $1/\lambda_2$, …, $1/\lambda_n$ are the eigenvalues of A^{-1}. Also $|A|/\lambda_1$, $|A|/\lambda_2$, …, $|A|/\lambda_n$ are the eigenvalues of* $\operatorname{Adj} A$.

(iv) *A and A^T have same eigenvalues.*
(v) *If A is a triangular matrix, the diagonal entries are the eigenvalues of A.*
(vi) *The sum of the eigenvalues of A is* $\operatorname{tr} A$. *Recall that the trace of a matrix A, denoted by* $\operatorname{tr} A$, *is the sum of the diagonal elements of a matrix A. The sum of the eigenvalues of a skew-symmetric matrix is zero.*
(vii) *The product of the eigenvalues of A is the determinant.*
(viii) *Similar matrices have same eigenvalues.*

PROOF. In the proof of the Theorem, we use Lemma **??**. (i) Since $A\mathbf{x}_i = \lambda_i\mathbf{x}_i$, we have $A^2\mathbf{x}_i = A(\lambda_i\mathbf{x}_i) = \lambda_i^2\mathbf{x}_i$. Therefore λ_i^2 is an eigenvalue of A^2. Similarly λ_i^m is an eigenvalue of A^m. Let

$$p(x) = a_0x^n + a_1x^{n-1} + \cdots + a_n$$

be the given polynomial. Then

$$p(A) = a_0A^n + a_1A^{n-1} + \cdots + a_nI$$

and therefore

$$\begin{aligned} p(A)\mathbf{x}_i &= a_0A^n\mathbf{x}_i + \cdots + a_n\mathbf{x}_i \\ &= a_0\lambda_i^n\mathbf{x}_i + a_1\lambda_i^{n-1}\mathbf{x}_i + \cdots + a_n\mathbf{x}_i \\ &= (a_0\lambda_i^n + a_1\lambda_i^{n-1} + \cdots + a_n)\mathbf{x}_i \\ &= p(\lambda_i)\mathbf{x}_i. \end{aligned}$$

Therefore $p(\lambda_i)$ is an eigenvalue of $p(A)$.

(ii) A is singular $\iff |A| = 0 \iff |A - 0I| = 0 \iff 0$ is an eigenvalue of A.

(iii) Since A is nonsingular, all eigenvalues are nonzero. Also $A\mathbf{x} = \lambda\mathbf{x}$ implies $A^{-1}A\mathbf{x} = A^{-1}\lambda\mathbf{x}$ or $A^{-1}\mathbf{x} = \frac{1}{\lambda}\mathbf{x}$.

(iv) Since

$$|A^T - \lambda I| = |(A - \lambda I)^T| = |A - \lambda I|,$$

the result follows.

(v) Let $A = [a_{ij}]$. Since $A - \lambda I$ is triangular and the determinant of triangular matrix is the product of its diagonal entries,

$$|A - \lambda I| = (a_{11} - \lambda)(a_{22} - \lambda)\dots(a_{nn} - \lambda).$$

Therefore the characteristic equation is

$$(a_{11} - \lambda)(a_{22} - \lambda)\dots(a_{nn} - \lambda) = 0$$

and the eigenvalues are the diagonal entries $a_{11}, \dots, a_{nn}$.

(vi) and (vii) can be proved by expanding the determinant $|A - \lambda I|$.

(viii) Let A and B be similar matrices. Then there is a nonsingular matrix P such that $B = P^{-1}AP$. Since

$$\begin{aligned} |B - \lambda I| &= |P^{-1}AP - \lambda P^{-1}IP| \\ &= |P^{-1}(A - \lambda I)P| \\ &= |P^{-1}||P||A - \lambda I| \\ &= |A - \lambda I|, \end{aligned}$$

the result follows. Note that we have used the fact that

$$|A^{-1}||A| = |A^{-1}A| = |I| = 1.$$

□

THEOREM 3.3.

(i) *Eigenvalues of real orthogonal matrices are real or complex conjugates in pairs and have absolute value one. Also if λ is an eigenvalue of an orthogonal matrix, so is $1/\lambda$.*
(ii) *All the eigenvalues of a real symmetric matrix are real.*
(iii) *All the eigenvalues of a real skew-symmetric matrix are pure imaginary.*

PROOF. (i) Let λ be an eigenvalue of a real orthogonal matrix A and let $\mathbf{x} \neq 0$ be the corresponding eigenvector. Then $A\mathbf{x} = \lambda\mathbf{x}$, $\overline{A} = A$, $AA^T = A^TA = I$. Now

$$\begin{aligned} \overline{\mathbf{x}}^T\mathbf{x} &= \overline{\mathbf{x}}^T A^T A\mathbf{x} = \overline{\mathbf{x}}^T\overline{A}^T A\mathbf{x} = (\overline{A}\overline{\mathbf{x}})^T A\mathbf{x} \\ &= (\overline{\lambda}\overline{\mathbf{x}})^T \lambda\mathbf{x} = \overline{\lambda}\lambda\overline{\mathbf{x}}^T\mathbf{x} = |\lambda|^2\overline{\mathbf{x}}^T\mathbf{x}. \end{aligned}$$

Since $\overline{\mathbf{x}}^T\mathbf{x}$ is positive, we have $|\lambda|^2 = 1$ or $|\lambda| = 1$.

Also the characteristic equation of real matrix has real coefficients, the roots are real or complex conjugate in pairs. Let λ be an eigenvalue of the orthogonal matrix A. Then $1/\lambda$ is an eigenvalue of $A^{-1} = A^T$. But A and A^T have same eigenvalues. Therefore $1/\lambda$ is an eigenvalue of A.

(ii) Since A is real symmetric, we have $\overline{A} = A$ and $A^T = A$. Let λ be an eigenvalue of A and $\mathbf{x}$ be the corresponding eigenvector. Then

$$\overline{A}\overline{\mathbf{x}} = \overline{\lambda}\overline{\mathbf{x}}, \quad \overline{\mathbf{x}}^T\overline{A}^T = \overline{\lambda}\overline{\mathbf{x}}^T$$

or

$$\overline{\mathbf{x}}^T A = \overline{\lambda}\overline{\mathbf{x}}^T.$$

Post multiplying by $\mathbf{x}$,

$$\overline{\mathbf{x}}^T A\mathbf{x} = \overline{\lambda}\overline{\mathbf{x}}^T\mathbf{x}$$

or

$$\lambda\overline{\mathbf{x}}^T\mathbf{x} = \overline{\lambda}\overline{\mathbf{x}}^T\mathbf{x}.$$

Since $\overline{\mathbf{x}}^T\mathbf{x}$ is positive, we have $\lambda = \overline{\lambda}$. Therefore λ is real.

Proof of (iii) is similar to proof of (ii). □

THEOREM 3.4. *An eigenvector cannot correspond to two different eigenvalues.*

PROOF. If $\mathbf{x}$ is an eigenvector of two different eigenvalues λ and μ of a matrix A, then $A\mathbf{x} = \lambda\mathbf{x}$ and $A\mathbf{x} = \mu\mathbf{x}$ and therefore $\lambda\mathbf{x} = \mu\mathbf{x}$ or $(\lambda - \mu)\mathbf{x} = 0$. Since this shows that $\mathbf{x} = 0$ which is not possible. Therefore a vector cannot corresponds to two different eigenvalues. □

EXAMPLE 3.4. *Find the eigenvalues of A^2 where*

$$A = \begin{pmatrix} 2 & -2 & 2 \\ 1 & 1 & 1 \\ 1 & 3 & -1 \end{pmatrix}.$$

The eigenvalues of A^2 are the squares of the eigenvalues of A. Since the eigenvalues of A are -2, 2, 2, the eigenvalues of the matrix A^2 are 4, 4, 4.

EXAMPLE 3.5. *Find the sum and product of the eigenvalues of the matrix*

$$\begin{pmatrix} 2 & 3 & -2 \\ -2 & 1 & 1 \\ 1 & 0 & 2 \end{pmatrix}.$$

The sum of eigenvalues is the trace of the matrix, that is, 2+1+2=5. The product of the eigenvalues is the determinant of the matrix

$$\begin{vmatrix} 2 & 3 & -2 \\ -2 & 1 & 1 \\ 1 & 0 & 2 \end{vmatrix} = 21.$$

2. Cayley-Hamilton Theorem

THEOREM 3.5 (Cayley-Hamilton Theorem). *Every square matrix satisfies its characteristic equation.*

PROOF. Note that the characteristic equation of a matrix A of order n is $|A-\lambda I| = 0$ is a polynomial equation of degree n. Let this polynomial equation be

$$p(x) = a_0\lambda^n + a_1\lambda^{n-1} + \cdots + a_n = 0.$$

We have to prove that

$$p(A) = a_0A^n + a_1A^{n-1} + \cdots + a_nI = \mathbf{0}.$$

Let B be the adjoint of $A - \lambda I$. The matrix B can be written as polynomial in λ of degree $n-1$ where each of the coefficient is a matrix of order n:

$$B = B_0\lambda^{n-1} + B_1\lambda^{n-2} + \cdots + B_{n-1}.$$

Since for any matrix A, $(\text{Adj } A)A = A(\text{Adj } A) = |A|I$, we have

$$(A - \lambda I)\,\text{Adj}(A - \lambda I) = |A - \lambda I|I.$$

Therefore

$$(A - \lambda I)(B_0\lambda^{n-1} + B_1\lambda^{n-2} + \cdots + B_{n-1})$$
$$= (a_0\lambda^n + a_1\lambda^{n-1} + \cdots + a_n)I.$$

Comparing the coefficient of powers of λ, we have

$$\begin{aligned} -B_0 &= a_0I \\ AB_0 - B_1 &= a_1I \\ AB_1 - B_2 &= a_2I \\ \cdots & \cdots \\ AB_{n-1} &= a_nI. \end{aligned}$$

Pre-multiplying the above equations by A^n, A^{n-1}, ..., I respectively and adding them we have

$$a_0A^n + a_1A^{n-1} + \cdots + a_nI = \mathbf{0}.$$

This proves the result. □

If A is nonsingular, then the constant term a_n is nonzero. Therefore, we have

$$I = \frac{-1}{a_n}(a_0A^n + a_1A^{n-1} + \cdots + a_{n-1}A)$$

and multiplying by A^{-1} we have

$$A^{-1} = \frac{-1}{a_n}(a_0A^{n-1} + a_1A^{n-2} + \cdots + a_{n-1}I).$$

Using this equation we can compute A^{-1} by using the powers of A.

EXAMPLE 3.6. *Verify the Cayley-Hamilton Theorem for the matrix* $A = \begin{pmatrix} 1 & 0 & -2 \\ 2 & 2 & 4 \\ 0 & 0 & 2 \end{pmatrix}$. *Find the inverse of A, if it exists.*

The characteristic equation of the above matrix is given by

$$\begin{vmatrix} 1-\lambda & 0 & -2 \\ 2 & 2-\lambda & 4 \\ 0 & 0 & 2-\lambda \end{vmatrix} = 0$$

and this becomes

$$\lambda^3 - 5\lambda^2 + 8\lambda - 4 = 0.$$

Since $A^2 = \begin{pmatrix} 1 & 0 & -6 \\ 6 & 4 & 12 \\ 0 & 0 & 4 \end{pmatrix}$ *and* $A^3 = \begin{pmatrix} 1 & 0 & -14 \\ 14 & 8 & 28 \\ 0 & 0 & 8 \end{pmatrix}$, *it is seen that*

$$A^3 - 5A^2 + 8A - 4I = \mathbf{0}.$$

Since $4I = A^3 - 5A^2 + 8A$, *we get after a multiplication by* $A^{-1}/4$,

$$\begin{aligned} A^{-1} &= \frac{1}{4}(A^2 - 5A + 8I) \\ &= \begin{pmatrix} 1 & 0 & 1 \\ -1 & 1/2 & -2 \\ 0 & 0 & 1/2 \end{pmatrix}. \end{aligned}$$

EXAMPLE 3.7. *Verify Cayley-Hamilton Theorem for the matrix J of order n where*

$$J = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \cdots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}.$$

The characteristic equation is

$$\begin{vmatrix} 1-\lambda & 1 & \cdots & 1 \\ 1 & 1-\lambda & \cdots & 1 \\ \vdots & \vdots & \cdots & \vdots \\ 1 & 1 & \cdots & 1-\lambda \end{vmatrix} = 0.$$

Writing the first column of this determinant as the sum of all the columns and taking the common factors in the first column we have

$$(n-\lambda)\begin{vmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1-\lambda & \cdots & 1 \\ \vdots & \vdots & \cdots & \vdots \\ 1 & 1 & \cdots & 1-\lambda \end{vmatrix} = 0.$$

By subtracting the first row from the remaining rows, we get

$$(n-\lambda)\begin{vmatrix} 1 & 1 & \cdots & 1 \\ 0 & -\lambda & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & -\lambda \end{vmatrix} = 0$$

which becomes

$$(n-\lambda)(-\lambda)^{n-1} = 0$$

or

$$\lambda^n - n\lambda^{n-1} = 0.$$

Note $J^2 = nJ$ *and hence* $J^k = n^{k-1}J = nJ^{k-1}$. *In particular we have* $J^n - nJ^{n-1} = 0$. *Thus* J *satisfies its characteristic equation.*

3. Diagonalization

A matrix is **diagonalizable** if it is similar to a diagonal matrix. That is, the matrix A is diagonalizable if there exists a nonsingular matrix M such that $M^{-1}AM$ is a diagonal matrix.

THEOREM 3.6. *A square matrix of order* n *is diagonalizable if and only if it has* n *linearly independent eigenvectors.*

PROOF. If A is diagonalizable, then there is a nonsingular matrix M such that $M^{-1}AM = D$ where $D = \operatorname{diag}(d_1, d_2, d_3, \ldots, d_n)$ is the diagonal matrix whose diagonal entries are d_1, d_2, …, d_n respectively. Let $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$, …, $\mathbf{x}_n$ be the columns of the matrix M. Since M is nonsingular, $\mathbf{x}_i$ are nonzero and also linearly independent. Since $AM = MD$, we have

$$\begin{aligned} [A\mathbf{x}_1\ A\mathbf{x}_2\ \ldots\ A\mathbf{x}_n] &= A[\mathbf{x}_1\ \mathbf{x}_2\ \ldots\ \mathbf{x}_n] \\ &= [\mathbf{x}_1\ \mathbf{x}_2\ \ldots\ \mathbf{x}_n][d_1\ \ldots\ d_n] \\ &= [d_1\mathbf{x}_1\ d_2\mathbf{x}_2\ \ldots\ d_n\mathbf{x}_n] \end{aligned}$$

and therefore $A\mathbf{x}_i = d_i\mathbf{x}_i$ for $i = 1,\ 2,\ \ldots,\ n$. Therefore $\mathbf{x}_i$ is an eigenvector corresponding to d_i by Lemma **??**.

To prove the converse, let us assume that A has n linearly independent eigenvectors $\mathbf{x}_i$, $i = 1,\ 2,\ \ldots,\ n$. Let λ_i be the eigenvalue corresponding to $\mathbf{x}_i$. Let $M = [\mathbf{x}_1\ \mathbf{x}_2\ \ldots\ \mathbf{x}_n]$ be the matrix formed by the eigenvectors. Since $A\mathbf{x}_i = \lambda_i\mathbf{x}_i$, we have

$$\begin{aligned} AM &= A[\mathbf{x}_1\ \mathbf{x}_2\ \ldots\ \mathbf{x}_n] \\ &= [A\mathbf{x}_1\ A\mathbf{x}_2\ \ldots\ A\mathbf{x}_n] \\ &= [\lambda_1\mathbf{x}_1\ \lambda_2\mathbf{x}_2\ \ldots\ \lambda_n\mathbf{x}_n] \\ &= M\operatorname{diag}(\lambda_1\ \lambda_2\ \ldots\ \lambda_n) \end{aligned}$$

we have $M^{-1}AM = \operatorname{diag}(\lambda_1\ \lambda_2\ \ldots\ \lambda_n)$. This prove that A is diagonalizable. □

A matrix is diagonalizable if there are n linearly independent eigenvectors and is invertible if there are n nonzero eigenvalues. Therefore diagonalizability is concerned with the eigenvectors and invertibility is concerned with eigenvalues. Since eigenvectors corresponding to distinct eigenvalues are linearly independent, diagonalization can (possibly) fail only when the matrix has repeated eigenvalues. (Note that case of identity matrix!

Here all the eigenvalues are equal but however there are n linearly independent eigenvectors.) If an eigenvalue is repeated for p times, what we need for diagonalizability is p linearly independent eigenvectors.

If A is diagonalizable, then $M^{-1}AM = D$ where D is diagonal and therefore

$$A = MDM^{-1}$$

and hence

$$\begin{aligned} A^2 &= MDM^{-1}MDM^{-1} \\ &= MDIDM^{-1} \\ &= MD^2M^{-1}. \end{aligned}$$

In general we have

$$A^m = MD^mM^{-1}.$$

Since $D = \operatorname{diag}(d_1 \ \dots \ d_n)$, we have

$$D^m = \operatorname{diag}(d_1^m \ \dots \ d_n^m),$$

the above formula for A^m is useful in computing higher powers of A.

Also if A is diagonalizable, then A can found from the corresponding eigenvalues and eigenvectors by using the formula $A = MDM^{-1}$ where D is the diagonal matrix whose diagonal entries are eigenvalues and M is the matrix whose columns are the corresponding eigenvectors.

Example 3.8. *Diagonalize the matrix*

$$A = \begin{pmatrix} -2 & 0 & 6 \\ -1 & 1 & 2 \\ -2 & 0 & 5 \end{pmatrix}.$$

The characteristic equation of A is given by

$$\begin{vmatrix} -2-\lambda & 0 & 6 \\ -1 & 1-\lambda & 2 \\ -2 & 0 & 5-\lambda \end{vmatrix} = 0,$$

or

$$(\lambda - 1)(\lambda^2 - 3\lambda + 2) = (\lambda - 1)^2(\lambda - 2) = 0.$$

Therefore the eigenvalues are $\lambda = 1, 1, 2$.

When $\lambda = 1$, the corresponding eigenvectors are the solution of

$$\begin{pmatrix} -3 & 0 & 6 \\ -1 & 0 & 2 \\ -2 & 0 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0.$$

The above system reduces to the following single equation:

$$-x_1 + 2x_3 = 0.$$

Therefore $x_1 = 2k, x_2 = l, x_3 = k$ where k and l are two constants. Since rank of $A - I$ is one, the above system of equation has two linearly independent solution. In fact, $\mathbf{x}_1 = [2, 0, 1]^T \quad \mathbf{x}_2 = [0, 1, 0]^T$ are two linearly independent eigenvectors.

When $\lambda = 2$, the corresponding eigenvector is the solution of

$$\begin{pmatrix} -4 & 0 & 6 \\ -1 & -1 & 2 \\ -2 & 0 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0.$$

The solution is given by $\mathbf{x}_3 = [3, 1, 2]^T$. Consider the matrix

$$M = [\mathbf{x}_1 \ \mathbf{x}_2 \ \mathbf{x}_3] = \begin{pmatrix} 2 & 0 & 3 \\ 0 & 1 & 1 \\ 1 & 0 & 2 \end{pmatrix}.$$

The inverse of M is given by

$$M^{-1} = \begin{pmatrix} 2 & 0 & -3 \\ 1 & 1 & -2 \\ -1 & 0 & 2 \end{pmatrix}.$$

Now

$$\begin{aligned} M^{-1}AM &= \begin{pmatrix} 2 & 0 & -3 \\ 1 & 1 & -2 \\ -1 & 0 & 2 \end{pmatrix} \begin{pmatrix} -2 & 0 & 6 \\ -1 & 1 & 2 \\ -2 & 0 & 5 \end{pmatrix} \begin{pmatrix} 2 & 0 & 3 \\ 0 & 1 & 1 \\ 1 & 0 & 2 \end{pmatrix} \\ &= \begin{pmatrix} 2 & 0 & -3 \\ 1 & 1 & -2 \\ -2 & 0 & 4 \end{pmatrix} \begin{pmatrix} 2 & 0 & -3 \\ 1 & 1 & -2 \\ -1 & 0 & 2 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}. \end{aligned}$$

EXAMPLE 3.9. *Find the eigenvalues and eigenvectors of*

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix}.$$

The characteristic equation of the given matrix is

$$\lambda^4 + \lambda^2 - 4\lambda - 4 = 0.$$

The eigenvalues are $\lambda = -2, -1, 2$. The corresponding eigenvectors are

$$\mathbf{x}_1 = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}, \quad \mathbf{x}_2 = \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}, \quad \mathbf{x}_3 = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}.$$

The matrix of eigenvectors is

$$M = \begin{pmatrix} 0 & 1 & 2 \\ -1 & -1 & 1 \\ 1 & -1 & 1 \end{pmatrix}$$

and its inverse is given by

$$M^{-1} = \frac{1}{6} \begin{pmatrix} 0 & -3 & 3 \\ 2 & -2 & -2 \\ 2 & 1 & 1 \end{pmatrix}.$$

A computation shows that

$$M^{-1}AM = \begin{pmatrix} -2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

Notice that the matrix A is symmetric and the eigenvectors $\mathbf{x}_1,\ \mathbf{x}_2,\ \mathbf{x}_3$ are orthogonal.

EXAMPLE 3.10. *Find a matrix A whose eigenvalues are 1, 2 and 3 and the corresponding eigenvectors are*

$$\begin{pmatrix}1\\0\\1\end{pmatrix}, \begin{pmatrix}1\\1\\-1\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}.$$

Consider the matrix M of eigenvectors and the diagonal matrix D of eigenvalues

$$M = \begin{pmatrix}1 & 1 & 0\\0 & 1 & 1\\1 & -1 & 0\end{pmatrix}, \quad D = \begin{pmatrix}1 & 0 & 0\\0 & 2 & 0\\0 & 0 & 3\end{pmatrix}.$$

The inverse of M is given by

$$M^{-1} = \begin{pmatrix}1/2 & 0 & 1/2\\1/2 & 0 & -1/2\\-1/2 & 1 & 1/2\end{pmatrix}$$

and therefore

$$A = MDM^{-1} = \begin{pmatrix}5/2 & 0 & 1/2\\1/2 & 1 & -1/2\\1/2 & 0 & 5/2\end{pmatrix}$$

is the required matrix with given eigenvalues and eigenvectors.

A matrix P is **orthogonal** if $PP^T = P^TP = I$. Thus for orthogonal matrices P, we have

$$P^T = P^{-1}.$$

A matrix A is orthogonally diagonalizable if there is an orthogonal matrix P such that the matrix P^TAP is diagonal. The matrix P is said to **orthogonally diagonalize the matrix A.**

A set of vectors $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ is **orthonormal** if

$$\|\mathbf{x}_i\| = 1 \quad \text{for all } i$$

and

$$\mathbf{x}_i.\mathbf{x}_j = 0 \quad \text{for all } i \neq j.$$

In other words, the vectors are orthogonal and of unit norm.

If A is orthogonally diagonalizable, then there is an orthogonal matrix P such that $P^TAP = D$, where D is a diagonal matrix. Thus we have

$$\begin{aligned}PDP^T &= P(P^TAP)P^T\\ &= (PP^T)A(PP^T)\\ &= IAI\\ &= A.\end{aligned}$$

Since the transpose of a diagonal matrix is itself, we have

$$\begin{aligned}A^T &= (PDP^T)^T\\ &= (P^T)^TD^TP^T\\ &= PDP^T\\ &= A\end{aligned}$$

and therefore A is symmetric.

THEOREM 3.7. *Let A be a square matrix of order n. Then the following are equivalent.*

(1) *A is orthogonally diagonalizable.*
(2) *A has an orthonormal set of n vectors.*
(3) *A is symmetric.*

PROOF. Proof is omitted. □

EXAMPLE 3.11. *Orthogonally diagonalize the matrix*

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix}.$$

The characteristic equation of the given matrix is

$$\lambda^4 + \lambda^2 - 4\lambda - 4 = 0.$$

The eigenvalues are $\lambda = -2, -1, 2$. The corresponding eigenvectors are

$$\mathbf{x}_1 = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}, \quad \mathbf{x}_2 = \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}, \quad \mathbf{x}_3 = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}.$$

The corresponding orthonormal eigenvectors are given by

$$\mathbf{v}_1 = \begin{pmatrix} 0 \\ -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \end{pmatrix}, \quad \mathbf{v}_3 = \begin{pmatrix} \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{pmatrix}.$$

The matrix of eigenvectors is

$$P = \begin{pmatrix} 0 & \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \end{pmatrix}.$$

Now it easy to see that

$$P^T P = P P^T = I$$

and

$$P^T A P = \begin{pmatrix} -2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$