CHAPTER 5

Vector Space $\mathbb{R}^n$

5.1 Subspaces and Spanning

In this chapter we investigate $\mathbb{R}^n$ in full generality, and introduce some of the most important concepts and methods in linear algebra.

Subspaces of $\mathbb{R}^n$

Definition 5.1.1. Subspace of $\mathbb{R}^n$: A set U of vectors in $\mathbb{R}^n$ is called a **subspace** of $\mathbb{R}^n$ if it satisfies the following properties:

S1. The zero vector $\mathbf{0} \in U$.

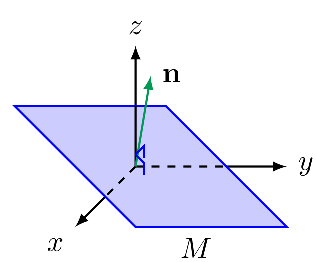
S2. If $\mathbf{x} \in U$ and $\mathbf{y} \in U$, then $\mathbf{x} + \mathbf{y} \in U$.

S3. If $\mathbf{x} \in U$, then $a\mathbf{x} \in U$ for every real number a.

We say that the subset U is **closed under addition** if S2 holds, and that U is **closed under scalar multiplication** if S3 holds.

Clearly $\mathbb{R}^n$ is a subspace of itself, and this chapter is about these subspaces and their properties. The set $U = \{\mathbf{0}\}$, consisting of only the zero vector, is also a subspace because $\mathbf{0} + \mathbf{0} = \mathbf{0}$ and $a\mathbf{0} = \mathbf{0}$ for each a in $\mathbb{R}$; it is called the **zero subspace**. Any subspace of $\mathbb{R}^n$ other than $\{\mathbf{0}\}$ or $\mathbb{R}^n$ is called a **proper** subspace.

Every plane M through the origin in $\mathbb{R}^3$ has equation $ax + by + cz = 0$ where a, b, and c are not all zero. Here $\mathbf{n} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ is a normal for the plane and $M = \{\mathbf{v} \text{ in } \mathbb{R}^3 \mid \mathbf{n} \cdot \mathbf{v} = 0\}$ where $\mathbf{v} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ and $\mathbf{n} \cdot \mathbf{v}$ denotes the dot product (see the diagram below). Then M is a subspace of $\mathbb{R}^3$.



Example 5.1.2. Planes and lines through the origin in $\mathbb{R}^3$ are all subspaces of $\mathbb{R}^3$.

Subspaces can also be used to describe important features of an $m \times n$ matrix A. The **null space** of A, denoted null A, and the **image space** of A, denoted im A, are defined by

$$\text{null } A = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}\} \quad \text{and} \quad \text{im } A = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\}$$

Example 5.1.3. If A is an $m \times n$ matrix, then:

(a) null A is a subspace of $\mathbb{R}^n$.

(b) im A is a subspace of $\mathbb{R}^m$.

Example 5.1.4. Show $E_\lambda(A) = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \lambda\mathbf{x}\} = \text{null}\,(\lambda I - A)$ is a subspace of $\mathbb{R}^n$ for each $n \times n$ matrix A and number λ.

$E_\lambda(A)$ is called the **eigenspace** of A corresponding to λ. Recall, λ is an **eigenvalue** of A if $E_\lambda(A) \neq \{\mathbf{0}\}$. In this case the nonzero vectors in $E_\lambda(A)$ are called the **eigenvectors** of A corresponding to λ.

Spanning Sets

Let $\mathbf{v}$ and $\mathbf{w}$ be two nonzero, nonparallel vectors in $\mathbb{R}^3$ with their tails at the origin. The plane M through the origin containing these vectors is described by saying that $\mathbf{n} = \mathbf{v} \times \mathbf{w}$ is a *normal* for M, and that M consists of all vectors $\mathbf{p}$ such that $\mathbf{n} \cdot \mathbf{p} = 0$. While this is a very useful way to look at planes, there is another approach that is at least as useful in $\mathbb{R}^3$ and, more importantly, works for all subspaces of $\mathbb{R}^n$ for any $n \geq 1$.

Definition 5.1.5. Linear Combinations and Span in $\mathbb{R}^n$: The set of all such linear combinations is called the **span** of the $\mathbf{x}_i$ and is denoted

$$\operatorname{span}\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\} = \{t_1\mathbf{x}_1 + t_2\mathbf{x}_2 + \cdots + t_k\mathbf{x}_k \mid t_i \text{ in } \mathbb{R}\}$$

If $V = \operatorname{span}\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$, we say that V is **spanned** by the vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$, and that the vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ **span** the space V.

Example 5.1.6. Let $\mathbf{x} = (2, -1, 2, 1)$ and $\mathbf{y} = (3, 4, -1, 1)$ in $\mathbb{R}^4$. Determine whether $\mathbf{p} = (0, -11, 8, 1)$ or $\mathbf{q} = (2, 3, 1, 2)$ are in $U = \operatorname{span}\{\mathbf{x}, \mathbf{y}\}$.

Theorem 5.1.7. *Span Theorem*

Let $U = span\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ in $\mathbb{R}^n$. Then:

(a) *U is a subspace of $\mathbb{R}^n$ containing each $\mathbf{x}_i$.*

(b) *If W is a subspace of $\mathbb{R}^n$ and each $\mathbf{x}_i \in W$, then $U \subseteq W$.*

Example 5.1.8. If $\mathbf{x}$ and $\mathbf{y}$ are in $\mathbb{R}^n$, show that $\text{span}\{\mathbf{x}, \mathbf{y}\} = \text{span}\{\mathbf{x}+\mathbf{y}, \mathbf{x}-\mathbf{y}\}$.

Column j of the $n \times n$ identity matrix I_n is denoted $\mathbf{e}_j$ and called the jth **coordinate vector** in $\mathbb{R}^n$, and the set $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is called the **standard basis** of $\mathbb{R}^n$.

Example 5.1.9. $\mathbb{R}^n = \text{span}\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ where $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ are the columns of I_n.

Example 5.1.10. Given an $m \times n$ matrix A, let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ denote the basic solutions to the system $A\mathbf{x} = \mathbf{0}$ given by the gaussian algorithm. Then

$$\text{null } A = \text{span}\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$$

Example 5.1.11. Let $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n$ denote the columns of the $m \times n$ matrix A. Then

$$\text{im } A = \text{span}\{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n\}$$

5.2 Independence and Dimension

Linear Independence

Definition 5.2.1. Linear Independence in $\mathbb{R}^n$: We call a set $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ of vectors **linearly independent** (or simply **independent**) if it satisfies the following condition:

$$\text{If } t_1\mathbf{x}_1 + t_2\mathbf{x}_2 + \cdots + t_k\mathbf{x}_k = \mathbf{0} \text{ then } t_1 = t_2 = \cdots = t_k = 0$$

Theorem 5.2.2. *If $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ is an independent set of vectors in $\mathbb{R}^n$, then every vector in span$\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ has a* ***unique*** *representation as a linear combination of the $\mathbf{x}_i$.*

It is useful to state the definition of independence in different language. Let us say that a linear combination **vanishes** if it equals the zero vector, and call a linear combination **trivial** if every coefficient is zero. Then the definition of independence can be compactly stated as follows:

> A set of vectors is independent if and only if the only linear combination that vanishes is the trivial one.

Hence we have a procedure for checking that a set of vectors is independent:

Theorem 5.2.3. *Independence Test*

To verify that a set $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ of vectors in $\mathbb{R}^n$ is independent, proceed as follows:

(a) *Set a linear combination equal to zero: $t_1\mathbf{x}_1 + t_2\mathbf{x}_2 + \cdots + t_k\mathbf{x}_k = \mathbf{0}$.*

(b) *Show that $t_i = 0$ for each i (that is, the linear combination is trivial).*

If some nontrivial linear combination vanishes, the vectors are not independent.

Example 5.2.4. Determine whether $\{(1, 0, -2, 5), (2, 1, 0, -1), (1, 1, 2, 1)\}$ is independent in $\mathbb{R}^4$.

Example 5.2.5. Show that the standard basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ of $\mathbb{R}^n$ is independent.

Example 5.2.6. If $\{\mathbf{x}, \mathbf{y}\}$ is independent, show that $\{2\mathbf{x} + 3\mathbf{y}, \mathbf{x} - 5\mathbf{y}\}$ is also independent.

Example 5.2.7. Show that the zero vector in $\mathbb{R}^n$ does not belong to any independent set.

Example 5.2.8. 014081 Given $\mathbf{x}$ in $\mathbb{R}^n$, show that $\{\mathbf{x}\}$ is independent if and only if $\mathbf{x} \neq \mathbf{0}$.

Example 5.2.9. Show that the nonzero rows of a row-echelon matrix R are independent.

A set of vectors in $\mathbb{R}^n$ is called **linearly dependent** (or simply **dependent**) if it is *not* linearly independent, equivalently if some nontrivial linear combination vanishes.

Example 5.2.10. If $\mathbf{v}$ and $\mathbf{w}$ are nonzero vectors in $\mathbb{R}^3$, show that $\{\mathbf{v}, \mathbf{w}\}$ is dependent if and only if $\mathbf{v}$ and $\mathbf{w}$ are parallel.

With this we can give a geometric description of what it means for a set $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ in $\mathbb{R}^3$ to be independent.

Example 5.2.11. Let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ be nonzero vectors in $\mathbb{R}^3$ where $\{\mathbf{v}, \mathbf{w}\}$ independent. Show that $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is independent if and only if $\mathbf{u}$ is not in the plane $M = \text{span}\{\mathbf{v}, \mathbf{w}\}$.

By the inverse theorem, the following conditions are equivalent for an $n \times n$ matrix A:r

(a) *A is invertible.*

(b) *If $A\mathbf{x} = \mathbf{0}$ where $\mathbf{x}$ is in $\mathbb{R}^n$, then $\mathbf{x} = \mathbf{0}$.*

(c) *$A\mathbf{x} = \mathbf{b}$ has a solution $\mathbf{x}$ for every vector $\mathbf{b}$ in $\mathbb{R}^n$.*

Theorem 5.2.12. *If A is an $m \times n$ matrix, let $\{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n\}$ denote the columns of A.*

(a) *$\{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n\}$ is independent in $\mathbb{R}^m$ if and only if $A\mathbf{x} = \mathbf{0}$, $\mathbf{x}$ in $\mathbb{R}^n$, implies $\mathbf{x} = \mathbf{0}$.*

(b) *$\mathbb{R}^m = span\{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n\}$ if and only if $A\mathbf{x} = \mathbf{b}$ has a solution $\mathbf{x}$ for every vector $\mathbf{b}$ in $\mathbb{R}^m$.*

Theorem 5.2.13. *The following are equivalent for an $n \times n$ matrix A:*

(a) *A is invertible.*

(b) *The columns of A are linearly independent.*

(c) *The columns of A span $\mathbb{R}^n$.*

(d) *The rows of A are linearly independent.*

(e) *The rows of A span the set of all $1 \times n$ rows.*

Example 5.2.14. Show that $S = \{(2, -2, 5), (-3, 1, 1), (2, 7, -4)\}$ is independent in $\mathbb{R}^3$.

Dimension

Theorem 5.2.15. *Fundamental Theorem*

Let U be a subspace of $\mathbb{R}^n$. If U is spanned by m vectors, and if U contains k linearly independent vectors, then $k \leq m$.

Definition 5.2.16. Basis of $\mathbb{R}^n$: If U is a subspace of $\mathbb{R}^n$, a set $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$ of vectors in U is called a **basis** of U if it satisfies the following two conditions:

(a) $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$ is linearly independent.

(b) $U = \text{span}\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$.

Theorem 5.2.17. *Invariance Theorem*

If $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$ and $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_k\}$ are bases of a subspace U of $\mathbb{R}^n$, then $m = k$.

Definition 5.2.18. Dimension of a Subspace of $\mathbb{R}^n$: If U is a subspace of $\mathbb{R}^n$ and $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$ is any basis of U, the number, m, of vectors in the basis is called the **dimension** of U, denoted

$$\dim U = m$$

Example 5.2.19. $\dim(\mathbb{R}^n) = n$ and $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ is a basis.

Returning to subspaces of $\mathbb{R}^n$, we define

$$\dim\{\mathbf{0}\} = 0$$

This amounts to saying $\{\mathbf{0}\}$ has a basis containing *no* vectors. This makes sense because $\mathbf{0}$ cannot belong to *any* independent set.

Example 5.2.20. Let $U = \left\{ \begin{bmatrix} r \\ s \\ r \end{bmatrix} \middle| r, s \text{ in } \mathbb{R} \right\}$. Show that U is a subspace of $\mathbb{R}^3$, find a basis, and calculate $\dim U$.

Example 5.2.21. Let $B = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ be a basis of $\mathbb{R}^n$. If A is an invertible $n \times n$ matrix, then $D = \{A\mathbf{x}_1, A\mathbf{x}_2, \ldots, A\mathbf{x}_n\}$ is also a basis of $\mathbb{R}^n$.

Theorem 5.2.22. *Let $U \neq \{\mathbf{0}\}$ be a subspace of $\mathbb{R}^n$. Then:*

(a) *U has a basis and dim$U \leq n$.*

(b) *Any independent set in U can be enlarged (by adding vectors from any fixed basis of U) to a basis of U.*

(c) *Any spanning set for U can be cut down (by deleting vectors) to a basis of U.*

Example 5.2.23. Find a basis of $\mathbb{R}^4$ containing $S = \{\mathbf{u}, \mathbf{v}\}$ where $\mathbf{u} = (0, 1, 2, 3)$ and $\mathbf{v} = (2, -1, 0, 1)$.

Theorem 5.2.24. *Let U be a subspace of $\mathbb{R}^n$ where $\dim U = m$ and let $B = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$ be a set of m vectors in U. Then B is independent if and only if B spans U.*

Theorem 5.2.25. *Let $U \subseteq W$ be subspaces of $\mathbb{R}^n$. Then:*

(a) *$\dim U \le \dim W$.*

(b) *If $\dim U = \dim W$, then $U = W$.*

It follows from Theorem 5.2.25 that if U is a subspace of $\mathbb{R}^n$, then $\dim U$ is one of the integers $0, 1, 2, \dots, n$, and that:

$$\begin{array}{lll} \dim U = 0 & \text{if and only if} & U = \{\mathbf{0}\}, \\ \dim U = n & \text{if and only if} & U = \mathbb{R}^n \end{array}$$

The other subspaces of $\mathbb{R}^n$ are called **proper**.

Example 5.2.26. (a) If U is a subspace of $\mathbb{R}^2$ or $\mathbb{R}^3$, then $\dim U = 1$ if and only if U is a line through the origin.

(b) If U is a subspace of $\mathbb{R}^3$, then $\dim U = 2$ if and only if U is a plane through the origin.

Note that this proof shows that if $\mathbf{v}$ and $\mathbf{w}$ are nonzero, nonparallel vectors in $\mathbb{R}^3$, then $\text{span}\{\mathbf{v}, \mathbf{w}\}$ is the plane with normal $\mathbf{n} = \mathbf{v} \times \mathbf{w}$.

5.3 Orthogonality

Dot Product, Length, and Distance

If $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ are two n-tuples in $\mathbb{R}^n$, recall that their **dot product** is:

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$

Definition 5.3.1. Length in $\mathbb{R}^n$: As in $\mathbb{R}^3$, the **length** $\|\mathbf{x}\|$ of the vector is defined by

$$\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

Where $\sqrt{(\quad)}$ indicates the positive square root.

A vector $\mathbf{x}$ of length 1 is called a **unit vector**. If $\mathbf{x} \neq \mathbf{0}$, then $\|\mathbf{x}\| \neq 0$ and it follows easily that $\frac{1}{\|\mathbf{x}\|}\mathbf{x}$ is a unit vector.

Example 5.3.2. If $\mathbf{x} = (1, -1, -3, 1)$ and $\mathbf{y} = (2, 1, 1, 0)$ in $\mathbb{R}^4$, find unit vectors using $\mathbf{x}$ and $\mathbf{y}$.

Theorem 5.3.3. *Let $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ denote vectors in $\mathbb{R}^n$. Then:*

(a) $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$.

(b) $\mathbf{x} \cdot (\mathbf{y} + \mathbf{z}) = \mathbf{x} \cdot \mathbf{y} + \mathbf{x} \cdot \mathbf{z}$.

(c) $(a\mathbf{x}) \cdot \mathbf{y} = a(\mathbf{x} \cdot \mathbf{y}) = \mathbf{x} \cdot (a\mathbf{y})$ *for all scalars* a.

(d) $\|\mathbf{x}\|^2 = \mathbf{x} \cdot \mathbf{x}$.

(e) $\|\mathbf{x}\| \geq 0$, *and* $\|\mathbf{x}\| = 0$ *if and only if* $\mathbf{x} = \mathbf{0}$.

(f) $\|a\mathbf{x}\| = |a|\|\mathbf{x}\|$ *for all scalars* a.

Example 5.3.4. Show that $\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + 2(\mathbf{x}\cdot\mathbf{y}) + \|\mathbf{y}\|^2$ for any $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$.

Example 5.3.5. Suppose that $\mathbb{R}^n = \operatorname{span}\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_k\}$ for some vectors $\mathbf{f}_i$. If $\mathbf{x}\cdot\mathbf{f}_i = 0$ for each i where $\mathbf{x}$ is in $\mathbb{R}^n$, show that $\mathbf{x} = \mathbf{0}$.

Theorem 5.3.6. *Cauchy Inequality*
If $\mathbf{x}$ *and* $\mathbf{y}$ *are vectors in* $\mathbb{R}^n$*, then*

$$|\mathbf{x}\cdot\mathbf{y}| \leq \|\mathbf{x}\|\|\mathbf{y}\|$$

Moreover $|\mathbf{x}\cdot\mathbf{y}| = \|\mathbf{x}\|\|\mathbf{y}\|$ *if and only if one of* $\mathbf{x}$ *and* $\mathbf{y}$ *is a multiple of the other.*

The Cauchy inequality is equivalent to $(\mathbf{x}\cdot\mathbf{y})^2 \leq \|\mathbf{x}\|^2\|\mathbf{y}\|^2$.

Corollary 5.3.7. *Triangle Inequality*
If $\mathbf{x}$ *and* $\mathbf{y}$ *are vectors in* $\mathbb{R}^n$*, then* $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

Definition 5.3.8. Distance in $\mathbb{R}^n$: If $\mathbf{x}$ and $\mathbf{y}$ are two vectors in $\mathbb{R}^n$, we define the **distance** $d(\mathbf{x},\mathbf{y})$ between $\mathbf{x}$ and $\mathbf{y}$ by

$$d(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|$$

Theorem 5.3.9. *If* $\mathbf{x}$*,* $\mathbf{y}$*, and* $\mathbf{z}$ *are three vectors in* $\mathbb{R}^n$ *we have:*

(a) $d(\mathbf{x},\mathbf{y}) \geq 0$ *for all* $\mathbf{x}$ *and* $\mathbf{y}$.

(b) $d(\mathbf{x},\mathbf{y}) = 0$ *if and only if* $\mathbf{x}=\mathbf{y}$.

(c) $d(\mathbf{x},\mathbf{y}) = d(\mathbf{y},\mathbf{x})$ *for all* $\mathbf{x}$ *and* $\mathbf{y}$.

(d) $d(\mathbf{x},\mathbf{z}) \leq d(\mathbf{x},\mathbf{y}) + d(\mathbf{y},\mathbf{z})$*for all* $\mathbf{x}$*,* $\mathbf{y}$*, and* $\mathbf{z}$. *Triangle inequality.*

Orthogonal Sets and the Expansion Theorem

Definition 5.3.10. Orthogonal and Orthonormal Sets: We say that two vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$ are **orthogonal** if $\mathbf{x}\cdot\mathbf{y}=0$, extending the terminology in $\mathbb{R}^3$. More generally, a set $\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_k\}$ of vectors in $\mathbb{R}^n$ is called an **orthogonal set** if

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

Note that $\{\mathbf{x}\}$ is an orthogonal set if $\mathbf{x}\neq\mathbf{0}$. A set $\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_k\}$ of vectors in $\mathbb{R}^n$ is called **orthonormal** if it is orthogonal and, in addition, each $\mathbf{x}_i$ is a unit vector:

$$\|\mathbf{x}_i\| = 1 \text{ for each } i.$$

Definition 5.3.11. Normalizing an Orthogonal Set: Hence if $\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_k\}$ is an orthogonal set, then $\{\frac{1}{\|\mathbf{x}_1\|}\mathbf{x}_1, \frac{1}{\|\mathbf{x}_2\|}\mathbf{x}_2,\cdots,\frac{1}{\|\mathbf{x}_k\|}\mathbf{x}_k\}$ is an orthonormal set, and we say that it is the result of **normalizing** the orthogonal set $\{\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_k\}$.

Example 5.3.12. If $\mathbf{f}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ -1 \end{bmatrix}$, $\mathbf{f}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix}$, $\mathbf{f}_3 = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$, and $\mathbf{f}_4 = \begin{bmatrix} -1 \\ 3 \\ -1 \\ 1 \end{bmatrix}$, verify $\{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3, \mathbf{f}_4\}$ is an orthogonal set in $\mathbb{R}^4$.

Theorem 5.3.13. *Pythagoras' Theorem*
If $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ is an orthogonal set in $\mathbb{R}^n$, then

$$\|\mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_k\|^2 = \|\mathbf{x}_1\|^2 + \|\mathbf{x}_2\|^2 + \cdots + \|\mathbf{x}_k\|^2.$$

Theorem 5.3.14. *Expansion Theorem*
Let $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_m\}$ be an orthogonal basis of a subspace U of $\mathbb{R}^n$. If $\mathbf{x}$ is any vector in U, we have

$$\mathbf{x} = \left(\frac{\mathbf{x} \cdot \mathbf{f}_1}{\|\mathbf{f}_1\|^2}\right) \mathbf{f}_1 + \left(\frac{\mathbf{x} \cdot \mathbf{f}_2}{\|\mathbf{f}_2\|^2}\right) \mathbf{f}_2 + \cdots + \left(\frac{\mathbf{x} \cdot \mathbf{f}_m}{\|\mathbf{f}_m\|^2}\right) \mathbf{f}_m$$

The expansion of $\mathbf{x}$ as a linear combination of the orthogonal basis $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_m\}$ is called the **Fourier expansion** of $\mathbf{x}$, and the coefficients $t_1 = \frac{\mathbf{x} \cdot \mathbf{f}_i}{\|\mathbf{f}_i\|^2}$ are called the **Fourier coefficients**. Note that if $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_m\}$ is actually orthonormal, then $t_i = \mathbf{x} \cdot \mathbf{f}_i$ for each i.

Example 5.3.15. Expand $\mathbf{x} = (a, b, c, d)$ as a linear combination of the orthogonal basis $\{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3, \mathbf{f}_4\}$ of $\mathbb{R}^4$ given in the previous example.

5.4 Rank of a Matrix

While it has been our custom to write the n-tuples in $\mathbb{R}^n$ as columns, in this section we will frequently write them as rows. Subspaces, independence, spanning, and dimension are defined for rows using matrix operations, just as for columns. If A is an $m \times n$ matrix, we define:

Definition 5.4.1. Column and Row Space of a Matrix

The **column space**, col$A$, of A is the subspace of $\mathbb{R}^m$ spanned by the columns of A.
The **row space**, row$A$, of A is the subspace of $\mathbb{R}^n$ spanned by the rows of A.

Lemma 5.4.2. *Let A and B denote $m \times n$ matrices.*

(a) *If $A \to B$ by elementary row operations, then row$A$ = row$B$.*

(b) *If $A \to B$ by elementary column operations, then col$A$ = col$B$.*

Lemma 5.4.3. *If R is a row-echelon matrix, then*

(a) *The nonzero rows of R are a basis of row$R$.*

(b) *The columns of R containing leading ones are a basis of col$R$.*

Example 5.4.4. Find a basis of $U = \text{span}\{(1, 1, 2, 3), (2, 4, 1, 0), (1, 5, -4, -9)\}$.

Theorem 5.4.5. *Rank Theorem*

Let A denote any $m \times n$ matrix of rank r. Then

$$dim(col\ A) = dim(row\ A) = r$$

Moreover, if A is carried to a row-echelon matrix R by row operations, then

(a) *The r nonzero rows of R are a basis of row$A$.*

(b) *If the leading 1s lie in columns $j_1, j_2, \ldots, j_r$ of R, then columns $j_1, j_2, \ldots, j_r$ of A are a basis of col$A$.*

Example 5.4.6. Compute the rank of $A = \begin{bmatrix} 1 & 2 & 2 & -1 \\ 3 & 6 & 5 & 0 \\ 1 & 2 & 1 & 2 \end{bmatrix}$ and find bases for row$A$ and col$A$.

Corollary 5.4.7. *If A is any matrix, then $rankA = rank(A^T)$.*

Corollary 5.4.8. *If A is an $m \times n$ matrix, then rank $A \leq m$ and rank $A \leq n$.*

Corollary 5.4.9. *$rankA = rank(UA) = rank(AV)$ whenever U and V are invertible.*

Lemma 5.4.10. *Let A, U, and V be matrices of sizes $m \times n$, $p \times m$, and $n \times q$ respectively.*

(a) *$col(AV) \subseteq colA$, with equality if $VV' = I_n$ for some V'.*

(b) *$row(UA) \subseteq rowA$, with equality if $U'U = I_m$ for some U'.*

Corollary 5.4.11. *If A is $m \times n$ and B is $n \times m$, then $rankAB \leq rankA$ and $rankAB \leq rankB$.*

Theorem 5.4.12. *Let A denote an $m \times n$ matrix of rank r. Then*

(a) *The $n - r$ basic solutions to the system $A\mathbf{x} = \mathbf{0}$ provided by the gaussian algorithm are a basis of null (A), so dim[null $(A)] = n - r$.*

(b) *Rank Theorem provides a basis of im $(A) = col(A)$, and dim[im $(A)] = r$.*

Example 5.4.13. If $A = \begin{bmatrix} 1 & -2 & 1 & 1 \\ -1 & 2 & 0 & 1 \\ 2 & -4 & 1 & 0 \end{bmatrix}$, find bases of null (A) and im (A), and so find their dimensions.

Theorem 5.4.14. *The following are equivalent for an $m \times n$ matrix A:*

(a) *$\text{rank}A = n$.*

(b) *The rows of A span $\mathbb{R}^n$.*

(c) *The columns of A are linearly independent in $\mathbb{R}^m$.*

(d) *The $n \times n$ matrix $A^T A$ is invertible.*

(e) *$CA = I_n$ for some $n \times m$ matrix C.*

(f) *If $A\mathbf{x} = \mathbf{0}$, $\mathbf{x}$ in $\mathbb{R}^n$, then $\mathbf{x} = \mathbf{0}$.*

Theorem 5.4.15. *The following are equivalent for an $m \times n$ matrix A:*

(a) *$\text{rank}A = m$.*

(b) *The columns of A span $\mathbb{R}^m$.*

(c) *The rows of A are linearly independent in $\mathbb{R}^n$.*

(d) *The $m \times m$ matrix AA^T is invertible.*

(e) *$AC = I_m$ for some $n \times m$ matrix C.*

(f) *The system $A\mathbf{x} = \mathbf{b}$ is consistent for every $\mathbf{b}$ in $\mathbb{R}^m$.*

Example 5.4.16. Show that $\begin{bmatrix} 3 & x+y+z \\ x+y+z & x^2+y^2+z^2 \end{bmatrix}$ is invertible if x, y, and z are not all equal.

5.5 Similarity and Diagonalization

Similar Matrices

Definition 5.5.1. Similar Matrices: If A and B are $n \times n$ matrices, we say that A and B are **similar**, and write $A \sim B$, if $B = P^{-1}AP$ for some invertible matrix P.

If $A \sim B$, then necessarily $B \sim A$. To see why, suppose that $B = P^{-1}AP$. Then $A = PBP^{-1} = Q^{-1}BQ$ where $Q = P^{-1}$ is invertible. This proves the second of the following properties of similarity:

1. *$A \sim A$ for all square matrices A.*
2. *If $A \sim B$, then $B \sim A$.*
3. *If $A \sim B$ and $B \sim C$, then $A \sim C$.*

These properties are often expressed by saying that the similarity relation $\sim$ is an **equivalence relation** on the set of $n \times n$ matrices.

Example 5.5.2. If A is similar to B and either A or B is diagonalizable, show that the other is also diagonalizable.

Similarity is compatible with inverses, transposes, and powers:

$$\text{If } A \sim B \text{ then} \quad A^{-1} \sim B^{-1}, \quad A^T \sim B^T, \quad \text{and} \quad A^k \sim B^k \text{ for all integers } k \geq 1.$$

Definition 5.5.3. Trace of a Matrix: The **trace** tr$A$ of an $n \times n$ matrix A is defined to be the sum of the main diagonal elements of A.

In other words:

$$\text{If } A = [a_{ij}], \text{ then tr } A = a_{11} + a_{22} + \cdots + a_{nn}.$$

Lemma 5.5.4. *Let A and B be $n \times n$ matrices. Then $tr(AB) = tr(BA)$.*

Theorem 5.5.5. *If A and B are similar $n \times n$ matrices, then A and B have the same determinant, rank, trace, characteristic polynomial, and eigenvalues.*

Example 5.5.6. Are the matrices $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ similar?

Diagonalization Revisited

Recall that a square matrix A is **diagonalizable** if there exists an invertible matrix P such that $P^{-1}AP = D$ is a diagonal matrix, that is if A is similar to a diagonal matrix D. Unfortunately, not all matrices are diagonalizable. Determining whether A is diagonalizable is closely related to the eigenvalues and eigenvectors of A. Recall that a number λ is called an **eigenvalue** of A if $A\mathbf{x} = \lambda\mathbf{x}$ for some nonzero column $\mathbf{x}$ in $\mathbb{R}^n$, and any such nonzero vector $\mathbf{x}$ is called an **eigenvector** of A corresponding to λ (or simply a λ-eigenvector of A). The eigenvalues and eigenvectors of A are closely related to the **characteristic polynomial** $c_A(x)$ of A, defined by

$$c_A(x) = \det(xI - A)$$

Theorem 5.5.7. *Let A be an $n \times n$ matrix.*

(a) *The eigenvalues λ of A are the roots of the characteristic polynomial $c_A(x)$ of A.*

(b) *The λ-eigenvectors $\mathbf{x}$ are the nonzero solutions to the homogeneous system*

$$(\lambda I - A)\mathbf{x} = \mathbf{0}$$

of linear equations with $\lambda I - A$ as coefficient matrix.

Example 5.5.8. Show that the eigenvalues of a triangular matrix are the main diagonal entries.

Theorem 5.5.9. *Let A be an $n \times n$ matrix.*

(a) *A is diagonalizable if and only if $\mathbb{R}^n$ has a basis $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ consisting of eigenvectors of A.*

(b) *When this is the case, the matrix $P = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix}$ is invertible and $P^{-1}AP = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ where, for each i, λ_i is the eigenvalue of A corresponding to $\mathbf{x}_i$.*

Theorem 5.5.10. *Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ be eigenvectors corresponding to distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$ of an $n \times n$ matrix A. Then $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ is a linearly independent set.*

Theorem 5.5.11. *If A is an $n \times n$ matrix with n distinct eigenvalues, then A is diagonalizable.*

Example 5.5.12. Show that $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 3 \\ -1 & 1 & 0 \end{bmatrix}$ is diagonalizable.

Lemma 5.5.13. *Let $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ be a linearly independent set of eigenvectors of an $n \times n$ matrix A, extend it to a basis $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k, \ldots, \mathbf{x}_n\}$ of $\mathbb{R}^n$, and let*

$$P = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix}$$

be the (invertible) $n \times n$ matrix with the $\mathbf{x}_i$ as its columns. If $\lambda_1, \lambda_2, \ldots, \lambda_k$ are the (not necessarily distinct) eigenvalues of A corresponding to $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ respectively, then $P^{-1}AP$ has block form

$$P^{-1}AP = \begin{bmatrix} diag(\lambda_1, \lambda_2, \ldots, \lambda_k) & B \\ 0 & A_1 \end{bmatrix}$$

where B has size $k \times (n-k)$ and A_1 has size $(n-k) \times (n-k)$.

Note that this Lemma(with $k = n$) shows that an $n \times n$ matrix A is diagonalizable if $\mathbb{R}^n$ has a basis of eigenvectors of A.

Definition 5.5.14. Eigenspace of a Matrix: If λ is an eigenvalue of an $n \times n$ matrix A, define the **eigenspace** of A corresponding to λ by

$$E_\lambda(A) = \{\mathbf{x} \text{ in } \mathbb{R}^n \mid A\mathbf{x} = \lambda\mathbf{x}\}$$

Recall that the **multiplicity** of an eigenvalue λ of A is the number of times λ occurs as a root of the characteristic polynomial $c_A(x)$ of A. In other words, the multiplicity of λ is the largest integer $m \geq 1$ such that

$$c_A(x) = (x - \lambda)^m g(x)$$

for some polynomial $g(x)$.

Lemma 5.5.15. *Let λ be an eigenvalue of multiplicity m of a square matrix A. Then $dim\,[E_\lambda(A)] \leq m$.*

When does $\dim\,[E_\lambda(A)] = m$ for each eigenvalue λ? It turns out that this characterizes the diagonalizable $n \times n$ matrices A for which $c_A(x)$ **factors completely** over $\mathbb{R}$. By this we mean that $c_A(x) = (x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_n)$, where the λ_i are *real* numbers (not necessarily distinct); in other words, every eigenvalue of A is real.

Theorem 5.5.16. *The following are equivalent for a square matrix A for which $c_A(x)$ factors completely.*

(a) *A is diagonalizable.*

(b) *$dim[E_\lambda(A)]$ equals the multiplicity of λ for every eigenvalue λ of the matrix A.*

Example 5.5.17. If $A = \begin{bmatrix} 5 & 8 & 16 \\ 4 & 1 & 8 \\ -4 & -4 & -11 \end{bmatrix}$ and $B = \begin{bmatrix} 2 & 1 & 1 \\ 2 & 1 & -2 \\ -1 & 0 & -2 \end{bmatrix}$ show that A is diagonalizable but B is not.

Complex Eigenvalues

All the matrices we have considered have had real eigenvalues. But this need not be the case: The matrix $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ has characteristic polynomial $c_A(x) = x^2 + 1$ which has no real roots. Nonetheless, this matrix is diagonalizable; the only difference is that we must use a larger set of scalars, the complex numbers.

Indeed, nearly everything we have done for real matrices can be done for complex matrices. The methods are the same; the only difference is that the arithmetic is carried out with complex numbers rather than real ones.

But the complex numbers are better than the real numbers in one respect: While there are polynomials like x^2+1 with real coefficients that have no real root, this problem does not arise with the complex numbers: *Every* nonconstant polynomial with complex coefficients has a complex root, and hence factors completely as a product of linear factors. This fact is known as the fundamental theorem of algebra.

Example 5.5.18. Diagonalize the matrix $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.

CHAPTER 6

Vector Spaces

In this chapter we introduce vector spaces in full generality. We deal with the notion of an *abstract* vector space. It turns out that there are many systems in which a natural addition and scalar multiplication are defined and satisfy the usual rules familiar from $\mathbb{R}^n$. The study of abstract vector spaces is a way to deal with all these examples simultaneously. The new aspect is that we are dealing with an abstract system in which *all we know* about the vectors is that they are objects that can be added and multiplied by a scalar and satisfy rules familiar from $\mathbb{R}^n$.

6.1 Examples and Basic Properties

Definition 6.1.1. Vector Spaces
A **vector space** consists of a nonempty set V of objects (called **vectors**) that can be added, that can be multiplied by a real number (called a **scalar** in this context), and for which certain axioms hold. If $\mathbf{v}$ and $\mathbf{w}$ are two vectors in V, their sum is expressed as $\mathbf{v}+\mathbf{w}$, and the scalar product of $\mathbf{v}$ by a real number a is denoted as $a\mathbf{v}$. These operations are called **vector addition** and **scalar multiplication**, respectively, and the following axioms are assumed to hold.

Axioms for vector addition

A1. *If* $\mathbf{u}$ *and* $\mathbf{v}$ *are in* V*, then* $\mathbf{u}+\mathbf{v}$ *is in* V.

A2. $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$ *for all* $\mathbf{u}$ *and* $\mathbf{v}$ *in* V.

A3. $\mathbf{u}+(\mathbf{v}+\mathbf{w})=(\mathbf{u}+\mathbf{v})+\mathbf{w}$ *for all* $\mathbf{u}$, $\mathbf{v}$, *and* $\mathbf{w}$ *in* V.

A4. *An element* $\mathbf{0}$ *in* V *exists such that* $\mathbf{v}+\mathbf{0}=\mathbf{v}=\mathbf{0}+\mathbf{v}$ *for every* $\mathbf{v}$ *in* V.

A5. *For each* $\mathbf{v}$ *in* V*, an element* $-\mathbf{v}$ *in* V *exists such that* $-\mathbf{v}+\mathbf{v}=\mathbf{0}$ *and* $\mathbf{v}+(-\mathbf{v})=\mathbf{0}$.

Axioms for scalar multiplication

S1. *If* $\mathbf{v}$ *is in* V*, then* $a\mathbf{v}$ *is in* V *for all* a *in* $\mathbb{R}$.

S2. $a(\mathbf{v}+\mathbf{w})=a\mathbf{v}+a\mathbf{w}$ *for all* $\mathbf{v}$ *and* $\mathbf{w}$ *in* V *and all* a *in* $\mathbb{R}$.

S3. $(a+b)\mathbf{v}=a\mathbf{v}+b\mathbf{v}$ *for all* $\mathbf{v}$ *in* V *and all* a *and* b *in* $\mathbb{R}$.

S4. $a(b\mathbf{v})=(ab)\mathbf{v}$ *for all* $\mathbf{v}$ *in* V *and all* a *and* b *in* $\mathbb{R}$.

S5. $1\mathbf{v}=\mathbf{v}$ *for all* $\mathbf{v}$ *in* V.

The content of axioms A1 and S1 is described by saying that V is **closed** under vector addition and scalar multiplication. The element $\mathbf{0}$ in axiom A4 is called the **zero vector**, and the vector $-\mathbf{v}$ in axiom A5 is called the **negative** of $\mathbf{v}$.

Example 6.1.2. $\mathbb{R}^n$ is a vector space using matrix addition and scalar multiplication.

In a general vector space, the vectors need not be n-tuples as in $\mathbb{R}^n$. They can be any kind of objects at all as long as the addition and scalar multiplication are defined and the axioms are satisfied.

Example 6.1.3. Show the set $\mathbf{M}_{mn}$ of all $m \times n$ matrices is a vector space using matrix addition and scalar multiplication.

Example 6.1.4. Show that every subspace of $\mathbb{R}^n$ is a vector space in its own right using the addition and scalar multiplication of $\mathbb{R}^n$.

Example 6.1.5. Let V denote the set of all ordered pairs (x, y) and define addition in V as in $\mathbb{R}^2$. However, define a new scalar multiplication in V by

$$a(x, y) = (ay, ax)$$

Determine if V is a vector space with these operations.

A **polynomial** in an indeterminate x is an expression

$$p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

where $a_0, a_1, a_2, \ldots, a_n$ are real numbers called the **coefficients** of the polynomial. If all the coefficients are zero, the polynomial is called the **zero polynomial** and is denoted simply as 0. If $p(x) \neq 0$, the highest power of x with a nonzero coefficient is called the **degree** of $p(x)$ denoted as $\deg p(x)$. The coefficient itself is called the **leading coefficient** of $p(x)$. The degree of the zero polynomial is not defined.

Let **P** denote the set of all polynomials and suppose that

$$\begin{aligned} p(x) &= a_0 + a_1x + a_2x^2 + \cdots \\ q(x) &= b_0 + b_1x + b_2x^2 + \cdots \end{aligned}$$

are two polynomials in **P** (possibly of different degrees). Then $p(x)$ and $q(x)$ are called **equal** [written $p(x) = q(x)$] if and only if all the corresponding coefficients are equal—that is, $a_0 = b_0$, $a_1 = b_1$, $a_2 = b_2$, and so on. In particular, $a_0 + a_1x + a_2x^2 + \cdots = 0$ means that $a_0 = 0$, $a_1 = 0$, $a_2 = 0$, $\ldots$, and this is the reason for calling x an **indeterminate**. The set **P** has an addition and scalar multiplication defined on it as follows: if $p(x)$ and $q(x)$ are as before and a is a real number,

$$\begin{aligned} p(x) + q(x) &= (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + \cdots \\ ap(x) &= aa_0 + (aa_1)x + (aa_2)x^2 + \cdots \end{aligned}$$

Evidently, these are again polynomials, so **P** is closed under these operations, called **pointwise** addition and scalar multiplication. The other vector space axioms are easily verified.

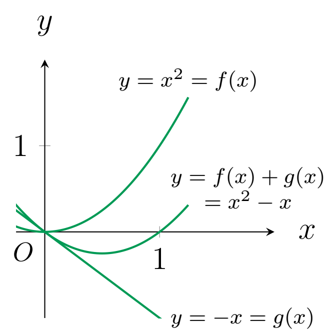
Example 6.1.6. The set $\mathbf{P}$ of all polynomials is a vector space with the foregoing addition and scalar multiplication. The zero vector is the zero polynomial, and the negative of a polynomial $p(x) = a_0 + a_1x + a_2x^2 + \dots$ is the polynomial $-p(x) = -a_0 - a_1x - a_2x^2 - \dots$ obtained by negating all the coefficients.

Example 6.1.7. Given $n \geq 1$, let $\mathbf{P}_n$ denote the set of all polynomials of degree at most n, together with the zero polynomial. That is

$$\mathbf{P}_n = \{a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \mid a_0, a_1, a_2, \dots, a_n \text{ in } \mathbb{R}\}.$$

Then $\mathbf{P}_n$ is a vector space. Indeed, sums and scalar multiples of polynomials in $\mathbf{P}_n$ are again in $\mathbf{P}_n$, and the other vector space axioms are inherited from $\mathbf{P}$. In particular, the zero vector and the negative of a polynomial in $\mathbf{P}_n$ are the same as those in $\mathbf{P}$.

If a and b are real numbers and $a < b$, the **interval** $[a, b]$ is defined to be the set of all real numbers x such that $a \leq x \leq b$. A (real-valued) **function** f on $[a, b]$ is a rule that associates to every number x in $[a, b]$ a real number denoted $f(x)$. The rule is frequently specified by giving a formula for $f(x)$ in terms of x. In fact, every polynomial $p(x)$ can be regarded as the formula for a function p.



The set of all functions on $[a, b]$ is denoted $\mathbf{F}[a, b]$. Two functions f and g in $\mathbf{F}[a, b]$ are **equal** if $f(x) = g(x)$ for every x in $[a, b]$, and we describe this by saying that f and g have the **same action**. Note that two polynomials are equal in $\mathbf{P}$ if and only if they are equal as functions.

If f and g are two functions in $\mathbf{F}[a, b]$, and if r is a real number, define the sum $f + g$ and the scalar product rf by

$$\begin{aligned} (f + g)(x) &= f(x) + g(x) && \text{for each } x \text{ in } [a, b] \\ (rf)(x) &= rf(x) && \text{for each } x \text{ in } [a, b] \end{aligned}$$

In other words, the action of $f+g$ upon x is to associate x with the number $f(x)+g(x)$, and rf associates x with $rf(x)$. The sum of $f(x) = x^2$ and $g(x) = -x$ is shown in the diagram. These operations on $\mathbf{F}[a, b]$ are called **pointwise addition and scalar multiplication** of functions and they are the usual operations familiar from elementary algebra and calculus.

Example 6.1.8. The set $\mathbf{F}[a,b]$ of all functions on the interval $[a,b]$ is a vector space using pointwise addition and scalar multiplication. The zero function (in axiom A4), denoted 0, is the constant function defined by

$$0(x) = 0 \quad \text{for each } x \text{ in } [a,b]$$

The negative of a function f is denoted $-f$ and has action defined by

$$(-f)(x) = -f(x) \quad \text{for each } x \text{ in } [a,b]$$

Axioms A1 and S1 are clearly satisfied because, if f and g are functions on $[a,b]$, then $f+g$ and rf are again such functions.

Theorem 6.1.9. *Cancellation*
Let $\mathbf{u}$, $\mathbf{v}$, *and* $\mathbf{w}$ *be vectors in a vector space* V. *If* $\mathbf{v}+\mathbf{u} = \mathbf{v}+\mathbf{w}$, *then* $\mathbf{u} = \mathbf{w}$.

To subtract a vector $\mathbf{v}$ from both sides of a vector equation, we added $-\mathbf{v}$ to both sides. With this in mind, we define **difference** $\mathbf{u}-\mathbf{v}$ of two vectors in V as

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

We shall say that this vector is the result of having **subtracted** $\mathbf{v}$ from $\mathbf{u}$.

Theorem 6.1.10. *If* $\mathbf{u}$ *and* $\mathbf{v}$ *are vectors in a vector space* V, *the equation*

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

has one and only one solution $\mathbf{x}$ *in* V *given by*

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

Similarly, cancellation shows that there is only one zero vector in any vector space and only one negative of each vector. Hence we speak of *the* zero vector and *the* negative of a vector.

Theorem 6.1.11. *Let* $\mathbf{v}$ *denote a vector in a vector space* V *and let* a *denote a real number.*

(a) $0\mathbf{v} = \mathbf{0}$.

(b) $a\mathbf{0} = \mathbf{0}$.

(c) *If* $a\mathbf{v} = \mathbf{0}$, *then either* $a = 0$ *or* $\mathbf{v} = \mathbf{0}$.

(d) $(-1)\mathbf{v} = -\mathbf{v}$.

(e) $(-a)\mathbf{v} = -(a\mathbf{v}) = a(-\mathbf{v})$.

The properties in the previous Theorem are familiar for matrices; the point here is that they hold in *every* vector space.

Example 6.1.12. If $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are vectors in a vector space V, simplify the expression

$$2(\mathbf{u}+3\mathbf{w})-3(2\mathbf{w}-\mathbf{v})-3[2(2\mathbf{u}+\mathbf{v}-4\mathbf{w})-4(\mathbf{u}-2\mathbf{w})]$$

Example 6.1.13. A set $\{\mathbf{0}\}$ with one element becomes a vector space if we define

$$\mathbf{0}+\mathbf{0}=\mathbf{0} \quad \text{and} \quad a\mathbf{0}=\mathbf{0} \text{ for all scalars } a.$$

The resulting space is called the **zero vector space** and is denoted $\{\mathbf{0}\}$.

6.2 Subspaces and Spanning Sets

Definition 6.2.1. Subspaces of a Vector Space
If V is a vector space, a nonempty subset $U \subseteq V$ is called a **subspace** of V if U is itself a vector space using the addition and scalar multiplication of V.

Theorem 6.2.2. *Subspace Test*
A subset U of a vector space is a subspace of V if and only if it satisfies the following three conditions:

(a) $\mathbf{0}$ *lies in U where* $\mathbf{0}$ *is the zero vector of V.*

(b) *If* $\mathbf{u}_1$ *and* $\mathbf{u}_2$ *are in U, then* $\mathbf{u}_1 + \mathbf{u}_2$ *is also in U.*

(c) *If* $\mathbf{u}$ *is in U, then $a\mathbf{u}$ is also in U for each scalar a.*

Example 6.2.3. If V is any vector space, show that $\{\mathbf{0}\}$ and V are subspaces of V.

The vector space $\{\mathbf{0}\}$ is called the **zero subspace** of V.

Example 6.2.4. Let $\mathbf{v}$ be a vector in a vector space V. Show that the set

$$\mathbb{R}\mathbf{v} = \{a\mathbf{v} \mid a \text{ in } \mathbb{R}\}$$

of all scalar multiples of $\mathbf{v}$ is a subspace of V.

Example 6.2.5. Let A be a fixed matrix in $\mathbf{M}_{nn}$. Show that $U = \{X \text{ in } \mathbf{M}_{nn} \mid AX = XA\}$ is a subspace of $\mathbf{M}_{nn}$.

Suppose $p(x)$ is a polynomial and a is a number. Then the number $p(a)$ obtained by replacing x by a in the expression for $p(x)$ is called the **evaluation** of $p(x)$ at a. If $p(a) = 0$, the number a is called a **root** of $p(x)$.

Example 6.2.6. Consider the set U of all polynomials in $\mathbf{P}$ that have 3 as a root:

$$U = \{p(x) \in \mathbf{P} \mid p(3) = 0\}$$

Show that U is a subspace of $\mathbf{P}$.

Recall that the space $\mathbf{P}_n$ consists of all polynomials of the form

$$a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

where $a_0, a_1, a_2, \ldots, a_n$ are real numbers, and so is closed under the addition and scalar multiplication in $\mathbf{P}$. Moreover, the zero polynomial is included in $\mathbf{P}_n$. Thus the subspace test gives

Example 6.2.7. $\mathbf{P}_n$ is a subspace of $\mathbf{P}$ for each $n \geq 0$.

Example 6.2.8. Show that the subset $\mathbf{D}[a, b]$ of all **differentiable functions** on $[a, b]$ is a subspace of the vector space $\mathbf{F}[a, b]$ of all functions on $[a, b]$.

Linear Combinations and Spanning Sets

Definition 6.2.9. Linear Combinations and Spanning
Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a set of vectors in a vector space V. As in $\mathbb{R}^n$, a vector $\mathbf{v}$ is called a **linear combination** of the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ if it can be expressed in the form

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n$$

where $a_1, a_2, \ldots, a_n$ are scalars, called the **coefficients** of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$. The set of all linear combinations of these vectors is called their **span**, and is denoted by

$$\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} = \{a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n \mid a_i \text{ in } \mathbb{R}\}$$

If it happens that $V = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$, these vectors are called a **spanning set** for V. For example, the span of two vectors $\mathbf{v}$ and $\mathbf{w}$ is the set

$$\text{span}\{\mathbf{v}, \mathbf{w}\} = \{s\mathbf{v} + t\mathbf{w} \mid s \text{ and } t \text{ in } \mathbb{R}\}$$

of all sums of scalar multiples of these vectors.

Example 6.2.10. Consider the vectors $p_1 = 1 + x + 4x^2$ and $p_2 = 1 + 5x + x^2$ in $\mathbf{P}_2$. Determine whether p_1 and p_2 lie in $\text{span}\{1 + 2x - x^2, 3 + 5x + 2x^2\}$.

Example 6.2.11. Show $\mathbf{M}_{mn}$ is the span of the set of all $m \times n$ matrices with exactly one entry equal to 1, and all other entries zero.

The fact that every polynomial in $\mathbf{P}_n$ has the form $a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$ where each a_i is in $\mathbb{R}$ shows that

Example 6.2.12. $\mathbf{P}_n = \text{span}\{1, x, x^2, \ldots, x^n\}$.

Theorem 6.2.13. *Let $U = span\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ in a vector space V. Then:*

(a) *U is a subspace of V containing each of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$.*

(b) *U is the "smallest" subspace containing these vectors in the sense that any subspace that contains each of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ must contain U.*

Example 6.2.14. Show that $\mathbf{P}_3 = \text{span}\{x^2 + x^3, x, 2x^2 + 1, 3\}$.

Example 6.2.15. Let $\mathbf{u}$ and $\mathbf{v}$ be two vectors in a vector space V. Show that

$$\text{span}\{\mathbf{u}, \mathbf{v}\} = \text{span}\{\mathbf{u} + 2\mathbf{v}, \mathbf{u} - \mathbf{v}\}$$

6.3 Linear Independence and Dimension

Definition 6.3.1. Linear Independence and Dependence
As in $\mathbb{R}^n$, a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ in a vector space V is called **linearly independent** (or simply **independent**) if it satisfies the following condition:

$$\text{If} \quad s_1\mathbf{v}_1 + s_2\mathbf{v}_2 + \cdots + s_n\mathbf{v}_n = \mathbf{0}, \quad \text{then} \quad s_1 = s_2 = \cdots = s_n = 0.$$

A set of vectors that is not linearly independent is said to be **linearly dependent** (or simply **dependent**).

The **trivial linear combination** of the vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ is the one with every coefficient zero:

$$0\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_n$$

This is obviously one way of expressing $\mathbf{0}$ as a linear combination of the vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$, and they are linearly independent when it is the *only* way.

Example 6.3.2. Show that $\{1 + x, 3x + x^2, 2 + x - x^2\}$ is independent in $\mathbf{P}_2$.

Example 6.3.3. Show that $\{\sin x, \cos x\}$ is independent in the vector space $\mathbf{F}[0, 2\pi]$ of functions defined on the interval $[0, 2\pi]$.

Example 6.3.4. Suppose that $\{\mathbf{u}, \mathbf{v}\}$ is an independent set in a vector space V. Show that $\{\mathbf{u} + 2\mathbf{v}, \mathbf{u} - 3\mathbf{v}\}$ is also independent.

Example 6.3.5. Show that any set of polynomials of distinct degrees is independent.

Example 6.3.6. Suppose that A is an $n \times n$ matrix such that $A^k = 0$ but $A^{k-1} \neq 0$. Show that $B = \{I, A, A^2, \dots, A^{k-1}\}$ is independent in $\mathbf{M}_{nn}$.

Example 6.3.7. Let V denote a vector space.

(a) If $\mathbf{v} \neq \mathbf{0}$ in V, then $\{\mathbf{v}\}$ is an independent set.

(b) No independent set of vectors in V can contain the zero vector.

Theorem 6.3.8. *Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a linearly independent set of vectors in a vector space V. If a vector $\mathbf{v}$ has two (ostensibly different) representations*

$$\begin{aligned}\mathbf{v} &= s_1\mathbf{v}_1 + s_2\mathbf{v}_2 + \cdots + s_n\mathbf{v}_n \\ \mathbf{v} &= t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_n\mathbf{v}_n\end{aligned}$$

as linear combinations of these vectors, then $s_1 = t_1, s_2 = t_2, \ldots, s_n = t_n$. In other words, every vector in V can be written in a unique way as a linear combination of the $\mathbf{v}_i$.

Theorem 6.3.9. *Fundamental Theorem*
Suppose a vector space V can be spanned by n vectors. If any set of m vectors in V is linearly independent, then $m \leq n$.

If $V = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$, and if $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ is an independent set in V, the above proof shows not only that $m \leq n$ but also that m of the (spanning) vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ can be replaced by the (independent) vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ and the resulting set will still span$V$. In this form the result is called the **Steinitz Exchange Lemma**.

Definition 6.3.10. Basis of a Vector Space
As in $\mathbb{R}^n$, a set $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ of vectors in a vector space V is called a **basis** of V if it satisfies the following two conditions:

(a) $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is linearly independent

(b) $V = \text{span}\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$

Theorem 6.3.11. *Invariance Theorem*
Let $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ *and* $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_m\}$ *be two bases of a vector space* V. *Then* $n = m$.

Definition 6.3.12. Dimension of a Vector Space
If $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is a basis of the nonzero vector space V, the number n of vectors in the basis is called the **dimension** of V, and we write

$$\dim V = n$$

The zero vector space $\{\mathbf{0}\}$ is defined to have dimension 0:

$$\dim \{\mathbf{0}\} = 0$$

Example 6.3.13. The space $\mathbf{M}_{mn}$ has dimension mn, and one basis consists of all $m \times n$ matrices with exactly one entry equal to 1 and all other entries equal to 0. We call this the **standard basis** of $\mathbf{M}_{mn}$.

Example 6.3.14. Show that $\dim \mathbf{P}_n = n+1$ and that $\{1, x, x^2, \ldots, x^n\}$ is a basis, called the **standard basis** of $\mathbf{P}_n$.

Example 6.3.15. If $\mathbf{v} \neq \mathbf{0}$ is any nonzero vector in a vector space V, show that $\operatorname{span}\{\mathbf{v}\} = \mathbb{R}\mathbf{v}$ has dimension 1.

Example 6.3.16. Let $A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ and consider the subspace

$$U = \{X \text{ in } \mathbf{M}_{22} \mid AX = XA\}$$

of $\mathbf{M}_{22}$. Show that $\dim U = 2$ and find a basis of U.

Example 6.3.17. Show that the set V of all symmetric 2×2 matrices is a vector space, and find the dimension of V.

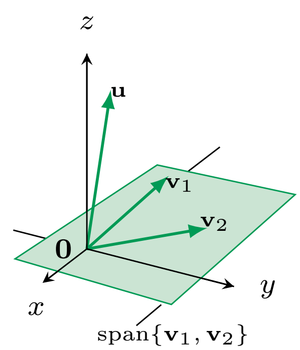
Example 6.3.18. Let $B = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ be nonzero vectors in a vector space V. Given nonzero scalars $a_1, a_2, \dots, a_n$, write $D = \{a_1\mathbf{v}_1, a_2\mathbf{v}_2, \dots, a_n\mathbf{v}_n\}$. If B is independent or spans V, the same is true of D. In particular, if B is a basis of V, so also is D.

6.4 Finite Dimensional Spaces

Lemma 6.4.1. *Independent Lemma*
Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ be an independent set of vectors in a vector space V. If $\mathbf{u} \in V$ but $\mathbf{u} \notin span\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$, then $\{\mathbf{u}, \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is also independent.

Note that the converse of the Independent Lemma is also true: if $\{\mathbf{u}, \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is independent, then $\mathbf{u}$ is not in span$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$.

As an illustration, suppose that $\{\mathbf{v}_1, \mathbf{v}_2\}$ is independent in $\mathbb{R}^3$. Then $\mathbf{v}_1$ and $\mathbf{v}_2$ are not parallel, so span$\{\mathbf{v}_1, \mathbf{v}_2\}$ is a plane through the origin (shaded in the diagram). By the Independent Lemma, $\mathbf{u}$ is not in this plane if and only if $\{\mathbf{u}, \mathbf{v}_1, \mathbf{v}_2\}$ is independent.



Definition 6.4.2. Finite Dimensional and Infinite Dimensional Vector Spaces
A vector space V is called **finite dimensional** if it is spanned by a finite set of vectors. Otherwise, V is called **infinite dimensional**.

Thus the zero vector space $\{\mathbf{0}\}$ is finite dimensional because $\{\mathbf{0}\}$ is a spanning set.

Lemma 6.4.3. *Let V be a finite dimensional vector space. If U is any subspace of V, then any independent subset of U can be enlarged to a finite basis of U.*

Theorem 6.4.4. *019430 Let V be a finite dimensional vector space spanned by m vectors.*

(a) *V has a finite basis, and dim$V \leq m$.*

(b) *Every independent set of vectors in V can be enlarged to a basis of V by adding vectors from any fixed basis of V.*

(c) *If U is a subspace of V, then*

   a. *U is finite dimensional and dim$U \leq$ dim$V$.*

   b. *If dim$U =$ dim$V$ then $U = V$.*

Example 6.4.5. Enlarge the independent set $D = \left\{ \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \right\}$ to a basis of $\mathbf{M}_{22}$.

Example 6.4.6. Find a basis of $\mathbf{P}_3$ containing the independent set $\{1+x, 1+x^2\}$.

Example 6.4.7. Show that the space $\mathbf{P}$ of all polynomials is infinite dimensional.

Example 6.4.8. If $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_k$ are independent columns in $\mathbb{R}^n$, show that they are the first k columns in some invertible $n \times n$ matrix.

Theorem 6.4.9. *Let U and W be subspaces of the finite dimensional space V.*

(a) *If $U \subseteq W$, then $dim U \leq dim W$.*

(b) *If $U \subseteq W$ and $dim U = dim W$, then $U = W$.*

Example 6.4.10. If a is a number, let W denote the subspace of all polynomials in $\mathbf{P}_n$ that have a as a root:

$$W = \{p(x) \mid p(x) \in \mathbf{P}_n \text{ and } p(a) = 0\}$$

Show that $\{(x-a), (x-a)^2, \ldots, (x-a)^n\}$ is a basis of W.

A set of vectors is called **dependent** if it is *not* independent, that is if some nontrivial linear combination vanishes.

Lemma 6.4.11. *Dependent Lemma*
A set $D = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ of vectors in a vector space V is dependent if and only if some vector in D is a linear combination of the others.

Theorem 6.4.12. *Let V be a finite dimensional vector space. Any spanning set for V can be cut down (by deleting vectors) to a basis of V.*

Example 6.4.13. Find a basis of $\mathbf{P}_3$ in the spanning set $S = \{1, x + x^2, 2x - 3x^2, 1 + 3x - 2x^2, x^3\}$.

Theorem 6.4.14. *Let V be a vector space with $\dim V = n$, and suppose S is a set of exactly n vectors in V. Then S is independent if and only if S spans V.*

Example 6.4.15. Consider the set $S = \{p_0(x), p_1(x), \ldots, p_n(x)\}$ of polynomials in $\mathbf{P}_n$. If $\deg p_k(x) = k$ for each k, show that S is a basis of $\mathbf{P}_n$.

Example 6.4.16. Let V denote the space of all symmetric 2×2 matrices. Find a basis of V consisting of invertible matrices.

Example 6.4.17. Let A be any $n \times n$ matrix. Show that there exist $n^2 + 1$ scalars $a_0, a_1, a_2, \ldots, a_{n^2}$ not all zero, such that

$$a_0 I + a_1 A + a_2 A^2 + \cdots + a_{n^2} A^{n^2} = 0$$

where I denotes the $n \times n$ identity matrix.

If U and W are subspaces of a vector space V, there are two related subspaces that are of interest, their **sum** $U + W$ and their **intersection** $U \cap W$, defined by

$$\begin{aligned} U + W &= \{\mathbf{u} + \mathbf{w} \mid \mathbf{u} \in U \text{ and } \mathbf{w} \in W\} \\ U \cap W &= \{\mathbf{v} \in V \mid \mathbf{v} \in U \text{ and } \mathbf{v} \in W\} \end{aligned}$$

It is routine to verify that these are indeed subspaces of V, that $U \cap W$ is contained in both U and W, and that $U + W$ contains both U and W.

Theorem 6.4.18. *Suppose that U and W are finite dimensional subspaces of a vector space V. Then $U + W$ is finite dimensional and*

$$dim(U + W) = dim\ U + dim\ W - dim(U \cap W).$$

If $U \cap W = \{\mathbf{0}\}$, then there are *no* vectors $\mathbf{x}_i$ in the above proof, and the argument shows that if $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ and $\{\mathbf{w}_1, \ldots, \mathbf{w}_p\}$ are bases of U and W respectively, then $\{\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{w}_1, \ldots, \mathbf{w}_p\}$ is a basis of $U + W$. In this case $U + W$ is said to be a **direct sum** (written $U \oplus W$).

CHAPTER 7

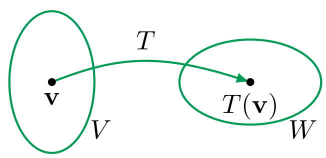
Linear Transformations

7.1 Examples and Elementary Properties

Definition 7.1.1. Linear Transformations of Vector Spaces
If V and W are two vector spaces, a function $T : V \to W$ is called a **linear transformation** if it satisfies the following axioms.

T1.	$T(\mathbf{v} + \mathbf{v}_1) = T(\mathbf{v}) + T(\mathbf{v}_1)$	for all $\mathbf{v}$ and $\mathbf{v}_1$ in V.
T2.	$T(r\mathbf{v}) = rT(\mathbf{v})$	for all $\mathbf{v}$ in V and r in $\mathbb{R}$.

A linear transformation $T : V \to V$ is called a **linear operator** on V. The situation can be visualized as in the diagram below.



Axiom T1 is just the requirement that T *preserves* vector addition. It asserts that the result $T(\mathbf{v} + \mathbf{v}_1)$ of adding $\mathbf{v}$ and $\mathbf{v}_1$ first and then applying T is the same as applying T first to get $T(\mathbf{v})$ and $T(\mathbf{v}_1)$ and then adding. Similarly, axiom T2 means that T *preserves* scalar multiplication. Note that, even though the additions in axiom T1 are both denoted by the same symbol $+$, the addition on the left forming $\mathbf{v} + \mathbf{v}_1$ is carried out in V, whereas the addition $T(\mathbf{v}) + T(\mathbf{v}_1)$ is done in W. Similarly, the scalar multiplications $r\mathbf{v}$ and $rT(\mathbf{v})$ in axiom T2 refer to the spaces V and W, respectively.

Example 7.1.2. If V and W are vector spaces, the following are linear transformations:

Identity operator $V \to V$	$1_V : V \to V$	where $1_V(\mathbf{v}) = \mathbf{v}$ for all $\mathbf{v}$ in V
Zero transformation $V \to W$	$0 : V \to W$	where $0(\mathbf{v}) = \mathbf{0}$ for all $\mathbf{v}$ in V
Scalar operator $V \to V$	$a : V \to V$	where $a(\mathbf{v}) = a\mathbf{v}$ for all $\mathbf{v}$ in V
		(Here a is any real number.)

The symbol 0 will be used to denote the zero transformation from V to W for *any* spaces V and W.

Example 7.1.3. Show that the transposition and trace are linear transformations. More precisely,

$$
\begin{array}{ll}
R : \mathbf{M}_{mn} \to \mathbf{M}_{nm} & \text{where } R(A) = A^T \text{ for all } A \text{ in } \mathbf{M}_{mn} \\
S : \mathbf{M}_{mn} \to \mathbb{R} & \text{where } S(A) = \text{tr } A \text{ for all } A \text{ in } \mathbf{M}_{nn}
\end{array}
$$

are both linear transformations.

Example 7.1.4. If a is a scalar, define $E_a : \mathbf{P}_n \to \mathbb{R}$ by $E_a(p) = p(a)$ for each polynomial p in $\mathbf{P}_n$. Show that E_a is a linear transformation (called **evaluation** at a).

Theorem 7.1.5. *Let $T : V \to W$ be a linear transformation.*

(a) $T(\mathbf{0}) = \mathbf{0}$.

(b) $T(-\mathbf{v}) = -T(\mathbf{v})$ *for all* $\mathbf{v}$ *in* V.

(c) $T(r_1\mathbf{v}_1 + r_2\mathbf{v}_2 + \cdots + r_k\mathbf{v}_k) = r_1T(\mathbf{v}_1) + r_2T(\mathbf{v}_2) + \cdots + r_kT(\mathbf{v}_k)$ *for all* $\mathbf{v}_i$ *in* V *and all* r_i *in* $\mathbb{R}$.

Example 7.1.6. Let $T : V \to W$ be a linear transformation. If $T(\mathbf{v} - 3\mathbf{v}_1) = \mathbf{w}$ and $T(2\mathbf{v} - \mathbf{v}_1) = \mathbf{w}_1$, find $T(\mathbf{v})$ and $T(\mathbf{v}_1)$ in terms of $\mathbf{w}$ and $\mathbf{w}_1$.

As for functions in general, two linear transformations $T : V \to W$ and $S : V \to W$ are called **equal** (written $T = S$) if they have the same **action**; that is, if $T(\mathbf{v}) = S(\mathbf{v})$ for all $\mathbf{v}$ in V.

Theorem 7.1.7. *Let $T : V \to W$ and $S : V \to W$ be two linear transformations. Suppose that $V = span\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$. If $T(\mathbf{v}_i) = S(\mathbf{v}_i)$ for each i, then $T = S$.*

Example 7.1.8. Let $V = \text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. Let $T : V \to W$ be a linear transformation. If $T(\mathbf{v}_1) = \cdots = T(\mathbf{v}_n) = \mathbf{0}$, show that $T = 0$, the zero transformation from V to W.

Theorem 7.1.9. *Let V and W be vector spaces and let $\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ be a basis of V. Given any vectors $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n$ in W (they need not be distinct), there exists a unique linear transformation $T : V \to W$ satisfying $T(\mathbf{b}_i) = \mathbf{w}_i$ for each $i = 1, 2, \ldots, n$. In fact, the action of T is as follows:*

Given $\mathbf{v} = v_1\mathbf{b}_1 + v_2\mathbf{b}_2 + \cdots + v_n\mathbf{b}_n$ in V, v_i in $\mathbb{R}$, then

$$T(\mathbf{v}) = T(v_1\mathbf{b}_1 + v_2\mathbf{b}_2 + \cdots + v_n\mathbf{b}_n) = v_1\mathbf{w}_1 + v_2\mathbf{w}_2 + \cdots + v_n\mathbf{w}_n.$$

Example 7.1.10. Find a linear transformation $T : \mathbf{P}_2 \to \mathbf{M}_{22}$ such that

$$T(1+x) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad T(x+x^2) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \text{and} \quad T(1+x^2) = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

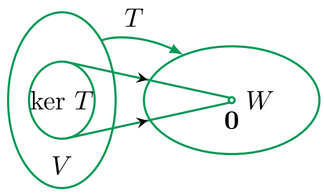
7.2 Kernel and Image of a Linear Transformation

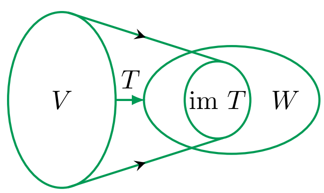
This section is devoted to two important subspaces associated with a linear transformation $T : V \to W$.

Definition 7.2.1. Kernel and Image of a Linear Transformation
The **kernel** of T (denoted ker T) and the **image** of T (denoted im T or $T(V)$) are defined by

$$\ker T = \{\mathbf{v} \text{ in } V \mid T(\mathbf{v}) = \mathbf{0}\}$$
$$\operatorname{im} T = \{T(\mathbf{v}) \mid \mathbf{v} \text{ in } V\} = T(V)$$

The kernel of T is often called the **nullspace** of T because it consists of all vectors $\mathbf{v}$ in V satisfying the *condition* that $T(\mathbf{v}) = \mathbf{0}$. The image of T is often called the **range** of T and consists of all vectors $\mathbf{w}$ in W of the *form* $\mathbf{w} = T(\mathbf{v})$ for some $\mathbf{v}$ in V. These subspaces are depicted in the diagrams below.





Example 7.2.2. Let $T_A : \mathbb{R}^n \to \mathbb{R}^m$ be the linear transformation induced by the $m \times n$ matrix A, that is $T_A(\mathbf{x}) = A\mathbf{x}$ for all columns $\mathbf{x}$ in $\mathbb{R}^n$. Then

$$\ker T_A = \{\mathbf{x} \mid A\mathbf{x} = \mathbf{0}\} = \operatorname{null} A \quad \text{and}$$
$$\operatorname{im} T_A = \{A\mathbf{x} \mid \mathbf{x} \text{ in } \mathbb{R}^n\} = \operatorname{im} A$$

Theorem 7.2.3. *Let $T : V \to W$ be a linear transformation.*

(a) *ker T is a subspace of V.*

(b) *im T is a subspace of W.*

Given a linear transformation $T : V \to W$:

dim(ker T) is called the **nullity** of T and denoted as nullity(T)
dim(im T) is called the **rank** of T and denoted as rank(T)

Example 7.2.4. Given an $m \times n$ matrix A, show that im $T_A =$ col A, so rank $T_A =$ rank A.

Example 7.2.5. Define a transformation $P : \mathbf{M}_{nn} \to \mathbf{M}_{nn}$ by $P(A) = A - A^T$ for all A in $\mathbf{M}_{nn}$. Show that P is linear and that:

a. ker P consists of all symmetric matrices.

b. im P consists of all skew-symmetric matrices.

One-to-One and Onto Transformations

Definition 7.2.6. One-to-one and Onto Linear Transformations
Let $T : V \to W$ be a linear transformation.

(a) T is said to be **onto** if im $T = W$.

(b) T is said to be **one-to-one** if $T(\mathbf{v}) = T(\mathbf{v}_1)$ implies $\mathbf{v} = \mathbf{v}_1$.

A vector $\mathbf{w}$ in W is said to be **hit** by T if $\mathbf{w} = T(\mathbf{v})$ for some $\mathbf{v}$ in V.

Theorem 7.2.7. *If $T : V \to W$ is a linear transformation, then T is one-to-one if and only if ker $T = \{\mathbf{0}\}$.*

Example 7.2.8. The identity transformation $1_V : V \to V$ is both one-to-one and onto for any vector space V.

Example 7.2.9. Consider the linear transformations

$$\begin{array}{ll} S : \mathbb{R}^3 \to \mathbb{R}^2 & \text{given by } S(x, y, z) = (x + y, x - y) \\ T : \mathbb{R}^2 \to \mathbb{R}^3 & \text{given by } T(x, y) = (x + y, x - y, x) \end{array}$$

Show that T is one-to-one but not onto, whereas S is onto but not one-to-one.

Example 7.2.10. Let U be an invertible $m \times m$ matrix and define

$$T : \mathbf{M}_{mn} \to \mathbf{M}_{mn} \quad \text{by} \quad T(X) = UX \text{ for all } X \text{ in } \mathbf{M}_{mn}$$

Show that T is a linear transformation that is both one-to-one and onto.

Theorem 7.2.11. *Let A be an $m \times n$ matrix, and let $T_A : \mathbb{R}^n \to \mathbb{R}^m$ be the linear transformation induced by A, that is $T_A(\mathbf{x}) = A\mathbf{x}$ for all columns $\mathbf{x}$ in $\mathbb{R}^n$.*

(a) *T_A is onto if and only if rank $A = m$.*

(b) *T_A is one-to-one if and only if rank $A = n$.*

The Dimension Theorem

Theorem 7.2.12. *Dimension Theorem*
Let $T : V \to W$ be any linear transformation and assume that ker T and im T are both finite dimensional. Then V is also finite dimensional and

$$\mathit{dim}\ V = \mathit{dim}(\mathit{ker}\ T) + \mathit{dim}(\mathit{im}\ T)$$

In other words, dim V = nullity(T) + rank(T).

Theorem 7.2.13. *Let $T : V \to W$ be a linear transformation, and let $\{\mathbf{e}_1, \dots, \mathbf{e}_r, \mathbf{e}_{r+1}, \dots, \mathbf{e}_n\}$ be a basis of V such that $\{\mathbf{e}_{r+1}, \dots, \mathbf{e}_n\}$ is a basis of ker T. Then $\{T(\mathbf{e}_1), \dots, T(\mathbf{e}_r)\}$ is a basis of im T, and hence r = rank T.*

The dimension theorem is one of the most useful results in all of linear algebra. It shows that if either dim(ker T) or dim(im T) can be found, then the other is automatically known. In many cases it is easier to compute one than the other, so the theorem is a real asset.

Example 7.2.14. Let A be an $m \times n$ matrix of rank r. Show that the space null A of all solutions of the system $A\mathbf{x} = \mathbf{0}$ of m homogeneous equations in n variables has dimension $n - r$.

Example 7.2.15. If $T : V \to W$ is a linear transformation where V is finite dimensional, then

$$\dim(\ker T) \leq \dim V \quad \text{and} \quad \dim(\text{im } T) \leq \dim V$$

Indeed, $\dim V = \dim(\ker T) + \dim(\text{im } T)$ by Dimension Theorem. Of course, the first inequality also follows because $\ker T$ is a subspace of V.

Example 7.2.16. Let $D : \mathbf{P}_n \to \mathbf{P}_{n-1}$ be the differentiation map defined by $D\,[p(x)] = p'(x)$. Compute $\ker D$ and hence conclude that D is onto.

Example 7.2.17. Given a in $\mathbb{R}$, the evaluation map $E_a : \mathbf{P}_n \to \mathbb{R}$ is given by $E_a\,[p(x)] = p(a)$. Show that E_a is linear and onto, and hence conclude that $\{(x-a), (x-a)^2, \ldots, (x-a)^n\}$ is a basis of ker E_a, the subspace of all polynomials $p(x)$ for which $p(a) = 0$.

Example 7.2.18. If A is any $m \times n$ matrix, show that rank $A =$ rank $A^T A =$ rank AA^T.

7.3 Isomorphisms and Composition

Often two vector spaces can consist of quite different types of vectors but, on closer examination, turn out to be the same underlying space displayed in different symbols. For example, consider the spaces

$$\mathbb{R}^2 = \{(a,b) \mid a,b \in \mathbb{R}\} \quad \text{and} \quad \mathbf{P}_1 = \{a + bx \mid a,b \in \mathbb{R}\}$$

Compare the addition and scalar multiplication in these spaces:

$$\begin{aligned} (a,b) + (a_1,b_1) &= (a+a_1, b+b_1) & \qquad (a+bx) + (a_1+b_1x) &= (a+a_1) + (b+b_1)x \\ r(a,b) &= (ra, rb) & r(a+bx) &= (ra) + (rb)x \end{aligned}$$

Clearly these are the *same* vector space expressed in different notation: if we change each (a,b) in $\mathbb{R}^2$ to $a+bx$, then $\mathbb{R}^2$ *becomes* $\mathbf{P}_1$, complete with addition and scalar multiplication. This can be expressed by noting that the map $(a,b) \mapsto a+bx$ is a linear transformation $\mathbb{R}^2 \to \mathbf{P}_1$ that is both one-to-one and onto.

Definition 7.3.1. Isomorphic Vector Spaces
A linear transformation $T : V \to W$ is called an **isomorphism** if it is both onto and one-to-one. The vector spaces V and W are said to be **isomorphic** if there exists an isomorphism $T : V \to W$, and we write $V \cong W$ when this is the case.

Example 7.3.2. The identity transformation $1_V : V \to V$ is an isomorphism for any vector space V.

Example 7.3.3. If $T : \mathbf{M}_{mn} \to \mathbf{M}_{nm}$ is defined by $T(A) = A^T$ for all A in $\mathbf{M}_{mn}$, then T is an isomorphism (verify). Hence $\mathbf{M}_{mn} \cong \mathbf{M}_{nm}$.

Example 7.3.4. Isomorphic spaces can "look" quite different. For example, $\mathbf{M}_{22} \cong \mathbf{P}_3$ because the map $T : \mathbf{M}_{22} \to \mathbf{P}_3$ given by $T\begin{bmatrix} a & b \\ c & d \end{bmatrix} = a + bx + cx^2 + dx^3$ is an isomorphism (verify).

The word *isomorphism* comes from two Greek roots: *iso*, meaning "same," and *morphos*, meaning "form." An isomorphism $T : V \to W$ induces a pairing

$$\mathbf{v} \leftrightarrow T(\mathbf{v})$$

between vectors $\mathbf{v}$ in V and vectors $T(\mathbf{v})$ in W that preserves vector addition and scalar multiplication. Hence, *as far as their vector space properties are concerned*, the spaces V and W are identical except for notation. Because addition and scalar multiplication in either space are completely determined by the same operations in the other space, all *vector space* properties of either space are completely determined by those of the other.

Theorem 7.3.5. *If V and W are finite dimensional spaces, the following conditions are equivalent for a linear transformation $T : V \to W$.*

(a) *T is an isomorphism.*

(b) *If $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is any basis of V, then $\{T(\mathbf{e}_1), T(\mathbf{e}_2), \ldots, T(\mathbf{e}_n)\}$ is a basis of W.*

(c) *There exists a basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ of V such that $\{T(\mathbf{e}_1), T(\mathbf{e}_2), \ldots, T(\mathbf{e}_n)\}$ is a basis of W.*

Theorem 7.3.6. *If V and W are finite dimensional vector spaces, then $V \cong W$ if and only if dim V = dim W.*

Corollary 7.3.7. *Let U, V, and W denote vector spaces. Then:*

(a) *$V \cong V$ for every vector space V.*

(b) *If $V \cong W$ then $W \cong V$.*

(c) *If $U \cong V$ and $V \cong W$, then $U \cong W$.*

Corollary 7.3.8. *If V is a vector space and dim $V = n$, then V is isomorphic to $\mathbb{R}^n$.*

If V is a vector space of dimension n, note that there are important explicit isomorphisms $V \to \mathbb{R}^n$. Fix a basis $B = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ of V and write $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ for the standard basis of $\mathbb{R}^n$. By Theorem **??** there is a unique linear transformation $C_B : V \to \mathbb{R}^n$ given by

$$C_B(v_1\mathbf{b}_1 + v_2\mathbf{b}_2 + \cdots + v_n\mathbf{b}_n) = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + \cdots + v_n\mathbf{e}_n = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

where each v_i is in $\mathbb{R}$. Moreover, $C_B(\mathbf{b}_i) = \mathbf{e}_i$ for each i so C_B is an isomorphism by Theorem 7.3.5, called the **coordinate isomorphism** corresponding to the basis B.

Example 7.3.9. Let V denote the space of all 2×2 symmetric matrices. Find an isomorphism $T : \mathbf{P}_2 \to V$ such that $T(1) = I$, where I is the 2×2 identity matrix.

Theorem 7.3.10. *If V and W have the same dimension n, a linear transformation $T : V \to W$ is an isomorphism if it is either one-to-one or onto.*

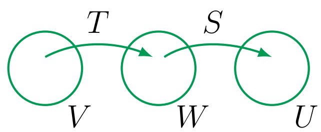
Composition

Definition 7.3.11. Composition of Linear Transformations

Given linear transformations $V \xrightarrow{T} W \xrightarrow{S} U$, the **composite** $ST : V \to U$ of T and S is defined by

$$ST(\mathbf{v}) = S\,[T(\mathbf{v})] \quad \text{for all } \mathbf{v} \text{ in } V$$

The operation of forming the new function ST is called **composition** (sometimes denoted $S \circ T$).



Example 7.3.12. Define: $S : \mathbf{M}_{22} \to \mathbf{M}_{22}$ and $T : \mathbf{M}_{22} \to \mathbf{M}_{22}$ by $S\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} c & d \\ a & b \end{bmatrix}$ and $T(A) = A^T$ for $A \in \mathbf{M}_{22}$. Describe the action of ST and TS, and show that $ST \neq TS$.

Theorem 7.3.13. *Let* $V \xrightarrow{T} W \xrightarrow{S} U \xrightarrow{R} Z$ *be linear transformations.*

(a) *The composite* ST *is again a linear transformation.*

(b) $T1_V = T$ *and* $1_W T = T$.

(c) $(RS)T = R(ST)$.

Theorem 7.3.14. *Let V and W be finite dimensional vector spaces. The following conditions are equivalent for a linear transformation $T : V \to W$.*

(a) *T is an isomorphism.*

(b) *There exists a linear transformation $S : W \to V$ such that $ST = 1_V$ and $TS = 1_W$.*

Moreover, in this case S is also an isomorphism and is uniquely determined by T:

$$\textit{If } \mathbf{w} \textit{ in } W \textit{ is written as } \mathbf{w} = T(\mathbf{v}), \textit{ then } S(\mathbf{w}) = \mathbf{v}.$$

Given an isomorphism $T : V \to W$, the unique isomorphism $S : W \to V$ satisfying condition (b) of the previous is called the **inverse** of T and is denoted by T^{-1}. Hence $T : V \to W$ and $T^{-1} : W \to V$ are related by the **fundamental identities**:

$$T^{-1}\left[T(\mathbf{v})\right] = \mathbf{v} \text{ for all } \mathbf{v} \text{ in } V \quad \text{and} \quad T\left[T^{-1}(\mathbf{w})\right] = \mathbf{w} \text{ for all } \mathbf{w} \text{ in } W$$

In other words, each of T and T^{-1} reverses the action of the other.

Example 7.3.15. Define $T : \mathbf{P}_1 \to \mathbf{P}_1$ by $T(a + bx) = (a - b) + ax$. Show that T has an inverse, and find the action of T^{-1}.

Example 7.3.16. If $B = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ is a basis of a vector space V, the coordinate transformation $C_B : V \to \mathbb{R}^n$ is an isomorphism defined by

$$C_B(v_1\mathbf{b}_1 + v_2\mathbf{b}_2 + \cdots + v_n\mathbf{b}_n) = (v_1, v_2, \ldots, v_n)^T$$

The way to reverse the action of C_B is clear: $C_B^{-1} : \mathbb{R}^n \to V$ is given by

$$C_B^{-1}(v_1, v_2, \ldots, v_n) = v_1\mathbf{b}_1 + v_2\mathbf{b}_2 + \cdots + v_n\mathbf{b}_n \quad \text{for all } v_i \text{ in } V$$

Example 7.3.17. Define $T : \mathbb{R}^3 \to \mathbb{R}^3$ by $T(x, y, z) = (z, x, y)$. Show that $T^3 = 1_{\mathbb{R}^3}$, and hence find T^{-1}.

Example 7.3.18. Define $T : \mathbf{P}_n \to \mathbb{R}^{n+1}$ by $T(p) = (p(0), p(1), \ldots, p(n))$ for all p in $\mathbf{P}_n$. Show that T^{-1} exists.

CHAPTER 8

Orthogonality

8.1 Orthogonal Complements and Projections

If $\{\mathbf{v}_1, \dots, \mathbf{v}_m\}$ is linearly independent in a general vector space, and if $\mathbf{v}_{m+1}$ is not in span$\{\mathbf{v}_1, \dots, \mathbf{v}_m\}$, then $\{\mathbf{v}_1, \dots, \mathbf{v}_m, \mathbf{v}_{m+1}\}$ is independent. Here is the analog for *orthogonal* sets in $\mathbb{R}^n$.

Lemma 8.1.1. *Orthogonal Lemma*
Let $\{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_m\}$ *be an orthogonal set in* $\mathbb{R}^n$. *Given* $\mathbf{x}$ *in* $\mathbb{R}^n$, *write*

$$\mathbf{f}_{m+1} = \mathbf{x} - \frac{\mathbf{x} \cdot \mathbf{f}_1}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 - \frac{\mathbf{x} \cdot \mathbf{f}_2}{\|\mathbf{f}_2\|^2}\mathbf{f}_2 - \cdots - \frac{\mathbf{x} \cdot \mathbf{f}_m}{\|\mathbf{f}_m\|^2}\mathbf{f}_m$$

Then:

(a) $\mathbf{f}_{m+1} \cdot \mathbf{f}_k = 0$ *for* $k = 1, 2, \dots, m$.

(b) *If* $\mathbf{x}$ *is not in span*$\{\mathbf{f}_1, \dots, \mathbf{f}_m\}$, *then* $\mathbf{f}_{m+1} \neq \mathbf{0}$ *and* $\{\mathbf{f}_1, \dots, \mathbf{f}_m, \mathbf{f}_{m+1}\}$ *is an orthogonal set.*

The orthogonal lemma has three important consequences for $\mathbb{R}^n$. The first is an extension for orthogonal sets of the fundamental fact that any independent set is part of a basis.

Theorem 8.1.2. *Let* U *be a subspace of* $\mathbb{R}^n$.

(a) *Every orthogonal subset* $\{\mathbf{f}_1, \dots, \mathbf{f}_m\}$ *in* U *is a subset of an orthogonal basis of* U.

(b) U *has an orthogonal basis.*

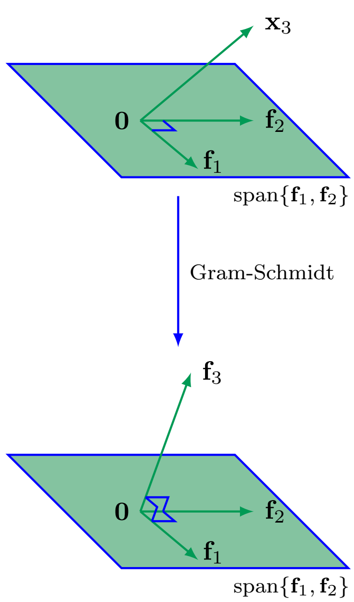
Theorem 8.1.3. *Gram-Schmidt Orthogonalization Algorithm*
If $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$ *is any basis of a subspace* U *of* $\mathbb{R}^n$, *construct* $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_m$ *in* U *successively as follows:*

$$\begin{aligned}
\mathbf{f}_1 &= \mathbf{x}_1 \\
\mathbf{f}_2 &= \mathbf{x}_2 - \tfrac{\mathbf{x}_2 \cdot \mathbf{f}_1}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 \\
\mathbf{f}_3 &= \mathbf{x}_3 - \tfrac{\mathbf{x}_3 \cdot \mathbf{f}_1}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 - \tfrac{\mathbf{x}_3 \cdot \mathbf{f}_2}{\|\mathbf{f}_2\|^2}\mathbf{f}_2 \\
&\vdots \\
\mathbf{f}_k &= \mathbf{x}_k - \tfrac{\mathbf{x}_k \cdot \mathbf{f}_1}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 - \tfrac{\mathbf{x}_k \cdot \mathbf{f}_2}{\|\mathbf{f}_2\|^2}\mathbf{f}_2 - \cdots - \tfrac{\mathbf{x}_k \cdot \mathbf{f}_{k-1}}{\|\mathbf{f}_{k-1}\|^2}\mathbf{f}_{k-1}
\end{aligned}$$

for each $k = 2, 3, \ldots, m$. *Then*

(a) $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_m\}$ *is an orthogonal basis of* U.

(b) *span*$\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_k\}$ = *span*$\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ *for each* $k = 1, 2, \ldots, m$.

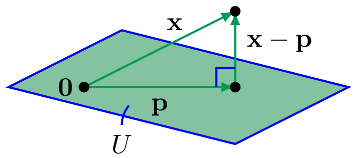


Remark
Observe that the vector $\frac{\mathbf{x} \cdot \mathbf{f}_i}{\|\mathbf{f}_i\|^2}\mathbf{f}_i$ is unchanged if a nonzero scalar multiple of $\mathbf{f}_i$ is used in place of $\mathbf{f}_i$. Hence, if a newly constructed $\mathbf{f}_i$ is multiplied by a nonzero scalar at some stage of the Gram-Schmidt algorithm, the subsequent $\mathbf{f}$s will be unchanged. This is useful in actual calculations.

Example 8.1.4. Find an orthogonal basis of the row space of $A = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 3 & 2 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$.

Projections

Suppose a point $\mathbf{x}$ and a plane U through the origin in $\mathbb{R}^3$ are given, and we want to find the point $\mathbf{p}$ in the plane that is closest to $\mathbf{x}$. Our geometric intuition assures us that such a point $\mathbf{p}$ exists. In fact (see the diagram), $\mathbf{p}$ must be chosen in such a way that $\mathbf{x}-\mathbf{p}$ is *perpendicular* to the plane.



Now we make two observations: first, the plane U is a *subspace of* $\mathbb{R}^3$ (because U contains the origin); and second, that the condition that $\mathbf{x}-\mathbf{p}$ is perpendicular to the plane U means that $\mathbf{x}-\mathbf{p}$ is *orthogonal* to every vector in U. In these terms the whole discussion makes sense in $\mathbb{R}^n$. Furthermore, the orthogonal lemma provides exactly what is needed to find $\mathbf{p}$ in this more general setting.

Definition 8.1.5. Orthogonal Complement of a Subspace of $\mathbb{R}^n$
If U is a subspace of $\mathbb{R}^n$, define the **orthogonal complement** $U^\perp$ of U (pronounced "U-perp") by

$$U^\perp = \{\mathbf{x} \text{ in } \mathbb{R}^n \mid \mathbf{x}\cdot\mathbf{y} = 0 \text{ for all } \mathbf{y} \text{ in } U\}$$

Lemma 8.1.6. *Let U be a subspace of $\mathbb{R}^n$.*

(a) *$U^\perp$ is a subspace of $\mathbb{R}^n$.*

(b) *$\{\mathbf{0}\}^\perp = \mathbb{R}^n$ and $(\mathbb{R}^n)^\perp = \{\mathbf{0}\}$.*

(c) *If $U = \text{span}\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$, then $U^\perp = \{\mathbf{x} \text{ in } \mathbb{R}^n \mid \mathbf{x}\cdot\mathbf{x}_i = 0 \text{ for } i = 1, 2, \ldots, k\}$.*

Example 8.1.7. Find $U^\perp$ if $U = \text{span}\{(1,-1,2,0), (1,0,-2,3)\}$ in $\mathbb{R}^4$.

Definition 8.1.8. Projection onto a Subspace of $\mathbb{R}^n$

Let U be a subspace of $\mathbb{R}^n$ with orthogonal basis $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_m\}$. If $\mathbf{x}$ is in $\mathbb{R}^n$, the vector

$$\text{proj}_U \mathbf{x} = \frac{\mathbf{x} \cdot \mathbf{f}_1}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 + \frac{\mathbf{x} \cdot \mathbf{f}_2}{\|\mathbf{f}_2\|^2}\mathbf{f}_2 + \cdots + \frac{\mathbf{x} \cdot \mathbf{f}_m}{\|\mathbf{f}_m\|^2}\mathbf{f}_m$$

is called the **orthogonal projection** of $\mathbf{x}$ on U. For the zero subspace $U = \{\mathbf{0}\}$, we define

$$\text{proj}_{\{\mathbf{0}\}} \mathbf{x} = \mathbf{0}$$

Theorem 8.1.9. *Projection Theorem*

If U is a subspace of $\mathbb{R}^n$ and $\mathbf{x}$ is in $\mathbb{R}^n$, write $\mathbf{p} = proj_U \mathbf{x}$. Then:

(a) $\mathbf{p}$ *is in U and* $\mathbf{x} - \mathbf{p}$ *is in* $U^\perp$.

(b) $\mathbf{p}$ *is the vector in U closest to* $\mathbf{x}$ *in the sense that*

$$\|\mathbf{x} - \mathbf{p}\| < \|\mathbf{x} - \mathbf{y}\| \quad \textit{for all } \mathbf{y} \in U, \mathbf{y} \neq \mathbf{p}$$

Example 8.1.10. Let $U = \text{span}\{\mathbf{x}_1, \mathbf{x}_2\}$ in $\mathbb{R}^4$ where $\mathbf{x}_1 = (1, 1, 0, 1)$ and $\mathbf{x}_2 = (0, 1, 1, 2)$. If $\mathbf{x} = (3, -1, 0, 2)$, find the vector in U closest to $\mathbf{x}$ and express $\mathbf{x}$ as the sum of a vector in U and a vector orthogonal to U.

Example 8.1.11. Find the point in the plane with equation $2x + y - z = 0$ that is closest to the point $(2, -1, -3)$.

Theorem 8.1.12. *Let U be a fixed subspace of $\mathbb{R}^n$. If we define $T : \mathbb{R}^n \to \mathbb{R}^n$ by*

$$T(\mathbf{x}) = proj_U \mathbf{x} \quad \textit{for all } \mathbf{x} \textit{ in } \mathbb{R}^n$$

(a) *T is a linear operator.*

(b) *$imT = U$ and $kerT = U^\perp$.*

(c) *$dimU + dimU^\perp = n$.*

8.2 Orthogonal Diagonalization

Recall that an $n \times n$ matrix A is diagonalizable if and only if it has n linearly independent eigenvectors. Moreover, the matrix P with these eigenvectors as columns is a diagonalizing matrix for A, that is

$$P^{-1}AP \text{ is diagonal.}$$

Theorem 8.2.1. *The following conditions are equivalent for an $n \times n$ matrix P.*

(1) P is invertible and $P^{-1} = P^T$.

(2) The rows of P are orthonormal.

(3) The columns of P are orthonormal.

Definition 8.2.2. Orthogonal Matrices
An $n \times n$ matrix P is called an **orthogonal matrix** if it satisfies one (and hence all) of the conditions in Theorem 8.2.1.

Example 8.2.3. The rotation matrix $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ is orthogonal for any angle θ.

Example 8.2.4. The matrix $\begin{bmatrix} 2 & 1 & 1 \\ -1 & 1 & 1 \\ 0 & -1 & 1 \end{bmatrix}$ has orthogonal rows but the columns are not orthogonal. However, if the rows are normalized, the resulting matrix $\begin{bmatrix} \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{-1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$ is orthogonal (so the columns are now orthonormal).

Example 8.2.5. If P and Q are orthogonal matrices, then PQ is also orthogonal, as is $P^{-1} = P^T$.

[0]In view of (2) and (3) of Theorem 8.2.1, *orthonormal matrix* might be a better name. But *orthogonal matrix* is standard.

Definition 8.2.6. Orthogonally Diagonalizable Matrices
An $n \times n$ matrix A is said to be **orthogonally diagonalizable** when an orthogonal matrix P can be found such that $P^{-1}AP = P^TAP$ is diagonal.

Theorem 8.2.7. *Principal Axes Theorem*
The following conditions are equivalent for an $n \times n$ matrix A.

(1) A has an orthonormal set of n eigenvectors.

(2) A is orthogonally diagonalizable.

(3) A is symmetric.

A set of orthonormal eigenvectors of a symmetric matrix A is called a set of **principal axes** for A. Because the eigenvalues of a (real) symmetric matrix are real, Theorem 8.2.7 is also called the **real spectral theorem**, and the set of distinct eigenvalues is called the **spectrum** of the matrix.

Example 8.2.8. Find an orthogonal matrix P such that $P^{-1}AP$ is diagonal, where $A = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ -1 & 2 & 5 \end{bmatrix}$.

Theorem 8.2.9. *If A is an $n \times n$ symmetric matrix, then*

$$(A\mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (A\mathbf{y})$$

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

Proof. Recall that $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T\mathbf{y}$ for all columns $\mathbf{x}$ and $\mathbf{y}$. Because $A^T = A$, we get

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

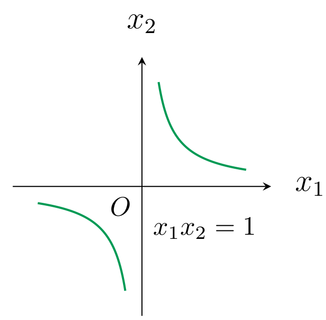
□

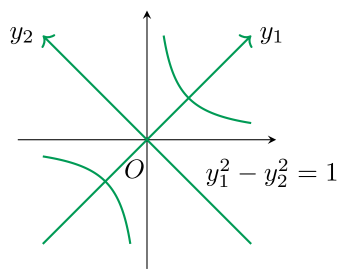
Theorem 8.2.10. *If A is a symmetric matrix, then eigenvectors of A corresponding to distinct eigenvalues are orthogonal.*

Example 8.2.11. Orthogonally diagonalize the symmetric matrix $A = \begin{bmatrix} 8 & -2 & 2 \\ -2 & 5 & 4 \\ 2 & 4 & 5 \end{bmatrix}$.

If A is symmetric and a set of orthogonal eigenvectors of A is given, the eigenvectors are called principal axes of A. The name comes from geometry. An expression $q = ax_1^2 + bx_1x_2 + cx_2^2$ is called a **quadratic form** in the variables x_1 and x_2, and the graph of the equation $q = 1$ is called a **conic** in these variables. For example, if $q = x_1x_2$, the graph of $q = 1$ is given in the first diagram.

But if we introduce new variables y_1 and y_2 by setting $x_1 = y_1 + y_2$ and $x_2 = y_1 - y_2$, then q becomes $q = y_1^2 - y_2^2$, a diagonal form with no cross term y_1y_2 (see the second diagram). Because of this, the y_1 and y_2 axes are called the principal axes for the conic (hence the name). Orthogonal diagonalization provides a systematic method for finding principal axes.





Example 8.2.12. Find principal axes for the quadratic form $q = x_1^2 - 4x_1x_2 + x_2^2$.

Theorem 8.2.13. *Triangulation Theorem*
If A is an $n \times n$ matrix with n real eigenvalues, an orthogonal matrix P exists such that $P^T AP$ is upper triangular.

Corollary 8.2.14. *If A is an $n \times n$ matrix with real eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ (possibly not all distinct), then $detA = \lambda_1\lambda_2 \ldots \lambda_n$ and $trA = \lambda_1 + \lambda_2 + \cdots + \lambda_n$.*

8.3 Positive Definite Matrices

All the eigenvalues of any symmetric matrix are real; this section is about the case in which the eigenvalues are positive.

Definition 8.3.1. Positive Definite Matrices
A square matrix is called **positive definite** if it is symmetric and all its eigenvalues λ are positive, that is $\lambda > 0$.

Because these matrices are symmetric, the principal axes theorem plays a central role in the theory.

Theorem 8.3.2. *If A is positive definite, then it is invertible and $detA > 0$.*

If $\mathbf{x}$ is a column in $\mathbb{R}^n$ and A is any real $n \times n$ matrix, we view the 1×1 matrix $\mathbf{x}^T A\mathbf{x}$ as a real number. With this convention, we have the following characterization of positive definite matrices.

Theorem 8.3.3. *A symmetric matrix A is positive definite if and only if $\mathbf{x}^T A\mathbf{x} > 0$ for every column $\mathbf{x} \neq \mathbf{0}$ in $\mathbb{R}^n$.*

Example 8.3.4. If U is any invertible $n \times n$ matrix, show that $A = U^T U$ is positive definite.

It is remarkable that the converse to Example 8.3.4 is also true. In fact every positive definite matrix A can be factored as $A = U^T U$ where U is an upper triangular matrix with positive elements on the main diagonal.

If A is any $n \times n$ matrix, let ${}^{(r)}A$ denote the $r \times r$ submatrix in the upper left corner of A; that is, ${}^{(r)}A$ is the matrix obtained from A by deleting the last $n - r$ rows and columns. The matrices ${}^{(1)}A, {}^{(2)}A, {}^{(3)}A, \ldots, {}^{(n)}A = A$ are called the **principal submatrices** of A.

Example 8.3.5. If $A = \begin{bmatrix} 10 & 5 & 2 \\ 5 & 3 & 2 \\ 2 & 2 & 3 \end{bmatrix}$ then ${}^{(1)}A = [10]$, ${}^{(2)}A = \begin{bmatrix} 10 & 5 \\ 5 & 3 \end{bmatrix}$ and ${}^{(3)}A = A$.

Lemma 8.3.6. *If A is positive definite, so is each principal submatrix ${}^{(r)}A$ for $r = 1, 2, \ldots, n$.*

If A is positive definite, Lemma 8.3.6 and Theorem 8.3.2 show that $\det({}^{(r)}A) > 0$ for every r.

Theorem 8.3.7. *The following conditions are equivalent for a symmetric $n \times n$ matrix A:*

(1) A is positive definite.

(2) $\det({}^{(r)}A) > 0$ for each $r = 1, 2, \ldots, n$.

(3) $A = U^T U$ where U is an upper triangular matrix with positive entries on the main diagonal.

The factorization in (3) is unique (called the ***Cholesky factorization*** *of A).*

Theorem 8.3.8. *Algorithm for the Cholesky Factorization*
If A is a positive definite matrix, the Cholesky factorization $A = U^T U$ can be obtained as follows:

Step 1. Carry A to an upper triangular matrix U_1 with positive diagonal entries using row operations each of which adds a multiple of a row to a lower row.

Step 2. Obtain U from U_1 by dividing each row of U_1 by the square root of the diagonal entry in that row.

Example 8.3.9. Find the Cholesky factorization of $A = \begin{bmatrix} 10 & 5 & 2 \\ 5 & 3 & 2 \\ 2 & 2 & 3 \end{bmatrix}$.

8.7 Complex Matrices

If A is an $n \times n$ matrix, the characteristic polynomial $c_A(x)$ is a polynomial of degree n and the eigenvalues of A are just the roots of $c_A(x)$. In most of our examples these roots have been *real* numbers (in fact, the examples have been carefully chosen so this will be the case!); but it need not happen, even when the characteristic polynomial has real coefficients. For example, if $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ then $c_A(x) = x^2 + 1$ has roots i and $-i$, where i is a complex number satisfying $i^2 = -1$. Therefore, we have to deal with the possibility that the eigenvalues of a (real) square matrix might be complex numbers.

The set of complex numbers is denoted $\mathbb{C}$. If $n \geq 1$, we denote the set of all n-tuples of complex numbers by $\mathbb{C}^n$. As with $\mathbb{R}^n$, these n-tuples will be written either as row or column matrices and will be referred to as **vectors**. We define vector operations on $\mathbb{C}^n$ as follows:

$$\begin{aligned} (v_1, v_2, \ldots, v_n) + (w_1, w_2, \ldots, w_n) &= (v_1 + w_1, v_2 + w_2, \ldots, v_n + w_n) \\ u(v_1, v_2, \ldots, v_n) &= (uv_1, uv_2, \ldots, uv_n) \quad \text{for } u \text{ in } \mathbb{C} \end{aligned}$$

With these definitions, $\mathbb{C}^n$ satisfies the axioms for a vector space (with complex scalars) given in Chapter 6. Thus we can speak of spanning sets for $\mathbb{C}^n$, of linearly independent subsets, and of bases. In all cases, the definitions are identical to the real case, except that the scalars are allowed to be complex numbers. In particular, the standard basis of $\mathbb{R}^n$ remains a basis of $\mathbb{C}^n$, called the **standard basis** of $\mathbb{C}^n$.

A matrix $A = [a_{ij}]$ is called a **complex matrix** if every entry a_{ij} is a complex number. If $z = a + bi$ is a complex number, the **conjugate** of z is the complex number, denoted $\overline{z}$, given by

$$\overline{z} = a - bi \quad \text{where } z = a + bi$$

Hence $\overline{z}$ is obtained from z by negating the imaginary part. The notion of conjugation for complex numbers extends to matrices as follows: Define the **conjugate** of $A = [a_{ij}]$ to be the matrix

$$\overline{A} = \left[\ \overline{a}_{ij}\ \right]$$

obtained from A by conjugating every entry.

$$\overline{A + B} = \overline{A} + \overline{B} \quad \text{and} \quad \overline{AB} = \overline{A}\,\overline{B}$$

holds for all (complex) matrices of appropriate size.

The Standard Inner Product

There is a natural generalization to $\mathbb{C}^n$ of the dot product in $\mathbb{R}^n$.

Definition 8.7.1. Standard Inner Product in $\mathbb{R}^n$
Given $\mathbf{z} = (z_1, z_2, \ldots, z_n)$ and $\mathbf{w} = (w_1, w_2, \ldots, w_n)$ in $\mathbb{C}^n$, define their **standard inner product** $\langle \mathbf{z}, \mathbf{w} \rangle$ by

$$\langle \mathbf{z}, \mathbf{w} \rangle = z_1\overline{w}_1 + z_2\overline{w}_2 + \cdots + z_n\overline{w}_n = \mathbf{z} \cdot \overline{\mathbf{w}}$$

where $\overline{w}$ is the conjugate of the complex number w.

Example 8.7.2. If $\mathbf{z} = (2, 1-i, 2i, 3-i)$ and $\mathbf{w} = (1-i, -1, -i, 3+2i)$, then

$$\langle \mathbf{z}, \mathbf{w} \rangle = 2(1+i) + (1-i)(-1) + (2i)(i) + (3-i)(3-2i) = 6-6i$$
$$\langle \mathbf{z}, \mathbf{z} \rangle = 2 \cdot 2 + (1-i)(1+i) + (2i)(-2i) + (3-i)(3+i) = 20$$

Theorem 8.7.3. *Let* $\mathbf{z}$, $\mathbf{z}_1$, $\mathbf{w}$, *and* $\mathbf{w}_1$ *denote vectors in* $\mathbb{C}^n$, *and let* λ *denote a complex number.*

(1) $\langle \mathbf{z} + \mathbf{z}_1, \mathbf{w} \rangle = \langle \mathbf{z}, \mathbf{w} \rangle + \langle \mathbf{z}_1, \mathbf{w} \rangle$ *and* $\langle \mathbf{z}, \mathbf{w} + \mathbf{w}_1 \rangle = \langle \mathbf{z}, \mathbf{w} \rangle + \langle \mathbf{z}, \mathbf{w}_1 \rangle$.

(2) $\langle \lambda \mathbf{z}, \mathbf{w} \rangle = \lambda \langle \mathbf{z}, \mathbf{w} \rangle$ *and* $\langle \mathbf{z}, \lambda \mathbf{w} \rangle = \overline{\lambda} \langle \mathbf{z}, \mathbf{w} \rangle$.

(3) $\langle \mathbf{z}, \mathbf{w} \rangle = \overline{\langle \mathbf{w}, \mathbf{z} \rangle}$.

(4) $\langle \mathbf{z}, \mathbf{z} \rangle \geq 0$, *and* $\langle \mathbf{z}, \mathbf{z} \rangle = 0$ *if and only if* $\mathbf{z} = \mathbf{0}$.

Definition 8.7.4. Norm and Length in $\mathbb{C}^n$
As for the dot product on $\mathbb{R}^n$, property (4) enables us to define the **norm** or **length** $\|\mathbf{z}\|$ of a vector $\mathbf{z} = (z_1, z_2, \ldots, z_n)$ in $\mathbb{C}^n$:

$$\|\mathbf{z}\| = \sqrt{\langle \mathbf{z}, \mathbf{z} \rangle} = \sqrt{|z_1|^2 + |z_2|^2 + \cdots + |z_n|^2}$$

Theorem 8.7.5. *If* $\mathbf{z}$ *is any vector in* $\mathbb{C}^n$, *then*

(a) $\|\mathbf{z}\| \geq 0$ *and* $\|\mathbf{z}\| = 0$ *if and only if* $\mathbf{z} = \mathbf{0}$.

(b) $\|\lambda \mathbf{z}\| = |\lambda| \|\mathbf{z}\|$ *for all complex numbers* λ.

A vector $\mathbf{u}$ in $\mathbb{C}^n$ is called a **unit vector** if $\|\mathbf{u}\| = 1$.

Example 8.7.6. In $\mathbb{C}^4$, find a unit vector $\mathbf{u}$ that is a positive real multiple of $\mathbf{z} = (1-i, i, 2, 3+4i)$.

Definition 8.7.7. Conjugate Transpose in $\mathbb{C}^n$
The **conjugate transpose** A^H of a complex matrix A is defined by

$$A^H = (\overline{A})^T = \overline{(A^T)}$$

Observe that $A^H = A^T$ when A is real.[1]

Example 8.7.8.

$$\begin{bmatrix} 3 & 1-i & 2+i \\ 2i & 5+2i & -i \end{bmatrix}^H = \begin{bmatrix} 3 & -2i \\ 1+i & 5-2i \\ 2-i & i \end{bmatrix}$$

Theorem 8.7.9. *Let A and B denote complex matrices, and let λ be a complex number.*

(1) $(A^H)^H = A$.

(2) $(A+B)^H = A^H + B^H$.

(3) $(\lambda A)^H = \overline{\lambda} A^H$.

(4) $(AB)^H = B^H A^H$.

Hermitian and Unitary Matrices

Definition 8.7.10. Hermitian Matrices
A square complex matrix A is called **hermitian** if $A^H = A$, equivalently if $\overline{A} = A^T$.

Hermitian matrices are easy to recognize because the entries on the main diagonal must be real, and the "reflection" of each nondiagonal entry in the main diagonal must be the conjugate of that entry.

Example 8.7.11. $\begin{bmatrix} 3 & i & 2+i \\ -i & -2 & -7 \\ 2-i & -7 & 1 \end{bmatrix}$ is hermitian, whereas $\begin{bmatrix} 1 & i \\ i & -2 \end{bmatrix}$ and $\begin{bmatrix} 1 & i \\ -i & i \end{bmatrix}$ are not.

Theorem 8.7.12. *An $n \times n$ complex matrix A is hermitian if and only if*

$$\langle A\mathbf{z}, \mathbf{w} \rangle = \langle \mathbf{z}, A\mathbf{w} \rangle$$

for all n-tuples $\mathbf{z}$ and $\mathbf{w}$ in $\mathbb{C}^n$.

Theorem 8.7.13. *Let A denote a hermitian matrix.*

(a) *The eigenvalues of A are real.*

(b) *Eigenvectors of A corresponding to distinct eigenvalues are orthogonal.*

[1] Other notations for A^H are A^* and $A^\dagger$.

Definition 8.7.14. Orthogonal and Orthonormal Vectors in $\mathbb{C}^n$
As in the real case, a set of nonzero vectors $\{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_m\}$ in $\mathbb{C}^n$ is called **orthogonal** if $\langle \mathbf{z}_i, \mathbf{z}_j \rangle = 0$ whenever $i \neq j$, and it is **orthonormal** if, in addition, $\|\mathbf{z}_i\| = 1$ for each i.

Theorem 8.7.15. *The following are equivalent for an $n \times n$ complex matrix A.*

(a) *A is invertible and $A^{-1} = A^H$.*

(b) *The rows of A are an orthonormal set in $\mathbb{C}^n$.*

(c) *The columns of A are an orthonormal set in $\mathbb{C}^n$.*

Definition 8.7.16. Unitary Matrices
A square complex matrix U is called **unitary** if $U^{-1} = U^H$.

Thus a real matrix is unitary if and only if it is orthogonal.

Example 8.7.17. The matrix $A = \begin{bmatrix} 1+i & 1 \\ 1-i & i \end{bmatrix}$ has orthogonal columns, but the rows are not orthogonal. Normalizing the columns gives the unitary matrix $\frac{1}{2}\begin{bmatrix} 1+i & \sqrt{2} \\ 1-i & \sqrt{2}i \end{bmatrix}$.

Example 8.7.18. Consider the hermitian matrix $A = \begin{bmatrix} 3 & 2+i \\ 2-i & 7 \end{bmatrix}$. Find the eigenvalues of A, find two orthonormal eigenvectors, and so find a unitary matrix U such that $U^H AU$ is diagonal.

Unitary Diagonalization

An $n \times n$ complex matrix A is called **unitarily diagonalizable** if $U^H AU$ is diagonal for some unitary matrix U.

Theorem 8.7.19. *Schur's Theorem*
If A is any $n \times n$ complex matrix, there exists a unitary matrix U such that

$$U^H AU = T$$

is upper triangular. Moreover, the entries on the main diagonal of T are the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of A (including multiplicities).

Corollary 8.7.20. *Let A be an $n \times n$ complex matrix, and let $\lambda_1, \lambda_2, \ldots, \lambda_n$ denote the eigenvalues of A, including multiplicities. Then*

$$\det A = \lambda_1 \lambda_2 \cdots \lambda_n \quad \text{and} \quad \operatorname{tr} A = \lambda_1 + \lambda_2 + \cdots + \lambda_n$$

Schur's theorem asserts that every complex matrix can be "unitarily triangularized." However, we cannot substitute "unitarily diagonalized" here. In fact, if $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, there is no invertible complex matrix U at all such that $U^{-1}AU$ is diagonal. However, the situation is much better for hermitian matrices.

Theorem 8.7.21. *Spectral Theorem*
If A is hermitian, there is a unitary matrix U such that $U^H AU$ is diagonal.

Example 8.7.22. Show that the non-hermitian matrix $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ is unitarily diagonalizable.

An $n \times n$ complex matrix N is called **normal** if $NN^H = N^H N$. It is clear that every hermitian or unitary matrix is normal, as is the matrix $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$.

Theorem 8.7.23. *An $n \times n$ complex matrix A is unitarily diagonalizable if and only if A is normal.*

Recall that the characteristic polynomial of a square matrix A is defined by $c_A(x) = \det(xI - A)$, and that the eigenvalues of A are just the roots of $c_A(x)$.

Theorem 8.7.24. *Cayley-Hamilton Theorem*
If A is an $n \times n$ complex matrix, then $c_A(A) = 0$; that is, A is a root of its characteristic polynomial.

8.9 An Application to Quadratic Forms

An expression like $x_1^2 + x_2^2 + x_3^2 - 2x_1x_3 + x_2x_3$ is called a quadratic form in the variables x_1, x_2, and x_3. In this section we show that new variables y_1, y_2, and y_3 can always be found so that the quadratic form, when expressed in terms of the new variables, has no cross terms y_1y_2, y_1y_3, or y_2y_3.

Definition 8.9.1. Quadratic Form
A **quadratic form** q in the n variables $x_1, x_2, \ldots, x_n$ is a linear combination of terms $x_1^2, x_2^2, \ldots, x_n^2$, and cross terms $x_1x_2, x_1x_3, x_2x_3, \ldots$.

Example 8.9.2. Write $q = x_1^2 + 3x_3^2 + 2x_1x_2 - x_1x_3$ in the form $q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$, where A is a symmetric 3×3 matrix.

Theorem 8.9.3. *Diagonalization Theorem*
Let $q = \mathbf{x}^T A\mathbf{x}$ be a quadratic form in the variables $x_1, x_2, \dots, x_n$, where $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ and A is a symmetric $n \times n$ matrix. Let P be an orthogonal matrix such that $P^T AP$ is diagonal, and define new variables $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$ by

$$\mathbf{x} = P\mathbf{y} \quad \textit{equivalently} \quad \mathbf{y} = P^T\mathbf{x}$$

If q is expressed in terms of these new variables $y_1, y_2, \dots, y_n$, the result is

$$q = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$$

where $\lambda_1, \lambda_2, \dots, \lambda_n$ are the eigenvalues of A repeated according to their multiplicities.

Example 8.9.4. Find new variables y_1, y_2, y_3, and y_4 such that

$$q = 3(x_1^2 + x_2^2 + x_3^2 + x_4^2) + 2x_1x_2 - 10x_1x_3 + 10x_1x_4 + 10x_2x_3 - 10x_2x_4 + 2x_3x_4$$

has diagonal form, and find the corresponding principal axes.

Congruence

Theorem 8.9.5. *If $q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ is a quadratic form given by a symmetric matrix A, then A is uniquely determined by q.*

Two $n \times n$ matrices A and B are called **congruent**, written $A \overset{c}{\sim} B$, if $B = U^T AU$ for some invertible matrix U. Here are some properties of congruence:

(1) *$A \overset{c}{\sim} A$ for all A.*

(2) *If $A \overset{c}{\sim} B$, then $B \overset{c}{\sim} A$.*

(3) *If $A \overset{c}{\sim} B$ and $B \overset{c}{\sim} C$, then $A \overset{c}{\sim} C$.*

(4) *If $A \overset{c}{\sim} B$, then A is symmetric if and only if B is symmetric.*

(5) *If $A \overset{c}{\sim} B$, then* $\text{rank}A = \text{rank}B$.

Example 8.9.6. The symmetric matrices $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ have the same rank but are not congruent. Indeed, if $A \overset{c}{\sim} B$, an invertible matrix U exists such that $B = U^T AU = U^T U$. But then $-1 = \det B = (\det U)^2$, a contradiction.

Theorem 8.9.7. *Sylvester's Law of Inertia*
If $A \overset{c}{\sim} B$, then A and B have the same number of positive eigenvalues, counting multiplicities.

The **index** of a symmetric matrix A is the number of positive eigenvalues of A. If $q = q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ is a quadratic form, the **index** and **rank** of q are defined to be, respectively, the index and rank of the matrix A.

Now let $q = q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ be any quadratic form in n variables, of index k and rank r, where A is symmetric. We claim that new variables $\mathbf{z}$ can be found so that q is **completely diagonalized**—that is,

$$q(\mathbf{z}) = z_1^2 + \cdots + z_k^2 - z_{k+1}^2 - \cdots - z_r^2$$

If $k \leq r \leq n$, let $D_n(k, r)$ denote the $n \times n$ diagonal matrix whose main diagonal consists of k ones, followed by $r - k$ minus ones, followed by $n - r$ zeros. Then we seek new variables $\mathbf{z}$ such that

$$q(\mathbf{z}) = \mathbf{z}^T D_n(k, r)\mathbf{z}$$

To determine $\mathbf{z}$, first diagonalize A as follows: Find an orthogonal matrix P_0 such that

$$P_0^T AP_0 = D = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r, 0, \ldots, 0)$$

is diagonal with the nonzero eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_r$ of A on the main diagonal (followed by $n-r$ zeros). By reordering the columns of P_0, if necessary, we may assume that $\lambda_1, \ldots, \lambda_k$

are positive and $\lambda_{k+1}, \ldots, \lambda_r$ are negative. This being the case, let D_0 be the $n \times n$ diagonal matrix

$$D_0 = \operatorname{diag}\left(\frac{1}{\sqrt{\lambda_1}}, \ldots, \frac{1}{\sqrt{\lambda_k}}, \frac{1}{\sqrt{-\lambda_{k+1}}}, \ldots, \frac{1}{\sqrt{-\lambda_r}}, 1, \ldots, 1\right)$$

Then $D_0^T D D_0 = D_n(k, r)$, so if new variables $\mathbf{z}$ are given by $\mathbf{x} = (P_0 D_0)\mathbf{z}$, we obtain

$$q(\mathbf{z}) = \mathbf{z}^T D_n(k, r)\mathbf{z} = z_1^2 + \cdots + z_k^2 - z_{k+1}^2 - \cdots - z_r^2$$

as required. Note that the change-of-variables matrix $P_0 D_0$ from $\mathbf{z}$ to $\mathbf{x}$ has orthogonal columns (in fact, scalar multiples of the columns of P_0).

Example 8.9.8. Completely diagonalize the quadratic form q in Example 8.9.4 and find the index and rank.

Theorem 8.9.9. *Let A and B be symmetric $n \times n$ matrices, and let $0 \leq k \leq r \leq n$.*

(1) A has index k and rank r if and only if $A \overset{c}{\sim} D_n(k, r)$.

(2) $A \overset{c}{\sim} B$ if and only if they have the same rank and index.

CHAPTER 9

Change of Basis

9.1 The Matrix of a Linear Transformation

Let $T : V \to W$ be a linear transformation where $\dim V = n$ and $\dim W = m$. The aim in this section is to describe the action of T as multiplication by an $m \times n$ matrix A. The idea is to convert a vector $\mathbf{v}$ in V into a column in $\mathbb{R}^n$, multiply that column by A to get a column in $\mathbb{R}^m$, and convert this column back to get $T(\mathbf{v})$ in W.

Converting vectors to columns is a simple matter, but one small change is needed. Up to now the *order* of the vectors in a basis has been of no importance. However, in this section, we shall speak of an **ordered basis** $\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$, which is just a basis where the order in which the vectors are listed is taken into account. Hence $\{\mathbf{b}_2, \mathbf{b}_1, \mathbf{b}_3\}$ is a different *ordered* basis from $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$.

If $B = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ is an ordered basis in a vector space V, and if

$$\mathbf{v} = v_1\mathbf{b}_1 + v_2\mathbf{b}_2 + \cdots + v_n\mathbf{b}_n, \quad v_i \in \mathbb{R}$$

is a vector in V, then the (uniquely determined) numbers $v_1, v_2, \ldots, v_n$ are called the **coordinates** of $\mathbf{v}$ with respect to the basis B.

Definition 9.1.1. Coordinate Vector $C_B(\mathbf{v})$ of $\mathbf{v}$ for a basis B
The **coordinate vector** of $\mathbf{v}$ with respect to B is defined to be

$$C_B(\mathbf{v}) = (v_1\mathbf{b}_1 + v_2\mathbf{b}_2 + \cdots + v_n\mathbf{b}_n) = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

Note that $C_B(\mathbf{b}_i) = \mathbf{e}_i$ is column i of I_n.

Example 9.1.2. The coordinate vector for $\mathbf{v} = (2, 1, 3)$ with respect to the ordered basis $B = \{(1, 1, 0), (1, 0, 1), (0, 1, 1)\}$ of $\mathbb{R}^3$ is $C_B(\mathbf{v}) = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}$ because

$$\mathbf{v} = (2, 1, 3) = 0(1, 1, 0) + 2(1, 0, 1) + 1(0, 1, 1)$$

Theorem 9.1.3. *If V has dimension n and $B = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$ is any ordered basis of V, the coordinate transformation $C_B : V \to \mathbb{R}^n$ is an isomorphism. In fact, $C_B^{-1} : \mathbb{R}^n \to V$ is given by*

$$C_B^{-1}\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = v_1\mathbf{b}_1 + v_2\mathbf{b}_2 + \cdots + v_n\mathbf{b}_n \quad \textit{for all} \quad \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \textit{ in } \mathbb{R}^n.$$

Definition 9.1.4. Matrix $M_{DB}(T)$ of $T : V \to W$ for bases D and B
This is called the **matrix of T corresponding to the ordered bases B and D**, and we use the following notation:

$$M_{DB}(T) = \begin{bmatrix} C_D[T(\mathbf{b}_1)] & C_D[T(\mathbf{b}_2)] & \cdots & C_D[T(\mathbf{b}_n)] \end{bmatrix}$$

Theorem 9.1.5. *Let $T : V \to W$ be a linear transformation where dim $V = n$ and dim $W = m$, and let $B = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ and D be ordered bases of V and W, respectively. Then the matrix $M_{DB}(T)$ just given is the unique $m \times n$ matrix A that satisfies*

$$C_D T = T_A C_B$$

Hence the defining property of $M_{DB}(T)$ is

$$C_D[T(\mathbf{v})] = M_{DB}(T)C_B(\mathbf{v}) \textit{ for all } \mathbf{v} \textit{ in } V$$

The matrix $M_{DB}(T)$ is given in terms of its columns by

$$M_{DB}(T) = \begin{bmatrix} C_D[T(\mathbf{b}_1)] & C_D[T(\mathbf{b}_2)] & \cdots & C_D[T(\mathbf{b}_n)] \end{bmatrix}$$

Example 9.1.6. Define $T : \mathbf{P}_2 \to \mathbb{R}^2$ by $T(a+bx+cx^2) = (a+c, b-a-c)$ for all polynomials $a + bx + cx^2$. If $B = \{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ and $D = \{\mathbf{d}_1, \mathbf{d}_2\}$ where

$$\mathbf{b}_1 = 1, \mathbf{b}_2 = x, \mathbf{b}_3 = x^2 \quad \text{and} \quad \mathbf{d}_1 = (1, 0), \mathbf{d}_2 = (0, 1)$$

compute $M_{DB}(T)$ and verify Theorem 9.1.5.

Example 9.1.7. Suppose $T : \mathbf{M}_{22}(\mathbb{R}) \to \mathbb{R}^3$ is linear with matrix $M_{DB}(T) = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$ where

$$B = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\} \text{ and } D = \{(1,0,0),(0,1,0),(0,0,1)\}$$

Compute $T(\mathbf{v})$ where $\mathbf{v} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

Example 9.1.8. Let A be an $m \times n$ matrix, and let $T_A : \mathbb{R}^n \to \mathbb{R}^m$ be the matrix transformation induced by $A : T_A(\mathbf{x}) = A\mathbf{x}$ for all columns $\mathbf{x}$ in $\mathbb{R}^n$. If B and D are the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively (ordered as usual), then

$$M_{DB}(T_A) = A$$

In other words, the matrix of T_A corresponding to the standard bases is A itself.

Example 9.1.9. Let V and W have ordered bases B and D, respectively. Let dim $V = n$.

(a) The identity transformation $1_V : V \to V$ has matrix $M_{BB}(1_V) = I_n$.

(b) The zero transformation $0 : V \to W$ has matrix $M_{DB}(0) = 0$.

Theorem 9.1.10. *Let $V \xrightarrow{T} W \xrightarrow{S} U$ be linear transformations and let B, D, and E be finite ordered bases of V, W, and U, respectively. Then*

$$M_{EB}(ST) = M_{ED}(S) \cdot M_{DB}(T)$$

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[rr, bend right, "ST"'] & W \arrow[r, "S"] & U \end{tikzcd}$$

Theorem 9.1.11. *Let $T : V \to W$ be a linear transformation, where dim $V =$ dim $W = n$. The following are equivalent.*

(a) *T is an isomorphism.*

(b) *$M_{DB}(T)$ is invertible for all ordered bases B and D of V and W.*

(c) *$M_{DB}(T)$ is invertible for some pair of ordered bases B and D of V and W.*

When this is the case, $[M_{DB}(T)]^{-1} = M_{BD}(T^{-1})$.

The rank of a linear transformation $T : V \to W$ is rank $T = \dim(\text{im } T)$. Moreover, if A is any $m \times n$ matrix and $T_A : \mathbb{R}^n \to \mathbb{R}^m$ is the matrix transformation, we showed that $\text{rank}(T_A) = \text{rank } A$. So it may not be surprising that rank T equals the rank of any matrix of T.

Theorem 9.1.12. *Let $T : V \to W$ be a linear transformation where dim $V = n$ and dim $W = m$. If B and D are any ordered bases of V and W, then rank $T =$ rank$[M_{DB}(T)]$.*

Example 9.1.13. Define $T : \mathbf{P}_2 \to \mathbb{R}^3$ by $T(a + bx + cx^2) = (a - 2b, 3c - 2a, 3c - 4b)$ for a, b, $c \in \mathbb{R}$. Compute rank T.

Example 9.1.14. Let $T : V \to W$ be a linear transformation where $\dim V = n$ and $\dim W = m$. Choose an ordered basis $B = \{\mathbf{b}_1, \ldots, \mathbf{b}_r, \mathbf{b}_{r+1}, \ldots, \mathbf{b}_n\}$ of V in which $\{\mathbf{b}_{r+1}, \ldots, \mathbf{b}_n\}$ is a basis of $\ker T$, possibly empty. Then $\{T(\mathbf{b}_1), \ldots, T(\mathbf{b}_r)\}$ is a basis of $\operatorname{im} T$, so extend it to an ordered basis $D = \{T(\mathbf{b}_1), \ldots, T(\mathbf{b}_r), \mathbf{f}_{r+1}, \ldots, \mathbf{f}_m\}$ of W. Because $T(\mathbf{b}_{r+1}) = \cdots = T(\mathbf{b}_n) = \mathbf{0}$, we have

$$M_{DB}(T) = \begin{bmatrix} C_D[T(\mathbf{b}_1)] & \cdots & C_D[T(\mathbf{b}_r)] & C_D[T(\mathbf{b}_{r+1})] & \cdots & C_D[T(\mathbf{b}_n)] \end{bmatrix} = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$$

Incidentally, this shows that $\operatorname{rank} T = r$ by Theorem 9.1.12.

9.2 Operators and Similarity

The central problem of linear algebra is to understand the structure of a linear transformation $T : V \to V$ from a space V to itself. Such transformations are called **linear operators**. If $T : V \to V$ is a linear operator where $\dim(V) = n$, it is possible to choose bases B and D of V such that the matrix $M_{DB}(T)$ has a very simple form: $M_{DB}(T) = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$ where $r = \operatorname{rank} T$. Consequently, only the rank of T is revealed by determining the simplest matrices $M_{DB}(T)$ of T where the bases B and D can be chosen arbitrarily. But if we insist that $B = D$ and look for bases B such that $M_{BB}(T)$ is as simple as possible, we learn a great deal about the operator T.

The B-matrix of an Operator

Definition 9.2.1. Matrix $M_{DB}(T)$ of $T : V \to W$ for basis B
If $T : V \to V$ is an operator on a vector space V, and if B is an ordered basis of V, define $M_B(T) = M_{BB}(T)$ and call this the ***B*-matrix** of T.

Recall that if $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear operator and $E = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is the standard basis of $\mathbb{R}^n$, then $C_E(\mathbf{x}) = \mathbf{x}$ for every $\mathbf{x} \in \mathbb{R}^n$, so $M_E(T) = [T(\mathbf{e}_1), T(\mathbf{e}_2), \ldots, T(\mathbf{e}_n)]$ is the **standard matrix** of the operator T.

Theorem 9.2.2. *Let $T : V \to V$ be an operator where $\dim V = n$, and let B be an ordered basis of V.*

(1) $C_B(T(\mathbf{v})) = M_B(T)C_B(\mathbf{v})$ for all $\mathbf{v}$ in V.

(2) If $S : V \to V$ is another operator on V, then $M_B(ST) = M_B(S)M_B(T)$.

(3) T is an isomorphism if and only if $M_B(T)$ is invertible. In this case $M_D(T)$ is invertible for every ordered basis D of V.

(4) If T is an isomorphism, then $M_B(T^{-1}) = [M_B(T)]^{-1}$.

(5) If $B = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$, then $M_B(T) = \begin{bmatrix} C_B[T(\mathbf{b}_1)] & C_B[T(\mathbf{b}_2)] & \cdots & C_B[T(\mathbf{b}_n)] \end{bmatrix}$.

Definition 9.2.3. Change Matrix $P_{D \leftarrow B}$ for bases B and D
With this in mind, define the **change matrix** $P_{D \leftarrow B}$ by

$$P_{D \leftarrow B} = M_{DB}(1_V) \quad \text{for any ordered bases } B \text{ and } D \text{ of } V$$

Theorem 9.2.4. *Let $B = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$ and D denote ordered bases of a vector space V. Then the change matrix $P_{D\leftarrow B}$ is given in terms of its columns by*

$$P_{D\leftarrow B} = \left[\begin{array}{cccc} C_D(\mathbf{b}_1) & C_D(\mathbf{b}_2) & \cdots & C_D(\mathbf{b}_n) \end{array}\right] \tag{9.1}$$

and has the property that

$$C_D(\mathbf{v}) = P_{D\leftarrow B} C_B(\mathbf{v}) \text{ for all } \mathbf{v} \text{ in } V \tag{9.2}$$

Moreover, if E is another ordered basis of V, we have

(1) $P_{B\leftarrow B} = I_n$

(2) $P_{D\leftarrow B}$ *is invertible and* $(P_{D\leftarrow B})^{-1} = P_{B\leftarrow D}$

(3) $P_{E\leftarrow D} P_{D\leftarrow B} = P_{E\leftarrow B}$

Example 9.2.5. In $\mathbf{P}_2$ find $P_{D\leftarrow B}$ if $B = \{1, x, x^2\}$ and $D = \{1, (1-x), (1-x)^2\}$. Then use this to express $p = p(x) = a + bx + cx^2$ as a polynomial in powers of $(1-x)$.

Theorem 9.2.6. *Similarity Theorem*
Let B_0 and B be two ordered bases of a finite dimensional vector space V. If $T : V \to V$ is any linear operator, the matrices $M_B(T)$ and $M_{B_0}(T)$ of T with respect to these bases are similar. More precisely,

$$M_B(T) = P^{-1}M_{B_0}(T)P$$

where $P = P_{B_0 \leftarrow B}$ is the change matrix from B to B_0.

Example 9.2.7. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be defined by $T(a, b, c) = (2a - b, b + c, c - 3a)$. If B_0 denotes the standard basis of $\mathbb{R}^3$ and $B = \{(1, 1, 0), (1, 0, 1), (0, 1, 0)\}$, find an invertible matrix P such that $P^{-1}M_{B_0}(T)P = M_B(T)$.

Theorem 9.2.8. *Let A be an $n \times n$ matrix and let E be the standard basis of $\mathbb{R}^n$.*

(1) Let A' be similar to A, say $A' = P^{-1}AP$, and let B be the ordered basis of $\mathbb{R}^n$ consisting of the columns of P in order. Then $T_A : \mathbb{R}^n \to \mathbb{R}^n$ is linear and

$$M_E(T_A) = A \quad \text{and} \quad M_B(T_A) = A'$$

(2) If B is any ordered basis of $\mathbb{R}^n$, let P be the (invertible) matrix whose columns are the vectors in B in order. Then

$$M_B(T_A) = P^{-1}AP$$

Example 9.2.9. Given $A = \begin{bmatrix} 10 & 6 \\ -18 & -11 \end{bmatrix}$, $P = \begin{bmatrix} 2 & -1 \\ -3 & 2 \end{bmatrix}$, and $D = \begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix}$, verify that $P^{-1}AP = D$ and use this fact to find a basis B of $\mathbb{R}^2$ such that $M_B(T_A) = D$.

Example 9.2.10. (a) If $T : V \to V$ is an operator where V is finite dimensional, show that $TST = T$ for some invertible operator $S : V \to V$.

(b) If A is an $n \times n$ matrix, show that $AUA = A$ for some invertible matrix U.

A property of $n \times n$ matrices is called a **similarity invariant** if, whenever a given $n \times n$ matrix A has the property, every matrix similar to A also has the property.

If $T : V \to V$ is a linear operator on a finite dimensional space V, define the **determinant** of T (denoted $\det T$) by

$$\det T = \det M_B(T), \quad B \text{ any basis of } V$$

This is independent of the choice of basis B because, if D is any other basis of V, the matrices $M_B(T)$ and $M_D(T)$ are similar and so have the same determinant. In the same way, the **trace** of T (denoted $\operatorname{tr} T$) can be defined by

$$\operatorname{tr} T = \operatorname{tr} M_B(T), \quad B \text{ any basis of } V$$

Example 9.2.11. Let S and T denote linear operators on the finite dimensional space V. Show that

$$\det (ST) = \det S \det T$$

If $T : V \to V$ is a linear operator on the finite dimensional space V, define the **characteristic polynomial** of T by

$$c_T(x) = c_A(x) \text{ where } A = M_B(T),\ B \text{ any basis of } V$$

In other words, the characteristic polynomial of an operator T is the characteristic polynomial of *any* matrix representing T.

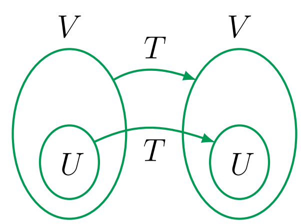
Example 9.2.12. Compute the characteristic polynomial $c_T(x)$ of the operator $T : \mathbf{P}_2 \to \mathbf{P}_2$ given by $T(a + bx + cx^2) = (b + c) + (a + c)x + (a + b)x^2$.

9.3 Invariant Subspaces and Direct Sums

A fundamental question in linear algebra is the following: If $T : V \to V$ is a linear operator, how can a basis B of V be chosen so the matrix $M_B(T)$ is as simple as possible? A basic technique for answering such questions will be explained in this section. If U is a subspace of V, write its image under T as

$$T(U) = \{T(\mathbf{u}) \mid \mathbf{u} \text{ in } U\}$$

Definition 9.3.1. T-invariant Subspace
Let $T : V \to V$ be an operator. A subspace $U \subseteq V$ is called ***T*-invariant** if $T(U) \subseteq U$, that is, $T(\mathbf{u}) \in U$ for every vector $\mathbf{u} \in U$. Hence T is a linear operator on the vector space U.



Example 9.3.2. Let $T : V \to V$ be any linear operator. Then:

(a) $\{\mathbf{0}\}$ and V are T-invariant subspaces.

(b) Both ker T and im $T = T(V)$ are T-invariant subspaces.

(c) If U and W are T-invariant subspaces, so are $T(U)$, $U \cap W$, and $U + W$.

Example 9.3.3. Define $T : \mathbb{R}^3 \to \mathbb{R}^3$ by $T(a, b, c) = (3a + 2b, b - c, 4a + 2b - c)$. Then $U = \{(a, b, a) \mid a, b \text{ in } \mathbb{R}\}$ is T-invariant because

$$T(a, b, a) = (3a + 2b, b - a, 3a + 2b)$$

is in U for all a and b (the first and last entries are equal).

Example 9.3.4. Let $T : V \to V$ be a linear operator, and suppose that $U = \text{span}\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ is a subspace of V. Show that U is T-invariant if and only if $T(\mathbf{u}_i)$ lies in U for each $i = 1, 2, \dots, k$.

Example 9.3.5. Define $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T(a,b) = (b,-a)$. Show that $\mathbb{R}^2$ contains no T-invariant subspace except 0 and $\mathbb{R}^2$.

Definition 9.3.6. Restriction of an Operator
Let $T : V \to V$ be a linear operator. If U is any T-invariant subspace of V, then

$$T : U \to U$$

is a linear operator on the subspace U, called the **restriction** of T to U.

Theorem 9.3.7. *Let $T : V \to V$ be a linear operator where V has dimension n and suppose that U is any T-invariant subspace of V. Let $B_1 = \{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$ be any basis of U and extend it to a basis $B = \{\mathbf{b}_1, \ldots, \mathbf{b}_k, \mathbf{b}_{k+1}, \ldots, \mathbf{b}_n\}$ of V in any way. Then $M_B(T)$ has the block triangular form*

$$M_B(T) = \begin{bmatrix} M_{B_1}(T) & Y \\ 0 & Z \end{bmatrix}$$

where Z is $(n-k) \times (n-k)$ and $M_{B_1}(T)$ is the matrix of the restriction of T to U.

Theorem 9.3.8. *Let A be a block upper triangular matrix, say*

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} & \cdots & A_{1n} \\ 0 & A_{22} & A_{23} & \cdots & A_{2n} \\ 0 & 0 & A_{33} & \cdots & A_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & A_{nn} \end{bmatrix}$$

where the diagonal blocks are square. Then:

(1) $\det A = (\det A_{11})(\det A_{22})(\det A_{33}) \cdots (\det A_{nn})$.

(2) $c_A(x) = c_{A_{11}}(x) c_{A_{22}}(x) c_{A_{33}}(x) \cdots c_{A_{nn}}(x)$.

Example 9.3.9. Consider the linear operator $T : \mathbf{P}_2 \to \mathbf{P}_2$ given by

$$T(a + bx + cx^2) = (-2a - b + 2c) + (a + b)x + (-6a - 2b + 5c)x^2$$

Show that $U = \operatorname{span}\{x, 1+2x^2\}$ is T-invariant, use it to find a block upper triangular matrix for T, and use that to compute $c_T(x)$.

Eigenvalues

Let $T : V \to V$ be a linear operator. A one-dimensional subspace $\mathbb{R}\mathbf{v}$, $\mathbf{v} \neq \mathbf{0}$, is T-invariant if and only if $T(r\mathbf{v}) = rT(\mathbf{v})$ lies in $\mathbb{R}\mathbf{v}$ for all r in $\mathbb{R}$. This holds if and only if $T(\mathbf{v})$ lies in $\mathbb{R}\mathbf{v}$; that is, $T(\mathbf{v}) = \lambda\mathbf{v}$ for some λ in $\mathbb{R}$. A real number λ is called an **eigenvalue** of an operator $T : V \to V$ if

$$T(\mathbf{v}) = \lambda\mathbf{v}$$

holds for some nonzero vector $\mathbf{v}$ in V. In this case, $\mathbf{v}$ is called an **eigenvector** of T corresponding to λ. The subspace

$$E_\lambda(T) = \{\mathbf{v} \text{ in } V \mid T(\mathbf{v}) = \lambda\mathbf{v}\}$$

is called the **eigenspace** of T corresponding to λ. If A is an $n \times n$ matrix, a real number λ is an eigenvalue of the matrix operator $T_A : \mathbb{R}^n \to \mathbb{R}^n$ if and only if λ is an eigenvalue of the matrix A. Moreover, the eigenspaces agree:

$$E_\lambda(T_A) = \{\mathbf{x} \text{ in } \mathbb{R}^n \mid A\mathbf{x} = \lambda\mathbf{x}\} = E_\lambda(A)$$

Theorem 9.3.10. *Let $T : V \to V$ be a linear operator where $\dim V = n$, let B denote any ordered basis of V, and let $C_B : V \to \mathbb{R}^n$ denote the coordinate isomorphism. Then:*

(1) The eigenvalues λ of T are precisely the eigenvalues of the matrix $M_B(T)$ and thus are the roots of the characteristic polynomial $c_T(x)$.

(2) In this case the eigenspaces $E_\lambda(T)$ and $E_\lambda[M_B(T)]$ are isomorphic via the restriction $C_B : E_\lambda(T) \to E_\lambda[M_B(T)]$.

Example 9.3.11. Find the eigenvalues and eigenspaces for $T : \mathbf{P}_2 \to \mathbf{P}_2$ given by

$$T(a + bx + cx^2) = (2a + b + c) + (2a + b - 2c)x - (a + 2c)x^2$$

Theorem 9.3.12. *Each eigenspace of a linear operator $T : V \to V$ is a T-invariant subspace of V.*

Direct Sums

If U and W are subspaces of V, their *sum* $U + W$ and their *intersection* $U \cap W$ are:

$$U + W = \{\mathbf{u} + \mathbf{w} \mid \mathbf{u} \text{ in } U \text{ and } \mathbf{w} \text{ in } W\}$$
$$U \cap W = \{\mathbf{v} \mid \mathbf{v} \text{ lies in both } U \text{ and } W\}$$

Definition 9.3.13. Direct Sum of Subspaces
A vector space V is said to be the **direct sum** of subspaces U and W if

$$U \cap W = \{\mathbf{0}\} \quad \text{and} \quad U + W = V$$

In this case we write $V = U \oplus W$. Given a subspace U, any subspace W such that $V = U \oplus W$ is called a **complement** of U in V.

Example 9.3.14. In the space $\mathbb{R}^5$, consider the subspaces $U = \{(a, b, c, 0, 0) \mid a, b, \text{ and } c \text{ in } \mathbb{R}\}$ and $W = \{(0, 0, 0, d, e) \mid d \text{ and } e \text{ in } \mathbb{R}\}$. Show that $\mathbb{R}^5 = U \oplus W$.

Example 9.3.15. If U is a subspace of $\mathbb{R}^n$, show that $\mathbb{R}^n = U \oplus U^{\perp}$.

Example 9.3.16. Let $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ be a basis of a vector space V, and partition it into two parts: $\{\mathbf{e}_1, \ldots, \mathbf{e}_k\}$ and $\{\mathbf{e}_{k+1}, \ldots, \mathbf{e}_n\}$. If $U = \text{span}\{\mathbf{e}_1, \ldots, \mathbf{e}_k\}$ and $W = \text{span}\{\mathbf{e}_{k+1}, \ldots, \mathbf{e}_n\}$, show that $V = U \oplus W$.

Theorem 9.3.17. *Let U and W be subspaces of a finite dimensional vector space V. The following three conditions are equivalent:*

(1) $V = U \oplus W$.

(2) Each vector $\mathbf{v}$ in V can be written uniquely in the form

$$\mathbf{v} = \mathbf{u} + \mathbf{w} \qquad \mathbf{u} \text{ in } U, \mathbf{w} \text{ in } W$$

(3) If $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ and $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ are bases of U and W, respectively, then $B = \{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{w}_1, \ldots, \mathbf{w}_m\}$ is a basis of V.

(The uniqueness in (2) means that if $\mathbf{v} = \mathbf{u}_1 + \mathbf{w}_1$ is another such representation, then $\mathbf{u}_1 = \mathbf{u}$ and $\mathbf{w}_1 = \mathbf{w}$.)

Theorem 9.3.18. *If a finite dimensional vector space V is the direct sum $V = U \oplus W$ of subspaces U and W, then*

$$\dim V = \dim U + \dim W$$

Theorem 9.3.19. *Let $T : V \to V$ be a linear operator where V has dimension n. Suppose $V = U_1 \oplus U_2$ where both U_1 and U_2 are T-invariant. If $B_1 = \{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$ and $B_2 = \{\mathbf{b}_{k+1}, \ldots, \mathbf{b}_n\}$ are bases of U_1 and U_2 respectively, then*

$$B = \{\mathbf{b}_1, \ldots, \mathbf{b}_k, \mathbf{b}_{k+1}, \ldots, \mathbf{b}_n\}$$

is a basis of V, and $M_B(T)$ has the block diagonal form

$$M_B(T) = \begin{bmatrix} M_{B_1}(T) & 0 \\ 0 & M_{B_2}(T) \end{bmatrix}$$

where $M_{B_1}(T)$ and $M_{B_2}(T)$ are the matrices of the restrictions of T to U_1 and to U_2 respectively.

Definition 9.3.20. Reducible Linear Operator
The linear operator $T : V \to V$ is said to be **reducible** if nonzero T-invariant subspaces U_1 and U_2 can be found such that $V = U_1 \oplus U_2$.

Example 9.3.21. Let $T : V \to V$ be a linear operator satisfying $T^2 = 1_V$ (such operators are called **involutions**). Define

$$U_1 = \{\mathbf{v} \mid T(\mathbf{v}) = \mathbf{v}\} \quad \text{and} \quad U_2 = \{\mathbf{v} \mid T(\mathbf{v}) = -\mathbf{v}\}$$

(a) Show that $V = U_1 \oplus U_2$.

(b) If $\dim V = n$, find a basis B of V such that $M_B(T) = \begin{bmatrix} I_k & 0 \\ 0 & -I_{n-k} \end{bmatrix}$ for some k.

(c) Conclude that, if A is an $n \times n$ matrix such that $A^2 = I$, then A is similar to $\begin{bmatrix} I_k & 0 \\ 0 & -I_{n-k} \end{bmatrix}$ for some k.

Example 9.3.22. Consider the operator $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by $T \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a+b \\ b \end{bmatrix}$. Show that $U_1 = \mathbb{R} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ is T-invariant but that U_1 has not T-invariant complement in $\mathbb{R}^2$.

CHAPTER 10

Inner Product Spaces

10.1 Inner Products and Norms

The plan in this chapter is to define an *inner product* on an arbitrary real vector space V (of which the dot product is an example in $\mathbb{R}^n$) and use it to introduce these concepts in V.

Definition 10.1.1. Inner Product Spaces
An **inner product** on a real vector space V is a function that assigns a real number $\langle \mathbf{v}, \mathbf{w} \rangle$ to every pair $\mathbf{v}$, $\mathbf{w}$ of vectors in V in such a way that the following axioms are satisfied.

P1. $\langle \mathbf{v}, \mathbf{w} \rangle$ *is a real number for all* $\mathbf{v}$ *and* $\mathbf{w}$ *in* V.

P2. $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$ *for all* $\mathbf{v}$ *and* $\mathbf{w}$ *in* V.

P3. $\langle \mathbf{v} + \mathbf{w}, \mathbf{u} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle + \langle \mathbf{w}, \mathbf{u} \rangle$ *for all* $\mathbf{u}$, $\mathbf{v}$, *and* $\mathbf{w}$ *in* V.

P4. $\langle r\mathbf{v}, \mathbf{w} \rangle = r\langle \mathbf{v}, \mathbf{w} \rangle$ *for all* $\mathbf{v}$ *and* $\mathbf{w}$ *in* V *and all* r *in* $\mathbb{R}$.

P5. $\langle \mathbf{v}, \mathbf{v} \rangle > 0$ *for all* $\mathbf{v} \neq \mathbf{0}$ *in* V.

A real vector space V with an inner product $\langle \ , \rangle$ will be called an **inner product space**. Note that every subspace of an inner product space is again an inner product space using the same inner product.

Example 10.1.2. $\mathbb{R}^n$ is an inner product space with the dot product as inner product:

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v} \cdot \mathbf{w} \quad \text{for all } \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$$

This is also called the **euclidean** inner product, and $\mathbb{R}^n$, equipped with the dot product, is called **euclidean** n**-space**.

Example 10.1.3. If A and B are $m \times n$ matrices, define $\langle A, B \rangle = \operatorname{tr}(AB^T)$ where $\operatorname{tr}(X)$ is the trace of the square matrix X. Show that $\langle\ ,\ \rangle$ is an inner product in $\mathbf{M}_{mn}$.

Theorem 10.1.4. *Let $\langle\ ,\ \rangle$ be an inner product on a space V; let $\mathbf{v}$, $\mathbf{u}$, and $\mathbf{w}$ denote vectors in V; and let r denote a real number.*

(a) $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$

(b) $\langle \mathbf{v}, r\mathbf{w} \rangle = r\langle \mathbf{v}, \mathbf{w} \rangle = \langle r\mathbf{v}, \mathbf{w} \rangle$

(c) $\langle \mathbf{v}, \mathbf{0} \rangle = 0 = \langle \mathbf{0}, \mathbf{v} \rangle$

(d) $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ *if and only if* $\mathbf{v} = \mathbf{0}$

Theorem 10.1.5. *If A is any $n \times n$ positive definite matrix, then*

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T A \mathbf{y} \textit{ for all columns } \mathbf{x}, \mathbf{y} \textit{ in } \mathbb{R}^n$$

defines an inner product on $\mathbb{R}^n$, and every inner product on $\mathbb{R}^n$ arises in this way.

Thus, just as every linear operator $\mathbb{R}^n \to \mathbb{R}^n$ corresponds to an $n \times n$ matrix, every inner product on $\mathbb{R}^n$ corresponds to a positive definite $n \times n$ matrix. In particular, the dot product corresponds to the identity matrix I_n.

Remark
If we refer to the inner product space $\mathbb{R}^n$ without specifying the inner product, we mean that the dot product is to be used.

Example 10.1.6. Let the inner product $\langle\ ,\rangle$ be defined on $\mathbb{R}^2$ by

$$\left\langle \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}, \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \right\rangle = 2v_1w_1 - v_1w_2 - v_2w_1 + v_2w_2$$

Find a symmetric 2×2 matrix A such that $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T A\mathbf{y}$ for all $\mathbf{x}$, $\mathbf{y}$ in $\mathbb{R}^2$.

Let $\langle\ ,\rangle$ be an inner product on $\mathbb{R}^n$ given as in Theorem 10.1.5 by a positive definite matrix A. If $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^T$, then $\langle \mathbf{x}, \mathbf{x} \rangle = \mathbf{x}^T A\mathbf{x}$ is an expression in the variables $x_1, x_2, \ldots, x_n$ called a **quadratic form** as we have seen before in Chapter 8.

Norm and Distance

Definition 10.1.7. Norm and Distance
As in $\mathbb{R}^n$, if $\langle\ ,\ \rangle$ is an inner product on a space V, the **norm** $\|\mathbf{v}\|$ of a vector $\mathbf{v}$ in V is defined by

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v}\rangle}$$

We define the **distance** between vectors $\mathbf{v}$ and $\mathbf{w}$ in an inner product space V to be

$$\mathrm{d}(\mathbf{v}, \mathbf{w}) = \|\mathbf{v} - \mathbf{w}\|$$

Note that axiom P5 guarantees that $\langle \mathbf{v}, \mathbf{v}\rangle \geq 0$, so $\|\mathbf{v}\|$ is a real number.

Example 10.1.8. Show that $\langle \mathbf{u}+\mathbf{v}, \mathbf{u}-\mathbf{v}\rangle = \|\mathbf{u}\|^2 - \|\mathbf{v}\|^2$ in any inner product space.

A vector $\mathbf{v}$ in an inner product space V is called a **unit vector** if $\|\mathbf{v}\| = 1$. The set of all unit vectors in V is called the **unit ball** in V. For example, if $V = \mathbb{R}^2$ (with the dot product) and $\mathbf{v} = (x, y)$, then

$$\|\mathbf{v}\|^2 = 1 \quad \text{if and only if} \quad x^2 + y^2 = 1$$

Hence the unit ball in $\mathbb{R}^2$ is the **unit circle** $x^2 + y^2 = 1$ with centre at the origin and radius 1. However, the shape of the unit ball varies with the choice of inner product.

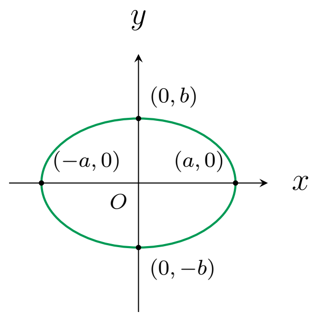
Example 10.1.9. Let $a > 0$ and $b > 0$. If $\mathbf{v} = (x, y)$ and $\mathbf{w} = (x_1, y_1)$, define an inner product on $\mathbb{R}^2$ by

$$\langle \mathbf{v}, \mathbf{w} \rangle = \frac{xx_1}{a^2} + \frac{yy_1}{b^2}$$

In this case

$$\|\mathbf{v}\|^2 = 1 \quad \text{if and only if} \quad \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

so the unit ball is the ellipse shown in the diagram.



Theorem 10.1.10. *030480 If $\mathbf{v} \neq \mathbf{0}$ is any vector in an inner product space V, then $\frac{1}{\|\mathbf{v}\|}\mathbf{v}$ is the unique unit vector that is a positive multiple of $\mathbf{v}$.*

Theorem 10.1.11. *Cauchy-Schwarz Inequality*
If $\mathbf{v}$ and $\mathbf{w}$ are two vectors in an inner product space V, then

$$\langle \mathbf{v}, \mathbf{w} \rangle^2 \leq \|\mathbf{v}\|^2 \|\mathbf{w}\|^2$$

Moreover, equality occurs if and only if one of $\mathbf{v}$ and $\mathbf{w}$ is a scalar multiple of the other.

Theorem 10.1.12. *If V is an inner product space, the norm $\|\cdot\|$ has the following properties.*

(a) $\|\mathbf{v}\| \geq 0$ *for every vector $\mathbf{v}$ in V.*

(b) $\|\mathbf{v}\| = 0$ *if and only if $\mathbf{v} = \mathbf{0}$.*

(c) $\|r\mathbf{v}\| = |r|\|\mathbf{v}\|$ *for every $\mathbf{v}$ in V and every r in $\mathbb{R}$.*

(d) $\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$ *for all $\mathbf{v}$ and $\mathbf{w}$ in V (**triangle inequality**).*

[0]Hermann Amandus Schwarz (1843–1921) was a German mathematician at the University of Berlin. He had strong geometric intuition, which he applied with great ingenuity to particular problems. A version of the inequality appeared in 1885.

Example 10.1.13. Let $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a spanning set for an inner product space V. If $\mathbf{v}$ in V satisfies $\langle \mathbf{v}, \mathbf{v}_i \rangle = 0$ for each $i = 1, 2, \dots, n$, show that $\mathbf{v} = \mathbf{0}$.

Theorem 10.1.14. *030545 Let V be an inner product space.*

(a) $d(\mathbf{v}, \mathbf{w}) \geq 0$ *for all* $\mathbf{v}$, $\mathbf{w}$ *in* V.

(b) $d(\mathbf{v}, \mathbf{w}) = 0$ *if and only if* $\mathbf{v} = \mathbf{w}$.

(c) $d(\mathbf{v}, \mathbf{w}) = d(\mathbf{w}, \mathbf{v})$ *for all* $\mathbf{v}$ *and* $\mathbf{w}$ *in* V.

(d) $d(\mathbf{v}, \mathbf{w}) \leq d(\mathbf{v}, \mathbf{u}) + d(\mathbf{u}, \mathbf{w})$ *for all* $\mathbf{v}$, $\mathbf{u}$, *and* $\mathbf{w}$ *in* V.

10.2 Orthogonal Sets of Vectors

Recall that two nonzero geometric vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$ are perpendicular (or orthogonal) if and only if $\mathbf{x} \cdot \mathbf{y} = 0$. In general, two vectors $\mathbf{v}$ and $\mathbf{w}$ in an inner product space V are said to be **orthogonal** if

$$\langle \mathbf{v}, \mathbf{w} \rangle = 0$$

A set $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n\}$ of vectors is called an **orthogonal set of vectors** if

(a) *Each* $\mathbf{f}_i \neq \mathbf{0}$.

(b) $\langle \mathbf{f}_i, \mathbf{f}_j \rangle = 0$ *for all* $i \neq j$.

If, in addition, $\|\mathbf{f}_i\| = 1$ for each i, the set $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n\}$ is called an **orthonormal** set.

Theorem 10.2.1. *Pythagoras' Theorem*
If $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n\}$ *is an orthogonal set of vectors, then*

$$\|\mathbf{f}_1 + \mathbf{f}_2 + \cdots + \mathbf{f}_n\|^2 = \|\mathbf{f}_1\|^2 + \|\mathbf{f}_2\|^2 + \cdots + \|\mathbf{f}_n\|^2$$

The proof of the next result is left to the reader.

Theorem 10.2.2. *Let* $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n\}$ *be an orthogonal set of vectors.*

(a) $\{r_1\mathbf{f}_1, r_2\mathbf{f}_2, \ldots, r_n\mathbf{f}_n\}$ *is also orthogonal for any* $r_i \neq 0$ *in* $\mathbb{R}$.

(b) $\left\{\frac{1}{\|\mathbf{f}_1\|}\mathbf{f}_1, \frac{1}{\|\mathbf{f}_2\|}\mathbf{f}_2, \ldots, \frac{1}{\|\mathbf{f}_n\|}\mathbf{f}_n\right\}$ *is an orthonormal set.*

As before, the process of passing from an orthogonal set to an orthonormal one is called **normalizing** the orthogonal set.

Theorem 10.2.3. *Every orthogonal set of vectors is linearly independent.*

Example 10.2.4. Show that $\left\{ \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix} \right\}$ is an orthogonal basis of $\mathbb{R}^3$ with inner product

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^T A \mathbf{w}, \text{ where } A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Theorem 10.2.5. *Expansion Theorem*
Let $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n\}$ *be an orthogonal basis of an inner product space* V*. If* $\mathbf{v}$ *is any vector in* V*, then*

$$\mathbf{v} = \frac{\langle \mathbf{v}, \mathbf{f}_1 \rangle}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 + \frac{\langle \mathbf{v}, \mathbf{f}_2 \rangle}{\|\mathbf{f}_2\|^2}\mathbf{f}_2 + \cdots + \frac{\langle \mathbf{v}, \mathbf{f}_n \rangle}{\|\mathbf{f}_n\|^2}\mathbf{f}_n$$

is the expansion of $\mathbf{v}$ *as a linear combination of the basis vectors.*

The coefficients $\frac{\langle \mathbf{v}, \mathbf{f}_1 \rangle}{\|\mathbf{f}_1\|^2}, \frac{\langle \mathbf{v}, \mathbf{f}_2 \rangle}{\|\mathbf{f}_2\|^2}, \ldots, \frac{\langle \mathbf{v}, \mathbf{f}_n \rangle}{\|\mathbf{f}_n\|^2}$ in the expansion theorem are sometimes called the **Fourier coefficients** of $\mathbf{v}$ with respect to the orthogonal basis $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n\}$.

Example 10.2.6. If $a_0, a_1, \ldots, a_n$ are distinct numbers and $p(x)$ and $q(x)$ are in $\mathbf{P}_n$, define

$$\langle p(x), q(x) \rangle = p(a_0)q(a_0) + p(a_1)q(a_1) + \cdots + p(a_n)q(a_n)$$

This is an inner product on $\mathbf{P}_n$. (Axioms P1–P4 are routinely verified, and P5 holds because 0 is the only polynomial of degree n with $n+1$ distinct roots.)

The **Lagrange polynomials** $\delta_0(x), \delta_1(x), \ldots, \delta_n(x)$ relative to the numbers $a_0, a_1, \ldots, a_n$ are defined as follows:

$$\delta_k(x) = \frac{\prod_{i \neq k}(x - a_i)}{\prod_{i \neq k}(a_k - a_i)} \quad k = 0, 1, 2, \ldots, n$$

where $\prod_{i \neq k}(x - a_i)$ means the product of all the terms

$$(x - a_0), (x - a_1), (x - a_2), \ldots, (x - a_n)$$

except that the kth term is omitted. Then $\{\delta_0(x), \delta_1(x), \ldots, \delta_n(x)\}$ is orthonormal with respect to $\langle\ ,\rangle$ because $\delta_k(a_i) = 0$ if $i \neq k$ and $\delta_k(a_k) = 1$. These facts also show that $\langle p(x), \delta_k(x) \rangle = p(a_k)$ so the expansion theorem gives

$$p(x) = p(a_0)\delta_0(x) + p(a_1)\delta_1(x) + \cdots + p(a_n)\delta_n(x)$$

for each $p(x)$ in $\mathbf{P}_n$. This is the **Lagrange interpolation expansion** of $p(x)$, which is important in numerical integration.

Lemma 10.2.7. *Orthogonal Lemma*
Let $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_m\}$ *be an orthogonal set of vectors in an inner product space* V*, and let* $\mathbf{v}$ *be any vector not in span*$\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_m\}$*. Define*

$$\mathbf{f}_{m+1} = \mathbf{v} - \frac{\langle \mathbf{v}, \mathbf{f}_1 \rangle}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 - \frac{\langle \mathbf{v}, \mathbf{f}_2 \rangle}{\|\mathbf{f}_2\|^2}\mathbf{f}_2 - \cdots - \frac{\langle \mathbf{v}, \mathbf{f}_m \rangle}{\|\mathbf{f}_m\|^2}\mathbf{f}_m$$

Then $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_m, \mathbf{f}_{m+1}\}$ *is an orthogonal set of vectors.*

Theorem 10.2.8. *Gram-Schmidt Orthogonalization Algorithm*
Let V be an inner product space and let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be any basis of V. Define vectors $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n$ in V successively as follows:

$$\begin{aligned}
\mathbf{f}_1 &= \mathbf{v}_1 \\
\mathbf{f}_2 &= \mathbf{v}_2 - \tfrac{\langle \mathbf{v}_2, \mathbf{f}_1 \rangle}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 \\
\mathbf{f}_3 &= \mathbf{v}_3 - \tfrac{\langle \mathbf{v}_3, \mathbf{f}_1 \rangle}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 - \tfrac{\langle \mathbf{v}_3, \mathbf{f}_2 \rangle}{\|\mathbf{f}_2\|^2}\mathbf{f}_2 \\
\vdots &\qquad \vdots \\
\mathbf{f}_k &= \mathbf{v}_k - \tfrac{\langle \mathbf{v}_k, \mathbf{f}_1 \rangle}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 - \tfrac{\langle \mathbf{v}_k, \mathbf{f}_2 \rangle}{\|\mathbf{f}_2\|^2}\mathbf{f}_2 - \cdots - \tfrac{\langle \mathbf{v}_k, \mathbf{f}_{k-1} \rangle}{\|\mathbf{f}_{k-1}\|^2}\mathbf{f}_{k-1}
\end{aligned}$$

for each $k = 2, 3, \ldots, n$. Then

(a) $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n\}$ *is an orthogonal basis of V.*

(b) *span*$\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_k\} =$ *span*$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ *holds for each $k = 1, 2, \ldots, n$.*

The purpose of the Gram-Schmidt algorithm is to convert a basis of an inner product space into an *orthogonal* basis. In particular, it shows that every finite dimensional inner product space *has* an orthogonal basis.

Example 10.2.9. Consider $V = \mathbf{P}_3$ with the inner product $\langle p, q \rangle = \int_{-1}^{1} p(x)q(x)dx$. If the Gram-Schmidt algorithm is applied to the basis $\{1, x, x^2, x^3\}$, show that the result is the orthogonal basis

$$\{1, x, \frac{1}{3}(3x^2 - 1), \frac{1}{5}(5x^3 - 3x)\}$$

Corollary 10.2.10. *If V is any n-dimensional inner product space, then V is isomorphic to $\mathbb{R}^n$ as inner product spaces. More precisely, if E is any orthonormal basis of V, the coordinate isomorphism*

$$C_E : V \to \mathbb{R}^n \text{ satisfies } \langle \mathbf{v}, \mathbf{w} \rangle = C_E(\mathbf{v}) \cdot C_E(\mathbf{w})$$

for all $\mathbf{v}$ and $\mathbf{w}$ in V.

Let U be a subspace of an inner product space V. As in $\mathbb{R}^n$, the **orthogonal complement** $U^\perp$ of U in V is defined by

$$U^\perp = \{\mathbf{v} \mid \mathbf{v} \in V, \langle \mathbf{v}, \mathbf{u} \rangle = 0 \text{ for all } \mathbf{u} \in U\}$$

Theorem 10.2.11. *Let U be a finite dimensional subspace of an inner product space V.*

(a) *$U^\perp$ is a subspace of V and $V = U \oplus U^\perp$.*

(b) *If $dim V = n$, then $dim U + dim U^\perp = n$.*

(c) *If $dim V = n$, then $U^{\perp\perp} = U$.*

Definition 10.2.12. Orthogonal Projection on a Subspace
The projection on U with kernel $U^\perp$ is called the **orthogonal projection** on U (or simply the **projection** on U) and is denoted $\text{proj}_U : V \to V$.

Theorem 10.2.13. *Projection Theorem*
Let U be a finite dimensional subspace of an inner product space V and let $\mathbf{v}$ be a vector in V.

(a) *$proj_U : V \to V$ is a linear operator with image U and kernel $U^\perp$.*

(b) *$proj_U \mathbf{v}$ is in U and $\mathbf{v} - proj_U \mathbf{v}$ is in $U^\perp$.*

(c) *If $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_m\}$ is any orthogonal basis of U, then*

$$proj_U \mathbf{v} = \frac{\langle \mathbf{v}, \mathbf{f}_1 \rangle}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 + \frac{\langle \mathbf{v}, \mathbf{f}_2 \rangle}{\|\mathbf{f}_2\|^2}\mathbf{f}_2 + \cdots + \frac{\langle \mathbf{v}, \mathbf{f}_m \rangle}{\|\mathbf{f}_m\|^2}\mathbf{f}_m$$

Note that there is no requirement that V is finite dimensional.

Example 10.2.14. Let U be a subspace of the finite dimensional inner product space V. Show that $\text{proj}_{U^\perp}\mathbf{v} = \mathbf{v} - \text{proj}_U\mathbf{v}$ for all $\mathbf{v} \in V$.

Theorem 10.2.15. *Approximation Theorem*
Let U be a finite dimensional subspace of an inner product space V. If $\mathbf{v}$ is any vector in V, then $proj_U\mathbf{v}$ is the vector in U that is closest to $\mathbf{v}$. Here ***closest*** *means that*

$$\|\mathbf{v} - proj_U\mathbf{v}\| < \|\mathbf{v} - \mathbf{u}\|$$

for all $\mathbf{u}$ in U, $\mathbf{u} \neq proj_U\mathbf{v}$.

Example 10.2.16. Consider the space $\mathbf{C}[-1, 1]$ of real-valued continuous functions on the interval $[-1, 1]$ with inner product $\langle f, g\rangle = \int_{-1}^{1} f(x)g(x)dx$. Find the polynomial $p = p(x)$ of degree at most 2 that best approximates the absolute-value function f given by $f(x) = |x|$.

10.3 Orthogonal Diagonalization

Theorem 10.3.1. *Let $T : V \to V$ be a linear operator on a finite dimensional space V. Then the following conditions are equivalent.*

(a) *V has a basis consisting of eigenvectors of T.*

(b) *There exists a basis B of V such that $M_B(T)$ is diagonal.*

Definition 10.3.2. Diagonalizable Linear Operators
A linear operator T on a finite dimensional space V is called **diagonalizable** if V has a basis consisting of eigenvectors of T.

Example 10.3.3. Let $T : \mathbf{P}_2 \to \mathbf{P}_2$ be given by

$$T(a + bx + cx^2) = (a + 4c) - 2bx + (3a + 2c)x^2$$

Find the eigenspaces of T and hence find a basis of eigenvectors.

Theorem 10.3.4. *Let $T : V \to V$ be a linear operator on an inner product space V. If $B = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ is an orthogonal basis of V, then*

$$M_B(T) = \left[\frac{\langle \mathbf{b}_i, T(\mathbf{b}_j)\rangle}{\|\mathbf{b}_i\|^2}\right]$$

Example 10.3.5. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be given by

$$T(a, b, c) = (a + 2b - c, 2a + 3c, -a + 3b + 2c)$$

If the dot product in $\mathbb{R}^3$ is used, find the matrix of T with respect to the standard basis $B = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ where $\mathbf{e}_1 = (1, 0, 0)$, $\mathbf{e}_2 = (0, 1, 0)$, $\mathbf{e}_3 = (0, 0, 1)$.

Theorem 10.3.6. *Let V be a finite dimensional inner product space. The following conditions are equivalent for a linear operator $T : V \to V$.*

(a) $\langle \mathbf{v}, T(\mathbf{w}) \rangle = \langle T(\mathbf{v}), \mathbf{w} \rangle$ *for all* $\mathbf{v}$ *and* $\mathbf{w}$ *in* V.

(b) *The matrix of T is symmetric with respect to every orthonormal basis of V.*

(c) *The matrix of T is symmetric with respect to some orthonormal basis of V.*

(d) *There is an orthonormal basis $B = \{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_n\}$ of V such that $\langle \mathbf{f}_i, T(\mathbf{f}_j) \rangle = \langle T(\mathbf{f}_i), \mathbf{f}_j \rangle$ holds for all i and j.*

A linear operator T on an inner product space V is called **symmetric** if $\langle \mathbf{v}, T(\mathbf{w}) \rangle = \langle T(\mathbf{v}), \mathbf{w} \rangle$ holds for all $\mathbf{v}$ and $\mathbf{w}$ in V.

Example 10.3.7. If A is an $n \times n$ matrix, let $T_A : \mathbb{R}^n \to \mathbb{R}^n$ be the matrix operator given by $T_A(\mathbf{v}) = A\mathbf{v}$ for all columns $\mathbf{v}$. If the dot product is used in $\mathbb{R}^n$, then T_A is a symmetric operator if and only if A is a symmetric matrix.

Theorem 10.3.8. *A symmetric linear operator on a finite dimensional inner product space has real eigenvalues.*

If U is a subspace of an inner product space V, recall that its orthogonal complement is the subspace $U^{\perp}$ of V defined by

$$U^{\perp} = \{\mathbf{v} \text{ in } V \mid \langle \mathbf{v}, \mathbf{u} \rangle = 0 \text{ for all } \mathbf{u} \text{ in } U\}$$

Theorem 10.3.9. *Let $T : V \to V$ be a symmetric linear operator on an inner product space V, and let U be a T-invariant subspace of V. Then:*

(a) *The restriction of T to U is a symmetric linear operator on U.*

(b) *$U^{\perp}$ is also T-invariant.*

Theorem 10.3.10. *Principal Axes Theorem*
The following conditions are equivalent for a linear operator T on a finite dimensional inner product space V.

(a) *T is symmetric.*

(b) *V has an orthogonal basis consisting of eigenvectors of T.*

Example 10.3.11. Let $T : \mathbf{P}_2 \to \mathbf{P}_2$ be given by

$$T(a + bx + cx^2) = (8a - 2b + 2c) + (-2a + 5b + 4c)x + (2a + 4b + 5c)x^2$$

Using the inner product $\langle a + bx + cx^2, a' + b'x + c'x^2 \rangle = aa' + bb' + cc'$, show that T is symmetric and find an orthonormal basis of $\mathbf{P}_2$ consisting of eigenvectors.

10.4 Isometries

If V is an inner product space, a transformation $S : V \to V$ (not necessarily linear) is said to be **distance preserving** if the distance between $S(\mathbf{v})$ and $S(\mathbf{w})$ is the same as the distance between $\mathbf{v}$ and $\mathbf{w}$ for all vectors $\mathbf{v}$ and $\mathbf{w}$; more formally, if

$$\|S(\mathbf{v}) - S(\mathbf{w})\| = \|\mathbf{v} - \mathbf{w}\| \quad \text{for all } \mathbf{v} \text{ and } \mathbf{w} \text{ in } V \tag{10.1}$$

Distance-preserving maps need not be linear. For example, if $\mathbf{u}$ is any vector in V, the transformation $S_{\mathbf{u}} : V \to V$ defined by $S_{\mathbf{u}}(\mathbf{v}) = \mathbf{v} + \mathbf{u}$ for all $\mathbf{v}$ in V is called **translation** by $\mathbf{u}$, and it is routine to verify that $S_{\mathbf{u}}$ is distance preserving for any $\mathbf{u}$. However, $S_{\mathbf{u}}$ is linear only if $\mathbf{u} = \mathbf{0}$ (since then $S_{\mathbf{u}}(\mathbf{0}) = \mathbf{0}$). Remarkably, distance-preserving operators that do fix the origin are necessarily linear.

Lemma 10.4.1. *Let V be an inner product space of dimension n, and consider a distance-preserving transformation $S : V \to V$. If $S(\mathbf{0}) = \mathbf{0}$, then S is linear.*

Definition 10.4.2. Isometries
Distance-preserving linear operators are called **isometries**.

Theorem 10.4.3. *If V is a finite dimensional inner product space, then every distance-preserving transformation $S : V \to V$ is the composite of a translation and an isometry.*

Theorem 10.4.4. *Let $T : V \to V$ be a linear operator on a finite dimensional inner product space V.*

The following conditions are equivalent:

1. *T is an isometry.* *(T preserves distance)*
2. *$\|T(\mathbf{v})\| = \|\mathbf{v}\|$ for all $\mathbf{v}$ in V.* *(T preserves norms)*
3. *$\langle T(\mathbf{v}), T(\mathbf{w})\rangle = \langle \mathbf{v}, \mathbf{w}\rangle$ for all $\mathbf{v}$ and $\mathbf{w}$ in V.* *(T preserves inner products)*
4. *If $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n\}$ is an orthonormal basis of V, then $\{T(\mathbf{f}_1), T(\mathbf{f}_2), \ldots, T(\mathbf{f}_n)\}$ is also an orthonormal basis.* *(T preserves orthonormal bases)*
5. *T carries some orthonormal basis to an orthonormal basis.*

Corollary 10.4.5. *Let V be a finite dimensional inner product space.*

1. *Every isometry of a finite dimensional V is an isomorphism.*
2. a. *$1_V : V \to V$ is an isometry.*
   b. *The composite of two isometries of V is an isometry.*
   c. *The inverse of an isometry of V is an isometry.*

The conditions in part (2) of the corollary assert that the set of isometries of a finite dimensional inner product space forms an algebraic system called a **group**.

Example 10.4.6. Rotations of $\mathbb{R}^2$ about the origin are isometries, as are reflections in lines through the origin: They clearly preserve distance and so are linear by Lemma 10.4.5. Similarly, rotations about lines through the origin and reflections in planes through the origin are isometries of $\mathbb{R}^3$.

Example 10.4.7. Let $T : \mathbf{M}_{nn} \to \mathbf{M}_{nn}$ be the transposition operator: $T(A) = A^T$. Then T is an isometry if the inner product is $\langle A, B\rangle = \operatorname{tr}(AB^T) = \sum_{i,j} a_{ij}b_{ij}$. In fact, T permutes the basis consisting of all matrices with one entry 1 and the other entries 0.

Theorem 10.4.8. *Let $T : V \to V$ be an operator where V is a finite dimensional inner product space. The following conditions are equivalent.*

(a) *T is an isometry.*

(b) *$M_B(T)$ is an orthogonal matrix for every orthonormal basis B.*

(c) *$M_B(T)$ is an orthogonal matrix for some orthonormal basis B.*

It is important that B is *orthonormal* in Theorem **??**. For example, $T : V \to V$ given by $T(\mathbf{v}) = 2\mathbf{v}$ preserves *orthogonal* sets but is not an isometry, as is easily checked.

If P is an orthogonal square matrix, then $P^{-1} = P^T$. Taking determinants yields $(\det P)^2 = 1$, so $\det P = \pm 1$. Hence:

Corollary 10.4.9. *If $T : V \to V$ is an isometry where V is a finite dimensional inner product space, then $\det T = \pm 1$.*

Example 10.4.10. If A is any $n \times n$ matrix, the matrix operator $T_A : \mathbb{R}^n \to \mathbb{R}^n$ is an isometry if and only if A is orthogonal using the dot product in $\mathbb{R}^n$. Indeed, if E is the standard basis of $\mathbb{R}^n$, then $M_E(T_A) = A$.

Theorem 10.4.11. *Let $T : V \to V$ be an isometry on the two-dimensional inner product space V. Then there are two possibilities.*

Either (1) *There is an orthonormal basis B of V such that*

$$M_B(T) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix},\ 0 \le \theta < 2\pi$$

or (2) *There is an orthonormal basis B of V such that*

$$M_B(T) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

Furthermore, type (1) occurs if and only if $\det T = 1$, and type (2) occurs if and only if $\det T = -1$.

Corollary 10.4.12. *An operator $T : \mathbb{R}^2 \to \mathbb{R}^2$ is an isometry if and only if T is a rotation or a reflection.*

Example 10.4.13. 032272 In each case, determine whether $T_A : \mathbb{R}^2 \to \mathbb{R}^2$ is a rotation or a reflection, and then find the angle or fixed line:

$$\text{(a) } A = \frac{1}{2}\begin{bmatrix} 1 & \sqrt{3} \\ -\sqrt{3} & 1 \end{bmatrix} \qquad \text{(b) } A = \frac{1}{5}\begin{bmatrix} -3 & 4 \\ 4 & 3 \end{bmatrix}$$

Lemma 10.4.14. *Let $T : V \to V$ be an isometry of a finite dimensional inner product space V. If U is a T-invariant subspace of V, then $U^\perp$ is also T-invariant.*

Lemma 10.4.15. *Let $T : V \to V$ be an isometry of the finite dimensional inner product space V. If λ is a complex eigenvalue of T, then $|\lambda| = 1$.*

Lemma 10.4.16. *Let $T : V \to V$ be an isometry of the n-dimensional inner product space V. If T has a nonreal eigenvalue, then V has a two-dimensional T-invariant subspace.*

Theorem 10.4.17. *Let $T : V \to V$ be an isometry of the n-dimensional inner product space V. Given an angle θ, write $R(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$. Then there exists an orthonormal basis B of V such that $M_B(T)$ has one of the following block diagonal forms, classified for convenience by whether n is odd or even:*

$$n = 2k+1 \quad \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & R(\theta_1) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R(\theta_k) \end{bmatrix} \quad \textit{or} \quad \begin{bmatrix} -1 & 0 & \cdots & 0 \\ 0 & R(\theta_1) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R(\theta_k) \end{bmatrix}$$

$$n = 2k \quad \begin{bmatrix} R(\theta_1) & 0 & \cdots & 0 \\ 0 & R(\theta_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R(\theta_k) \end{bmatrix} \quad \textit{or} \quad \begin{bmatrix} -1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & R(\theta_1) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & R(\theta_{k-1}) \end{bmatrix}$$

Theorem 10.4.18. *If $T : \mathbb{R}^3 \to \mathbb{R}^3$ is an isometry, there are three possibilities.*

a. T is a rotation, and $M_B(T) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}$ for some orthonormal basis B.

b. T is a reflection, and $M_B(T) = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ for some orthonormal basis B.

c. $T = QR = RQ$ where Q is a reflection, R is a rotation about an axis perpendicular to the fixed plane of Q and $M_B(T) = \begin{bmatrix} -1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}$ for some orthonormal basis B.

Hence T is a rotation if and only if $\det T = 1$.

A useful way of analyzing a given isometry $T : \mathbb{R}^3 \to \mathbb{R}^3$ comes from computing the eigenvalues of T. Because the characteristic polynomial of T has degree 3, it must have a real root. Hence, there must be at least one real eigenvalue, and the only possible real eigenvalues are ± 1 by Lemma 10.4.15. Thus Table 10.1 includes all possibilities.

Table 10.1

Eigenvalues of T	**Action of T**
(1) 1, no other real eigenvalues	Rotation about the line $\mathbb{R}\mathbf{f}$ where $\mathbf{f}$ is an eigenvector corresponding to 1. [Case (a) of Theorem **??**.]
(2) -1, no other real eigenvalues	Rotation about the line $\mathbb{R}\mathbf{f}$ followed by reflection in the plane $(\mathbb{R}\mathbf{f})^{\perp}$ where $\mathbf{f}$ is an eigenvector corresponding to -1. [Case (c) of Theorem **??**.]
(3) $-1, 1, 1$	Reflection in the plane $(\mathbb{R}\mathbf{f})^{\perp}$ where $\mathbf{f}$ is an eigenvector corresponding to -1. [Case (b) of Theorem **??**.]
(4) $1, -1, -1$	This is as in (1) with a rotation of π.
(5) $-1, -1, -1$	Here $T(\mathbf{x}) = -\mathbf{x}$ for all x. This is (2) with a rotation of π.
(6) $1, 1, 1$	Here T is the identity isometry.

Example 10.4.19. Analyze the isometry $T : \mathbb{R}^3 \to \mathbb{R}^3$ given by $T \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} y \\ z \\ -x \end{bmatrix}$.

Let V be an n-dimensional inner product space. A subspace of V of dimension $n-1$ is called a **hyperplane** in V. Thus the hyperplanes in $\mathbb{R}^3$ and $\mathbb{R}^2$ are, respectively, the planes and lines through the origin. Let $Q : V \to V$ be an isometry with matrix

$$M_B(Q) = \begin{bmatrix} -1 & 0 \\ 0 & I_{n-1} \end{bmatrix}$$

for some orthonormal basis $B = \{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_n\}$. Then $Q(\mathbf{f}_1) = -\mathbf{f}_1$ whereas $Q(\mathbf{u}) = \mathbf{u}$ for each $\mathbf{u}$ in $U = \text{span}\{\mathbf{f}_2, \dots, \mathbf{f}_n\}$. Hence U is called the **fixed hyperplane** of Q, and Q is called **reflection** in U. Note that each hyperplane in V is the fixed hyperplane of a (unique) reflection of V. Clearly, reflections in $\mathbb{R}^2$ and $\mathbb{R}^3$ are reflections in this more general sense.

Continuing the analogy with $\mathbb{R}^2$ and $\mathbb{R}^3$, an isometry $T : V \to V$ is called a **rotation** if there exists an orthonormal basis $\{\mathbf{f}_1, \dots, \mathbf{f}_n\}$ such that

$$M_B(T) = \begin{bmatrix} I_r & 0 & 0 \\ 0 & R(\theta) & 0 \\ 0 & 0 & I_s \end{bmatrix}$$

in block form, where $R(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$, and where either I_r or I_s (or both) may be missing. If $R(\theta)$ occupies columns i and $i+1$ of $M_B(T)$, and if $W = \text{span}\{\mathbf{f}_i, \mathbf{f}_{i+1}\}$, then W is T-invariant and the matrix of $T : W \to W$ with respect to $\{\mathbf{f}_i, \mathbf{f}_{i+1}\}$ is $R(\theta)$. Clearly, if W is viewed as a copy of $\mathbb{R}^2$, then T is a rotation in W. Moreover, $T(\mathbf{u}) = \mathbf{u}$ holds for all vectors $\mathbf{u}$ in the $(n-2)$-dimensional subspace $U = \text{span}\{\mathbf{f}_1, \dots, \mathbf{f}_{i-1}, \mathbf{f}_{i+1}, \dots, \mathbf{f}_n\}$, and U is called the **fixed axis** of the rotation T. In $\mathbb{R}^3$, the axis of any rotation is a line (one-dimensional), whereas in $\mathbb{R}^2$ the axis is $U = \{\mathbf{0}\}$.

Theorem 10.4.20. *Let $T : V \to V$ be an isometry of a finite dimensional inner product space V. Then there exist isometries $T_1, \dots, T$ such that*

$$T = T_k T_{k-1} \cdots T_2 T_1$$

where each T_i is either a rotation or a reflection, at most one is a reflection, and $T_i T_j = T_j T_i$ holds for all i and j. Furthermore, T is a composite of rotations if and only if $detT = 1$.