Linear Algebra 2

Course Notes for MATH 235

Edition 1.1

D. Wolczuk

Contents

Chapter 7

Fundamental Subspaces

The main purpose of this chapter is to review a few important concepts from the first six chapters. These concepts include subspaces, bases, dimension, and linear mappings. As you will soon see the rest of the book relies heavily on these and other concepts from the first six chapters.

7.1 Bases of Fundamental Subspaces

Recall from Math 136 the four fundamental subspaces of a matrix.

DEFINITION
Fundamental Subspaces

Let A be an $m \times n$ matrix. The four fundamental subspaces of A are

1. The columnspace of A is $\text{Col}(A) = \{A\vec{x} \in \mathbb{R}^m \mid \vec{x} \in \mathbb{R}^n\}$.
2. The rowspace of A is $\text{Row}(A) = \{A^T\vec{x} \in \mathbb{R}^n \mid \vec{x} \in \mathbb{R}^m\}$.
3. The nullspace of A is $\text{Null}(A) = \{\vec{x} \in \mathbb{R}^n \mid A\vec{x} = \vec{0}\}$.
4. The left nullspace of A is $\text{Null}(A^T) = \{\vec{x} \in \mathbb{R}^m \mid A^T\vec{x} = \vec{0}\}$.

THEOREM 1

Let A be an $m\times n$ matrix. Then $\text{Col}(A)$ and $\text{Null}(A^T)$ are subspaces of $\mathbb{R}^m$ and $\text{Row}(A)$ and $\text{Null}(A)$ are subspaces of $\mathbb{R}^n$.

Our goal now is to find an easy way to determine a basis for each of the four fundamental subspaces.

REMARK

To help you understand the following two proofs, you may wish to pick a simple 3×2 matrix A and follow the steps of the proof with your matrix A.

THEOREM 2 Let A be an $m \times n$ matrix. The columns of A which correspond to leading ones in the reduced row echelon form of A form a basis for $\mathrm{Col}(A)$. Moreover,

$$\dim \mathrm{Col}(A) = \mathrm{rank}\, A$$

Proof: We first observe that if A is the zero matrix, then the result is trivial. Hence, we can assume that $\mathrm{rank}\, A = r > 0$.

Denote the columns of the reduced row echelon form R of A by $\vec{r}_1, \dots, \vec{r}_n$. Since $\mathrm{rank}\, A = r$, R contains r leading ones. Let $t_1, \dots, t_r$ denote the indexes of the columns of R which contain leading ones. We will first show that $\mathcal{B} = \{\vec{r}_{t_1}, \dots, \vec{r}_{t_r}\}$ is a basis for $\mathrm{Col}(R)$.

Observe that by definition of the reduced row echelon form the vectors $\vec{r}_{t_1}, \dots, \vec{r}_{t_r}$ are distinct standard basis vectors of $\mathbb{R}^m$ and hence form a linearly independent set. Additionally, every column of R which does not contain a leading one can be written as a linear combination of the columns which do contain leading ones, so $\mathrm{Span}\,\mathcal{B} = \mathrm{Col}(R)$. Therefore, $\mathcal{B}$ is a basis for $\mathrm{Col}(R)$ as claimed.

Denote the columns of A by $\vec{a}_1, \dots, \vec{a}_n$. We will now show that $C = \{\vec{a}_{t_1}, \dots, \vec{a}_{t_r}\}$ is a basis for $\mathrm{Col}(A)$ by using the fact that $\mathcal{B}$ is a basis for $\mathrm{Col}(R)$. To do this, we first need to find a relationship between the vectors in $\mathcal{B}$ and C.

Since R is the reduced row echelon form of A there exists a sequence of elementary matrices $E_1, \dots, E_k$ such that $E_k \cdots E_1 A = R$. Let $E = E_k \cdots E_1$. Recall that every elementary matrix is invertible, hence $E^{-1} = E_1^{-1} \cdots E_k^{-1}$ exists. Then

$$R = EA = [E\vec{a}_1 \ \cdots \ E\vec{a}_n]$$

Consequently, $\vec{r}_i = E\vec{a}_i$, or $\vec{a}_i = E^{-1}\vec{r}_i$.

To prove that C is linearly independent, we use the definition of linear independence. Consider

$$c_1\vec{a}_{t_1} + \cdots + c_r\vec{a}_{t_r} = \vec{0}$$

Multiply both sides by E to get

$$\begin{aligned} E(c_1\vec{a}_{t_1} + \cdots + c_r\vec{a}_{t_r}) &= E\vec{0} \\ c_1E\vec{a}_{t_1} + \cdots + c_rE\vec{a}_{t_r} &= \vec{0} \\ c_1\vec{r}_{t_1} + \cdots + c_r\vec{r}_{t_r} &= \vec{0} \end{aligned}$$

Thus, $c_1 = \cdots = c_r = 0$ since $\{\vec{r}_{t_1}, \dots, \vec{r}_{t_r}\}$ is linearly independent. Thus, C is linearly independent.

To show that C spans $\mathrm{Col}(A)$, we need to show that if we pick any $\vec{b} \in \mathrm{Col}(A)$, we can write it as a linear combination of the vectors $\vec{a}_{t_1}, \dots, \vec{a}_{t_r}$. Pick $\vec{b} \in \mathrm{Col}(A)$. Then, by definition of the columnspace, there exists a $\vec{x} \in \mathbb{R}^n$ such that $A\vec{x} = \vec{b}$. Then we get

$$\begin{aligned} \vec{b} &= A\vec{x} \\ \vec{b} &= E^{-1}R\vec{x} \\ E\vec{b} &= R\vec{x} \end{aligned}$$

Therefore, $E\vec{b}$ is in the columnspace of R and hence can be written as a linear combination of the basis vectors $\{\vec{r}_{t_1}, \dots, \vec{r}_{t_r}\}$. Hence, we have

$$\begin{aligned} E\vec{b} &= d_1\vec{r}_{t_1} + \cdots + d_r\vec{r}_{t_r} \\ \vec{b} &= E^{-1}(d_1\vec{r}_{t_1} + \cdots + d_r\vec{r}_{t_r}) \\ \vec{b} &= d_1E^{-1}\vec{r}_{t_1} + \cdots + d_rE^{-1}\vec{r}_{t_r} \\ \vec{b} &= d_1\vec{a}_{t_1} + \cdots + d_r\vec{a}_{t_r} \end{aligned}$$

as required. Thus, $\operatorname{Span}\{\vec{a}_{t_1}, \dots, \vec{a}_{t_r}\} = \operatorname{Col}(A)$. Hence, $\{\vec{a}_{t_1}, \dots, \vec{a}_{t_r}\}$ is a basis for $\operatorname{Col}(A)$.

Recall that the dimension of a vector space is the number of vectors in any basis. Thus, $\dim \operatorname{Col}(A) = r = \operatorname{rank} A$. □

THEOREM 3 Let A be an $m \times n$ matrix. The set of all non-zero rows in the reduced row echelon form of A form a basis for $\operatorname{Row}(A)$. Hence,

$$\dim \operatorname{Row}(A) = \operatorname{rank} A$$

Proof: We first observe that if A is the zero matrix, then the result is trivial. Hence, we can assume that $\operatorname{rank} A = r > 0$.

Let R be the reduced row echelon form of A. By definition of the reduced row echelon form, we have that the set of all non-zero rows of R form a basis for the rowspace of R. We will now show that they also form a basis for the rowspace of A.

As in the proof above, denote the columns of A by $\vec{a}_1, \dots, \vec{a}_n$. Let $E = E_k \cdots E_1$ where $E_1, \dots, E_k$ are elementary matrices such that $E_k \cdots E_1 A = R$. We then have $A = E^{-1}R$.

Let $\vec{b} \in \operatorname{Row}(A)$. Then

$$\begin{aligned} \vec{b} &= A^T\vec{x} \\ &= (E^{-1}R)^T\vec{x} \\ &= R^T(E^{-1})^T\vec{x} \end{aligned}$$

Let $\vec{y} = (E^{-1})^T\vec{x}$. Then, we have $\vec{b} = R^T\vec{y}$, so $\vec{b} \in \operatorname{Row}(R)$. Thus, $\operatorname{Row}(A) \subseteq \operatorname{Row}(R)$. Hence, the non-zero rows of R spans $\operatorname{Row}(A)$. Since they are also linearly independent, they form a basis for $\operatorname{Row}(A)$, and the result follows. □

We know how to find a basis for the nullspace of a matrix from our work on finding bases of eigenspaces in Chapter 6. You may have noticed in Chapter 6 that the standard procedure for finding a spanning set of a nullspace always led to a linearly independent set. We could prove this directly, but it would be awkward. Instead, we first prove the following important result.

THEOREM 4 **(Dimension Theorem)**

Let A be an $m \times n$ matrix. Then

$$\operatorname{rank} A + \dim \operatorname{Null}(A) = n$$

Proof: Let $\{\vec{v}_1, \dots, \vec{v}_k\}$ be a basis for $\operatorname{Null}(A)$ so that $\dim \operatorname{Null}(A) = k$. Then we can extend this to a basis $\{\vec{v}_1, \dots, \vec{v}_k, \vec{v}_{k+1}, \dots, \vec{v}_n\}$ for $\mathbb{R}^n$. We will prove that $\{A\vec{v}_{k+1}, \dots, A\vec{v}_n\}$ is a basis for $\operatorname{Col}(A)$.

Consider

$$\vec{0} = c_{k+1}A(\vec{v}_{k+1}) + \cdots + c_nA(\vec{v}_n) = A(c_{k+1}\vec{v}_{k+1} + \cdots + c_n\vec{v}_n)$$

Then, $c_{k+1}\vec{v}_{k+1} + \cdots + c_n\vec{v}_n \in \operatorname{Null}(A)$. Hence, we can write it as a linear combination of the basis vectors for $\operatorname{Null}(A)$. So, we get

$$\begin{aligned} c_{k+1}\vec{v}_{k+1} + \cdots + c_n\vec{v}_n &= d_1\vec{v}_1 + \cdots + d_k\vec{v}_k \\ -d_1\vec{v}_1 - \cdots - d_k\vec{v}_k + c_{k+1}\vec{v}_{k+1} + \cdots + c_n\vec{v}_n &= \vec{0} \end{aligned}$$

Thus, $d_1 = \cdots = d_k = c_{k+1} = \cdots = c_n = 0$ since $\{\vec{v}_1, \dots, \vec{v}_n\}$ is linearly independent. Therefore, $\{A\vec{v}_{k+1}, \dots, A\vec{v}_n\}$ is linearly independent.

Let $\vec{b} \in \operatorname{Col}(A)$. Then $\vec{b} = A\vec{x}$ for some $\vec{x} \in \mathbb{R}^n$. Writing $\vec{x}$ as a linear combination of the basis vectors $\{\vec{v}_1, \dots, \vec{v}_n\}$ gives

$$\begin{aligned} \vec{b} &= A(c_1\vec{v}_1 + \cdots + c_k\vec{v}_k + c_{k+1}\vec{v}_{k+1} + \cdots + c_n\vec{v}_n) \\ &= c_1A\vec{v}_1 + \cdots + c_kA\vec{v}_k + c_{k+1}A\vec{v}_{k+1} + \cdots + c_nA\vec{v}_n \end{aligned}$$

But, $\vec{v}_i \in \operatorname{Null}(A)$ for $1 \le i \le k$, so $A\vec{v}_i = \vec{0}$. Hence, we have

$$\begin{aligned} \vec{b} &= \vec{0} + \cdots + \vec{0} + c_{k+1}A\vec{v}_{k+1} + \cdots + c_nA\vec{v}_n \\ &= c_{k+1}A\vec{v}_{k+1} + \cdots + c_nA\vec{v}_n \end{aligned}$$

Thus, $\operatorname{Span}\{A\vec{v}_{k+1}, \dots, A\vec{v}_n\} = \operatorname{Col}(A)$.

Therefore, we have show that $\{A\vec{v}_{k+1}, \dots, A\vec{v}_n\}$ is a basis for $\operatorname{Col}(L)$ and hence

$$\operatorname{rank} A = \dim \operatorname{Col}(A) = n - k = n - \dim \operatorname{Null}(A)$$

as required. $\square$

COROLLARY 5 Let A be an $m \times n$ matrix with rank $A = r$. Then

$$\begin{aligned} \dim \operatorname{Null}(A) &= n - r \\ \dim \operatorname{Null}(A^T) &= m - r \end{aligned}$$

We now can find a basis for the four fundamental subspaces of a matrix. We demonstrate this with a couple examples.

EXAMPLE 1 Let $A = \begin{bmatrix} 1 & 2 & -1 \\ 1 & 1 & 2 \\ 1 & 0 & 5 \end{bmatrix}$. Find a basis for the four fundamental subspaces.

Solution: Row reducing A we get

$$\begin{bmatrix} 1 & 2 & -1 \\ 1 & 1 & 2 \\ 1 & 0 & 5 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 5 \\ 0 & 1 & -3 \\ 0 & 0 & 0 \end{bmatrix} = R$$

A basis for Row(A) is the set of non-zero rows of R. Thus, a basis is $\left\{ \begin{bmatrix} 1 \\ 0 \\ 5 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -3 \end{bmatrix} \right\}$ (remember that in linear algebra, vectors in $\mathbb{R}^n$ are always written as column vectors).

A basis for Col(A) is the columns of A which correspond to the columns of R which contain leading ones. Hence, a basis for Col(A) is $\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} \right\}$.

To find a basis for Null(A), we solve $A\vec{x} = \vec{0}$ as normal. That is, we rewrite the system $R\vec{x} = \vec{0}$ as a system of equations:

$$\begin{aligned} x_1 + 5x_3 &= 0 \\ x_2 - 3x_3 &= 0 \end{aligned}$$

Hence, x_3 is a free variable and the general solution is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -5x_3 \\ 3x_3 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} -5 \\ 3 \\ 1 \end{bmatrix}$$

Consequently, a basis for Null(A) is $\left\{ \begin{bmatrix} -5 \\ 3 \\ 1 \end{bmatrix} \right\}$.

To find a basis for Null(A^T), we can just use our method for finding the nullspace of a matrix on A^T. That is, we row reduce A^T to get

$$\begin{bmatrix} 1 & 1 & 1 \\ 2 & 1 & 0 \\ -1 & 2 & 5 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

So, we have

$$\begin{aligned} x_1 - x_3 &= 0 \\ x_2 + 2x_3 &= 0 \end{aligned}$$

Thus, x_3 is a free variable and the general solution is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_3 \\ -2x_3 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

Therefore, a basis for the left nullspace of A is $\left\{ \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \right\}$.

EXAMPLE 2 Let $A = \begin{bmatrix} 2 & 4 & -4 & 2 \\ 1 & 2 & 2 & -7 \\ -1 & -2 & 0 & 3 \end{bmatrix}$. Find a basis for the four fundamental subspaces.

Solution: Row reducing A we get $\begin{bmatrix} 2 & 4 & -4 & 2 \\ 1 & 2 & 2 & -7 \\ -1 & -2 & 0 & 3 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 0 & -3 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 \end{bmatrix} = R$

The set of non-zero rows $\left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \\ -3 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ -2 \end{bmatrix} \right\}$ of R forms a basis for Row(A).

The columns of A which correspond to the columns of R which contain leading ones form a basis for Col(A). Hence, a basis for Col(A) is $\left\{ \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} -4 \\ 2 \\ 0 \end{bmatrix} \right\}$.

To find a basis for Null(A), we solve $A\vec{x} = \vec{0}$ as normal. That is, we rewrite the system $R\vec{x} = \vec{0}$ as a system of equations:

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

Hence, x_2 and x_4 are free variables and the general solution is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -2x_2 + 3x_4 \\ x_2 \\ 2x_4 \\ x_4 \end{bmatrix} = x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 3 \\ 0 \\ 2 \\ 1 \end{bmatrix}$$

Hence, a basis for Null(A) is $\left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ 2 \\ 1 \end{bmatrix} \right\}$.

We row reduce A^T to get $\begin{bmatrix} 2 & 1 & -1 \\ 4 & 2 & -2 \\ -4 & 2 & 0 \\ 2 & -7 & 3 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -1/4 \\ 0 & 1 & -1/2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

Thus, we have

$$\begin{aligned} x_1 - \frac{1}{4}x_3 &= 0 \\ x_2 - \frac{1}{2}x_3 &= 0 \end{aligned}$$

Consequently, x_3 is a free variable and the general solution is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{4}x_3 \\ \frac{1}{2}x_3 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} 1/4 \\ 1/2 \\ 1 \end{bmatrix}$$

Therefore, a basis for the left nullspace of A is $\left\{ \begin{bmatrix} 1/4 \\ 1/2 \\ 1 \end{bmatrix} \right\}$.

REMARK

It is not actually necessary to row reduce A^T to find a basis for the left nullspace of A. We can use the fact that $EA = R$ where E is the product of elementary matrices used to bring A to its reduced row echelon form R to find a basis for the left nullspace. The derivation of this procedure is left as an exercise.

Section 7.1 Problems

1. Find a basis for the four fundamental subspace of each matrix.

(a) $\begin{bmatrix} 1 & 2 & -1 \\ 2 & 4 & 3 \end{bmatrix}$ (b) $\begin{bmatrix} 2 & -1 & 5 \\ 3 & 4 & 2 \\ 1 & 1 & 1 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & -2 & 3 & 5 \\ -2 & 4 & 0 & -4 \\ 3 & -6 & -5 & 1 \end{bmatrix}$ (d) $\begin{bmatrix} 3 & 2 & 5 & 3 \\ -1 & 0 & -3 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 4 & -5 & 1 \end{bmatrix}$

2. Let A be an $n \times n$ matrix. Prove that $\mathrm{Null}(A) = \{\vec{0}\}$ if and only if $\det A = 0$.

3. Let B be an $m \times n$ matrix.

   (a) Prove that if $\vec{x}$ is any vector in the left nullspace of B, then $\vec{x}^T B = \vec{0}^T$.

   (b) Prove that if $\vec{x}$ is any vector in the left nullspace of B and $\vec{y}$ is any vector in the columnspace of B, then $\vec{x} \cdot \vec{y} = 0$.

4. Invent a 2×2 matrix A such that $\mathrm{Null}(A) = \mathrm{Col}(A)$.

7.2 Subspaces of Linear Mappings

Recall from Math 136 that we defined a linear mapping $L : \mathbb{R}^n \to \mathbb{R}^m$ to be a function with domain $\mathbb{R}^n$ and codomain $\mathbb{R}^m$ such that

$$L(s\vec{x} + t\vec{y}) = sL(\vec{x}) + tL(\vec{y})$$

for all $\vec{x}, \vec{y} \in \mathbb{R}^n$ and $s, t \in \mathbb{R}$. We defined the range and kernel (nullspace) of a linear mapping $L : \mathbb{R}^n \to \mathbb{R}^m$ by

$$\text{Range}(L) = \{L(\vec{x}) \in \mathbb{R}^m \mid \vec{x} \in \mathbb{R}^n\}$$
$$\text{Ker}(L) = \{\vec{x} \in \mathbb{R}^n \mid L(\vec{x}) = \vec{0}\}$$

It is easy to verify using the Subspace Test that Range(L) is a subspace of $\mathbb{R}^m$ and Ker(L) is a subspace of $\mathbb{R}^n$.

Additionally, we saw that every linear mapping $L : \mathbb{R}^n \to \mathbb{R}^m$ is a matrix mapping. In particular, we defined the standard matrix $[L]$ of L by

$$[L] = \begin{bmatrix} L(\vec{e}_1) & \cdots & L(\vec{e}_n) \end{bmatrix}$$

where $\{\vec{e}_1, \dots, \vec{e}_n\}$ is the standard basis for $\mathbb{R}^n$. It satisfies

$$L(\vec{x}) = [L]\vec{x}$$

for all $\vec{x} \in \mathbb{R}^n$. This relationship between linear mappings and matrix mappings is very important, and we will continue to look at this during the remainder of the book. The purpose of the next two theorems is to further demonstrate this relationship.

THEOREM 1 If $L : \mathbb{R}^n \to \mathbb{R}^m$ is a linear mapping, then Range$(L) = \text{Col}([L])$ and rank$[L] = \dim \text{Range}(L)$.

THEOREM 2 If $L : \mathbb{R}^n \to \mathbb{R}^m$ is a linear mapping, then Ker$(L) = \text{Null}([L])$ and $\dim(\text{Ker}(L)) = n - \text{rank}[L]$.

EXAMPLE 1 Let $L : \mathbb{R}^4 \to \mathbb{R}^3$ be the linear mapping with standard matrix

$$[L] = \begin{bmatrix} 0 & 1 & 0 & -2 \\ 1 & 2 & 1 & -1 \\ 2 & 4 & 3 & -1 \end{bmatrix}$$

Find a basis for the range of L and kernel of L.

Solution: To find a basis for the range of L, we can just find a basis for the columnspace of $[L]$. Row reducing $[L]$ gives

$$\begin{bmatrix} 0 & 1 & 0 & -2 \\ 1 & 2 & 1 & -1 \\ 2 & 4 & 3 & -1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

Since the first three columns of the reduced row echelon form a basis for the columnspace of the reduced row echelon form of $[L]$, the first three columns of $[L]$ form a basis for $\mathrm{Col}([L])$. Thus, a basis for the range of L is $\left\{\begin{bmatrix}0\\1\\2\end{bmatrix}, \begin{bmatrix}1\\2\\4\end{bmatrix}, \begin{bmatrix}0\\1\\3\end{bmatrix}\right\}$.

To find a basis for $\mathrm{Ker}(L)$, we will find a basis for the nullspace of $[L]$. Thus, we need to find a basis for the solution space of the homogeneous system $[L]\vec{x} = \vec{0}$. We found above that the RREF of $[L]$ is $\begin{bmatrix}1 & 0 & 0 & 2\\0 & 1 & 0 & -2\\0 & 0 & 1 & 1\end{bmatrix}$. Thus, we have

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

Then the general solution of $[L]\vec{x} = \vec{0}$ is

$$\vec{x} = \begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix} = \begin{bmatrix}-2x_4\\2x_4\\-x_4\\x_4\end{bmatrix} = x_4\begin{bmatrix}-2\\2\\-1\\1\end{bmatrix}$$

So, a basis for $\mathrm{Ker}(L)$ is $\left\{\begin{bmatrix}-2\\2\\-1\\1\end{bmatrix}\right\}$.

The Dimension Theorem for matrices gives us the following result.

THEOREM 3 Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear mapping. Then,

$$\dim \mathrm{Range}(L) + \dim \mathrm{Ker}(L) = \dim(\mathbb{R}^n)$$

Section 7.2 Problems

1. Let $L(x_1, x_2) = (x_1, -2x_1 + x_2, 0)$.
   (a) Prove that L is linear.
   (b) Find the standard matrix of L.
   (c) Find a basis for $\mathrm{Ker}(L)$ and $\mathrm{Range}(L)$.
2. Prove Theorem 1 and Theorem 2.

Chapter 8

Linear Mappings

8.1 General Linear Mappings

Linear Mappings $L : \mathbb{V} \to \mathbb{W}$

We now observe that we can extend our definition of a linear mapping to the case where the domain and codomain are general vectors spaces instead of just $\mathbb{R}^n$.

DEFINITION
Linear Mapping

Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces. A mapping $L : \mathbb{V} \to \mathbb{W}$ is called **linear** if

$$L(s\vec{x} + t\vec{y}) = sL(\vec{x}) + tL(\vec{y})$$

for all $\vec{x}, \vec{y} \in \mathbb{V}$ and $s, t \in \mathbb{R}$.

REMARKS

1. As before, two linear mappings L and M are equal if and only if they have the same domain, the same codomain, and $L(\vec{v}) = M(\vec{v})$ for all $\vec{v}$ in the domain.
2. The definition of a linear mapping above still makes sense because of the closure properties of vector spaces.
3. As before, we sometimes call a linear mapping $L : \mathbb{V} \to \mathbb{V}$ a **linear operator**.
4. It is important not to assume that any results that held for linear mappings $L : \mathbb{R}^n \to \mathbb{R}^m$ also hold for linear mappings $L : \mathbb{V} \to \mathbb{W}$.

EXAMPLE 1 Let $L : P_3(\mathbb{R}) \to \mathbb{R}^2$ be defined by $L(a + bx + cx^2 + dx^3) = \begin{bmatrix} a+d \\ b+c \end{bmatrix}$.

(a) Evaluate $L(1 + 2x^2 - x^3)$.

Solution: $L(1 + 2x^2 - x^3) = \begin{bmatrix} 1+(-1) \\ 0+2 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$.

(b) Prove that L is linear.

Solution: Let $a_1 + b_1x + c_1x^2 + d_1x^3, a_2 + b_2x + c_2x^2 + d_2x^3 \in P_3(\mathbb{R})$ and $s, t \in \mathbb{R}$, then

$$\begin{aligned}
L\big(s(a_1 + b_1x + c_1x^2 + d_1x^3) &+ t(a_2 + b_2x + c_2x^2 + d_2x^3)\big) \\
&= L\big((sa_1 + ta_2) + (sb_1 + tb_2)x + (sc_1 + tc_2)x^2 + (sd_1 + td_2)x^3\big) \\
&= \begin{bmatrix} sa_1 + ta_2 + sd_1 + td_2 \\ sb_1 + tb_2 + sc_1 + tc_2 \end{bmatrix} \\
&= s\begin{bmatrix} a_1 + d_1 \\ b_1 + c_1 \end{bmatrix} + t\begin{bmatrix} a_2 + d_2 \\ b_2 + c_2 \end{bmatrix} \\
&= sL(a_1 + b_1x + c_1x^2 + d_1x^3) + tL(a_2 + b_2x + c_2x^2 + d_2x^3)
\end{aligned}$$

Thus, L is linear.

EXAMPLE 2 Let $\operatorname{tr} : M_{n\times n}(\mathbb{R}) \to \mathbb{R}$ be defined by $\operatorname{tr} A = \sum_{i=1}^{n} (A)_{ii}$ (called the **trace** of a matrix). Prove that tr is linear.

Solution: Let $A, B \in M_{n\times n}(\mathbb{R})$ and $s, t \in \mathbb{R}$. Then

$$\begin{aligned}
\operatorname{tr}(sA + tB) &= \sum_{i=1}^{n} (sA + tB)_{ii} \\
&= \sum_{i=1}^{n} (s(A)_{ii} + t(B)_{ii}) \\
&= s\sum_{i=1}^{n} (A)_{ii} + t\sum_{i=1}^{n} (B)_{ii} \\
&= s \operatorname{tr} A + t \operatorname{tr} B
\end{aligned}$$

Thus, tr is linear.

EXAMPLE 3

Prove that the mapping $L : P_2(\mathbb{R}) \to M_{2\times 2}(\mathbb{R})$ defined by $L(a + bx + cx^2) = \begin{bmatrix} a & bc \\ 0 & abc \end{bmatrix}$ is not linear.

Solution: Observe that

$$L(1 + x + x^2) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

But,

$$L(2(1 + x + x^2)) = L(2 + 2x + 2x^2) = \begin{bmatrix} 2 & 4 \\ 0 & 8 \end{bmatrix} \neq 2L(1 + x + x^2)$$

We now begin to show that many of the results we had for linear mappings $L : \mathbb{R}^n \to \mathbb{R}^m$ also hold for linear mappings $L : \mathbb{V} \to \mathbb{W}$.

THEOREM 1

Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces and let $L : \mathbb{V} \to \mathbb{W}$ be a linear mapping. Then,

$$L(\vec{0}) = \vec{0}$$

DEFINITION

Addition

Scalar Multiplication

Let $L : \mathbb{V} \to \mathbb{W}$ and $M : \mathbb{V} \to \mathbb{W}$ be linear mappings. Then we define $L + M$ by

$$(L + M)(\vec{v}) = L(\vec{v}) + M(\vec{v})$$

and for any $t \in \mathbb{R}$ we define tL by

$$(tL)(\vec{v}) = tL(\vec{v})$$

EXAMPLE 4

Let $L : M_{2\times 2}(\mathbb{R}) \to P_2(\mathbb{R})$ be defined by $L\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = a + bx + (a + c + d)x^2$ and $M : M_{2\times 2}(\mathbb{R}) \to P_2(\mathbb{R})$ be defined by $M\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = (b + c) + dx^2$. Then L and M are both linear and $L + M$ is the mapping defined by

$$\begin{aligned}(L + M)\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) &= L\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) + M\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) \\ &= [a + bx + (a + c + d)x^2] + [(b + c) + dx^2] \\ &= (a + b + c) + bx + (a + c + 2d)x^2\end{aligned}$$

Similarly, $4L$ is the mapping defined by

$$(4L)\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = 4L\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = 4a + 4bx + (4a + 4c + 4d)x^2$$

THEOREM 2

Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces. The set $\mathbb{L}$ of all linear mappings $L : \mathbb{V} \to \mathbb{W}$ with standard addition and scalar multiplication of mappings is a vector space.

Proof: To prove that $\mathbb{L}$ is a vector space, we need to show that it satisfies all ten vector spaces axioms. We will prove V1 and V2 and leave the rest as exercises. Let $L, M \in \mathbb{L}$.

V1 To prove that $\mathbb{L}$ is closed under addition, we need to show that $L + M$ is a linear mapping with domain $\mathbb{V}$ and codomain $\mathbb{W}$.

By definition, the domain of $L + M$ is $\mathbb{V}$ and for any $\vec{v} \in \mathbb{V}$ we have

$$(L + M)(\vec{v}) = L(\vec{v}) + M(\vec{v}) \in \mathbb{W}$$

since $L(\vec{v}) \in \mathbb{W}$, $M(\vec{v}) \in \mathbb{W}$, and $\mathbb{W}$ is closed under addition. Moreover, since L and M are linear for any $\vec{v}_1, \vec{v}_2 \in \mathbb{V}$ and $s, t \in \mathbb{R}$ we have

$$\begin{aligned}(L + M)(s\vec{v}_1 + t\vec{v}_2) &= L(s\vec{v}_1 + t\vec{v}_2) + M(s\vec{v}_1 + t\vec{v}_2)\\ &= sL(\vec{v}_1) + tL(\vec{v}_2) + sM(\vec{v}_1) + tM(\vec{v}_2)\\ &= s[L(\vec{v}_1) + M(\vec{v}_1)] + t[L(\vec{v}_2) + M(\vec{v}_2)]\\ &= s(L + M)(\vec{v}_1) + t(L + M)(\vec{v}_2)\end{aligned}$$

Thus, $L + M$ is linear so $L + M \in \mathbb{L}$.

V2 For any $\vec{v} \in \mathbb{V}$ we have

$$(L + M)(\vec{v}) = L(\vec{v}) + M(\vec{v}) = M(\vec{v}) + L(\vec{v}) = (M + L)(\vec{v})$$

since addition in $\mathbb{W}$ is commutative. Hence $L + M = M + L$.

□

We also define the composition of mappings in the expected way.

DEFINITION

Composition

Let $L : \mathbb{V} \to \mathbb{W}$ and $M : \mathbb{W} \to \mathbb{U}$ be linear mappings. Then we define $M \circ L$ by

$$(M \circ L)(\vec{v}) = M(L(\vec{v}))$$

for all $\vec{v} \in \mathbb{V}$.

EXAMPLE 5 Let $L : P_2(\mathbb{R}) \to \mathbb{R}^3$ be defined by $L(a+bx+cx^2) = \begin{bmatrix} a+b \\ c \\ 0 \end{bmatrix}$ and let $M : \mathbb{R}^3 \to M_{2\times 2}(\mathbb{R})$ be defined by $M\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_1 & 0 \\ x_2 + x_3 & 0 \end{bmatrix}$. Then $M \circ L$ is the mapping defined by

$$(M \circ L)(a + bx + cx^2) = M(L(a + bx + cx^2)) = M\left(\begin{bmatrix} a+b \\ c \\ 0 \end{bmatrix}\right) = \begin{bmatrix} a+b & 0 \\ c & 0 \end{bmatrix}$$

Observe that $M \circ L$ is in fact a linear mapping from $P_2(\mathbb{R})$ to $M_{2\times 2}(\mathbb{R})$.

Section 8.1 Problems

1. Determine which of the following mappings are linear. If it is linear, prove it. If not, give a counterexample to show that it is not linear.

   (a) $L : \mathbb{R}^2 \to M_{2\times 2}(\mathbb{R})$ defined by $L(a, b) = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$

   (b) $T : P_2(\mathbb{R}) \to P_2(\mathbb{R})$ defined by $T(a + bx + cx^2) = (a - b) + (bc)x^2$.

   (c) $L : P_2(\mathbb{R}) \to M_{2\times 2}(\mathbb{R})$ defined by $L(a + bx + cx^2) = \left(\begin{bmatrix} 1 & 0 \\ a+c & b+c \end{bmatrix}\right)$.

   (d) $T : M_{2\times 2}(\mathbb{R}) \to \mathbb{R}^3$ defined by $T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} a+b \\ b+c \\ c-a \end{bmatrix}$

   (e) $D : P_3(\mathbb{R}) \to P_2(\mathbb{R})$ defined by $D(a + bx + cx^2 + dx^3) = b + 2cx + 3dx^2$

   (f) $L : M_{2\times 2}(\mathbb{R}) \to \mathbb{R}$ defined by $L(A) = \det A$.

2. Let $\mathcal{B}$ be a basis for an n-dimensional vector space $\mathbb{V}$. Prove that the mapping $L : \mathbb{V} \to \mathbb{R}^n$ defined by $L(\vec{v}) = [\vec{v}]_{\mathcal{B}}$ for all $\vec{v} \in \mathbb{V}$ is linear.

3. Let $L : \mathbb{V} \to \mathbb{W}$ be a linear mapping. Prove that $L(\vec{0}) = \vec{0}$.

4. Let $L : \mathbb{V} \to \mathbb{W}$ and $M : \mathbb{W} \to \mathbb{U}$ be linear mappings. Prove that $M \circ L$ is a linear mapping from $\mathbb{V}$ to $\mathbb{U}$.

5. Let $\mathbb{L}$ be the set of all linear mappings from $\mathbb{V}$ to $\mathbb{W}$ with standard addition and scalar multiplication of linear mappings. Prove that

   (a) $tL \in \mathbb{L}$ for all $L \in \mathbb{L}$ and $t \in \mathbb{R}$.

   (b) $t(L + M) = tL + tM$ for all $L, M \in \mathbb{L}$ and $t \in \mathbb{R}$.

8.2 Rank-Nullity Theorem

Our goal in this section is not only to extend the definitions of the range and kernel of a linear mapping to general linear mappings, but to also generalize Theorem 8.2.3 to general linear mappings. We will see that this generalization, called the Rank-Nullity Theorem, is extremely usefully.

DEFINITION

Range

Kernel

Let $L : \mathbb{V} \to \mathbb{W}$ be a linear mapping, then the **kernel** of L is

$$\mathrm{Ker}(L) = \{\vec{v} \in \mathbb{V} \mid L(\vec{v}) = \vec{0}_{\mathbb{W}}\}$$

and the **range** of L is

$$\mathrm{Range}(L) = \{L(\vec{v}) \mid \vec{v} \in \mathbb{V}\}$$

THEOREM 1

Let $L : \mathbb{V} \to \mathbb{W}$ be a linear mapping. Then $\mathrm{Ker}(L)$ is a subspace of $\mathbb{V}$ and $\mathrm{Range}(L)$ is a subspace of $\mathbb{W}$.

The procedure for finding a basis for the range and kernel of a general linear mapping is, of course, exactly the same as we saw for linear mappings from $\mathbb{R}^n$ to $\mathbb{R}^m$.

EXAMPLE 1

Find a basis for the range and kernel of $L : P_3(\mathbb{R}) \to \mathbb{R}^2$ defined by

$$L(a + bx + cx^2 + dx^3) = \begin{bmatrix} a+d \\ b+c \end{bmatrix}$$

Solution: If $a + bx + cx^2 + dx^3 \in \mathrm{Ker}(L)$, then

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = L(a + bx + cx^2 + dx^3) = \begin{bmatrix} a+d \\ b+c \end{bmatrix}$$

Therefore, $a = -d$ and $b = -c$. Thus, every vector in $\mathrm{Ker}(L)$ has the form

$$a + bx + cx^2 + dx^3 = -d - cx + cx^2 + dx^3 = d(-1 + x^3) + c(-x + x^2)$$

Since $\{-1 + x^3, -x + x^2\}$ is clearly linearly independent, it is a basis for $\mathrm{Ker}(L)$.

The range of L contains all vectors of the form

$$\begin{bmatrix} a+d \\ b+c \end{bmatrix} = (a+d)\begin{bmatrix} 1 \\ 0 \end{bmatrix} + (b+c)\begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

So a basis for $\mathrm{Range}(L)$ is $\left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\}$.

EXAMPLE 2

Determine the dimension of the range and kernel of $L : P_2(\mathbb{R}) \to M_{3\times 2}(\mathbb{R})$ defined by

$$L(a + bx + cx^2) = \begin{bmatrix} a & b \\ c & b \\ a+b & a+c-b \end{bmatrix}$$

Solution: If $a + bx + cx^2 \in \text{Ker}(L)$, then

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = L(a + bx + cx^2) = \begin{bmatrix} a & b \\ c & b \\ a+b & a+c-b \end{bmatrix}$$

Hence, we must have $a = b = c = 0$. Thus, $\text{Ker}(L) = \{\vec{0}\}$, so a basis for $\text{Ker}(L)$ is the empty set. Hence, $\dim(\text{Ker}(L)) = 0$.

The range of L contains all vectors of the form

$$\begin{bmatrix} a & b \\ c & b \\ a+b & a+c-b \end{bmatrix} = a\begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 1 & 1 \end{bmatrix} + b\begin{bmatrix} 0 & 1 \\ 0 & 1 \\ 1 & -1 \end{bmatrix} + c\begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

We can easily verify that the set $\left\{\begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ 1 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}\right\}$ is linearly independent and hence a basis for $\text{Range}(L)$. Consequently, $\dim(\text{Range}(L)) = 3$.

EXERCISE 1

Find a basis for the range and kernel of $L : \mathbb{R}^2 \to M_{2\times 2}(\mathbb{R})$ defined by

$$L\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 & x_1 + x_2 \\ 0 & x_1 + x_2 \end{bmatrix}$$

Observe in all of the examples above that $\dim(\text{Range}(L)) + \dim(\text{Ker}(L)) = \dim \mathbb{V}$ which matches the result we had for linear mappings $L : \mathbb{R}^n \to \mathbb{R}^m$. Before we extend this to the general case, we make a couple of definitions.

DEFINITION

Rank

Nullity

Let $L : \mathbb{V} \to \mathbb{W}$ be a linear mapping. Then we define the **rank** of L by

$$\text{rank}(L) = \dim(\text{Range}(L))$$

We define the **nullity** of L to be

$$\text{nullity}(L) = \dim(\text{Ker}(L))$$

THEOREM 2 **(Rank-Nullity Theorem)**

Let $\mathbb{V}$ be an n-dimensional vector space and let $\mathbb{W}$ be a vector space. If $L : \mathbb{V} \to \mathbb{W}$ is linear, then

$$\operatorname{rank}(L) + \operatorname{nullity}(L) = n$$

Proof: Assume that nullity$(L) = k$ and let $\{\vec{v}_1, \cdots, \vec{v}_k\}$ be a basis for Ker(L). Then we can extend $\{\vec{v}_1, \cdots, \vec{v}_k\}$ to a basis $\{\vec{v}_1, \cdots, \vec{v}_k, \vec{v}_{k+1}, \cdots, \vec{v}_n\}$ for $\mathbb{V}$. We claim that $\{L(\vec{v}_{k+1}), \cdots, L(\vec{v}_n)\}$ is a basis for Range(L).

Let $\vec{y} \in$ Range(L). Then by definition, there exists $\vec{v} \in \mathbb{V}$ such that

$$\begin{aligned}
\vec{y} = L(\vec{v}) &= L(c_1\vec{v}_1 + \cdots + c_k\vec{v}_k + c_{k+1}\vec{v}_{k+1} + \cdots + c_n\vec{v}_n) \\
&= c_1L(\vec{v}_1) + \cdots + c_kL(\vec{v}_k) + c_{k+1}L(\vec{v}_{k+1}) + \cdots + c_nL(\vec{v}_n) \\
&= \vec{0} + \cdots + \vec{0} + c_{k+1}L(\vec{v}_{k+1}) + \cdots + c_nL(\vec{v}_n) \\
&= c_{k+1}L(\vec{v}_{k+1}) + \cdots + c_nL(\vec{v}_n)
\end{aligned}$$

Hence

$$\vec{y} \in \operatorname{Span}\{L(\vec{v}_{k+1}), \ldots, L(\vec{v}_n)\}$$

Consider

$$\begin{aligned}
c_{k+1}L(\vec{v}_{k+1}) + \cdots + c_nL(\vec{v}_n) &= \vec{0} \\
L(c_{k+1}\vec{v}_{k+1} + \cdots + c_n\vec{v}_n) &= \vec{0}
\end{aligned}$$

This implies that $c_{k+1}\vec{v}_{k+1} + \cdots + c_n\vec{v}_n \in$ Ker(L). Thus we can write

$$c_{k+1}\vec{v}_{k+1} + \cdots + c_n\vec{v}_n = d_1\vec{v}_1 + \cdots + d_k\vec{v}_k$$

But, then we have

$$-d_1\vec{v}_1 - \cdots - d_k\vec{v}_k + c_{k+1}\vec{v}_{k+1} + \cdots + c_n\vec{v}_n = \vec{0}$$

Hence, $d_1 = \cdots = d_k = c_{k+1} = \cdots = c_n = 0$ since $\{\vec{v}_1, \cdots, \vec{v}_n\}$ is linearly independent as it is a basis for $\mathbb{V}$.

Therefore, we have shown that $\{L(\vec{v}_{k+1}), \cdots, L(\vec{v}_n)\}$ is a basis for Range(L) and hence

$$\operatorname{rank}(L) = n - k = n - \operatorname{nullity}(L)$$

□

REMARK

The proof of the Rank-Nullity Theorem should seem very familiar. It is essentially identical to that of the Dimension Theorem in Chapter 7. In this book there will be quite a few times where proofs are essentially repeated, so spending time to make sure you understand proofs when you first see them can have real long term benefits.

EXAMPLE 3 Let $L : P_2(\mathbb{R}) \to M_{2\times 3}(\mathbb{R})$ be defined by

$$L(a + bx + cx^2) = \begin{bmatrix} a & a & c \\ a & a & c \end{bmatrix}$$

Find the rank and nullity of L.

Solution: If $a + bx + cx^2 \in \operatorname{Ker}(L)$ then

$$\begin{bmatrix} a & a & c \\ a & a & c \end{bmatrix} = L(a + bx + cx^2) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Hence $a = 0$ and $c = 0$. So every vector in $\operatorname{Ker}(L)$ has the form bx so a basis for $\operatorname{Ker}(L)$ is $\{x\}$. Thus $\operatorname{nullity}(L) = 1$ and so by the Rank-Nullity Theorem we get that

$$\operatorname{rank}(L) = \dim P_2(\mathbb{R}) - \operatorname{nullity}(L) = 3 - 1 = 2$$

In the example above, we could have instead found a basis for the range and then used the Rank-Nullity Theorem to find the nullity of L.

Section 8.2 Problems

1. Find the rank and nullity of the following linear mappings.

   (a) $L : \mathbb{R}^2 \to M_{2\times 2}(\mathbb{R})$ defined by $L(a + bx + cx^2) = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$

   (b) $L : P_2(\mathbb{R}) \to P_2(\mathbb{R})$ defined by $L(a + bx + cx^2) = (a - b) + (bc)x^2$.

   (c) $T : P_2(\mathbb{R}) \to M_{2\times 2}(\mathbb{R})$ defined by $T(a + bx + cx^2) = \begin{bmatrix} 0 & 0 \\ a + c & b + c \end{bmatrix}$.

   (d) $L : \mathbb{R}^3 \to P_1(\mathbb{R})$ defined by $L\left(\begin{bmatrix} a \\ b \\ c \end{bmatrix}\right) = (a + b) + (a + b + c)x$.

2. Prove that if $\{L(\vec{v}_1), \dots, L(\vec{v}_k)\}$ spans $\mathbb{W}$, then $\dim \mathbb{V} \geq \dim \mathbb{W}$.

3. Find a linear mapping $L : \mathbb{V} \to \mathbb{W}$ where $\dim \mathbb{V} \geq \dim \mathbb{W}$, but $\operatorname{Range}(L) \neq \mathbb{W}$.

4. Find a linear mapping $L : \mathbb{V} \to \mathbb{V}$ such that $\operatorname{Ker}(L) = \operatorname{Range}(L)$.

5. Let $\mathbb{V}$ and $\mathbb{W}$ be n-dimensional vector spaces and let $L : \mathbb{V} \to \mathbb{W}$ be a linear mapping. Prove that $\operatorname{Range}(L) = \mathbb{W}$ if and only if $\operatorname{Ker}(L) = \{\vec{0}\}$.

6. Let $\mathbb{U}, \mathbb{V}, \mathbb{W}$ be finite dimensional vector spaces and let $L : \mathbb{V} \to \mathbb{U}$ and $M : \mathbb{U} \to \mathbb{W}$ be linear mappings.

   (a) Prove that $\operatorname{rank}(M \circ L) \leq \operatorname{rank}(M)$.

   (b) Prove that $\operatorname{rank}(M \circ L) \leq \operatorname{rank}(L)$.

   (c) Prove that if M is invertible, then $\operatorname{rank}(M \circ L) = \operatorname{rank} L$.

8.3 Matrix of a Linear Mapping

We now show that every linear mapping $L : \mathbb{V} \to \mathbb{W}$ can also be represented as a matrix mapping. However, we must be careful when dealing with general vector spaces as our domain and codomain. For example, it is certainly impossible to represent a linear mapping $L : P_2(\mathbb{R}) \to M_{2\times 2}(\mathbb{R})$ as a matrix mapping $L(\vec{x}) = A\vec{x}$ since we can not multiply a matrix by a polynomial $\vec{x} \in P_2(\mathbb{R})$. Moreover, we require the output to be a 2×2 matrix.

Thus, if we are going to define a matrix representation of a general linear mapping, we need to convert vectors from $\mathbb{V}$ to vectors in $\mathbb{R}^n$. Recall that the coordinate vector of $\vec{v} \in \mathbb{V}$ with respect to an ordered basis $\mathcal{B}$ is a vector in $\mathbb{R}^n$. In particular, if $\mathcal{B} = \{\vec{v}_1, \dots, \vec{v}_n\}$ and $\vec{v} = b_1\vec{v}_1 + \cdots + b_n\vec{v}_n$, then the coordinate vector of $\vec{v}$ with respect to $\mathcal{B}$ is defined to be

$$[\vec{v}]_{\mathcal{B}} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

Using coordinates, we can write a matrix mapping representation for a linear mapping $L : \mathbb{V} \to \mathbb{W}$. That is, we want to find a matrix A such that

$$[L(\vec{x})]_{\mathcal{C}} = A[\vec{x}]_{\mathcal{B}}$$

for every $\vec{x} \in \mathbb{V}$, where $\mathcal{B}$ is a basis for $\mathbb{V}$ and $\mathcal{C}$ is a basis for $\mathbb{W}$.

Consider the left-hand side $[L(\vec{x})]_{\mathcal{C}}$. Using properties of linear mappings and coordinates, we get

$$\begin{aligned} [L(\vec{x})]_{\mathcal{C}} &= [L(b_1\vec{v}_1 + \cdots + b_n\vec{v}_n)]_{\mathcal{C}} \\ &= [b_1L(\vec{v}_1) + \cdots + b_nL(\vec{v}_n)]_{\mathcal{C}} \\ &= b_1[L(\vec{v}_1)]_{\mathcal{C}} + \cdots + b_n[L(\vec{v}_n)]_{\mathcal{C}} \\ &= \begin{bmatrix} [L(\vec{v}_1)]_{\mathcal{C}} & \cdots & [L(\vec{v}_n)]_{\mathcal{C}} \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} \\ &= A[\vec{x}]_{\mathcal{B}} \end{aligned}$$

Thus, we see the desired matrix is

$$A = \begin{bmatrix} [L(\vec{v}_1)]_{\mathcal{C}} & \cdots & [L(\vec{v}_n)]_{\mathcal{C}} \end{bmatrix}$$

DEFINITION

Matrix of a Linear Mapping

Suppose $\mathcal{B} = \{\vec{v}_1, \dots, \vec{v}_n\}$ is any basis for a vector space $\mathbb{V}$ and $\mathcal{C}$ is any basis for a finite dimensional vector space $\mathbb{W}$ and let $L : \mathbb{V} \to \mathbb{W}$ be a linear mapping. Then the **matrix of L with respect to bases $\mathcal{B}$ and $\mathcal{C}$** is

$${}_{\mathcal{C}}[L]_{\mathcal{B}} = \begin{bmatrix} [L(\vec{v}_1)]_{\mathcal{C}} & \cdots & [L(\vec{v}_n)]_{\mathcal{C}} \end{bmatrix}$$

It satisfies

$$[L(\vec{x})]_{\mathcal{C}} = {}_{\mathcal{C}}[L]_{\mathcal{B}}[\vec{x}]_{\mathcal{B}}$$

for all $\vec{x} \in \mathbb{V}$.

EXAMPLE 1 Let $L : P_2(\mathbb{R}) \to \mathbb{R}^2$ be the linear mapping defined by $L(a + bx + cx^2) = \begin{bmatrix} a+c \\ b-c \end{bmatrix}$, $\mathcal{B} = \{1 + x^2, 1 + x, -1 + x + x^2\}$, $C = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$. Find ${}_C[L]_{\mathcal{B}}$.

Solution: To find ${}_C[L]_{\mathcal{B}}$ we need to determine the C-coordinates of the images of the vectors in $\mathcal{B}$ under L. We have

$$\begin{aligned} L(1 + x^2) &= \begin{bmatrix} 2 \\ -1 \end{bmatrix} = 3\begin{bmatrix} 1 \\ 0 \end{bmatrix} + (-1)\begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ L(1 + x) &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 0\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 1\begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ L(-1 + x + x^2) &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0\begin{bmatrix} 1 \\ 1 \end{bmatrix} \end{aligned}$$

Hence,

$${}_C[L]_{\mathcal{B}} = \begin{bmatrix} [L(1 + x^2)]_C & [L(1 + x)]_C & [L(-1 + x + x^2)]_C \end{bmatrix} = \begin{bmatrix} 3 & 0 & 0 \\ -1 & 1 & 0 \end{bmatrix}$$

We can check our answer. Let $p(x) = a + bx + cx^2$. To use ${}_C[L]_{\mathcal{B}}$ we need $[p(x)]_{\mathcal{B}}$. Hence, we need to write $p(x)$ as a linear combination of the vectors in $\mathcal{B}$. In particular, we need to solve

$$a+bx+cx^2 = t_1(1+x^2)+t_2(1+x)+t_3(-1+x+x^2) = (t_1+t_2-t_3)+(t_2+t_3)x+(t_1+t_3)x^3$$

Row reducing the corresponding augmented matrix gives

$$\left[\begin{array}{ccc|c} 1 & 1 & -1 & a \\ 0 & 1 & 1 & b \\ 1 & 0 & 1 & c \end{array}\right] \sim \left[\begin{array}{c} (a-b+3c)/3 \\ (a+2b-c)/3 \\ (-a+b+c)/3 \end{array}\right]$$

Thus,

$$[p(x)]_{\mathcal{B}} = \frac{1}{3}\begin{bmatrix} a-b+3c \\ a+2b-c \\ -a+b+c \end{bmatrix}$$

So we get

$$\begin{aligned} [L(p(x))]_C &= {}_C[L]_{\mathcal{B}}[p(x)]_{\mathcal{B}} \\ &= \begin{bmatrix} 3 & 0 & 0 \\ -1 & 1 & 0 \end{bmatrix} \frac{1}{3}\begin{bmatrix} a-b+2c \\ a+2b-c \\ -a+b+c \end{bmatrix} \\ &= \begin{bmatrix} a-b+2c \\ b-c \end{bmatrix} \end{aligned}$$

Therefore, by definition of C-coordinates

$$L(p(x)) = (a-b+2c)\begin{bmatrix} 1 \\ 0 \end{bmatrix} + (b-c)\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} a+c \\ b-c \end{bmatrix}$$

as required.

EXAMPLE 2

Let $T : \mathbb{R}^2 \to M_{2\times 2}(\mathbb{R})$ be the linear mapping defined by $T\left(\begin{bmatrix} a \\ b \end{bmatrix}\right) = \begin{bmatrix} a+b & 0 \\ 0 & a-b \end{bmatrix}$, and let $\mathcal{B} = \left\{\begin{bmatrix} 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix}\right\}$ and $\mathcal{C} = \left\{\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}\right\}$. Determine ${}_{\mathcal{C}}[T]_{\mathcal{B}}$ and use it to calculate the $T(\vec{v})$ where $[\vec{v}]_{\mathcal{B}} = \begin{bmatrix} 2 \\ -3 \end{bmatrix}$.

Solution: To find ${}_{\mathcal{C}}[L]_{\mathcal{B}}$ we need to determine the $\mathcal{C}$-coordinates of the images of the vectors in $\mathcal{B}$ under L. We have

$$T(2,-1) = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} = -2\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} + 1\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + 2\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} + 0\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$

$$T(1,2) = \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix} = 4\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} + 3\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + (-4)\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} + 0\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$

Hence,

$${}_{\mathcal{C}}[T]_{\mathcal{B}} = \begin{bmatrix} [T(2,-1)]_{\mathcal{C}} & [T(1,2)]_{\mathcal{C}} \end{bmatrix} = \begin{bmatrix} -2 & 4 \\ 1 & 3 \\ 2 & -4 \\ 0 & 0 \end{bmatrix}$$

Thus, we get

$$[T(\vec{v})]_{\mathcal{C}} = \begin{bmatrix} -2 & 4 \\ 1 & 3 \\ 2 & -4 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 2 \\ -3 \end{bmatrix} = \begin{bmatrix} -16 \\ -7 \\ 16 \\ 0 \end{bmatrix}$$

so

$$T(\vec{v}) = (-16)\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} + (-7)\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + 16\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} + (0)\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} -7 & 0 \\ 0 & 9 \end{bmatrix}$$

In the special case of a linear operator L acting on a finite-dimensional vector space $\mathbb{V}$ with basis $\mathcal{B}$, we often wish to find the matrix ${}_{\mathcal{B}}[L]_{\mathcal{B}}$.

DEFINITION

Matrix of a Linear Operator

Suppose $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_n\}$ is any basis for a vector space $\mathbb{V}$ and let $L : \mathbb{V} \to \mathbb{V}$ be a linear operator. Then the **$\mathcal{B}$-matrix of** L (or the matrix of L with respect to the basis $\mathcal{B}$) is

$$[L]_{\mathcal{B}} = \begin{bmatrix} [L(\vec{v}_1)]_{\mathcal{B}} & \cdots & [L(\vec{v}_n)]_{\mathcal{B}} \end{bmatrix}$$

It satisfies

$$[L(\vec{x})]_{\mathcal{B}} = [L]_{\mathcal{B}}[\vec{x}]_{\mathcal{B}}$$

for all $\vec{x} \in \mathbb{V}$.

EXAMPLE 3 Let $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \right\}$ and let $A = \begin{bmatrix} -6 & -2 & 9 \\ -5 & -1 & 7 \\ -7 & -2 & 10 \end{bmatrix}$. Find the $\mathcal{B}$-matrix of the linear mapping defined by $L(\vec{x}) = A\vec{x}$.

Solution: By definition, the columns of the $\mathcal{B}$-matrix are the $\mathcal{B}$-coordinates of the images of the vectors in $\mathcal{B}$ under L. We find that

$$\begin{bmatrix} -6 & -2 & 9 \\ -5 & -1 & 7 \\ -7 & -2 & 10 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 1\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + 0\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + 0\begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$$

$$\begin{bmatrix} -6 & -2 & 9 \\ -5 & -1 & 7 \\ -7 & -2 & 10 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + 1\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + 0\begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$$

$$\begin{bmatrix} -6 & -2 & 9 \\ -5 & -1 & 7 \\ -7 & -2 & 10 \end{bmatrix}\begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 6 \\ 6 \\ 7 \end{bmatrix} = 3\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + 2\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + 1\begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$$

Hence, $[L]_{\mathcal{B}} = \begin{bmatrix} [L(1,1,1)]_{\mathcal{B}} & [L(1,0,1)]_{\mathcal{B}} & [L(1,3,2)]_{\mathcal{B}} \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$

EXAMPLE 4 Let $L : P_2(\mathbb{R}) \to P_2(\mathbb{R})$ be defined by $L(a+bx+cx^2) = (2a+b)+(a+b+c)x+(b+2c)x^2$. Find $[L]_{\mathcal{B}}$ where $\mathcal{B} = \{-1+x^2, 1-2x+x^2, 1+x+x^2\}$ and find $[L]_{\mathcal{S}}$ where $\mathcal{S} = \{1, x, x^2\}$.

Solution: To determine $[L]_{\mathcal{B}}$ we need to find the $\mathcal{B}$-coordinates of the images of the vectors in $\mathcal{B}$ under L.

$$\begin{aligned} L(-1+x^2) &= -2+0x+2x^2 = 2(-1+x^2)+0(1-2x+x^2)+0(1+x+x^2) \\ L(1-2x+x^2) &= 0+0x+0x^2 = 0(-1+x^2)+0(1-2x+x^2)+0(1+x+x^2) \\ L(1+x+x^2) &= 3+3x+3x^2 = 0(-1+x^2)+0(1-2x+x^2)+3(1+x+x^2) \end{aligned}$$

Thus,

$$[L]_{\mathcal{B}} = \begin{bmatrix} [L(-1+x^2)]_{\mathcal{B}} & [L(1-2x+x^2)]_{\mathcal{B}} & [L(1+x+x^2)]_{\mathcal{B}} \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

Similarly, for $[L]_{\mathcal{S}}$ we calculate the $\mathcal{S}$-coordinates of the image of the vectors in $\mathcal{S}$ under L.

$$\begin{aligned} L(1) &= 2+x = 2(1)+1(x)+0x^2 \\ L(x) &= 1+x+x^2 = 1(1)+1(x)+1(x^2) \\ L(x^2) &= x+2x^2 = 0(1)+1(x)+2(x^2) \end{aligned}$$

Hence,

$$[L]_{\mathcal{S}} = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

In Example 4 we found the matrix of L with respect to the basis $\mathcal{B}$ is diagonal while the matrix of L with respect to the basis $\mathcal{S}$ is not. This is perhaps not surprising since we see that the vectors in $\mathcal{B}$ are **eigenvectors** of L (that is, they satisfy $L(\vec{v}_i) = \lambda_i \vec{v}_i$), while the vectors in $\mathcal{S}$ are not. Recall from Math 136, that we called such a basis a geometrically natural basis.

REMARK

The relationship between diagonalization and the matrix of a linear operator is extremely important. We will review this a little later in the text. You may also find it useful at this point to review Section 6.1 in the Math 136 course notes.

Section 8.3 Problems

1. Find the matrix of each linear mapping with respect to the bases $\mathcal{B}$ and C.

   (a) $D : P_2(\mathbb{R}) \to P_1(\mathbb{R})$ defined by $D(a + bx + cx^2) = b + 2cx$, $\mathcal{B} = \{1, x, x^2\}$, $C = \{1, x\}$

   (b) $L : \mathbb{R}^2 \to P_2(\mathbb{R})$ defined by $L(a_1, a_2) = (a_1 + a_2) + a_1x^2$, $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$, $C = \{1 + x^2, 1 + x, -1 - x + x^2\}$

   (c) $T : M_{2\times2}(\mathbb{R}) \to \mathbb{R}^2$ defined by $T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} a + c \\ b + c \end{bmatrix}$,
   $\mathcal{B} = \left\{ \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix} \right\}$, $C = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$

2. Find the $\mathcal{B}$-matrix of each linear mapping with respect to the give basis $\mathcal{B}$.

   (a) $D : P_2(\mathbb{R}) \to P_2(\mathbb{R})$ defined by $D(a + bx + cx^2) = b + 2cx$, $\mathcal{B} = \{1, x, x^2\}$

   (b) $L : M_{2\times2}(\mathbb{R}) \to M_{2\times2}(\mathbb{R})$ defined by $L\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} a & b + c \\ 0 & d \end{bmatrix}$,
   $\mathcal{B} = \left\{ \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right\}$

   (c) $T : \mathbb{U} \to \mathbb{U}$, where $\mathbb{U}$ is the subspace of diagonal matrices in $M_{2\times2}(\mathbb{R})$, defined by $T\left(\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}\right) = \begin{bmatrix} a + b & 0 \\ 0 & 2a + b \end{bmatrix}$, $\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \right\}$

3. Let $\mathbb{V}$ be an n-dimensional vector space and let $L : \mathbb{V} \to \mathbb{V}$ by the linear operator defined by $L(\vec{v}) = \vec{v}$. Prove that $[L]_{\mathcal{B}} = I$ for any basis $\mathcal{B}$ of $\mathbb{V}$.

4. Invent a mapping $L : \mathbb{R}^2 \to \mathbb{R}^2$ and a basis $\mathcal{B}$ for $\mathbb{R}^2$, such that $\operatorname{Col}([L]_{\mathcal{B}}) \neq \operatorname{Range}(L)$. Justify your answer.

8.4 Isomorphisms

Recall that the ten vector space axioms define a "structure" for the set based on the operations of addition and scalar multiplication. Since all vector spaces satisfy the same ten properties, we would expect that all n-dimensional vector spaces should have the same structure. This idea seems to be confirmed by our work with coordinate vectors in Math 136. No matter what n-dimensional vector space $\mathbb{V}$ we use and which basis we use for the vector space, we have a nice way of relating the vectors in $\mathbb{V}$ to vectors in $\mathbb{R}^n$. Moreover, with respect to this basis the operations of addition and scalar multiplication are preserved. Consider the following calculations:

$$\begin{bmatrix}1\\2\\3\end{bmatrix} + \begin{bmatrix}4\\5\\6\end{bmatrix} = \begin{bmatrix}5\\7\\9\end{bmatrix}$$

$$(1 + 2x + 3x^2) + (4 + 5x + 6x^2) = 5 + 7x + 9x^2$$

$$(\vec{v}_1 + 2\vec{v}_2 + 3\vec{v}_3) + (4\vec{v}_1 + 5\vec{v}_2 + 6\vec{v}_3) = 5\vec{v}_1 + 7\vec{v}_2 + 9\vec{v}_3$$

No matter which vector space we are using and which basis for that vector space, any linear combination of vectors is really just performed on the coordinates of the vectors with respect to the basis.

We will now look at how to use general linear mappings to mathematically prove these observations. We begin with some familiar concepts.

One-To-One and Onto

DEFINITION
One-To-One
Onto

Let $L : \mathbb{V} \to \mathbb{W}$ be a linear mapping. L is called **one-to-one** (injective) if for every $\vec{u}, \vec{v} \in \mathbb{V}$ such that $L(\vec{u}) = L(\vec{v})$ we must have $\vec{u} = \vec{v}$.
L is called **onto** (surjective) if for every $\vec{w} \in \mathbb{W}$, there exists $\vec{v} \in \mathbb{V}$ such that $L(\vec{v}) = \vec{w}$.

EXAMPLE 1

Let $L : \mathbb{R}^2 \to P_2(\mathbb{R})$ be the linear mapping defined by $L(a_1, a_2) = a_1 + a_2x^2$. Then L is one-to-one, since if $L(a_1, a_2) = L(b_1, b_2)$, then $a_1 + a_2x^2 = b_1 + b_2x^2$ and hence $a_1 = b_1$ and $a_2 = b_2$. L is not onto, since there is no $\vec{a} \in \mathbb{R}^2$ such that $L(\vec{a}) = x$.

EXAMPLE 2

Let $L : M_{2\times2}(\mathbb{R}) \to \mathbb{R}$ be the linear mapping defined by $L(A) = \operatorname{tr} A$. Then L is not one-to-one, since $L\left(\begin{bmatrix}1 & 0\\0 & -1\end{bmatrix}\right) = L\left(\begin{bmatrix}0 & 0\\0 & 0\end{bmatrix}\right)$, but $\begin{bmatrix}1 & 0\\0 & -1\end{bmatrix} \neq \begin{bmatrix}0 & 0\\0 & 0\end{bmatrix}$. L is onto, since if we pick any $x \in \mathbb{R}$, then we can find a matrix $A \in M_{2\times2}(\mathbb{R})$, for example $A = \begin{bmatrix}x & 0\\0 & 0\end{bmatrix}$, such that $L(A) = x$.

Observe that a linear mapping being one-to-one means that each $\vec{w} \in \operatorname{Range}(L)$ is mapped to by exactly one $\vec{v} \in \mathbb{V}$ while L being onto means that $\operatorname{Range}(L) = \mathbb{W}$. So there is a relationship between a mapping being onto and its range. We now establish a relationship between a mapping being one-to-one and its kernel. These relationships will be exploited in some proofs below.

LEMMA 1 Let $L : \mathbb{V} \to \mathbb{W}$ be a linear mapping. L is one-to-one if and only if $\mathrm{Ker}(L) = \{\vec{0}\}$.

Proof: Assume L is one-to-one. Let $\vec{x} \in \mathrm{Ker}(L)$. Then $L(\vec{x}) = \vec{0} = L(\vec{0})$ which implies that $\vec{x} = \vec{0}$ since L is one-to-one. Hence $\mathrm{Ker}(L) = \{\vec{0}\}$.

Assume $\mathrm{Ker}(L) = \{\vec{0}\}$ and consider $L(\vec{u}) = L(\vec{v})$. Then

$$\vec{0} = L(\vec{u}) - L(\vec{v}) = L(\vec{u} - \vec{v})$$

Hence, $\vec{u} - \vec{v} \in \mathrm{Ker}(L)$, so $\vec{u} - \vec{v} = \vec{0}$. Therefore, $\vec{u} = \vec{v}$ and so L is one-to-one. $\square$

EXAMPLE 3 Let $L : \mathbb{R}^3 \to P_5(\mathbb{R})$ be the linear mapping defined by

$$L(a, b, c) = a(1 + x^3) + b(1 + x^2 + x^5) + cx^4$$

Prove that L is one-to-one, but not onto.

Solution: Let $\vec{a} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \in \mathrm{Ker}(L)$. Then,

$$0 = L(a, b, c) = a(1 + x^3) + b(1 + x^2 + x^5) + cx^4$$

This gives $a = b = c = 0$, so $\mathrm{Ker}(L) = \left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \right\}$. Hence, L is one-to-one by Lemma 1.

Observe that a basis for $\mathrm{Range}(L)$ is $\{1 + x^3, 1 + x^2 + x^5, x^4\}$. Hence, $\dim(\mathrm{Range}(L)) = 3 < \dim P_5(\mathbb{R})$. Thus, $\mathrm{Range}(L) \neq P_5(\mathbb{R})$, so L is not onto.

EXAMPLE 4 Let $L : M_{2\times 2}(\mathbb{R}) \to \mathbb{R}^2$ be the linear mapping defined by

$$L\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} a \\ b \end{bmatrix}$$

Prove that L is onto, but not one-to-one.

Solution: Observe that $\begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix} \in \mathrm{Ker}(L)$. Thus, L is not one-to-one by Lemma 1.

Let $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2$. Then observe that $L\left(\begin{bmatrix} x_1 & x_2 \\ 0 & 0 \end{bmatrix}\right) = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, so L is onto.

EXERCISE 1 Let $L : \mathbb{R}^3 \to P_2(\mathbb{R})$ be the linear mapping defined by $L(a, b, c) = a + bx + cx^2$. Prove that L is one-to-one and onto.

Isomorphisms

To prove that two vector spaces $\mathbb{V}$ and $\mathbb{W}$ have the same structure we need to relate each vector $\vec{v}_i \in \mathbb{V}$ with a unique vector $\vec{w}_i \in \mathbb{W}$ such that if $a\vec{v}_1 + b\vec{v}_2 = \vec{v}_3$, then $a\vec{w}_1 + b\vec{w}_2 = \vec{w}_3$. Thus, we will require a linear mapping from $\mathbb{V}$ to $\mathbb{W}$ which is both one-to-one and onto.

DEFINITION
Isomorphism
Isomorphic

Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces. Then $\mathbb{V}$ is said to be **isomorphic** to $\mathbb{W}$ if there exists a linear mapping $L : \mathbb{V} \to \mathbb{W}$ which is one-to-one and onto. L is called an **isomorphism** from $\mathbb{V}$ to $\mathbb{W}$.

EXAMPLE 5

Show that $\mathbb{R}^4$ is isomorphic to $M_{2\times 2}(\mathbb{R})$.

Solution: We need to invent a linear mapping L from $\mathbb{R}^4$ to $M_{2\times 2}(\mathbb{R})$ that is one-to-one and onto. To define such a linear mapping, we think about how to relate vectors in $\mathbb{R}^4$ to vectors in $M_{2\times 2}(\mathbb{R})$. Keeping in mind our work at the beginning of this section, we define the mapping $L : \mathbb{R}^4 \to M_{2\times 2}(\mathbb{R})$ by

$$L(a_0, a_1, a_2, a_3) = \begin{bmatrix} a_0 & a_1 \\ a_2 & a_3 \end{bmatrix}$$

Linear: Let $\vec{a} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix}, \vec{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} \in \mathbb{R}^4$ and $s, t \in \mathbb{R}$. Then L is linear since

$$\begin{aligned} L(s\vec{a} + t\vec{b}) &= L(sa_0 + tb_0, sa_1 + tb_1, sa_2 + tb_2, sa_3 + tb_3) \\ &= \begin{bmatrix} sa_0 + tb_0 & sa_1 + tb_1 \\ sa_2 + tb_2 & sa_3 + tb_3 \end{bmatrix} \\ &= s\begin{bmatrix} a_0 & a_1 \\ a_2 & a_3 \end{bmatrix} + t\begin{bmatrix} b_0 & b_1 \\ b_2 & b_3 \end{bmatrix} = sL(\vec{a}) + tL(\vec{b}) \end{aligned}$$

One-To-One: If $L(\vec{a}) = L(\vec{b})$, then $\begin{bmatrix} a_0 & a_1 \\ a_2 & a_3 \end{bmatrix} = \begin{bmatrix} b_0 & b_1 \\ b_2 & b_3 \end{bmatrix}$ which implies $a_i = b_i$ for $i = 0, 1, 2, 3$. Therefore, $\vec{a} = \vec{b}$ and so L is one-to-one.

Onto: Pick $\begin{bmatrix} a_0 & a_1 \\ a_2 & a_3 \end{bmatrix} \in M_{2\times 2}(\mathbb{R})$. Then $L(a_0, a_1, a_2, a_3) = \begin{bmatrix} a_0 & a_1 \\ a_2 & a_3 \end{bmatrix}$, so L is onto.

Hence, L is an isomorphism and so $\mathbb{R}^4$ is isomorphic to $M_{2\times 2}(\mathbb{R})$.

It is instructive to think carefully about the isomorphism in the example above. Observe that the images of the standard basis vectors for $\mathbb{R}^4$ are the standard basis vectors for $M_{2\times 2}(\mathbb{R})$. That is, the isomorphism is mapping basis vectors to basis vectors. We will keep this in mind when constructing an isomorphism in the next example.

EXAMPLE 6 Let $\mathbb{T} = \{p(x) \in P_2(\mathbb{R}) \mid p(1) = 0\}$ and $\mathbb{S} = \left\{ \begin{bmatrix} a \\ b \\ c \end{bmatrix} \in \mathbb{R}^4 \mid a + b + c = 0 \right\}$. Prove that $\mathbb{T}$ is isomorphic to $\mathbb{S}$.

Solution: By the factor theorem, every vector $p(x) \in \mathbb{T}$ has the form $(x-1)(a+bx)$. Hence, a basis for $\mathbb{T}$ is $\mathcal{B} = \{1 - x, x - x^2\}$. Also, every vector $\vec{x} \in \mathbb{S}$ has the form

$$\vec{x} = \begin{bmatrix} a \\ b \\ -a-b \end{bmatrix} = a\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + b\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$$

Therefore, a basis for $\mathbb{S}$ is $\mathcal{C} = \left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \right\}$. So, to define an isomorphism we want to map the vectors in $\mathcal{B}$ to the vectors in $\mathcal{C}$. In particular, we define $L : \mathbb{T} \to \mathbb{S}$ by

$$L\Big(a(1-x) + b(x-x^2)\Big) = a\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + b\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$$

<u>Linear</u>: Let $p(x), q(x) \in \mathbb{T}$ with $p(x) = a_1(1-x) + b_1(x-x^2)$ and $q(x) = a_2(1-x) + b_2(x-x^2)$. Then for any $s, t \in \mathbb{R}$ we get

$$\begin{aligned} L(sp(x) + tq(x)) &= L\Big(s(a_1 + ta_2)(1-x) + (sb_1 + tb_2)(x-x^2)\Big) \\ &= s(a_1 + ta_2)\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + (sb_1 + tb_2)\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \\ &= s\left[a_1\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + b_1\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}\right] + t\left[a_2\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + b_2\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}\right] \\ &= sL(p(x)) + tL(q(x)) \end{aligned}$$

Thus L is linear.

<u>One-To-One</u>: Let $a(1-x) + b(x-x^2) \in \mathrm{Ker}(L)$. Then

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = L\Big(a(1-x) + b(x-x^2)\Big) = a\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + b\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} a \\ b \\ -a-b \end{bmatrix}$$

Hence, $a = b = 0$, so $\mathrm{Ker}(L) = \{0\}$. Therefore, by Lemma 1, L is one-to-one.

<u>Onto</u>: Let $a\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + b\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \in \mathbb{S}$. Then

$$L\Big(a(1-x) + b(x-x^2)\Big) = a\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + b\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$$

so L is onto.

Thus, L is an isomorphism and $\mathbb{T}$ is isomorphic to $\mathbb{S}$.

EXERCISE 2 Prove that $P_2(\mathbb{R})$ is isomorphic to $\mathbb{R}^3$.

Observe in the examples above that the isomorphic vector spaces have the same dimension. This makes sense since if we are mapping basis vectors to basis vectors, then both vector spaces need to have the same number of basis vectors. We now prove that this is indeed always the case.

THEOREM 2 Let $\mathbb{V}$ and $\mathbb{W}$ be finite dimensional vector spaces. Then $\mathbb{V}$ is isomorphic to $\mathbb{W}$ if and only if $\dim \mathbb{V} = \dim \mathbb{W}$.

Proof: On one hand, assume that $\dim \mathbb{V} = \dim \mathbb{W} = n$. Let $\mathcal{B} = \{\vec{v}_1, \dots, \vec{v}_n\}$ be a basis for $\mathbb{V}$ and $\mathcal{C} = \{\vec{w}_1, \dots, \vec{w}_n\}$ be a basis for $\mathbb{W}$. As in our examples above, we define a mapping $L : \mathbb{V} \to \mathbb{W}$ which maps the basis vectors in $\mathcal{B}$ to the basis vectors in $\mathcal{C}$. So, we define

$$L(t_1\vec{v}_1 + \cdots + t_n\vec{v}_n) = t_1\vec{w}_1 + \cdots + t_n\vec{w}_n$$

<u>Linear</u>: Let $\vec{x} = a_1\vec{v}_1 + \cdots + a_n\vec{v}_n, \vec{y} = b_1\vec{v}_1 + \cdots + b_n\vec{v}_n \in \mathbb{V}$ and $s, t \in \mathbb{R}$. Then

$$\begin{aligned} L(s\vec{x} + t\vec{y}) &= L((sa_1 + tb_1)\vec{v}_1 + \cdots + (sa_n + tb_n)\vec{v}_n) \\ &= (sa_1 + tb_1)\vec{w}_1 + \cdots + (sa_n + tb_n)\vec{w}_n \\ &= s(a_1\vec{w}_1 + \cdots + a_n\vec{w}_n) + t(b_1\vec{w}_1 + \cdots + b_n\vec{w}_n) \\ &= sL(\vec{x}) + tL(\vec{y}) \end{aligned}$$

Therefore, L is linear.

<u>One-To-One</u>: If $\vec{v} \in \operatorname{Ker}(L)$, then

$$\vec{0} = L(\vec{v}) = L(c_1\vec{v}_1 + \cdots + c_n\vec{v}_n) = c_1L(\vec{v}_1) + \cdots + c_nL(\vec{v}_n) = c_1\vec{w}_1 + \cdots + c_n\vec{w}_n$$

$\mathcal{C}$ is linearly independent so we get $c_1 = \cdots = c_n = 0$. Therefore $\operatorname{Ker}(L) = \{\vec{0}\}$ and hence L is one-to-one.

<u>Onto</u>: We have $\operatorname{Ker}(L) = \{\vec{0}\}$ since L is one-to-one. Thus, we get $\operatorname{rank}(L) = \dim \mathbb{V} - 0 = n$ by the Rank-Nullity Theorem. Consequently, $\operatorname{Range}(L)$ is an n-dimensional subspace of $\mathbb{W}$ which implies that $\operatorname{Range}(L) = \mathbb{W}$, so L is onto.

Thus, L is an isomorphism from $\mathbb{V}$ to $\mathbb{W}$ and so $\mathbb{V}$ is isomorphic to $\mathbb{W}$.

On the other hand, assume that $\mathbb{V}$ is isomorphic to $\mathbb{W}$. Then there exists an isomorphism L from $\mathbb{V}$ to $\mathbb{W}$. Since L is one-to-one and onto, we get $\operatorname{Range}(L) = \mathbb{W}$ and $\operatorname{Ker}(L) = \{\vec{0}\}$. Thus, the Rank-Nullity Theorem gives

$$\dim \mathbb{W} = \dim(\operatorname{Range}(L)) = \operatorname{rank}(L) = \dim \mathbb{V} - \operatorname{nullity}(L) = \dim \mathbb{V}$$

as required. $\square$

The proof not only proves the theorem, but it demonstrates a few additional facts. First, it shows the intuitively obvious fact that if $\mathbb{V}$ is isomorphic to $\mathbb{W}$, then $\mathbb{W}$ is isomorphic to $\mathbb{V}$. Typically we just say that $\mathbb{V}$ and $\mathbb{W}$ are isomorphic. Second, it confirms that we can make an isomorphism from $\mathbb{V}$ to $\mathbb{W}$ by mapping basis vectors of $\mathbb{V}$ to basis vectors to $\mathbb{W}$. Finally, observe that once we had proven that L was one-to-one, we could exploit the Rank-Nullity Theorem and that $\dim \mathbb{V} = \dim \mathbb{W}$ to immediately get that the mapping was onto. In particular, it shows us how to prove the following theorem.

THEOREM 3 If $\mathbb{V}$ and $\mathbb{W}$ are isomorphic vector spaces and $L : \mathbb{V} \to \mathbb{W}$ is linear, then L is one-to-one if and only if L is onto.

REMARK

Note that if $\mathbb{V}$ and $\mathbb{W}$ are both n-dimensional, then it does **not** mean every linear mapping $L : \mathbb{V} \to \mathbb{W}$ must be an isomorphism. For example, $L : \mathbb{R}^4 \to M_{2\times 2}(\mathbb{R})$ defined by $L(x_1, x_2, x_3, x_4) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ is definitely not one-to-one nor onto!

We saw that we can make an isomorphism by mapping basis vectors to basis vectors. The following theorem shows that this property actually characterizes isomorphisms.

THEOREM 4 Let $\mathbb{V}$ and $\mathbb{W}$ be isomorphic vector spaces and let $\{\vec{v}_1, \dots, \vec{v}_n\}$ be a basis for $\mathbb{V}$. Then a linear mapping $L : \mathbb{V} \to \mathbb{W}$ is an isomorphism if and only if $\{L(\vec{v}_1), \dots, L(\vec{v}_n)\}$ is a basis for $\mathbb{W}$.

The following exercise is intended to be quite challenging. It requires complete understanding of what we have done in this chapter.

EXERCISE 3 Use the following steps to find a basis for the vector space $\mathbb{L}$ of all linear operators $L : \mathbb{V} \to \mathbb{V}$ on an 2-dimensional vector space $\mathbb{V}$.

1. Prove that $\mathbb{L}$ is isomorphic to $M_{2\times 2}(\mathbb{R})$ by constructing an isomorphism T.
2. Find an isomorphism T^{-1} from $M_{2\times 2}(\mathbb{R})$ to $\mathbb{L}$. Let $\{\vec{e}_1, \vec{e}_2, \vec{e}_3, \vec{e}_4\}$ denote the standard basis vectors for $M_{2\times 2}(\mathbb{R})$. Calculate $T^{-1}(\vec{e}_1)$, $T^{-1}(\vec{e}_2)$, $T^{-1}(\vec{e}_3)$, $T^{-1}(\vec{e}_4)$. By Theorem 4 these vectors for a basis for $\mathbb{L}$.
3. Demonstrate that you are correct by proving that the set $\{T^{-1}(\vec{e}_1), T^{-1}(\vec{e}_2), T^{-1}(\vec{e}_3), T^{-1}(\vec{e}_4)\}$ is linearly independent and spans $\mathbb{L}$.

Section 8.4 Problems

1. For each of the following pairs of vector spaces, define an explicit isomorphism to establish that the spaces are isomorphic. Prove that your map is an isomorphism.

   (a) $P_1(\mathbb{R})$ and $\mathbb{R}^2$

   (b) $P_3(\mathbb{R})$ and $M_{2\times 2}(\mathbb{R})$

   (c) $L : P_2(\mathbb{R}) \to P_2(\mathbb{R})$ defined by $L(a + bx + cx^2) = (a - b) + (bc)x^2$

   (d) The vector space $\mathbb{P} = \{p(x) \in P_3 \mid p(1) = 0\}$ and the vector space $\mathbb{U}$ of 2×2 upper triangular matrices

   (e) The vector space $\mathbb{S} = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ 0 \\ x_3 \end{bmatrix} \in \mathbb{R}^4 \mid x_1 - x_2 + x_3 = 0 \right\}$ and the vector space $\mathbb{U} = \{a + bx + cx^2 \in P_2 \mid a = b\}$

   (f) Any n-dimensional vector space $\mathbb{V}$ and $P_{n-1}(\mathbb{R})$.

2. Let $\mathbb{V}$ and $\mathbb{W}$ be n-dimensional vector spaces and let $L : \mathbb{V} \to \mathbb{W}$ be a linear mapping. Prove that L is one-to-one if and only if L is onto.

3. Let $L : \mathbb{V} \to \mathbb{W}$ be an isomorphism and let $\{\vec{v}_1, \dots, \vec{v}_n\}$ be a basis for $\mathbb{V}$. Prove that $\{L(\vec{v}_1), \dots, L(\vec{v}_n)\}$ is a basis for $\mathbb{W}$.

4. Let $L : \mathbb{V} \to \mathbb{U}$ and $M : \mathbb{U} \to \mathbb{W}$ be linear mappings.

   (a) Prove that if L and M are onto, then $M \circ L$ is onto.

   (b) Give an example where L is not onto, but $M \circ L$ is onto.

   (c) Is it possible to give an example where M is not onto, but $M \circ L$ is onto? Explain.

5. Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces with $\dim \mathbb{V} = n$ and $\dim \mathbb{W} = m$, let $L : \mathbb{V} \to \mathbb{W}$ be a linear mapping, and let A be the matrix of L with respect to bases $\mathcal{B}$ for $\mathbb{V}$ and $\mathcal{C}$ for $\mathbb{W}$.

   (a) Define an explicit isomorphism from $\operatorname{Range}(L)$ to $\operatorname{Col}(A)$. Prove that your map is an isomorphism.

   (b) Use (a) to prove that $\operatorname{rank}(L) = \operatorname{rank}(A)$.

6. Let $\mathbb{U}$ and $\mathbb{V}$ be finite dimensional vector spaces and $L : \mathbb{U} \to \mathbb{V}$ and $M : \mathbb{V} \to \mathbb{U}$ both be one-to-one linear mappings. Prove that $\mathbb{U}$ and $\mathbb{V}$ are isomorphic.

Chapter 9

Inner Products

In Math 136, we looked at the dot product function for vectors in $\mathbb{R}^n$. We saw that the dot product function has an important relationship to the length of a vector and to the angle between two vectors. Moreover, the dot product gave us an easy way of finding the projection of a vector onto a plane. In Chapter 8, we saw that every n-dimensional vector space is identical to $\mathbb{R}^n$ in terms of being a vector space. Thus, we should be able to extend these ideas to general vector spaces.

9.1 Inner Product Spaces

Recall that we defined a vector space so that the operations of addition and scalar multiplication had the essential properties of addition and scalar multiplication of vectors in $\mathbb{R}^n$. Thus, to generalize the concept of the dot product to general vector spaces, it makes sense to include the essential properties of the dot product in our definition.

DEFINITION

Inner Product

Let $\mathbb{V}$ be a vector space. An **inner product** on $\mathbb{V}$ is a function $\langle\, , \rangle : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ that has the following properties: for every $\vec{v}, \vec{u}, \vec{w} \in \mathbb{V}$ and $s, t \in \mathbb{R}$ we have

I1 $\langle \vec{v}, \vec{v} \rangle \geq 0$ and $\langle \vec{v}, \vec{v} \rangle = 0$ if and only if $\vec{v} = \vec{0}$ (Positive Definite)
I2 $\langle \vec{v}, \vec{w} \rangle = \langle \vec{w}, \vec{v} \rangle$ (Symmetric)
I3 $\langle s\vec{v} + t\vec{u}, \vec{w} \rangle = s\langle \vec{v}, \vec{w} \rangle + t\langle \vec{u}, \vec{w} \rangle$ (Bilinear)

DEFINITION

Inner Product Space

A vector space $\mathbb{V}$ with an inner product $\langle\, , \rangle$ on $\mathbb{V}$ is called an **inner product space**.

In the same way that a vector space is dependent on the defined operations of addition and scalar multiplication, an inner product space is dependent on the definitions of addition, scalar multiplication, and the inner product. This will be demonstrated in the examples below.

EXAMPLE 1 The dot product is an inner product on $\mathbb{R}^n$, called the standard inner product on $\mathbb{R}^n$.

EXAMPLE 2 Which of the following defines an inner product on $\mathbb{R}^3$?

(a) $\langle \vec{x}, \vec{y} \rangle = x_1y_1 + 2x_2y_2 + 4x_3y_3$.

Solution: We have
$$\langle \vec{x}, \vec{x} \rangle = x_1^2 + 2x_2^2 + 4x_3^2 \geq 0$$
and $\langle \vec{x}, \vec{x} \rangle = 0$ if and only if $\vec{x} = \vec{0}$. Hence, it is positive definite.

$$\langle \vec{x}, \vec{y} \rangle = x_1y_1 + 2x_2y_2 + 4x_3y_3 = y_1x_1 + 2y_2x_2 + 4y_3x_3 = \langle \vec{y}, \vec{x} \rangle$$

Therefore, it is symmetric. Also, it is bilinear since

$$\begin{aligned}\langle s\vec{x} + t\vec{y}, \vec{z} \rangle &= (sx_1 + ty_1)z_1 + 2(sx_2 + ty_2)z_2 + 4(sx_3 + ty_3)z_3 \\ &= s(x_1z_1 + 2x_2z_2 + 4x_3z_3) + t(y_1z_1 + 2y_2z_2 + 4y_3z_3) \\ &= s\langle \vec{x}, \vec{z} \rangle + t\langle \vec{y}, \vec{z} \rangle\end{aligned}$$

Thus $\langle\, , \rangle$ is an inner product.

(b) $\langle \vec{x}, \vec{y} \rangle = x_1y_1 - x_2y_2 + x_3y_3$.

Solution: Observe that if $\vec{x} = \begin{bmatrix}0\\1\\0\end{bmatrix}$, then

$$\langle \vec{x}, \vec{x} \rangle = 0(0) - 1(1) + 0(0) = -1 < 0$$

So, it is not positive definite and thus it is not an inner product.

(c) $\langle \vec{x}, \vec{y} \rangle = x_1y_1 + x_2y_2$.

Solution: If $\vec{x} = \begin{bmatrix}0\\0\\1\end{bmatrix}$, then

$$\langle \vec{x}, \vec{x} \rangle = 0(0) + 0(0) = 0$$

but $\vec{x} \neq \vec{0}$. So, it is not positive definite and thus it is not an inner product.

REMARK

A vector space can have infinitely many different inner products. However, different inner products do not necessarily differ in an interesting way. For example, in $\mathbb{R}^n$ all inner products behave like the dot product with respect to some basis.

EXAMPLE 3 On the vector space $M_{2\times 2}(\mathbb{R})$, define $\langle\, ,\rangle$ by

$$\langle A, B\rangle = \operatorname{tr}(B^T A)$$

where tr is the trace of a matrix defined as the sum of the diagonal entires. Show that $\langle\, ,\rangle$ is an inner product on $M_{2\times 2}(\mathbb{R})$.

Solution: Let $A, B, C \in M_{2\times 2}(\mathbb{R})$ and $s, t \in \mathbb{R}$. Then:

$$\begin{aligned}\langle A, A\rangle = \operatorname{tr}(A^T A) &= \operatorname{tr}\left(\begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22}\end{bmatrix}\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22}\end{bmatrix}\right) \\ &= \operatorname{tr}\left(\begin{bmatrix} a_{11}^2 + a_{21}^2 & a_{11}a_{12} + a_{21}a_{22} \\ a_{11}a_{12} + a_{21}a_{22} & a_{12}^2 + a_{22}^2\end{bmatrix}\right) \\ &= a_{11}^2 + a_{21}^2 + a_{12}^2 + a_{22}^2 \geq 0\end{aligned}$$

Moreover, $\langle A, A\rangle = 0$ if and only if A is the zero matrix. Hence, $\langle\, ,\rangle$ is positive definite.

$$\begin{aligned}\langle A, B\rangle = \operatorname{tr}(B^T A) &= \operatorname{tr}\left(\begin{bmatrix} b_{11} & b_{21} \\ b_{12} & b_{22}\end{bmatrix}\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22}\end{bmatrix}\right) \\ &= \operatorname{tr}\left(\begin{bmatrix} a_{11}b_{11} + a_{21}b_{21} & b_{11}a_{12} + b_{21}a_{22} \\ a_{11}b_{12} + a_{21}b_{22} & a_{12}b_{12} + a_{22}b_{22}\end{bmatrix}\right) \\ &= a_{11}b_{11} + a_{12}b_{12} + a_{21}b_{21} + a_{22}b_{22}\end{aligned}$$

With a similar calculation we find that

$$\langle B, A\rangle = a_{11}b_{11} + a_{12}b_{12} + a_{21}b_{21} + a_{22}b_{22} = \langle A, B\rangle$$

Therefore, $\langle\, ,\rangle$ is symmetric. In Section 8.1 we proved that tr is a linear operator. Hence, we get

$$\begin{aligned}\langle sA + tB, C\rangle = \operatorname{tr}(C^T(sA + tB)) &= \operatorname{tr}(sC^T A + tC^T B) \\ &= s\operatorname{tr}(C^T A) + t\operatorname{tr}(C^T B) = s\langle A, C\rangle + t\langle B, C\rangle\end{aligned}$$

Thus, $\langle\, ,\rangle$ is also bilinear, and so it is an inner product on $M_{2\times 2}(\mathbb{R})$.

REMARKS

1. Of course, this can be generalized to an inner product $\langle A, B\rangle = \operatorname{tr}(B^T A)$ on $M_{m\times n}(\mathbb{R})$. This is called the standard inner product on $M_{m\times n}(\mathbb{R})$.
2. Observe that calculating $\langle A, B\rangle$ corresponds exactly to finding the dot product on the isomorphic vectors in $\mathbb{R}^{mn}$. As a result, when using this inner product you do not actually have to compute $B^T A$.

EXAMPLE 4 Prove that the function $\langle\, ,\rangle$ defined by

$$\langle p(x), q(x)\rangle = p(-1)q(-1) + p(0)q(0) + p(1)q(1)$$

is an inner product on $P_2(\mathbb{R})$.

Solution: Let $p(x) = a + bx + cx^2, q(x), r(x) \in P_2(\mathbb{R})$. Then:

$$\begin{aligned}\langle p(x), p(x)\rangle &= [p(-1)]^2 + [p(0)]^2 + [p(1)]^2 \\ &= [a - b + c]^2 + [a]^2 + [a + b + c]^2 \geq 0\end{aligned}$$

Moreover, $\langle p(x), p(x)\rangle = 0$ if and only if $a - b + c = 0$, $a = 0$, and $a + b + c = 0$, which implies $a = b = c = 0$. Hence, $\langle\, ,\rangle$ is positive definite. We have

$$\begin{aligned}\langle p(x), q(x)\rangle &= p(-1)q(-1) + p(0)q(0) + p(1)q(1) \\ &= q(-1)p(-1) + q(0)p(0) + q(1)p(1) \\ &= \langle q(x), p(x)\rangle\end{aligned}$$

Hence, $\langle\, ,\rangle$ is symmetric. Finally, we get

$$\begin{aligned}\langle (sp + tq)(x), r(x)\rangle =& (sp + tq)(-1)r(-1) + (sp + tq)(0)r(0) + (sp + tq)(1)r(1) \\ =& [sp(-1) + tq(-1)]r(-1) + [sp(0) + tq(0)]r(0) + \\ & + [sp(1) + tq(1)]r(1) \\ =& s[p(-1)r(-1) + p(0)r(0) + p(1)r(1)] + \\ & + t[q(-1)r(-1) + q(0)r(0) + q(1)r(1)] \\ =& s\,\langle p(x), r(x)\rangle + t\,\langle q(x), r(x)\rangle\end{aligned}$$

Therefore, $\langle\, ,\rangle$ is also bilinear and hence is an inner product on $P_2(\mathbb{R})$.

EXAMPLE 5 An extremely important inner product in applied mathematics, physics, and engineering is the inner product

$$\langle f(x), g(x)\rangle = \int_{-\pi}^{\pi} f(x)g(x)\, dx$$

on the vector space $C[-\pi, \pi]$ of continuous functions defined on the closed interval from $-\pi$ to π. This is the foundation for Fourier Series.

THEOREM 1 Let $\mathbb{V}$ be an inner product space with inner product $\langle\, ,\rangle$. Then for any $\vec{v} \in \mathbb{V}$ we have

$$\left\langle \vec{v}, \vec{0}\right\rangle = 0$$

REMARK

Whenever one considers an inner product space one must define which inner product they are using. However, in $\mathbb{R}^n$ and $M_{m\times n}(\mathbb{R})$ one generally uses the standard inner product. Thus, whenever we are working in $\mathbb{R}^n$ or $M_{m\times n}(\mathbb{R})$, if no other inner product is mentioned, we will take this to mean that the standard inner product is being used.

Section 9.1 Problems

1. Calculate the following inner products under the standard inner product on $M_{2\times 2}(\mathbb{R})$.

   (a) $\left\langle \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix} \right\rangle$ (b) $\left\langle \begin{bmatrix} 3 & -4 \\ 2 & 5 \end{bmatrix}, \begin{bmatrix} 4 & 1 \\ 2 & -1 \end{bmatrix} \right\rangle$ (c) $\left\langle \begin{bmatrix} 2 & 3 \\ -4 & 1 \end{bmatrix}, \begin{bmatrix} -5 & 3 \\ 1 & 5 \end{bmatrix} \right\rangle$

2. Calculate the following given that the function
$$\langle p(x), q(x)\rangle = p(0)q(0) + p(1)q(1) + p(2)q(2)$$
defines an inner product on $P_2(\mathbb{R})$.

   (a) $\langle x, 1 + x + x^2\rangle$ (b) $\langle 1 + x^2, 1 + x^2\rangle$ (c) $\langle x + x^2, 1 + x + x^2\rangle$

3. Determine, with proof, which of the following functions defines an inner product on the given vector space.

   (a) On $\mathbb{R}^3$, $\langle \vec{x}, \vec{y}\rangle = 2x_1y_1 + 3x_2y_2 + 4x_3y_3$
   (b) On $P_2(\mathbb{R})$, $\langle p(x), q(x)\rangle = p(-1)q(-1) + p(1)q(1)$
   (c) On $M_{2\times 2}(\mathbb{R})$, $\left\langle \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix}, \begin{bmatrix} b_1 & b_2 \\ b_3 & b_4 \end{bmatrix} \right\rangle = a_1b_1 - a_4b_4 + a_2b_2 - a_3b_3$
   (d) On $P_2(\mathbb{R})$, $\langle p(x), q(x)\rangle = p(-2)q(-2) + p(1)q(1) + p(2)q(2)$
   (e) On $\mathbb{R}^3$, $\langle \vec{x}, \vec{y}\rangle = 2x_1y_1 - x_1y_2 - x_2y_1 + 2x_2y_2 + x_3y_3$
   (f) On $P_2(\mathbb{R})$, $\langle p(x), q(x)\rangle = 2p(-1)q(-1)+2p(0)q(0)+2p(1)q(1)-p(-1)q(1)-p(1)q(-1)$

4. Let $\mathbb{V}$ be an inner product space with inner product $\langle\, ,\rangle$. Prove that $\langle \vec{v}, \vec{0}\rangle = 0$ for any $\vec{v} \in \mathbb{V}$.

5. Let $\langle\, ,\rangle$ denote the standard inner product on $\mathbb{R}^n$ and let $A \in M_{n\times n}(\mathbb{R})$. Prove that
$$\langle A\vec{x}, \vec{y}\rangle = \langle \vec{x}, A^T\vec{y}\rangle$$
for all $\vec{x}, \vec{y} \in \mathbb{R}^n$. [HINT: Recall that $\vec{x}^T\vec{y} = \vec{x}\cdot\vec{y}$.]

9.2 Orthogonality and Length

The concepts of length and orthogonality are fundamental in geometry and have many real world applications. Thus, we now extend these concepts to general inner product spaces.

Length

DEFINITION
Length

Let $\mathbb{V}$ be an inner product space with inner product $\langle\, ,\rangle$. Then for any $\vec{v} \in \mathbb{V}$ we define the **length** (or norm) of $\vec{v}$ by

$$\|\vec{v}\| = \sqrt{\langle \vec{v}, \vec{v}\rangle}$$

Observe that the length of a vector in an inner product space is dependent on the definition of the inner product being used in the same way that the sum of two vectors is dependent on the definition of addition being used. This is demonstrated in the examples below.

EXAMPLE 1

Find the length of $\vec{x} = \begin{bmatrix}1\\0\\2\end{bmatrix}$ in $\mathbb{R}^3$ with the inner product $\langle \vec{x}, \vec{y}\rangle = x_1y_1 + 2x_2y_2 + 4x_3y_3$.

Solution: We have

$$\|\vec{x}\| = \sqrt{\langle \vec{x}, \vec{x}\rangle} = \sqrt{(1)(1) + 2(0)(0) + 4(2)(2)} = \sqrt{17}$$

Observe that under the standard inner product the length of $\vec{x}$ would have been $\sqrt{5}$.

EXAMPLE 2

Let $p(x) = x$ and $q(x) = 2 - 3x^2$ be vectors in $P_2(\mathbb{R})$ with inner product

$$\langle p(x), q(x)\rangle = p(-1)q(-1) + p(0)q(0) + p(1)q(1)$$

Calculate the length of $p(x)$ and the length of $q(x)$.

Solution: We have

$$\begin{aligned}
\|p(x)\| &= \sqrt{\langle p(x), p(x)\rangle} = \sqrt{p(-1)p(-1) + p(0)p(0) + p(1)p(1)} \\
&= \sqrt{(-1)(-1) + 0(0) + 1(1)} = \sqrt{2} \\
\|q(x)\| &= \sqrt{\langle q(x), q(x)\rangle} = \sqrt{q(-1)q(-1) + q(0)q(0) + q(1)q(1)} \\
&= \sqrt{(-1)(-1) + 2(2) + (-1)(-1)} = \sqrt{6}
\end{aligned}$$

EXAMPLE 3 Find the length of $p(x) = x$ in $P_2(\mathbb{R})$ with inner product

$$\langle p(x), q(x)\rangle = p(0)q(0) + p(1)q(1) + p(2)q(2)$$

Solution: We have

$$\begin{aligned} \|p(x)\| &= \sqrt{\langle p(x), p(x)\rangle} = \sqrt{p(0)p(0) + p(1)p(1) + p(2)p(2)} \\ &= \sqrt{0(0) + 1(1) + 2(2)} = \sqrt{5} \end{aligned}$$

EXAMPLE 4 On $C[-\pi, \pi]$ find the length of $f(x) = x$ under the inner product

$$\langle f(x), g(x)\rangle = \int_{-\pi}^{\pi} f(x)g(x)\, dx$$

Solution: We have

$$\|x\|^2 = \int_{-\pi}^{\pi} x^2\, dx = \frac{1}{3}x^3\Big|_{-\pi}^{\pi} = \frac{2}{3}\pi^3$$

Thus, $\|x\| = \sqrt{2\pi^3/3}$.

EXAMPLE 5 Find the length of $A = \begin{bmatrix} 1/2 & 1/2 \\ -1/2 & -1/2 \end{bmatrix}$ in $M_{2\times 2}(\mathbb{R})$.

Solution: Using the standard inner product we get

$$\|A\| = \sqrt{\langle A, A\rangle} = \sqrt{(1/2)^2 + (1/2)^2 + (-1/2)^2 + (-1/2)^2} = 1$$

In many cases it is useful (and sometimes necessary) to use vectors which have length one.

DEFINITION
Unit Vector

Let $\mathbb{V}$ be an inner product space with inner product $\langle\, ,\rangle$. If $\vec{v} \in \mathbb{V}$ is a vector such that $\|\vec{v}\| = 1$, then $\vec{v}$ is called a **unit vector**.

In many situations, we will be given a general vector $\vec{v} \in \mathbb{V}$ and need to find a unit vector in the direction of $\vec{v}$. This is called **normalizing** the vector. The following theorem shows us how to do this.

THEOREM 1 Let $\mathbb{V}$ be an inner product space with inner product $\langle\, ,\rangle$. Then for any $\vec{v} \in \mathbb{V}$ and $t \in \mathbb{R}$, we have

$$\|t\vec{v}\| = |t|\,\|\vec{v}\|$$

Moreover, if $\vec{v} \neq \vec{0}$, then $\hat{v} = \dfrac{1}{\|\vec{v}\|}\vec{v}$ is a unit vector in the direction of $\vec{v}$.

EXAMPLE 6

Find a unit vector in the direction of $p(x) = x$ in $P_2(\mathbb{R})$ with inner product

$$\langle p(x), q(x)\rangle = p(0)q(0) + p(1)q(1) + p(2)q(2)$$

Solution: We have

$$\|p(x)\| = \sqrt{p(0)p(0) + p(1)p(1) + p(2)p(2)} = \sqrt{0(0) + 1(1) + 2(2)} = \sqrt{5}$$

Thus, a unit vector in the direction of p is

$$\hat{p}(x) = \frac{1}{\sqrt{5}}x$$

Orthogonality

DEFINITION

Orthogonal Vectors

Let $\mathbb{V}$ be an inner product space with inner product $\langle\, ,\rangle$. If $\vec{x}, \vec{y} \in \mathbb{V}$ such that

$$\langle \vec{x}, \vec{y}\rangle = 0$$

then $\vec{x}$ and $\vec{y}$ are said to be orthogonal.

EXAMPLE 7

The standard basis vectors for $\mathbb{R}^2$ are not orthogonal under the inner product $\langle \vec{x}, \vec{y}\rangle = 2x_1y_1 + x_1y_2 + x_2y_1 + 2x_2y_2$ for $\mathbb{R}^2$ since

$$\left\langle \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\rangle = 2(1)(0) + 1(1) + 0(0) + 2(0)(1) = 1$$

EXAMPLE 8

Show that $\begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix}$ is orthogonal to $\begin{bmatrix} -1 & 2 \\ -1 & 0 \end{bmatrix}$ in $M_{2\times 2}(\mathbb{R})$.

Solution: $\left\langle \begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix}, \begin{bmatrix} -1 & 2 \\ -1 & 0 \end{bmatrix} \right\rangle = (-1)(1) + 2(2) + (-1)(3) + 0(-1) = 0.$

So, they are orthogonal.

DEFINITION

Orthogonal Set

If $\mathcal{S} = \{\vec{v}_1, \dots, \vec{v}_k\}$ is a set of vectors in an inner product space $\mathbb{V}$ with inner product $\langle\, ,\rangle$ such that $\left\langle \vec{v}_i, \vec{v}_j \right\rangle = 0$ for all $i \neq j$, then $\mathcal{S}$ is called an **orthogonal set**.

EXAMPLE 9

Prove that $\{1, x, 3x^2 - 2\}$ is an orthogonal set of vectors in $P_2(\mathbb{R})$ under the inner product $\langle p(x), q(x)\rangle = p(-1)q(-1) + p(0)q(0) + p(1)q(1)$.

Solution: We have

$$\begin{aligned}\langle 1, x\rangle &= 1(-1) + 1(0) + 1(1) = 0\\ \left\langle 1, 3x^2 - 2\right\rangle &= 1(1) + 1(-2) + 1(1) = 0\\ \left\langle x, 3x^2 - 2\right\rangle &= (-1)(1) + 0(-2) + 1(1) = 0\end{aligned}$$

Hence, each vector is orthogonal to all other vectors so the set is orthogonal.

EXAMPLE 10

Prove that $\{1, x, 3x^2 - 2\}$ is not an orthogonal set of vectors in $P_2(\mathbb{R})$ under the inner product $\langle p(x), q(x)\rangle = p(0)q(0) + p(1)q(1) + p(2)q(2)$.

Solution: We have $\langle 1, x\rangle = 1(0) + 1(1) + 1(2) = 3$ so 1 and x are not orthogonal. Thus, the set is not orthogonal.

EXAMPLE 11

Prove that $\left\{\begin{bmatrix}1 & 2\\ 0 & 1\end{bmatrix}, \begin{bmatrix}0 & 1\\ 0 & -2\end{bmatrix}, \begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix}\right\}$ is an orthogonal set in $M_{2\times 2}(\mathbb{R})$.

Solution: We have

$$\begin{aligned}\left\langle \begin{bmatrix}1 & 2\\ 0 & 1\end{bmatrix}, \begin{bmatrix}0 & 1\\ 0 & -2\end{bmatrix}\right\rangle &= 1(0) + 2(1) + 0(0) + 1(-2) = 0\\ \left\langle \begin{bmatrix}1 & 2\\ 0 & 1\end{bmatrix}, \begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix}\right\rangle &= 1(0) + 2(0) + 0(1) + 1(0) = 0\\ \left\langle \begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix}, \begin{bmatrix}0 & 1\\ 0 & -2\end{bmatrix}\right\rangle &= 0(0) + 0(1) + 1(0) + 0(-2) = 0\end{aligned}$$

Hence, the set is orthogonal.

EXERCISE 1

Determine if $\left\{\begin{bmatrix}3 & 1\\ -1 & 2\end{bmatrix}, \begin{bmatrix}1 & -3\\ 2 & 1\end{bmatrix}, \begin{bmatrix}1 & 2\\ 3 & -1\end{bmatrix}\right\}$ is an orthogonal set in $M_{2\times 2}(\mathbb{R})$.

One important property of orthogonal vectors in $\mathbb{R}^2$ is the Pythagorean Theorem. This extends to an orthogonal set in an inner product space.

THEOREM 2

If $\{\vec{v}_1, \dots, \vec{v}_k\}$ is an orthogonal set in an inner product space $\mathbb{V}$, then

$$\|\vec{v}_1 + \cdots + \vec{v}_k\|^2 = \|\vec{v}_1\|^2 + \cdots + \|\vec{v}_k\|^2$$

The following theorem gives us an important result about orthogonal sets which do not contain the zero vector.

THEOREM 3 Let $S = \{\vec{v}_1, \dots, \vec{v}_k\}$ be an orthogonal set in an inner product space $\mathbb{V}$ with inner product $\langle\,,\rangle$ such that $\vec{v}_i \neq \vec{0}$ for all $1 \leq i \leq k$. Then S is linearly independent.

Proof: Consider

$$\vec{0} = c_1\vec{v}_1 + \cdots + c_k\vec{v}_k$$

Take the inner product of both sides with $\vec{v}_i$ to get

$$\begin{aligned}\left\langle \vec{v}_i, \vec{0}\right\rangle &= \langle \vec{v}_i, c_1\vec{v}_1 + \cdots + c_k\vec{v}_k\rangle \\ 0 &= c_1\langle \vec{v}_i, \vec{v}_1\rangle + \cdots + c_i\langle \vec{v}_i, \vec{v}_i\rangle + \cdots + c_k\langle \vec{v}_i, \vec{v}_k\rangle \\ &= 0 + \cdots + 0 + c_i\|\vec{v}_i\|^2 + 0 + \cdots + 0 \\ &= c_i\|\vec{v}_i\|^2\end{aligned}$$

Since $\vec{v}_i \neq \vec{0}$ we have that $\|\vec{v}_i\| \neq 0$, and so $c_i = 0$ for $1 \leq i \leq k$. Therefore, S is linearly independent. □

Consequently, if we have an orthogonal set of non-zero vectors which spans an inner product space $\mathbb{V}$, then it will be a basis for $\mathbb{V}$. We will call this an **orthogonal basis**. Since the vectors in the basis are all orthogonal to each other, our geometric intuition tells us that it should be quite easy to find the coordinates of any vector with respect to this basis. The following theorem demonstrates this.

THEOREM 4 If $\mathcal{S} = \{\vec{v}_1, \dots, \vec{v}_n\}$ is an orthogonal basis for an inner product space $\mathbb{V}$ with inner product $\langle\,,\rangle$ and $\vec{v} \in \mathbb{V}$, then the coefficient of $\vec{v}_i$ when $\vec{v}$ is written as a linear combination of the vectors in $\mathcal{S}$ is $\dfrac{\langle \vec{v}, \vec{v}_i\rangle}{\|\vec{v}_i\|^2}$. In particular,

$$\vec{v} = \frac{\langle \vec{v}, \vec{v}_1\rangle}{\|\vec{v}_1\|^2}\vec{v}_1 + \cdots + \frac{\langle \vec{v}, \vec{v}_n\rangle}{\|\vec{v}_n\|^2}\vec{v}_n$$

Proof: Since $\vec{v} \in \mathbb{V}$ we can write

$$\vec{v} = c_1\vec{v}_1 + \cdots + c_n\vec{v}_n$$

Taking the inner product of both sides with $\vec{v}_i$ gives

$$\begin{aligned}\langle \vec{v}_i, \vec{v}\rangle &= \langle \vec{v}_i, c_1\vec{v}_1 + \cdots + c_n\vec{v}_n\rangle \\ &= c_1\langle \vec{v}_i, \vec{v}_1\rangle + \cdots + c_i\langle \vec{v}_i, \vec{v}_i\rangle + \cdots + c_n\langle \vec{v}_i, \vec{v}_n\rangle \\ &= 0 + \cdots + 0 + c_i\|\vec{v}_i\|^2 + 0 + \cdots + 0\end{aligned}$$

since $\mathcal{S}$ is orthogonal. Also, $\vec{v}_i \neq \vec{0}$ which implies that $\|\vec{v}_i\|^2 \neq 0$. Therefore, we get

$$c_i = \frac{\langle \vec{v}, \vec{v}_i\rangle}{\|\vec{v}_i\|^2}$$

This is valid for $1 \leq i \leq n$ and so the result follows. □

EXAMPLE 12 Find the coordinates of $\vec{x} = \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}$ in $\mathbb{R}^3$ with respect to the orthogonal basis $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \right\}$.

Solution: By Theorem 4 we have

$$\begin{aligned} c_1 &= \frac{\langle \vec{x}, \vec{v}_1 \rangle}{\|\vec{v}_1\|^2} = \frac{0}{3} = 0 \\ c_2 &= \frac{\langle \vec{x}, \vec{v}_2 \rangle}{\|\vec{v}_2\|^2} = \frac{4}{2} = 2 \\ c_3 &= \frac{\langle \vec{x}, \vec{v}_3 \rangle}{\|\vec{v}_3\|^2} = \frac{-6}{6} = -1 \end{aligned}$$

Thus, $[\vec{x}]_{\mathcal{B}} = \begin{bmatrix} 0 \\ 2 \\ -1 \end{bmatrix}$.

Note that it is easy to verify that

$$\begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix} = 0 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + 2 \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + (-1) \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

EXAMPLE 13 Given that $\mathcal{B} = \left\{ \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} -2 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix} \right\}$ is an orthogonal basis for a subspace $\mathbb{S}$ of $M_{2\times 2}(\mathbb{R})$, find the coordinates of $A = \begin{bmatrix} 4 & 2 \\ 3 & -1 \end{bmatrix}$ with respect to $\mathcal{B}$.

Solution: We have

$$\begin{aligned} c_1 &= \frac{\langle A, \vec{v}_1 \rangle}{\|\vec{v}_1\|^2} = \frac{4(1) + 2(1) + 3(1) + (-1)(1)}{1^2 + 1^2 + 1^2 + 1^2} = \frac{8}{4} = 2 \\ c_2 &= \frac{\langle A, \vec{v}_2 \rangle}{\|\vec{v}_2\|^2} = \frac{4(-2) + 2(0) + 3(1) + (-1)(1)}{(-2)^2 + 0^2 + 1^2 + 1^2} = \frac{-6}{6} = -1 \\ c_3 &= \frac{\langle A, \vec{v}_3 \rangle}{\|\vec{v}_3\|^2} = \frac{4(0) + 2(0) + 3(1) + (-1)(-1)}{(0)^2 + 0^2 + 1^2 + (-1)^2} = \frac{4}{2} = 2 \end{aligned}$$

Thus, $[A]_{\mathcal{B}} = \begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix}$.

Orthonormal Bases

Observe that the formula for the coordinates with respect to an orthogonal basis would be even easier if all the basis vectors were unit vectors. Thus, it is desirable to consider such bases.

DEFINITION
Orthonormal Set

If $S = \{\vec{v}_1, \dots, \vec{v}_k\}$ is an orthogonal set in an inner product space $\mathbb{V}$ such that $\|\vec{v}_i\| = 1$ for $1 \leq i \leq k$, then S is called an **orthonormal set**.

By Theorem 3 an orthonormal set is necessarily linearly independent.

DEFINITION
Orthonormal Basis

A basis for an inner product space $\mathbb{V}$ which is an orthonormal set is called an **orthonormal basis** of $\mathbb{V}$.

EXAMPLE 14

The standard basis of $\mathbb{R}^n$ is an orthonormal basis under the standard inner product.

EXAMPLE 15

The standard basis for $\mathbb{R}^3$ is not an orthonormal basis under the inner product

$$\langle \vec{x}, \vec{y} \rangle = x_1y_1 + 2x_2y_2 + x_3y_3$$

since $\vec{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ is not a unit vector under this inner product.

EXAMPLE 16

Let $\mathcal{B} = \left\{ \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \frac{1}{\sqrt{6}}\begin{bmatrix} 1 \\ -2 \\ -1 \end{bmatrix} \right\}$. Verify that $\mathcal{B}$ is an orthonormal basis in $\mathbb{R}^3$.

Solution: We first verify that each vector is a unit vector. We have

$$\left\langle \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \right\rangle = \frac{1}{3} + \frac{1}{3} + \frac{1}{3} = 1$$

$$\left\langle \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\rangle = \frac{1}{2} + 0 + \frac{1}{2} = 1$$

$$\left\langle \frac{1}{\sqrt{6}}\begin{bmatrix} 1 \\ -2 \\ -1 \end{bmatrix}, \frac{1}{\sqrt{6}}\begin{bmatrix} 1 \\ -2 \\ -1 \end{bmatrix} \right\rangle = \frac{1}{6} + \frac{4}{6} + \frac{1}{6} = 1$$

We now verify that $\mathcal{B}$ is orthogonal.

$$\left\langle \frac{1}{\sqrt{3}}\begin{bmatrix}1\\1\\-1\end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\1\end{bmatrix}\right\rangle = \frac{1}{\sqrt{6}}(1+0-1) = 0$$
$$\left\langle \frac{1}{\sqrt{3}}\begin{bmatrix}1\\1\\-1\end{bmatrix}, \frac{1}{\sqrt{6}}\begin{bmatrix}1\\-2\\-1\end{bmatrix}\right\rangle = \frac{1}{\sqrt{18}}(1-2+1) = 0$$
$$\left\langle \frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\1\end{bmatrix}, \frac{1}{\sqrt{6}}\begin{bmatrix}1\\-2\\-1\end{bmatrix}\right\rangle = \frac{1}{\sqrt{12}}(1+0-1) = 0$$

Therefore, $\mathcal{B}$ is an orthonormal set of 3 vectors in $\mathbb{R}^3$ which implies that it is a basis for $\mathbb{R}^3$.

EXERCISE 2

Show that $\mathcal{B} = \left\{\begin{bmatrix}1/\sqrt{2} & 1/\sqrt{2}\\ 0 & 0\end{bmatrix}, \begin{bmatrix}0 & 0\\ 1/\sqrt{2} & 1/\sqrt{2}\end{bmatrix}, \begin{bmatrix}1/2 & -1/2\\ 1/2 & -1/2\end{bmatrix}\right\}$ is an orthonormal set in $M_{2\times 2}(\mathbb{R})$.

EXERCISE 3

Show that $\mathcal{B} = \{1, x, x^2\}$ is not an orthonormal set for $P_2(\mathbb{R})$ under the inner product

$$\langle p(x), q(x)\rangle = p(-2)q(-2) + p(0)q(0) + p(2)q(2)$$

EXAMPLE 17

Turn $\{1, x, 3x^2-2\}$ into an orthonormal basis for $P_2(\mathbb{R})$ under the inner product

$$\langle p(x), q(x)\rangle = p(-1)q(-1) + p(0)q(0) + p(1)q(1)$$

Solution: In Example 9 we showed that $\{1, x, 3x^2-2\}$ is orthogonal. Hence, it is a linearly independent set of three vectors in $P_2(\mathbb{R})$ and thus is a basis for $P_2(\mathbb{R})$. To turn it into an orthonormal basis, we just need to normalize each vector. We have

$$\begin{aligned}
\|1\| &= \sqrt{1^2+1^2+1^2} = \sqrt{3}\\
\|x\| &= \sqrt{(-1)^2+0^2+1^2} = \sqrt{2}\\
\|3x^2-2\| &= \sqrt{1^2+(-2)^2+1^2} = \sqrt{6}
\end{aligned}$$

Hence, an orthonormal basis for $P_2(\mathbb{R})$ is $\left\{\frac{1}{\sqrt{3}}, \frac{x}{\sqrt{2}}, \frac{3x^2-2}{\sqrt{6}}\right\}$

EXAMPLE 18 Show that the set $\{1, \sin x, \cos x\}$ is an orthogonal set in $C[-\pi, \pi]$ under $\langle f(x), g(x)\rangle = \int_{-\pi}^{\pi} f(x)g(x)\,dx$, then make it an orthonormal set.

Solution:

$$\begin{aligned}\langle 1, \sin x\rangle &= \int_{-\pi}^{\pi} 1 \cdot \sin x\,dx = 0\\ \langle 1, \cos x\rangle &= \int_{-\pi}^{\pi} 1 \cdot \cos x\,dx = 0\\ \langle \sin x, \cos x\rangle &= \int_{-\pi}^{\pi} \sin x \cdot \cos x\,dx = 0\end{aligned}$$

Thus, the set is orthogonal. Next we find that

$$\begin{aligned}\langle 1, 1\rangle &= \int_{-\pi}^{\pi} 1^2\,dx = 2\pi\\ \langle \sin x, \sin x\rangle &= \int_{-\pi}^{\pi} \sin^2 x\,dx = \pi\\ \langle \cos x, \cos x\rangle &= \int_{-\pi}^{\pi} \cos^2 x\,dx = \pi\end{aligned}$$

Hence, $\|1\| = \sqrt{2\pi}$, $\|\sin x\| = \sqrt{\pi}$, and $\|\cos x\| = \sqrt{\pi}$. Therefore, $\left\{\frac{1}{\sqrt{2\pi}}, \frac{\sin x}{\sqrt{\pi}}, \frac{\cos x}{\sqrt{\pi}}\right\}$ is an orthonormal set in $C[-\pi, \pi]$.

The following theorem shows, as predicted, that it is very easy to find coordinates of a vector with respect to an orthonormal basis.

THEOREM 5 If $\vec{v}$ is any vector in an inner product space $\mathbb{V}$ with inner product $\langle\, , \rangle$ and $\mathcal{B} = \{\vec{v}_1, \dots, \vec{v}_n\}$ is an orthonormal basis for $\mathbb{V}$, then

$$\vec{v} = \langle \vec{v}, \vec{v}_1\rangle \vec{v}_1 + \cdots + \langle \vec{v}, \vec{v}_n\rangle \vec{v}_n$$

Proof: By Theorem 4 we have

$$\vec{v} = \frac{\langle \vec{v}, \vec{v}_1\rangle}{\|\vec{v}_1\|^2}\vec{v}_1 + \cdots + \frac{\langle \vec{v}, \vec{v}_n\rangle}{\|\vec{v}_n\|^2}\vec{v}_n = \langle \vec{v}, \vec{v}_1\rangle \vec{v}_1 + \cdots + \langle \vec{v}, \vec{v}_n\rangle \vec{v}_n$$

since $\|\vec{v}_i\| = 1$ for $1 \le i \le n$. $\square$

REMARK

The formula for finding coordinates with respect to an orthonormal basis looks nicer that that for an orthogonal basis. However, in practice, is it not necessarily easier to use as the vectors in an orthonormal basis often contain square roots. Compare the following example to Example 12.

EXAMPLE 19

Find $\begin{bmatrix}1\\2\\3\end{bmatrix}_{\mathcal{B}}$ given that $\mathcal{B} = \left\{\frac{1}{\sqrt{3}}\begin{bmatrix}1\\1\\-1\end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\1\end{bmatrix}, \frac{1}{\sqrt{6}}\begin{bmatrix}1\\-2\\-1\end{bmatrix}\right\}$ is an orthonormal basis for $\mathbb{R}^3$.

Solution: We have

$$c_1 = \langle \vec{x}, \vec{v}_1 \rangle = \left\langle \begin{bmatrix}1\\2\\3\end{bmatrix}, \frac{1}{\sqrt{3}}\begin{bmatrix}1\\1\\-1\end{bmatrix} \right\rangle = 0$$

$$c_2 = \langle \vec{x}, \vec{v}_2 \rangle = \left\langle \begin{bmatrix}1\\2\\3\end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\1\end{bmatrix} \right\rangle = \frac{4}{\sqrt{2}} = 2\sqrt{2}$$

$$c_3 = \langle \vec{x}, \vec{v}_3 \rangle = \left\langle \begin{bmatrix}1\\2\\3\end{bmatrix}, \frac{1}{\sqrt{6}}\begin{bmatrix}1\\-2\\-1\end{bmatrix} \right\rangle = \frac{-6}{\sqrt{6}} = -\sqrt{6}$$

Thus, $[\vec{x}]_{\mathcal{B}} = \begin{bmatrix}0\\2\sqrt{2}\\-\sqrt{6}\end{bmatrix}$.

EXAMPLE 20

Let $\langle p(x), q(x) \rangle = p(-1)q(-1) + p(0)q(0) + p(1)q(1)$ be an inner product for $P_2(\mathbb{R})$. Write $f(x) = 1 + x + x^2$ as a linear combination of the vectors in the orthonormal basis $\mathcal{B} = \left\{\frac{1}{\sqrt{3}}, \frac{x}{\sqrt{2}}, \frac{3x^2-2}{\sqrt{6}}\right\}$.

Solution: By Theorem 5, the coordinates of $\vec{x}$ with respect to $\mathcal{B}$ are

$$c_1 = \langle f, \vec{v}_1 \rangle = 1\left(\frac{1}{\sqrt{3}}\right) + 1\left(\frac{1}{\sqrt{3}}\right) + 3\left(\frac{1}{\sqrt{3}}\right) = \frac{5}{\sqrt{3}}$$

$$c_2 = \langle f, \vec{v}_2 \rangle = 1\left(\frac{-1}{\sqrt{2}}\right) + 1(0) + 3\left(\frac{1}{\sqrt{2}}\right) = \sqrt{2}$$

$$c_3 = \langle f, \vec{v}_3 \rangle = 1\left(\frac{1}{\sqrt{6}}\right) + 1\left(\frac{-2}{\sqrt{6}}\right) + 3\left(\frac{1}{\sqrt{6}}\right) = \frac{2}{\sqrt{6}}$$

Thus,

$$1 + x + x^2 = \frac{5}{\sqrt{3}}\left(\frac{1}{\sqrt{3}}\right) + \sqrt{2}\left(\frac{x}{\sqrt{2}}\right) + \frac{2}{\sqrt{6}}\left(\frac{3x^2-2}{\sqrt{6}}\right)$$

Orthogonal Matrices

We end this section by looking at a very important application of orthonormal bases of $\mathbb{R}^n$.

THEOREM 6

For an $n \times n$ matrix P, the following are equivalent.

(1) The columns of P form an orthonormal basis for $\mathbb{R}^n$.
(2) $P^T = P^{-1}$
(3) The rows of P form an orthonormal basis for $\mathbb{R}^n$.

Proof: Let $P = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix}$ where $\mathcal{B} = \{\vec{v}_1, \dots, \vec{v}_n\}$ is an orthonormal basis for $\mathbb{R}^n$.

(1) $\Leftrightarrow$ (2): By definition of matrix multiplication we have

$$P^T P = \begin{bmatrix} \vec{v}_1^T \\ \vdots \\ \vec{v}_n^T \end{bmatrix} \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix} = \begin{bmatrix} \vec{v}_1 \cdot \vec{v}_1 & \vec{v}_1 \cdot \vec{v}_2 & \cdots & \vec{v}_1 \cdot \vec{v}_n \\ \vec{v}_2 \cdot \vec{v}_1 & \vec{v}_2 \cdot \vec{v}_2 & \cdots & \vec{v}_2 \cdot \vec{v}_n \\ \vdots & \vdots & \ddots & \vdots \\ \vec{v}_n \cdot \vec{v}_1 & \vec{v}_n \cdot \vec{v}_2 & \cdots & \vec{v}_n \cdot \vec{v}_n \end{bmatrix}$$

Therefore $P^T P = I$ if and only if $\vec{v}_i \cdot \vec{v}_j = 0$ whenever $i \neq j$ and $\vec{v}_i \cdot \vec{v}_i = 1$.

(2) $\Leftrightarrow$ (3): The rows of P form an orthonormal basis for $\mathbb{R}^n$ if and only if the columns of P^T from an orthonormal basis for $\mathbb{R}^n$. We proved above that this is true if and only if

$$\begin{aligned} (P^T)^T &= (P^T)^{-1} \\ P &= (P^{-1})^T \\ P^T &= \left((P^{-1})^T\right)^T \\ P^T &= P^{-1} \end{aligned}$$

as required. $\square$

DEFINITION

Orthogonal Matrix

Let P be an $n \times n$ matrix such that the columns of P form an orthonormal basis for $\mathbb{R}^n$. Then P is called an **orthogonal matrix**.

REMARKS

1. Be very careful to remember that an orthogonal matrix has *orthonormal rows and columns*. A matrix whose columns form an orthogonal set but not an orthonormal set is not orthogonal!

2. By Theorem 6 we could have defined an orthogonal matrix as a matrix P such that $P^T = P^{-1}$ instead. We will use this property of orthogonal matrices frequently.

EXAMPLE 21 Which of the following matrices are orthogonal.

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \qquad B \begin{bmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{bmatrix}, \quad C = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/\sqrt{6} & -1/\sqrt{6} & 2/\sqrt{6} \\ -1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \end{bmatrix}$$

Solution: The columns of A are the standard basis vectors for $\mathbb{R}^3$ which we know is an orthonormal basis for $\mathbb{R}^3$. Thus, A is orthogonal.

Although the columns of B are clearly orthogonal, we have that

$$\left\| \begin{bmatrix} 1/2 \\ -1/2 \end{bmatrix} \right\| = \sqrt{(1/2)^2 + (-1/2)^2} = \sqrt{1/2} \neq 1$$

Thus, the first column is not a unit vector, so B is not orthogonal.

By matrix multiplication, we find that

$$CC^T = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/\sqrt{6} & -1/\sqrt{6} & 2/\sqrt{6} \\ -1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{6} & -1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3} \end{bmatrix} = I$$

Hence C is orthogonal.

We now look at some useful properties of orthogonal matrices.

THEOREM 7 Let P and Q be $n \times n$ orthogonal matrices and $\vec{x}, \vec{y} \in \mathbb{R}^n$. Then

(1) $(P\vec{x}) \cdot (P\vec{y}) = \vec{x} \cdot \vec{y}$.
(2) $\|P\vec{x}\| = \|\vec{x}\|$.
(3) $\det P = \pm 1$.
(4) All real eigenvalues of P are 1 or -1.
(5) PQ is an orthogonal matrix.

Proof: We will prove (1) and (2) and leave the others as exercises. We have

$$(P\vec{x}) \cdot (P\vec{y}) = (P\vec{x})^T (P\vec{y}) = \vec{x}^T P^T P\vec{y} = \vec{x}^T \vec{y} = \vec{x} \cdot \vec{y}$$

Thus,

$$\|P\vec{x}\|^2 = (P\vec{x}) \cdot (P\vec{x}) = \vec{x} \cdot \vec{x} = \|\vec{x}\|^2$$

as required. $\square$

Section 9.2 Problems

1. Consider the subspace Span $\mathcal{B}$ in $M_{2\times 2}(\mathbb{R})$ where

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}, \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}, \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \right\}$$

   (a) Prove that $\mathcal{B}$ is an orthogonal basis for $\mathbb{S}$.

   (b) Turn $\mathcal{B}$ into an orthonormal basis C by normalizing the vectors in $\mathcal{B}$.

   (c) Find the coordinates of $\vec{x} = \begin{bmatrix} -3 & 6 \\ -2 & -1 \end{bmatrix}$ with respect to the orthonormal basis C.

2. Consider the inner product $\langle\, , \rangle$ on $P_2(\mathbb{R})$ defined by

$$\langle p(x), q(x)\rangle = p(-1)q(-1) + p(0)q(0) + p(1)q(1)$$

   Let $\mathcal{B} = \{1 + x, -2 + 3x, 2 - 3x^2\}$

   (a) Prove that $\mathcal{B}$ is an orthogonal basis for $P_2(\mathbb{R})$.

   (b) Find the coordinates of $5 + x - 3x^2$ with respect to $\mathcal{B}$.

3. Consider the inner product $\langle\, , \rangle$ on defined $\mathbb{R}^3$ by $\langle \vec{x}, \vec{y}\rangle = x_1y_1 + 2x_2y_2 + x_3y_3$. Let $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 3 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 7 \end{bmatrix} \right\}$

   (a) Prove that $\mathcal{B}$ is an orthogonal basis for $\mathbb{R}^3$.

   (b) Turn $\mathcal{B}$ into an orthonormal basis C by normalizing the vectors in $\mathcal{B}$.

   (c) Find the coordinates of $\vec{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ with respect to the orthonormal basis C.

4. Which of the following matrices are orthogonal.

   (a) $\begin{bmatrix} 1/\sqrt{10} & -3/\sqrt{10} \\ 3/\sqrt{10} & 1/\sqrt{10} \end{bmatrix}$ (b) $\begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}$ (c) $\begin{bmatrix} 2/3 & 1/3 & 2/3 \\ -1/\sqrt{18} & 4/\sqrt{18} & -1/\sqrt{18} \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix}$

   (d) $\begin{bmatrix} 1 & 1 & 1 \\ 2 & -1 & 0 \\ 1 & 1 & -1 \end{bmatrix}$ (e) $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$ (f) $\begin{bmatrix} 2/\sqrt{20} & 3/\sqrt{11} & 0 \\ 3/\sqrt{20} & -1/\sqrt{11} & -1/\sqrt{2} \\ 3/\sqrt{20} & -1/\sqrt{11} & 1/\sqrt{2} \end{bmatrix}$

5. Consider $P_2(\mathbb{R})$ with inner product $\langle p(x), q(x)\rangle = p(-1)q(-1) + p(0)q(0) + p(1)q(1)$. Given that $B = \{1 - x^2, \frac{1}{2}(x - x^2)\}$ is an orthonormal set, extend B to find an orthonormal basis for $P_2(\mathbb{R})$.

6. Let $\mathbb{V}$ be an inner product space with inner product $\langle\, , \rangle$. Prove that for any $\vec{v} \in \mathbb{V}$ and $t \in \mathbb{R}$, we have

$$\|t\vec{v}\| = |t|\, \|\vec{v}\|$$

7. Let P and Q be $n \times n$ orthogonal matrices

   (a) Prove that $\det P = \pm 1$.

   (b) Prove that all real eigenvalues of P are ± 1.

   (c) Give an orthogonal matrix P whose eigenvalues are not ± 1.

   (d) Prove that PQ is an orthogonal matrix.

8. Given that $\mathcal{B} = \left\{ \begin{bmatrix} 1/\sqrt{6} \\ -2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}, \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}, \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{bmatrix} \right\}$ is an orthonormal basis for $\mathbb{R}^3$. Using $\mathcal{B}$, determine another orthonormal basis for $\mathbb{R}^3$ which includes the vector $\begin{bmatrix} 1/\sqrt{6} \\ 1/\sqrt{3} \\ 1/\sqrt{2} \end{bmatrix}$, and briefly explain why your basis is orthonormal.

9. Let $\{\vec{v}_1, \ldots, \vec{v}_m\}$ be a set of m orthonormal vectors in $\mathbb{R}^n$ with $m < n$. Let $Q = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_m \end{bmatrix}$. Prove that $Q^T Q = I_m$.

10. Let $\{\vec{v}_1, \ldots, \vec{v}_n\}$ be an orthonormal basis for an inner product space $\mathbb{V}$ with inner product $\langle\,,\rangle$ and let $\vec{x} = c_1\vec{v}_1 + \cdots + c_n\vec{v}_n$ and $\vec{y} = d_1\vec{v}_1 + \cdots + d_n\vec{v}_n$. Show that

$$\langle \vec{x}, \vec{y} \rangle = c_1 d_1 + \cdots + c_n d_n \quad \text{and} \quad ||\vec{x}||^2 = c_1^2 + \cdots + c_n^2$$

9.3 The Gram-Schmidt Procedure

We saw in Math 136 that every finite dimensional vector space has a basis. In this section, we derive an algorithm for changing a basis for a given inner product space into an orthogonal basis. Once we have an orthogonal basis, it is easy to normalize each vector to make it orthonormal.

Let $\mathbb{W}$ be a n-dimensional inner product space and let $\{\vec{w}_1, \ldots, \vec{w}_n\}$ be a basis $\mathbb{W}$. We want to find an orthogonal basis $\{\vec{v}_1, \ldots, \vec{v}_n\}$ for $\mathbb{W}$.

To develop a standard procedure, we will look at a few simple cases and then generalize.

First, consider the case where $\{\vec{w}_1\}$ is a basis for $\mathbb{W}$. In this case, we can take $\vec{v}_1 = \vec{w}_1$ so that $\{\vec{v}_1\}$ is an orthogonal basis for $\mathbb{W}$.

Now, consider the case where $\{\vec{w}_1, \vec{w}_2\}$ is a basis for $\mathbb{W}$. Starting as in the case above we let $\vec{v}_1 = \vec{w}_1$. We then need to pick $\vec{v}_2$ so that it is orthogonal to $\vec{v}_1$ and $\{\vec{v}_1, \vec{v}_2\}$ spans $\mathbb{W}$. To see how to find such a vector $\vec{v}_2$ we work backwards. Assume that we have an orthogonal set $\{\vec{v}_1, \vec{v}_2\}$ which spans $\mathbb{W}$, then we must have that $\vec{w}_2 \in \operatorname{Span}\{\vec{v}_1, \vec{v}_2\}$. If this is true, then, from our work with coordinates with respect to an orthogonal basis, we find that

$$\vec{w}_2 = \frac{< \vec{w}_2, \vec{v}_1 >}{||\vec{v}_1||^2}\vec{v}_1 + \frac{< \vec{w}_2, \vec{v}_2 >}{||\vec{v}_2||^2}\vec{v}_2$$

where $\frac{< \vec{w}_2, \vec{v}_2 >}{||\vec{v}_2||^2} \neq 0$ since $\vec{w}_2$ is not a scalar multiple of $\vec{w}_1 = \vec{v}_1$. Rearranging gives

$$\frac{< \vec{w}_2, \vec{v}_2 >}{||\vec{v}_2||^2}\vec{v}_2 = \vec{w}_2 - \frac{< \vec{w}_2, \vec{v}_1 >}{||\vec{v}_1||^2}\vec{v}_1$$

Multiplying by a non-zero scalar does not change orthogonality or spanning, so we can take any scalar multiple of $\vec{v}_2$. For simplicity, we take

$$\vec{v}_2 = \vec{w}_2 - \frac{< \vec{w}_2, \vec{v}_1 >}{||\vec{v}_1||^2}\vec{v}_1$$

Consequently, we have that $\{\vec{v}_1, \vec{v}_2\}$ is an orthogonal basis for $\mathbb{W}$.

For the case where $\{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$ is a basis for $\mathbb{W}$, we can repeat the same procedure. We start by picking $\vec{v}_1$ and $\vec{v}_2$ as above. We then need to find $\vec{v}_3$ such that $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is orthogonal and spans $\mathbb{W}$. Using the same argument as in the previous case, we get

$$\vec{v}_3 = \vec{w}_3 - \frac{< \vec{w}_3, \vec{v}_1 >}{||\vec{v}_1||^2}\vec{v}_1 - \frac{< \vec{w}_3, \vec{v}_2 >}{||\vec{v}_2||^2}\vec{v}_2$$

and $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is an orthogonal basis for $\mathbb{W}$. Repeating this algorithm for the case $\mathbb{W} = \operatorname{Span}\{\vec{w}_1, \ldots, \vec{w}_k\}$ gives us the following result.

THEOREM 1 **(Gram-Schmidt Orthogonalization)**

Let $\mathbb{W}$ be a subspace of an inner product space with basis $\{\vec{w}_1, \ldots, \vec{w}_k\}$. Define $\vec{v}_1, \ldots, \vec{v}_k$ successively as follows:

$$\begin{aligned}
\vec{v}_1 &= \vec{w}_1 \\
\vec{v}_2 &= \vec{w}_2 - \frac{< \vec{w}_2, \vec{v}_1 >}{||\vec{v}_1||^2}\vec{v}_1 \\
\vec{v}_i &= \vec{w}_i - \frac{< \vec{w}_i, \vec{v}_1 >}{||\vec{v}_1||^2}\vec{v}_1 - \frac{< \vec{w}_i, \vec{v}_2 >}{||\vec{v}_2||^2}\vec{v}_2 - \cdots - \frac{< \vec{w}_i, \vec{v}_{i-1} >}{||\vec{v}_{i-1}||^2}\vec{v}_{i-1}
\end{aligned}$$

for $3 \leq i \leq k$. Then $\operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_i\} = \operatorname{Span}\{\vec{w}_1, \ldots, \vec{w}_i\}$ for $1 \leq i \leq k$. In particular, $\{\vec{v}_1, \ldots, \vec{v}_k\}$ is an orthogonal basis of $\mathbb{W}$.

EXAMPLE 1 Use the Gram-Schmidt procedure to transform $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right\}$ into an orthonormal basis for $\mathbb{R}^3$.

Solution: Denote the vectors in $\mathcal{B}$ by $\vec{w}_1$, $\vec{w}_2$, and $\vec{w}_3$ respectively. We first take $\vec{v}_1 = \vec{w}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. Then

$$\vec{w}_2 - \frac{\langle \vec{w}_2, \vec{v}_1 \rangle}{||\vec{v}_1||^2}\vec{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} - \frac{1}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2/3 \\ -4/3 \\ 2/3 \end{bmatrix}$$

We can take any scalar multiple of this, so we take $\vec{v}_2 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$ to simplify the next calculation. Next,

$$\begin{aligned} \vec{v}_3 &= \vec{w}_3 - \frac{\langle \vec{w}_3, \vec{v}_1 \rangle}{\|\vec{v}_1\|^2}\vec{v}_1 - \frac{\langle \vec{w}_3, \vec{v}_2 \rangle}{\|\vec{v}_2\|^2}\vec{v}_2 \\ &= \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - \frac{2}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \frac{1}{6}\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 0 \\ -1/2 \end{bmatrix} \end{aligned}$$

Then $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is an orthogonal basis for $\mathbb{R}^3$. To get an orthonormal basis, we normalize each vector. We get

$$\hat{v}_1 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \qquad \hat{v}_2 = \frac{1}{\sqrt{6}}\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \qquad \hat{v}_3 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$

So, $\{\hat{v}_1, \hat{v}_2, \hat{v}_3\}$ is an orthonormal basis for $\mathbb{R}^3$.

EXAMPLE 2

Use the Gram-Schmidt procedure to transform $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right\}$ into an orthogonal basis for $\mathbb{R}^3$ with inner product

$$\langle \vec{x}, \vec{y} \rangle = 2x_1y_1 + x_2y_2 + 3x_3y_3$$

Solution: Take $\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. Then

$$\vec{w}_2 - \frac{\langle \vec{w}_2, \vec{v}_1 \rangle}{\|\vec{v}_1\|^2}\vec{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} - \frac{4}{6}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/3 \\ -5/3 \\ 1/3 \end{bmatrix}$$

To simplify calculations we take $\vec{v}_2 = \begin{bmatrix} 1 \\ -5 \\ 1 \end{bmatrix}$. Next,

$$\begin{aligned} \vec{v}_3 &= \vec{w}_3 - \frac{\langle \vec{w}_3, \vec{v}_1 \rangle}{\|\vec{v}_1\|^2}\vec{v}_1 - \frac{\langle \vec{w}_3, \vec{v}_2 \rangle}{\|\vec{v}_2\|^2}\vec{v}_2 \\ &= \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - \frac{3}{6}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \frac{3}{30}\begin{bmatrix} 1 \\ -5 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{3}{5} \\ 0 \\ -\frac{2}{5} \end{bmatrix} \end{aligned}$$

Thus, an orthogonal basis is $\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -5 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ -2 \end{bmatrix} \right\}$.

EXAMPLE 3 Find an orthogonal basis of $P_3(\mathbb{R})$ with inner product $\langle p, q\rangle = \int_{-1}^{1} p(x)q(x)\, dx$ by applying the Gram-Schmidt procedure to the basis $\{1, x, x^2, x^3\}$.

Solution: Take $\vec{v}_1 = 1$. Then

$$\vec{v}_2 = x - \frac{\langle x, 1\rangle}{\|1\|^2}1 = x - \frac{0}{2}1 = x$$

$$\vec{v}_3 = x^2 - \frac{\left\langle x^2, 1\right\rangle}{\|1\|^2}1 - \frac{\left\langle x^2, x\right\rangle}{\|x\|^2}x = x^2 - \frac{\frac{2}{3}}{2}1 - \frac{0}{\frac{2}{3}}x = x^2 - \frac{1}{3}$$

We instead take $\vec{v}_3 = 3x^2 - 1$ to make calculations easier. Finally, we get

$$\vec{v}_4 = x^3 - \frac{\left\langle x^3, 1\right\rangle}{\|1\|^2}1 - \frac{\left\langle x^3, x\right\rangle}{\|x\|^2}x - \frac{\left\langle x^3, 3x^2-1\right\rangle}{\|3x^2-1\|^2}(3x^2-1) = x^3 - \frac{3}{5}x$$

Hence, an orthogonal basis for this inner product space $\left\{1, x, x^2 - \frac{1}{3}, x^3 - \frac{3}{5}x\right\}$.

EXAMPLE 4 Use the Gram-Schmidt procedure to find an orthogonal basis for the subspace $\mathbb{W}$ of $M_{2\times 2}(\mathbb{R})$ spanned by

$$\{\vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{w}_4\} = \left\{\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right\}$$

under the standard inner product.

Solution: Take $\vec{v}_1 = \vec{w}_1$. Then

$$\vec{w}_2 - \frac{\langle \vec{w}_2, \vec{v}_1\rangle}{\|\vec{v}_1\|^2}\vec{v}_1 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} - \frac{2}{3}\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1/3 & -2/3 \\ 1 & 1/3 \end{bmatrix}$$

So, we take $\vec{v}_2 = \begin{bmatrix} 1 & -2 \\ 3 & 1 \end{bmatrix}$. Next we get

$$\vec{v}_3 = \vec{w}_3 - \frac{\langle \vec{w}_3, \vec{v}_1\rangle}{\|\vec{v}_1\|^2}\vec{v}_1 - \frac{\langle \vec{w}_3, \vec{v}_2\rangle}{\|\vec{v}_2\|^2}\vec{v}_2 = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} - \frac{5}{3}\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} - \frac{5}{15}\begin{bmatrix} 1 & -2 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

At first glance, we may think something has gone wrong since the zero vector cannot be a member of the basis. However, if we look closely, we see that this implies that $\vec{w}_3$ is a linear combination of $\vec{w}_1$ and $\vec{w}_2$. Hence, we have

$$\operatorname{Span}\{\vec{w}_1, \vec{w}_2, \vec{w}_4\} = \operatorname{Span}\{\vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{w}_4\}$$

Therefore, we can ignore $\vec{w}_3$ and continue the procedure using $\vec{w}_4$.

$$\vec{v}_3 = \vec{w}_4 - \frac{\langle \vec{w}_4, \vec{v}_1\rangle}{\|\vec{v}_1\|^2}\vec{v}_1 - \frac{\langle \vec{w}_4, \vec{v}_2\rangle}{\|\vec{v}_2\|^2}\vec{v}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \frac{2}{3}\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} - \frac{2}{15}\begin{bmatrix} 1 & -2 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 1/5 & -2/5 \\ -2/5 & 1/5 \end{bmatrix}$$

Hence an orthogonal basis for $\mathbb{W}$ is $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$.

Section 9.3 Problems

1. Use the Gram-Schmidt procedure to transform $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} \right\}$ into an orthogonal basis for $\mathbb{R}^3$.

2. Consider $P_2(\mathbb{R})$ with the inner product
$$\langle p(x), q(x) \rangle = p(-1)q(-1) + p(0)q(0) + p(1)q(1)$$
Use the Gram-Schmidt procedure to transform $\mathcal{S} = \{1, x, x^2\}$ into an orthonormal basis for $P_2(\mathbb{R})$.

3. Let $\mathcal{B} = \left\{ \begin{bmatrix} -1 & 1 \\ 2 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ -1 & -1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right\}$. Find an orthogonal basis for the subspace $\mathbb{S} = \operatorname{Span} \mathcal{B}$ of $M_{2\times 2}(\mathbb{R})$.

4. Suppose $P_2(\mathbb{R})$ has an inner product $\langle\, , \rangle$ which satisfies the following:
$$\langle 1, 1 \rangle = 2 \quad \langle 1, x \rangle = 2 \quad \langle 1, x^2 \rangle = -2$$
$$\langle x, x \rangle = 4 \quad \langle x, x^2 \rangle = -2 \quad \langle x^2, x^2 \rangle = 3$$
Given that $\mathcal{B} = \{x, 2x^2 + x, 2\}$ is a basis of $P_2(\mathbb{R})$. Apply the Gram-Schmidt procedure to this basis to find an orthogonal basis of $P_2(\mathbb{R})$.

9.4 General Projections

In Math 136 we saw how to calculate the projection of a vector onto a plane in $\mathbb{R}^3$. Such projections have many important uses. So, our goal now is to extend projections to general inner product spaces.

We will do this by mimicking what we did with projections in $\mathbb{R}^n$. We first recall from Math 136 that given a vector $\vec{x} \in \mathbb{R}^3$ and a plane P in $\mathbb{R}^3$ which passes through the origin, we wanted to write $\vec{x}$ as a the sum of a vector in P and a vector orthogonal to every vector in P. Therefore, given a subspace $\mathbb{W}$ of an inner product space $\mathbb{V}$ and any vector $\vec{v} \in \mathbb{V}$, we want to find a vector $\operatorname{proj}_{\mathbb{W}} \vec{v}$ in $\mathbb{W}$ and a vector $\operatorname{perp}_{\mathbb{W}} \vec{v}$ which is orthogonal to every vector in $\mathbb{W}$ such that
$$\vec{v} = \operatorname{proj}_{\mathbb{W}} \vec{v} + \operatorname{perp}_{\mathbb{W}} \vec{v}$$

In the case of the plane in $\mathbb{R}^3$, we knew that the orthogonal vector was a scalar multiple of the normal vector. In the general case, we need to start by looking at the set of vectors which are orthogonal to every vector in $\mathbb{W}$.

DEFINITION
Orthogonal Complement

Let $\mathbb{W}$ be a subspace of an inner product space $\mathbb{V}$. The **orthogonal complement** $\mathbb{W}^\perp$ of $\mathbb{W}$ in $\mathbb{V}$ is defined by
$$\mathbb{W}^\perp = \{\vec{v} \in \mathbb{V} \mid \langle \vec{w}, \vec{v} \rangle = 0 \text{ for all } \vec{w} \in \mathbb{W}\}$$

EXAMPLE 1 Let $\mathbb{W} = \operatorname{Span}\left\{\begin{bmatrix}1\\0\\0\\1\end{bmatrix}, \begin{bmatrix}1\\1\\0\\0\end{bmatrix}\right\}$. Find $\mathbb{W}^{\perp}$ in $\mathbb{R}^4$.

Solution: We want to find all $\vec{v} = \begin{bmatrix}v_1\\v_2\\v_3\\v_4\end{bmatrix}$ such that

$$0 = \left\langle \begin{bmatrix}v_1\\v_2\\v_3\\v_4\end{bmatrix}, \begin{bmatrix}1\\0\\0\\1\end{bmatrix} \right\rangle = v_1 + v_4$$

$$0 = \left\langle \begin{bmatrix}v_1\\v_2\\v_3\\v_4\end{bmatrix}, \begin{bmatrix}1\\1\\0\\0\end{bmatrix} \right\rangle = v_1 + v_2$$

Solving this homogeneous system of two equations we find that the general solution is $\vec{v} = s\begin{bmatrix}0\\0\\1\\0\end{bmatrix} + t\begin{bmatrix}-1\\1\\0\\1\end{bmatrix}$, $s, t \in \mathbb{R}$. Thus,

$$\mathbb{W}^{\perp} = \operatorname{Span}\left\{\begin{bmatrix}0\\0\\1\\0\end{bmatrix}, \begin{bmatrix}-1\\1\\0\\1\end{bmatrix}\right\}$$

EXAMPLE 2 Let $\mathbb{W} = \operatorname{Span}\{x\}$ in $P_2(\mathbb{R})$ with $\langle p(x), q(x)\rangle = \int_0^1 p(x)q(x)\,dx$. Find $\mathbb{W}^{\perp}$.

Solution: We want to find all $p(x) = a + bx + cx^2$ such that $\langle p, x\rangle = 0$. This gives

$$0 = \int_0^1 ax + bx^2 + cx^3 = \frac{1}{2}a + \frac{1}{3}b + \frac{1}{4}c$$

So $c = -2a - \frac{4}{3}b$. Thus every vector in the orthogonal compliment has the form $a + bx - \left(2a - \frac{4}{3}b\right)x^2 = a(1 - 2x^2) + b\left(x - \frac{4}{3}x^2\right)$. Thus we get

$$\mathbb{W}^{\perp} = \operatorname{Span}\{1 - 2x^2, 3x - 4x^2\}$$

The following theorem shows that for a vector $\vec{x}$ to be in $\mathbb{W}^{\perp}$ it only need be orthogonal to each of the basis vectors of $\mathbb{W}$.

THEOREM 1 Let $\{\vec{v}_1, \ldots, \vec{v}_k\}$ be a basis for a subspace $\mathbb{W}$ of an inner product space $\mathbb{V}$. If $\langle \vec{x}, \vec{v}_i \rangle = 0$ for $1 \leq i \leq k$, then $\vec{x} \in \mathbb{W}^\perp$.

The following theorem gives some important properties of the othogonal compliment which we will require later.

THEOREM 2 Let $\mathbb{W}$ be a finite dimensional subspace of an inner product space $\mathbb{V}$. Then

(1) $\mathbb{W}^\perp$ is a subspace of $\mathbb{V}$.
(2) If $\dim \mathbb{V} = n$, then $\dim \mathbb{W}^\perp = n - \dim \mathbb{W}$.
(3) If $\dim \mathbb{V} = n$, then $(\mathbb{W}^\perp)^\perp = \mathbb{W}$.
(4) $\mathbb{W} \cap \mathbb{W}^\perp = \{\vec{0}\}$.
(5) If $\dim \mathbb{V} = n$, $\{\vec{v}_1, \ldots, \vec{v}_k\}$ is an orthogonal basis for $\mathbb{W}$, and $\{\vec{v}_{k+1}, \ldots, \vec{v}_n\}$ is an orthogonal basis for $\mathbb{W}^\perp$, then $\{\vec{v}_1, \ldots, \vec{v}_k, \vec{v}_{k+1}, \ldots, \vec{v}_n\}$ is an orthogonal basis for $\mathbb{V}$.

Proof: For (1), we apply the Subspace Test. By definition $\mathbb{W}$ is a subset of $\mathbb{V}$. Also, $\vec{0} \in \mathbb{W}^\perp$ since $\langle \vec{0}, \vec{w} \rangle = 0$ for all $\vec{w} \in \mathbb{W}$ by Theorem 3.2.1. Let $\vec{u}, \vec{v} \in \mathbb{W}^\perp$ and $t \in \mathbb{R}$. Then for any $\vec{w} \in \mathbb{W}$ we have

$$\langle \vec{u} + \vec{v}, \vec{w} \rangle = \langle \vec{u}, \vec{w} \rangle + \langle \vec{v}, \vec{w} \rangle = 0 + 0 = 0$$

and

$$\langle t\vec{u}, \vec{w} \rangle = t \langle \vec{u}, \vec{w} \rangle = t(0) = 0$$

Thus, $\vec{u} + \vec{v} \in \mathbb{W}^\perp$ and $t\vec{u} \in \mathbb{W}^\perp$, so $\mathbb{W}^\perp$ is a subspace of $\mathbb{V}$.

For (2), let $\{\vec{v}_1, \ldots, \vec{v}_k\}$ be an orthonormal basis for $\mathbb{W}$. We can extend this to a basis for $\mathbb{V}$ and then apply the Gram-Schmidt procedure to get an orthonormal basis $\{\vec{v}_1, \ldots, \vec{v}_n\}$ for $\mathbb{V}$. Then, for any $\vec{x} \in \mathbb{W}^\perp$ we have $\vec{x} \in \mathbb{V}$ so we can write

$$\begin{aligned}
\vec{x} &= \langle \vec{x}, \vec{v}_1 \rangle \vec{v}_1 + \cdots + \langle \vec{x}, \vec{v}_k \rangle \vec{v}_k + \langle \vec{x}, \vec{v}_{k+1} \rangle \vec{v}_{k+1} + \cdots + \langle \vec{x}, \vec{v}_n \rangle \vec{v}_n \\
&= \vec{0} + \cdots + \vec{0} + \langle \vec{x}, \vec{v}_{k+1} \rangle \vec{v}_{k+1} + \cdots + \langle \vec{x}, \vec{v}_n \rangle \vec{v}_n
\end{aligned}$$

Hence, $\mathbb{W}^\perp = \operatorname{Span}\{\vec{v}_{k+1}, \ldots, \vec{v}_n\}$. Moreover, $\{\vec{v}_{k+1}, \ldots, \vec{v}_n\}$ is linearly independent since it is orthonormal. Thus, $\dim \mathbb{W}^\perp = n - k = n - \dim \mathbb{W}$.

For (3), if $\vec{w} \in \mathbb{W}$, then $\langle \vec{w}, \vec{v} \rangle = 0$ for all $\vec{v} \in \mathbb{W}^\perp$ by definition of $\mathbb{W}^\perp$. Hence $\mathbb{W} \in (\mathbb{W}^\perp)^\perp$. So $\mathbb{W} \subseteq (\mathbb{W}^\perp)^\perp$ Also, $\dim \mathbb{W}^{\perp\perp} = n - \dim \mathbb{W}^\perp = n - (n - \dim \mathbb{W}) = \dim \mathbb{W}$. Hence $\mathbb{W} = (\mathbb{W}^\perp)^\perp$.

For (4), if $\vec{x} \in \mathbb{W} \cap \mathbb{W}^\perp$, then $\langle \vec{x}, \vec{x} \rangle = 0$ and hence $\vec{x} = \vec{0}$. Therefore, $\mathbb{W} \cap \mathbb{W}^\perp = \{\vec{0}\}$.

For (5), we have that $\{\vec{v}_1, \ldots, \vec{v}_k, \vec{v}_{k+1}, \ldots, \vec{v}_n\}$ is an orthogonal set of non-zero vectors in $\mathbb{V}$ since every vector in $\mathbb{W}^\perp$ is orthogonal to every vector $\mathbb{W}$. Moreover, it contains n vectors by (4). Hence, we have that it is linearly independent set of n vectors. Thus, it is an orthogonal basis for $\mathbb{V}$. □

Property (3) may seem obvious at first glance which may make us wonder why the condition that $\dim \mathbb{V} = n$ was necessary. This is because this property may not be true in an infinite dimensional inner product space! That is, we can have $(\mathbb{W}^\perp)^\perp \neq \mathbb{W}$. We demonstrate this with an example.

EXAMPLE 3 Let $\mathbb{P}$ be the inner product space of all real polynomials with inner product

$$\left\langle a_0 + a_1 x + \cdots + a_k x^k, b_0 + b_1 x + \cdots + b_l x^l \right\rangle = a_0 b_0 + a_1 b_1 + \cdots + a_p b_p$$

where $p = \min(k, l)$. Now, let $\mathbb{U}$ be the subspace

$$\mathbb{U} = \{g(x) \in \mathbb{P} \mid g(1) = 0\}$$

We claim that $\mathbb{U}^\perp = \{0\}$. Let $f(x) \in \mathbb{U}^\perp$, say $f(x) = a_0 + a_1 x + \cdots + a_n x^n$. Note that $\langle f, x^{n+1} \rangle = 0$. Define $g(x) = f(x) - f(1)x^{n+1}$. We get that $g(1) = 0$ and hence $g \in \mathbb{U}$. Therefore, we have

$$0 = \langle f, g \rangle = \langle f, f - f(1)x^{n+1} \rangle = \langle f, f \rangle - f(1)\langle f, x^{n+1} \rangle = \langle f, f \rangle$$

and hence $f(x) = 0$.

Of course $(\mathbb{U}^\perp)^\perp = \mathbb{P}$, since $\langle \vec{0}, \vec{v} \rangle = 0$ for every $\vec{v} \in \mathbb{V}$. Therefore, $(\mathbb{U}^\perp)^\perp \neq \mathbb{U}$.

Notice that, as in the proof of property (3) above, we do have that $\mathbb{U} \subset (\mathbb{U}^\perp)^\perp$.

We now return to our purpose of looking at orthogonal complements, to define the projection of a vector $\vec{v}$ onto a subspace $\mathbb{W}$ of an inner product space $\mathbb{V}$. We stated that we want to find $\operatorname{proj}_{\mathbb{W}} \vec{v}$ and $\operatorname{perp}_{\mathbb{W}} \vec{v}$ such that $\vec{v} = \operatorname{proj}_{\mathbb{W}} \vec{v} + \operatorname{perp}_{\mathbb{W}} \vec{v}$ with $\operatorname{proj}_{\mathbb{W}} \vec{v} \in \mathbb{W}$ and $\operatorname{perp}_{\mathbb{W}} \vec{v} \in \mathbb{W}^\perp$.

Suppose that $\dim \mathbb{V} = n$, $\{\vec{v}_1, \ldots, \vec{v}_k\}$ is an orthogonal basis for $\mathbb{W}$, and $\{\vec{v}_{k+1}, \ldots, \vec{v}_n\}$ is an orthogonal basis for $\mathbb{W}^\perp$. By Theorem 2 we have that $\{\vec{v}_1, \ldots, \vec{v}_k, \vec{v}_{k+1}, \ldots, \vec{v}_n\}$ is an orthogonal basis for $\mathbb{V}$. So, for any $\vec{v} \in \mathbb{V}$ we find the coordinates with respect to this orthogonal basis are:

$$\vec{v} = \frac{\langle \vec{v}, \vec{v}_1 \rangle}{\|\vec{v}_1\|^2}\vec{v}_1 + \cdots + \frac{\langle \vec{v}, \vec{v}_k \rangle}{\|\vec{v}_k\|^2}\vec{v}_k + \frac{\langle \vec{v}, \vec{v}_{k+1} \rangle}{\|\vec{v}_{k+1}\|^2}\vec{v}_{k+1} + \cdots + \frac{\langle \vec{v}, \vec{v}_n \rangle}{\|\vec{v}_n\|^2}\vec{v}_n$$

This gives us what we want! Taking

$$\operatorname{proj}_{\mathbb{W}} \vec{v} = \frac{\langle \vec{v}, \vec{v}_1 \rangle}{\|\vec{v}_1\|^2}\vec{v}_1 + \cdots + \frac{\langle \vec{v}, \vec{v}_k \rangle}{\|\vec{v}_k\|^2}\vec{v}_k \in \mathbb{W}$$

and

$$\operatorname{perp}_{\mathbb{W}} \vec{v} = \frac{\langle \vec{v}, \vec{v}_{k+1} \rangle}{\|\vec{v}_{k+1}\|^2}\vec{v}_{k+1} + \cdots + \frac{\langle \vec{v}, \vec{v}_n \rangle}{\|\vec{v}_n\|^2}\vec{v}_n \in \mathbb{W}^\perp$$

we get $\vec{v} = \operatorname{proj}_{\mathbb{W}} \vec{v} + \operatorname{perp}_{\mathbb{W}} \vec{v}$ with $\operatorname{proj}_{\mathbb{W}} \vec{v} \in \mathbb{W}$ and $\operatorname{perp}_{\mathbb{W}} \vec{v} \in \mathbb{W}^\perp$.

DEFINITION

Projection

Perpendicular

Suppose $\mathbb{W}$ is a k-dimensional subspace of an inner product space $\mathbb{V}$ and $\{\vec{v}_1, \dots, \vec{v}_k\}$ is an orthogonal basis for $\mathbb{W}$. For any $\vec{v} \in \mathbb{V}$ we define the **projection** of $\vec{v}$ onto $\mathbb{W}$ by

$$\text{proj}_{\mathbb{W}} \vec{v} = \frac{\langle \vec{v}, \vec{v}_1 \rangle}{\|\vec{v}_1\|^2}\vec{v}_1 + \cdots + \frac{\langle \vec{v}, \vec{v}_k \rangle}{\|\vec{v}_k\|^2}\vec{v}_k$$

and the **perpendicular** of the projection by

$$\text{perp}_{\mathbb{W}} \vec{v} = \vec{v} - \text{proj}_{\mathbb{W}} \vec{v}$$

We have defined the perpendicular of the projection in such a way that we do not need an orthogonal basis (or any basis) for $\mathbb{W}^\perp$. To ensure this is valid, we need to verify that $\text{perp}_{\mathbb{W}} \vec{v} \in \mathbb{W}^\perp$.

THEOREM 3

Suppose $\mathbb{W}$ is a k-dimensional subspace of an inner product space $\mathbb{V}$. Then for any $\vec{v} \in \mathbb{V}$, we have

$$\text{perp}_{\mathbb{W}} \vec{v} = \vec{v} - \text{proj}_{\mathbb{W}} \vec{v} \in \mathbb{W}^\perp$$

Proof: Let $\{\vec{v}_1, \dots, \vec{v}_k\}$ be an orthogonal basis for $\mathbb{W}$. Then we can write any $\vec{w} \in \mathbb{W}$ as $\vec{w} = c_1\vec{v}_1 + \cdots + c_k\vec{v}_k$. This gives

$$\langle \text{perp}_{\mathbb{W}} \vec{v}, \vec{w} \rangle = \langle \vec{v} - \text{proj}_{\mathbb{W}} \vec{v}, \vec{w} \rangle = \langle \vec{v}, \vec{w} \rangle - \langle \text{proj}_{\mathbb{W}} \vec{v}, \vec{w} \rangle$$

Observe that

$$\begin{aligned}\langle \vec{v}, \vec{w} \rangle &= \langle \vec{v}, c_1\vec{v}_1 + \cdots + c_k\vec{v}_k \rangle \\ &= c_1 \langle \vec{v}, \vec{v}_1 \rangle + \cdots + c_k \langle \vec{v}, \vec{v}_k \rangle\end{aligned}$$

and, using the fact that $\{\vec{v}_1, \dots, \vec{v}_k\}$ is an orthogonal set, we get

$$\begin{aligned}\langle \text{proj}_{\mathbb{W}} \vec{v}, \vec{w} \rangle &= \left\langle \frac{\langle \vec{v}, \vec{v}_1 \rangle}{\|\vec{v}_1\|^2}\vec{v}_1 + \cdots + \frac{\langle \vec{v}, \vec{v}_k \rangle}{\|\vec{v}_k\|^2}\vec{v}_k, c_1\vec{v}_1 + \cdots + c_k\vec{v}_k \right\rangle \\ &= c_1\frac{\langle \vec{v}, \vec{v}_1 \rangle}{\|\vec{v}_1\|^2} \langle \vec{v}_1, \vec{v}_1 \rangle + \cdots + c_k\frac{\langle \vec{v}, \vec{v}_k \rangle}{\|\vec{v}_k\|^2} \langle \vec{v}_k, \vec{v}_k \rangle \\ &= c_1 \langle \vec{v}, \vec{v}_1 \rangle + \cdots + c_k \langle \vec{v}, \vec{v}_k \rangle\end{aligned}$$

Thus,

$$\langle \vec{v}, \vec{w} \rangle - \langle \text{proj}_{\mathbb{W}} \vec{v}, \vec{w} \rangle = 0$$

and hence $\text{perp}_{\mathbb{W}} \vec{v}$ is orthogonal to every $\vec{w} \in \mathbb{W}$ and so $\text{perp}_{\mathbb{W}} \vec{v} \in \mathbb{W}^\perp$. $\square$

Observe that this implies that $\text{proj}_{\mathbb{W}} \vec{v}$ and $\text{perp}_{\mathbb{W}} \vec{v}$ are orthogonal for any $\vec{v} \in \mathbb{V}$ as we saw in Math 136. Additionally, these functions satisfy other properties which we saw for projections in Math 136.

THEOREM 4 Suppose $\mathbb{W}$ is a k-dimensional subspace of an inner product space $\mathbb{V}$. Then, $\text{proj}_{\mathbb{W}}$ is a linear operator on $\mathbb{V}$ with kernel $\mathbb{W}^{\perp}$.

THEOREM 5 Suppose that $\mathbb{W}$ is a subspace of a finite dimensional subspace $\mathbb{V}$. Then, for any $\vec{v} \in \mathbb{V}$ we have

$$\text{proj}_{\mathbb{W}^{\perp}} \vec{v} = \text{perp}_{\mathbb{W}} \vec{v}$$

EXAMPLE 4 Let $\mathbb{W} = \text{Span}\left\{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\right\}$ be a subspace of $M_{2\times 2}(\mathbb{R})$. Find $\text{proj}_{\mathbb{W}} \begin{bmatrix} 1 & -1 \\ 2 & 3 \end{bmatrix}$.

Solution: Denote the vectors in the basis for $\mathbb{W}$ by $\vec{v}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\vec{v}_2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, and observe that $\{\vec{v}_1, \vec{v}_2\}$ is orthogonal. Let $\vec{x} = \begin{bmatrix} 1 & -1 \\ 2 & 3 \end{bmatrix}$, then

$$\text{proj}_{\mathbb{W}} \begin{bmatrix} 1 & -1 \\ 2 & 3 \end{bmatrix} = \frac{\langle \vec{x}, \vec{v}_1 \rangle}{\|\vec{v}_1\|^2}\vec{v}_1 + \frac{\langle \vec{x}, \vec{v}_2 \rangle}{\|\vec{v}_2\|^2}\vec{v}_2 = \frac{4}{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{-2}{2}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$$

EXAMPLE 5 Let $\mathcal{B} = \left\{\begin{bmatrix} 1 \\ 0 \\ -1 \\ 1, \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \\ -1 \end{bmatrix}\right\}$ and let $\vec{x} = \begin{bmatrix} 4 \\ 3 \\ -2 \\ 5 \end{bmatrix}$. Find the projection of $\vec{x}$ onto the subspace $\mathbb{S} = \text{Span}\,\mathcal{B}$ of $\mathbb{R}^4$.

Solution: To find the projection, we must have an orthogonal basis for $\mathbb{S}$. Thus, our first step must be to perform the Gram-Schmidt procedure on $\mathcal{B}$. Denote the given basis by $\vec{z}_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 1, \end{bmatrix}$, $\vec{z}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}$, $\vec{z}_3 = \begin{bmatrix} -1 \\ 1 \\ 0 \\ -1 \end{bmatrix}$. Let $\vec{w}_1 = \vec{z}_1$. Then, we get

$$\vec{w}_2 = \vec{z}_2 - \frac{\langle \vec{z}_2, \vec{w}_1 \rangle}{\|\vec{w}_1\|^2}\vec{w}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} - \frac{2}{3}\begin{bmatrix} 1 \\ 0 \\ -1 \\ 1, \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 1 \\ 3 \\ 2 \\ 1 \end{bmatrix}$$

To simplify calculations we use $\vec{w}_2 = \begin{bmatrix} 1 \\ 3 \\ 2 \\ 1 \end{bmatrix}$ instead. Then, we get

$$\vec{w}_3 = \vec{z}_3 - \frac{\langle \vec{z}_3, \vec{w}_1 \rangle}{\|\vec{w}_1\|^2}\vec{w}_1 - \frac{\langle \vec{z}_3, \vec{w}_2 \rangle}{\|\vec{w}_2\|^2}\vec{w}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \\ -1 \end{bmatrix} + \frac{2}{3}\begin{bmatrix} 1 \\ 0 \\ -1 \\ 1, \end{bmatrix} - \frac{1}{15}\begin{bmatrix} 1 \\ 3 \\ 2 \\ 1 \end{bmatrix} = \frac{2}{5}\begin{bmatrix} -1 \\ 2 \\ -2 \\ -1 \end{bmatrix}$$

We take $\vec{w}_3 = \begin{bmatrix} -1 \\ 2 \\ -2 \\ -1 \end{bmatrix}$. Thus, the set $\{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$ is an orthogonal basis for $\mathbb{S}$. We can now determine the projection.

$$\operatorname{proj}_{\mathbb{S}} \vec{x} = \frac{\langle \vec{x}, \vec{w}_1 \rangle}{\|\vec{w}_1\|^2}\vec{w}_1 + \frac{\langle \vec{x}, \vec{w}_2 \rangle}{\|\vec{w}_2\|^2}\vec{w}_2 + \frac{\langle \vec{x}, \vec{w}_3 \rangle}{\|\vec{w}_3\|^2}\vec{w}_3 = \frac{11}{3}\vec{w}_1 + \frac{14}{15}\vec{w}_2 + \frac{1}{10}\vec{w}_3 = \begin{bmatrix} 9/2 \\ 3 \\ -2 \\ 9/2 \end{bmatrix}$$

REMARK

Observe that the iterative step in the Gram-Schmidt procedure is just calculating a perpendicular of a projection. In particular, the Gram-Schmidt procedure can be restated as follows: If $\{\vec{w}_1, \dots, \vec{w}_k\}$ is a basis for an inner product space $\mathbb{W}$, then let $\vec{v}_1 = \vec{w}_1$, and for $2 \leq i \leq k$ recursively define $\mathbb{W}_{i-1} = \operatorname{Span}\{\vec{v}_1, \dots, \vec{v}_{i-1}\}$ and $\vec{v}_i = \operatorname{perp}_{\mathbb{W}_{i-1}} \vec{w}_i$. Then $\{\vec{v}_1, \dots, \vec{v}_k\}$ is an orthogonal basis for $\mathbb{W}$.

If $\{\vec{v}_1, \dots, \vec{v}_k\}$ is an orthonormal basis for $\mathbb{W}$, then the formula for calculating a projection simplifies to

$$\operatorname{proj}_{\mathbb{W}} \vec{v} = \langle \vec{v}, \vec{v}_1 \rangle \vec{v}_1 + \cdots + \langle \vec{v}, \vec{v}_k \rangle \vec{v}_k$$

EXAMPLE 6 Let $\mathbb{W} = \operatorname{Span}\left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -\frac{4}{5} \\ 0 \\ \frac{3}{5} \end{bmatrix} \right\}$ be a subspace of $\mathbb{R}^3$. Find $\operatorname{proj}_{\mathbb{W}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and $\operatorname{perp}_{\mathbb{W}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

Solution: Observe that $\left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -\frac{4}{5} \\ 0 \\ \frac{3}{5} \end{bmatrix} \right\}$ is an orthonormal basis for $\mathbb{W}$. Hence,

$$\operatorname{proj}_{\mathbb{W}} \vec{x} = \left\langle \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\rangle \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \left\langle \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -4/5 \\ 0 \\ 3/5 \end{bmatrix} \right\rangle \begin{bmatrix} -4/5 \\ 0 \\ 3/5 \end{bmatrix} = \begin{bmatrix} 4/25 \\ 1 \\ -3/25 \end{bmatrix}$$

$$\operatorname{perp}_{\mathbb{W}} \vec{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 4/25 \\ 1 \\ -3/25 \end{bmatrix} = \begin{bmatrix} 21/25 \\ 0 \\ 28/25 \end{bmatrix}$$

Section 9.4 Problems

1. Consider $\mathbb{S} = \operatorname{Span}\left\{\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix}\right\}$ in $M_{2\times 2}(\mathbb{R})$. Find a basis for $\mathbb{S}^\perp$.

2. Find an orthogonal basis for the orthogonal compliment of each subspace of $P_2(\mathbb{R})$ under the inner product $\langle p(x), q(x)\rangle = p(-1)q(-1) + p(0)q(0) + p(1)q(1)$.

   (a) $\mathbb{S} = \operatorname{Span}\left\{x^2 + 1\right\}$ (b) $\mathbb{S} = \operatorname{Span}\{x^2\}$

3. Let $\vec{w}_1 = \begin{bmatrix} 1 \\ 2 \\ 0 \\ 1 \end{bmatrix}$, $\vec{w}_2 = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 2 \end{bmatrix}$, $\vec{w}_3 = \begin{bmatrix} 2 \\ 1 \\ 3 \\ 2 \end{bmatrix}$, and let $\mathbb{S} = \operatorname{Span}\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$.

   (a) Find an orthonormal basis for $\mathbb{S}$.

   (b) Let $\vec{y} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 12 \end{bmatrix}$. Determine $\operatorname{proj}_{\mathbb{S}} \vec{y}$.

4. Consider the subspace $\mathbb{S} = \operatorname{Span}\left\{\begin{bmatrix} 1 & 0 \\ -1 & 1, \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix}\right\}$ of $M_{2\times 2}(\mathbb{R})$.

   (a) Find an orthogonal basis for $\mathbb{S}$.

   (b) Determine the projection of $\vec{x} = \begin{bmatrix} 4 & 3 \\ -2 & 5 \end{bmatrix}$ on to $\mathbb{S}$.

5. On $P_2(\mathbb{R})$ define the inner product $\langle p, q\rangle = p(-1)q(-1) + p(0)q(0) + p(1)q(1)$ and let $\mathbb{S} = \operatorname{Span}\left\{1, x - x^2\right\}$.

   (a) Use the Gram-Schmidt procedure to determine an orthogonal basis for $\mathbb{S}$.

   (b) Determine $\operatorname{proj}_{\mathbb{S}}(1 + x + x^2)$.

6. Define the inner product $\langle \vec{x}, \vec{y}\rangle = 2x_1y_1 + x_2y_2 + 3x_3y_3$ on $\mathbb{R}^3$. Extend the set $\left\{\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}\right\}$ to an orthogonal basis for $\mathbb{R}^3$.

7. Suppose $\mathbb{W}$ is a k-dimensional subspace of an inner product space $\mathbb{V}$. Prove that $\operatorname{proj}_{\mathbb{W}}$ is a linear operator on $\mathbb{V}$ with kernel $\mathbb{W}^\perp$.

8. In $P_3(\mathbb{R})$ using the inner product $\langle p(x), q(x)\rangle = p(-1)q(-1)+p(0)q(0)+p(1)q(1)+p(2)q(2)$ the following four vectors are orthogonal:

$$\begin{aligned} p_1(x) &= 9 - 13x - 15x^2 + 10x^3 \\ p_2(x) &= 1 + x - x^2 \\ p_3(x) &= 1 - 2x \\ p_4(x) &= 1 \end{aligned}$$

   Let $\mathbb{S} = \operatorname{Span}\{p_1(x), p_2(x), p_3(x)\}$. Determine the projection of $-1 + x + x^2 - x^3$ onto $\mathbb{S}$. [HINT: If you do an excessive number of calculations, you have missed the point of the question.]

9.5 The Fundamental Theorem

Let $\mathbb{W}$ be a subspace of a finite dimensional inner product space $\mathbb{V}$. In the last section we saw that any $\vec{v} \in \mathbb{V}$ we can be written as $\vec{v} = \vec{x} + \vec{y}$ where $\vec{x} \in \mathbb{W}$ and $\vec{y} \in \mathbb{W}^\perp$. We now invent some notation for this.

DEFINITION
Direct Sum

Let $\mathbb{V}$ be a vector space and $\mathbb{U}$ and $\mathbb{W}$ be subspaces of a vector space $\mathbb{V}$ such that $\mathbb{U} \cap \mathbb{W} = \{\vec{0}\}$. The **direct sum** of $\mathbb{U}$ and $\mathbb{W}$ is

$$\mathbb{U} \oplus \mathbb{W} = \{\vec{u} + \vec{w} \in \mathbb{V} \mid \vec{u} \in \mathbb{U}, \vec{w} \in \mathbb{W}\}$$

EXAMPLE 1

Let $\mathbb{U} = \operatorname{Span}\left\{\begin{bmatrix}1\\0\\1\end{bmatrix}\right\}$ and $\mathbb{W} = \operatorname{Span}\left\{\begin{bmatrix}1\\1\\1\end{bmatrix}\right\}$. What is $\mathbb{U} \oplus \mathbb{W}$?

Solution: Since $\mathbb{U} \cap \mathbb{W} = \{\vec{0}\}$ we have

$$\begin{aligned}\mathbb{U} \oplus \mathbb{W} = \{\vec{u} + \vec{w} \in \mathbb{V} \mid \vec{u} \in \mathbb{U}, \vec{w} \in \mathbb{W}\} &= \left\{ s\begin{bmatrix}1\\0\\1\end{bmatrix} + t\begin{bmatrix}1\\1\\1\end{bmatrix} \mid s, t \in \mathbb{R}\right\} \\ &= \operatorname{Span}\left\{\begin{bmatrix}1\\0\\1\end{bmatrix}, \begin{bmatrix}1\\1\\1\end{bmatrix}\right\}\end{aligned}$$

THEOREM 1

If $\mathbb{U}$ and $\mathbb{W}$ are subspaces of a vector space $\mathbb{V}$ such that $\mathbb{U} \cap \mathbb{W} = \{\vec{0}\}$, then $\mathbb{U} \oplus \mathbb{W}$ is a subspace of $\mathbb{V}$. Moreover, if $\{\vec{v}_1, \dots, \vec{v}_k\}$ is a basis for $\mathbb{U}$ and $\{\vec{w}_1, \dots, \vec{w}_\ell\}$ is a basis for $\mathbb{W}$, then $\{\vec{v}_1, \dots, \vec{v}_k, \vec{w}_1, \dots, \vec{w}_\ell\}$ is a basis for $\mathbb{U} \oplus \mathbb{W}$.

EXAMPLE 2

Let $\mathbb{U} = \operatorname{Span}\{x^2 + 1, x^4\}$ and $\mathbb{W} = \operatorname{Span}\{x^3 - x, x^3 + x^2 + 1\}$. Find a basis for $\mathbb{U} \oplus \mathbb{W}$.

Solution: Since $\mathbb{U} \cap \mathbb{W} = \{\vec{0}\}$ a basis for $\mathbb{U} \oplus \mathbb{W}$ is $\{x^2 + 1, x^4, x^3 - x, x^3 + x^2 + 1\}$.

Theorem 9.4.2 tells us that for any subspace $\mathbb{W}$ of an n-dimensional inner product space $\mathbb{V}$ that $\mathbb{W} \cap \mathbb{W}^\perp = \{\vec{0}\}$ and $\dim \mathbb{W}^\perp = n - \dim \mathbb{W}$. Combining this with Theorem 1 we get the following result.

THEOREM 2

Let $\mathbb{V}$ be a finite dimensional inner product space and let $\mathbb{W}$ be a subspace of $\mathbb{V}$. Then

$$\mathbb{W} \oplus \mathbb{W}^\perp = \mathbb{V}$$

We now look at the related result for matrices. We start by looking at an example.

EXAMPLE 3 Find a basis for each of the four fundamental subspaces of $A = \begin{bmatrix} 1 & 2 & 0 & -1 \\ 3 & 6 & 1 & -1 \\ -2 & -4 & 2 & 6 \end{bmatrix}$. Then determine the orthogonal complement of the rowspace of A and the orthogonal complement of the columnspace of A.

Solution: To find a basis for each of the four fundamental subspaces, we use the method derived in Section 7.1. Row reducing A to reduced row echelon form gives $R = \begin{bmatrix} 1 & 2 & 0 & -1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

Hence, a basis for Row(A) is $\left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 2 \end{bmatrix} \right\}$, a basis for Null($A$) is $\left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \end{bmatrix} \right\}$,

and a basis for Col(A) is $\left\{ \begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \right\}$.

Row reducing A^T gives $\begin{bmatrix} 1 & 0 & -8 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. Thus, a basis for the left nullspace is $\left\{ \begin{bmatrix} 8 \\ -2 \\ 1 \end{bmatrix} \right\}$.

Observe that the basis vectors for Null(A) are orthogonal to the basis vectors for Row(A). Additionally, we know that $\dim(\text{Row}\,A)^\perp = 4 - 2 = 2$ by Theorem 9.4.1 (2). Therefore, since $\dim(\text{Null}(A)) = 2$ we have that $(\text{Row}\,A)^\perp = \text{Null}(A)$.

Similarly, we see that $(\text{Col}\,A)^\perp = \text{Null}(A^T)$.

The result of this example seems pretty amazing. Moreover, applying parts (4) and (5) of Theorem 9.4.1 and our notation for direct sums, we get that

$$\mathbb{R}^n = \text{Row}(A) \oplus \text{Null}(A) \quad \text{and} \quad \mathbb{R}^m = \text{Col}(A) \oplus \text{Null}(A^T)$$

This tells us that every vector in $\mathbb{R}^4$ is the sum of a vector in the rowspace of A and a vector in the nullspace of A.

The generalization of this result to any $m \times n$ matrix is extremely important.

THEOREM 3 **(The Fundamental Theorem of Linear Algebra)**

Let A be an $m \times n$ matrix, then $\operatorname{Col}(A)^{\perp} = \operatorname{Null}(A^T)$ and $\operatorname{Row}(A)^{\perp} = \operatorname{Null}(A)$. In particular,

$$\mathbb{R}^n = \operatorname{Row}(A) \oplus \operatorname{Null}(A) \quad \text{and} \quad \mathbb{R}^m = \operatorname{Col}(A) \oplus \operatorname{Null}(A^T)$$

Proof: Let $A = \begin{bmatrix} \vec{v}_1^T \\ \vdots \\ \vec{v}_n^T \end{bmatrix}$.

If $\vec{x} \in \operatorname{Row}(A)^{\perp}$, then $\vec{x}$ is orthogonal to each column of A^T. Hence, $\vec{v}_i \cdot \vec{x} = 0$ for $1 \leq i \leq n$. Thus we get

$$A\vec{x} = \begin{bmatrix} \vec{v}_1^T \vec{x} \\ \vdots \\ \vec{v}_n^T \vec{x} \end{bmatrix} = \begin{bmatrix} \vec{v}_1 \cdot \vec{x} \\ \vdots \\ \vec{v}_n \cdot \vec{x} \end{bmatrix} = \vec{0}$$

hence $\vec{x} \in \operatorname{Null}(A)$.

On the other hand, let $\vec{x} \in \operatorname{Null}(A)$, then $A\vec{x} = \vec{0}$ so $\vec{v}_i \cdot \vec{x} = 0$ for all i. Now, pick any $\vec{b} \in \operatorname{Row}(A)$. Then $\vec{b} = c_1\vec{v}_1 + \cdots + c_n\vec{v}_n$. Therefore,

$$\vec{b} \cdot \vec{x} = (c_1\vec{v}_1 + \cdots + c_n\vec{v}_n) \cdot \vec{x} = c_1(\vec{v}_1 \cdot \vec{x}) + \cdots + c_n(\vec{v}_n \cdot \vec{x}) = 0$$

Hence $\vec{x} \in \operatorname{Row}(A)^{\perp}$.

Applying what we proved above to A^T we get

$$(\operatorname{Col}(A))^{\perp} = (\operatorname{Row}(A^T))^{\perp} = \operatorname{Null}(A^T)$$

$\square$

Observe that the Fundamental Theorem of Linear Algebra tells us more than the relationship between the four fundamental subspaces. For example, if A is an $m \times n$ matrix, then we know that the rank and nullity of the linear mapping $L(\vec{x}) = A\vec{x}$ is $\operatorname{rank} L = \operatorname{rank} A = \dim \operatorname{Row} A$ and $\operatorname{nullity} L = \dim(\operatorname{Null} A)$. Since $\mathbb{R}^n = \operatorname{Row}(A) \oplus \operatorname{Null}(A)$ we get by Theorem 9.4.1 (2) that

$$n = \operatorname{rank} L + \operatorname{nullity}(L)$$

That is, the Fundamental Theorem of Linear Algebra implies the Rank-Nullity Theorem.

It can also be shown that the Fundamental Theorem of Linear Algebra implies a lot of our results about solving systems of linear equations. We will also find it useful in a couple of proofs in future sections.

Section 9.5 Problems

1. Let $A = \begin{bmatrix} 1 & 1 & 3 & 1 \\ 2 & 2 & 6 & 2 \\ -1 & 0 & -2 & -3 \\ 3 & 1 & 7 & 7 \end{bmatrix}$. Find a basis for $\text{Row}(A)^{\perp}$.

2. Prove that if $\mathbb{U}$ and $\mathbb{W}$ are subspaces of a vector space $\mathbb{V}$ such that $\mathbb{U} \cap \mathbb{W} = \{\vec{0}\}$, then $\mathbb{U} \oplus \mathbb{W}$ is a subspace of $\mathbb{V}$. Moreover, if $\{\vec{v}_1, \dots, \vec{v}_k\}$ is a basis for $\mathbb{U}$ and $\{\vec{w}_1, \cdots, \vec{w}_\ell\}$ is a basis for $\mathbb{W}$, then $\{\vec{v}_1, \dots, \vec{v}_k, \vec{w}_1, \dots, \vec{w}_\ell\}$ is a basis for $\mathbb{U} \oplus \mathbb{W}$.

9.6 The Method of Least Squares

In the sciences one often tries to find a correlation between two quantities by collecting data from repeated experimentation. Say, for example, a scientist is comparing quantities x and y which is known to satisfy a quadratic relation $y = a_0 + a_1 x + a_2 x^2$. The scientist would like to find the values of a_0, a_1, and a_2 which best fits their experimental data.

Observe that the data collected from each experiment will correspond to an equation which can be used to determine values of the three unknowns a_0, a_1, and a_2. To get as accurate of a result as possible, the scientists will perform the experiment many times. We thus end up with a system of linear equations which has more equations than unknowns. Such a system of equations is called an **overdetermined system**. Also notice that due to experimentation error, the system is very likely to be inconsistent. So, the scientist needs to find the values of a_0, a_1, and a_2 which best approximates the data they collected.

To solve this problem, we rephrase it in terms of linear algebra. Let A be an $m \times n$ matrix with $m > n$ and let the system $A\vec{x} = \vec{b}$ be inconsistent. We want to find a vector $\vec{x}$ that minimizes the distance between $A\vec{x}$ and $\vec{b}$. Hence, we need to minimize $||A\vec{x} - \vec{b}||$. The following theorem tells us how to do this.

THEOREM 1 **(Approximation Theorem)**

Let $\mathbb{W}$ be a finite dimensional subspace of an inner product space $\mathbb{V}$. If $\vec{v} \in \mathbb{V}$, then the vector closest to $\vec{v}$ in $\mathbb{W}$ is $\text{proj}_{\mathbb{W}} \vec{v}$. That is,

$$||\vec{v} - \text{proj}_{\mathbb{W}} \vec{v}|| < ||\vec{v} - \vec{w}||$$

for all $\vec{w} \in \mathbb{W}$, $\vec{w} \neq \text{proj}_{\mathbb{W}} \vec{v}$.

Proof: Consider $\vec{v} - \vec{w} = (\vec{v} - \text{proj}_W \vec{v}) + (\text{proj}_W \vec{v} - \vec{w})$. Then observe that

$$< \vec{v} - \text{proj}_W \vec{v}, \text{proj}_W \vec{v} - \vec{w} >=< \text{perp}_W \vec{v}, \text{proj}_W \vec{v} > - < \text{perp}_W \vec{v}, \vec{w} >= 0 - 0 = 0$$

Hence, $\{\vec{v}-\operatorname{proj}_W \vec{v}, \operatorname{proj}_W \vec{v}-\vec{w}\}$ is an orthogonal set. Therefore, since the Pythagorean Theorem holds in an orthogonal set, we get

$$\begin{aligned} \|\vec{v}-\vec{w}\|^2 &= \|(\vec{v}-\operatorname{proj}_W \vec{v}) + (\operatorname{proj}_W \vec{v}-\vec{w})\|^2 \\ &= \|\vec{v}-\operatorname{proj}_W \vec{v}\|^2 + \|\operatorname{proj}_W \vec{v}-\vec{w}\|^2 \\ &> \|\vec{v}-\operatorname{proj}_W \vec{v}\|^2 \end{aligned}$$

since $\|\operatorname{proj}_W \vec{v}-\vec{w}\|^2 > 0$ if $\vec{w} \neq \operatorname{proj}_{\mathbb{W}} \vec{v}$. The result now follows. □

Notice that $A\vec{x}$ is in the columnspace of A. Thus, the Approximation Theorem tells us that we can minimize $\|A\vec{x}-\vec{b}\|$ by finding the projection of $\vec{b}$ onto the columnspace of A. Therefore, if we solve the consistent system

$$A\vec{x} = \operatorname{proj}_{\operatorname{Col} A} \vec{b}$$

we will find the desired vector $\vec{x}$. This might seem quite simple, but we can make it even easier. Subtracting both sides of this equation from $\vec{b}$ we get

$$\vec{b} - A\vec{x} = \vec{b} - \operatorname{proj}_{\operatorname{Col} A} \vec{b} = \operatorname{perp}_{\operatorname{Col} A} \vec{b}$$

Thus, $\vec{b} - A\vec{x}$ is in the orthogonal complement of the columnspace of A which, by the Fundamental Theorem of Linear Algebra, means $\vec{b} - A\vec{x}$ is in the nullspace of A^T. Hence

$$A^T(\vec{b} - A\vec{x}) = \vec{0}$$

or equivalently

$$A^TA\vec{x} = A^T\vec{b}$$

This is called the **normal system** and the individual equations are called the **normal equations**. This system will be consistent by construction. However, it need not have a unique solution. If it does have infinitely many solutions, then each of the solutions will minimize $\|A\vec{x}-\vec{b}\|$.

EXAMPLE 1 Determine the vector $\vec{x}$ that minimizes $\|A\vec{x}-\vec{b}\|$ for the system

$$\begin{aligned} 3x_1 - x_2 &= 4 \\ x_1 + 2x_2 &= 0 \\ 2x_1 + x_2 &= 1 \end{aligned}$$

Solution: We have $A = \begin{bmatrix} 3 & -1 \\ 1 & 2 \\ 2 & 1 \end{bmatrix}$ so $A^TA = \begin{bmatrix} 14 & 1 \\ 1 & 6 \end{bmatrix}$, $A^T\vec{b} = \begin{bmatrix} 14 \\ -3 \end{bmatrix}$. Since A^TA is invertible, we find that

$$\vec{x} = (A^TA)^{-1}A^T\vec{b} = \frac{1}{83}\begin{bmatrix} 6 & -1 \\ -1 & 14 \end{bmatrix}\begin{bmatrix} 14 \\ -3 \end{bmatrix} = \frac{1}{83}\begin{bmatrix} 87 \\ -56 \end{bmatrix}$$

Thus, $x_1 = \frac{87}{83}$ and $x_2 = \frac{-56}{83}$.

Note that these give $3x_1 - x_2 = 3.82$, $x_1 + 2x_2 = -0.3$, $2x_1 + 3x_2 = 1.42$, so we have found a fairly good approximation.

EXAMPLE 2 Determine the vector $\vec{x}$ that minimizes $\|A\vec{x}-\vec{b}\|$ for the system

$$\begin{aligned} 2x_1 + x_2 &= -5 \\ -2x_1 + x_2 &= 8 \\ 2x_1 + 3x_2 &= 1 \end{aligned}$$

Solution: We have $A = \begin{bmatrix} 2 & 1 \\ -2 & 1 \\ 2 & 3 \end{bmatrix}$ so

$$A^TA = \begin{bmatrix} 12 & 6 \\ 6 & 11 \end{bmatrix}, \qquad A^T\vec{b} = \begin{bmatrix} -24 \\ 6 \end{bmatrix}$$

Since A^TA is invertible, we find that

$$\vec{x} = (A^TA)^{-1}A^T\vec{b} = \frac{1}{96}\begin{bmatrix} 11 & -6 \\ -6 & 12 \end{bmatrix}\begin{bmatrix} -24 \\ 6 \end{bmatrix} = \begin{bmatrix} -300 \\ 216 \end{bmatrix}$$

Thus, $x_1 = \frac{87}{83}$ and $x_2 = \frac{-56}{83}$.

This method of finding an approximate solution is called the **method of least squares**. It is called this because we are minimizing

$$\|A\vec{x}-\vec{b}\| = \left\| \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix} \right\| = \sqrt{v_1^2 + \cdots + v_m^2}$$

which is equivalent to minimizing $v_1^2 + \cdots + v_m^2$.

We now return to our problem of finding the curve of best fit for a set of data points.

Say in an experiment we get a set of data points $(x_1, y_1), \ldots, (x_m, y_m)$ and we want to find the values of $a_0, \ldots, a_n$ such that $p(x) = a_0 + a_1x + \cdots + a_nx^n$ is the polynomial of best fit. That is, we want the values of $a_0, \ldots, a_n$ such that the values of y_i are approximated as well as possible by $p(x_i)$.

Let $\vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ and define $p(\vec{x}) = \begin{bmatrix} p(x_1) \\ \vdots \\ p(x_n) \end{bmatrix}$. To make this look like the method of least squares we write

$$p(\vec{x}) = \begin{bmatrix} p(x_1) \\ \vdots \\ p(x_n) \end{bmatrix} = \begin{bmatrix} a_0 + a_1(x_1) + \cdots + a_n(x_1)^n \\ \vdots \\ a_0 + a_1(x_m) + \cdots + a_n(x_m)^n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & \cdots & x_1^n \\ \vdots & \vdots & & \vdots \\ 1 & x_m & \cdots & x_m^n \end{bmatrix}\begin{bmatrix} a_0 \\ \vdots \\ a_n \end{bmatrix} = X\vec{a}$$

Thus, we are trying to minimize $\|X\vec{a}-\vec{y}\|$ and we can use the method of least squares above.

This can be summarized in the following theorem.

THEOREM 2 Let n data points $(x_1, y_1), \dots, (x_m, y_n)$ be given and write

$$\vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}, \qquad X = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & \cdots & x_2^n \\ \vdots & & & & \vdots \\ 1 & x_m & x_m^2 & \cdots & x_m^n \end{bmatrix}$$

Then, if $\vec{a} = \begin{bmatrix} a_0 \\ \vdots \\ a_n \end{bmatrix}$ is any solution to the normal system

$$X^T X \vec{a} = X^T \vec{y}$$

then the polynomial

$$p(x) = a_0 + a_1 x + \cdots + a_n x^n$$

is the best fitting polynomial of degree n for the given data. Moreover, if at least $n+1$ of the numbers $x_1, \dots, x_m$ are distinct, then the matrix $X^T X$ is invertible and thus $\vec{a}$ is unique with

$$\vec{a} = (X^T X)^{-1} X^T \vec{y}$$

Proof: We prove the first part above, so we just prove the second part.

Suppose that $n + 1$ of the x_i are distinct and consider a linear combination of the columns of X.

$$c_0 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + c_1 \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} + \cdots + c_n \begin{bmatrix} x_1^n \\ x_2^n \\ \vdots \\ x_m^n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Now let $q(x) = c_0 + c_1 x + \cdots + c_n x^n$. Equating coefficients we see that $x_1, x_2, \dots, x_m$ are all roots of $q(x)$. Then $q(x)$ is a degree n polynomial with at least $n + 1$ distinct roots which contradicts the Fundamental Theorem of Algebra. Therefore, $q(x)$ must be the zero polynomial, and so $c_i = 0$ for $0 \leq i \leq n$. Thus, the columns of X are linearly independent.

To show this implies that $X^T X$ is invertible we consider $X^T X \vec{v} = \vec{0}$. We have

$$\|X\vec{v}\|^2 = (X\vec{v})^T X\vec{v} = \vec{v}^T X^T X \vec{v} = \vec{v}^T \vec{0} = 0$$

Hence, $X\vec{v} = \vec{0}$, so $\vec{v} = \vec{0}$ since the columns of X are linearly independent. Thus $X^T X$ is invertible as required. □

EXAMPLE 3 Find a_0 and a_1 to obtain the best fitting equation of the form $y = a_0 + a_1 x$ for the given data.

x	1	3	4	6	7
y	1	2	3	4	5

Solution: We let $X = \begin{bmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 4 \\ 1 & 6 \\ 1 & 7 \end{bmatrix}$ and $\vec{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{bmatrix}$. We then get

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 3 & 4 & 6 & 7 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 4 \\ 1 & 6 \\ 1 & 7 \end{bmatrix} = \begin{bmatrix} 5 & 21 \\ 21 & 111 \end{bmatrix}$$

$$X^T \vec{y} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 3 & 4 & 6 & 7 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{bmatrix} = \begin{bmatrix} 15 \\ 78 \end{bmatrix}$$

By Theorem 2 we get that $\vec{a} = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}$ is unique and satisfies

$$\begin{aligned} \vec{a} &= (X^T X)^{-1} X^T \vec{y} \\ &= \begin{bmatrix} 5 & 21 \\ 21 & 111 \end{bmatrix}^{-1} \begin{bmatrix} 15 \\ 78 \end{bmatrix} \\ &= \frac{1}{114} \begin{bmatrix} 111 & -21 \\ -21 & 5 \end{bmatrix} \begin{bmatrix} 15 \\ 78 \end{bmatrix} \\ &= \frac{1}{38} \begin{bmatrix} 9 \\ 25 \end{bmatrix} \end{aligned}$$

So, the line of best fit is $y = \frac{9}{38} + \frac{25}{38} x$.

EXAMPLE 4 Find a_0, a_1, a_2 to obtain the best fitting equation of the form $y = a_0 + a_1x + a_2x^2$ for the given data.

x	-3	-1	0	1	3
y	3	1	1	2	4

Solution: We have $\vec{y} = \begin{bmatrix} 3 \\ 1 \\ 1 \\ 2 \\ 4 \end{bmatrix}$ and $X = \begin{bmatrix} 1 & -3 & 9 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 3 & 9 \end{bmatrix}$. Hence,

$$X^TX = \begin{bmatrix} 5 & 0 & 20 \\ 0 & 20 & 0 \\ 20 & 0 & 164 \end{bmatrix}, \qquad X^T\vec{y} = \begin{bmatrix} 11 \\ 4 \\ 66 \end{bmatrix}$$

By Theorem 2 we get that $\vec{a} = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}$ is unique and satisfies

$$\begin{aligned} \vec{a} &= (X^TX)^{-1}X^T\vec{y} \\ &= \begin{bmatrix} 5 & 0 & 20 \\ 0 & 20 & 0 \\ 20 & 0 & 164 \end{bmatrix}^{-1} \begin{bmatrix} 11 \\ 4 \\ 66 \end{bmatrix} \\ &= \frac{1}{210}\begin{bmatrix} 242 \\ 42 \\ 55 \end{bmatrix} \end{aligned}$$

So the best fitting parabola is

$$p(x) = \frac{121}{105} + \frac{1}{5}x + \frac{11}{42}x^2$$

Section 9.6 Problems

1. Find the vector $\vec{x}$ that minimizes $||A\vec{x} - \vec{b}||$ for each of the following systems.

(a)
$$\begin{aligned} x_1 + 2x_2 &= 1 \\ x_1 + x_2 &= 2 \\ x_1 - x_2 &= 5 \end{aligned}$$

(b)
$$\begin{aligned} x_1 + 5x_2 &= 1 \\ -2x_1 - 7x_2 &= 1 \\ x_1 + 2x_2 &= 0 \end{aligned}$$

(c)
$$\begin{aligned} 2x_1 + x_2 &= 4 \\ 2x_1 - x_2 &= -1 \\ 3x_1 + 2x_2 &= 8 \end{aligned}$$

(d)
$$\begin{aligned} x_1 + 2x_2 &= 2 \\ x_1 + 2x_2 &= 3 \\ x_1 + 3x_2 &= 2 \\ x_1 + 3x_2 &= 3 \end{aligned}$$

(e)
$$\begin{aligned} x_1 - x_2 \phantom{{}+x_3} &= 2 \\ 2x_1 - x_2 + x_3 &= -2 \\ 2x_1 + 2x_2 + x_3 &= 3 \\ x_1 - x_2 \phantom{{}+x_3} &= -2 \end{aligned}$$

(f)
$$\begin{aligned} 2x_1 - 2x_2 &= 1 \\ -x_1 + x_2 &= 2 \\ 3x_1 + x_2 &= 1 \\ 2x_1 - x_2 &= 2 \end{aligned}$$

2. Find a_0 and a_1 to obtain the best fitting equation of the form $y = a_0 + a_1x$ for the given data.

(a)

x	-1	0	1
y	-3	2	2

(b)

x	-1	0	1	2
y	-3	-1	0	1

3. Find a_0, a_1, a_2 to obtain the best fitting equation of the form $y = a_0 + a_1x + a_2x^2$ for the given data.

(a)

x	-2	-1	1	2
y	0	-2	0	1

(b)

x	-2	-1	0	1	2
y	0	1	-1	3	1

Chapter 10

Applications of Orthogonal Matrices

10.1 Orthogonal Similarity

In Math 136, we saw that two matrices A and B are similar if there exists an invertible matrix P such that $P^{-1}AP = B$. Moreover, we saw that this corresponds to applying a change of coordinates to the standard matrix of a linear operator. In particular, if $L : \mathbb{R}^n \to \mathbb{R}^n$ is a linear operator and $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_n\}$ is a basis for $\mathbb{R}^n$, then taking $P = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix}$ we get

$$[L]_{\mathcal{B}} = P^{-1}[L]P$$

We then looked for a basis $\mathcal{B}$ of $\mathbb{R}^n$ such that $[L]_{\mathcal{B}}$ was diagonal. We found that such a basis is made up of the eigenvectors of $[L]$.

In the last chapter, we saw that orthonormal bases have some very nice properties. For example, if the columns of P form an orthonormal basis for $\mathbb{R}^n$, then P is orthogonal and it is very easy to find P^{-1}. In particular, we have $P^{-1} = P^T$. Hence, we now turn our attention to trying to find an orthonormal basis $\mathcal{B}$ of eigenvectors such that $[L]_{\mathcal{B}}$ is diagonal. Since the corresponding matrix P will be orthogonal, we will get

$$[L]_{\mathcal{B}} = P^T[L]P$$

DEFINITION
Orthogonally Similar

Two matrices A and B are said to be **orthogonally similar** if there exists an orthogonal matrix P such that

$$P^T AP = B$$

Since $P^T = P^{-1}$ we have that if A and B are orthogonally similar, then they are similar. Therefore, all the properties of similar matrices still hold. In particular, if A and B are orthogonally similar, then $\operatorname{rank} A = \operatorname{rank} B$, $\operatorname{tr} A = \operatorname{tr} B$, $\det A = \det B$, and A and B have the same eigenvalues.

We saw in Math 136 that not every square matrix A has a set of eigenvectors which forms a basis for $\mathbb{R}^n$. Since we are now going to require the additional condition that the basis is orthonormal, we expect that even fewer matrices will have this property. So, we first just look for matrices A such that there exists an orthogonal matrix P for which $P^T AP$ is upper triangular.

THEOREM 1 **(Triangularization Theorem)**

If A is an $n \times n$ matrix with real eigenvalues, then A is orthogonally similar to an upper triangular matrix T.

Proof: We prove the result by induction on n. If $n = 1$, then A is upper triangular. Therefore, we can take P to be the orthogonal matrix $P = [1]$ and the result follows. Now assume the result holds for all $(n-1) \times (n-1)$ matrices and consider an $n \times n$ matrix A with all real eigenvalues.

Let $\vec{v}_1$ be a unit eigenvector of one of A's eigenvalues λ_1. Extend the set $\{\vec{v}_1\}$ to a basis for $\mathbb{R}^n$ and apply the Gram-Schmidt procedure to produce an orthonormal basis $\{\vec{v}_1, \vec{w}_2, \dots, \vec{w}_n\}$ for $\mathbb{R}^n$. Then the matrix $P_1 = \begin{bmatrix} \vec{v}_1 & \vec{w}_2 & \dots & \vec{w}_n \end{bmatrix}$ is orthogonal and

$$P_1^T A P_1 = \begin{bmatrix} \vec{v}_1^T \\ \vec{w}_2^T \\ \vdots \\ \vec{w}_n^T \end{bmatrix} A \begin{bmatrix} \vec{v}_1 & \vec{w}_2 & \dots & \vec{w}_n \end{bmatrix} = \begin{bmatrix} \vec{v}_1 \cdot A\vec{v}_1 & \vec{v}_1 \cdot A\vec{w}_2 & \cdots & \vec{v}_1 A\vec{w}_n \\ \vec{w}_2 \cdot A\vec{v}_1 & \vec{w}_2 \cdot A\vec{w}_2 & \dots & \vec{w}_2 \cdot A\vec{w}_n \\ \vdots & & \ddots & \vdots \\ \vec{w}_n \cdot A\vec{v}_1 & \vec{w}_n \cdot A\vec{w}_2 & \dots & \vec{w}_n \cdot A\vec{w}_n \end{bmatrix}$$

Consider the entries in the first column. Since $A\vec{v}_1 = \lambda_1 \vec{v}_1$ we get

$$\vec{w}_i \cdot A\vec{v}_1 = \vec{w}_i \cdot \lambda_1 \vec{v}_1 = \lambda_1(\vec{w}_i \cdot \vec{v}_1) = 0$$

for $2 \le i \le n$ and $\vec{v}_1 \cdot A\vec{v}_1 = \lambda_1$. Hence

$$P_1^T A P_1 = \begin{bmatrix} \lambda_1 & \vec{b}^T \\ \vec{0} & A_1 \end{bmatrix}$$

where A_1 is an $(n-1) \times (n-1)$ matrix, $\vec{b} \in \mathbb{R}^{n-1}$, and $\vec{0}$ is the zero vector in $\mathbb{R}^{n-1}$. A is similar to $\begin{bmatrix} \lambda_1 & \vec{b}^T \\ \vec{0} & A_1 \end{bmatrix}$ so all the eigenvalues of A_1 are eigenvalues of A and hence all the eigenvalues of A_1 are real. Therefore, by the inductive hypothesis, there exists an $(n-1) \times (n-1)$ orthogonal matrix Q such that $Q^T A_1 Q = T_1$ is upper triangular.

Let $P_2 = \begin{bmatrix} 1 & \vec{0}^T \\ \vec{0} & Q \end{bmatrix}$. Since the columns of Q are orthonormal, the columns of P_2 are also orthonormal and hence P_2 is an orthogonal matrix. Consequently, the matrix $P = P_1 P_2$ is also orthogonal by Theorem 7. Then by block multiplication we get

$$\begin{aligned} P^T A P &= (P_1 P_2)^T A (P_1 P_2) = P_2^T P_1^T A P_1 P_2 \\ &= \begin{bmatrix} 1 & \vec{0}^T \\ \vec{0} & Q^T \end{bmatrix} \begin{bmatrix} \lambda_1 & \vec{b}^T \\ \vec{0} & A_1 \end{bmatrix} \begin{bmatrix} 1 & \vec{0}^T \\ \vec{0} & Q \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 & \vec{b}^T Q \\ \vec{0} & Q^T A_1 Q \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 & \vec{b}^T Q \\ \vec{0} & T_1 \end{bmatrix} \end{aligned}$$

is upper triangular. Thus, the result follows by induction. $\square$

If A is orthogonally similar to an upper triangular matrix T, then A and T must share the same eigenvalues. Thus, since T is upper triangular, the eigenvalues must appear along the main diagonal of T.

Notice that the proof of the theorem gives us a method for finding an orthogonal matrix P so that $P^TAP = T$ is upper triangular. However, since the proof is by induction, this leads to a recursive algorithm. We demonstrate this with an example.

EXAMPLE 1 Let $A = \begin{bmatrix} 2 & 1 & 1 \\ 0 & 3 & -2 \\ 0 & 2 & -1 \end{bmatrix}$. Find an orthogonal matrix P such that P^TAP is upper triangular.

Solution: As in the proof, we first need to find one of the real eigenvalues of A. By inspection, we see that 2 is an eigenvalue of A with unit eigenvector $\vec{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$. Next we extend $\{\vec{v}_1\}$ to an orthonormal basis for $\mathbb{R}^3$. This is very easy in this case as we can take $\vec{w}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ and $\vec{w}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. Thus we get $P_1 = I$ and

$$P_1^TAP_1 = \begin{bmatrix} 2 & 1 & 1 \\ 0 & 3 & -2 \\ 0 & 2 & -1 \end{bmatrix} = \begin{bmatrix} 2 & \vec{b}^T \\ \vec{0} & A_1 \end{bmatrix}$$

where $A_1 = \begin{bmatrix} 3 & -2 \\ 2 & -1 \end{bmatrix}$ and $\vec{b}^T = \begin{bmatrix} 1 & 1 \end{bmatrix}$.

We now need to apply the inductive hypothesis on A_1. Thus, we need to find a 2×2 orthogonal matrix Q such that $Q^TA_1Q = T_1$ is upper triangular. We start by finding a real eigenvalue of A_1. Consider $A_1 - \lambda I = \begin{bmatrix} 3-\lambda & -2 \\ 2 & -1-\lambda \end{bmatrix}$. Then $C(\lambda) = \det(A_1 - \lambda I) = (\lambda - 1)^2$, so we have one eigenvalue $\lambda = 1$. This gives $A_1 - \lambda I = \begin{bmatrix} 2 & -2 \\ 2 & -2 \end{bmatrix}$ so a corresponding unit eigenvector is $\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$.

We extend $\left\{ \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \right\}$ to a orthonormal basis for $\mathbb{R}^2$ by picking $\frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \end{bmatrix}$. Therefore, $Q = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ is an orthogonal matrix and

$$Q^TA_1Q = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 3 & -2 \\ 2 & -1 \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -4 \\ 0 & 1 \end{bmatrix} = T_1$$

Now that we have Q and T_1, the proof of the theorem tells us to choose

$$P_2 = \begin{bmatrix} 1 & \vec{0}^T \\ \vec{0} & Q \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

and $P = P_1P_2 = P_2$. Then

$$P^TAP = \begin{bmatrix} 2 & \vec{b}^T Q \\ \vec{0} & T_1 \end{bmatrix} = \begin{bmatrix} 2 & \sqrt{2} & 0 \\ 0 & 1 & -4 \\ 0 & 0 & 1 \end{bmatrix}$$

as required.

Since the procedure is recursive, it would be very inefficient for larger matrices. There is a much better algorithm for doing this, although we will not cover it in this book. The purpose of the example above is not to teach one how to find such an orthogonal matrix, but to help one understand the proof.

Section 10.1 Problems

1. Prove that if A is orthogonal similar to B and B is orthogonally similar to C, then A is orthogonal similar to C.

2. By following the steps in the proof of the Triangularization Theorem find an orthogonal matrix P and upper triangular matrix T such that $P^TAP = T$ for each of the following matrices.

   (a) $A = \begin{bmatrix} 2 & 4 \\ 1 & 5 \end{bmatrix}$ (b) $A = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 2 & 1 \\ 4 & 0 & 5 \end{bmatrix}$

10.2 Orthogonal Diagonalization

We now know that we can find an orthonormal basis $\mathcal{B}$ such that the $\mathcal{B}$-matrix of any linear operator L on $\mathbb{R}^n$ with real eigenvalues is upper triangular. However, as we saw in Math 136, having a diagonal matrix would be even better. So, we now look at which matrices are orthogonally similar to a diagonal matrix.

DEFINITION

Orthogonally Diagonalizable

An $n \times n$ matrix A is said to be **orthogonally diagonalizable** if there exists an orthogonal matrix P and diagonal matrix D such that

$$P^TAP = D$$

that is, if A is orthogonally similar to a diagonal matrix.

Our goal is to determine all matrices which are orthogonally diagonalizable. At first glance, it may not be clear of how to even start trying to find matrices which are orthogonally diagonalizable, if any exist. One way to approach this problem is to start by looking at what condition a matrix must have if it is orthogonally diagonalizable.

THEOREM 1 If A is orthogonally diagonalizable, then $A^T = A$.

Proof: If A is orthogonally diagonalizable, then there exists an orthogonal matrix P such that $P^TAP = D$ with D diagonal. Since $P^T = P^{-1}$ we can write this as $A = PDP^T$. Taking transposes gives

$$A^T = (PDP^T)^T = (P^T)^T D^T P^T = PDP^T = A$$

since $D = D^T$ as D is diagonal. Thus, $A = A^T$. □

Observe that this theorem does not even guarantee the existence of an orthogonally diagonalizable matrix. It only shows that if one exists, then it must satisfy $A^T = A$. But, on the other hand, it gives us a place to start looking for orthogonally diagonalizable matrices. Lets first consider the set of 2×2 such matrices.

EXAMPLE 1 Prove that every 2×2 matrix $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ is orthogonally diagonalizable.

Solution: We will show that we can find an orthonormal basis of eigenvectors for A so that we can find an orthogonal matrix P which diagonalizes A. We have $C(\lambda) = \det \begin{bmatrix} a-\lambda & b \\ b & c-\lambda \end{bmatrix} = \lambda^2 - (a+c)\lambda + ac - b^2$. Thus, by the quadratic formula, the eigenvalues of A are

$$\begin{aligned}\lambda_+ &= \frac{a+c+\sqrt{(a+c)^2 - 4(ac-b^2)}}{2} = \frac{a+c+\sqrt{(a-c)^2+4b^2}}{2}\\ \lambda_- &= \frac{a+c-\sqrt{(a+c)^2 - 4(ac-b^2)}}{2} = \frac{a+c-\sqrt{(a-c)^2+4b^2}}{2}\end{aligned}$$

If $a = c$ and $b = 0$, then A is already diagonal and thus can be orthogonally diagonalized by I. Otherwise, A has two real eigenvalues with algebraic multiplicity 1. We have

$$A - \lambda_+ I = \begin{bmatrix} \frac{a-c-\sqrt{(a-c)^2+4b^2}}{2} & b \\ b & \frac{c-a-\sqrt{(a-c)^2+4b^2}}{2} \end{bmatrix} \sim \begin{bmatrix} 1 & \frac{2b}{a-c-\sqrt{(a-c)^2+4b^2}} \\ 0 & 0 \end{bmatrix}$$

Hence, an eigenvector for λ_+ is $\vec{v}_1 = \begin{bmatrix} \frac{-2b}{a-c-\sqrt{(a-c)^2+4b^2}} \\ 1 \end{bmatrix}$. Also,

$$A - \lambda_- I = \begin{bmatrix} \frac{a-c+\sqrt{(a-c)^2+4b^2}}{2} & b \\ b & \frac{c-a+\sqrt{(a-c)^2+4b^2}}{2} \end{bmatrix} \sim \begin{bmatrix} 1 & \frac{2b}{a-c+\sqrt{(a-c)^2+4b^2}} \\ 0 & 0 \end{bmatrix}$$

so, an eigenvector for λ_- is $\vec{v}_2 = \begin{bmatrix} \frac{-2b}{a-c+\sqrt{(a-c)^2+4b^2}} \\ 1 \end{bmatrix}$.

Observe that

$$\begin{aligned}\vec{v}_1 \cdot \vec{v}_2 &= \left(\frac{-2b}{a-c-\sqrt{(a-c)^2+4b^2}}\right)\left(\frac{-2b}{a-c+\sqrt{(a-c)^2+4b^2}}\right) + 1(1) \\ &= \frac{4b^2}{(a-c)^2-(a-c)^2-4b^2} + 1 = 0\end{aligned}$$

Thus, the eigenvectors are orthogonal and hence we can normalize them to get an orthonormal basis for $\mathbb{R}^2$. So, A can be diagonalized by an orthogonal matrix.

It is clear that the condition $A^T = A$ is going to be very important. Thus, we make the following definition.

DEFINITION

Symmetric Matrix

A matrix A such that $A^T = A$ is said to be **symmetric**.

It is clear from the definition that a matrix must be square to be symmetric. Moreover, observe that a matrix A is symmetric if and only if $a_{ij} = A_{ji}$ for all $1 \leq i, j \leq n$.

EXAMPLE 2

Determine which of the following matrices is symmetric.

(a) $A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 0 & 1 \\ 1 & 1 & -3 \end{bmatrix}$ (b) $B = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$ (c) $C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$

Solution: It is easy to verify that $A^T = A$ and $C^T = C$ so they are both symmetric. However,

$$B^T = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \neq B$$

so B is not symmetric.

Example 1 shows that every 2×2 symmetric matrix is orthogonally diagonalizable. Thanks to our work in the last section proving that every symmetric matrix is orthogonally diagonalizable is not difficult. We start by stating an amazing result about symmetric matrices.

LEMMA 2

If A is a symmetric matrix with real entries, then all of its eigenvalues are real.

This proof requires properties of vectors and matrices with complex entries and so will be delayed until Chapter 11.

THEOREM 3 **(Principal Axis Theorem)**

Every symmetric matrix is orthogonally diagonalizable.

Proof: Let A be a symmetric matrix. By Lemma 2 all eigenvalues of A are real. Therefore, we can apply the Triangularization Theorem to get that there exists an orthogonal matrix P such that $P^T AP = T$ is upper triangular. Since A is symmetric, we have that $A^T = A$ and hence

$$T^T = (P^T AP)^T = P^T A^T (P^T)^T = P^T AP = T$$

Therefore, T is an upper triangular symmetric matrix. But, if T is upper triangular, then T^T is lower triangular, and so T is both upper and lower triangular and hence T is diagonal. □

Although proving this important theorem is quite easy thanks to the Triangularization Theorem, it does not give us a nice way of orthogonally diagonalizing a symmetric matrix. So, we instead refer to our solution of Example 1. In the example we saw that the eigenvectors corresponding to different eigenvalues were naturally orthogonal. To prove that this is always the case, we first prove an important property of symmetric matrices.

THEOREM 4 A matrix A is symmetric if and only if $\vec{x} \cdot (A\vec{y}) = (A\vec{x}) \cdot \vec{y}$ for all $\vec{x}, \vec{y} \in \mathbb{R}^n$.

Proof: Suppose that A is symmetric, then for any $\vec{x}, \vec{y} \in \mathbb{R}^n$ we have

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

Conversely, if $\vec{x} \cdot (A\vec{y}) = (A\vec{x}) \cdot \vec{y}$ for all $\vec{x}, \vec{y} \in \mathbb{R}^n$, then

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

Since this is valid for all $\vec{y} \in \mathbb{R}^n$ we get that $\vec{x}^T A = \vec{x}^T A^T$ by Theorem 3.1.4 from the Math 136 course notes. Taking transposes of both sides gives $A^T \vec{x} = A\vec{x}$ and this is still valid for all $\vec{x} \in \mathbb{R}^n$. Hence, applying Theorem 3.1.4 again gives $A^T = A$ as required. □

THEOREM 5 If $\vec{v}_1, \vec{v}_2$ are eigenvectors of a symmetric matrix A corresponding to distinct eigenvalues λ_1, λ_2, then $\vec{v}_1$ is orthogonal to $\vec{v}_2$.

Proof: We are assuming that $A\vec{v}_1 = \lambda_1\vec{v}_1$ and $A\vec{v}_2 = \lambda_2\vec{v}_2$, $\lambda_1 \neq \lambda_2$. Theorem 4 gives

$$\lambda_1(\vec{v}_1 \cdot \vec{v}_2) = (\lambda_1\vec{v}_1) \cdot \vec{v}_2 = (A\vec{v}_1) \cdot \vec{v}_2 = \vec{v}_1 \cdot (A\vec{v}_2) = \vec{v}_1 \cdot (\lambda_2\vec{v}_2) = \lambda_2(\vec{v}_1 \cdot \vec{v}_2)$$

Hence, $(\lambda_1 - \lambda_2)(\vec{v}_1 \cdot \vec{v}_2) = 0$. But, $\lambda_1 \neq \lambda_2$, so $\vec{v}_1 \cdot \vec{v}_2 = 0$ as required. □

Consequently, if a symmetric matrix A has n distinct eigenvalues, the basis of eigenvectors which diagonalizes A will naturally be orthogonal. Hence, to orthogonally diagonalize such a matrix A, we just need to normalize these eigenvectors to form an orthonormal basis for $\mathbb{R}^n$ of eigenvectors of A.

EXAMPLE 3 Find an orthogonal matrix P such that P^TAP is diagonal, where $A = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ -1 & 2 & 5 \end{bmatrix}$.

Solution: The characteristic polynomial is

$$C(\lambda) = \begin{vmatrix} 1-\lambda & 0 & -1 \\ 0 & 1-\lambda & 2 \\ -1 & 2 & 5-\lambda \end{vmatrix} = (1-\lambda)[(1-\lambda)(5-\lambda)-4]-(1-\lambda) = -(\lambda-1)\lambda(\lambda-6)$$

Thus the eigenvalues of A are $\lambda_1 = 0$, $\lambda_2 = 1$, and $\lambda_3 = 6$. For $\lambda_1 = 0$ we get

$$A - 0I = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ -1 & 2 & 5 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

Hence, a basis for E_{λ_1}, the eigenspace of λ_1, is $\left\{ \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \right\}$. For $\lambda_2 = 1$, we get

$$A - I \sim \begin{bmatrix} 1 & -2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Hence, a basis for E_{λ_2} is $\left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} \right\}$. For $\lambda_3 = 6$, we get

$$A - 6I \sim \begin{bmatrix} 1 & 0 & 1/5 \\ 0 & 1 & -2/5 \\ 0 & 0 & 0 \end{bmatrix}$$

Hence, a basis for E_{λ_3} is $\left\{ \begin{bmatrix} -1 \\ 2 \\ 5 \end{bmatrix} \right\}$.

As predicted by the theorem, we can easily verify that these are orthogonal to each other. After normalizing we find that an orthonormal basis of eigenvectors for A is

$$\left\{ \begin{bmatrix} 1/\sqrt{6} \\ -2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}, \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \\ 0 \end{bmatrix}, \begin{bmatrix} -1/\sqrt{30} \\ 2/\sqrt{30} \\ 5/\sqrt{30} \end{bmatrix} \right\}$$

Thus, we take

$$P = \begin{bmatrix} 1/\sqrt{6} & 2\sqrt{5} & -1/\sqrt{30} \\ -2/\sqrt{6} & 1/\sqrt{5} & 2/\sqrt{30} \\ 1/\sqrt{6} & 0 & 5/\sqrt{30} \end{bmatrix}$$

Since the columns of P form an orthonormal basis for $\mathbb{R}^3$ we get that P is orthogonal. Moreover, since the columns of P form a basis of eigenvectors for A, we get that P diagonalizes A. In particular,

$$P^T AP = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 6 \end{bmatrix}$$

Notice that Theorem 5 does not say anything about different eigenvectors corresponding to the same eigenvalue. In particular, if an eigenvalue of a symmetric matrix A has geometric multiplicity 2, then there is no guarantee that a basis for the eigenspace of the eigenvalue will be orthogonal. Of course, this is easily fixed as we can just apply the Gram-Schmidt procedure to find an orthonormal basis for the eigenspace.

EXAMPLE 4

Orthogonally diagonalize $A = \begin{bmatrix} 8 & -2 & 2 \\ -2 & 5 & 4 \\ 2 & 4 & 5 \end{bmatrix}$.

Solution: The characteristic polynomial of A is

$$C(\lambda) = \begin{vmatrix} 8-\lambda & -2 & 2 \\ -2 & 5-\lambda & 4 \\ 2 & 4 & 5-\lambda \end{vmatrix} = \begin{bmatrix} 8-\lambda & -4 & 2 \\ 0 & 0 & 9-\lambda \\ 2 & \lambda-1 & 5-\lambda \end{bmatrix} = -\lambda(\lambda-9)^2$$

Thus, the eigenvalues are $\lambda_1 = 0$ and $\lambda_2 = 9$ where the algebraic multiplicity of λ_1 is 1 and the algebraic multiplicity of λ_2 is 2. For $\lambda_1 = 0$ we get

$$A - 0I \sim \begin{bmatrix} 1 & 0 & 1/2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Hence, a basis for E_{λ_1} is $\left\{ \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix} \right\} = \{\vec{v}_1\}$. For $\lambda_2 = 9$ we get

$$A - 9I \sim \begin{bmatrix} 1 & 2 & -2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Hence, a basis for E_{λ_2} is $\left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \right\} = \{\vec{v}_2, \vec{v}_3\}$.

Observe that $\vec{v}_2$ and $\vec{v}_3$ are both orthogonal to $\vec{v}_1$, but not to each other. Thus, we apply the Gram-Schmidt procedure on the basis $\{\vec{v}_2, \vec{v}_3\}$ to get an orthonormal basis for E_{λ_2}.

$$\vec{w}_2 = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}$$

$$\vec{w}_3 = \vec{v}_3 - \frac{\langle \vec{v}_3, \vec{w}_2 \rangle}{\|\vec{w}_2\|^2}\vec{w}_2 = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} - \frac{-4}{5}\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2/5 \\ 4/5 \\ 1 \end{bmatrix}$$

Instead, we take $\vec{w}_3 = \begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix}$. Thus, $\{\vec{w}_2, \vec{w}_3\}$ is an orthogonal basis for the eigenspace of λ_2 and hence $\{\vec{v}_1, \vec{w}_2, \vec{w}_3\}$ is an orthogonal basis for $\mathbb{R}^3$ of eigenvectors of A.

We normalize the vectors to get the orthonormal basis

$$\left\{ \begin{bmatrix} 1/3 \\ 2/3 \\ -2/3 \end{bmatrix}, \begin{bmatrix} -2/\sqrt{5} \\ 1/\sqrt{5} \\ 0 \end{bmatrix}, \begin{bmatrix} 2/\sqrt{45} \\ 4/\sqrt{45} \\ 5/\sqrt{45} \end{bmatrix} \right\}$$

Hence, we get

$$P = \begin{bmatrix} 1/3 & -2/\sqrt{5} & 2/\sqrt{45} \\ 2/3 & 1/\sqrt{5} & 4/\sqrt{45} \\ -2/3 & 0 & 5/\sqrt{45} \end{bmatrix} \quad \text{and} \quad D = P^T AP = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 9 \end{bmatrix}$$

EXAMPLE 5 Orthogonally diagonalize $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$.

Solution: The characteristic polynomial of A is

$$C(\lambda) = \begin{vmatrix} 1-\lambda & 1 & 1 \\ 1 & 1-\lambda & 1 \\ 1 & 1 & 1-\lambda \end{vmatrix} = \begin{bmatrix} 1-\lambda & 1 & 2 \\ 1 & 1-\lambda & 2-\lambda \\ 0 & \lambda & 0 \end{bmatrix} = -\lambda^2(\lambda - 3)$$

Thus, the eigenvalues are $\lambda_1 = 0$ with algebraic multiplicity 2 and $\lambda_2 = 3$ with algebraic multiplicity 1. For $\lambda_1 = 0$ we get

$$A - 0I \sim \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Hence, a basis for E_{λ_1} is $\left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}$.

Since this is not an orthogonal basis for E_{λ_1}, we need to apply the Gram-Schmidt procedure. We get

$$\vec{w}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$$

$$\vec{w}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -1/2 \\ -1/2 \\ 1 \end{bmatrix}$$

Instead, we take $\vec{w}_2 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$. Then, $\{\vec{w}_1, \vec{w}_2\}$ is an orthogonal basis for E_{λ_1}. For $\lambda_2 = 3$ we get

$$A - 3I \sim \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

Hence, a basis for E_{λ_2} is $\left\{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right\}$.

We normalize the basis vectors to get the orthonormal basis

$$\left\{\begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}, \begin{bmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ -2/\sqrt{6} \end{bmatrix}, \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}\right\}$$

Hence, we take

$$P = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ 0 & -2/\sqrt{6} & 1/\sqrt{3} \end{bmatrix}$$

and get

$$P^TAP = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

Section 10.2 Problems

1. Orthogonally diagonalize each of the following symmetric matrices.

(a) $\begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$ (b) $\begin{bmatrix} 4 & -2 \\ -2 & 7 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & -2 \\ -2 & -2 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ (e) $\begin{bmatrix} 0 & 2 & -2 \\ 2 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix}$ (f) $\begin{bmatrix} 2 & -2 & 5 \\ -2 & -5 & -2 \\ -5 & -2 & 2 \end{bmatrix}$

(g) $\begin{bmatrix} 2 & -4 & -4 \\ -4 & 2 & -4 \\ -4 & -4 & 2 \end{bmatrix}$ (h) $\begin{bmatrix} 5 & -1 & 1 \\ -1 & 5 & 1 \\ 1 & 1 & 5 \end{bmatrix}$ (i) $\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$

2. Prove that if A is invertible and orthogonally diagonalizable, then A^{-1} is orthogonally diagonalizable.

3. Determine whether each statement is true or false. Justify your answer with a proof or a counter example.

(a) Every orthogonal matrix is orthogonally diagonalizable.

(b) If A and B are orthogonally diagonalizable, then AB is orthogonally diagonalizable.

(c) If A is orthogonally similar to a symmetric matrix B, then A is orthogonally diagonalizable.

10.3 Quadratic Forms

We now use the results of the last section to study a very important class of functions called quadratic forms. Quadratic forms are not only important in linear algebra, but also in number theory, group theory, differential geometry, and many other areas.

Simply put, a quadratic form is a function defined as a linear combination of all possible terms x_ix_j for $1 \leq i \leq j \leq n$. For example, for two variables x_1, x_2 we have $a_1x_1^2 + a_2x_1x_2 + a_3x_2^2$, and for three variables x_1, x_2, x_3 we get $a_1x_1^2 + a_2x_1x_2 + a_3x_1x_3 + a_4x_2^2 + a_5x_2x_3 + a_6x_3^2$.

Notice that a quadratic form is certainly not a linear function. Instead, they are tied to linear algebra by matrix multiplication. For example, if $B = \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix}$ and $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, then we have

$$\begin{aligned} Q(\vec{x}) = \vec{x}^T B\vec{x} &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} x_1 + 2x_2 \\ -x_2 \end{bmatrix} \\ &= x_1(x_1 + 2x_2) + x_2(-x_2) = x_1^2 + 2x_1x_2 - x_2^2 \end{aligned}$$

which is a quadratic form. We now use this to precisely define a quadratic form.

DEFINITION
Quadratic Form

A **quadratic form** on $\mathbb{R}^n$ with corresponding matrix $n \times n$ matrix A is defined by

$$Q(\vec{x}) = \vec{x}^T A\vec{x}, \qquad \text{for any } \vec{x} \in \mathbb{R}^n$$

There is a connection between the dot product and quadratic forms. Observe that

$$Q(\vec{x}) = \vec{x}^T A\vec{x} = \vec{x} \cdot (A\vec{x}) = (A\vec{x}) \cdot \vec{x} = (A\vec{x})^T \vec{x} = \vec{x}^T A^T \vec{x}$$

Hence, A and A^T give the same quadratic form (this does not imply that $A = A^T$). Of course, this is most natural when A is symmetric. In fact, it can be shown that every quadratic form can be written as $Q(\vec{x}) = \vec{x}^T A\vec{x}$ where A is symmetric. Moreover, each symmetric matrix A uniquely determines a quadratic form. Thus, as we will see, we often deal with a quadratic form and its corresponding symmetric matrix in the same way.

EXAMPLE 1

Let $B = \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix}$. Find the symmetric matrix corresponding to $Q(\vec{x}) = \vec{x}^T B\vec{x}$.

Solution: From our work above we have

$$Q(\vec{x}) = \vec{x}^T B\vec{x} = x_1^2 + 2x_1x_2 - x_2^2$$

We want to find a symmetric matrix $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ such that

$$x_1^2 + 2x_1x_2 - x_2^2 = \vec{x}^T A\vec{x}$$

We have that

$$\begin{aligned} x_1^2 + 2x_1x_2 - x_2^2 = \vec{x}^T A\vec{x} &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} ax_1 + bx_2 \\ bx_1 + cx_2 \end{bmatrix} \\ &= x_1(ax_1 + bx_2) + x_2(bx_1 + cx_2) \\ &= a_1x_1^2 + 2a_2x_1x_2 + a_3x_2^2 \end{aligned}$$

Hence, we take $a_1 = 1$, $a_2 = 1$, and $a_3 = -1$. Consequently, the symmetric matrix corresponding to $Q(\vec{x})$ is $A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$.

For a given quadratic form on $\mathbb{R}^n$, an easy way to think of the corresponding symmetric matrix is as a grid with the rows and columns corresponding to the variables $x_1, \ldots, x_n$. So the 11 entry of the matrix is the x_1x_1 grid and so must have the coefficient of x_1^2. The 12 entry of the matrix is the x_1x_2 grid, but the 21 entry is also a x_1x_2 grid, so they have to equally split the coefficient of x_1x_2 between them. Of course, this also works in reverse.

EXAMPLE 2 Let $A = \begin{bmatrix} 1 & 4 \\ 4 & 3 \end{bmatrix}$. Then the quadratic form corresponding to A is

$$Q(x_1, x_2) = x_1^2 + 2(4)x_1x_2 + 3x_2^2 = x_1^2 + 8x_1x_2 + 3x_2^2$$

EXAMPLE 3 Let $A = \begin{bmatrix} 1 & 2 & -3 \\ 2 & -4 & 0 \\ -3 & 0 & -1 \end{bmatrix}$. Then the quadratic form corresponding to A is

$$\begin{aligned} Q(x_1, x_2, x_3) &= x_1^2 + 2(2)x_1x_2 + 2(-3)x_1x_3 - 4x_2^2 + 2(0)x_2x_3 - x_3^2 \\ &= x_1^2 + 4x_1x_2 - 6x_1x_3 - 4x_2^2 - x_3^2 \end{aligned}$$

EXERCISE 1 Let $A = \begin{bmatrix} -3 & 1/2 & 1 \\ 1/2 & 3/2 & 2 \\ 1 & 2 & 0 \end{bmatrix}$. Find the quadratic form corresponding to A.

EXAMPLE 4 Let $Q(x_1, x_2) = 2x_1^2 + 3x_1x_2 - 4x_2^2$. Find the corresponding symmetric matrix A.

Solution: We get $a_{11} = 2$, $2a_{12} = 3$, and $a_{22} = 4$. Hence,

$$A = \begin{bmatrix} 2 & 3/2 \\ 3/2 & -4 \end{bmatrix}$$

EXAMPLE 5 Let $Q(x_1, x_2, x_3) = x_1^2 - 2x_1x_2 + 8x_1x_3 + 3x_2^2 + x_2x_3 - 5x_3^2$, Find the corresponding symmetric matrix A.

Solution: We have $a_{11} = 1$, $2a_{12} = -2$, $2a_{13} = 8$, $a_{22} = 3$, $2a_{23} = 1$, and $a_{33} = -5$. Hence,

$$A = \begin{bmatrix} 1 & -1 & 4 \\ -1 & 3 & 1/2 \\ 4 & 1/2 & -5 \end{bmatrix}$$

EXERCISE 2 Let $Q(x_1, x_2, x_3) = 3x_1^2 + \frac{1}{2}x_1x_2 - x_2^2 + 2x_2x_3 + 2x_3^2$, Find the corresponding symmetric matrix A.

EXAMPLE 6 Let $Q(x_1, x_2, x_3) = x_1^2 - 2x_2^2 - 3x_3^2$, then the corresponding symmetric matrix is

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -3 \end{bmatrix}$$

The terms $x_i x_j$ with $i \neq j$ are called cross-terms. Thus, a quadratic form $Q(\vec{x})$ has no cross terms if and only if its corresponding symmetric matrix is diagonal. We will call such a quadratic form a **diagonal quadratic form**.

Classifying Quadratic Forms

DEFINITION Let $Q(\vec{x})$ be a quadratic form. Then

1. $Q(\vec{x})$ is positive definite if $Q(\vec{x}) > 0$ for all $\vec{x} \neq \vec{0}$.
2. $Q(\vec{x})$ is negative definite if $Q(\vec{x}) < 0$ for all $\vec{x} \neq \vec{0}$.
3. $Q(\vec{x})$ is indefinite if $Q(\vec{x}) > 0$ for some $\vec{x}$ and $Q(\vec{x}) < 0$ for some $\vec{x}$.
4. $Q(\vec{x})$ is positive semidefinite if $Q(\vec{x}) \geq 0$ for all $\vec{x}$ and $Q(\vec{x}) = 0$ for some $\vec{x} \neq \vec{0}$.
5. $Q(\vec{x})$ is negative semidefinite if $Q(\vec{x}) \leq 0$ for all $\vec{x}$ and $Q(\vec{x}) = 0$ for some $\vec{x} \neq \vec{0}$.

EXAMPLE 7 $Q(x_1, x_2) = x_1^2 + x_2^2$ is positive definite, $Q(x_1, x_2) = -x_1^2 - x_2^2$ is negative definite, $Q(x_1, x_2) = x_1^2 - x_2^2$ is indefinite, $Q(x_1, x_2) = x_1^2$ is positive semidefinite, and $Q(x_1, x_2) = -x_1^2$ is negative semidefinite.

We classify symmetric matrices in the same way that we classify their corresponding quadratic forms.

EXAMPLE 8 $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ corresponds to the quadratic form $Q(x_1, x_2) = x_1^2 + x_2^2$, so A is positive definite.

EXAMPLE 9 Classify the symmetric matrix $A = \begin{bmatrix} 4 & -2 \\ -2 & 4 \end{bmatrix}$ and the corresponding quadratic form $Q(\vec{x}) = \vec{x}^T A\vec{x}$.

Solution: Observe that we have

$$Q(\vec{x}) = 4x_1^2 - 4x_1x_2 + 4x_2^2 = (2x_1 - x_2)^2 + 3x_2^2 > 0$$

whenever $\vec{x} \neq \vec{0}$. Thus, $Q(\vec{x})$ and A are both positive definite.

REMARK

The definitions, of course, can be generalized to any function and matrix. For example, the matrix $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ is positive definite since

$$\vec{x}^T A\vec{x} = x_1^2 + x_1x_2 + x_2^2 = (x_1 + \frac{1}{2}x_2)^2 + \frac{3}{4}x_2^2 > 0$$

for all $\vec{x} \neq \vec{0}$. In this book we will just be dealing with quadratic forms and symmetric matrices. However, it is important to remember that "Assume A is positive definite" does not imply that A is symmetric.

Each of the quadratic forms in the example above was simple to classify since it either had no cross-terms or we were easily able to complete a square. To classify more difficult quadratic forms, we use our theory from the last section and the connection between quadratic forms and symmetric matrices.

THEOREM 1 Let A be the symmetric matrix corresponding to a quadratic form $Q(\vec{x}) = \vec{x}^T A\vec{x}$. If P is an orthogonal matrix that diagonalizes A, then $Q(\vec{x})$ can be expressed as

$$\lambda_1 y_1^2 + \cdots \lambda_n y_n^2$$

where $\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \vec{y} = P^T\vec{x}$ and where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of A corresponding to the columns of P.

Proof: Since P is orthogonal, if $\vec{y} = P^T\vec{x}$, then $\vec{x} = P\vec{y}$. Moreover, we have that $P^T AP = D$ is diagonal where the diagonal entries of D are the eigenvalues $\lambda_1, \dots, \lambda_n$ of A. Hence,

$$Q(\vec{x}) = \vec{x}^T A\vec{x} = (P\vec{y})^T A(P\vec{y}) = \vec{y}^T P^T AP\vec{y} = \vec{y}^T D\vec{y} = \lambda_1 y_1^2 + \cdots \lambda_n y_n^2$$

□

This theorem shows that every quadratic form $Q(\vec{x})$ is equivalent to a diagonal quadratic form. Moreover, $Q(\vec{x})$ can be brought into this form by the change of variables $\vec{y} = P^T\vec{x}$ where P is an orthogonal matrix which diagonalizes the corresponding symmetric matrix A.

EXAMPLE 10 Find new variables y_1, y_2, y_3, y_4 such that

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

has diagonal form. Use the diagonal form to classify $Q(\vec{x})$.

Solution: We have corresponding symmetric matrix $A = \begin{bmatrix} 3 & 1 & -5 & 5 \\ 1 & 3 & 5 & -5 \\ -5 & 5 & 3 & 1 \\ 5 & -5 & 1 & 3 \end{bmatrix}$.

We first find an orthogonal matrix P which diagonalizes A. We have $C(\lambda) = (\lambda - 12)(\lambda + 8)(\lambda - 4)^2$ and finding the corresponding normalized eigenvectors gives

$$P = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & 1 & -1 \end{bmatrix}$$

with $P^TAP = \begin{bmatrix} 12 & 0 & 0 & 0 \\ 0 & -8 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$. Thus, if

$$\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = P^T \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} x_1 - x_2 - x_3 + x_4 \\ x_1 - x_2 + x_3 - x_4 \\ x_1 + x_2 + x_3 + x_4 \\ x_1 + x_2 - x_3 + x_4 \end{bmatrix}$$

then we get

$$Q = 12y_1^2 - 8y_2^2 + 4y_3^2 + 4y_4^2$$

Therefore, $Q(\vec{x})$ clearly takes positive and negative values, so $Q(\vec{x})$ is indefinite.

REMARKS

1. The eigenvectors we used to make up P are the principal axes of A. We will see a geometric interpretation of this in the next section.

2. By changing the order of the eigenvectors in P we also change the order of the eigenvalues in D and hence the coefficients of the corresponding y_i. For example, if we took

$$P = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ -1 & 1 & -1 & 1 \end{bmatrix}$$

then we would get $Q = -8y_1^2 + 4y_2^2 + 4y_3^2 + 12y_4^2$. Notice, that since we can pick any vector $\vec{y} \in \mathbb{R}^4$, this does in fact give us exactly the same set of values as our choice above. Alternatively, you can think of just doing another change of variables $z_1 = y_2$, $z_2 = y_3$, $z_3 = y_4$, and $z_4 = y_1$.

Generalizing the method of the example above gives us the following theorem.

THEOREM 2

Let A be a symmetric matrix. Then the quadratic form $Q(\vec{x}) = \vec{x}^T A\vec{x}$ is

(1) positive definite if and only if the eigenvalues of A are all positive.
(2) negative definite if and only if the eigenvalues of A are all negative.
(3) indefinite if and only if the some of the eigenvalues of A are positive and some are negative.
(4) positive semidefinite if and only if it has some zero eigenvalues and the rest positive.
(5) negative semidefinite if and only if has some zero eigenvalues and the rest negative.

Proof: We will prove (1). The proofs of the others are similar. By the Principal Axis Theorem A is orthogonally diagonalizable, so there exists an orthogonal matrix P such that

$$Q(\vec{x}) = \vec{x}^T A\vec{x} = \vec{y}^T D\vec{y} = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2$$

where $\vec{y} = P^T\vec{x}$ and $\lambda_1, \dots, \lambda_n$ are the eigenvalues of A. Consequently, $Q(\vec{x}) > 0$ for all $\vec{x} \neq \vec{0}$ if and only if all λ_i are positive. Thus, $Q(\vec{x})$ is positive definite if and only if the eigenvalues of A are all positive. □

EXAMPLE 11

Classify the following:

(a) $Q(x_1, x_2) = 4x_1^2 - 6x_1x_2 + 2x_2^2$.

Solution: We have $A = \begin{bmatrix} 4 & -3 \\ -3 & 2 \end{bmatrix}$ so

$$\det(A - \lambda I) = \begin{vmatrix} 4-\lambda & -3 \\ -3 & 2-\lambda \end{vmatrix} = \lambda^2 - 6\lambda - 1$$

Applying the quadratic formula we get $\lambda = \frac{6 \pm \sqrt{40}}{2}$. So, we have both positive and negative eigenvalues so Q is indefinite.

(b) $Q(x_1, x_2, x_3) = 2x_1^2 + 2x_1x_2 + 2x_1x_3 + 2x_2^2 + 2x_2x_3 + 2x_3^2$.

Solution: We have $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$. We can find that the eigenvalues are 1, 1 and 4. Hence all the eigenvalues of A are positive so A is positive definite.

(c) $A = \begin{bmatrix} 3 & 2 & 0 \\ 2 & 2 & 2 \\ 0 & 2 & 1 \end{bmatrix}$.

Solution: The eigenvalues of A are 5, 2, and -1. Therefore, A is indefinite.

Section 10.3 Problems

1. For each of the following quadratic forms $Q(\vec{x})$:

   (i) Find the corresponding symmetric matrix.

   (ii) Classify the quadratic form and its corresponding symmetric matrix.

   (iii) Find a corresponding diagonal form of $Q(\vec{x})$ and the change of variables which brings it into this form.

   (a) $Q(\vec{x}) = x_1^2 + 3x_1x_2 + x_2^2$

   (b) $Q(\vec{x}) = 8x_1^2 + 4x_1x_2 + 11x_2^2$

   (c) $Q(\vec{x}) = x_1^2 + 8x_1x_2 - 5x_2^2$

   (d) $Q(\vec{x}) = -x_1^2 + 4x_1x_2 + 2x_2^2$

   (e) $Q(\vec{x}) = 4x_1^2 + 4x_1x_2 + 4x_1x_3 + 4x_2^2 + 4x_2x_3 + 4x_3^2$

   (f) $Q(\vec{x}) = 4x_1x_2 - 4x_1x_3 + x_2^2 - x_3^2$

   (g) $Q(\vec{x}) = 4x_1^2 - 2x_1x_2 + 2x_1x_3 + 4x_2^2 + 2x_2x_3 + 4x_3^2$

   (h) $Q(\vec{x}) = -4x_1^2 + 2x_1x_2 + 2x_1x_3 - 4x_2^2 - 2x_2x_3 - 4x_3^2$

2. Let $Q(\vec{x}) = \vec{x}^T A\vec{x}$ with $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ and $\det A \neq 0$.

   (a) Prove that Q is positive definite if $\det A > 0$ and $a > 0$.

   (b) Prove that Q is negative definite if $\det A > 0$ and $a < 0$.

   (c) Prove that Q is indefinite if $\det A < 0$.

3. Let A be an $n \times n$ matrix and let $\vec{x}, \vec{y} \in \mathbb{R}^n$. Define $< \vec{x}, \vec{y} >= \vec{x}^T A\vec{y}$. Prove that $< , >$ is an inner product on $\mathbb{R}^n$ if and only if A is a positive definite, symmetric matrix.

4. Let A be a positive definite symmetric matrix. Prove that:

   (a) the diagonal entries of A are all positive.

   (b) A is invertible.

5. Let A and B be symmetric $n \times n$ matrices whose eigenvalues are all positive. Show that the eigenvalues of $A + B$ are all positive.

10.4 Graphing Quadratic Forms

In many applications of quadratic forms it is important to be able to sketch the graph of a quadratic form $Q(\vec{x}) = k$ for some k. Observe that in general it is not easy to identify the shape of the graph of a general equation $ax_1^2 + bx_1x_2 + cx_2^2 = k$ by inspection. On the other hand, if you are familiar with conic sections, it is easy to determine the shape of the graph of an equation of the form $\lambda_1 x_1^2 + \lambda_2 x_2^2 = k$. We demonstrate this in a table.

The graph of $\lambda_1 x_1^2 + \lambda_2 x_2^2 = k$ looks like:

	$k > 0$	$k = 0$	$k < 0$
$\lambda_1 > 0, \lambda_2 > 0$	ellipse	point$(0,0)$	dne
$\lambda_1 < 0, \lambda_2 < 0$	dne	point$(0,0)$	ellipse
$\lambda_1\lambda_2 < 0$	hyperbola	asymptotes for hyperbola	hyperbola
$\lambda_1 = 0, \lambda_2 > 0$	parallel lines	line $x_2 = 0$	dne
$\lambda_1 = 0, \lambda_2 < 0$	dne	line $x_2 = 0$	parallel lines

In the last section we saw that we could bring a quadratic form into diagonal form by performing the change variables $\vec{y} = P^T\vec{x}$ where P orthogonally diagonalizes the corresponding symmetric matrix. Therefore, to sketch an equation of the form $ax_1^2 + bx_1x_2 + cx_2^2 = k$ we could first apply the change of variables to bring it into the form $\lambda_1 y_1^2 + \lambda_2 y_2^2 = k$ which will be easy to sketch. However, we must determine how performing the change of variables will affect the graph.

THEOREM 1 Let $Q(\vec{x}) = ax_1^2 + bx_1x_2 + cx_2^2$. Then there exists an orthogonal matrix P which corresponds to a rotation such that the change of variables $\vec{y} = P^T\vec{x}$ brings $Q(\vec{x})$ into diagonal form.

Proof: Let $A = \begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix}$. Since A is symmetric we can apply the Principal Axis Theorem to get that there exists an orthonormal basis $\{\vec{v}_1, \vec{v}_2\}$ of $\mathbb{R}^2$ of eigenvectors of A. Let $\vec{v}_1 = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$ and $\vec{v}_2 = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$. Since $\vec{v}_1$ is a unit vector we must have $a_1^2 + a_2^2 = 1$. Hence, the components lie on the unit circle and so there exists an angle θ such that $a_1 = \cos\theta$ and $a_2 = \sin\theta$. Moreover, since $\vec{v}_2$ is a unit vector orthogonal to $\vec{v}_1$ we can pick $b_1 = -\sin\theta$ and $b_2 = \cos\theta$. Hence we have

$$P = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

which corresponds to a rotation by θ. Finally, from our work above, we know that this change of basis matrix brings Q into diagonal form. $\square$

REMARK

Observe that our choices for $\vec{v}_1$ and $\vec{v}_2$ in the proof above were not unique. For example, we could have picked $\vec{v}_1 = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$ and $\vec{v}_2 = \begin{bmatrix} \sin\theta \\ \cos\theta \end{bmatrix}$. In this case the matrix $P = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 \end{bmatrix}$ would correspond to a rotation and a reflection. This does not contradict the theorem though. The theorem only says that we can choose P to correspond to a rotation, not that a rotation is the only choice. We will demonstrate this in an example below.

EXAMPLE 1

Consider the equation $x_1^2 + x_1x_2 + x_2^2 = 1$. Find a rotation so that the equation has no cross terms.

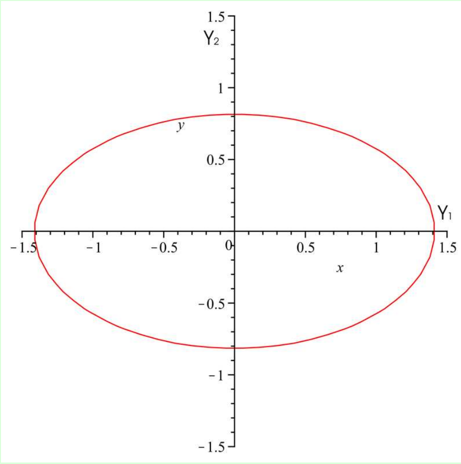
Solution: We have $A = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1 \end{bmatrix}$. The eigenvalues are $\lambda_1 = \frac{1}{2}$ and $\lambda_2 = \frac{3}{2}$. Therefore, the graph of $x_1^2 + x_1x_2 + x_2^2 = 1$ is an ellipse. We find that the corresponding unit eigenvectors are $\vec{v}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and $\vec{v}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ -1 \end{bmatrix}$. So, from our work in the proof of the theorem, we can pick $\cos\theta = \dfrac{-1}{\sqrt{2}}$ and $\sin\theta = \dfrac{1}{\sqrt{2}}$. Hence, $\theta = \dfrac{3\pi}{4}$ and

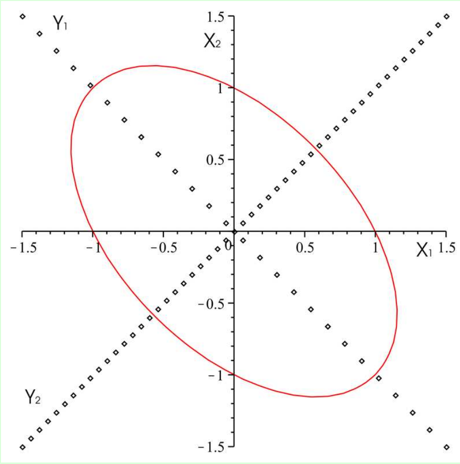
$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} (-y_1 - y_2)/\sqrt{2} \\ (y_1 - y_2)/\sqrt{2} \end{bmatrix}$$

Substituting these into the equation gives

$$\begin{aligned} 1 = x_1^2 + x_1x_2 + x_2^2 &= \left(\frac{-y_1 - y_2}{\sqrt{2}}\right)^2 + \left(\frac{-y_1 - y_2}{\sqrt{2}}\right)\left(\frac{y_1 - y_2}{\sqrt{2}}\right) + \left(\frac{y_1 - y_2}{\sqrt{2}}\right)^2 \\ &= \frac{1}{2}(y_1 + 2y_1y_2 + y_2^2) + \frac{1}{2}(-y_1^2 + y_2^2) + \frac{1}{2}(y_1^2 - 2y_1y_2 + y_2^2) \\ &= \frac{1}{2}y_1^2 + \frac{3}{2}y_2^2 \end{aligned}$$

Thus, the graph of $x_1^2 + x_1x_2 + x_2^2 = 1$ is the graph of $1 = \dfrac{1}{2}y_1^2 + \dfrac{3}{2}y_2^2$ rotated by $\frac{3\pi}{4}$ degrees. This is shown below.





Let us look at this procedure in general. Assume that we want to sketch $Q(x_1, x_2) = k$ where $Q(x_1, x_2)$ is a quadratic form with corresponding symmetric matrix A. We first write $Q(x_1, x_2)$ in diagonal form $\lambda_1 y_1^2 + \lambda_2 y_2^2$ by finding the orthogonal matrix $P = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 \end{bmatrix}$ which diagonalizes A and performing the change of variables $\vec{y} = P^T\vec{x}$. We can then easily sketch $\lambda_1 y_1^2 + \lambda_2 y_2^2 = k$ (finding the equations of the asymptotes if it is a hyperbola) in the y_1y_2-plane. Then using the fact that

$$\vec{x} = P\vec{y} = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = y_1\vec{v}_1 + y_2\vec{v}_2$$

we can find the y_1-axis and the y_2-axis in the x_1x_2-plane. In particular, the y_1-axis in the y_1y_2-plane is spanned by $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$, thus the y_1-axis in the x_1x_2-plane is spanned by

$$\vec{x} = 1\vec{v}_1 + 0\vec{v}_2 = \vec{v}_1$$

Hence, performing the change of variables rotates the y_1-axis to be in the direction of $\vec{v}_1$. Similarly, the y_2 axis is rotated to the $\vec{v}_2$ direction. Therefore, we can sketch the graph of $Q(x_1, x_2) = k$ in the x_1x_2-plane by drawing the y_1 and y_2 axes in the x_1x_2-plane and sketching the graph of $\lambda_1 y_1^2 + \lambda_2 y_2^2 = k$ on these axes as we did in the y_1y_2-plane. For this reason, the orthogonal eigenvectors $\vec{v}_1, \vec{v}_2$ of A are called the **principal axes** for $Q(x_1, x_2)$.

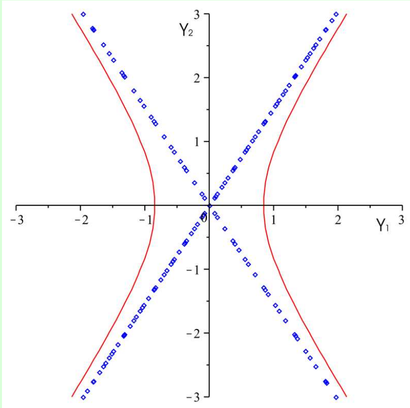
EXAMPLE 2 Sketch $6x_1^2 + 6x_1x_2 - 2x_2^2 = 5$.

Solution: We start by finding an orthogonal matrix which diagonalizes the corresponding symmetric matrix A.

We have $A = \begin{bmatrix} 6 & 3 \\ 3 & -2 \end{bmatrix}$, so $\det(A - \lambda I) = \begin{vmatrix} 6-\lambda & 3 \\ 3 & -2-\lambda \end{vmatrix} = (\lambda - 7)(\lambda + 3)$. We find corresponding eigenvectors of A are $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$ for $\lambda = 7$ and $\begin{bmatrix} -1 \\ 3 \end{bmatrix}$ for $\lambda = -3$. Thus, $P = \begin{bmatrix} 3/\sqrt{10} & -1/\sqrt{10} \\ 1/\sqrt{10} & 3/\sqrt{10} \end{bmatrix}$ orthogonally diagonalizes A.

Then, from our work above, we know that the change of variables $\vec{y} = P^T\vec{x}$ changes $6x_1^2 + 6x_1x_2 - 2x_2^2 = 5$ into $7y_1^2 - 3y_2^2 = 5$. To sketch the hyperbola $5 = 7y_1^2 - 3y_2^2$ more accurately, we first find the equations of its asymptotes. We get

$$\begin{aligned} 0 &= 7y_1^2 - 3y_2^2 \\ y_2 &= \pm\frac{\sqrt{21}}{3}y_1 \\ &\approx \pm 1.528y_1 \end{aligned}$$

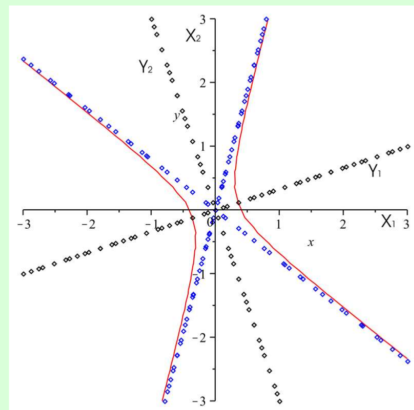


Now to sketch $6x_1^2 + 6x_1x_2 - 2x_2^2 = 5$ in the x_1x_2-plane we first draw the y_1-axis in the direction of $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$ and the y_2-axis in the direction of $\begin{bmatrix} -1 \\ 3 \end{bmatrix}$. Next, to be more precise, we use the change of variables $\vec{y} = P^T\vec{x}$ to convert the equations of the asymptotes of the hyperbola to equations in the x_1x_2-plane. We have

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \frac{1}{\sqrt{10}}\begin{bmatrix} 3 & 1 \\ -1 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{\sqrt{10}}\begin{bmatrix} 3x_1 + x_2 \\ -x_1 + 3x_2 \end{bmatrix}$$

Hence, we get the equations of the asymptotes are

$$\begin{aligned} y_2 &= \pm\frac{\sqrt{21}}{3}y_1 \\ \frac{1}{\sqrt{10}}(-x_1 + 3x_2) &= \pm\frac{\sqrt{21}}{3}\frac{1}{\sqrt{10}}(3x_1 + x_2) \\ -3x_1 + 9x_2 &= \pm 3\sqrt{21}x_1 \pm \sqrt{21}x_2 \\ x_2(9 \mp \sqrt{21}) &= (3 \pm 3\sqrt{21})x_1 \\ x_2 &= \frac{3 \pm 3\sqrt{21}}{9 \mp \sqrt{21}}x_1 \end{aligned}$$



Plotting the asymptotes and copying the graph from above onto the y_1 and y_2 axes gives the picture to the right.

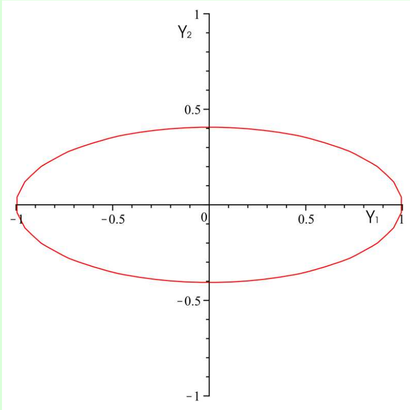
REMARK

The advantage of finding the principal axes over just calculating the rotation is that we actually get precise equations of the axes and the asymptotes which are required in some applications.

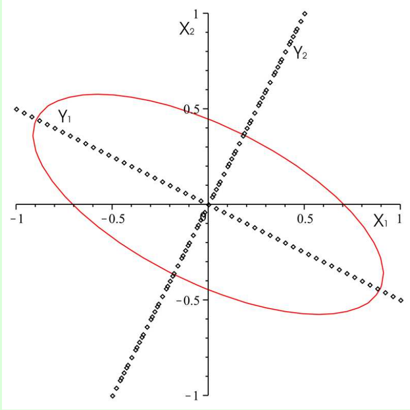
EXAMPLE 3 Sketch $2x_1^2 + 4x_1x_2 + 5x_2^2 = 1$.

Solution: We have $A = \begin{bmatrix} 2 & 2 \\ 2 & 5 \end{bmatrix}$, so $\det(A - \lambda I) = \begin{vmatrix} 2-\lambda & 2 \\ 2 & 5-\lambda \end{vmatrix} = (\lambda - 6)(\lambda - 1)$. We find corresponding eigenvectors of A are $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ for $\lambda = 6$ and $\begin{bmatrix} -2 \\ 1 \end{bmatrix}$ for $\lambda = 1$. To demonstrate a rotation and a reflection, we will pick $P = \begin{bmatrix} -2/\sqrt{5} & 1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}$.

We find corresponding eigenvectors of A are $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ for $\lambda = 6$ and $\begin{bmatrix} -2 \\ 1 \end{bmatrix}$ for $\lambda = 1$. To demonstrate a rotation and a reflection, we will pick $P = \begin{bmatrix} -2/\sqrt{5} & 1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}$. The change of variables $\vec{y} = P^T\vec{x}$ changes $2x_1^2 + 4x_1x_2 + 5x_2^2 = 1$ into the ellipse $y_1^2 + 6y_2^2 = 1$. Graphing gives the picture to the right.



To sketch $2x_1^2+4x_1x_2+5x_2^2 = 1$ in the x_1x_2-plane we draw the y_1-axis in the direction of $\begin{bmatrix} -2 \\ 1 \end{bmatrix}$ and the y_2-axis in the direction of $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and copy the picture above onto the axes appropriately to get the picture to the right.

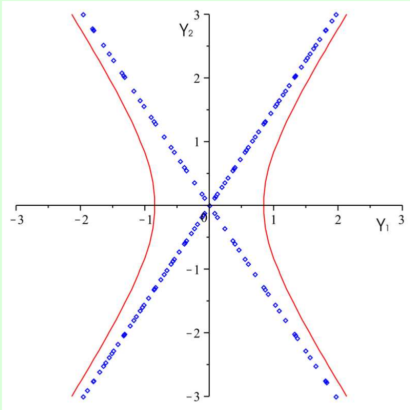


EXAMPLE 4 Sketch $x_1^2 + 4x_1x_2 - 2x_2^2 = 1$.

Solution: We have $A = \begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix}$ so $\det(A-\lambda I) = \begin{vmatrix} 1-\lambda & 2 \\ 2 & -2-\lambda \end{vmatrix} = (\lambda+3)(\lambda-2)$. We find corresponding eigenvectors of A are $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$ for $\lambda = -3$ and $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ for $\lambda = 2$. Thus, $P = \begin{bmatrix} -1/\sqrt{5} & 2/\sqrt{5} \\ 2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix}$ orthogonally diagonalizes A.

Then, the change of variables $\vec{y} = P^T\vec{x}$ changes $x_1^2 + 4x_1x_2 - 2x_2^2 = 1$ into $-3y_1^2 + 2y_2^2 = 5$. This is a hyperbola, so we find the equations of its asymptotes. We get

$$\begin{aligned} 0 &= -3y_1^2 + 2y_2^2 \\ y_2 &= \pm\frac{\sqrt{6}}{2}y_1 \\ &\approx \pm 1.225y_1 \end{aligned}$$

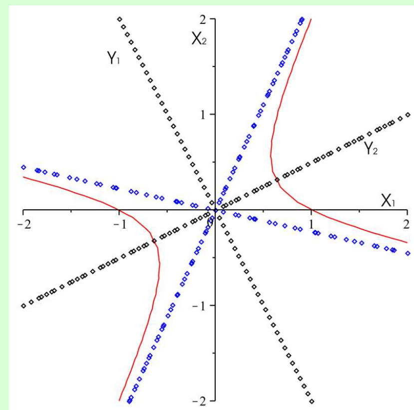


To sketch $x_1^2 + 4x_1x_2 - 2x_2^2 = 1$ in the x_1x_2-plane we first draw the y_1-axis in the direction of $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$ and the y_2-axis in the direction of $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$. Next, we use the change of variables $\vec{y} = P^T\vec{x}$ to convert the equations of the asymptotes of the hyperbola to equations in the x_1x_2-plane. We have

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \frac{1}{\sqrt{5}}\begin{bmatrix} -1 & 2 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{\sqrt{5}}\begin{bmatrix} -x_1 + 2x_2 \\ 2x_1 + x_2 \end{bmatrix}$$

Hence, we get the equations of the asymptotes are

$$\begin{aligned} y_2 &= \pm\frac{\sqrt{6}}{2}x_1 \\ \frac{1}{\sqrt{5}}(2x_1 + x_2) &= \pm\frac{\sqrt{6}}{2}\frac{1}{\sqrt{5}}(-x_1 + 2x_2) \\ 4x_1 + 2x_2 &= \mp\sqrt{6}x_1 \pm 2\sqrt{6}x_2 \\ x_2(2 \mp 2\sqrt{6}) &= (-4 \mp \sqrt{6})x_1 \\ x_2 &= \frac{-4 \mp \sqrt{6}}{2 \mp 2\sqrt{6}}x_1 \end{aligned}$$



Plotting the asymptotes and copying the graph from above onto the y_1 and y_2 axes gives the picture to the right.

Notice that in the last two examples, the change of variables we used actually corresponded to a reflection and a rotation.

Section 10.4 Problems

1. Sketch the graph of each of the following equations showing both the original and new axes. For any hyperbola, find the equation of the asymptotes.

   (a) $x_1^2 + 8x_1x_2 + x_2^2 = 6$ (b) $3x_1^2 - 2x_1x_2 + 3x_2^2 = 12$
   (c) $-4x_1^2 + 4x_1x_2 - 7x_2^2 = 8$ (d) $-3x_1^2 - 4x_1x_2 = 4$
   (e) $-x_1^2 + 4x_1x_2 + 2x_2^2 = 6$ (f) $4x_1^2 + 4x_1x_2 + 4x_2^2 = 12$

10.5 Optimizing Quadratic Forms

In calculus we use quadratic forms to classify critical points as local minimums and maximums. However, in many other applications of quadratic forms, we actually want to find the maximum and/or minimum of the quadratic form subject to a constraint. Most of the time, it is possible to use a change of variables so that the constraint is $\|\vec{x}\| = 1$. That is, given a quadratic form $Q(\vec{x}) = \vec{x}^T A\vec{x}$ on $\mathbb{R}^n$ we want to find the maximum and minimum value of $Q(\vec{x})$ subject to $\|\vec{x}\| = 1$. For ease, we instead use the equivalent constraint

$$1 = \|\vec{x}\|^2 = \vec{x}_1^2 + \cdots \vec{x}_n^2$$

To develop a procedure for solving this problem in general, we begin by looking at a couple of examples.

EXAMPLE 1 Find the maximum and minimum of $Q(x_1, x_2) = 2x_1^2 + 3x_2^2$ subject to the constraint $x_1^2 + x_2^2 = 1$.

Solution: Since we don't have a general method for solving this type of problem, we resort to the basics. We will first try a few points (x_1, x_2) that satisfy the constraint. We get

$$\begin{aligned}
Q(1,0) &= 2 \\
Q\left(\frac{\sqrt{3}}{2}, \frac{1}{2}\right) &= \frac{9}{4} \\
Q\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) &= \frac{5}{2} \\
Q\left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right) &= \frac{11}{4} \\
Q(0,1) &= 3
\end{aligned}$$

Although we have only tried a few points, we may guess that 3 is the maximum value. Since we have already found that $Q(0, 1) = 3$, to prove that 3 is the maximum we just need to show that 3 is an upper bound for $Q(x_1, x_2)$ subject to $x_1^2 + x_2^2 = 1$. Indeed, we have that

$$Q(x_1, x_2) = 2x_1^2 + 3x_2^2 \leq 3x_1^2 + 3x_2^2 = 3(x_1^2 + x_2^2) = 3$$

Hence, 3 is the maximum.

Similarly, we have

$$Q(x_1, x_2) = 2x_1^2 + 3x_2^2 \geq 2x_1^2 + 2x_2^2 = 2(x_1^2 + x_2^2) = 2$$

Hence, 2 is a lower bound for $Q(x_1, x_2)$ subject to the constraint and $Q(1, 0) = 2$, so 2 is the minimum.

This example was extremely easy since we had no cross-terms in the quadratic form. We now look at what happens if we have cross-terms.

EXAMPLE 2 Find the maximum and minimum of $Q(x_1, x_2) = 7x_1^2 + 8x_1x_2 + x_2^2$ subject to the constraint $x_1^2 + x_2^2 = 1$.

Solution: We first try some values of (x_1, x_2) that satisfy the constraint. We have

$$\begin{aligned}
Q(1,0) &= 7\\
Q\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) &= 8\\
Q(0,1) &= 1\\
Q\left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) &= 0\\
Q(-1,0) &= 7
\end{aligned}$$

From this we might think that the maximum is 8 and the minimum is 0, although we again realize that we have tested very few points. Indeed, if we try more points, we quickly find that these are not the maximum and minimum. So, instead of testing points, we need to try to think of where the maximum and minimum should occur. In the previous example, they occurred at $(1,0)$ and $(0,1)$ which are on the principal axes. Thus, it makes sense to consider the principal axes of this quadratic form as well.

We find that $Q(x_1, x_2)$ has eigenvalues $\lambda = 9$ and $\lambda = -1$ with corresponding unit eigenvectors $\vec{v}_1 = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}$ and $\vec{v}_2 = \begin{bmatrix} 1/\sqrt{5} \\ -2/\sqrt{5} \end{bmatrix}$. Taking these for (x_1, x_2) we get

$$Q\left(\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right) = 9, \qquad Q\left(\frac{1}{\sqrt{5}}, \frac{-2}{\sqrt{5}}\right) = -1$$

Moreover, taking

$$\vec{y} = P^T\vec{x} = \frac{1}{\sqrt{5}}\begin{bmatrix} 2 & 1 \\ 1 & -2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{\sqrt{5}}\begin{bmatrix} 2x_1 + x_2 \\ x_1 - 2x_2 \end{bmatrix}$$

we get

$$\begin{aligned}
Q(x_1, x_2) = \vec{y}^T D\vec{y} &= 9\left(\frac{2}{\sqrt{5}}x_1 + \frac{1}{\sqrt{5}}x_2\right)^2 - \left(\frac{1}{\sqrt{5}}x_1 - \frac{2}{\sqrt{5}}x_2\right)^2\\
&\leq 9\left(\frac{2}{\sqrt{5}}x_1 + \frac{1}{\sqrt{5}}x_2\right)^2 + 9\left(\frac{1}{\sqrt{5}}x_1 - \frac{2}{\sqrt{5}}x_2\right)^2\\
&= 9\left[\left(\frac{2}{\sqrt{5}}x_1 + \frac{1}{\sqrt{5}}x_2\right)^2 + \left(\frac{1}{\sqrt{5}}x_1 - \frac{2}{\sqrt{5}}x_2\right)^2\right] = 9
\end{aligned}$$

since

$$\left(\frac{2}{\sqrt{5}}x_1 + \frac{1}{\sqrt{5}}x_2\right)^2 + \left(\frac{1}{\sqrt{5}}x_1 - \frac{2}{\sqrt{5}}x_2\right)^2 = x_1^2 + x_2^2 = 1$$

Similarly, we can show that $Q(x_1, x_2) \geq -1$. Thus, the maximum of $Q(x_1, x_2)$ subject to $\|\vec{x}\| = 1$ is 9 and the minimum is -1.

We generalize what we learned in the examples above to get the following theorem.

THEOREM 1 Let $Q(\vec{x})$ be a quadratic form on $\mathbb{R}^n$ with corresponding symmetric matrix A. The maximum value and minimum value of $Q(\vec{x})$ subject to the constraint $\|\vec{x}\| = 1$ are the greatest and least eigenvalues of A respectively. Moreover, these values occur when $\vec{x}$ is taken to be a corresponding unit eigenvector of the eigenvalue.

Proof: Since A is symmetric, we can find an orthogonal matrix $P = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix}$ which orthogonally diagonalizes A where $\vec{v}_1, \dots, \vec{v}_n$ are arranged so that the corresponding eigenvalues $\lambda_1, \dots, \lambda_n$ satisfy $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$.

Hence, there exists a diagonal matrix D such that

$$\vec{x}^T A \vec{x} = \vec{y}^T D \vec{y}$$

where $\vec{x} = P\vec{y}$. Since P is orthogonal we have that

$$\|\vec{y}\| = \|P\vec{y}\| = \|\vec{x}\| = 1$$

Thus, the quadratic form $\vec{y}^T D\vec{y}$ subject to $\|\vec{y}\| = 1$ takes the same set of values as the quadratic form $\vec{x}^T A\vec{x}$ subject to $\|\vec{x}\| = 1$. Hence, we just need to find the maximum and minimum of $\vec{y}^T D\vec{y}$ subject to $\|\vec{y}\| = 1$.

Since $y_1^2 + \cdots + y_n^2 = \|\vec{y}\|^2 = 1$ we have

$$\vec{y}^T D\vec{y} = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2 \leq \lambda_1 y_1^2 + \cdots + \lambda_1 y_n^2 = \lambda_1(y_1^2 + \cdots + y_n^2) = \lambda_1$$

and $\vec{y}^T D\vec{y} = \lambda_1$ with $\vec{y} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$. Hence λ_1 is the maximum value.

Similarly,

$$\vec{y}^T D\vec{y} = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2 \geq \lambda_n y_1^2 + \cdots + \lambda_n y_n^2 = \lambda_n(y_1^2 + \cdots + y_n^2) = \lambda_n$$

and $\vec{y}^T D\vec{y} = \lambda_n$ with $\vec{y} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$. Hence λ_n is the minimum value. $\square$

EXAMPLE 3 Let $A = \begin{bmatrix} -3 & 0 & -1 \\ 0 & 3 & 0 \\ -1 & 0 & -3 \end{bmatrix}$. Find the maximum and minimum value of the quadratic form $Q(\vec{x}) = \vec{x}^T A\vec{x}$ subject to $x_1^2 + x_2^2 + x_3^2 = 1$.

Solution: The characteristic polynomial of A is

$$C(\lambda) = \begin{vmatrix} -3-\lambda & 0 & -1 \\ 0 & 3-\lambda & 0 \\ -1 & 0 & -3-\lambda \end{vmatrix} = -(\lambda - 3)(\lambda + 4)(\lambda + 2)$$

Thus, by Theorem 1 the maximum of $Q(\vec{x})$ subject to the constraint is 3 and the minimum is -4.

EXAMPLE 4 Find the maximum and minimum value of the quadratic form

$$Q(x_1, x_2, x_3) = 4x_1^2 - 2x_1x_2 + 2x_1x_3 + 4x_2^2 - 2x_2x_3 + 4x_3^2$$

subject to $x_1^2 + x_2^2 + x_3^2 = 1$. Find vectors at which the maximum and minimum occur.

Solution: The characteristic polynomial of the corresponding symmetric matrix is

$$C(\lambda) = \begin{vmatrix} 4-\lambda & 1 & 1 \\ 1 & 4-\lambda & -1 \\ 1 & -1 & 4-\lambda \end{vmatrix} = -(\lambda-5)^2(\lambda-2)$$

We find that a unit eigenvectors corresponding to $\lambda_1 = 5$ and $\lambda_2 = 2$ are $\vec{v}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}$ and $\vec{v}_2 = \begin{bmatrix} -1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$ respectively. Thus, by Theorem 1 the maximum of $Q(\vec{x})$ subject to $x_1^2 + x_2^2 + x_3^2 = 1$ is 5 and occurs when $\vec{x} = \vec{v}_1$, and the minimum is 2 and occurs when $\vec{x} = \vec{v}_2$.

Section 10.5 Problems

1. Find the maximum and minimum of each quadratic form subject to $\|\vec{x}\| = 1$.

   (a) $Q(\vec{x}) = 4x_1^2 - 2x_1x_2 + 4x_2^2$

   (b) $Q(\vec{x}) = 3x_1^2 + 10x_1x_2 - 3x_2^2$

   (c) $Q(\vec{x}) = -x_1^2 + 8x_2^2 + 4x_2x_3 + 11x_3^2$

   (d) $Q(\vec{x}) = 3x_1^2 - 4x_1x_2 + 8x_1x_3 + 6x_2^2 + 4x_2x_3 + 3x_3^2$

   (e) $Q(\vec{x}) = 2x_1^2 + 2x_1x_2 + 2x_1x_3 + 2x_2^2 + 2x_2x_3 + 2x_3^2$

10.6 Singular Value Decomposition

We have now seen that finding the maximum and minimum of a quadratic form subject to $\|\vec{x}\| = 1$ is very easy. However, in many applications we are not lucky enough to get a symmetric matrix; in fact, we often do not even get a square matrix. Thus, it is very desirable to derive a similar method for finding the maximum and minimum of $\|A\vec{x}\|$ for an $m \times n$ matrix A subject to $\|\vec{x}\| = 1$. This derivation will lead us to a very important matrix factorization known as the singular value decomposition. We again begin by looking at an example.

EXAMPLE 1

Consider a linear mapping $f : \mathbb{R}^3 \to \mathbb{R}^2$ defined by $f(\vec{x}) = A\vec{x}$ with $A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & -1 \end{bmatrix}$. Since the range is now a subspace of $\mathbb{R}^2$, we want to find the maximum length of $A\vec{x}$ subject to $\|\vec{x}\| = 1$. That is, we want to maximize $\|A\vec{x}\|$. Observe that

$$\|A\vec{x}\|^2 = (A\vec{x})^T(A\vec{x}) = \vec{x}^T A^T A\vec{x}$$

Since A^TA is symmetric this is a quadratic form. Hence, using Theorem 10.5.1, to maximize $\|A\vec{x}\|$ we just need to find the square root of the largest eigenvalue of A^TA.

We have

$$A^TA = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 1 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

The characteristic polynomial of A^TA is $C(\lambda) = -\lambda(\lambda - 3)(\lambda - 2)$. Thus, the eigenvalues of A^TA are 3, 2 and 0. Hence, the maximum of $\|A\vec{x}\|$ subject to $\|\vec{x}\| = 1$ is $\sqrt{3}$ and the minimum is 0. Moreover, these occur at the corresponding eigenvectors $\begin{bmatrix} -1/\sqrt{3} \\ -1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$ and $\begin{bmatrix} -1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}$ of A^TA.

In the example, we found that the eigenvalues of A^TA were all non-negative. This was important as we had to take the square root of them to maximize/minimize $\|A\vec{x}\|$. We now prove that this is always the case.

THEOREM 1

Let A be an $m\times n$ matrix and let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of A^TA with corresponding unit eigenvectors $\vec{v}_1, \dots, \vec{v}_n$. Then $\lambda_1, \dots, \lambda_n$ are all non-negative. In particular,

$$\|A\vec{v}_i\| = \sqrt{\lambda_i}$$

Proof: For $1 \le i \le n$ we have $A^TA\vec{v}_i = \lambda_i\vec{v}_i$ and hence

$$\|A\vec{v}_i\|^2 = (A\vec{v}_i)^T A\vec{v}_i = v_i^T A^T A\vec{v}_i = \vec{v}_i^T(\lambda_i\vec{v}_i) = \lambda_i\vec{v}_i^T\vec{v}_i = \lambda_i\|\vec{v}_i\| = \lambda_i$$

since $\|\vec{v}_i\| = 1$. $\square$

Observe from the example above that the square root of the largest and smallest eigenvalues of A^TA are the maximum and minimum of values of $\|A\vec{x}\|$ subject to $\|\vec{x}\| = 1$. So, these are behaving like the eigenvalues of a symmetric matrix. This motivates the following definition.

DEFINITION
Singular Values

The **singular values** $\sigma_1, \dots, \sigma_n$ of an $m \times n$ matrix A are the square roots of the eigenvalues of A^TA arranged so that $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_n \ge 0$.

EXAMPLE 2 Find the singular values of $A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \end{bmatrix}$.

Solution: We have $A^TA = \begin{bmatrix} 5 & 5 & 5 \\ 5 & 5 & 5 \\ 5 & 5 & 5 \end{bmatrix}$. The characteristic polynomial is $C(\lambda) = -\lambda^2(\lambda - 15)$ so the eigenvalues are (from greatest to least) $\lambda_1 = 15$, $\lambda_2 = 0$, and $\lambda_3 = 0$. Thus, the singular values of A are $\sigma_1 = \sqrt{15}$, $\sigma_2 = 0$, and $\sigma_3 = 0$.

EXAMPLE 3 Find the singular values of $B = \begin{bmatrix} 1 & 1 \\ -1 & 2 \\ -1 & 1 \end{bmatrix}$.

Solution: We have $B^TB = \begin{bmatrix} 3 & -2 \\ -2 & 6 \end{bmatrix}$. The characteristic polynomial is $C(\lambda) = (\lambda - 2)(\lambda - 7)$. Hence, the eigenvalues of B^TB are $\lambda_1 = 7$ and $\lambda_2 = 2$, so the singular values of B are $\sigma_1 = \sqrt{7}$ and $\sigma_2 = \sqrt{2}$.

EXAMPLE 4 Find the singular values of $A = \begin{bmatrix} 2 & -1 \\ 1 & 0 \\ 1 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 2 & 1 & 1 \\ -1 & 0 & 2 \end{bmatrix}$.

Solution: We have $A^TA = \begin{bmatrix} 6 & 0 \\ 0 & 5 \end{bmatrix}$. Hence, the eigenvalues of A^TA are $\lambda_1 = 6$ and $\lambda_2 = 5$. Thus, the singular values of A are $\sigma_1 = \sqrt{6}$ and $\sigma_2 = \sqrt{5}$.

We have $B^TB = \begin{bmatrix} 5 & 2 & 0 \\ 2 & 1 & 1 \\ 0 & 1 & 5 \end{bmatrix}$. The characteristic polynomial of B^TB is

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

Hence, the eigenvalues of B^TB are $\lambda_1 = 6$, $\lambda_2 = 5$, and $\lambda_3 = 0$. Thus, the singular values of B are $\sigma_1 = \sqrt{6}$, $\sigma_2 = \sqrt{5}$, and $\sigma_3 = 0$.

Recall that 0 is an eigenvalue of an $n \times n$ matrix A if and only if $\operatorname{rank} A \neq n$. In particular, we know that the number of non-zero eigenvalues of A^TA is the rank of A^TA. Thus, it is natural to ask if there is a relationship between the number of non-zero singular values of an $m \times n$ matrix A and the rank of A. In the problems above, we see that we have the rank of A equals the number of non-zero singular values. To prove this result in general, we use the following lemma.

THEOREM 2 Let A be an $m \times n$ matrix. Then $\operatorname{rank}(A^T A) = \operatorname{rank}(A)$.

Proof: If $A\vec{x} = \vec{0}$, then $A^T A\vec{x} = A^T\vec{0} = \vec{0}$. Hence the nullspace of A is a subset of the nullspace of $A^T A$. On the other hand, consider $A^T A\vec{x} = \vec{0}$. Then

$$\|A\vec{x}\|^2 = (A\vec{x}) \cdot (A\vec{x}) = (A\vec{x})^T(A\vec{x}) = \vec{x}^T A^T A\vec{x} = \vec{x}^T\vec{0} = 0$$

Thus, $A\vec{x} = \vec{0}$. Hence, the nullspace of $A^T A$ is a subset of the nullspace of A.

Therefore, $\dim(\operatorname{Null}(A^T A)) = \dim(\operatorname{Null}(A))$ and so by the Rank-Nullity Theorem

$$\operatorname{rank}(A^T A) = n - \dim(\operatorname{Null}(A^T A)) = n - \dim(\operatorname{Null}(A)) = \operatorname{rank}(A)$$

□

COROLLARY 3 If A is an $m \times n$ matrix and $\operatorname{rank}(A) = r$, then A has r non-zero singular values.

We now want to extend the similarity of singular values and eigenvalues by defining singular vectors. Observe that if A is an $m \times n$ matrix with $m \neq n$, then we cannot have $A\vec{v} = \sigma\vec{v}$ since $A\vec{v} \in \mathbb{R}^m$. Hence, the best we can do to match the definition of eigenvalues and eigenvectors is to pick suitable non-zero vectors $\vec{v} \in \mathbb{R}^n$ and $\vec{u} \in \mathbb{R}^m$ such that $A\vec{v} = \sigma\vec{u}$.

By definition, for any non-zero singular value σ of A there is a vector $\vec{v} \neq \vec{0}$ such that $A^T A\vec{v} = \sigma^2\vec{v}$. Thus, if we have $A\vec{v} = \sigma\vec{u}$, then we have $A^T A\vec{v} = A^T(\sigma\vec{u})$ so $\sigma^2\vec{v} = \sigma A^T\vec{u}$. Dividing by σ, we see that we must also have $A^T\vec{u} = \sigma\vec{v}$. Moreover, by Theorem 1, if $\vec{v}$ is a unit eigenvector of $A^T A$, then we get that $\vec{u} = \frac{1}{\sigma}A\vec{v}$ is also a unit vector.

Therefore, for a non-zero singular value σ of A, we will want unit vectors $\vec{v}$ and $\vec{u}$ such that

$$A\vec{v} = \sigma\vec{u} \qquad \text{and} \qquad A^T\vec{u} = \sigma\vec{v}$$

However, our derivation does not work for $\sigma = 0$. In this case, we will see that we are satisfied with just one of these conditions being satisfied.

DEFINITION
Singular Vectors

Let A be an $m \times n$ matrix. If $\vec{v} \in \mathbb{R}^n$ and $\vec{u} \in \mathbb{R}^m$ are unit vectors and $\sigma \neq 0$ is a singular value of A such that

$$A\vec{v} = \sigma\vec{u} \qquad \text{and} \qquad A^T\vec{u} = \sigma\vec{v}$$

then we say that $\vec{u}$ is a **left singular vector** of A and $\vec{v}$ is a **right singular vector** of A. Additionally, if $\vec{u}$ is a unit vector such that $A^T\vec{u} = \vec{0}$, then $\vec{u}$ is a **left singular vector** of A. If $\vec{v}$ is a unit vector such that $A\vec{v} = \vec{0}$, then $\vec{v}$ is a **right singular vector** of A.

This definition of right singular vectors, not only preserves our relationship of these vectors and the eigenvectors of $A^T A$, but we also get the corresponding result for the left singular vectors and the eigenvectors of AA^T.

THEOREM 4 Let A be an $m \times n$ matrix. If $\vec{v}$ is a right singular vector of A, then $\vec{v}$ is an eigenvector of A^TA. If $\vec{u}$ is a left singular vector of A, then $\vec{u}$ is an eigenvector of AA^T.

Now that we have singular values and singular vectors mimicking eigenvalues and eigenvectors, we wish to try to mimic orthogonal diagonalization. Let A be an $m \times n$ matrix. As was the case with singular vectors, if $m \neq n$, then P^TAP is not defined for a square matrix P. Thus, we want to find an $m \times m$ orthogonal matrix U and an $n \times n$ orthogonal matrix V such that $U^TAV = \Sigma$, where Σ is an $m \times n$ "diagonal" matrix.

To do this, we require that for any $m \times n$ matrix A there exists an orthonormal basis for $\mathbb{R}^n$ of right singular vectors and an orthonormal basis for $\mathbb{R}^m$ of left singular vectors. We now prove that these always exist.

LEMMA 5 Let A be an $m \times n$ matrix with $\operatorname{rank}(A) = r$ and suppose that $\{\vec{v}_1, \ldots, \vec{v}_n\}$ is an orthonormal basis for $\mathbb{R}^n$ consisting of the eigenvectors of A^TA arranged so that the corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$ are arranged from greatest to least and let $\sigma_1, \ldots, \sigma_n$ be the singular values of A. Then $\{\frac{A\vec{v}_1}{\sigma_1}, \ldots, \frac{A\vec{v}_r}{\sigma_r}\}$ is an orthonormal basis for $\operatorname{Col} A$.

Proof: Since A has rank r, we know that A^TA has rank r and so A^TA has r non-zero eigenvalues. Thus, $\sigma_1, \ldots, \sigma_r$ are the r non-zero singular values of A.

Now observe that for $1 \leq i, j \leq r$, $i \neq j$ we have

$$A\vec{v}_i \cdot A\vec{v}_j = (A\vec{v}_i)^T A\vec{v}_j = \vec{v}_i A^T A\vec{v}_j = \vec{v}_i^T \lambda_j \vec{v}_j = \lambda_j \vec{v}_i^T \vec{v}_j = 0$$

since $\{\vec{v}_1, \ldots, \vec{v}_r\}$ is an orthonormal set. Hence, $\{A\vec{v}_1, \ldots, A\vec{v}_r\}$ is orthogonal and so $\{\frac{A\vec{v}_1}{\sigma_1}, \ldots, \frac{A\vec{v}_r}{\sigma_r}\}$ is orthonormal by Theorem 1. Moreover, we know that $\dim \operatorname{Col}(A) = r$, so this is an orthonormal basis for $\operatorname{Col} A$. $\square$

THEOREM 6 Let A be an $m \times n$ matrix with rank r. There exists an orthonormal basis $\{\vec{v}_1, \ldots, \vec{v}_n\}$ of $\mathbb{R}^n$ of right singular vectors of A and an orthonormal basis $\{\vec{u}_1, \ldots, \vec{u}_m\}$ of $\mathbb{R}^m$ of left singular vectors of A.

Proof: By definition eigenvectors of A^TA are right singular vectors of A. Hence, since A^TA is symmetric, we can find an orthonormal basis $\{\vec{v}_1, \ldots, \vec{v}_n\}$ of right singular vector of A.

By Lemma 5, the left singular vector $\vec{u}_i = \frac{1}{\sigma_i} A\vec{v}_i$, $1 \leq i \leq r$, form an orthonormal basis for $\operatorname{Col}(A)$. Also, by definition, a left singular vector $\vec{u}_j$ of A corresponding to $\sigma = 0$ lies in the nullspace of A^T. Hence, by the Fundamental Theorem of Linear Algebra, if $\{\vec{u}_{r+1}, \ldots, \vec{u}_m\}$ is an orthonormal basis for the nullspace of A^T, then $\{\vec{u}_1, \ldots, \vec{u}_m\}$ is an orthonormal basis for $\mathbb{R}^m$ of left singular vectors of A. $\square$

Thus, for any $m \times n$ matrix A with rank r we have an orthonormal basis $\{\vec{v}_1, \dots, \vec{v}_n\}$ for $\mathbb{R}^n$ of right singular vectors of A corresponding to the n singular values $\sigma_1, \dots, \sigma_n$ of A and an orthonormal basis $\{\vec{u}_1, \dots, \vec{u}_m\}$ for $\mathbb{R}^m$ of left singular vectors such that

$$\begin{aligned} A\begin{bmatrix}\vec{v}_1 & \cdots & \vec{v}_n\end{bmatrix} &= \begin{bmatrix}A\vec{v}_1 & \cdots A\vec{v}_r & A\vec{v}_{r+1} & \cdots & A\vec{v}_n\end{bmatrix} \\ &= \begin{bmatrix}\sigma_1\vec{u}_1 & \cdots & \sigma_r\vec{u}_r & \vec{0} & \cdots & \vec{0}\end{bmatrix} \\ &= \begin{bmatrix}\vec{u}_1 & \cdots & \vec{u}_m\end{bmatrix}\Sigma \end{aligned}$$

where Σ is the $m \times n$ matrix with $(\Sigma)_{ii} = \sigma_i$ for $1 \le i \le r$ and all other entries of Σ are 0.

Hence, we have that there exists an orthogonal matrix V and orthogonal matrix U such that $AV = U\Sigma$. Instead of writing this as $U^T AV = \Sigma$, we typically, write this as $A = U\Sigma V^T$ to get a matrix decomposition of A.

DEFINITION

Singular Value Decomposition (SVD)

A **singular value decomposition** of an $m \times n$ matrix A is a factorization of the form

$$A = U\Sigma V^T$$

where U is an orthogonal matrix containing left singular vectors of A, V is an orthogonal matrix containing right singular vectors of A, and Σ is the $m \times n$ matrix with $(\Sigma)_{ii} = \sigma_i$ for $1 \le i \le r$ and all other entries of Σ are 0.

ALGORITHM

To find a singular value decomposition of a matrix, we follow what we did above.

(1) Find the eigenvalues $\lambda_1, \dots, \lambda_n$ of $A^T A$ arranged from greatest to least and a corresponding set of orthonormal eigenvectors $\{\vec{v}_1, \dots, \vec{v}_n\}$.

(2) Let $V = \begin{bmatrix}\vec{v}_1 & \cdots & \vec{v}_n\end{bmatrix}$ and let Σ be the $m \times n$ matrix whose first r diagonal entries are the r singular values of A arranged from greatest to least and all other entries 0.

(3) Find left singular vectors of A by computing $\vec{u}_i = \frac{1}{\sigma_i} A\vec{v}_i$ for $1 \le i \le r$. Then extend the set $\{\vec{u}_1, \dots, \vec{u}_r\}$ to an orthonormal basis for $\mathbb{R}^m$. Take $U = \begin{bmatrix}\vec{u}_1 & \cdots & \vec{u}_m\end{bmatrix}$.

Then $A = U\Sigma V^T$.

REMARK

As indicated in the proof of Theorem 5. One way to extend $\{\vec{u}_1, \dots, \vec{u}_r\}$ to an orthonormal basis for $\mathbb{R}^m$ is by finding an orthonormal basis for Null(A^T).

EXAMPLE 5

Find a singular value decomposition of $A = \begin{bmatrix} 1 & -1 & 3 \\ 3 & 1 & 1 \end{bmatrix}$.

Solution: We find that $A^TA = \begin{bmatrix} 10 & 2 & 6 \\ 2 & 2 & -2 \\ 6 & -2 & 10 \end{bmatrix}$ has eigenvalues (ordered from greatest to least) of $\lambda_1 = 16$, $\lambda_2 = 6$ and $\lambda_3 = 0$. Corresponding orthonormal eigenvectors are

$$\vec{v}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ -1/\sqrt{3} \end{bmatrix}, \quad \vec{v}_3 = \begin{bmatrix} 1/\sqrt{6} \\ -2/\sqrt{6} \\ -1/\sqrt{6} \end{bmatrix}$$

So we let $V = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \\ 0 & 1/\sqrt{3} & -2/\sqrt{6} \\ 1/\sqrt{2} & -1/\sqrt{3} & -1/\sqrt{6} \end{bmatrix}$.

The singular values of A are $\sigma_1 = \sqrt{16} = 4$, and $\sigma_2 = \sqrt{6}$, so we let $\Sigma = \begin{bmatrix} 4 & 0 & 0 \\ 0 & \sqrt{6} & 0 \end{bmatrix}$.

Next, we compute

$$\vec{u}_1 = \frac{1}{\sigma_1} A\vec{v}_1 = \frac{1}{4}\begin{bmatrix} 4/\sqrt{2} \\ 4/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

$$\vec{u}_2 = \frac{1}{\sigma_2} A\vec{v}_2 = \frac{1}{\sqrt{6}}\begin{bmatrix} -3/\sqrt{3} \\ 3/\sqrt{3} \end{bmatrix} = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

Since this forms a basis for $\mathbb{R}^2$ we take $U = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$.

Then, we have a singular value decomposition $A = U\Sigma V^T$.

EXAMPLE 6

Find a singular value decomposition of $B = \begin{bmatrix} 2 & -4 \\ 2 & 2 \\ -4 & 0 \\ 1 & 4 \end{bmatrix}$.

Solution: We have $B^TB = \begin{bmatrix} 25 & 0 \\ 0 & 36 \end{bmatrix}$. The eigenvalues are $\lambda_1 = 36$ and $\lambda_2 = 25$ with corresponding orthonormal eigenvectors $\begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. So $V = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

The singular values of B are $\sigma_1 = 6$ and $\sigma_2 = 5$ so $\Sigma = \begin{bmatrix} 6 & 0 \\ 0 & 5 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$.

We compute

$$\vec{u}_1 = \frac{1}{\sigma_1}B\vec{v}_1 = \frac{1}{6}\begin{bmatrix} -4 \\ 2 \\ 0 \\ 4 \end{bmatrix}$$

$$\vec{u}_2 = \frac{1}{\sigma_2}B\vec{v}_2 = \frac{1}{5}\begin{bmatrix} 2 \\ 2 \\ -4 \\ 1 \end{bmatrix}$$

But, we only have 2 orthogonal vectors in $\mathbb{R}^4$, so we need to extend $\{\vec{u}_1, \vec{u}_2\}$ to an orthonormal basis for $\mathbb{R}^4$. We know that we can complete a basis for $\mathbb{R}^4$ by finding an orthonormal basis for $\operatorname{Null}(B^T)$.

Applying the Gram-Schmidt algorithm to a basis for $\operatorname{Null}(B^T)$ we get vectors

$$\vec{u}_3 = \frac{1}{3}\begin{bmatrix} 1 \\ -2 \\ 0 \\ 2 \end{bmatrix}, \quad \vec{u}_4 = \frac{1}{15}\begin{bmatrix} 8 \\ 8 \\ 9 \\ 4 \end{bmatrix}.$$

We let $U = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \vec{u}_3 & \vec{u}_4 \end{bmatrix}$ and then we have a singular value decomposition $U\Sigma V^T$ of B.

EXAMPLE 7 Find a singular value decomposition for $A = \begin{bmatrix} 1 & 1 \\ 2 & 2 \\ -1 & -1 \end{bmatrix}$.

Solution: We have $A^TA = \begin{bmatrix} 6 & 6 \\ 6 & 6 \end{bmatrix}$. Thus, by inspection, the eigenvalues of A $\lambda_1 = 12$ and $\lambda_2 = 0$, with corresponding unit eigenvectors $\vec{v}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ and $\vec{v}_2 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ respectively. Hence, the singular values of A are $\sigma_1 = \sqrt{12}$ and $\sigma_2 = 0$. Thus, we have

$$V = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \sqrt{12} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

We compute

$$\vec{u}_1 = \frac{1}{\sigma_1}A\vec{v}_1 = \frac{1}{\sqrt{25}}\begin{bmatrix} 2 \\ 4 \\ -2 \end{bmatrix}$$

Since we only have one vectors in $\mathbb{R}^3$, we need to extend $\{\vec{u}_1\}$ to an orthonormal basis for $\mathbb{R}^3$. To do this, we find an orthonormal basis for $\operatorname{Null}(A^T)$.

We observe that a basis for Null(A^T) is $\left\{\begin{bmatrix}-2\\1\\0\end{bmatrix}, \begin{bmatrix}1\\0\\1\end{bmatrix}\right\}$. Applying the Gram-Schmidt algorithm to this basis and normalizing gives

$$\vec{u}_2 = \frac{1}{\sqrt{6}}\begin{bmatrix}-2\\1\\0\end{bmatrix}, \quad \vec{u}_3 = \frac{1}{\sqrt{30}}\begin{bmatrix}1\\2\\5\end{bmatrix}$$

We let $U = \begin{bmatrix}\vec{u}_1 & \vec{u}_2 & \vec{u}_3\end{bmatrix}$ and then we have a singular value decomposition $U\Sigma V^T$ of A.

Section 10.6 Problems

1. Find a singular value decomposition for the following matrices.

   (a) $\begin{bmatrix}2 & 2\\-1 & 2\end{bmatrix}$ (b) $\begin{bmatrix}3 & -4\\8 & 6\end{bmatrix}$ (c) $\begin{bmatrix}2 & 3\\0 & 2\end{bmatrix}$

   (d) $\begin{bmatrix}1 & 2\\1 & 2\\1 & 2\\1 & 2\end{bmatrix}$ (e) $\begin{bmatrix}1 & 1 & 1 & 1\\2 & 2 & 2 & 2\end{bmatrix}$ (f) $\begin{bmatrix}4 & 4 & -2\\3 & 2 & 4\end{bmatrix}$

2. Let A be an $m \times n$ matrix and let P be an $m \times m$ orthogonal matrix. Show that PA has the same singular values as A.

3. Let $U\Sigma V^T$ be a singular value decomposition for an $m \times n$ matrix A with rank r. Find, with proof, an orthonormal basis for Row(A), Col(A), Null(A) and Null(A^T) from the columns of U and V.

4. Let $A = \begin{bmatrix}2 & 1 & 1\\0 & 2 & 0\\0 & 0 & 2\end{bmatrix}$ be the standard matrix of a linear mapping L. (Observe that A is not diagonalizable).

   (a) Find a basis $\mathcal{B}$ for $\mathbb{R}^3$ of right singular vectors of A.

   (b) Find a basis $\mathcal{C}$ for $\mathbb{R}^3$ of left singular vectors of A.

   (c) Determine ${}_{\mathcal{C}}[L]_{\mathcal{B}}$.

Chapter 11

Complex Vector Spaces

Our goal in this chapter is to extend everything we have done with real vector spaces to complex vector spaces. That is, vector spaces which use complex numbers for their scalars instead of just real numbers. We begin with a very brief review of the basic operations on complex numbers.

11.1 Complex Number Review

Recall that a complex number is a number of the form $z = a + bi$ where $a, b \in \mathbb{R}$ and $i^2 = -1$. The set of all complex numbers is denoted $\mathbb{C}$. We call a the real part of z and denote it by

$$\operatorname{Re} z = a$$

and we call b the imaginary part of z and denote it by

$$\operatorname{Im} z = b$$

It is very important to notice that the real numbers are a subset of the complex numbers. In particular, every real number a is the complex number $a + 0i$. If the real part of a complex number z is 0, then we say that z is imaginary. Any complex number which has non-zero imaginary part is said to be non-real.

We define addition and subtraction of complex numbers by

$$(a + bi) \pm (c + di) = (a \pm c) + (b \pm d)i$$

EXAMPLE 1 We have

$$\begin{aligned}(3 + 4i) + (2 - 3i) &= (3 + 2) + (4 - 3)i = 5 + i\\ (1 - 3i) - (4 + 2i) &= (1 - 4) + (-3 - 2)i = -3 - 5i\end{aligned}$$

Observe that addition of complex numbers is defined by adding the real parts and adding the imaginary parts of the complex numbers. That is, we add complex numbers component wise just like vectors in $\mathbb{R}^n$.

In fact, the set $\mathbb{C}$ with this rule for addition and with scalar multiplication by a real scalar defined by

$$t(a + bi) = ta + tbi, \quad t \in \mathbb{R}$$

forms a vector space over $\mathbb{R}$. Let's find a basis for this real vector space. Notice that every vector in $\mathbb{C}$ can be written as $a + bi$ with $a, b \in \mathbb{R}$, so $\{1, i\}$ spans $\mathbb{C}$. Also, $\{1, i\}$ is clearly linearly independent as the only solution to

$$t_1 1 + t_2 i = 0 + 0i$$

is $t_1 = t_2 = 0$ since t_1 and t_2 are real scalars. Hence, $\mathbb{C}$ forms a two dimensional real vector space. Consequently, it is isomorphic to every other two dimensional real vector space. In particular, it is isomorphic to the plane $\mathbb{R}^2$. As a result, the set $\mathbb{C}$ is often called the complex plane. We generally think of having an isomorphism which maps the complex number $z = a + bi$ to the point (a, b) in $\mathbb{R}^2$.

We can use this isomorphism to define the absolute value of a complex number $z = a + bi$. Recall that the absolute value of a real number x measures the distance that x is from the origin. So, we define the absolute value of z as the distance from the point (a, b) in $\mathbb{R}^2$ to the origin.

DEFINITION
Absolute Value

The **absolute value** or **modulus** of a complex number $z = a + bi$ is defined by

$$|z| = \sqrt{a^2 + b^2}$$

EXAMPLE 2

We have

$$|2 - 3i| = \sqrt{2^2 + (-3)^2} = \sqrt{13}$$

$$\left|\frac{1}{2} + \frac{1}{2}i\right| = \sqrt{\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2} = \sqrt{\frac{1}{2}}$$

Of course, this has the same properties as the real absolute value.

THEOREM 1

Let $w, z \in \mathbb{C}$. Then

(1) $|z|$ is a non-negative real number.
(2) $|z| = 0$ if and only if $z = 0$.
(3) $|w + z| \leq |w| + |z|$

REMARK

If $z = a + bi$ is real, then $b = 0$, so $|z| = |a + 0i| = \sqrt{a^2 + 0^2} = |a|$. Hence, this is a direct generalization of the real absolute value function.

The ability to represent complex numbers as points in the plane is very important in the study of complex numbers. However, viewing $\mathbb{C}$ as a real vector space has one serious limitation; it only defines the multiplication of a real scalar by a complex number and does not immediately give us a way of multiplying two complex numbers together. So, we return to the standard form of complex numbers and define multiplication of complex numbers by using the normal distributive property and the fact that $i^2 = -1$. We get

$$(a + bi)(c + di) = ac + adi + bci + bdi^2 = ac - bd + (ad + bc)i$$

EXAMPLE 3

We have

$$(3 + 4i)(2 - 3i) = 3(2) - 4(-3) + [4(2) + 3(-3)]i = 18 - i$$
$$(1 - 3i)(4 + 2i) = 1(4) - (-3)(2) + [(-3)(4) + 1(2)]i = 10 - 10i$$

We get the following familiar properties.

THEOREM 2

Let $z_1, z_2, z_3 \in \mathbb{C}$, then

(1) $z_1 + z_2 = z_2 + z_1$
(2) $z_1 z_2 = z_2 z_1$
(3) $z_1 + (z_2 + z_3) = (z_1 + z_2) + z_3$
(4) $z_1(z_2 z_3) = (z_1 z_2) z_3$
(5) $z_1(z_2 + z_3) = z_1 z_2 + z_1 z_3$
(6) $|z_1 z_2| = |z_1||z_2|$

We look at one more important example of complex multiplication.

EXAMPLE 4

$$(a + bi)(a - bi) = a^2 + b^2 + [b(a) - a(b)]i = a^2 + b^2 = |a + bi|^2$$

This example motivates the following definition.

DEFINITION
Complex Conjugate

Let $z = a + bi$ be a complex number. Then, the **complex conjugate** of z is defined by

$$\overline{z} = a - bi$$

EXAMPLE 5

We have

$$\begin{aligned}
\overline{4-3i} &= 4+3i\\
\overline{3i} &= -3i\\
\overline{4} &= 4\\
\overline{2+2i} &= 2-2i
\end{aligned}$$

We have many important properties of complex conjugates.

THEOREM 3

Let $w, z \in \mathbb{C}$ with $z = a + bi$, then

(1) $\overline{\overline{z}} = z$
(2) z is real if and only if $\overline{z} = z$
(3) z is imaginary if and only if $\overline{z} = -z$
(4) $\overline{z \pm w} = \overline{z} \pm \overline{w}$
(5) $\overline{zw} = \overline{z}\,\overline{w}$
(6) $z + \overline{z} = 2\operatorname{Re}(z) = 2a$
(7) $z - \overline{z} = i2\operatorname{Im}(z) = 2bi$
(8) $z\overline{z} = a^2 + b^2 = |z|^2$

Since the absolute value of a non-zero complex number is a positive real number, we can use property (8) to divide complex numbers. We demonstrate this with some examples.

EXAMPLE 6

Calculate $\dfrac{2-3i}{1+2i}$.

Solution: To calculate this, we make the denominator real by multiplying the numerator and the denominator by the complex conjugate of the denominator. We get

$$\frac{2-3i}{1+2i} = \frac{2-3i}{1+2i} \cdot \frac{1-2i}{1-2i} = \frac{2-6+(-3-4)i}{1^2+2^2} = \frac{-4+i}{5} = \frac{-4}{5} - \frac{7}{5}i$$

We can easily check this answer. We have

$$(1+2i)\left(\frac{-4}{5} - \frac{7}{5}i\right) = -\frac{4}{5} + \frac{14}{5} + \left[-\frac{8}{5} - \frac{7}{5}\right]i = 2-3i$$

EXAMPLE 7

Calculate $\dfrac{i}{1+i}$.

Solution: We have

$$\frac{i}{1+i} = \frac{i}{1+i} \cdot \frac{1-i}{1-i} = \frac{1+i}{1+1} = \frac{1}{2} + \frac{1}{2}i$$

EXERCISE 1 Calculate $\dfrac{5}{3-4i}$.

In some real world applications, such as problems involving electric circuits, one needs to solve systems of linear equations involving complex numbers. The procedure for solving systems of linear equations is the same, except that we can now multiply a row by a non-zero complex number and we can a complex multiple of one row to another. We demonstrate this with a few examples.

EXAMPLE 8 Solve the system of linear equations

$$\begin{aligned} z_1 + iz_2 + (-1+2i)z_3 &= -1+2i \\ z_2 + 2z_3 &= 2+2i \\ 2z_1 + (-1+2i)z_2 + (-6+4i)z_3 &= -4 \end{aligned}$$

Solution: We row reduce the augmented matrix of the system to RREF.

$$\left[\begin{array}{ccc|c} 1 & i & -1+2i & -1+2i \\ 0 & 1 & 2 & 2+2i \\ 2 & -1+2i & -6+4i & -4 \end{array}\right] \begin{array}{l} \\ \\ R_3 - 2R_1 \end{array} \sim$$

$$\left[\begin{array}{ccc|c} 1 & i & -1+2i & -1+2i \\ 0 & 1 & 2 & 2+2i \\ 0 & -1 & -4 & -2-4i \end{array}\right] \begin{array}{l} R_1 - iR_2 \\ \\ R_3 + R_2 \end{array} \sim$$

$$\left[\begin{array}{ccc|c} 1 & 0 & -1 & 1 \\ 0 & 1 & 2 & 2+2i \\ 0 & 0 & -2 & -2i \end{array}\right] \begin{array}{l} \\ \\ -\frac{1}{2}R_3 \end{array} \sim$$

$$\left[\begin{array}{ccc|c} 1 & 0 & -1 & 1 \\ 0 & 1 & 2 & 2+2i \\ 0 & 0 & 1 & i \end{array}\right] \begin{array}{l} R_1 + R_3 \\ R_2 - 2R_3 \\ \\ \end{array} \sim \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1+i \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & i \end{array}\right]$$

Hence, the solution is $z_1 = 1+i$, $z_2 = 2$, and $z_3 = i$.

REMARK

It is very easy to make calculation mistakes when working with complex numbers. Thus, it is highly recommended that you check your answer whenever possible.

EXAMPLE 9

Solve the system of linear equation

$$\begin{aligned} z_1 + z_2 + iz_3 &= 1 \\ -2iz_1 + z_2 + (1-2i)z_3 &= -2i \\ (1+i)z_1 + (2+2i)z_2 - 2z_3 &= 2 \end{aligned}$$

Solution: We row reduce the augmented matrix of the system to RREF.

$$\left[\begin{array}{ccc|c} 1 & 1 & i & 1 \\ -2i & 1 & 1-2i & -2i \\ 1+i & 2+2i & -2 & 2 \end{array}\right] \begin{array}{c} \\ R_2 + 2iR_1 \\ R_3 + (1+i)R_1 \end{array} \sim$$

$$\left[\begin{array}{ccc|c} 1 & 1 & i & 1 \\ 0 & 1+2i & -1-2i & 0 \\ 0 & 1+i & -1-i & 1-i \end{array}\right] \tfrac{1}{1+2i}R_2 \sim$$

$$\left[\begin{array}{ccc|c} 1 & 1 & i & 1 \\ 0 & 1 & -1 & 0 \\ 0 & 1+i & -1-i & 1-i \end{array}\right] \begin{array}{c} R_1 - R_2 \\ \\ R_3 - (1+i)R_2 \end{array} \sim \left[\begin{array}{ccc|c} 1 & 0 & 1+i & 1 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1-i \end{array}\right]$$

Hence, the system is inconsistent.

EXAMPLE 10

Find the general solution of the system of linear equations

$$\begin{aligned} z_1 - iz_2 &= 1+i \\ (1+i)z_1 + (1-i)z_2 + (1-i)z_3 &= 1+3i \\ 2z_1 - 2iz_2 + (3+i)z_3 &= 1+5i \end{aligned}$$

Solution: We row reduce the augmented matrix of the system to RREF.

$$\left[\begin{array}{ccc|c} 1 & -i & 0 & 1+i \\ 1+i & 1-i & 1-i & 1+3i \\ 2 & -2i & 3+i & 1+5i \end{array}\right] \begin{array}{c} \\ R_2 - (1+i)R_1 \\ R_3 - 2R_1 \end{array} \sim$$

$$\left[\begin{array}{ccc|c} 1 & -i & 0 & 1+i \\ 0 & 0 & 1-i & 1+i \\ 0 & 0 & 3+i & -1+3i \end{array}\right] \tfrac{1}{1-i}R_2 \sim$$

$$\left[\begin{array}{ccc|c} 1 & -i & 0 & 1+i \\ 0 & 0 & 1 & i \\ 0 & 0 & 3+i & -1+3i \end{array}\right] \begin{array}{c} \\ \\ R_3 - (3+i)R_2 \end{array} \sim \left[\begin{array}{ccc|c} 1 & -i & 0 & 1+i \\ 0 & 0 & 1 & i \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Thus, z_2 is a free variable. Since we are solving this system over $\mathbb{C}$, this means that z_2 can take any *complex* value. Hence, we let $z_2 = \alpha \in \mathbb{C}$. Then, the general solution is $z_1 = 1 + i + \alpha i$, $z_2 = \alpha$, and $z_3 = i$.

Section 11.1 Problems

1. Calculate the following

(a) $(3+4i)-(-2+6i)$ (b) $(-1+2i)(3+2i)$ (c) $-3i(-2+3i)$

(d) $|(1+i)(1-2i)(3+4i)|$ (e) $|1+6i|$ (f) $\left|\frac{2}{3}-\sqrt{2}i\right|$

(g) $\frac{2}{1-i}$ (h) $\frac{4-3i}{3-4i}$ (i) $\frac{2+5i}{-3-6i}$

2. Solving the following systems of linear equations over $\mathbb{C}$.

(a)
$$\begin{aligned} z_1+(2+i)z_2+iz_3 \qquad\quad &= 1+i\\ iz_1+(-1+2i)z_2 \qquad\quad +2iz_4 &= -i\\ z_1+(2+i)z_2+(1+i)z_3+2iz_4 &= 2-i \end{aligned}$$

(b)
$$\begin{aligned} iz_1+2z_2-(3+i)z_3 &= 1\\ (1+i)z_1+(2-2i)z_2-4z_3 &= i\\ iz_1+2z_2-(3+3i)z_3 &= 1+2i \end{aligned}$$

(c)
$$\begin{aligned} iz_1+(1+i)z_2+z_3 &= 2i\\ (1-i)z_1+(1-2i)z_2+(-2+i)z_3 &= -2+i\\ 2iz_1+2iz_2+2z_3 &= 4+2i \end{aligned}$$

(d)
$$\begin{aligned} z_1-z_2+iz_3 &= 2i\\ (1+i)z_1-iz_2+iz_3 &= -2+i\\ (1-i)z_1+(-1+2i)z_2+(1+2i)z_3 &= 3+2i \end{aligned}$$

3. Prove that for any $z_1, z_2, z_3 \in \mathbb{C}$ that we have

$$z_1(z_2+z_3) = z_1z_2+z_1z_3$$

4. Prove that for any positive integer n and $z \in \mathbb{C}$ we have

$$\overline{z^n} = (\overline{z})^n$$

11.2 Complex Vector Spaces

We saw in the last section that the set of all complex numbers $\mathbb{C}$ can be thought of as a two dimensional real vector space. However, we see this is inappropriate for the solution space of Example 5.1.10 since it requires multiplication by complex scalars. Thus, we need to extend the definition of a real vector space to a complex vector space.

DEFINITION

Complex Vector Space

A set $\mathbb{V}$ is called a **vector space over** $\mathbb{C}$ if there is an operation of addition and an operation of scalar multiplication such that for any $\vec{v}, \vec{z}, \vec{w} \in \mathbb{V}$ and $\alpha, \beta \in \mathbb{C}$ we have:

V1 $\vec{z} + \vec{w} \in \mathbb{V}$
V2 $(\vec{z} + \vec{w}) + \vec{v} = \vec{z} + (\vec{w} + \vec{v})$
V3 $\vec{z} + \vec{w} = \vec{w} + \vec{z}$
V4 There is a vector denoted $\vec{0}$ in $\mathbb{V}$ such that $\vec{z} + \vec{0} = \vec{z}$. It is called the **zero vector**.
V5 For each $\vec{z} \in \mathbb{V}$ there exists an element $-\vec{z}$ such that $\vec{z} + (-\vec{z}) = \vec{0}$. $-\vec{z}$ is called the **additive inverse** of $\vec{z}$.
V6 $\alpha\vec{z} \in \mathbb{V}$
V7 $\alpha(\beta\vec{z}) = (\alpha\beta)\vec{z}$
V8 $(\alpha + \beta)\vec{z} = \alpha\vec{z} + \beta\vec{z}$
V9 $\alpha(\vec{z} + \vec{w}) = \alpha\vec{z} + \alpha\vec{w}$
V10 $1\vec{z} = \vec{z}$.

This definition looks a lot like the definition of a real vector space. In fact, the only difference is that we now allow the scalars to be any complex number, instead of just any real number. In the rest of this section we will generalize most of our concepts from real vector spaces to complex vector spaces.

EXAMPLE 1

The set $\mathbb{C}^n = \left\{ \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} \mid z_i \in \mathbb{C} \right\}$ is a vector space over $\mathbb{C}$ with addition defined by

$$\begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} + \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} z_1 + w_1 \\ \vdots \\ z_n + w_n \end{bmatrix}$$

and scalar multiplication defined by

$$\alpha \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} = \begin{bmatrix} \alpha z_1 \\ \vdots \\ \alpha z_n \end{bmatrix}$$

for any $\alpha \in \mathbb{C}$.

Just like a complex number $z \in \mathbb{C}$, we can split a vector in $\mathbb{C}^n$ into a real and imaginary part.

THEOREM 1

If $\vec{z} \in \mathbb{C}^n$, then there exists vectors $\vec{x}, \vec{y} \in \mathbb{R}^n$ such that

$$\vec{z} = \vec{x} + i\vec{y}$$

Other familiar real vector spaces can be turned into complex vector spaces.

EXAMPLE 2

The set $M_{m\times n}(\mathbb{C})$ of all $m \times n$ matrices with complex entries is a complex vector space with standard addition and complex scalar multiplication of matrices.

We now extend all of our theory from real vector spaces to complex vectors space.

DEFINITION
Subspace

If $\mathbb{S}$ is a subset of a complex vector space $\mathbb{V}$ and $\mathbb{S}$ is a complex vector space under the same operations as $\mathbb{V}$, then $\mathbb{S}$ is said to be a **subspace** of $\mathbb{V}$.

As in the real case, we use the Subspace Test to prove a non-empty subset $\mathbb{S}$ of a complex vector space $\mathbb{V}$ is a subspace of $\mathbb{V}$.

EXAMPLE 3

Prove that $\mathbb{S} = \left\{ \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \mid z_1 + iz_2 = -z_3 \right\}$ is a subspace of $\mathbb{C}^3$.

Solution: By definition $\mathbb{S}$ is a subset of $\mathbb{C}^3$. Moreover, $\mathbb{S}$ is non-empty since $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ satisfies $0 + i(0) = -0$. Let $\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}, \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} \in \mathbb{S}$. Then $z_1 + iz_2 = -z_3$ and $w_1 + iw_2 = -w_3$. Hence,

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} z_1 + w_1 \\ z_2 + w_2 \\ z_3 + w_3 \end{bmatrix} \in \mathbb{S}$$

since $(z_1 + w_1) + i(z_2 + w_2) = z_1 + iz_2 + w_1 + iw_2 = -z_3 - w_3 = -(z_3 + w_3)$, and

$$\alpha \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} \alpha z_1 \\ \alpha z_2 \\ \alpha z_3 \end{bmatrix} \in \mathbb{S}$$

since $\alpha z_1 + i(\alpha z_2) = \alpha(z_1 + iz_2) = \alpha(-z_3) = -(\alpha z_3)$. Thus, $\mathbb{S}$ is a subspace of $\mathbb{C}^3$ by the subspace test.

EXAMPLE 4 Is $\mathbb{R}$ a subspace of $\mathbb{C}$ as a vector space over $\mathbb{C}$?

Solution: By definition $\mathbb{R}$ is a non-empty subset of $\mathbb{C}$. Let $x, y \in \mathbb{R}$. Then, $x + y$ is also a real number, so the set is closed under addition. But, the difference between a real vector space and a complex vector space is that we have scalar multiplication by complex numbers in a complex vector space. So, if we take $2 \in \mathbb{R}$ and the scalar $i \in \mathbb{C}$, then we get $i(2) \notin \mathbb{R}$. Thus, $\mathbb{R}$ is not closed under complex scalar multiplication, and hence it is not a subspace of $\mathbb{C}$.

The concepts of linear independence, spanning, bases, and dimension are defined the same as in the real case except that we now have complex scalar multiplication.

EXAMPLE 5 Find a basis for $\mathbb{C}$ as a vector space over $\mathbb{C}$.

Solution: It may be tempting to think that a basis for $\mathbb{C}$ is $\{1, i\}$, but this would be incorrect since this set is linearly dependent in $\mathbb{C}$. In particular, the equation $\alpha_1 1 + \alpha_2 i = 0$ has solution $\alpha_1 = 1$ and $\alpha_2 = i$ since $1(1) + i(i) = 1 - 1 = 0$. A basis for $\mathbb{C}$ is $\{1\}$ since any complex number $a + bi$ is just equal to $(a + bi)(1)$.

EXERCISE 1 What is the standard basis for $\mathbb{C}^n$?

EXAMPLE 6 Determine the dimension of the subspace $\mathbb{S} = \left\{ \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \mid z_1 + iz_2 = -z_3 \right\}$ of $\mathbb{C}^3$.

Solution: Every vector in $\mathbb{S}$ can be written in the form

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ -z_1 - iz_2 \end{bmatrix} = z_1 \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + z_2 \begin{bmatrix} 0 \\ 1 \\ -i \end{bmatrix}$$

Thus, $\left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -i \end{bmatrix} \right\}$ spans $\mathbb{S}$ and is clearly linearly independent since neither vector is a scalar multiple of the other, so it is a basis for $\mathbb{S}$. Thus, $\dim \mathbb{S} = 2$.

EXERCISE 2 Find a basis for the four fundamental subspaces of $A = \begin{bmatrix} 1 & -1 & -i \\ i & 1 & -1 + 2i \\ -i & 1 + 2i & -3 + 2i \\ 2 & 0 & 2i \end{bmatrix}$.

REMARK

Be warned that it is possible for two complex vectors to be scalar multiples of each other without looking like they are. For example, can you determine by inspection which of the following vectors is a scalar multiple of $\begin{bmatrix} 1-i \\ -2i \end{bmatrix}$?

$$\begin{bmatrix} 1 \\ -1-i \end{bmatrix}, \quad \begin{bmatrix} -i \\ -1-i \end{bmatrix}, \quad \begin{bmatrix} 7-i \\ 8+6i \end{bmatrix}$$

Coordinates with respect to a basis, linear mappings, and matrices of linear mappings are also defined in exactly the same way.

EXAMPLE 7 Given that $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 1+i \\ 1-i \end{bmatrix}, \begin{bmatrix} -1 \\ -i \\ -1+2i \end{bmatrix}, \begin{bmatrix} i \\ i \\ 2+2i \end{bmatrix} \right\}$ is a basis of $\mathbb{C}^3$, find the $\mathcal{B}$-coordinates of $\vec{z} = \begin{bmatrix} 2i \\ -2+i \\ 3+2i \end{bmatrix}$.

Solution: We need to find complex numbers α_1, α_2, and α_3 such that

$$\begin{bmatrix} 2i \\ -2+i \\ 3+2i \end{bmatrix} = \alpha_1 \begin{bmatrix} 1 \\ 1+i \\ 1-i \end{bmatrix} + \alpha_2 \begin{bmatrix} -1 \\ -i \\ -1+2i \end{bmatrix} + \alpha_3 \begin{bmatrix} i \\ i \\ 2+2i \end{bmatrix}$$

Row reducing the corresponding augmented matrix gives

$$\left[\begin{array}{ccc|c} 1 & -1 & i & 2i \\ 1+i & -i & i & -2+i \\ 1-i & -1+2i & 2+2i & 3+2i \end{array}\right] \begin{array}{l} \\ R_2-(1+i)R_1 \\ R_3-(1-i)R_1 \end{array} \sim$$

$$\left[\begin{array}{ccc|c} 1 & -1 & i & 2i \\ 0 & 1 & 1 & -i \\ 0 & i & 1+i & 1 \end{array}\right] \begin{array}{l} R_1+R_2 \\ \\ R_3-iR_2 \end{array} \sim$$

$$\left[\begin{array}{ccc|c} 1 & 0 & 1+i & i \\ 0 & 1 & 1 & -i \\ 0 & 0 & 1 & 0 \end{array}\right] \begin{array}{l} R_1-(1+i)R_3 \\ R_2-R_3 \\ \\ \end{array} \sim \left[\begin{array}{ccc|c} 1 & 0 & 0 & i \\ 0 & 1 & 0 & -i \\ 0 & 0 & 1 & 0 \end{array}\right]$$

Thus, $[\vec{z}]_{\mathcal{B}} = \begin{bmatrix} i \\ -i \\ 0 \end{bmatrix}$.

EXAMPLE 8 Find the standard matrix of $L : \mathbb{C}^2 \to \mathbb{C}^3$ given by $L(z_1, z_2) = \begin{bmatrix} iz_1 \\ z_2 \\ z_1 + z_2 \end{bmatrix}$.

Solution: The standard basis for $\mathbb{C}^2$ is $\left\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right\}$. We have

$L(1,0) = \begin{bmatrix} i \\ 0 \\ 1 \end{bmatrix}$ and $L(0,1) = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$. Hence, $[L] = \begin{bmatrix} L(1,0) & L(0,1) \end{bmatrix} = \begin{bmatrix} i & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$.

Of course, the concepts of determinants and invertibility are also the same.

EXAMPLE 9 We have

$$\det \begin{bmatrix} 1+i & 2 & 3 \\ 0 & -2i & -2+3i \\ 0 & 0 & 1-i \end{bmatrix} = (1+i)(-2i)(1-i) = -4i$$

$$\det \begin{bmatrix} i & i & 2i \\ 1 & 1 & 2 \\ 1+i & 1-i & 3i \end{bmatrix} = 0$$

EXAMPLE 10 Let $A = \begin{bmatrix} 1+i & 1 \\ 1 & 2-i \end{bmatrix}$ and $B = \begin{bmatrix} i & 0 & 1 \\ 1 & i & 1-i \\ 0 & 1 & 1+i \end{bmatrix}$. Show that A and B are invertible and find their inverse.

Solution: We have $\det A = \begin{vmatrix} 1+i & 1 \\ 1 & 2-i \end{vmatrix} = 2+i \neq 0$, so A is invertible. Using the formula for the inverse of a 2×2 matrix we found in Math 136 gives

$$A^{-1} = \frac{1}{2+i}\begin{bmatrix} 2-i & -1 \\ -1 & 1+i \end{bmatrix}$$

Row reducing $[B \mid I\,]$ to RREF gives

$$\left[\begin{array}{ccc|ccc} i & 0 & 1 & 1 & 0 & 0 \\ 1 & i & 1-i & 0 & 1 & 0 \\ 0 & 1 & 1+i & 0 & 0 & 1 \end{array}\right] \sim \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 2-2i & -1 & i \\ 0 & 1 & 0 & 3-i & -1-i & i \\ 0 & 0 & 1 & -1-2i & i & 1 \end{array}\right]$$

Since the RREF of B is I, B is invertible. Moreover, $B^{-1} = \begin{bmatrix} 2-2i & -1 & i \\ 3-i & -1-i & i \\ -1-2i & i & 1 \end{bmatrix}$.

EXERCISE 3 Let $A = \begin{bmatrix} 1+i & 2 & i \\ 2 & 3-2i & 1+i \\ 1-i & 1-3i & 1+i \end{bmatrix}$. Find the determinant of A and A^{-1}.

Complex Conjugates

We saw in Section 11.1 that the complex conjugate was a very important and useful operation on complex numbers. Thus, we should expect that it will also be useful and important in the study of general complex vector spaces. Therefore, we extend the definition of the complex conjugate to vectors in $\mathbb{C}^n$ and matrices in $M_{m\times n}(\mathbb{C})$.

DEFINITION
Complex Conjugate in $\mathbb{C}^n$

Let $\vec{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} \in \mathbb{C}^n$. Then, we define $\overline{\vec{z}}$ by $\overline{\vec{z}} = \begin{bmatrix} \overline{z_1} \\ \vdots \\ \overline{z_n} \end{bmatrix}$.

DEFINITION
Complex Conjugate in $M_{m\times n}(\mathbb{C})$

Let $A = \begin{bmatrix} z_{11} & \cdots & z_{1n} \\ \vdots & & \vdots \\ z_{m1} & \cdots & z_{mn} \end{bmatrix} \in M_{m\times n}(\mathbb{C})$. Then, we define $\overline{A}$ by $\overline{A} = \begin{bmatrix} \overline{z_{11}} & \cdots & \overline{z_{1n}} \\ \vdots & & \vdots \\ \overline{z_{m1}} & \cdots & \overline{z_{mn}} \end{bmatrix}$.

THEOREM 2 Let $A \in M_{m\times n}(\mathbb{C})$ and $\vec{z} \in \mathbb{C}^n$. Then $\overline{A\vec{z}} = \overline{A}\,\overline{\vec{z}}$.

EXAMPLE 11 Let $\vec{w} = \begin{bmatrix} 1 \\ 2i \\ 1-i \end{bmatrix}$, $\vec{z} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$, and $A = \begin{bmatrix} 1 & 1-3i \\ -i & 2i \end{bmatrix}$. Then

$$\overline{\vec{w}} = \begin{bmatrix} 1 \\ -2i \\ 1+i \end{bmatrix} \qquad \overline{\vec{z}} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \qquad \overline{A} = \begin{bmatrix} 1 & 1+3i \\ i & -2i \end{bmatrix}$$

EXAMPLE 12 Is the mapping $L : \mathbb{C}^2 \to \mathbb{C}^2$ defined by $L(\vec{z}) = \overline{\vec{z}}$ a linear mapping?

Solution: Consider $L(\alpha\vec{z}_1 + \vec{z}_2) = \overline{\alpha\vec{z}_1 + \vec{z}_2} = \overline{\alpha}\overline{\vec{z}_1} + \overline{\vec{z}_2}$, but $\overline{\alpha} \neq \alpha$ if α is not real. So, it is not linear. For example

$$L\left((1+i)\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = L\left(\begin{bmatrix} 1+i \\ 0 \end{bmatrix}\right) = \overline{\begin{bmatrix} 1+i \\ 0 \end{bmatrix}} = \begin{bmatrix} 1-i \\ 0 \end{bmatrix} \neq \begin{bmatrix} 1+i \\ 0 \end{bmatrix} = (1+i)L\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right)$$

Even though the complex conjugate of a vector in $\mathbb{C}^n$ does not define a linear mapping, we will see that complex conjugates of vectors in $\mathbb{C}^n$ and matrices in $M_{m\times n}(\mathbb{C})$ occur naturally in the study of complex vector spaces.

Section 11.2 Problems

1. Determine which of the following sets is a subspace of $\mathbb{C}^3$. Find a basis and the dimension of each subspace.

   (a) $\mathbb{S}_1 = \left\{ \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \in \mathbb{C}^3 \mid iz_1 = z_3 \right\}$

   (b) $\mathbb{S}_2 = \left\{ \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \in \mathbb{C}^3 \mid z_1 z_2 = 0 \right\}$

   (c) $\mathbb{S}_3 = \left\{ \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \in \mathbb{C}^3 \mid z_1 + z_2 + z_3 = 0 \right\}$

2. Find a basis for the four fundamental subspaces of $A = \begin{bmatrix} 1 & i \\ 1+i & -1+i \\ -1 & i \end{bmatrix}$.

3. Let $\vec{z} \in \mathbb{C}^n$. Prove that there exists vectors $\vec{x}, \vec{y} \in \mathbb{R}^n$ such that $\vec{z} = \vec{x} + i\vec{y}$.

4. Let $L : \mathbb{C}^3 \to \mathbb{C}^2$ be defined by $L(z_1, z_2, z_3) = \begin{bmatrix} -iz_1 + (1+i)z_2 + (1+2i)z_3 \\ (-1+i)z_1 - 2iz_2 - 3iz_3 \end{bmatrix}$.

   (a) Prove that L is linear and find the standard matrix of L.

   (b) Determine a basis for the range and kernel of L.

5. Let $L : \mathbb{C}^2 \to \mathbb{C}^2$ be defined by $L(z_1, z_2) = \begin{bmatrix} z_1 + (1+i)z_2 \\ (1+i)z_1 + 2iz_2 \end{bmatrix}$.

   (a) Prove that L is linear.

   (b) Determine a basis for the range and kernel of L.

   (c) Use the result of part (b) to define a basis $\mathcal{B}$ for $\mathbb{C}^2$ such that $[L]_{\mathcal{B}}$ is diagonal.

6. Calculate the determinant of $\begin{bmatrix} 1 & -1 & i \\ 1+i & -i & i \\ 1-i & -1+2i & 1+2i \end{bmatrix}$.

7. Find the inverse of $\begin{bmatrix} 1 & 1 & 2 \\ 0 & 1+i & 1 \\ i & 1+i & 1+2i \end{bmatrix}$.

8. Prove that for any $n \times n$ matrix A, we have $\det \overline{A} = \overline{\det A}$.

11.3 Complex Diagonalization

In Math 136 we saw that some matrices were not diagonalizable over $\mathbb{R}$ because they had complex eigenvalues. But, if we extend all of our concepts of eigenvalues, eigenvectors, and diagonalization to complex vector spaces, then we will be able to diagonalize these matrices.

DEFINITION
Eigenvalue
Eigenvector

Let $A \in M_{n\times n}(\mathbb{C})$. If there exists $\lambda \in \mathbb{C}$ and $\vec{z} \in \mathbb{C}^n$ with $\vec{z} \neq \vec{0}$ such that $A\vec{z} = \lambda\vec{z}$, then λ is called an **eigenvalue** of A and $\vec{z}$ is called an **eigenvector** of A corresponding to λ. We call $(\lambda, \vec{z})$ an **eigenpair**.

All of the theory of similar matrices, eigenvalues, eigenvectors, and diagonalization we did in Math 136 still applies in the complex case, except that we now allow the use of complex eigenvalues and eigenvectors.

EXAMPLE 1

Diagonalize $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ over $\mathbb{C}$.

Solution: The characteristic polynomial is

$$C(\lambda) = \begin{vmatrix} -\lambda & 1 \\ -1 & -\lambda \end{vmatrix} = \lambda^2 + 1$$

Therefore, the eigenvalues of A are $\lambda_1 = i$ and $\lambda_2 = -i$. For $\lambda_1 = i$ we get

$$A - iI = \begin{bmatrix} -i & 1 \\ -1 & -i \end{bmatrix} \sim \begin{bmatrix} 1 & i \\ 0 & 0 \end{bmatrix}$$

Hence, a basis for E_{λ_1} is $\left\{\begin{bmatrix} -i \\ 1 \end{bmatrix}\right\}$. For $\lambda_2 = -i$ we get

$$A + iI = \begin{bmatrix} i & 1 \\ -1 & i \end{bmatrix} \sim \begin{bmatrix} 1 & -i \\ 0 & 0 \end{bmatrix}$$

So, a basis for E_{λ_2} is $\left\{\begin{bmatrix} i \\ 1 \end{bmatrix}\right\}$. Thus, A is diagonalized by $P = \begin{bmatrix} -i & i \\ 1 & 1 \end{bmatrix}$ to $D = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}$.

EXAMPLE 2

Diagonalize $B = \begin{bmatrix} i & 1+i \\ 1+i & 1 \end{bmatrix}$ over $\mathbb{C}$.

Solution: The characteristic polynomial is

$$C(\lambda) = \begin{vmatrix} i-\lambda & 1+i \\ 1+i & 1-\lambda \end{vmatrix} = \lambda^2 - (1+i)\lambda - i$$

Using the quadratic formula we get

$$\lambda = \frac{(1+i) \pm \sqrt{6i}}{2} = \frac{(1+i) \pm \sqrt{6}(\frac{1+i}{\sqrt{2}})}{2} = \frac{1 \pm \sqrt{3}}{2}(1+i)$$

So, the eigenvalues are $\lambda_1 = \frac{1+\sqrt{3}}{2}(1+i)$ and $\lambda_2 = \frac{1-\sqrt{3}}{2}(1+i)$. For λ_1 we get

$$B - \lambda_1 I = \begin{bmatrix} i - \frac{1+\sqrt{3}}{2}(1+i) & 1+i \\ 1+i & 1 - \frac{1+\sqrt{3}}{2}(1+i) \end{bmatrix} \sim \begin{bmatrix} -\sqrt{3}+i & 2 \\ 2 & -\sqrt{3}-i \end{bmatrix}$$
$$\sim \begin{bmatrix} 1 & -2/(\sqrt{3}-i) \\ 0 & 0 \end{bmatrix}$$

Then a basis for E_{λ_1} is $\vec{v}_1 = \begin{bmatrix} 2 \\ \sqrt{3}-i \end{bmatrix}$. For λ_2 we get

$$B - \lambda_2 I = \begin{bmatrix} i - \frac{1-\sqrt{3}}{2}(1+i) & 1+i \\ 1+i & 1 - \frac{1-\sqrt{3}}{2}(1+i) \end{bmatrix} \sim \begin{bmatrix} \sqrt{3}+i & 2 \\ 2 & \sqrt{3}-i \end{bmatrix}$$
$$\sim \begin{bmatrix} 1 & 2/(\sqrt{3}+i) \\ 0 & 0 \end{bmatrix}$$

Hence, a basis for E_{λ_2} is $\vec{v}_2 = \begin{bmatrix} -2 \\ \sqrt{3}+i \end{bmatrix}$. Thus, B is diagonalized by $P = \begin{bmatrix} 2 & -2 \\ \sqrt{3}-i & \sqrt{3}+i \end{bmatrix}$ to $D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$.

EXAMPLE 3

Diagonalize $F = \begin{bmatrix} 3 & 2+i \\ 2-i & 7 \end{bmatrix}$ over $\mathbb{C}$.

Solution: The characteristic polynomial is

$$C(\lambda) = \begin{vmatrix} 3-\lambda & 2+i \\ 2-i & 7-\lambda \end{vmatrix} = \lambda^2 - 10\lambda + 16 = (\lambda-2)(\lambda-8)$$

So the eigenvalues are $\lambda_1 = 2$ and $\lambda_2 = 8$. For λ_1 we get

$$F - \lambda_1 I = \begin{bmatrix} 1 & 2+i \\ 2-i & 5 \end{bmatrix} \sim \begin{bmatrix} 1 & 2+i \\ 0 & 0 \end{bmatrix}$$

Then a basis for E_{λ_1} is $\vec{v}_1 = \begin{bmatrix} -2-i \\ 1 \end{bmatrix}$. For λ_2 we get

$$F - \lambda_2 I = \begin{bmatrix} -5 & 2+i \\ 2-i & -1 \end{bmatrix} \sim \begin{bmatrix} 1 & -1/(2-i) \\ 0 & 0 \end{bmatrix}$$

Hence, a basis for E_{λ_2} is $\vec{v}_2 = \begin{bmatrix} -1 \\ -2+i \end{bmatrix}$. Thus, F is diagonalized by $P = \begin{bmatrix} -2-i & -1 \\ 1 & -2+i \end{bmatrix}$ to $D = \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix}$.

As usual, these examples teach us more than just how to diagonalize complex matrices. In the first example we see that when the matrix has only real entries, then the eigenvalues come in complex conjugate pairs. The second example shows that when working with matrices with non-real entries our theory of symmetric matrices for real matrices does not apply. In particular, we had $B^T = B$, but the eigenvalues of B are not real. On the other hand, observe that the matrix F has non-real entires but the eigenvalues of F are all real.

THEOREM 1 Let $A \in M_{n\times n}(\mathbb{R})$. If λ is a non-real eigenvalue of A with corresponding eigenvector $\vec{z}$, then $\overline{\lambda}$ is also an eigenvalue of A with eigenvector $\overline{\vec{z}}$.

Proof: We have $A\vec{z} = \lambda\vec{z}$. Hence,

$$\overline{\lambda}\,\overline{\vec{z}} = \overline{\lambda\vec{z}} = \overline{A\vec{z}} = \overline{A}\,\overline{\vec{z}} = A\overline{\vec{z}}$$

since A is a real matrix. Thus, $\overline{\lambda}$ is an eigenvalue of A with eigenvector $\overline{\vec{z}}$. □

COROLLARY 2 If $A \in M_{n\times n}(\mathbb{R})$ with n odd, then A has at least one real eigenvalue.

EXAMPLE 4 Find all eigenvalues of $A = \begin{bmatrix} 0 & 2 & 1 \\ -2 & 3 & 0 \\ 1 & 0 & 2 \end{bmatrix}$ given that $\lambda_1 = 2 + i$ is an eigenvalue of A.

Solution: Since A is a real matrix and $\lambda_1 = 2 + i$ is an eigenvalue of A we have by Theorem 1 that $\lambda_2 = \overline{\lambda_1} = 2 - i$ is also an eigenvalue of A. Finally, we know that the sum of the eigenvalues is the trace of the matrix, so the remaining eigenvalue, which must be real, is

$$\lambda_3 = \operatorname{tr} A - \lambda_1 - \lambda_2 = 5 - (2 + i) - (2 - i) = 1$$

It is easy to verify this result by checking that $\det(A - I) = 0$.

Section 11.3 Problems

1. For each of the following matrices, either diagonalize the matrix over $\mathbb{C}$, or show that it is not diagonalizable.

(a) $\begin{bmatrix} 2 & 1+i \\ 1-i & 1 \end{bmatrix}$ (b) $\begin{bmatrix} 3 & 5 \\ -5 & -3 \end{bmatrix}$ (c) $\begin{bmatrix} 2 & 2 & -1 \\ -4 & 1 & 2 \\ 2 & 2 & -1 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & i \\ i & -1 \end{bmatrix}$ (e) $\begin{bmatrix} 2 & 1 & -1 \\ 2 & 1 & 0 \\ 3 & -1 & 2 \end{bmatrix}$ (f) $\begin{bmatrix} 1+i & 1 & 0 \\ 1 & 1 & -i \\ 1 & 0 & 1 \end{bmatrix}$

2. For any $\theta \in \mathbb{R}$, let $R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$.
   (a) Diagonalize R_θ over $\mathbb{C}$.
   (b) Verify your answer in (a) is correct, by showing the matrix P and diagonal matrix D from part (a) satisfy $P^{-1}R_\theta P = D$ for $\theta = 0$ and $\theta = \frac{\pi}{4}$.

3. Let $A \in M_{n\times n}(\mathbb{C})$ and let $\vec{z}$ be an eigenvector of A. Prove that $\overline{\vec{z}}$ is an eigenvector of $\overline{A}$. What is the corresponding eigenvalue?

4. Suppose that a real 2×2 matrix A has $2+i$ as an eigenvalue with a corresponding eigenvector $\begin{bmatrix} 1+i \\ i \end{bmatrix}$. Determine A.

11.4 Complex Inner Products

The concepts of orthogonality and orthonormal bases were very useful in real vector spaces. Hence, it makes sense to extend these concepts to complex vector spaces. In particular, we would like to be able to define the concepts of orthogonality and length in complex vector spaces to match those in the real case.

REMARK

In this section we leave the proofs of the theorems to the reader (or homework assignments/tests) as they are essentially the same as in the real case.

We start by looking at how to define an inner product for $\mathbb{C}^n$.

We first observe that if we extend the standard dot product to vectors in $\mathbb{C}^n$, then this does not define an inner product. Indeed, if $\vec{z} = \vec{x} + i\vec{y}$, then we have

$$\begin{aligned} \vec{z}\cdot\vec{z} &= z_1^2 + \cdots + z_n^2 \\ &= (x_1^2 + \cdots + x_n^2 - y_1^2 - \cdots - y_n^2) + 2i(x_1y_1 + \cdots + x_ny_n) \end{aligned}$$

But, to be an inner product we require that $\langle \vec{z}, \vec{z}\rangle$ be a non-negative real number so that we can define the length of a vector. Since $\vec{z}\cdot\vec{z}$ does not even have to be real, it cannot define an inner product.

Thinking of properties of complex numbers, the only way we can ensure that we get a non-negative real number when multiplying complex numbers is to multiply the complex number by its conjugate. In particular, if we define

$$\langle \vec{z}, \vec{w}\rangle = z_1\overline{w_1} + \cdots + z_n\overline{w_n} = \vec{z}\cdot\overline{\vec{w}}$$

then, we get

$$< \vec{z}, \vec{z} >= z_1\overline{z_1} + \cdots + z_n\overline{z_n} = |z_1|^2 + \cdots + |z_n|^2$$

which is not only a non-negative real number, but it is 0 if and only if $\vec{z} = \vec{0}$. Thus, this is what we are looking for.

DEFINITION

Standard Inner product on $\mathbb{C}^n$

The **standard inner product** $<,>$ on $\mathbb{C}^n$ is defined by

$$\langle \vec{z}, \vec{w} \rangle = \vec{z} \cdot \overline{\vec{w}} = z_1\overline{w_1} + \cdots + z_n\overline{w_n}$$

for any $\vec{z}, \vec{w} \in \mathbb{C}^n$.

REMARK

In some other fields, for example engineering, they use

$$\langle \vec{z}, \vec{w} \rangle = \overline{\vec{w}} \cdot \vec{z}$$

for the definition of the standard inner product for $\mathbb{C}^n$. Be warned that many computer programs use the engineering definition of the inner product for $\mathbb{C}^n$.

EXAMPLE 1

Let $\vec{z} = \begin{bmatrix} 1 \\ 2+i \\ 1-i \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} i \\ 2 \\ 1+3i \end{bmatrix}$. Then

$$< \vec{z}, \vec{z} > = \vec{z} \cdot \overline{\vec{z}} = \begin{bmatrix} 1 \\ 2+i \\ 1-i \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2-i \\ 1+i \end{bmatrix} = 1^2 + (2+i)(2-i) + (1-i)(1+i) = 8$$

$$< \vec{z}, \vec{w} > = \vec{z} \cdot \overline{\vec{w}} = \begin{bmatrix} 1 \\ 2+i \\ 1-i \end{bmatrix} \cdot \begin{bmatrix} -i \\ 2 \\ 1-3i \end{bmatrix} = 1(-i) + (2+i)(2) + (1-i)(1-3i) = 2-3i$$

$$< \vec{w}, \vec{z} > = \vec{w} \cdot \overline{\vec{z}} = \begin{bmatrix} i \\ 2 \\ 1+3i \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2-i \\ 1+i \end{bmatrix} = i(1) + 2(2-i) + (1+3i)(1+i) = 2+3i$$

This example shows that the standard inner product for $\mathbb{C}^n$ is not symmetric. The next example shows that it is also not bilinear.

EXAMPLE 2

Let $\vec{z}, \vec{w} \in \mathbb{C}^n$. Then, for any $\alpha \in \mathbb{C}$ we have $\langle \alpha\vec{z}, \vec{w} \rangle = \alpha \langle \vec{z}, \vec{w} \rangle$, but $\langle \vec{z}, \alpha\vec{w} \rangle = \overline{\alpha} \langle \vec{z}, \vec{w} \rangle$.

Solution: We have

$$\begin{aligned} \langle \alpha\vec{z}, \vec{w} \rangle &= (\alpha\vec{z}) \cdot \overline{\vec{w}} = \alpha(\vec{z} \cdot \overline{\vec{w}}) = \alpha \langle \vec{z}, \vec{w} \rangle \\ \langle \vec{z}, \alpha\vec{w} \rangle &= \vec{z} \cdot \overline{\alpha\vec{w}} = \vec{z} \cdot (\overline{\alpha}\,\overline{\vec{w}}) = \overline{\alpha}(\vec{z} \cdot \overline{\vec{w}}) = \overline{\alpha} \langle \vec{z}, \vec{w} \rangle \end{aligned}$$

EXERCISE 1

Let $\vec{z} = \begin{bmatrix} 1+i \\ 1-i \\ 2 \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} 1-i \\ 2i \\ -2 \end{bmatrix}$. Find $\langle \vec{z}, \vec{w} \rangle$ and $\langle \vec{w}, \vec{z} \rangle$.

THEOREM 1

Let $\langle \, , \rangle$ denote the standard inner product for $\mathbb{C}^n$. Then

(1) $\langle \vec{z}, \vec{z} \rangle \in \mathbb{R}$ and $\langle \vec{z}, \vec{z} \rangle \geq 0$ for all $\vec{z} \in \mathbb{V}$, and $\langle \vec{z}, \vec{z} \rangle = 0$ if and only if $\vec{z} = \vec{0}$
(2) $\langle \vec{z}, \vec{w} \rangle = \overline{\langle \vec{w}, \vec{z} \rangle}$
(3) $\langle \vec{v} + \vec{z}, \vec{w} \rangle = \langle \vec{v}, \vec{w} \rangle + \langle \vec{z}, \vec{w} \rangle$
(4) $\langle \alpha\vec{z}, \vec{w} \rangle = \alpha \langle \vec{z}, \vec{w} \rangle$

As we saw in the examples, this theorem shows that the standard inner product on $\mathbb{C}^n$ does not satisfy the same properties as a real inner product. So, to define an inner product on a general complex vector space, we should use these properties instead of those for a real inner product.

DEFINITION

Hermitian Inner Product

Let $\mathbb{V}$ be a complex vector space. A **Hermitian inner product** on $\mathbb{V}$ is a function $\langle \, , \rangle : \mathbb{V} \times \mathbb{V} \to \mathbb{C}$ such that for all $\vec{v}, \vec{w}, \vec{z} \in \mathbb{V}$ and $\alpha \in \mathbb{C}$ we have

(1) $\langle \vec{z}, \vec{z} \rangle \in \mathbb{R}$ and $\langle \vec{z}, \vec{z} \rangle \geq 0$ for all $\vec{z} \in \mathbb{V}$, and $\langle \vec{z}, \vec{z} \rangle = 0$ if and only if $\vec{z} = \vec{0}$
(2) $\langle \vec{z}, \vec{w} \rangle = \overline{\langle \vec{w}, \vec{z} \rangle}$
(3) $\langle \vec{v} + \vec{z}, \vec{w} \rangle = \langle \vec{v}, \vec{w} \rangle + \langle \vec{z}, \vec{w} \rangle$
(4) $\langle \alpha\vec{z}, \vec{w} \rangle = \alpha \langle \vec{z}, \vec{w} \rangle$

A complex vector space with a Hermitian inner product is called a **Hermitian inner product space**.

THEOREM 2

If $\langle \, , \rangle$ is a Hermitian inner product on a complex vector space $\mathbb{V}$, then for all $\vec{v}, \vec{w}, \vec{z} \in \mathbb{V}$ and $\alpha \in \mathbb{C}$ we have

$$\begin{aligned} \langle \vec{z}, \vec{v} + \vec{w} \rangle &= \langle \vec{z}, \vec{v} \rangle + \langle \vec{z}, \vec{w} \rangle \\ \langle \vec{z}, \alpha\vec{w} \rangle &= \overline{\alpha} \langle \vec{z}, \vec{w} \rangle \end{aligned}$$

It is important to observe that this is a true generalization of the real inner product. That is, we could use this for the definition of the real inner product since if $\langle \vec{z}, \vec{w} \rangle$ and α are strictly real, then this will satisfy the definition of a real inner product.

THEOREM 3

The function $\langle A, B \rangle = \operatorname{tr}(\overline{B}^T A)$ defines a Hermitian inner product on $M_{m\times n}(\mathbb{C})$.

EXAMPLE 3

Let $A = \begin{bmatrix} 1 & i \\ 2 & 2+i \end{bmatrix}$ and $B = \begin{bmatrix} -3 & 3+i \\ -2i & 0 \end{bmatrix}$. Find $\langle A, B\rangle$ under the inner product

$$\langle A, B\rangle = \operatorname{tr}(\overline{B^T} A)$$

Solution: We have

$$\begin{aligned}\langle A, B\rangle = \operatorname{tr}(\overline{B^T} A) &= \operatorname{tr}\left(\begin{bmatrix} -3 & 2i \\ 3-i & 0 \end{bmatrix}\begin{bmatrix} 1 & i \\ 2 & 2+i \end{bmatrix}\right) \\ &= \operatorname{tr}\begin{bmatrix} -3+4i & -2+i \\ 3-i & 1+3i \end{bmatrix} = -3+4i+1+3i = -2+7i\end{aligned}$$

EXAMPLE 4

If $A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$ and $B = \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{m1} & \cdots & b_{mn} \end{bmatrix}$, then

$$\overline{B}^T A = \begin{bmatrix} \overline{b_{11}} & \cdots & \overline{b_{m1}} \\ \vdots & & \vdots \\ \overline{b_{1n}} & \cdots & \overline{b_{mn}} \end{bmatrix}\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

Hence, we get

$$\operatorname{tr}(\overline{B}^T A) = \sum_{\ell=1}^{m} a_{\ell 1}\overline{b_{\ell 1}} + \sum_{\ell=1}^{m} a_{\ell 2}\overline{b_{\ell 2}} + \cdots + \sum_{\ell=1}^{m} a_{\ell n}\overline{b_{\ell n}} = \sum_{j=1}^{n}\sum_{\ell=1}^{m} a_{\ell j}\overline{b_{\ell j}}$$

which corresponds to the standard inner product of the corresponding vectors under the obvious isomorphism with $\mathbb{C}^{mn}$.

REMARK

As in the real case, whenever we are working with $\mathbb{C}^n$ or $M_{m\times n}(\mathbb{C})$, if no other inner product is specified, we will assume we are working with the standard inner product.

EXERCISE 2

Let $A, B \in M_{2\times 2}(\mathbb{C})$ with $A = \begin{bmatrix} 1 & 1-i \\ 2+2i & i \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 1-i \\ 2+2i & i \end{bmatrix}$. Determine $\langle A, B\rangle$ and $\langle B, A\rangle$.

Length and Orthogonality

We can now define length and orthogonality to match the definitions in the real case.

DEFINITION
Length
Unit Vector

Let $\mathbb{V}$ be a Hermitian inner product space with inner product $\langle\, ,\rangle$. Then, for any $\vec{z} \in \mathbb{V}$ we define the **length** of $\vec{z}$ by

$$\|\vec{z}\| = \sqrt{\langle \vec{z}, \vec{z}\rangle}$$

If $\|\vec{z}\| = 1$, then $\vec{z}$ is called a **unit vector**.

DEFINITION
Orthogonality

Let $\mathbb{V}$ be a Hermitian inner product space with inner product $\langle\, ,\rangle$. Then, for any $\vec{z}, \vec{w} \in \mathbb{V}$ we say that $\vec{z}$ and $\vec{w}$ are **orthogonal** if $\langle \vec{z}, \vec{w}\rangle = 0$.

EXAMPLE 5

Let $\vec{u} = \begin{bmatrix} i \\ 1+i \\ 2-i \end{bmatrix} \in \mathbb{C}^3$.

$$\begin{aligned} \|\vec{u}\| &= \sqrt{\langle \vec{u}, \vec{u}\rangle} = \sqrt{\vec{u} \cdot \overline{\vec{u}}} \\ &= \sqrt{i(-i) + (1+i)(1-i) + (2-i)(2+i)} = \sqrt{1+2+5} = \sqrt{8} \end{aligned}$$

EXAMPLE 6

Let $\vec{z} = \begin{bmatrix} 1 \\ i \\ 2+3i \end{bmatrix} \in \mathbb{C}^3$. Which of the following vectors are orthogonal to $\vec{z}$:

$$\vec{u} = \begin{bmatrix} -i \\ 1 \\ 0 \end{bmatrix}, \qquad \vec{v} = \begin{bmatrix} i \\ 1 \\ 0 \end{bmatrix}, \qquad \vec{w} = \begin{bmatrix} 1-2i \\ -7 \\ 1-i \end{bmatrix}$$

Solution: We have

$$\langle \vec{z}, \vec{u}\rangle = \vec{z} \cdot \overline{\vec{u}} = \begin{bmatrix} 1 \\ i \\ 2+3i \end{bmatrix} \cdot \begin{bmatrix} i \\ 1 \\ 0 \end{bmatrix} = i + i + 0 = 2i$$

$$\langle \vec{z}, \vec{v}\rangle = \vec{z} \cdot \overline{\vec{v}} = \begin{bmatrix} 1 \\ i \\ 2+3i \end{bmatrix} \cdot \begin{bmatrix} -i \\ 1 \\ 0 \end{bmatrix} = i - i + 0 = 0$$

$$\langle \vec{z}, \vec{w}\rangle = \vec{z} \cdot \overline{\vec{w}} = \begin{bmatrix} 1 \\ i \\ 2+3i \end{bmatrix} \cdot \begin{bmatrix} 1+2i \\ -7 \\ 1+i \end{bmatrix} = (1+2i) - 7i + (-1+5i) = 0$$

Hence, $\vec{v}$ and $\vec{w}$ are orthogonal to $\vec{z}$, and $\vec{u}$ is not orthogonal to $\vec{z}$.

Of course, these satisfy all of our familiar properties of length and orthogonality.

THEOREM 4

Let $\mathbb{V}$ be a Hermitian inner product space with inner product $\langle\,,\rangle$. Then, for any $\vec{z}, \vec{w} \in \mathbb{V}$ and $\alpha \in \mathbb{C}$ we have

(1) $\|\alpha\vec{z}\| = |\alpha|\,\|\vec{z}\|$
(2) $\|\vec{z} + \vec{w}\| \leq \|\vec{z}\| + \|\vec{w}\|$
(3) $\dfrac{1}{\|\vec{z}\|}\vec{z}$ is a unit vector in the direction of $\vec{z}$.

DEFINITION

Orthogonal

Orthonormal

Let $S = \{\vec{z}_1, \dots, \vec{z}_k\}$ be a set in a Hermitian inner product space with Hermitian inner product $\langle\,,\rangle$. S is said to be **orthogonal** if $\left\langle \vec{z}_\ell, \vec{z}_j \right\rangle = 0$ for all $\ell \neq j$. S is said to be **orthonormal** if it is orthogonal and $\|\vec{z}_j\| = 1$ for all j.

THEOREM 5

If $S = \{\vec{z}_1, \dots, \vec{z}_k\}$ is an orthogonal set in a Hermitian inner product space $\mathbb{V}$, then

$$\|\vec{z}_1 + \cdots + \vec{z}_k\|^2 = \|\vec{z}_1\|^2 + \cdots + \|\vec{z}_k\|^2$$

THEOREM 6

If $S = \{\vec{z}_1, \dots, \vec{z}_k\}$ is an orthogonal set of non-zero vectors in a Hermitian inner product space $\mathbb{V}$, then S is linearly independent.

The concept of projections and, of course, the Gram-Schmidt procedure are still valid for Hermitian inner products.

EXAMPLE 7

Use the Gram-Schmidt procedure to find an orthogonal basis for the subspace $\mathbb{S}$ of $\mathbb{C}^4$ defined by

$$\mathbb{S} = \operatorname{Span}\left\{\begin{bmatrix}1\\0\\1\\0\end{bmatrix}, \begin{bmatrix}i\\1\\i\\1\end{bmatrix}, \begin{bmatrix}1+i\\-1\\i\\1+i\end{bmatrix}\right\} = \{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$$

Solution: First, we let $\vec{v}_1 = \vec{w}_1$. Then we have

$$\vec{v}_2 = \vec{w}_2 - \frac{\langle \vec{w}_2, \vec{v}_1\rangle}{\|\vec{v}_1\|^2}\vec{v}_1 = \begin{bmatrix}i\\1\\i\\1\end{bmatrix} - \frac{2i}{2}\begin{bmatrix}1\\0\\1\\0\end{bmatrix} = \begin{bmatrix}0\\1\\0\\1\end{bmatrix}$$

$$\begin{aligned}\vec{v}_3 &= \vec{w}_3 - \frac{\langle \vec{w}_3, \vec{v}_1\rangle}{\|\vec{v}_1\|^2}\vec{v}_1 - \frac{\langle \vec{w}_3, \vec{v}_2\rangle}{\|\vec{v}_2\|^2}\vec{v}_2 \\ &= \begin{bmatrix}1+i\\-1\\i\\1+i\end{bmatrix} - \frac{1+2i}{2}\begin{bmatrix}1\\0\\1\\0\end{bmatrix} - \frac{i}{2}\begin{bmatrix}0\\1\\0\\1\end{bmatrix} = \begin{bmatrix}1/2\\-1-i/2\\-1/2\\1+i/2\end{bmatrix}\end{aligned}$$

Consequently, $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is an orthogonal basis for $\mathbb{S}$.

We recall that a real square matrix with orthonormal columns is an orthogonal matrix. Since orthogonal matrices were so useful in the real case, we generalize them as well.

DEFINITION
Unitary Matrix

If the columns of a matrix U form an orthonormal basis for $\mathbb{C}^n$, then U is called **unitary**.

THEOREM 7

Let $U \in M_{n\times n}(\mathbb{C})$. Then the following are equivalent:

(1) the columns of U form an orthonormal basis for $\mathbb{C}^n$.
(2) $\overline{U^T} = U^{-1}$
(3) the rows of U form an orthonormal basis for $\mathbb{C}^n$.

EXAMPLE 8

Determine which of the following matrices are unitary:

$$A = \begin{bmatrix} 1 & 1+2i \\ -1 & 1+2i \end{bmatrix}, \quad B = \begin{bmatrix} (1+i)/\sqrt{2} & 0 & 0 \\ 0 & 0 & 1 \\ 0 & i & 0 \end{bmatrix}, \quad C = \begin{bmatrix} \frac{1+i}{2} & \frac{1-i}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{i}{2} & 0 & \frac{-3-3i}{\sqrt{24}} \\ \frac{1}{2} & \frac{2i}{\sqrt{6}} & \frac{1-i}{\sqrt{24}} \end{bmatrix}$$

Solution: The columns of A are not unit vectors, so A is not unitary.

The columns of B are clearly unit vectors and orthogonal to each other, so the columns of B form an orthonormal basis for $\mathbb{C}^3$. Hence, B is unitary.

We have

$$\overline{C^T}C = \begin{bmatrix} \frac{1-i}{2} & \frac{-i}{2} & \frac{1}{2} \\ \frac{1+i}{\sqrt{6}} & 0 & \frac{-2i}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} & \frac{-3+3i}{\sqrt{24}} & \frac{1+i}{\sqrt{24}} \end{bmatrix} \begin{bmatrix} \frac{1+i}{2} & \frac{1-i}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{i}{2} & 0 & \frac{-3-3i}{\sqrt{24}} \\ \frac{1}{2} & \frac{2i}{\sqrt{6}} & \frac{1-i}{\sqrt{24}} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

so C is unitary.

REMARK

Observe that if the entries of A are all real, then A is unitary if and only if it is orthogonal.

THEOREM 8

If U and W are unitary matrices, then UW is a unitary matrix.

Notice this is the second time we have seen the conjugate of the transpose of a matrix. So, we make the following definition.

DEFINITION

Conjugate Transpose

Let $A \in M_{m\times n}(\mathbb{C})$. Then the **conjugate transpose** of A is

$$A^* = \overline{A}^T$$

EXAMPLE 9

If $A = \begin{bmatrix} 1 & 0 & 1+i & -2i \\ 1-3i & 2 & 2+i & i \\ i & -3i & 4-i & 1-5i \end{bmatrix}$, then $A^* = \begin{bmatrix} 1 & 1+3i & -i \\ 0 & 2 & 3i \\ 1-i & 2-i & 4+i \\ 2i & -i & 1+5i \end{bmatrix}$.

REMARK

Observe that if A is a real matrix, then $A^* = A^T$. Hence, the conjugate transpose is the complex version of the transpose.

THEOREM 9

Let $A, B \in M_{m\times n}(\mathbb{C})$ and let $\alpha \in \mathbb{C}$. Then

(1) $\langle A\vec{z}, \vec{w}\rangle = \langle \vec{z}, A^*\vec{w}\rangle$ for all $\vec{z}, \vec{w} \in \mathbb{C}^n$
(2) $(A^*)^* = A$
(3) $(A+B)^* = A^* + B^*$
(4) $(\alpha A)^* = \overline{\alpha} A^*$
(5) $(AB)^* = B^*A^*$

We end this section by proving Lemma 10.2.2.

LEMMA 10

If A is a symmetric matrix with real entries, then all of its eigenvalues are real.

Proof: Let λ be any eigenvalue of A with corresponding unit eigenvector $\vec{z} \in \mathbb{C}^n$. Then, we have

$$\lambda = \lambda\langle \vec{z}, \vec{z}\rangle = \langle \lambda\vec{z}, \vec{z}\rangle = \langle A\vec{z}, \vec{z}\rangle = A\vec{z}\cdot\overline{\vec{z}} = \vec{z}\cdot A\overline{\vec{z}}$$

by Theorem 10.2.4. Since A has all real entries, we get

$$A\overline{\vec{z}} = \overline{A\vec{z}} = \overline{\lambda\vec{z}} = \overline{\lambda}\overline{\vec{z}}$$

Thus,

$$\lambda = \vec{z}\cdot\overline{\lambda}\overline{\vec{z}} = \overline{\lambda}\vec{z}\cdot\overline{\vec{z}} = \overline{\lambda}\langle \vec{z}, \vec{z}\rangle = \overline{\lambda}$$

Consequently, λ is real. $\square$

Section 11.4 Problems

1. Let $\vec{u} = \begin{bmatrix} 1 \\ 2i \\ 1-i \end{bmatrix}$, $\vec{z} = \begin{bmatrix} 2 \\ -2i \\ 1-i \end{bmatrix}$, and $\vec{w} = \begin{bmatrix} 1+i \\ 1-2i \\ i \end{bmatrix}$ be vectors in $\mathbb{C}^3$. Calculate the following.

   (a) $\|\vec{u}\|$ (b) $\|\vec{w}\|$ (c) $\langle \vec{u}, \vec{z} \rangle$ (d) $\langle \vec{z}, \vec{u} \rangle$
   (e) $\langle \vec{u}, (2+i)\vec{z} \rangle$ (f) $\langle \vec{z}, \vec{w} \rangle$ (g) $\langle \overline{\vec{w}}, \overline{\vec{z}} \rangle$ (h) $\langle \vec{u}+\vec{z}, 2i\vec{w} - i\vec{z} \rangle$

2. Determine which of the following matrices is unitary.

   (a) $\begin{bmatrix} i/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -i/\sqrt{2} \end{bmatrix}$ (b) $\begin{bmatrix} (1+i)/2 & (1-i)/\sqrt{6} & (1-i)/\sqrt{12} \\ 1/2 & 0 & 3i/\sqrt{12} \\ 1/2 & 2i/\sqrt{6} & -i/\sqrt{12} \end{bmatrix}$

   (c) $\begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$ (d) $\begin{bmatrix} 0 & i & 0 \\ i & 0 & 0 \\ 0 & 0 & (1+i)/2 \end{bmatrix}$

3. Consider $\mathbb{C}^3$ with its standard inner product. Let $\vec{z} = \begin{bmatrix} 1+i \\ 2-i \\ -1+i \end{bmatrix}$, $\vec{w} = \begin{bmatrix} 1-i \\ -2-3i \\ -1 \end{bmatrix}$.

   (a) Evaluate $\langle \vec{z}, \vec{w} \rangle$ and $\langle \vec{w}, 2i\vec{z} \rangle$.
   (b) Find a vector in Span$\{\vec{z}, \vec{w}\}$ that is orthogonal to $\vec{z}$.
   (c) Write the formula for the projection of $\vec{u}$ onto $S = \text{Span}\{\vec{z}, \vec{w}\}$ and then find the projection.

4. Let $A, B \in M_{m\times n}(\mathbb{C})$.

   (a) Prove that $(A^*)^* = A$.
   (b) $\langle A\vec{z}, \vec{w} \rangle = \langle \vec{z}, A^*\vec{w} \rangle$ for all $\vec{z}, \vec{w} \in \mathbb{C}^n$.

5. Let $U \in M_{n\times n}(\mathbb{C})$. Prove that if the columns of U form an orthonormal basis for $\mathbb{C}^n$, then $U^* = U^{-1}$.

6. Let U be an $n \times n$ unitary matrix.

   (a) Show that $\langle U\vec{z}, U\vec{w} \rangle = \langle \vec{z}, \vec{w} \rangle$ for any $\vec{z}, \vec{w} \in \mathbb{C}^n$.
   (b) Suppose λ is an eigenvalue of U. Use part (a) to show that $|\lambda| = 1$.
   (c) Find a unitary matrix U which has eigenvalue λ where $\lambda \neq \pm 1$.

7. Let $\mathbb{V}$ be a complex inner product space, with complex inner product $\langle\, , \rangle$. Prove that if $\langle \vec{u}, \vec{v} \rangle = 0$, then $\|\vec{u}+\vec{v}\|^2 = \|\vec{u}\|^2 + \|\vec{v}\|^2$. Is the converse true?

11.5 Unitary Diagonalization

In Section 10.2, we saw that every real symmetric matrix is orthogonally diagonalizable. It is natural to ask if there is a comparable result in the case of matrices with complex entries.

First, we observe that if A is a real symmetric matrix, then the condition $A^T = A$ is equivalent to $A^* = A$. Hence, this condition should be our equivalent of symmetric.

DEFINITION
Hermitian matrix

A matrix $A \in M_{n\times n}(\mathbb{C})$ such that $A^* = A$ is called **Hermitian**.

EXAMPLE 1

Which of the following matrices are Hermitian?

$$A = \begin{bmatrix} 2 & 3-i \\ 3+i & 4 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 2i \\ -2i & 3-i \end{bmatrix}, \qquad C = \begin{bmatrix} 0 & i & i \\ -i & 0 & i \\ -i & i & 0 \end{bmatrix}$$

Solution: We have $A^* = \begin{bmatrix} 2 & 3-i \\ 3+i & 4 \end{bmatrix} = A$, so A is Hermitian.

$B^* = \begin{bmatrix} 1 & 2i \\ -2i & 3+i \end{bmatrix} \neq B$, so B is not Hermitian.

$C^* = \begin{bmatrix} 0 & i & i \\ -i & 0 & -i \\ -i & -i & 0 \end{bmatrix} \neq C$, so C is not Hermitian.

Observe that if A is Hermitian then we have $\overline{(A)_{\ell j}} = A_{j\ell}$, so the diagonal entries of A must be real, and for $\ell \neq j$ the ℓj-th entry must be the complex conjugate of the $j\ell$-th entry. Moreover, we see that every real symmetric matrix is Hermitian. Thus, we expect that Hermitian matrices satisfy the same properties as symmetric matrices.

THEOREM 1

An $n \times n$ matrix is Hermitian if and only if for all $\vec{z}, \vec{w} \in \mathbb{C}^n$ we have

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

Proof: If A is Hermitian, then $A^* = A$ and hence by Theorem 11.4.9 we get

$$\langle A\vec{z}, \vec{w}\rangle = \langle \vec{z}, A^*\vec{w}\rangle = \langle \vec{z}, A\vec{w}\rangle$$

If $< \vec{z}, A\vec{w} >=< A\vec{z}, \vec{w} >$, then we have

$$\vec{z}^T\,\overline{A}\,\overline{\vec{w}} = \vec{z}^T\,\overline{A\vec{w}} = \langle \vec{z}, A\vec{w}\rangle = \langle A\vec{z}, \vec{w}\rangle = (A\vec{z})^T\overline{\vec{w}} = \vec{z}^T A^T \overline{\vec{w}}$$

Since this is valid for all $\vec{z}, \vec{w} \in \mathbb{C}^n$, we have that $\overline{A} = A^T$, and hence $A = A^*$. Thus A is Hermitian. $\square$

THEOREM 2

Suppose A is an $n \times n$ Hermitian matrix. Then:

(1) all the eigenvalues of A are real.
(2) if λ_1 and λ_2 are distinct eigenvalues with corresponding eigenvectors $\vec{v}_1$ and $\vec{v}_2$ respectively, then $\vec{v}_1$ and $\vec{v}_2$ are orthogonal.

From this result we expect to get something very similar to the Principal Axis Theorem for Hermitian matrices. But, instead of orthogonally diagonalizing, we now look at the complex equivalent... unitarily diagonalizing.

DEFINITION
Unitarily Similar

If A and B are matrices such that $U^*AU = B$, where U is a unitary matrix, then we say that A and B are **unitarily similar**.

If A and B are unitarily similar, then they are similar. Consequently, all of our properties of similarity still applies.

DEFINITION
Unitarily Diagonalizable

An $n \times n$ matrix A is said to be **unitarily diagonalizable** if it is unitarily similar to a diagonal matrix D.

EXAMPLE 2

Let $A = \begin{bmatrix} -4 & 1-3i \\ 1+3i & 5 \end{bmatrix}$. Verify that A is Hermitian and show that A is unitarily diagonalizable.

Solution: We see that $A^* = \begin{bmatrix} -4 & 1-3i \\ 1+3i & 5 \end{bmatrix} = A$, so A is Hermitian. We have

$$C(\lambda) = \begin{vmatrix} -4-\lambda & 1-3i \\ 1+3i & 5-\lambda \end{vmatrix} = \lambda^2 - \lambda - 30 = (\lambda+5)(\lambda-6)$$

Hence, the eigenvalues are $\lambda_1 = -5$ and $\lambda_2 = 6$. We have

$$A + 5I = \begin{bmatrix} 1 & 1-3i \\ 1+3i & 10 \end{bmatrix} \sim \begin{bmatrix} 1 & 1-3i \\ 0 & 0 \end{bmatrix}$$

So a basis for E_{λ_1} is $\left\{\begin{bmatrix} -1+3i \\ 1 \end{bmatrix}\right\}$. We also have

$$A - 6I = \begin{bmatrix} -10 & 1-3i \\ 1+3i & -1 \end{bmatrix} \sim \begin{bmatrix} 1 & -(1-3i)/10 \\ 0 & 0 \end{bmatrix}$$

So a basis for E_{λ_2} is $\left\{\begin{bmatrix} 1-3i \\ 10 \end{bmatrix}\right\}$. Observe that

$$\left\langle \begin{bmatrix} -1+3i \\ 1 \end{bmatrix}, \begin{bmatrix} 1-3i \\ 10 \end{bmatrix} \right\rangle = (-1+3i)(1+3i) + 1(10) = 0$$

Thus, the eigenvectors are orthogonal. Normalizing them, we get that A is diagonalized by the unitary matrix $U = \begin{bmatrix} -(1+3i)/\sqrt{11} & (1-3i)/\sqrt{110} \\ 1/\sqrt{11} & 10/\sqrt{110} \end{bmatrix}$ to $D = \begin{bmatrix} -5 & 0 \\ 0 & 6 \end{bmatrix}$.

We expect that every Hermitian matrix is unitarily diagonalizable. We will prove this exactly the same way that we did in Section 10.2 for symmetric matrices; by first proving that every square matrix is unitarily similar to an upper triangular matrix. However, we now get the additional benefit of not having to restrict ourselves to real eigenvalues as we did with the Triangularization Theorem.

THEOREM 3 **(Schur's Theorem)**

If A is an $n \times n$ matrix, then A is unitarily similar to an upper triangular matrix whose diagonal entries are the eigenvalues of A.

Proof: We prove this by induction. If $n = 1$, then A is upper triangular. Assume the result holds for all $(n-1) \times (n-1)$ matrices. Let λ_1 be an eigenvalue of A with corresponding unit eigenvector $\vec{z}_1$. Extend $\{\vec{z}_1\}$ to an orthonormal basis $\{\vec{z}_1, \dots, \vec{z}_n\}$ of $\mathbb{C}^n$ and let $U_1 = \begin{bmatrix} \vec{z}_1 & \cdots & \vec{z}_n \end{bmatrix}$. Then, U_1 is unitary and we have

$$U_1^* A U_1 = \begin{bmatrix} \overline{\vec{z}_1}^T \\ \vdots \\ \overline{\vec{z}_n}^T \end{bmatrix} A \begin{bmatrix} \vec{z}_1 & \cdots & \vec{z}_n \end{bmatrix} = \begin{bmatrix} \overline{\vec{z}_1}^T \lambda_1 \vec{z}_1 & \overline{\vec{z}_1}^T A\vec{z}_2 & \cdots & \overline{\vec{z}_1}^T A\vec{z}_n \\ \overline{\vec{z}_2}^T \lambda_1 \vec{z}_1 & \overline{\vec{z}_2}^T A\vec{z}_2 & \cdots & \overline{\vec{z}_2}^T A\vec{z}_n \\ \vdots & \vdots & & \vdots \\ \overline{\vec{z}_n}^T \lambda_1 \vec{z}_1 & \overline{\vec{z}_n}^T A\vec{z}_2 & \cdots & \overline{\vec{z}_n}^T A\vec{z}_n \end{bmatrix}$$

For $1 \leq j \leq n$ we have

$$\overline{\vec{z}_j}^T \lambda_1 \vec{z}_1 = \lambda_1 \overline{\vec{z}_j}^T \vec{z}_1 = \lambda_1 \overline{\vec{z}_j^T} \vec{z}_1 = \lambda_1 \overline{\vec{z}_j} \cdot \vec{z}_1 = \lambda_1 \vec{z}_1 \cdot \overline{\vec{z}_j} = \lambda_1 \langle \vec{z}_1, \vec{z}_j \rangle$$

Since $\{\vec{z}_1, \dots, \vec{z}_n\}$ is orthonormal, we get that $\overline{\vec{z}_1}^T \lambda_1 \vec{z}_1 = \lambda_1$ and $\overline{\vec{z}_j}^T \lambda_1 \vec{z}_1 = 0$ for $2 \leq j \leq n$. Hence, we can write

$$U_1^* A U_1 = \begin{bmatrix} \lambda_1 & \vec{b}^T \\ 0 & A_1 \end{bmatrix}$$

where A_1 is an $(n-1) \times (n-1)$ matrix and $\vec{b} \in \mathbb{C}^{n-1}$. Thus, by the inductive hypothesis we get that there exits a unitary matrix Q such that $Q^* A_1 Q = T_1$ is upper triangular. Let $U_2 = \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix}$, then U_2 is clearly unitary and hence $U = U_1 U_2$ is unitary. Thus,

$$U^* A U = (U_1 U_2)^* A (U_1 U_2) = U_2^* U_1^* A U_1 U_2 = \begin{bmatrix} 1 & 0 \\ 0 & Q^* \end{bmatrix} \begin{bmatrix} \lambda_1 & \vec{b}^T \\ 0 & A_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix} = \begin{bmatrix} \lambda_1 & \vec{b}^T Q \\ 0 & T_1 \end{bmatrix}$$

which is upper triangular. Since unitarily similar matrices have the same eigenvalues and the eigenvalues of a triangular matrix are its diagonal entries, we get that the eigenvalues of A are along the main diagonal of $U^* A U$ as required. $\square$

REMARK

Schur's Theorem shows that for every $n \times n$ matrix A there exists a unitary matrix U such that $U^*AU = T$. Since $U^T = U^{-1}$ we can solve this for A to get $A = UTU^*$. This is called a Schur decomposition of A.

We can now prove the result corresponding to the Principal Axis Theorem.

THEOREM 4 **(Spectral Theorem for Hermitian Matrices)**

If A is Hermitian, then it is unitarily diagonalizable.

Proof: By Schur's Theorem, there exists a unitary matrix U and an upper triangular matrix T such that $U^*AU = T$. Since $A^* = A$ we get that

$$T^* = (U^*AU)^* = U^*A^*U^{**} = U^*AU = T$$

Since T is upper triangular, we get that T^* is lower triangular. Thus, T is both upper and lower triangular and hence is diagonal. Consequently, U unitarily diagonalizes A. □

In the real case, we also had that every orthogonally diagonalizable matrix was symmetric. Unfortunately, the corresponding result is not true in the complex case as the next example demonstrates.

EXAMPLE 3 Prove that $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ is unitarily diagonalizable, but not Hermitian.

Solution: Observe that $A^* = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \neq A$, so A is not Hermitian.

In Example 11.3.1, we showed that A is diagonalized by $P = \begin{bmatrix} -i & i \\ 1 & 1 \end{bmatrix}$. Observe that

$$\left\langle \begin{bmatrix} -i \\ 1 \end{bmatrix}, \begin{bmatrix} i \\ 1 \end{bmatrix} \right\rangle = (-i)(-i) + 1(1) = 0$$

so the columns of P are orthogonal. Hence, if we normalize them, we get that A is unitarily diagonalized by $U = \begin{bmatrix} -\frac{i}{\sqrt{2}} & \frac{i}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$.

The matrix in Example 3 satisfies $A^* = -A$ and so is called **skew-Hermitian**. Thus, we see that there are matrices which are not Hermitian but whose eigenvectors form an orthonormal basis for $\mathbb{C}^n$ and hence are unitarily diagonalizable. In particular, in the proof of the Spectral Theorem for Normal Matrices, we can observe that the reverse direction fails because if A is not Hermitian, we cannot guarantee that $T^* = T$ since some of the eigenvalues may not be real.

So, as we did in the real case, we should look for a necessary condition for a matrix to be unitarily diagonalizable.

Assume that eigenvectors of A form an orthonormal basis $\{\vec{z}_1, \dots, \vec{z}_n\}$ for $\mathbb{C}^n$. Then, we know that $U = \begin{bmatrix} \vec{z}_1 & \cdots & \vec{z}_n \end{bmatrix}$ unitarily diagonalizes A. That is, we have $U^*AU = D$, where D is diagonal. Hence, $D^* = U^*A^*U$. Now, notice that $DD^* = D^*D$ since D and D^* are diagonal. This gives

$$(U^*AU)(U^*A^*U) = (U^*A^*U)(U^*AU) \Rightarrow U^*AA^*U = U^*A^*AU$$

Consequently, if A is unitarily diagonalizable, then we must have $AA^* = A^*A$.

DEFINITION
Normal Matrix

An $n \times n$ matrix A is called **normal** if $AA^* = A^*A$.

THEOREM 5

(Spectral Theorem for Normal Matrices)

A matrix A is unitarily diagonalizable if and only if A is normal.

Proof: We proved that every unitarily diagonalizable matrix is normal above. So, we just need to prove that every normal matrix is unitarily diagonalizable. Of course, we do this using Schur's Theorem.

By Schur's Theorem, there exists an upper triangular matrix T and a unitary matrix U such that $U^*AU = T$. We just need to prove that T is in fact diagonal. Observe that

$$TT^* = (U^*AU)(U^*A^*U) = U^*AA^*U = U^*A^*AU = (U^*A^*U)(U^*AU) = T^*T.$$

Hence T is also normal and if we compare the diagonal entries of TT^* and T^*T we get

$$\begin{aligned} |t_{11}|^2 + |t_{12}|^2 + \cdots + |t_{1n}|^2 &= |t_{11}|^2 \\ |t_{22}|^2 + \cdots + |t_{2n}|^2 &= |t_{12}|^2 + |t_{22}|^2 \\ &\vdots \\ |t_{nn}|^2 &= |t_{1n}|^2 + |t_{2n}|^2 + \cdots + |t_{nn}|^2 \end{aligned}$$

Hence, we must have $t_{\ell j} = 0$ for all $\ell \neq j$, so T is diagonal as required. □

As a result normal matrices are very important. We now look at some useful properties of normal matrices.

THEOREM 6

If A is an $n \times n$ normal matrix, then

(1) $\|A\vec{z}\| = \|A^*\vec{z}\|$, for all $\vec{z} \in \mathbb{C}^n$.
(2) $A - \lambda I$ is normal for every $\lambda \in \mathbb{C}$.
(3) If $A\vec{z} = \lambda\vec{z}$, then $A^*\vec{z} = \overline{\lambda}\vec{z}$.

Proof: For (1) observe that if $AA^* = A^*A$, then $\overline{A}A^T = A^T\overline{A}$. Then

$$\begin{aligned}||A\vec{z}||^2 &=< A\vec{z}, A\vec{z} >= (A\vec{z})^T \overline{A\vec{z}} = \vec{z}^T A^T \overline{A}\overline{\vec{z}} = \vec{z}^T \overline{A} A^T \overline{\vec{z}} \\ &= \vec{z}^T (A^*)^T \overline{A^*\vec{z}} = (A^*\vec{z})^T \overline{A^*\vec{z}} =< A^*\vec{z}, A^*\vec{z} >= ||A^*\vec{z}||^2\end{aligned}$$

for any $\vec{z} \in \mathbb{C}^n$.

For (2) we observe that

$$\begin{aligned}(A - \lambda I)(A - \lambda I)^* &= (A - \lambda I)(A^* - \overline{\lambda}I) = AA^* - \lambda A^* - \overline{\lambda}A + |\lambda|^2 I \\ &= A^*A - \overline{\lambda}A - \lambda A^* + |\lambda|^2 I = (A^* - \overline{\lambda}I)(A - \lambda I) \\ &= (A - \lambda I)^*(A - \lambda I)\end{aligned}$$

For (3) suppose that $A\vec{z} = \lambda\vec{z}$ for some $\vec{z} \in \mathbb{C}^n$ and let $B = A - \lambda I$. Then B is normal by (2) and

$$B\vec{z} = (A - \lambda I)\vec{z} = A\vec{z} - \lambda\vec{z} = \vec{0}.$$

So, by (1) we get

$$0 = ||B\vec{z}|| = ||B^*\vec{z}|| = ||(A^* - \overline{\lambda}I)\vec{z}|| = ||A^*\vec{z} - \overline{\lambda}\vec{z}||$$

Thus, $A^*\vec{z} = \overline{\lambda}\vec{z}$, as required. □

Section 11.5 Problems

1. Unitarily diagonalize the following matrices.

   (a) $\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$ (b) $B = \begin{bmatrix} 4i & 1+3i \\ -1+3i & i \end{bmatrix}$ (c) $C = \begin{bmatrix} 1 & 0 & 1+i \\ 0 & 2 & 0 \\ 1-i & 0 & 0 \end{bmatrix}$

2. Prove that all the eigenvalues of a Hermitian matrix A are real.

3. Prove that all the eigenvalues of a skew-Hermitian matrix A are purely imaginary.

4. Let $A \in M_{n\times n}(\mathbb{C})$ satisfy $A^* = iA$.

   (a) Prove that A is normal.

   (b) Show that every eigenvalue λ of A must satisfy $\lambda = -i\overline{\lambda}$.

5. Let $A \in M_{n\times n}(\mathbb{C})$ be normal and invertible. Prove that $B = A^*A^{-1}$ is unitary.

6. Let A and B be Hermitian matrices. Prove that AB is Hermitian if and only if $AB = BA$.

7. Let $A = \begin{bmatrix} 2i & -2+i \\ 1-i & 3 \end{bmatrix}$.

   (a) Prove that $\lambda_1 = 2 + i$ is an eigenvalue of A.

   (b) Find a unitary matrix U such that $U^*AU = T$ is upper triangular.

11.6 Cayley-Hamilton Theorem

We now use Schur's theorem to prove an important result about the relationship of a matrix with its characteristic polynomial.

THEOREM 1 **(Cayley-Hamilton Theorem)**

If $A \in M_{n\times n}(\mathbb{C})$, then A is a root of its characteristic polynomial $C(\lambda)$.

Proof: By Schur's theorem, there exists a unitary matrix U and an upper triangular matrix T such that $U^*AU = T$. Let

$$C(\lambda) = c_n\lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0 = (\lambda - \lambda_1)\cdots(\lambda - \lambda_n)$$

For any $n \times n$ matrix X we define

$$C(X) = c_nX^n + c_{n-1}X^{n-1} + \cdots + c_1X + c_0I$$

Observe that

$$\begin{aligned} C(T) &= C(U^*AU) \\ &= c_n(U^*AU)^n + \cdots + c_1(U^*AU) + c_0I \\ &= c_nU^*A^nU + \cdots + c_1U^*AU + c_0U^*U \\ &= U^*c_nA^nU + \cdots + U^*c_1AU + U^*c_0U \\ &= U^*(c_nA^n + \cdots + c_1A + c_0I)U \\ &= U^*C(A)U \end{aligned}$$

We can complete the proof by showing $C(T)$ is the zero matrix.

Since the eigenvalues $\lambda_1, \dots, \lambda_n$ of A are the diagonal entries of T we have

$$C(X) = (-1)^n(X - \lambda_1I)(X - \lambda_2I)\cdots(X - \lambda_nI)$$

so

$$C(T) = (-1)^n(T - \lambda_1I)(T - \lambda_2I)\cdots(T - \lambda_nI). \tag{11.1}$$

Observe that the first column of $T - \lambda_1I$ is zero since the first column of T is $\begin{bmatrix}\lambda_1\\0\\\vdots\\0\end{bmatrix}$.

Then, since the second column of $T - \lambda_2I$ is $\begin{bmatrix}c\\0\\\vdots\\0\end{bmatrix}$, we get that $(T - \lambda_1I)(T - \lambda_2I)$ has the first two columns zero since the first two columns of $T - \lambda_1I$ have the form $\begin{bmatrix}b\\0\\\vdots\\0\end{bmatrix}$.

Continuing, we see that all the columns of Equation 11.1 are zero as required. □

EXAMPLE 1

Show that $A = \begin{bmatrix} a & b \\ 0 & c \end{bmatrix}$ is a root of its characteristic polynomial.

Solution: Since A is upper triangular, we have $C(\lambda) = (\lambda - a)(\lambda - c) = \lambda^2 - (a+c)\lambda + ac$. Then,

$$\begin{aligned} C(A) &= A^2 - (a+c)A + acI \\ &= \begin{bmatrix} a^2 & ab+bc \\ 0 & c^2 \end{bmatrix} - \begin{bmatrix} a^2+ac & ab+cb \\ 0 & ac+c^2 \end{bmatrix} + \begin{bmatrix} ac & 0 \\ 0 & ac \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \end{aligned}$$

EXERCISE 1

Show that $A = \begin{bmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & f \end{bmatrix}$ is a root of its characteristic polynomial in two ways. First, show it directly by using the method of Example 1. Second, show it by evaluating

$$C(A) = (A - aI)(A - dI)(A - fI)$$

by multiplying from left to right as in the proof.