Linear Algebra II

Gregg Waterman
Oregon Institute of Technology

Contents

1 Vector Spaces

Outcome/Performance Criteria:

1. Understand and work with general vector spaces.
   (a) Determine whether or not an object is in a given vector space.
   (b) Prove that a condition for a vector space holds, or provide a counterexample that it doesn't hold.
   (c) Find a linear combination of vectors; find a linear combination of some vectors that equals a given vector.
   (d) Determine whether a vector is in the span of a given set of vectors.
   (e) Describe the span of a set of vectors.
   (f) Determine whether a subset of a vector space is a subspace. If so, prove that it is; if not, show/explain that it does not contain the zero vector or give a *specific* example showing that it is not closed under one of addition or scalar multiplication.
   (g) Determine whether a set of vectors spans a vector space or subspace. If it doesn't, give a vector not in the span and prove that it is not.
   (h) Determine whether a set of vectors is linearly independent; if so, give a non-trivial linear combination of them that equals the zero vector or give one as a linear combination of the others.
   (i) Determine whether a set of vectors is a basis for a vector space; if not, tell why.
   (j) Give the representation of a vector with respect to a given basis, and give its coordinate vector in the appropriate $\mathbb{R}^n$.
   (k) Extend a set to a basis or reduce a spanning set to a basis.
   (l) For a vector space V with bases $\mathcal{B}$ and $\mathcal{C}$, given the coordinate vector of a vector $\mathbf{v} \in V$ with respect to $\mathcal{B}$, find the coordinate vector of v with respect to c.

In Math 341 you worked with vectors in $\mathbb{R}^n$, which generalize your earlier concept of a vector as a quantity in $\mathbb{R}^2$ or $\mathbb{R}^3$ having both magnitude and direction. Careful examination of the way we worked with vectors show that the vectors themselves, along with the way that we add them and multiply them by scalars, have certain properties. We might ask if there are other objects that behave in the same way, and there are! Since these other objects act in the same way as vectors in $\mathbb{R}^n$, we will call them vectors as well.

Before giving a formal definition of a vector space, we need to make brief mention of the scalars associated with a set of vectors. When considering the vector spaces $\mathbb{R}^2, \mathbb{R}^3$ and $\mathbb{R}^n$, the scalars used are always real numbers $\mathbb{R}$, with the two operations of addition and multiplication (subtraction and division are then defined in terms of addition and multiplication). The real numbers are the most common example of a mathematical structure called a **field**. We will not take the time to define a field more formally, but the other fields we will be mildly interested in are the field of complex numbers, denoted by $\mathbb{C}$, and the integers modulo two, denoted by $\mathbb{Z}_2$. This last field is described in more detail in Appendix A.

1.1 Vector Spaces

Performance Criterion:

1. (a) Determine whether or not an object is in a given vector space.
   (b) Prove that a condition for a vector space holds, or provide a counterexample that it doesn't hold.

The real numbers are the most common example of a mathematical structure called a **field**. We will not take the time to define a field more formally, but the other fields we will be mildly interested in are the field of complex numbers, denoted by $\mathbb{C}$, and the integers modulo two, denoted by $\mathbb{Z}_2$. This last field is described in more detail in Appendix A. *Assume that the field of scalars for any vector space is the real numbers, unless told otherwise.*

Definition 1.1: Let $\mathbb{F}$ be a field and let V be a set of objects which can be added to each other and multiplied by numbers from $\mathbb{F}$. If, for any $\mathbf{u}$ and $\mathbf{v}$ in V and any c and d in $\mathbb{F}$, all of the following properties hold, then we say that V is a **vector space** over the field $\mathbb{F}$. We call the elements of V **vectors** and the elements of $\mathbb{F}$ **scalars**.

(1) $\mathbf{u}+\mathbf{v}$ is in V — *closure under addition*
(2) $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$ — *commutativity*
(3) $(\mathbf{u}+\mathbf{v})+\mathbf{w}=\mathbf{u}+(\mathbf{v}+\mathbf{w})$ — *associativity*
(4) There exists an element $\mathbf{0}$ in V such that $\mathbf{u}+\mathbf{0}=\mathbf{u}$ — *additive identity*
(5) For each $\mathbf{u}$ in V, there is an element $-\mathbf{u}$ in V such that $\mathbf{u}+(-\mathbf{u})=\mathbf{0}$ — *additive inverses*
(6) $c\mathbf{u}$ is in V — *closure under scalar multiplication*
(7) $c(\mathbf{u}+\mathbf{v})=c\mathbf{u}+c\mathbf{v}$ — *scalar distributivity*
(8) $(c+d)\mathbf{u}=c\mathbf{u}+d\mathbf{u}$ — *vector distributivity*
(9) $c(d\mathbf{u})=(cd)\mathbf{u}$
(10) $1\mathbf{u}=\mathbf{u}$

There are a few classic examples of vector spaces that you should be familiar with.

◇ **Example 1.1(a):** Consider the familiar set of all vectors $\mathbf{u}=\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$, whose components u_1 and u_2 are real numbers, with the standard componentwise addition and scalar multiplication (with real number scalars). This is the familiar Euclidean space $\mathbb{R}^2$.

◇ **Example 1.1(b):** The **Euclidean spaces** $\mathbb{R}^n$, defined as in the previous example but with n components, are the most fundamental vector spaces.

Since the above definition allows things that we aren't used to thinking of as vectors to be vectors, we sometimes call vector spaces other than the Euclidean spaces $\mathbb{R}^n$ *abstract vector spaces* or *general vector spaces*. The remaining examples are all abstract vector spaces.

◇ **Example 1.1(c):** The complex vector spaces $\mathbb{C}^n$ are just like $\mathbb{R}^n$, except that both the components of vectors and the scalars are complex numbers.

◇ **Example 1.1(d):** The spaces $\mathbb{Z}_2^n$ are also like $\mathbb{R}^n$, but the vector components and scalars are numbers in $\mathbb{Z}_2$, the integers modulo two.

◇ **Example 1.1(e):** Suppose that we consider any two 2×3 matrices

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \qquad \text{and} \qquad B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix},$$

where all of the entries are real numbers. If we use the standard matrix addition and multiplication by a real scalar c defined by

$$A + B = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & a_{13}+b_{13} \\ a_{21}+b_{21} & a_{22}+b_{22} & a_{23}+b_{23} \end{bmatrix} \qquad \text{and} \qquad cA = \begin{bmatrix} ca_{11} & ca_{12} & ca_{13} \\ ca_{21} & ca_{22} & ca_{23} \end{bmatrix},$$

then all of the above properties hold. We denote this vector space by M_{23}. Of course the dimensions of the matrices don't matter, and the general spaces M_{mn} for any natural numbers m and n are defined the same way.

◇ **Example 1.1(f):** If we take any two functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ and any real number c, we can define new functions $f + g : \mathbb{R} \to \mathbb{R}$ and $cf : \mathbb{R} \to \mathbb{R}$ by

$$(f+g)(x) = f(x) + g(x) \qquad \text{and} \qquad (cf)(x) = c[f(x)]$$

for every real number x. Again, all of the vector space properties hold, so these functions form a vector space. We refer to this as the space of real valued functions, denoted by $\mathscr{F}$ or sometimes $\mathscr{F}(\mathbb{R})$.

◇ **Example 1.1(g):** If we consider real valued functions defined only on an interval $[a,b]$, such functions behave exactly as those in the previous example (again with real number scalars), so they are a vector space that we denote by $\mathscr{F}[a,b]$.

◇ **Example 1.1(h):** All polynomial functions of degree n or less, where n is any natural number, can be added and multiplies by scalars in the manner described in Example 1.1(d), and constitute a vector space denoted by $\mathscr{P}_n$. The set of all polynomials of any degree is also a vector space, denoted by just $\mathscr{P}$.

All of the above examples of vector spaces are fairly "ordinary." Here is a more unusual one:

◇ **Example 1.1(i):** Here we let our vectors be the set of positive real numbers, and the scalars are all real numbers. For any positive real numbers x and y and any real number c we define addition $\oplus$ and scalar multiplication $\odot$ by

$$x \oplus y = xy \qquad \text{and} \qquad c \odot x = x^c.$$

These two sets of vectors and scalars, along with the defined addition $\oplus$ and scalar multiplication $\odot$ do indeed meet all of the conditions needed to be a vector space. Let's prove condition (8):

Let x be any positive real number and let c and d be any real numbers. Then

$$(c+d) \odot x = x^{c+d} = x^c x^d = x^c \oplus x^d = c \odot x \oplus d \odot x.$$

Section 1.1 Exercises

1. Which of the following are in $\mathscr{F}$? For any that are not, give a closed interval $[a,b]$ for which the function *IS* in $\mathscr{F}[a,b]$.

   (a) $f(x) = x^2 - x - 2$ (b) $g(x) = \dfrac{1}{x^2 - x - 2}$ (c) $h(x) = \sin x$

   (d) $y = \tan x$ (e) $f(x) = e^x$ (f) $g(x) = |x|$

2. The set of all functions $\mathscr{F}$ that are continuous for all real numbers is denoted by $\mathscr{C}$. It is itself a vector space with the same addition and scalar multiplication as defined for $\mathscr{F}$.

   (a) Which of the functions from Exercise 1 are in $\mathscr{C}$?

   (b) Consider the function $f(x) = \begin{cases} 0 & \text{if } x = 0, \\ \dfrac{x}{|x|} & \text{if } x \neq 0 \end{cases}$. Is f in $\mathscr{F}$? Is it in $\mathscr{C}$?

   (c) The functions that are continuous on an interval $[a,b]$ are also a vector space, denoted by $\mathscr{C}[a,b]$. Redefine the function f from part (b) at $x = 0$ so that it is in $\mathscr{C}[0,1]$.

3. The set of all functions with continuous first derivatives on $\mathbb{R}$ or a closed interval $[a,b]$ are denoted by $\mathscr{C}^1$ and $\mathscr{C}^1[a,b]$, respectively. Similarly, $\mathscr{C}^n$ and $\mathscr{C}^n[a,b]$ are the sets of with continuous nth derivatives on $\mathbb{R}$ or $[a,b]$. With the addition and scalar multiplication from $\mathscr{F}$, each of these is a vector space.

   (a) Which of the functions from Exercise 1 are in $\mathscr{C}^1$?

   (b) For which values of n is the function $y = x^{7/3}$ in $\mathscr{C}^n[-1,1]$?

4. Consider the set of vectors from $\mathbb{R}^2$ with real number scalars, but define the addition of two vectors by

$$\mathbf{u} + \mathbf{v} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} u_1 + v_2 \\ u_2 + v_1 \end{bmatrix}.$$

   (a) Use a sequence of steps like those shown in Example 1.1(i) to show that condition (7) of a vector space holds.

   (b) Give a counterexample showing that condition (2) of a vector space does not hold.

5. Again consider the set of vectors from $\mathbb{R}^2$ with real number scalars, but this time define the addition of two vectors by

$$\mathbf{u} + \mathbf{v} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} u_1 + v_1 + 1 \\ u_2 + v_2 + 1 \end{bmatrix}.$$

   Prove that this is not a vector space by giving a counterexample showing that one of the conditions for a vector space does not hold.

6. (a) Determine the zero vector for Example 1.1(i).

   (b) What is the additive inverse of x for the vector space of Example 1.1(i)?

   (c) Prove that conditions (2) and (7) of a vector space hold for the vector space of Example 1.1(i).

7. Suppose that for a vector space V there are two zero vectors $\mathbf{0}_1$ and $\mathbf{0}_2$ satisfying condition (4) of the definition of a vector space. Show that it must be the case that $\mathbf{0}_1 = \mathbf{0}_2$.

8. Consider the set of vectors from $\mathbb{R}^2$ with real number scalars, but define the addition of two vectors by

$$\mathbf{u} + \mathbf{v} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} u_1 + v_1 \\ 0 \end{bmatrix}.$$

   This appears to meet all of the conditions for a vector space, but it is in fact not a vector space. Why do you suppose that is?

1.2 Linear Combinations and Span of a Set of Vectors

Performance Criterion:

1. (c) Find a linear combination of vectors; find a linear combination of some vectors that equals a given vector.
   (d) Determine whether a vector is in the span of a given set of vectors.
   (e) Describe the span of a set of vectors.

The two main operations with vector spaces are addition of vectors and multiplication of vectors by scalars. When we combine the two we get the most essential concept in linear algebra:

Definition 1.2: A **linear combination** of the vectors $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n$ is a vector

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n,$$

where $c_1, c_2, ..., c_n$ are scalars.

◇ **Example 1.2(a):** Let $p_1(x) = x + 1$, $p_2(x) = x^2 + 1$ and $p_3(x) = x^2 + x$. Give the linear combination $5p_1 - 2p_2 + 4p_3$.

$$5p_1 - 2p_2 + 4p_3 = 5(x+1) - 2(x^2+1) + 4(x^2+x) = 5x + 5 - 2x^2 - 2 + 4x^2 + 4x = 2x^2 + 9x + 3$$

More often it is the case that we wish to find a linear combination of some vectors equalling a given vector, as demonstrated in the next example.

◇ **Example 1.2(b):** Find a linear combination of the vectors $\mathbf{v}_1 = \begin{bmatrix} 3 \\ -4 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 7 \\ -3 \end{bmatrix}$ that equals the vector $\mathbf{w} = \begin{bmatrix} 1 \\ -14 \end{bmatrix}$.

We are looking for two scalars c_1 and c_2 such that $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = \mathbf{w}$. By the method of Example 3.3(d) we have

$$c_1 \begin{bmatrix} 3 \\ -4 \end{bmatrix} + c_2 \begin{bmatrix} 7 \\ -3 \end{bmatrix} = \begin{bmatrix} 1 \\ -14 \end{bmatrix}$$

$$\begin{bmatrix} 3c_1 \\ -4c_1 \end{bmatrix} + \begin{bmatrix} 7c_2 \\ -3c_2 \end{bmatrix} = \begin{bmatrix} 1 \\ -14 \end{bmatrix}$$

$$\begin{bmatrix} 3c_1 + 7c_2 \\ -4c_1 - 3c_2 \end{bmatrix} = \begin{bmatrix} 1 \\ -14 \end{bmatrix}$$

In the last line above we have two vectors that are equal. It should be intuitively obvious that this can only happen if the individual components of the two vectors are equal. This results in the system $\begin{array}{rcr} 3c_1 + 7c_2 & = & 1 \\ -4c_1 - 3c_2 & = & -14 \end{array}$ of two equations in the two unknowns c_1 and c_2. Solving, we arrive at $c_1 = 5$, $c_2 = -2$. It is easily verified that these are correct:

$$5\begin{bmatrix} 3 \\ -4 \end{bmatrix} - 2\begin{bmatrix} 7 \\ -3 \end{bmatrix} = \begin{bmatrix} 15 \\ -20 \end{bmatrix} - \begin{bmatrix} 14 \\ -6 \end{bmatrix} = \begin{bmatrix} 1 \\ -14 \end{bmatrix}$$

Definition 1.3: The **span** of a set $\mathcal{S} = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ of vectors is the set of all linear combinations of those vectors. It is denoted $\text{span}(\mathcal{S})$.

Note that the set $\mathcal{S}$ is finite (with n elements), but $\text{span}(\mathcal{S})$ is a set of infinitely many vectors. Generally we will want to know whether a vector $\mathbf{v}$ is in the span of a set $\mathcal{S} = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$. To determine this we attempt to find scalars $c_1, c_2, ..., c_n$ for which

$$\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n.$$

This will always entail solving a system of equations. We give one example here:

◇ **Example 1.2(c):** Determine whether $p_1(x) = -x^2 + 4x - 6$ and $p_2(x) = x^2 - 3x + 2$ are in the span of $\mathcal{S} = \{x^2 + x + 1, x^2 - x + 3\}$.

For p_1, we wish to know whether there are scalars c_1 and c_2 for which

$$-x^2 + 4x - 6 = c_1(x^2 + x + 1) + c_2(x^2 - x + 3) = (c_1 + c_2)x^2 + (c_1 - c_2)x + (c_1 + 3c_2)$$

Equating coefficients of like terms of $-x^2 + 4x - 6$ and $(c_1 + c_2)x^2 + (c_1 - c_2)x + (c_1 + 3c_2)$ gives the system

$$\begin{array}{rcr} c_1 + c_2 & = & -1 \\ c_1 - c_2 & = & 4 \\ c_1 + 3c_2 & = & -6 \end{array}.$$

The solution to this system is $c_1 = \frac{3}{2}$ and $c_2 = -\frac{5}{2}$, so p_1 is in the span of $\mathcal{S}$, with $p_1(x) = \frac{3}{2}(x^2 + x + 1) - \frac{5}{2}(x^2 - x + 3)$. A similar procedure shows that p_2 is *NOT* in $\text{span}(\mathcal{S})$. (See Exercise 5).

We will sometimes want to describe in some way the span of a set of vectors. An example of this is shown here:

◇ **Example 1.2(d):** Describe the span of the set $\mathcal{S} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$.

$\text{span}(\mathcal{S})$ is all vectors of the form $a\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + b\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} a \\ b \\ 0 \end{bmatrix}$, or all vectors in $\mathbb{R}^3$ whose third coordinate is zero. This is the xy-pane in $\mathbb{R}^3$.

Section 1.2 Exercises

1. Find a linear combination of $p_1(x) = x+1$, $p_2(x) = x^2+1$ and $p_3(x) = x^2+x$ that equals $p(x) = 4x^2-4x+2$. Give your final answer in the form $p(x) = c_1p_1(x)+c_2p_2(x)+c_3p_3(x)$ with the values of c_1, c_2 and c_3 inserted.

2. Let $\mathcal{S} = \{A_1, A_2, A_3\} = \left\{ \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right\}$. Determine whether $B = \begin{bmatrix} -4 & 1 \\ 2 & -7 \end{bmatrix}$ is in $\text{span}(\mathcal{S})$; if it is, give a linear combination of A_1, A_2, A_3 equalling B.

3. Determine whether $C = \begin{bmatrix} 0 & 1 \\ 3 & 6 \end{bmatrix}$ is in the span of the set $\mathcal{S}$ given in Exercise 2, and give a linear combination of A_1, A_2, A_3 equalling C if it is.

4. Consider the set $\mathcal{S} = \{x, x^2\}$. Give a polynomial $p_1(x)$ that *IS* in $\text{span}(\mathcal{S})$ and another polynomial $p_2(x)$ that *IS NOT* in $\text{span}(\mathcal{S})$.

5. For the vector space of Example 1.1(i), give the linear combination $(-3) \odot 2 \oplus 2 \odot 5$ as a single number. Give your answer as a mixed number (integer and fraction).

6. For each of the following, determine whether the vector $\mathbf{v}$ is in the span of the set S. If it is, give $\mathbf{v}$ as a linear combination of the vectors in S.

(a) $\mathbf{v} = \begin{bmatrix} 3 \\ -2 \\ -4 \\ 1 \end{bmatrix}$, $S = \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ -3 \\ 1 \end{bmatrix} \right\}$ (b) $\mathbf{v} = \begin{bmatrix} 19 \\ 10 \\ -1 \end{bmatrix}$, $S = \left\{ \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} -5 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 7 \\ -4 \end{bmatrix} \right\}$

7. Without doing any computations, explain briefly (use complete sentences!) why the vector $\mathbf{v} = \begin{bmatrix} -4 \\ 2 \\ 5 \end{bmatrix}$ is not in the span of $S = \left\{ \begin{bmatrix} -3 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} -5 \\ 0 \\ 1 \end{bmatrix} \right\}$.

8. Consider the vectors $\mathbf{v}_1 = \begin{bmatrix} 4 \\ 1 \\ -3 \\ 2 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 6 \\ 0 \\ -1 \\ -5 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$, $\mathbf{v}_4 = \begin{bmatrix} 1 \\ 5 \\ -4 \\ 20 \end{bmatrix}$, $\mathbf{w} = \begin{bmatrix} 6.5 \\ 3.0 \\ -4.5 \\ 2.5 \end{bmatrix}$. Determine whether $\mathbf{w}$ can be written as a linear combination of $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$ and $\mathbf{v}_4$ by doing the following:

   (a) Give the vector equation describing what you are attempting to do.
   (b) Give the augmented matrix for the system to be solved.
   (c) Give the row-reduced form of the augmented matrix.
   (d) Conclude by one of the following:
      - Writing that there is no linear combination equalling $\mathbf{w}$.
      - Giving the unique representation of $\mathbf{w}$ in terms of $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$ and $\mathbf{v}_4$.
      - If there is more than one representation of $\mathbf{w}$ in terms of $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$ and $\mathbf{v}_4$, **give two specific representations and the general representation in terms of a parameter t or s.**

9. Do parts (c) and (d) above for $\mathbf{v}_1 = \begin{bmatrix} 0 \\ 3 \\ -1 \\ 4 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 2 \\ -5 \\ 0 \\ -2 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 7 \\ 10 \\ -4 \\ 1 \end{bmatrix}$, $\mathbf{v}_4 = \begin{bmatrix} 6 \\ 3 \\ 5 \\ 1 \end{bmatrix}$, $\mathbf{w} = \begin{bmatrix} 11.5 \\ 4.5 \\ 8.0 \\ -7.0 \end{bmatrix}$

10. Do parts (c) and (d) above for $\mathbf{v}_1 = \begin{bmatrix} 4 \\ 1 \\ -3 \\ 2 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 6 \\ 0 \\ -1 \\ -5 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$, $\mathbf{v}_4 = \begin{bmatrix} 1 \\ 5 \\ -4 \\ 20 \end{bmatrix}$, $\mathbf{w} = \begin{bmatrix} 7.5 \\ 3.0 \\ 3.5 \\ 1.0 \end{bmatrix}$

1.3 Subspaces

Performance Criterion:

1. (f) Determine whether a subset of a vector space is a subspace. If so, prove that it is; if not, show/explain that it does not contain the zero vector or give a *specific* example showing that it is not closed under one of addition or scalar multiplication.

Suppose that we have a set $\mathcal{S} = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ of vectors, and that both $\mathbf{v}$ and $\mathbf{w}$ are in $\text{span}(\mathcal{S})$. Then there are scalars $c_1, c_2, ..., c_n$ and $d_1, d_2, ..., d_n$ such that

$$\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n \qquad \text{and} \qquad \mathbf{w} = d_1\mathbf{u}_1 + d_2\mathbf{u}_2 + \cdots + d_n\mathbf{u}_n.$$

Then

$$\begin{aligned} \mathbf{v} + \mathbf{w} &= c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n + \mathbf{w} + d_1\mathbf{u}_1 + d_2\mathbf{u}_2 + \cdots + d_n\mathbf{u}_n \\ &= (c_1 + d_1)\mathbf{u}_1 + (c_2 + d_2)\mathbf{u}_2 + \cdots + (c_n + d_n)\mathbf{u}_n \end{aligned}$$

From this we can see that $\mathbf{v} + \mathbf{w}$ is also in $\text{span}(\mathcal{S})$; we say that $\text{span}(\mathcal{S})$ is **closed under addition**. A similar computation would also show that if we multiply any vector in $\text{span}(\mathcal{S})$ by a scalar, the result is also in $\text{span}(\mathcal{S})$. This leads us to make the following definition:

Definition 1.4: Let V be a vector space and W a non-empty subset of V. Then W is a **subspace** of V if and only if the following conditions hold:

a) the zero vector is in W

b) if $\mathbf{u}$ and $\mathbf{v}$ are in W, then $\mathbf{u} + \mathbf{v}$ is in W

c) if $\mathbf{u}$ is in W and c is a scalar, then $c\mathbf{u}$ is in W

The implication of the above is that W is itself a "smaller" vector space contained in V. W "inherits" vector space properties 2, 3, 7, 8, 9, 10 from V, and the other properties follow as a result of the above definition. The first condition above is actually not necessary as part of the definition of a subspace, because it is automatically true if the other two are, but we state it because it offers a convenient way to quickly identify subsets that are not subspaces:

If a subset W of a vector space V does not contain the zero vector, then W is not a subspace of V.

If a subset *does* contain the zero vector but you do not think it is a subspace, you must prove it by finding a specific example where showing the set is not closed under either scalar multiplication or addition.

There are two main ways to tell whether a given subset of a vector space V *IS* a subspace:

1) Show that the set is closed under addition and scalar multiplication, OR

2) Show that the set is the span of a finite set $\mathcal{S}$ of *specific* vectors in V.

If either of those are true, then the zero vector will automatically be included in the subset, making it a subspace.

◇ **Example 1.3(a):** Determine whether the set $\mathcal{S}$ all vectors of the form $\begin{bmatrix} 0 \\ a \\ b \end{bmatrix}$, where a and b represent real numbers, is a subspace of $\mathbb{R}^3$.

Let's determine first whether we think $\mathcal{S}$ is a subspace. We note first that if $a = b = 0$ we get the zero vector, so $\mathcal{S}$ contains the zero vector. This doesn't prove $\mathcal{S}$ is a subspace, but we should always perform that check first, because if it fails we are done. We should be able to see that if we add two vectors of the given form, the first component of the resulting vector will be zero, so the set is closed under addition. Similarly, if we multiply a vector in $\mathcal{S}$ by a scalar we will again get a vector in $\mathcal{S}$, so the set is closed under scalar multiplication also. Therefore it is a subspace.

Although we have accomplished what we wished to, let's show that $\mathcal{S}$ is the span of a set of vectors. If we let a and b be any scalars, we can see that

$$\begin{bmatrix} 0 \\ a \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ a \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ b \end{bmatrix} = a\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + b\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

so $\mathcal{S}$ is the span of the set $\left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$, so it is a subspace.

◇ **Example 1.3(b):** Consider the set of all vectors in $\mathbb{R}^3$ for which at least two of the components are the same as each other. Is that set a subspace of $\mathbb{R}^3$?

We first note that all three components of the zero vector are the same, so the set in question contains the zero vector. It should also be clear that if we add two vectors for which all three components of each are the same, we'll get a vector in which all three components will also be the same. Let's add two vectors in which exactly two of the components are the same, but different components for each of the two vectors:

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

This shows that the set in question is not closed under addition, so it is not a subspace. (Note that the set *IS* closed under scalar multiplication.)

◇ **Example 1.3(c):** Any function for which $f(-x) = f(x)$ is called an even function. Are the even functions a subspace of $\mathscr{F}$?

Suppose that f and g are even and c is any scalar. We need to show that the functions $f+g$ and cf are even. By definition, $(f+g)(-x) = f(-x)+g(-x)$. But f and g are even, so $f(-x)+g(-x) = f(x)+g(x)$ which, in turn, equals $(f+g)(x)$. Therefore $(f+g)(-x) = (f+g)(x)$ and the set of even functions is closed under addition. By similar reasoning,

$$(cf)(-x) = c[f(-x)] = c[f(x)] = (cf)(x),$$

so the even functions are closed under scalar multiplication as well. Thus the set of even functions is a subspace of $\mathscr{F}$.

Section 1.3 Exercises

1. Determine whether all vectors of each of the following forms is a subspace of $\mathbb{R}^3$. For those that are, prove it by showing that the set satisfies conditions (b) and (c) of the definition. For those that aren't, argue why the set does not contain the zero vector or give a *specific* counterexample showing that the set is not closed under addition or scalar multiplication.

   (a) $\begin{bmatrix} a \\ 1 \\ b \end{bmatrix}$ (b) $\begin{bmatrix} a \\ 3a \\ b \end{bmatrix}$ (c) $\begin{bmatrix} a \\ a^2 \\ a^3 \end{bmatrix}$ (d) $\begin{bmatrix} a \\ b \\ c \end{bmatrix}$ (e) $\begin{bmatrix} 0 \\ a \\ 2a \end{bmatrix}$

2. For each of the sets in the previous exercise that *ARE* subspaces, show that the set is the span of a finite set of *specific* vectors.

3. Is the set of vectors of the form $\begin{bmatrix} a+bi \\ a-bi \end{bmatrix}$ a subspace of $\mathbb{C}^2$? *Keep in mind that for this space the scalars are complex numbers.*

4. Is the set of polynomials of degree exactly two a subspace of $\mathscr{P}_2$?

5. A 2×2 matrix A is symmetric if it has the form $A = \begin{bmatrix} a & c \\ c & b \end{bmatrix}$. Is the set of symmetric matrices in M_{22} a subspace of M_{22}?

6. Is the set of polynomials of the form $x^2 + bx + c$ a subspace of $\mathscr{P}_2$?

7. Consider the set of all functions f in $\mathscr{F}$ for which $f(0) = 2$. Is this a subspace of $\mathscr{F}$?

8. Consider all of the functions f for which the second derivative exists and $f'' + f = 0$. Is this a subspace of $\mathscr{F}$?

9. Is the set of all vectors $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ with $3x + z = 0$ a subspace of $\mathbb{R}^3$?

10. Is the set of vectors of the form $\begin{bmatrix} a+bi \\ c \end{bmatrix}$ a subspace of $\mathbb{C}^2$? (c is a real number.)

11. Is the set of all functions in $\mathscr{F}$ for which $f(x) \geq 0$ for all x a subspace of $\mathscr{F}$?

12. Is the set of polynomials of the form $ax^2 + bx$ a subspace of $\mathscr{P}_2$?

13. Is the set of upper triangular matrices in M_{22} a subspace of M_{22}?

14. Is the set of matrices in M_{22} with determinant zero a subspace of M_{22}?

15. We will use the symbol $\mathscr{D}$ to represent differentiable functions defined on all real numbers. Is $\mathscr{D}$ a subspace of $\mathscr{F}$?

16. Are periodic functions with period 2π a subspace of $\mathscr{F}$? You may never have seen a formal definition of periodic, but in this case it means that $f(x + 2\pi) = f(x)$ for all values of x.

17. Any function for which $f(-x) = -f(x)$ is called an odd function. Are the odd functions a subspace of $\mathscr{F}$?

18. Consider the set of all functions f in $\mathscr{F}$ for which $f(2) = 0$. Is this a subspace of $\mathscr{F}$?

19. Consider all of the functions f that are continuous on the interval $[0, 2\pi]$ and for which the second derivative exists on $(0, 2\pi)$. Amongst those functions consider the ones for which $f'' + f = 0$ and $f(0) = f(2\pi) = 0$. Is this a subspace of $\mathscr{F}$?

20. Suppose that U and W are subspaces of a vector space V. Show that $U \cap W$ is a subspace of V.

21. Give two subspaces U and W of $\mathbb{R}^2$ for which $U \cup W$ is *NOT* a subspace of $\mathbb{R}^2$.

1.4 Spanning Sets and Linear Independence

Performance Criteria:

1. (g) Determine whether a set of vectors spans a vector space or subspace. If it doesn't, give a vector not in the span and prove that it is not.

   (h) Determine whether a set of vectors is linearly independent; if not, give a nontrivial linear combination of them that equals the zero vector or give one as a linear combination of the others.

We return now to the idea of the span of a set of vectors, and add a bit to the language.

- Suppose that we have a set $\mathcal{S} = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n$ of vectors in some vector space V. If every vector in some subset W of V can be formed as a linear combination of vectors in $\mathcal{S}$, we say that $\mathcal{S}$ **spans** W.
- If $W = \text{span}(\mathcal{S})$, we say that $\mathcal{S}$ is a **spanning set** for W.

As demonstrated at the beginning of this section, the span of any set of vectors in a vector space V is a subspace of V. We can say more: the span of a set $\mathcal{S} = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ of vectors in a vector space V is the *smallest subspace containing all of the vectors in* $\mathcal{S}$.

As mathematicians, engineers or scientists, we are often trying to solve problems involving vectors in some specific subspace W of a larger vector space V. It is usually desirable to have some "small" set of vectors out of which all of the vectors in our subspace can be "built" using linear combinations. If the set is $\mathcal{S} = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ and we can form some other vector $\mathbf{v}$ as a linear combination

$$\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n,$$

we call $c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n$ the **representation of $\mathbf{v}$ in terms of $\mathcal{S}$**. Using this language, we can now state our objective more precisely: Given a subspace W of a vector space V, we would like to find a subset $\mathcal{S}$ of W such that

1) every vector in W has a representation in terms of $\mathcal{S}$, and

2) that representation is *unique*, meaning it is the only one.

The first of these conditions says that the set $\mathcal{S}$ spans W. The second condition will be met if the vectors in $\mathcal{S}$ are what we call **linearly independent**. If those two conditions are met, then $\mathcal{S}$ is called a **basis** for the subspace W.

Let's consider first the question of whether a given set spans a particular space.

◇ **Example 1.4(a):** The set of matrices of the form $\begin{bmatrix} 0 & a \\ b & c \end{bmatrix}$ is a subspace of M_{22}. Does the set $\mathcal{S} = \left\{ \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right\}$ span the subspace? If not, give a matrix in the subspace that cannot be obtained as a linear combination of elements of $\mathcal{S}$. Then, give any additional matrices needed to span the subspace.

A general linear combination of the vectors in $\mathcal{S}$ has the form

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

so, for example, the matrix $\begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix}$ cannot be obtained by a linear combination of the matrices in the set $\mathcal{S}$. Therefore $\mathcal{S}$ does not span the described set of matrices. It should be clear that if we include the matrix $\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ in $\mathcal{S}$, then $\mathcal{S}$ will span the described set of matrices.

◇ **Example 1.4(b):** Does the set $\mathcal{S} = \{1, x, x^2, x^3, x + x^3\}$ span $\mathscr{P}_3$? If not, give a polynomial in $\mathscr{P}_3$ that cannot be obtained as a linear combination of elements of $\mathcal{S}$.

Any polynomial in $\mathscr{P}_3$ has the form $p(x) = ax^3 + bx^2 + cx + d$ for scalars a, b, c and d. But this polynomial is a linear combination of the first four elements of $\mathcal{S}$, so $\mathcal{S}$ spans $\mathscr{P}_3$.

◇ **Example 1.4(c):** Does the set $\mathcal{S} = \{-1 + x + 2x^2, 1 + x + x^2, -3 + x + 3x^2\}$ span $\mathscr{P}_2$?

This example is not as straightforward as the previous two. Let's suppose that the set *does* span $\mathscr{P}_2$ and see if anything goes wrong. Let $ax^2 + bx + c$ be any polynomial in $\mathscr{P}_2$. If $\mathcal{S}$ does span $\mathscr{P}_2$ there must be constants c_1, c_2 and c_3 for which

$$\begin{aligned} ax^2 + bx + c &= c_1(-1 + x + 2x^2) + c_2(1 + x + x^2) + c_3(-3 + x + 3x^2) \\ &= -c_1 + c_1x + 2c_1x^2 + c_2 + c_2x + c_2x^2 - 3c_3 + c_3x + 3c_3x^2 \\ &= (2c_1 + c_2 + 3c_3)x^2 + (c_1 + c_2 + c_3)x + (-c_1 + c_2 - 3c_3) \end{aligned}$$

This gives us the system

$$\begin{aligned} 2c_1 + c_2 + 3c_3 &= a \\ c_1 + c_2 + c_3 &= b \ , \\ -c_1 + c_2 - 3c_3 &= c \end{aligned}$$

whose augmented matrix and reduced augmented matrices are shown below, with $\hat{a}$, $\hat{b}$ and $\hat{c}$ representing constants depending on the original constants a, b and c.

$$\begin{bmatrix} 2 & 1 & 3 & a \\ 1 & 1 & 1 & b \\ -1 & 1 & -3 & c \end{bmatrix} \quad \stackrel{rref}{\Longrightarrow} \quad \begin{bmatrix} 1 & 0 & 2 & \hat{a} \\ 0 & 1 & -1 & \hat{b} \\ 0 & 0 & 0 & \hat{c} \end{bmatrix}$$

The last row of the reduced matrix represents the equation $0c_1 + 0c_2 + 0c_3 = \hat{c}$, which is only true if $\hat{c} = 0$. However, $\hat{c}$ can be nonzero for the right choices of a, b and c, so the system has no solution. Therefore the set $\mathcal{S}$ does not span $\mathscr{P}_3$. In the exercises you wil be asked to give a specific polymonial that is not in the sapn of $\mathcal{S}$.

◇ **Example 1.4(d):** Does the set $\mathcal{S} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$ span $\mathbb{C}^2$? If not, give a vector in $\mathbb{C}^2$ that cannot be obtained as a linear combination of elements of $\mathcal{S}$. Then, give any additional vectors needed to span $\mathbb{C}^2$.

Remembering again that the scalars for $\mathbb{C}^2$ are complex numbers, we see that we can represent any vector

$$\begin{bmatrix} a + bi \\ c + di \end{bmatrix} \quad \text{in } \mathbb{C}^2 \text{ by} \quad (a + bi)\begin{bmatrix} 1 \\ 0 \end{bmatrix} + (c + di)\begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Therefore $\mathcal{S}$ spans $\mathbb{C}^2$.

Linear Independence

We now focus on our desire to have unique representations of vectors in a vector space or subspace in terms of a set $\mathcal{S}$. Assuming the set $\mathcal{S}$ spans the space, in order to have unique representations the vectors in $\mathcal{S}$ must be what we call linearly independent.

> **Definition 1.5:** A set of vectors $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ in a vector space V is **linearly independent** if the only scalars $c_1, c_2, ..., c_k$ for which
>
> $$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k = \mathbf{0} \tag{1}$$
>
> are $c_1 = c_2 = \cdots = c_k = 0$. If there are scalars $c_1, c_2, ..., c_k$, not all of which are zero, such that (1) is true, then the set $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ is **linearly dependent**.

Note that equation (1) is *always* true for $c_1 = c_2 = \cdots = c_n = 0$. When working with (1) we will always be led to a system of equations. If that system has more than one (so infinitely many) solutions, the vectors u_1, u_2, ..., u_n are linearly dependent. If $c_1 = c_2 = \cdots = c_n = 0$ is the only solution, then the vectors are linearly independent.

Now suppose that a set of vectors is linearly dependent. Then by the above, statement (1) is true for some scalars $c_1, c_2, ..., c_k$ with at least one c_i being non-zero. Therefore we can isolate the term $c_i\mathbf{u}_i$ on one side of the equation and then divide both sides by c_i (this is where we need it to be non-zero) to obtain $\mathbf{u}_i$ as a linear combination of all the other vectors in $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$. This is expressed in the following theorem.

Theorem 1.6: A set of vectors $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ in a vector space V is linearly dependent if and only if at least one of the vectors can be expressed as a linear combination of the others.

◇ **Example 1.4(e):** Determine whether the matrices $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ and $\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ in M_{22} are linearly independent or dependent. If they are linearly dependent, give one of them as a linear combination of the others.

Equation (1) from the definition in this case becomes

$$c_1\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + c_2\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & c_1 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ c_2 & 0 \end{bmatrix} = \begin{bmatrix} 0 & c_1 \\ c_2 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

which implies that only $c_1 = 0$ and $c_2 = 0$ makes equation (1) true. Therefore the matrices are linearly independent.

◇ **Example 1.4(f):** Determine whether the polynomials $-1 + x + 2x^2$, $1 + x + x^2$ and $-3 + x + 3x^2$ are linearly independent or dependent. If they are linearly dependent, give one of them as a linear combination of the others.

Here we must solve

$$c_1(-1 + x + 2x^2) + c_2(1 + x + x^2) + c_3(-3 + x + 3x^2) = 0. \tag{2}$$

(2) leads to the system of equations and reduced matrix below:

$$\begin{aligned} -c_1 + c_2 - 3c_3 &= 0 \\ c_1 + c_2 + \ \ c_3 &= 0 \\ 2c_1 + c_2 + 3c_3 &= 0 \end{aligned} \qquad\qquad \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Here c_3 is a free variable (meaning it can have any value). If we let it equal one, then we get $c_2 = 1$ and $c_1 = -2$. Thus the linear combination

$$-2(-1 + x + 2x^2) + (1 + x + x^2) + (-3 + x + 3x^2) \tag{3}$$

is equal to zero, so the three polynomials are linearly *dependent*. setting (3) equal to zero (which it is) and solving for the second polynomial gives us

$$1 + x + x^2 = 2(-1 + x + 2x^2) + (-3 + x + 3x^2).$$

This gives the second polynomial as a linear combination of the first and third.

In many cases where a set of vectors is linearly dependent, each of them can be written as a linear combination of the others. In the exercises you will give each of the other two polynomials from the previous example as a linear combination of the other two.

◇ **Example 1.4(g):** Determine whether the vectors $\begin{bmatrix} 1 \\ -2 \end{bmatrix}$ and $\begin{bmatrix} -3 \\ 6 \end{bmatrix}$ in $\mathbb{R}^2$ are linearly independent or dependent. If they are linearly dependent, give one of them as a linear combination of the other.

It is not hard to see that

$$3\begin{bmatrix} 1 \\ -2 \end{bmatrix} + \begin{bmatrix} -3 \\ 6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

so the vectors are linearly dependent. Moving the first term of the above equation to the other side gives us

$$\begin{bmatrix} -3 \\ 6 \end{bmatrix} = 3\begin{bmatrix} 1 \\ -2 \end{bmatrix},$$

which is the second vector as a (one term) linear combination of the first. Note that we also have

$$\begin{bmatrix} 1 \\ -2 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} -3 \\ 6 \end{bmatrix}.$$

The last example illustrates the following: Suppose that we have two vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ that are linearly dependent. By the definition there must then exist scalars c_1 and c_2, not both zero, such that $c_1\mathbf{u}_1 + c_2\mathbf{u}_2 = \mathbf{0}$. But if one of the scalars was zero, then the other would have to be as well, so they must *both* be non-zero. We can then get $c_1\mathbf{u}_1 = -c_2\mathbf{u}_2$ and divide both sides by either c_1 or $-c_2$ to get either of $\mathbf{u}_1$ or $\mathbf{u}_2$ as a scalar multiple of the other. This proves the following:

> **Theorem 1.7:** Two vectors are linearly dependent if and only if they are scalar multiples of each other. Conversely, two vectors are linearly independent if and only if they are not scalar multiples of each other.

WARNING: When considering a set of more than two vectors,

- if any pair of them are scalar multiples of each other, then the set as a whole is linearly dependent, *BUT*
- if no pairs of them are scalar multiples of each other, the vectors may still be linearly dependent.

This last point is illustrated by the three polynomials of Example 1.4(f).

The following definition seems a bit strange, but its purpose is to allow us the have the concepts of linear independence and dependence for infinite sets.

> **Definition 1.8:** A set S of vectors in a vector space V is **linearly dependent** if it contains finitely many linearly dependent vectors. S is **linearly independent** if *every* finite subset of S is linearly independent.

We conclude with a couple comments about language. Any set of just one vector is linearly independent, but that is not a particularly useful fact. Because of this, it rarely or never makes sense to say that just one vector is linearly independent (or dependent, since it can't be!). We will say that some vectors (more than one) are linearly independent or dependent, or that a set is linearly independent or dependent. Also, as there is no other kind of independence or dependence when working with vectors, we'll sometimes drop the word "linearly."

Section 1.4 Exercises

1. Follow a process like that of Example 1.4(c) to determine whether the set $\mathcal{S}$ below spans M_{22}.

$$\mathcal{S}=\left\{\begin{bmatrix}1&2\\3&4\end{bmatrix},\begin{bmatrix}1&1\\1&1\end{bmatrix},\begin{bmatrix}1&0\\0&1\end{bmatrix},\begin{bmatrix}3&3\\4&6\end{bmatrix}\right\}$$

2. In Example 1.4(b) it was (easily) shown that any polynomial in $\mathscr{P}_3$ can easily be represented as a linear combination of the first four elements of the set $\mathcal{S}=\{1,x,x^2,x^3,x+x^3\}$. For example,

$$5x^3-x^2+3x+2=5(x^3)+(-1)(x^2)+3(x)+2(1).$$

   (a) Give the representation of the same polynomial in terms of 1, x, x^2 and $x+x^3$.

   (b) Give yet another representation of the same polynomial in terms of 1, x^2, x^3 and $x+x^3$.

   (c) Parts (a) and (b) illustrates the problem with a set that spans a given space or subspace - each vector in in the space or subspace may not have a *unique* representation in terms of the spanning set. The problem here is that the given set $\mathcal{S}$ is not a linearly independent set. Find nonzero values of some of c_1, c_2, c_3, c_4 and c_5 for which equation (1) of Definition 1.5 holds.

   (d) Demonstrate that Theorem 1.6 holds for the set $\mathcal{S}=\{1,x,x^2,x^3,x+x^3\}$. *You should be able to do this without solving a system of equations.*

3. Determine whether the three polynomials $p_1(x)=x^2+2x+5$, $p_2(x)=x^2-2x+1$ and $p_3(x)=2x^2+x+4$ are linearly independent or dependent. If they are dependent, give a linear combination of them that is equal to the zero polynomial and give one of them as a linear combination of the others.

4. Determine whether the three polynomials $p(x)=2x^2-x+1$, $q(x)=x^2+3x+2$ and $r(x)=x^2+10x+5$ are linearly independent or dependent. If they are dependent, give a linear combination of them that is equal to the zero polynomial and give one of them as a linear combination of the others.

5. Determine whether the matrices below are linearly independent or dependent. If they are dependent, give a linear combination of them that is equal to the zero matrix and give one of them as a linear combination of the others.

$$A_1=\begin{bmatrix}5&7\\5&-10\end{bmatrix},\qquad A_2=\begin{bmatrix}1&2\\3&-4\end{bmatrix},\qquad A_3=\begin{bmatrix}0&3\\1&2\end{bmatrix},\qquad A_4=\begin{bmatrix}1&2\\0&0\end{bmatrix}$$

6. Determine whether the matrices below are linearly independent or dependent. If they are dependent, give a linear combination of them that is equal to the zero matrix and give one of them as a linear combination of the others.

$$B_1=\begin{bmatrix}1&2\\3&1\end{bmatrix},\qquad B_2=\begin{bmatrix}1&1\\1&1\end{bmatrix},\qquad B_3=\begin{bmatrix}2&1\\4&2\end{bmatrix}$$

7. For this exercise you will be working with the vector space $\mathscr{F}$. You will not be able to use matrix methods - you'll have to simply set up the standard equation and determine whether the desired scalars exist. One of the two following sets of vectors is linearly dependent. For the set that is, give **each function** as a linear combination of the other two.

   - $f(x)=3$, $g(x)=-\sin^2 x$, $h(x)=5\cos^2 x$
   - $f(x)=x$, $g(x)=\sin x$, $h(x)=e^x$

8. For the previous exercise you should suspect that one of the sets of functions is linearly independent, but you can't use matrix methods in the way that you have previously to show that. Instead you have to do the following.

   (a) Set up the standard equation for determining dependence or independence. It will have three unknown scalars.

   (b) The equation from (a) must hold for *all* values of x. Choose three values of x and substitute each into your equation from (a), evaluating the functions to three places past the decimal. This will give you three equations in three unknowns (the scalars). (Write them down, of course!)

   (c) Now you *have* reduced this to a problem that *can* be solved by matrix methods. Try solving it and show/tell what happens.

9. See Example 1.4(f) and the comment following it. Give

   (a) $-1+x+2x^2$ as a linear combination of $1+x+x^2$ and $-3+x+3x^2$.

   (b) $-3+x+3x^2$ as a linear combination of $-1+x+2x^2$ and $1+x+x^2$.

   You should use expression (3) of Example 1.4(f) to help you obtain your answers.

10. Determine whether each of the following pairs of vectors in $\mathbb{R}^2$ are linearly independent or dependent. If they are dependent, give each as a scalar multiple of the other (see Theorem 1.7).

   (a) $\begin{bmatrix} -6 \\ 4 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$ (b) $\begin{bmatrix} -6 \\ 4 \end{bmatrix}$ and $\begin{bmatrix} 9 \\ -6 \end{bmatrix}$

11. Three sets of vectors, in the spaces $\mathbb{R}^2$, $\mathscr{P}_2$ and M_{22}, are shown below. You should be able to determine by inspection that two of them are linearly dependent (one is harder to see than the other). Test the third to see if it is linearly independent or dependent.

$$\mathcal{S}_1 = \left\{ \begin{bmatrix} -3 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 \\ 7 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 6 \\ 2 \end{bmatrix} \right\} \qquad \mathcal{S}_2 = \{x^2-5x+2, 3x-1, 7\}$$

$$\mathcal{S}_3 = \left\{ \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 3 & 6 \\ 9 & 12 \end{bmatrix} \right\}$$

12. Give a polynomial in $\mathscr{P}_2$ that is not in the span of the set $\mathcal{S} = \{-1+x+2x^2, 1+x+x^2, -3+x+3x^2\}$ of Example 1.4(c) and demonstrate that it is not in the span of the set.

13. Does the set $\mathcal{S} = \{1-2x+x^2, -1+x-x^2, 2+x-x^2\}$ span $\mathscr{P}_2$? If not, give a polynomial in $\mathscr{P}_2$ that cannot be obtained as a linear combination of elements of $\mathcal{S}$. Then, give any additional polynomials needed to span $\mathscr{P}_2$.

1.5 Bases of Spaces and Subspaces

Performance Criteria:

1. (i) Determine whether a set of vectors is a basis for a vector space; if not, tell why.
   (j) Give the representation of a vector with respect to a given basis, and give its coordinate vector in the appropriate $\mathbb{R}^n$.
   (k) Extend a set to a basis or reduce a spanning set to a basis.

Although $\mathbb{R}^2$ and $\mathbb{R}^3$ both contain infinitely many points, $\mathbb{R}^3$ is in some sense "bigger" than $\mathbb{R}^2$. This idea is made more precise by the fact that all vectors in $\mathbb{R}^2$ can be obtained by linear combinations of two non-parallel vectors in $\mathbb{R}^2$, but in $\mathbb{R}^3$ it takes *three vectors not in the same plane* to obtain all vectors in $\mathbb{R}^3$ as linear combinations of those vectors. This thinking leads us directly to the concepts of a basis for a vector space and the dimension of a vector space. A basis for a vector space V is a "minimal" set of vectors for which the span (set of all linear combinations) is all of V. Such a set must be linearly independent.

Definition 1.9: A subset $\mathcal{B}$ of a vector space V is a **basis** for V if

1) $\mathcal{B}$ spans V and

2) $\mathcal{B}$ is linearly independent

◇ **Example 1.5(a):** The vectors $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ are a basis for $\mathbb{R}^2$. The computation $\begin{bmatrix} a \\ b \end{bmatrix} = a\begin{bmatrix} 1 \\ 0 \end{bmatrix} + b\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ shows that they span $\mathbb{R}^2$, and Theorem 1.7 tells us that they are linearly independent. This basis is called the **standard basis** for $\mathbb{R}^2$.

◇ **Example 1.5(b):** The vectors $\mathbf{u}_1 = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$ and $\mathbf{u}_2 = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$ can be shown to span $\mathbb{R}^2$, and Theorem 1.7 again indicates that they are linearly independent. Therefore $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2\}$ is a basis for $\mathbb{R}^2$.

The above examples indicate that *a vector space can have more than one basis!*

◇ **Example 1.5(c):** The polynomials $p_1(x) = x^2 - 7$ and $p_2 = x^2 + 1$ are not a basis for $\mathscr{P}_2$ because they don't span $\mathscr{P}_2$. It is easily seen that the polynomial $p_3(x) = 5x$ in $\mathscr{P}_2$ cannot be written as a linear combination of p_1 and p_2.

◇ **Example 1.5(d):** The polynomials $p_1(x) = 2x + 1$ and $p_2 = x - 3$ and $p_3(x) = 5x$ are not a basis for $\mathscr{P}_1$. They span $\mathscr{P}_1$, but they can be shown to be linearly dependent.

Later we will see that there are some other ways to determine whether a set is a basis for a given space without actually checking to see if the set spans the space and is a linearly independent set. Before going on, we should mention why the concept of a basis is so important. Our goal when working in a particular vector space or subspace is to find a "small" set of vectors out of which all vectors of interest can be formed as linear combinations. That is not all that we want, however. When we go to represent a vector in the space of interest as a linear combination of vectors in the "small" set, *we want there to be only one way to do it.* The following tells us that a basis meets that requirement.

Theorem 1.10: Let V be a vector space and let $\mathcal{B}$ be a basis for V. For every vector $\mathbf{v}$ in V, there is *exactly one* way to write $\mathbf{v}$ as a linear combination of the basis vectors in $\mathcal{B}$.

The following theorem establishes a relationship between a vector space with a basis containing n elements and the Euclidean space $\mathbb{R}^n$.

Definition 1.11: Let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ be a basis for a vector space V. Let $\mathbf{v}$ be a vector in V, and write $\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$. Then $c_1, c_2, ..., c_n$ are called the **coordinates of $\mathbf{v}$ with respect to $\mathcal{B}$**, and the column vector

$$[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$$

is called the **coordinate vector of $\mathbf{v}$ with respect to the basis $\mathcal{B}$.**

A vector space for which the scalars come from the field $\mathbb{R}$ is called a **real vector space**, and if the scalars for a vector space come from $\mathbb{C}$ the vector space is called a **complex vector space**. The coordinate vector of a vector in a real vector space is an element of some $\mathbb{R}^n$ and the coordinate vector of a vector in a complex vector space is an element of $\mathbb{C}^n$ for some value of n.

The following theorem can be used to easily rule out certain sets as bases for a vector space.

Theorem 1.12: Let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ be a basis for a vector space V. Then

a) any set of more than n vectors in V must be linearly dependent

b) any set of fewer than n vectors in V cannot span V

◇ **Example 1.5(e):** The set

$$\mathcal{S}_1 = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right\}$$

is easily shown to span M_{22} and to be linearly independent, so it is a basis for M_{22}. (In fact, it is the standard basis for that space.) The set

$$\mathcal{S}_2 = \left\{ \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} -3 & 4 \\ 1 & 7 \end{bmatrix} \right\}$$

has fewer than four elements, so it cannot be a basis for M_{22}. Note that this set *IS* linearly independent - the problem is that they don't span M_{22}.

◇ **Example 1.5(f):** The set $\mathcal{B} = \{1, x, x^2, x^3\}$ is the standard basis for $\mathscr{P}_3$. The set $\mathcal{S}_2 = \{1, x, x^2, x^3, x + x^3\}$ is not a basis because it contains more polynomials that the standard basis so it is a linearly dependent set.

Dimension of a Vector Space

We all have an intuitive sense of what the dimensions of the vector spaces $\mathbb{R}^n$ are, but we need to define it formally if we are going to use that language. We want to say that it is the number of vectors in a basis, but that could be ambiguous because a vector space can have many bases. The following theorem resolves that problem.

Theorem 1.13: If a vector space V has a basis with n vectors, then every basis for V has exactly n vectors.

Now we can make the definition we want:

Definition 1.14: If a basis for a vector space V consists of n vectors, we say that V has dimension n, or that it is an n-dimensional vector space

There are a couple special cases which we define as follows:

- The dimension of the zero vector space $\{\mathbf{0}\}$ is defined to be zero.
- A vector space that has no finite basis is called **infinite-dimensional**.

The following theorem allows us to determine whether or not a set is a basis for a vector space without checking both conditions of the definition, *assuming we know the dimension of the space.*

Theorem 1.15: Let V be a vector space with dim $V = n$. Then

a) any linearly independent set of vectors in V contains at most n vectors

b) any spanning set of V contains at least n vectors

c) any linearly independent set of exactly n vectors in V is a basis for V

d) any spanning set for V containing exactly n vectors is a basis for V

e) any linearly independent set in V can be extended to a basis in V

d) any spanning set for V can be reduced to a basis for V

Theorem 1.16: Let W be a subspace of a finite-dimensional vector space V. Then

a) W is finite-dimensional and $\dim(W) \leq \dim(V)$

b) $\dim(W) = \dim(V)$ if and only if $W = V$

Before continuing, let's make an important point: Other than the trivial subspace, a subspace of a vector space whose field is the real or complex numbers always contains infinitely many vectors. However, *a basis for any subspace of a finite-dimensional space contains finitely many vectors.*

Section 1.5 Exercises

1. For each of the following sets in the given vector space or subspace, determine whether the set is a basis. If it is not, tell why not. Answer with brief sentences like "$\mathcal{S}$ is a basis" or "$\mathcal{S}$ is not a basis because ..."

   (a) The set $\mathcal{S} = \{x^2, x^2+x, x^2+x+1\}$ in $\mathscr{P}_2$.

   (b) The set $\mathcal{S}\left\{\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} -3 & 4 \\ 1 & 7 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right\}$ in M_{22}.

   (c) The set $\mathcal{S} = \{x^2-2x+1, -x^2+2x-1, x^2+4x-5\}$ in $\mathscr{P}_2$.

   (d) The set $\mathcal{S} = \left\{\begin{bmatrix} -3 \\ 0 \\ 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \\ -7 \\ -2 \end{bmatrix}, \begin{bmatrix} 9 \\ 10 \\ 3 \\ 6 \end{bmatrix}\right\}$ in $\mathbb{R}^4$.

   (e) The set $\mathcal{S} = \left\{\begin{bmatrix} -3 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \\ -7 \end{bmatrix}, \begin{bmatrix} 9 \\ 10 \\ 3 \end{bmatrix}\right\}$ in $\mathbb{R}^3$.

2. (a) What is the dimension of $\mathbb{R}^n$? (b) What is the dimension of $\mathscr{P}_n$?

   (c) What is the dimension of M_{mn}?

3. (a) There is an obvious basis for $\mathscr{P}$, the set of polynomials of all degrees. Give it, using ... "notation."

   (b) $\mathcal{C} = \{1, 1+x, 1+x+x^2, 1+x+x^2+x^3, ...\}$ is also a basis for $\mathscr{P}$. Give the representation of

   $$p(x) = 4x^3 - x^2 + 5x + 2$$

   in terms of $\mathcal{C}$. This means to write your answer as

   $$4x^3 - x^2 + 5x + 2 = c_1(1) + c_2(1+x) + \cdots.$$

   You should be able to do this without solving a system of equations - see if you can figure out how.

   (c) Give the coordinates $[p(x)]_{\mathcal{C}}$ of $p(x)$ with respect to $\mathcal{C}$.

   (d) What is the dimension of $\mathscr{P}$?

4. Earlier we determined that the set of all upper triangular matrices in M_{22} is a subspace of M_{22}.

   (a) Give a basis $\mathcal{B}_1$ for that subspace. Make it so that all the entries of every matrix in the basis are either one or zero.

   (b) Give another basis $\mathcal{B}_2$ for the subspace, also consisting of matrices only containing ones and zeros.

   (c) Give yet another basis $\mathcal{B}_3$ for the subspace, also consisting of matrices only containing ones and zeros.

   (d) Give the coordinate vectors of the matrix $A = \begin{bmatrix} -4 & 1 \\ 0 & 2 \end{bmatrix}$ with respect to each of the three bases you have given. You should label each answer as $[A]_{\mathcal{B}_k}$, for $k = 1, 2, 3$.

   (e) What is the dimension of the set of upper triangular matrices in M_{22}?

5. We also determined that the set of all symmetric matrices in M_{22} is a subspace of M_{22}.

   (a) Find a basis $\mathcal{C}_1$ for that subspace.

(b) Find another basis $\mathcal{C}_2$ for that subspace.

(c) Find one more basis $\mathcal{C}_3$ for that subspace.

(d) Give the coordinate vectors for the vector $B = \begin{bmatrix} 3 & -2 \\ -2 & 1 \end{bmatrix}$ with respect to each of the three bases you have given.

(e) What is the dimension of the subspace?

6. Consider the set $\{1+x, 1+x^2, x+x^2, 1+x+x^2\}$ of polynomials in $\mathcal{P}_2$.

   (a) **Using a sentence or two**, give a brief explanation of why this set must be linearly dependent. Be sure to use the word *dimension* in your answer.

   (b) Give one of the polynomials as a linear combination of the others.

7. Look at Exercises 18, 20, 22, 24 of Section 6.2 (page 461). For each that *IS NOT* a basis, write a short sentence or two telling why it is not.

8. The set $\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \right\}$ is a basis for M_{22}. Find the coordinate vector for the matrix $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ with respect to $\mathcal{B}$. Denote your answer by the correct notation $[A]_{\mathcal{B}}$.

1.6 Change of Basis

Performance Criteria:

1. (l) For a vector space V with bases $\mathcal{B}$ and $\mathcal{C}$, given the coordinate vector of a vector $\mathbf{v} \in V$ with respect to $\mathcal{B}$, find the coordinate vector of v with respect to c.

The following example will result in some coordinate vectors that we'll use to illustrate the main idea of this section.

◇ **Example 1.6(a):** Find the coordinate vectors of $p(x) = 2x^2 - 5x + 1$ with respect to the two bases $\mathcal{B} = \{x^2 + x, x^2 + 1, x + 1\}$ and $\mathcal{C} = \{x^2, x^2 + x, x^2 + x + 1\}$.

We begin by writing $p(x)$ as linear combinations of the basis elements of each basis and regrouping:

$$2x^2 - 5x + 1 = b_1(x^2 + x) + b_2(x^2 + 1) + b_3(x + 1) = (b_1 + b_2)x^2 + (b_1 + b_3)x + (b_2 + b_3)$$

$$2x^2 - 5x + 1 = c_1(x^2) + c_2(x^2 + x) + c_3(x^2 + x + 1) = (c_1 + c_2 + c_3)x^2 + (c_2 + c_3)x + c_3$$

Setting like coefficients equal gives us the systems

$$\begin{array}{rcr} b_1 + b_2 & = & 2 \\ b_1 + b_3 & = & -5 \\ b_2 + b_3 & = & 1 \end{array} \qquad\qquad \begin{array}{rcr} c_1 + c_2 + c_3 & = & 2 \\ c_2 + c_3 & = & -5 \\ c_3 & = & 1 \end{array}$$

Solving these we obtain

$$[p(x)]_{\mathcal{B}} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} -2 \\ 4 \\ -3 \end{bmatrix} \qquad \text{and} \qquad [p(x)]_{\mathcal{C}} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 7 \\ -6 \\ 1 \end{bmatrix}$$

Suppose that for some vector $\mathbf{u}$ in an n-dimensional vector space V we have a coordinate vector $[\mathbf{u}]_{\mathcal{B}}$ with respect to a basis $\mathcal{B}$, and we wish to obtain the coordinate vector $[\mathbf{u}]_{\mathcal{C}}$ with respect to another basis $\mathcal{C}$. Given that we are trying to transform one vector into another, it would seem reasonable that there is an $n \times n$ matrix A such that $A[\mathbf{u}]_{\mathcal{B}} = [\mathbf{u}]_{\mathcal{C}}$. The key to figuring out how to build the matrix A is the following:

Linear Combination Form of a Matrix Times a Vector

The product of a matrix A and a vector $\mathbf{x}$ is a linear combination of the columns of A, with the scalars being the corresponding components of $\mathbf{x}$. Letting $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n$ denote the columns of A,

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

Let's let $[\mathbf{u}]_{\mathcal{B}} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$. Applying the the result in the box to $A[\mathbf{u}]_{\mathcal{B}} = [\mathbf{u}]_{\mathcal{C}}$ we have

$$A[\mathbf{u}]_{\mathcal{B}} = b_1\,\mathbf{a}_1 + b_2\,\mathbf{a}_2 + \cdots + b_n\,\mathbf{a}_n = [\mathbf{u}]_{\mathcal{C}} \tag{1}$$

for any vector $\mathbf{u} \in V$. Letting $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$, we note that $\mathbf{v}_1 = 1\,\mathbf{v}_1 + 0\,\mathbf{v}_2 + \cdots + 0\,\mathbf{v}_n$, so $[\mathbf{v}_1]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$.

Replacing $\mathbf{u}$ in (1) with $\mathbf{v}_1$, we have $b_1 = 1,\, b_2 = b_3 = \cdots = b_n = 0$ and

$$A[\mathbf{v}_1]_{\mathcal{B}} = 1\,\mathbf{a}_1 + 0\,\mathbf{a}_2 + \cdots + 0\,\mathbf{a}_n = [\mathbf{v}_1]_{\mathcal{C}}.$$

Thus $\mathbf{a}_1 = [\mathbf{v}_1]_{\mathcal{C}}$ and we would find that, similarly, $\mathbf{a}_k = [\mathbf{v}_k]_{\mathcal{C}}$ for $k = 1, 2, 3, ..., n$. We have proved the following:

Theorem 1.17: Let V be an n-dimensional vector space with bases $\mathcal{B}$ and $\mathcal{C}$, where $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$. The matrix $[P]_{\mathcal{B},\mathcal{C}}$ whose columns are the vectors $[\mathbf{v}_1]_{\mathcal{C}}$, $[\mathbf{v}_2]_{\mathcal{C}}$, $\cdots$, $[\mathbf{v}_n]_{\mathcal{C}}$ is the **change of basis matrix** from $\mathcal{B}$ to $\mathcal{C}$. That is, for any $\mathbf{u} \in V$

$$[P]_{\mathcal{B},\mathcal{C}}[\mathbf{u}]_{\mathcal{B}} = [\mathbf{u}]_{\mathcal{C}}.$$

◇ **Example 1.6(b):** Find the change of basis matrix with respect to the two bases $\mathcal{B} = \{x^2+x, x^2+1, x+1\}$ and $\mathcal{C} = \{x^2, x^2+x, x^2+x+1\}$ for $\mathscr{P}_2$. Then apply it to the coordinate vector of $p(x) = 2x^2 - 5x + 1$ with respect to $\mathcal{B}$ found in Example 1.6(a) to find $[2x^2 - 5x + 1]_{\mathcal{C}}$.

We need to find the coordinate vector with respect to $\mathcal{C}$ for each element if $\mathcal{B}$. We can see that the first element of $\mathcal{B}$ is the same as the second element of $\mathcal{C}$, so $[x^2 + x]_{\mathcal{C}} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$. Then we set

$$x^2 + 1 = c_1(x^2) + c_2(x^2 + x) + c_3(x^2 + x + 1) = (c_1 + c_2 + c_3)x^2 + (c_2 + c_3)x + c_3,$$

from which we can see that $c_3 = 1$, so $c_2 = -1$ and $c_1 = 1$. Thus $[x^2 + 1]_{\mathcal{C}} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$. The same process gives us $[x + 1]_{\mathcal{C}} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$, so we now have $[P]_{\mathcal{B},\mathcal{C}} = \begin{bmatrix} 0 & 1 & -1 \\ 1 & -1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$ and

$$[2x^2 - 5x + 1]_{\mathcal{C}} = [P]_{\mathcal{B},\mathcal{C}}[2x^2 - 5x + 1]_{\mathcal{B}} = \begin{bmatrix} 0 & 1 & -1 \\ 1 & -1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ 4 \\ -3 \end{bmatrix} = \begin{bmatrix} 7 \\ -6 \\ 1 \end{bmatrix}.$$

The last result agrees with the coordinate vector $[2x^2 - 5x + 1]_{\mathcal{C}}$ found in Example 1.6(a).

It would seem that if the matrix $[P]_{\mathcal{B},\mathcal{C}}$ "converts" a coordinate vector with respect to $\mathcal{B}$ to the corresponding coordinate vector with respect to $\mathcal{C}$, then the inverse of that matrix should convert a coordinate vector with respect to $\mathcal{C}$ to the corresponding coordinate vector with respect to $\mathcal{B}$. Of course not every matrix has an inverse, but in this case there is an inverse and it does what we would think it should.

Theorem 1.18: Let V be an n-dimensional vector space with bases $\mathcal{B}$ and $\mathcal{C}$ and change of basis matrix $[P]_{\mathcal{B},\mathcal{C}}$ from $\mathcal{B}$ to $\mathcal{C}$. Then $[P]_{\mathcal{B},\mathcal{C}}$ is invertible, and its inverse is the change of basis matrix from $\mathcal{C}$ to $\mathcal{B}$. That is,

$$[P]^{-1}_{\mathcal{B},\mathcal{C}} = [P]_{\mathcal{C},\mathcal{B}}.$$

Section 1.6 Exercises

1. The sets $\mathcal{B} = \{1, 1+x, 1+x^2\}$ and $\mathcal{C} = \{x, 1+x, x-x^2\}$ are bases for $\mathscr{P}_2$. Let $q(x) = 2 - 5x + 3x^2$.

   (a) Find $[q(x)]_{\mathcal{B}}$. This can be done by some arithmetic, so no work need be shown.

   (b) Find $[q(x)]_{\mathcal{C}}$, showing a linear combination equation and a system of equations that you are using to find the coordinate vector.

   (c) Find the change of basis matrix $[P]_{\mathcal{C},\mathcal{B}}$. You should be able to find the columns of the matrix in the same manner that you did part (a). Text your answer by seeing if it changes your answer to (b) to your answer to (a).

   (d) Use some sort of technology and Theorem 1.18 to find the matrix $[P]_{\mathcal{B},\mathcal{C}}$. Verify that it changes your answer to (a) to your answer to (b).

2. For this exercise let $\mathcal{B}$ be the standard basis $\mathcal{B} = \{1, x, x^2\}$ for $\mathscr{P}_2$, in that order, and let $\mathcal{C} = \{x, 1+x, x-x^2\}$. Find $[P]_{\mathcal{C},\mathcal{B}}$.

3. The sets $\mathcal{B} = \left\{ \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right\}$ and $\mathcal{C} = \left\{ \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \right\}$ are bases for the upper triangular matrices in M_{22}. Find the change of basis matrices $[P]_{\mathcal{B},\mathcal{C}}$ and $[P]_{\mathcal{C},\mathcal{B}}$.

4. Consider the three sets $\mathcal{B}_1 = \left\{ \begin{bmatrix} 1 \\ -2 \end{bmatrix}, \begin{bmatrix} -3 \\ 6 \end{bmatrix} \right\}$, $\mathcal{B}_2 = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right\}$, $\mathcal{B}_3 = \left\{ \begin{bmatrix} -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 \\ 1 \end{bmatrix} \right\}$ in $\mathbb{R}^2$.

   (a) One of the sets is not a basis for $\mathbb{R}^2$. Which one is it, and why? How can you tell?

   (b) Find both change of basis matrices between the two sets that are bases for $\mathbb{R}^2$.

1.7 Chapter 1 Exercises

We say that a function f is an **even function** if $f(-x) = f(x)$ for all values x in the domain of the function. A function is an **odd function** if $f(-x) = -f(x)$ for all x in the domain of f.

1. Determine whether each of the following functions is even, odd, or neither by computing $f(-x)$ and seeing if it is equal to $f(x)$, $-f(x)$ or neither.

$$f(x) = 8x^3 - 2x \ , \qquad g(x) = 2x^3 + x^2 + 5 \ , \qquad h(x) = x^2 + 5$$

2. Use a few test values to speculate as to whether the sine and cosine functions are even, odd, or neither. Show your results and give a sentence or two telling how you arrived at your conclusion.

3. Graph sine, cosine, and any of the functions from Exercise 1 that you found to be even or odd. What is special about the graphs of even functions? Of odd functions?

4. A natural question to ask is whether a function can be even and odd at the same time. Can you find such a function?

5. Suppose that a is some number greater than zero. Integrate all odd functions in Exercises 1 and 2 from $-a$ to a. State a conjecture about the integral of *any* odd function from $-a$ to a, and explain how the results from Exercise 3 support your conjecture.

6. Integrate all even functions from Exercises 1 and 2 from $-a$ to a, then from 0 to a. Explain your results in terms of Exercise 3 again.

7. Do some experimenting with your functions from Exercises 1 and 2 to make conjectures about

   (a) the product of two even functions,

   (b) the product of two odd functions,

   (c) the product of an even function and an odd function.

   How does this relate to even and odd numbers? *Think this through a bit - can you give some sort of explanation of why this might be the case?*

8. Let f be any function defined for all real numbers, and define two more functions

$$g(x) = \frac{f(x) + f(-x)}{2} \qquad \text{and} \qquad h(x) = \frac{f(x) - f(-x)}{2}. \tag{1}$$

   (a) Verify each of the following algebraically:

   - $g(x) + h(x) = f(x)$
   - g is even
   - h is odd

   (b) Go online and open an application called *Desmos* (just type that in the bar at the top and you should get it). Click *Start Graphing* and do the following:

   - Enter the function $f(x) = (x^3 + 4x + 2)e^{-x^2}$. Zoom in until the interesting part of the graph takes up most of the window. Is f odd? Even?
   - Enter g on the next line, exactly as given in (1) above. Does it appear to be even, as claimed/proven?
   - Enter h on the next line. Does it appear to be odd?
   - On the next line enter $y = g(x) + h(x)$. What happens?

2 Linear Transformations

Outcome/Performance Criteria:

2. Understand and work with linear transformations of general vector spaces.

   (a) Determine whether a given transformation is linear.

   (b) Given the action of a linear transformation on basis vectors, find the linear transformation of any vector.

   (c) Determine whether a given vector is in the kernel or range of a linear transformation. Describe the kernel and range of a linear transformation.

   (d) Determine whether a transformation is one-to-one; determine whether a transformation is onto.

   (e) Given formulas for two transformations, find the composition of them applied to a specific vector, find a formula for the composition.

   (f) Determine whether two transformations are inverses.

   (g) Determine whether a transformation is invertible, and find its inverse if it is.

   (h) Determine the matrix of a linear transformation $T : V \to W$ with respect to bases $\mathcal{B}$ and $\mathcal{C}$ of V and W, respectively.

   (i) Determine the matrix of a composition of linear transformations, determine the matrix of the inverse of a transformation.

When you were young your mathematical education consisted primarily of learning about numbers and their operations. Later you took algebra, followed by trigonometry and some calculus. All of those courses were about functions, which take in numbers and give out numbers. Numbers are the objects, functions are the things that act on those objects.

In the study of linear algebra, vectors are analogous to numbers; they are the objects that we work with. When studying linear algebra, the objects that take the place of functions are called transformations. Really, all that transformations are is functions that act on vectors rather than numbers. It is useful to have a different name for them, especially in the case that the "vectors" under consideration are really functions, like those in the vector space $\mathscr{F}$.

We are particularly interested in a kind of transformation called a *linear* transformation. We found in the last chapter that any "vector" in a finite dimensional vector space can in fact be represented by a true vector in some $\mathbb{R}^n$, called a coordinate vector. In this chapter we will see that if a transformation on a finite dimensional space is in fact linear, then its action can be defined by multiplication of coordinate vectors by a matrix.

2.1 Transformations and Linear Transformations

Performance Criteria:

2. (a) Determine whether a given transformation is linear.
   (b) Given the action of a linear transformation on basis vectors, find the linear transformation of any vector.

Any process (function) that takes one vector and creates from it another vector (not necessarily in the same vector space) is called a **transformation**. We are most interested in transformations that behave "nicely," as described below. *The study of linear transformations is really the "central core" of the subject of linear algebra.*

Definition 2.1: A **linear transformation** from a vector space V to a vector space W is a mapping $T : V \rightarrow W$ such that, for all $\mathbf{u}$ and $\mathbf{v}$ and all scalars c,

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

b) $T(c\mathbf{u}) = cT(\mathbf{u})$

It is worth taking a moment to think about what the above is saying. The first part says that if we add two vectors in V and transform the result to W, we could obtain the same final result by transforming the two vectors to W first, *then* adding them. The second part says that if we multiply by a scalar in V and then transform to W, the same result can be obtained by first transforming the vector to W and *then* multiplying by the same scalar.

◇ **Example 2.1(a):** Determine whether $T : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ defined by $T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_1 + x_2 \\ x_2 - x_3 \end{bmatrix}$ is linear. If it is, prove it in general; if it isn't, give a specific counterexample.

First let's try a specific scalar $c = 2$ and two specific vectors $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 4 \\ -5 \\ 6 \end{bmatrix}$. (I threw the negative in there just in case something funny happens when everything is positive.) Then

$$T\left(\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 4 \\ -5 \\ 6 \end{bmatrix}\right) = T\begin{bmatrix} 5 \\ -3 \\ 9 \end{bmatrix} = \begin{bmatrix} 2 \\ -12 \end{bmatrix}$$

and

$$T\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + T\begin{bmatrix} 4 \\ -5 \\ 6 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \end{bmatrix} + \begin{bmatrix} -1 \\ -11 \end{bmatrix} = \begin{bmatrix} 2 \\ -12 \end{bmatrix}$$

so the first condition of linearity *appears* to hold. Let's prove it in general. Let $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$ be arbitrary (that is, randomly selected) vectors in $\mathbb{R}^3$. Then

$$T(\mathbf{u}+\mathbf{v}) = T\left(\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}\right) = T\begin{bmatrix} u_1+v_1 \\ u_2+v_2 \\ u_3+v_3 \end{bmatrix} = \begin{bmatrix} u_1+v_1+u_2+v_2 \\ (u_2+v_2)-(u_3+v_3) \end{bmatrix} =$$

$$\begin{bmatrix} u_1+u_2+v_1+v_2 \\ (u_2-u_3)+(v_2-v_3) \end{bmatrix} = \begin{bmatrix} u_1+u_2 \\ u_2-u_3 \end{bmatrix} + \begin{bmatrix} v_1+v_2 \\ v_2-v_3 \end{bmatrix} = T\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} + T\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = T(\mathbf{u})+T(\mathbf{v})$$

This shows that the first condition of linearity holds in general. Let $\mathbf{u}$ again be arbitrary, along with the scalar c. Then

$$T(c\mathbf{u}) = T\left(c\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}\right) = T\begin{bmatrix} cu_1 \\ cu_2 \\ cu_3 \end{bmatrix} = \begin{bmatrix} cu_1+cu_2 \\ cu_2-cu_3 \end{bmatrix} =$$

$$\begin{bmatrix} c(u_1+u_2) \\ c(u_2-u_3) \end{bmatrix} = c\begin{bmatrix} u_1+u_2 \\ u_2-u_3 \end{bmatrix} = cT\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = cT(\mathbf{u})$$

so the second condition holds as well, and T is a linear transformation.

◇ **Example 2.1(b):** Define $T: M_{22} \to M_{22}$ by $T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$. Is T linear?

All that T does is switch two entries of a matrix, so it would seem we could either add, then switch, or switch then add. Same with multiplying by a scalar. Let's prove T is linear. Let $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ and $B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$ be in M_{22}. Then

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

Thus T satisfies the first condition of a linear transformation. Taking c to be any scalar and A as above,

$$T(cA) = T\left(c\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right) = T\left(\begin{bmatrix} ca_{11} & ca_{12} \\ ca_{21} & ca_{22} \end{bmatrix}\right) =$$

$$\begin{bmatrix} ca_{11} & ca_{21} \\ ca_{12} & ca_{22} \end{bmatrix} = c\begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix} = cT(A),$$

showing that T meets the second condition of a linear transformation. Thus T is in fact a linear transformation.

◇ **Example 2.1(c):** Define $T: \mathscr{P}_2 \to \mathbb{R}^2$ by $T(ax^2+bx+c) = \begin{bmatrix} ab \\ bc \end{bmatrix}$. Is T linear?

It is a bit harder to visualize how T would interact with addition in this case, so let's just try in on

$$p(x) = 3x^2-5x+1 \qquad \text{and} \qquad q(x) = x^2+2x+3.$$

We see that

$$T[p(x)+q(x)] = T[(3x^2-5x+1)+(x^2+2x+3)] = T(4x^2-3x+4) = \begin{bmatrix} -12 \\ -12 \end{bmatrix}$$

and

$$T[p(x)] + T[q(x)] = T(3x^2 - 5x + 1) + T(x^2 + 2x + 3) = \begin{bmatrix} -15 \\ -5 \end{bmatrix} + \begin{bmatrix} 2 \\ 6 \end{bmatrix} = \begin{bmatrix} -13 \\ 1 \end{bmatrix},$$

so $T[p(x) + q(x)] \neq T[p(x)] + T[q(x)]$ and T is not a linear transformation.

The facts given by the following theorem are things we would likely take for granted, but they can be proven from the definition.

> **Theorem 2.2:** Let $T : V \to W$ be a linear transformation. Then
>
> a) $T(\mathbf{0}) = \mathbf{0}$
>
> b) $T(-\mathbf{v}) = -T(\mathbf{v})$ for all $\mathbf{v}$ in V
>
> c) $T(\mathbf{u} - \mathbf{v}) = T(\mathbf{u}) - T(\mathbf{v})$

Together the two conditions in the definition of a linear transformation give us the fact that any linear combination of vectors can be formed in V and transformed to W, or all the vectors can first be transformed to vectors in W before the same linear combination is formed. Either way, the result is the same. The next theorem says this a bit more formally!

> **Theorem 2.3:** $T : V \to W$ is a linear transformation if and only if
>
> $$T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k) = c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2) + \cdots + c_kT(\mathbf{v}_k)$$
>
> for all $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k$ in V and all scalars $c_1, c_2, ..., c_k$.

Suppose, in particular, that the vectors $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k$ in the above theorem constitute a basis or V. Then every vector in V can be written as a unique linear combination of those vectors. If we have a linear transformation T and we know what it does to each of $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k$, then the above theorem says that we then know what T does to every vector in V.

◇ **Example 2.1(d):** The set $\mathcal{S}\left\{\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\right\}$ is a basis for M_{22}. Suppose that $T : M_{22} \to \mathscr{P}_2$ is a linear transformation such that

$$T\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\right) = 3x, \qquad T\left(\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right) = x^2 + x, \qquad T\left(\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}\right) = x - 2, \qquad T\left(\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\right) = 5.$$

Find $T\left(\begin{bmatrix} 3 & -1 \\ 5 & 2 \end{bmatrix}\right)$.

We can see that

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

Therefore, by Theorem 2.3,

$$\begin{aligned} T\left(\begin{bmatrix} 3 & -1 \\ 5 & 2 \end{bmatrix}\right) &= T\left(3\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + 5\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + 2\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\right) \\ &= 3T\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\right) - T\left(\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right) + 5T\left(\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}\right) + 2T\left(\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\right) \\ &= 3(3x) - (x^2 + x) + 5(x - 2) + 2(5) \\ &= -x^2 + 13x \end{aligned}$$

◇ **Example 2.1(e):** The set $\mathcal{B} = \{x^2+x, x^2+1, x+1\}$ is a basis for $\mathscr{P}_2$. Suppose that for the linear transformation $T : \mathscr{P}_2 \to \mathbb{R}^2$ we have

$$T(x^2+x) = \begin{bmatrix} -2 \\ 5 \end{bmatrix}, \qquad T(x^2+1) = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \qquad T(x+1) = \begin{bmatrix} 4 \\ 4 \end{bmatrix}.$$

Find $T(8x^2+x+3)$.

We must first find a representation of $8x^2+x+3$ in terms of the basis polynomials. To do that we set up the equation

$$c_1(x^2+x) + c_2(x^2+1) + c_3(x+1) = 8x^2+x+3,$$

which leads to the system of equations

$$\begin{aligned} c_1 + c_2 + \quad\;\; &= 8 \\ c_1 + \qquad c_3 &= 1 \\ c_2 + c_3 &= 3 \end{aligned}$$

The solution to the system is $c_1 = 3$, $c_2 = 5$ and $c_3 = -2$. Therefore

$$\begin{aligned} T(8x^2+x+3) &= T[3(x^2+x) + 5(x^2+1) - 2(x+1)] \\ &= 3\,T(x^2+x) + 5\,T(x^2+1) - 2\,T(x+1)] \\ &= 3\begin{bmatrix} -2 \\ 5 \end{bmatrix} + 5\begin{bmatrix} 1 \\ 3 \end{bmatrix} - 2\begin{bmatrix} 4 \\ 4 \end{bmatrix} \\ &= \begin{bmatrix} -6 \\ 15 \end{bmatrix} + \begin{bmatrix} 5 \\ 15 \end{bmatrix} - \begin{bmatrix} 8 \\ 8 \end{bmatrix} \\ &= \begin{bmatrix} -14 \\ 22 \end{bmatrix} \end{aligned}$$

Section 2.1 Exercises

1. Define $T : \mathscr{P}_2 \to \mathscr{P}_2$ by $T(ax^2+bx+c) = ax^2 + (b+c)x + (a+b+c)$. Is T linear?

2. Let $\mathscr{D}$ represent all the functions in $\mathscr{F}$ that are differentiable at all values of x. Define $T : \mathscr{D} \to \mathscr{F}$ by $[T(f)](x) = f'(x)$. Is T linear?

3. Define $T : M_{22} \to \mathbb{R}$ by $T(A) = \det(A)$. Is T linear?

4. Define $T : \mathscr{F} \to \mathscr{F}$ by $[T(f)](x) = f(x) + 3$. Is T linear?

5. Define $T : \mathscr{F} \to \mathscr{F}$ by $[T(f)](x) = 3f(x)$. Is T linear?

6. Define $T : \mathscr{C}[a,b] \to \mathbb{R}$ by $T(f) = \displaystyle\int_a^b f(x)\,dx$. Is T linear?

7. Recall that the transpose A^T of a matrix A is the matrix whose columns are the rows of A. Define $T : M_{23} \to M_{32}$ by $T(A) = A^T$. Is T linear?

8. The **trace** of a square matrix is the sum of the diagonal entries of the matrix. Define $T : M_{22} \to \mathbb{R}$ by $T(A) = trace(A)$. Is T linear?

9. Define $T : \mathscr{P}_2 \to \mathbb{R}$ by $T[p(x)] = p(1)$. Is T linear?

10. Define $T : \mathscr{P}_2 \to \mathbb{R}$ by $T[p(x)] = p(0)$. Is T linear?

11. Let $A = \begin{bmatrix} 3 & -1 \\ -4 & 0 \end{bmatrix}$ and define a transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T\mathbf{x} = A\mathbf{x}$. Prove that T is a linear transformation.

12. $\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \right\}$ is a basis for the upper triangular matrices in M_{22}, and $T : M_{22} \to \mathbb{R}^2$ is a linear transformation for which

$$T\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\right) = \begin{bmatrix} -1 \\ 5 \end{bmatrix}, \qquad T\left(\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \qquad T\left(\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\right) = \begin{bmatrix} 4 \\ -6 \end{bmatrix}.$$

Find $T\left(\begin{bmatrix} 7 & -4 \\ 0 & 3 \end{bmatrix}\right)$.

13. $\mathcal{C} = \{x^2 + x, x^2 + 1, x + 1\}$ is a basis for $\mathscr{P}_2$ and $S : \mathscr{P}_2 \to \mathscr{P}_2$ is a linear transformation for which

$$S(x^2 + x) = 3x + 1, \qquad S(x^2 + 1) = 7, \qquad S(x + 1) = x^2 - 3.$$

Find $S(2x^2 - 5x + 2)$.

14. $\mathcal{B} = \left\{ \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right\}$ is a basis for $\mathbb{R}^3$ and $T : \mathbb{R}^3 \to \mathbb{R}^3$ is a linear transformation for which

$$T\left(\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \qquad T\left(\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} -4 \\ 5 \\ -2 \end{bmatrix}, \qquad T\left(\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 5 \\ 3 \\ -1 \end{bmatrix}.$$

Find $T\left(\begin{bmatrix} 7 \\ 2 \\ -5 \end{bmatrix}\right)$.

For any n, the derivative $\dfrac{d^n}{dx^n}$ is a linear transformation from $\mathscr{C}^n$ to $\mathscr{C}^{n-1}$. We can form an n**th-order differential operator** by taking a linear combination of derivatives of order n and lower (including the "zeroth" derivative, which is just the function being operated on). Such differential operators are linear transformations, and we usually denote them by D. In your differential equations course you worked with a number of first- and second-order differential operators like $D = \dfrac{d^2}{dx^2} + 4\dfrac{d}{dx} + 3$, whose action on a function $y \in \mathscr{C}^2$ is defined by

$$D(y) = \left(\frac{d^2}{dx^2} + 4\frac{d}{dx} + 3\right)(y) = \frac{d^2y}{dx^2} + 4\frac{dy}{dx} + 3y. \tag{1}$$

15. For the differential operator (1), find $D(x^2 + 3x - 1)$, $D(e^{5x})$ and $D(\sin 2x)$.

16. When we attempt to solve the ODE (ordinary differential equation) $y'' + 4y' + 3y = 5\sin 3x$ we are looking for a function y for which $D(y) = 5\sin 3x$. Do the following to find y:

   (a) Assume that $y = A\sin 3x + B\cos 3x$ for some constants A and B. Apply D to this guess, being sure to use the chain rule!

   (b) Get your answer to (a) in the form $E\sin 3x + F\cos 3x$ (where E and F are expressions containing A and B) and set it equal to what we want it to be, $5\sin 3x$.

   (c) Equating like terms in your result from (b), we see that $E = 5$ and $F = 0$, giving two equations in the unknowns A and B solve for A and B, giving your answers as fractions.

   (d) Give the function y for which $D(y) = 5\sin 3x$.

17. Continue to use the operator D given in (1).

    (a) Letting $y = Ce^{rt}$ for some constants C and r, find $D(y)$.

    (b) Is(Are) there any value(s) of r for which $D(y) = 0$? If so, what are they?

    (c) Give three different solutions to the differential equation $y'' + 4y' + 3y = 0$.

18. Let $A = \begin{bmatrix} 3 & -1 \\ -4 & 0 \end{bmatrix}$ and define the transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T(\mathbf{x}) = A\mathbf{x}$. Prove that T is linear.

19. For this exercise T is the same as in the previous exercise.

    (a) Let $\mathbf{u} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $\mathbf{w} = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$. Find $T(\mathbf{u})$, $T(\mathbf{v})$ and $T(\mathbf{w})$.

    (b) Given a transformation T, any vector $\mathbf{x}$ for which $T(\mathbf{x}) = \lambda\mathbf{x}$ for some constant λ is called an **eigenvector** of T with **eigenvalue** λ. Are any of $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ eigenvectors? If so, what is the corresponding eigenvalue for each that is an eigenvector?

20. If D is a differential operator, any function f for which $D(f) = \lambda f$ for some constant λ is called an **eigenfunction** with **eigenvalue** λ.

    (a) Let D be the first derivative. Give an eigenfunction for that operator with eigenvalue -3.

    (b) Now let D be the second derivative. Find *two* eigenfunctions with eigenvalue four.

    (c) Again letting D be the second derivative, find two eigenfunctions with eigenvalue -4.

21. A very important ODE in many applications is $y'' + \lambda^2 y = 0$. Note that this can be rearranged to get $y'' = -\lambda^2 y$, which says that any y that is a solution to the differential equation is an eigenfunction with eigenvalue $-\lambda^2$.

    (a) Give the eigenfunctions of the second derivative with eigenvalue $-\lambda^2$.

    (b) (Challenge) Let $\mathcal{S}$ be the subset of $\mathscr{C}^2[0, 2\pi]$ consisting of functions $y = f(x)$ that have continuous second derivatives and for which $f(0) = f(2\pi) = 0$. It can be shown that this set is a subspace of $\mathscr{C}^2[0, 2\pi]$. Determine *ALL* eigenfunctions of the second derivative with eigenvalue $-\lambda^2$ that are in $\mathcal{S}$.

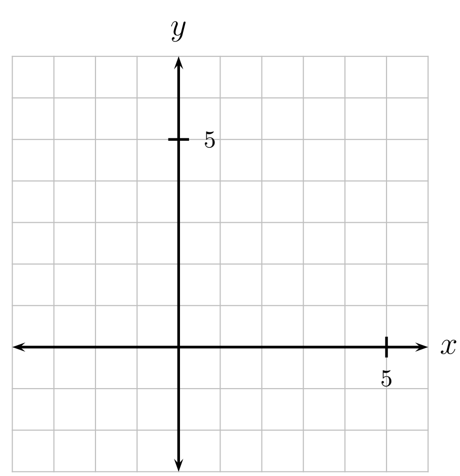
22. Let $B = \begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}$ and define $S : \mathbb{R}^2 \to \mathbb{R}^2$ by $S(\mathbf{x}) = B\mathbf{x}$. Let $\mathbf{u} = \begin{bmatrix} 4 \\ 0 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} -3 \\ 1 \end{bmatrix}$, $\mathbf{w} = \begin{bmatrix} 3 \\ 4.5 \end{bmatrix}$.

    (a) Find $B\mathbf{u}$, $B\mathbf{v}$ and $B\mathbf{w}$, with their components as decimals to the nearest tenth.

    (b) Plot and label $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$, $B\mathbf{u}$, $B\mathbf{v}$ and $B\mathbf{w}$ on the $\mathbb{R}^2$ coordinate grid to the right.

    (c) What does the matrix B seem to do to every vector?

    (d) The entries of B should look familiar to you. What is special about $\frac{1}{2}$ and $\frac{\sqrt{3}}{2}$?



2.2 Kernel and Range of a Linear Transformation

Performance Criteria:

2. (c) Determine whether a given vector is in the kernel or range of a linear transformation. Describe the kernel and range of a linear transformation.

   (d) Determine whether a transformation is one-to-one; determine whether a transformation is onto.

When working with transformations $T : \mathbb{R}^m \to \mathbb{R}^n$, you found that any *linear* transformation can be represented by multiplication by a matrix. At some point after that you were introduced to the concepts of the *null space* and *column space* of a matrix. In this section we present the analogous ideas for general vector spaces.

Definition 2.4: Let V and W be vector spaces, and let $T : V \to W$ be a transformation. We will call V the **domain** of T, and W is the **codomain** of T.

Definition 2.5: Let V and W be vector spaces, and let $T : V \to W$ be a *linear* transformation.

- The set of all vectors $\mathbf{v} \in V$ for which $T\mathbf{v} = \mathbf{0}$ is a subspace of V. It is called the **kernel** of T, and we will denote it by $\ker(T)$.
- The set of all vectors $\mathbf{w} \in W$ such that $\mathbf{w} = T\mathbf{v}$ for some $\mathbf{v} \in V$ is called the **range** of T. It is a subspace of W, and is denoted $\text{ran}(T)$.

It is worth making a few comments about the above:

- The kernel and range "belong to" the transformation, not the vector spaces V and W. If we had another linear transformation $S : V \to W$, it would most likely have a different kernel and range.
- The kernel of T is a subspace of V, and the range of T is a subspace of W. The kernel and range "live in different places."
- The fact that T is linear is essential to the kernel and range being subspaces.

Time for some examples!

◇ **Example 2.2(a):** $T : M_{22} \to \mathbb{R}^2$ defined by $T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} a+b \\ c+d \end{bmatrix}$ is linear. Describe its kernel and range and give the dimension of each.

It should be clear that $T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ if, and only if, $b = -a$ and $d = -c$. The kernel of T is therefore all matrices of the form

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

The two matrices $\begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}$ and $\begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix}$ are not scalar multiples of each other, so they must be linearly independent. Therefore the dimension of $\ker(T)$ is two.

Now suppose that we have any vector $\begin{bmatrix} a \\ b \end{bmatrix} \in \mathbb{R}^2$. Clearly $T\left(\begin{bmatrix} a & 0 \\ b & 0 \end{bmatrix}\right) = \begin{bmatrix} a \\ b \end{bmatrix}$, so the range of T is all of $\mathbb{R}^2$. Thus the dimension of $\text{ran}(T)$ is two.

◇ **Example 2.2(b):** $T : \mathscr{P}_1 \to \mathscr{P}_1$ defined by $T(ax + b) = 2bx - a$ is linear. Describe its kernel and range and give the dimension of each.

$T(ax + b) = 2bx - a = 0$ if, and only if, both a and b are zero. Therefore the kernel of T is only the zero polynomial. By definition, the dimension of the subspace consisting of only the zero vector is zero, so $\ker(T)$ has dimension zero.

Suppose that we take a random polynomial $cx + d$ in the codomain. If we consider the polynomial $-dx + \frac{1}{2}c$ in the domain we see that

$$T(-dx + \tfrac{1}{2}c) = 2(\tfrac{1}{2}c)x - (-d) = cx + d.$$

This shows that the range of T is all of $\mathscr{P}_1$, and it has dimension two.

◇ **Example 2.2(c):** $T : \mathscr{P}_2 \to \mathscr{P}_2$ defined by $T(ax^2 + bx + c) = ax^2 + (b+c)x + (a+b+c)$ is linear. Describe its kernel and range and give the dimension of each.

If $T(ax^2 + bx + c) = ax^2 + (b+c)x + (a+b+c) = 0$, then clearly $a = 0$ and $c = -b$. Thus the kernel of T is the set of all polynomials of the form $bx - b = b(x - 1)$. This set has dimension one ($x - 1$ is a basis).

The range of T is all polynomials of the form $ax^2 + (b+c)x + (a+b+c)$. If we let $b + c = d$, this is then the polynomials of the form $ax^2 + dx + (a+d) = a(x^2+1) + d(x+1)$. The range of T therefore has dimension two.

Definition 2.6: Let $T : V \to W$ be a linear transformation. The **nullity** of T is the dimension of the kernel of T, and the **rank** of T is the dimension of the range of T. They are denoted by $\text{nullity}(T)$ and $\text{rank}(T)$, respectively.

Examples 2.2(a),(b) and (c) illustrate the following important theorem, usually referred to as the **rank theorem**.

Theorem 2.7: Let $T : V \to W$ be a linear transformation. Then

$$\text{rank}(T) + \text{nullity}(T) = \dim(V),$$

where $\dim(V)$ is the dimension of V.

The next theorem can be useful in determining the rank of a transformation.

Theorem 2.8: Let $T : V \to W$ be a linear transformation and let $\mathcal{B} = \{\mathbf{v}_1, ..., \mathbf{v}_n\}$ be a spanning set for V. Then $T(\mathcal{B}) = \{T(\mathbf{v}_1), ..., T(\mathbf{v}_n)\}$ is a spanning set for the range of T.

Definition 2.9: Let $T : V \to W$ be a transformation. T is said to be **onto** if for every $\mathbf{w} \in W$ there is a $\mathbf{v} \in V$ such that $T(\mathbf{v}) = \mathbf{w}$.

◇ **Example 2.2(d):** Consider the transformation $T : \mathscr{P}_2 \to \mathscr{P}_2$ defined by $T(ax^2 + bx + c) = ax^2 + (b + c)x + (a + b + c)$ is linear. Is this transformation onto?

Consider the polynomial $x^2 + x + 1$ in the codomain $\mathscr{P}_2$, and suppose that there is a polynomial $p(x) = ax^2 + bx + c$ such that $T[p(x)] = x^2 + x + 1$. Then

$$T[p(x)] = T(ax^2 + bx + c) = ax^2 + (b + c)x + (a + b + c) = x^2 + x + 1,$$

resulting in the system of equations $a = 1$, $b + c = 1$ and $a + b + c = 1$. Clearly, adding the first two equations gives a result that contradicts the third equation, so this system has no solution. Therefore there is no polynomial $p(x)$ in the domain such that $T[p(x)] = x^2 + x + 1$, so T is not onto.

◇ **Example 2.2(e):** Define $T : \mathscr{P}_1 \to \mathbb{R}^2$ by $T(ax + b) = \begin{bmatrix} 3a - b \\ a + 2b \end{bmatrix}$. Is T onto?

Let $\begin{bmatrix} e \\ f \end{bmatrix}$ be given. We see that

$$T\left[\left(\frac{2e+f}{7}\right)x + \left(\frac{3f-e}{7}\right)\right] = \begin{bmatrix} 3\left(\frac{2e+f}{7}\right) - \left(\frac{3f-e}{7}\right) \\ \left(\frac{2e+f}{7}\right) + 2\left(\frac{3f-e}{7}\right) \end{bmatrix} = \begin{bmatrix} \dfrac{6e + 3f - 3f + e}{7} \\ \dfrac{2e + f + 6f - 2e}{7} \end{bmatrix} = \begin{bmatrix} e \\ f \end{bmatrix}.$$

This shows that if we are given a $\mathbf{w} \in \mathbb{R}^2$ we can find a polynomial $p(x) \in \mathscr{P}_1$ for which $T[p(x)] = \mathbf{w}$, so T is onto.

One might wonder how to even know whether a transformation is onto. The following theorem, which is essentially just a restatement of the definition, can be useful:

Theorem 2.10: Let $T : V \to W$ be linear transformation. T is onto if and only if $\operatorname{ran}(T) = W$.

If it is clear that the range of a transformation is all of the codomain, then we can tell if it is onto. Also, if we know what a typical element of the range looks like, then we can more easily construct examples showing that a transformation is not onto. In Example 2.2(c) we found that the range of the transformation from Example 2.2(d) was polynomials of the form $ax^2 + dx + (a + d)$. To get our counterexample we merely needed to find a polynomial in $\mathscr{P}_2$ such that the constant term was not the sum of the coefficients of x^2 and x.

Definition 2.11: Let $T : V \to W$ be a transformation. T is said to be **one-to-one** if $T\mathbf{u} = T\mathbf{v}$ implies $\mathbf{u} = \mathbf{v}$. Equivalently, $\mathbf{u} \neq \mathbf{v}$ implies $T\mathbf{u} \neq T\mathbf{v}$.

The second of the above two conditions is perhaps easier to use to get a feeling for whether a transformation is one-to-one and is easiest for presenting a counterexample for a transformation that is not one-to-one, but the first is generally easier for proving that a transformation is one-to-one. We'll see this in the following examples.

$\diamond$ **Example 2.2(f):** $T : \mathscr{P}_1 \to \mathscr{P}_1$ defined by $T(ax+b) = 2bx - a$ is linear. Is it one-to-one?

It should be fairly clear that if we have two different elements in the domain, their transformations should be two different elements in the codomain. Now suppose that $T(ax+b) = T(cx+d)$. Applying T to each side gives us $2bx - a = 2dx - c$ and equating coefficients gives us $2b = 2d$ and $-a = -c$. These imply that $b = d$ and $a = c$, so $ax + b = cx + d$. Therefore T is one-to-one.

$\diamond$ **Example 2.2(g):** $T : M_{22} \to \mathbb{R}^2$ defined by $T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} a+b \\ c+d \end{bmatrix}$ is linear. Is it one-to-one?

It seems like two different matrices in M_{22} could give us the same vector in $\mathbb{R}^2$, so T is probably not linear. In fact, if $A = \begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 1 \\ 0 & 2 \end{bmatrix}$, then $A \neq B$ but $T(A) = \begin{bmatrix} 1 \\ 2 \end{bmatrix} = T(B)$, so T is not one-to-one.

Now suppose that $T : V \to W$ is linear, $\ker(T) = \mathbf{0}$ and $T\mathbf{u} = T\mathbf{v}$. Subtracting $T\mathbf{v}$ from both sides gives us $T\mathbf{u} - T\mathbf{v} = \mathbf{0}$ and, because T is linear, $T(\mathbf{u} - \mathbf{v}) = \mathbf{0}$. but $\ker(T) = \mathbf{0}$, so $\mathbf{u} - \mathbf{v}$ must be zero, giving us $\mathbf{u} = \mathbf{v}$. This shows that if $\ker(T) = \mathbf{0}$, then T is one-to-one.

Conversely, suppose that (for the same linear transformation T) T is one-to-one, and let $\mathbf{v}$ be some vector in $\ker(T)$. Because $\mathbf{v} \in \ker(T)$, $T\mathbf{v} = \mathbf{0}$. By Theorem 2.2(a), $T(\mathbf{0}) = \mathbf{0}$ also, so $T\mathbf{v} = T(\mathbf{0})$. But because T is one-to one, we must have $\mathbf{v} = \mathbf{0}$, so the kernel of T consists of only the zero vector.

The above two paragraphs prove this:

> **Theorem 2.12:** A linear transformation $T : V \to W$ is one-to-one if and only if $\ker(T) = \mathbf{0}$.

Section 2.2 Exercises

1. Define $T : V \to W$, where $V = W = M_{22}$, by $T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix}$. T is linear.

   (a) Give a matrix A that is in the kernel of T, and a matrix B that is in the range of T.

   (b) Give the general form of a matrix in $\ker(T)$, then write it as a linear combination of specific matrices, as done in Example 2.2(a). What is the dimension of $\ker(T)$? Give a basis for $\ker(T)$.

   (c) Repeat part (b) for $\text{ran}(T)$.

   (d) What are $\dim(V)$, $\text{rank}(T)$ and $\text{nullity}(T)$? Is Theorem 2.7 satisfied?

   (e) Give two matrices A and B with $A \neq B$, but $T(A) = T(B)$. What does this tell us about T?

   (f) Give a matrix B in M_{22} for which there is no A in $V = M_{22}$ such that $T(A) = B$. What does this tell us about T?

2. Let's define the set D_{22} to be the set of diagonal 2×2 matrices in M_{22}. That is

$$D_{22} = \left\{ \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \;\middle|\; a, b \in \mathbb{R} \right\}.$$

   It is not hard to show that D_{22} is a subspace. If we define T as in Exercise 1 with $V = M_{22}$ and $W = D_{22}$, is T one-to-one? Is T onto?

3. Now define $T : V \to W$ as in Exercise 1, but let $W = M_{22}$ and

$$V = \left\{ \begin{bmatrix} a & b \\ a & b \end{bmatrix} \;\middle|\; a, b \in \mathbb{R} \right\}.$$

Is T one-to-one? Is T onto?

The previous two exercises illustrate the fact that whether a transformation $T : V \to W$ is one-to-one or onto depends not only on the "action" of T, but on the sets V and W as well. The same situation exists when working with functions. Consider the function $f(x) = x^2$, which we generally consider to be not one-to one because, for example $f(-2) = f(2) = 4$, and not onto because there is no x for which $f(x) = -5$. This is because we are thinking f as being defined from the real numbers to the real numbers. Note, however, that if we define $f : \mathbb{R} \to [0, \infty)$ then f becomes onto, and it can be made one-to-one by defining $f : [0, \infty) \to [0, \infty)$.

4. Let $V = W = \mathscr{P}_2$ and let $T : V \to W$ be the derivative operator. That is, for any $p \in \mathscr{P}_2$, $[T(p)](x) = p'(x)$.
   (a) Suppose that we know $p(x)$. Do we then know $p'(x)$ for certain?
   (b) If we know $p'(x)$ do we know $p(x)$ for certain?
   (c) Give a *specific* polynomial in $\ker(T)$. Then give the form of a general element of $\ker(T)$ and write it as a linear combination of specific elements of $\mathscr{P}_2$. Then give a basis $\mathcal{B}$ for $\ker(T)$ and give its dimension, which can be denoted $\dim(\ker(T))$.
   (d) Repeat (c) for $\operatorname{ran}(T)$.
   (e) Is T one-to-one? If not, give two elements $p, q \in \mathscr{P}_2$ for which $p \neq q$ but $T(p) = T(q)$.
   (f) Is T onto? If not, give an element q of $\mathscr{P}_2$ for which there is no $p \in \mathscr{P}_2$ with $[T(p)](x) = q(x)$.
   (g) Do your answers to (e) and (f) change if $W = \mathscr{P}_1$? If so, how?

5. Define $T : M_{22} \to M_{22}$ by $T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} d & c \\ b & a \end{bmatrix}$. In this exercise you will be led through proofs that T is both one-to-one and onto.
   (a) Let $B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be in the codomain M_{22}. Demonstrate that there is a matrix A in the domain M_{22} for which $T(A) = B$. Begin by saying "Let $A = ...$ This shows that any element of the codomain comes from applying T to some element of the domain M_{22}, so T is onto.
   (b) Now begin with "Let $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ and $B = \cdot$ in M_{22} be such that $T(A) = T(B)$." Proceed from there to show that $A = B$. Here are two hints:
      - What do you need to show in order for A and B to be equal?
      - Find $T(A)$ and $T(B)$ and set them equal - then proceed from there.

For Exercises 6 through 10, discuss the kernel and range of the given transformation. Give the rank and nullity for each and tell whether they are one-to one or onto, backing yourself with a proof, counterexample, or reference to a theorem.

6. Define $T : \mathscr{P}_2 \to \mathscr{P}_2$ by $[T(p)](x) = x^2 p''(x)$.

7. Define $T : \mathscr{P}_2 \to \mathbb{R}^2$ by $T(p) = \begin{bmatrix} p(0) \\ p''(0) \end{bmatrix}$.

8. Define $T : M_{22} \to \mathbb{R}$ by $T(A) = trace(A)$. (Recall that the trace is the sum of the diagonal elements.)

9. Define $T : M_{22} \to \mathbb{R}^2$ by $T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} a+d \\ b+d \end{bmatrix}$.

10. Define the linear differential operator $D : \mathscr{P}_2 \to \mathscr{P}_2$ by $D = \dfrac{d}{dx} + 3$.

11. Answer each of the following as its own exercise, with no connection to the previous ones.

    (a) Suppose that $T : V \to W$ is linear, and that $\dim(V) = 5$, $\dim(W) = 3$ and $\text{nullity}(T) = 2$. What is $\text{rank}(T)$? If possible, tell whether T is one-to-one, and whether it is onto.

    (b) Suppose that $T : V \to W$ is linear, and that $\dim(V) = 5$ and $\text{rank}(T) = 4$. What is $\text{nullity}(T)$? Give an inequality telling what you know about $\dim(W)$. If possible, tell whether T is one-to-one, and whether it is onto.

    (c) $T : V \to W$ is linear and $\text{rank}(T) = 4$. Give two inequalities telling what you then know about $\dim(V)$ and $\dim(W)$. If possible, tell whether T is one-to-one, and whether it is onto.

    (d) $T : V \to W$ is linear and $\text{rank}(T) = 4$. Give any inequalities that you can about $\dim(V)$ and $\dim(W)$. If possible, tell whether T is one-to-one, and whether it is onto.

We now return to the concept of differential operators. In Exercise 17 of the previous section you found that, for the differential operator $D : \mathscr{C}^2 \to \mathscr{C}^2$ defined by

$$D(y) = \left(\frac{d^2}{dx^2} + 4\frac{d}{dx} + 3\right)(y) = \frac{d^2y}{dx^2} + 4\frac{dy}{dx} + 3y, \tag{1}$$

both $D(e^{-3x}) = 0$ and $D(e^{-x}) = 0$. We now have a new way of describing this: *Both* e^{-3x} *and* e^{-x} *are in the kernel of* D. It is easily shown that any function of the form

$$y = C_1 e^{-3x} + C_2 e^{-x}, \tag{2}$$

where C_1 and C_2 are constants, is in the kernel of D. The functions e^{-3x} and e^{-x} are independent, and there is a theorem in differential equations that tells us that a second order differential operator like D has exactly two independent solutions. Thus those two functions form a basis for the kernel of D and *every* solution of $D(y) = 0$ has the form (2).

In Exercise 16 of the previous section you looked for a function y for which $D(y) = 5\sin 3x$ (D still being the operator given in (1) above), and found that $y = -\frac{1}{6}\sin 3x - \frac{1}{3}\cos 3x$ is such a function. This tells us that $5\sin 3x$ *is in the range of the operator* D.

12. For the following, take D to be the differential operator defined in (1).

    (a) Show that $4e^{2x}$ is in the range of D by applying D to $y = Ce^{2x}$ to find a value of C for which $D(Ce^{2x}) = 5e^{2x}$.

    (b) Show that $2x - 5$ is in the range of D by assuming that $D(Ax + B) = 2x - 5$ for some values of A and B.

13. Now take D to be the differential operator defined by

$$D(y) = \left(\frac{d}{dx} + 2\right)(y) = \frac{dy}{dx} + 2y \tag{3}$$

for any $y \in \mathscr{C}^1$. The kernel of D in this case consists of only one linearly independent function. Find a basis for the kernel by assuming that $D(y) = 0$ for $y = e^{rt}$, where r is some to be determined constant. When done, give the form of any element of the kernel of D.

2.3 Compositions and Inverses of Transformations, Isomorphisms

Performance Criteria:

2. (e) Given formulas for two transformations, find the composition of them applied to a specific vector, find a formula for the composition.
   (f) Determine whether two transformations are inverses.
   (g) Determine whether a transformation is invertible, and find its inverse if it is.

For almost any mathematical operation, the inverse is of prime importance to us. Since linear transformations are functions, we must treat the concept of the composition of two linear transformations before talking about inverses.

Definition 2.13: If $T : U \to V$ and $S : V \to W$ are transformations, then the **composition of** S **with** T is the transformation denoted by $S \circ T$ and defined by

$$(S \circ T)(\mathbf{u}) = S(T(\mathbf{u}))$$

for every $\mathbf{u}$ in U.

◇ **Example 2.3(a):** Let $S : \mathbb{R}^3 \to \mathscr{P}_2$ defined by $S\left(\begin{bmatrix} a \\ b \\ c \end{bmatrix}\right) = ax^2 + (a+b)x + (a+b+c)$ and $T : \mathscr{P}_2 \to \mathscr{P}_1$ be defined by $T[p(x)] = p'(x)$. Find any of $(S \circ T)(x^2 + 5x - 1)$, $(T \circ S)(x^2 + 5x - 1)$, $(S \circ T)\left(\begin{bmatrix} 5 \\ -2 \\ 3 \end{bmatrix}\right)$, $(T \circ S)\left(\begin{bmatrix} 5 \\ -2 \\ 3 \end{bmatrix}\right)$ that are possible.

We note in the above definition that the codomain of the first transformation to act must be the domain of the second transformation to act. Because of that, $T \circ S$ exists and $S \circ T$ does not. Note, however, that because S acts first, $T \circ S$ is only defined for elements of $\mathbb{R}^3$. Therefore $(T \circ S)(x^2 + 5x - 1)$ is meaningless and does not exist. We *CAN* find

$$(T \circ S)\left(\begin{bmatrix} 5 \\ -2 \\ 3 \end{bmatrix}\right) = T(5x^2 + 3x + 6) = 10x + 3. \tag{1}$$

◇ **Example 2.3(b):** Let S and T be as in the previous example. Find a formula for $(T \circ S)\left(\begin{bmatrix} a \\ b \\ c \end{bmatrix}\right)$.

Here we simply carry out the computation (1) above for a general element of $\mathbb{R}^3$:

$$(T \circ S)\left(\begin{bmatrix} a \\ b \\ c \end{bmatrix}\right) = T\big[ax^2 + (a+b)x + (a+b+c)\big] = 2ax + (a+b).$$

Thus, for any $\begin{bmatrix} a \\ b \\ c \end{bmatrix}$ we have $(T \circ S)\left(\begin{bmatrix} a \\ b \\ c \end{bmatrix}\right) = 2ax + (a+b)$.

Definition 2.13 is meaningful even if one (or both) of S and T is not linear. In the case that both are linear, it should be no surprise that the composition is as well:

> **Theorem 2.14:** If $T: U \to V$ and $S: V \to W$ are linear transformations, then $S \circ T : U \to W$ is a linear transformation.

Now we can define an inverse of a linear transformation:

> **Definition 2.15:** A linear transformation $T: V \to W$ is **invertible** if there is a linear transformation $T': W \to V$ such that
>
> $$(T' \circ T)\mathbf{v} = \mathbf{v} \text{ for all } \mathbf{v} \in V \qquad \text{and} \qquad (T \circ T')\mathbf{w} = \mathbf{w} \text{ for all } \mathbf{w} \in W$$
>
> In this case, T' is called an **inverse** for T.

◇ **Example 2.3(c):** Let $S: \mathbb{R}^2 \to \mathscr{P}_1$ by $S\left(\begin{bmatrix} a \\ b \end{bmatrix}\right) = (b-a)x + (2a-b)$ and $T: \mathscr{P}_1 \to \mathbb{R}^2$ by $T\big[p(x)\big] = \begin{bmatrix} p(1) \\ p(2) \end{bmatrix}$. Determine whether S and T are inverses.

Let $p \in \mathscr{P}_1$. Then $p(x) = ax + b$ for some $a, b \in \mathbb{R}$ and

$$\big[(S \circ T)(p)\big](x) = S[T(p(x)] = S\left(\begin{bmatrix} a+b \\ 2a+b \end{bmatrix}\right) = \big[(2a+b)-(a+b)\big]x + \big[2(a+b)-(2a+b)\big] = ax+b$$

Now let $\begin{bmatrix} a \\ b \end{bmatrix} \in \mathbb{R}^2$. We see that

$$(T \circ S)\left(\begin{bmatrix} a \\ b \end{bmatrix}\right) = T\left[S\left(\begin{bmatrix} a \\ b \end{bmatrix}\right)\right] = T\big[(b-a)x + (2a-b)\big] = \begin{bmatrix} (b-a)+(2a-b) \\ 2(b-a)+(2a-b) \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}.$$

The above two computations show that S and T are inverses of each other.

The next theorem allows us to replace *an* inverse of T with *the* inverse of T.

> **Theorem 2.16:** If T is an invertible linear transformation, then its inverse is unique. That is, there is only one inverse of T and we denote it by T^{-1}.

We generally wish to know the inverses of any transformations of interest to us, if they happen to be invertible. In some cases it is fairly easy to find the inverse of a transformation, as shown below.

◇ **Example 2.3(d):** $T: \mathscr{P}_1 \to \mathscr{P}_1$ defined by $T(ax+b) = 2bx - a$ is linear and invertible. Give the inverse transformation T^{-1}.

Because in this case $V = W = \mathscr{P}_1$, T^{-1} is a mapping from $\mathscr{P}_1$ to $\mathscr{P}_1$ and we can write $T^{-1}(ax+b) = cx+d$. But then $T(cx+d) = ax+b$ and, by applying T we have

$$T(cx+d) = 2dx - c = ax + b.$$

Equating like terms we get $a = 2d$ and $b = -c$. We then solve for c and d to get the output $cx+d$ of T^{-1}:

$$T^{-1}(ax+b) = -bx + \tfrac{a}{2}.$$

Let's summarize the method used in the last example to find the inverse $T^{-1}: W \to V$ of a transformation $T: V \to W$. This works when V and W are any of $\mathscr{P}_n$, $\mathbb{R}^n$ and M_{mn}:

(1) Create a general $\mathbf{w} \in W$ and a general $\mathbf{v} \in V$ for which $T^{-1}\mathbf{w} = \mathbf{v}$.

(2) It must be the case that, for the general $\mathbf{v}$ and $\mathbf{w}$ of (1), that $T\mathbf{v} = \mathbf{w}$. Apply T to $\mathbf{v}$ and set the result equal to $\mathbf{w}$.

(3) You should now be able to equate like terms/entries to solve for constants in $\mathbf{v}$.

(4) $T^{-1}\mathbf{w} = \mathbf{v}$ for the general $\mathbf{w}$ of part (1) and the $\mathbf{v}$ determined in (3) is the desired inverse.

Before seeking an inverse, we might wish to make sure that one exists. Here's a theorem that can be helpful in that regard:

Theorem 2.17: A linear transformation $T: V \to W$ is **invertible** if and only if T is one-to-one and onto.

The following theorem will soon be handy.

Theorem 2.18: Suppose that $\dim(V) = \dim(W) = n$ and $T: V \to W$ is linear. Then T is one-to-one if and only if it is onto.

Definition 2.19: A one-to-one and onto linear transformation $T: V \to W$ is called an **isomorphism**. If there is an isomorphic from one vector space to another, the two spaces are said to be **isomorphic**.

◇ **Example 2.3(e):** Show that $\mathscr{P}_2$ and $\mathbb{R}^3$ are isomorphic by showing that there is an isomorphism $T: \mathscr{P}_2 \to \mathbb{R}^3$.

First we note that $\mathcal{B} = \{x^2, x, 1\}$ is a basis for $\mathscr{P}_2$ and $\mathcal{C} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$ is a basis for $\mathbb{R}^2$, so both have dimension three. Define $T: \mathscr{P}_2 \to \mathbb{R}^2$ by $T(ax^2 + bx + c) = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$; T can easily be shown to be linear. For any $\mathbf{v} = \begin{bmatrix} d \\ e \\ f \end{bmatrix}$ in $\mathbb{R}^3$, $T(dx^2 + ex + f) = \mathbf{v}$, so T is onto. By Theorem 2.18, T is also one-to-one, so it is an isomorphism.

As you might guess, the above example illustrates a more general result:

Theorem 2.20: A vector space V has dimension n if and only if it is isomorphic to $\mathbb{R}^n$.

In particular, if V is an n-dimensional vector space with basis $\mathcal{B}$, the mapping that takes $\mathbf{v}$ to its coordinate vector $[\mathbf{v}]_{\mathcal{B}}$ is an isomorphism from V to $\mathbb{R}^n$.

Section 2.3 Exercises

1. Define the transformations $R:\mathbb{R}^2 \to \mathscr{P}_1$, $S:\mathscr{P}_1 \to \mathbb{R}^2$ and $T:\mathscr{P}_1 \to \mathscr{P}_1$ by

$$R\left(\begin{bmatrix} a \\ b \end{bmatrix}\right) = (b-a)x + (2a-b), \qquad S\big[p(x)\big] = \begin{bmatrix} p(1) \\ p(2) \end{bmatrix}, \qquad T(ax+b) = 2bx - a.$$

   (a) $T\circ R$ is from what space to what space? How about $S\circ T$? Give your answers in the form $T\circ R : V \to W$.

   (b) Give a formula for $T\circ R$.

   (c) Give a formula for $S\circ T$.

2. Define $R:\mathscr{P}_2 \to \mathscr{P}_2$, $S:M_{22} \to \mathscr{P}_2$ and $T:\mathbb{R}^3 \to \mathscr{P}_2$ by $[R(p)](x) = x^2p''(x)$,

$$S\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = (a+b)x^2 + (b+c)x + (c+d), \qquad T\left(\begin{bmatrix} a \\ b \\ c \end{bmatrix}\right) = ax^2 + (a+b)x + (a+b+c)$$

   (a) There are six potentially possible compositions of the two transformations. List them all in the form $R\circ S$.

   (b) For any of the compositions that *ARE* possible, write a statement of the form $R\circ S : V \to W$ indicating the spaces that the composition maps from and to.

   (c) Give formulas for each of the compositions that do exist.

3. Consider the transformations T and T^{-1} (both from $\mathscr{P}_1$ to $\mathscr{P}_1$) of Example 2.3(d), defined by $T(ax+b) = 2bx - a$ and $T^{-1}(ax+b) = -bx + \frac{a}{2}$. Compute both $(T\circ T^{-1})(ax+b)$ and $(T^{-1}\circ T)(ax+b)$ to verify that the transformations are in fact inverses. Show all steps as done in Example 2.3(c).

4. Let $S:\mathbb{R}^3 \to \mathscr{P}_2$ and $T:\mathscr{P}_2 \to \mathbb{R}^3$ be defined by

$$S\left(\begin{bmatrix} a \\ b \\ c \end{bmatrix}\right) = \frac{1}{2}cx^2 + (b-c)x + (a-b-c), \qquad T\big[p(x)\big] = \begin{bmatrix} p(1) \\ p'(1) \\ p''(1) \end{bmatrix}.$$

   Determine whether S and T are inverses.

5. Let $A = \begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}$ and $B = \begin{bmatrix} 3 & -5 \\ -1 & 2 \end{bmatrix}$, and define $S:\mathbb{R}^2 \to \mathbb{R}^2$ by $S(\mathbf{x}) = A\mathbf{x}$ and $T:\mathbb{R}^2 \to \mathbb{R}^2$ by $T(\mathbf{x}) = B\mathbf{x}$. Determine whether S and T are inverses.

6. Let $T:\mathbb{R}^3 \to \mathscr{P}_2$ defined by $T\left(\begin{bmatrix} a \\ b \\ c \end{bmatrix}\right) = ax^2 + (a+b)x + (a+b+c)$. Find a formula for the inverse function T^{-1}.

7. Find a formula for the inverse of $S:\mathscr{P}_2 \to \mathbb{R}^3$ defined by $S\big[p(x)\big] = \begin{bmatrix} p(0) \\ p'(0) \\ p''(0) \end{bmatrix}$.

8. How do we know that $S:M_{22} \to \mathscr{P}_2$ defined by $S\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = (a+b)x^2 + (b+c)x + (c+d)$ is not invertible?

2.4 The Matrix of a Linear Transformation

Performance Criteria:

2. (h) Determine the matrix of a linear transformation $T : V \to W$ with respect to bases $\mathcal{B}$ and $\mathcal{C}$ of V and W, respectively.

   (i) Determine the matrix of a composition of linear transformations, determine the matrix of the inverse of a transformation.

When you were in Math 341 you found that every linear transformation $T : \mathbb{R}^m \to \mathbb{R}^n$ could be performed by multiplying elements in $\mathbb{R}^m$ by an $n \times m$ matrix A. When dealing with more general vector spaces as we have been, the "vectors" aren't necessarily vectors in the "standard" sense of a vector in some $\mathbb{R}^n$. However, suppose that we have two vector spaces V and W of dimensions m and n, and having bases $\mathcal{B}$ and $\mathcal{C}$, respectively. Every vector $\mathbf{v} \in V$ has a coordinate vector $[\mathbf{v}]_{\mathcal{B}} \in \mathbb{R}^m$ and every $\mathbf{w} \in W$ has a coordinate vector $[\mathbf{w}]_{\mathcal{C}} \in \mathbb{R}^n$ also. Given a linear transformation $T : V \to W$, there is then a matrix A that performs the action of T in the sense given by the following theorem.

Theorem 2.21: Suppose that V and W are finite dimensional vector spaces with bases $\mathcal{B}$ and $\mathcal{C}$, and $T : V \to W$ is a linear transformation. Then there exists a matrix A such that for every $\mathbf{v} \in V$

$$A[\mathbf{v}]_{\mathcal{B}} = [T\mathbf{v}]_{\mathcal{C}}$$

If $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$, then the columns of A are the coordinate vectors $[T\mathbf{v}_1]_{\mathcal{C}}$, $[T\mathbf{v}_2]_{\mathcal{C}}$, ..., $[T\mathbf{v}_n]_{\mathcal{C}}$. The matrix A is often denoted $[T]_{\mathcal{B},\mathcal{C}}$.

◇ **Example 2.4(a):** The transformation $S : \mathscr{P}_2 \to \mathbb{R}^2$ defined by $[S(p)](x) = \begin{bmatrix} p(0) \\ p(1) \end{bmatrix}$ is linear, and $\mathcal{B} = \{1, x, x^2\}$ and $\mathcal{C} = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$ are bases for $\mathscr{P}_2$ and $\mathbb{R}^2$. Find $[T]_{\mathcal{B},\mathcal{C}}$.

First we see that $T(1) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $T(x) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and $T(x^2) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, and we can determine that $[T(1)]_{\mathcal{C}} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $[T(x)]_{\mathcal{C}} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and $[T(x^2)]_{\mathcal{C}} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. Therefore $[T]_{\mathcal{B},\mathcal{C}} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & -1 & -1 \end{bmatrix}$.

◇ **Example 2.4(b):** Let $p(x) = 3x^2 - 6x + 2$. Verify that the matrix from the previous example works as described in the theorem, by finding $[T]_{\mathcal{B},\mathcal{C}}[p(x)]_{\mathcal{B}}$ and $[T[p(x)]]_{\mathcal{C}}$ and seeing that they are equal.

We see that $[p(x)]_{\mathcal{B}} = \begin{bmatrix} 2 \\ -6 \\ 3 \end{bmatrix}$ and $[T]_{\mathcal{B},\mathcal{C}}[p(x)]_{\mathcal{B}} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & -1 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ -6 \\ 3 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$. We also have $T[p(x)] = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$ and we can see that $(-1)\begin{bmatrix} 1 \\ 1 \end{bmatrix} + (3)\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$, so $[T[p(x)]]_{\mathcal{C}} = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$.

A couple comments are in order:

- The matrix $[T]_{\mathcal{B},\mathcal{C}}$ is unique.
- If $W = V$ and $\mathcal{C} = \mathcal{B}$, we write $[T]_{\mathcal{B}}$.

The next example illustrates that sometimes we can determine the matrix of the transformation without using the method described in Theorem 2.21.

◇ **Example 2.4(c):** For the transformation $D : \mathscr{P}_2 \to \mathscr{P}_1$ given by $D[p(x)] = p'(x)$ and the basis $\mathcal{B} = \{1, x, x^2\}$ for $\mathscr{P}_2$ and $\mathcal{C} = \{1, x\}$ for $\mathscr{P}_1$, find $[T]_{\mathcal{B},\mathcal{C}}$.

If $p(x) = ax^2 + bx + c$, then $p'(x) = 2ax + b$ and $[p(x)]_{\mathcal{B}} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$, $[p'(x)]_{\mathcal{C}} = \begin{bmatrix} 2a \\ b \end{bmatrix}$. We then need to find a matrix A such that $A[p(x)]_{\mathcal{B}} = [p'(x)]_{\mathcal{C}}$. A must be 2×3 and we can think of

$$A[p(x)]_{\mathcal{B}} = \begin{bmatrix} * & * & * \\ * & * & * \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 2a \\ b \end{bmatrix}.$$

Now we need only to visualize what the first row would look like to obtain $2a$, the first coordinate of $[p'(x)]_{\mathcal{C}}$, and similarly for the second row. From this we can see that the desired matrix is $[T]_{\mathcal{B},\mathcal{C}} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$.

The following theorem says that

the matrix of a composition of two transformations is the product of the matrices of the individual transformations.

Theorem 2.22: Suppose that U, V and W are finite dimensional vector spaces with bases $\mathcal{B}, \mathcal{C}$ and $\mathcal{D}$, respectively. If $T : U \to V$ and $S : V \to W$ are linear, then

$$[S \circ T]_{\mathcal{B},\mathcal{D}} = [S]_{\mathcal{C},\mathcal{D}}\, [T]_{\mathcal{B},\mathcal{C}}$$

And this one says that

the matrix of the inverse of a transformation is the inverse of the matrix of the transformation.

Theorem 2.23: Suppose that V and W are n-dimensional vector spaces with bases $\mathcal{B}$ and $\mathcal{C}$, and $T : V \to W$ is linear. Then T is invertible if and only if $[T]_{\mathcal{B},\mathcal{C}}$ is, and

$$[T^{-1}]_{\mathcal{C},\mathcal{B}} = \big([T]_{\mathcal{B},\mathcal{C}}\big)^{-1}$$

Section 2.4 Exercises

1. Define $T : \mathscr{P}_1 \to \mathbb{R}^3$ by $T(ax+b) = \begin{bmatrix} a+b \\ a-b \\ b-a \end{bmatrix}$. T is linear, and $\mathcal{B} = \{x+2, 2x-1\}$ and $\mathcal{C} = \left\{ \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$ are bases for $\mathscr{P}_1$ and $\mathbb{R}^3$, respectively. In this exercise you will find the matrix of T with respect to $\mathcal{B}$ and $\mathcal{C}$ and verify that it "works."

   (a) Find $[3x-4]_{\mathcal{B}}$.

   (b) Find $T(3x-4)$ and give the coordinate vector of your answer with repsect to the basis $\mathcal{C}$. Give your answer using the notation $[T(3x-4)]_{\mathcal{C}}$, using fractions for the components.

   (c) Find $[T]_{\mathcal{B},\mathcal{C}}$, the matrix of T with respect to $\mathcal{B}$ and $\mathcal{C}$.

   (d) Find the product $[T]_{\mathcal{B},\mathcal{C}}[3x-4]_{\mathcal{B}}$. Is the result what you expected?

The point of exercises 2 through 6 is to explore the idea of "the" matrix of a linear transformation. See Exercise 6 for what the quotes are about.

2. Let $T : \mathscr{P}_3 \to \mathscr{P}_2$ be the derivative transformation, which is of course linear.

   (a) There is a very obvious map from $\mathscr{P}_3$ to $\mathbb{R}^4$. What vector in $\mathbb{R}^4$ would the polynomial $p(x) = 5 + \frac{1}{2}x - x^2 + 2x^3$ map to? ("Go" in the order given.)

   (b) The vector you gave in (a) is really just the coordinate vector $[p(x)]_{\mathcal{B}}$ with respect to some basis $\mathcal{B}$ for $\mathscr{P}_3$. Give that basis.

   (c) Find the derivative $p'(x)$ of $p(x) = 5+\frac{1}{2}x-x^2+2x^3$ and give the vector $[p'(x)]_{\mathcal{B}}$ in $\mathbb{R}^3$ that corresponds to it. (Technically, the basis here is not the basis you gave in (b), but it is "almost" the same.)

   (d) Find a matrix A such that $A[p(x)]_{\mathcal{B}} = [p'(x)]_{\mathcal{B}}$. This is the matrix of the transformation T when the bases for $\mathscr{P}_3$ and $\mathscr{P}_2$ are the standard bases for those spaces. If we call those two bases $\mathcal{E}_3$ and $\mathcal{E}_2$, we denote this matrix by $[T]_{\mathcal{E}_3,\mathcal{E}_2}$.

3. Find $[p(x)]_{\mathcal{B}}$, where $p(x) = 5 + \frac{1}{2}x - x^2 + 2x^3$ and $\mathcal{B} = \{1+x, x+x^2, x^2+x^3, x^3+1\}$. .

4. The set $\mathcal{C} = \{1,\ 1+x,\ 1+x+x^2\}$ is a basis for $\mathscr{P}_2$. For the polynomial $p(x) = 5 + \frac{1}{2}x - x^2 + 2x^3$, find $[p'(x)]_{\mathcal{C}}$.

5. For this exercise, let $\mathcal{B}$ and $\mathcal{C}$ be the bases from Exercises 2 and 3 above, and let $p(x) = 5 + \frac{1}{2}x - x^2 + 2x^3$.

   (a) Use Theorem 6.26 to find the matrix $[T]_{\mathcal{B},\mathcal{C}}$ of the derivative transformation *with respect to the bases* $\mathcal{B}$ *and* $\mathcal{C}$.

   (b) Check your results by finding $[T]_{\mathcal{B},\mathcal{C}}[p(x)]_{\mathcal{B}}$. The result should be $[p'(x)]_{\mathcal{C}}$, so check your answer against what you got for Exercise 3.

6. In the course of doing these exercises you found two different matrices for the same linear transformation $T : \mathscr{P}_3 \to \mathscr{P}_2$. Why can there be different matrices for the same transformation?

2.5 Chapter 2 Exercises

1. Define $T: M_{22} \to \mathbb{R}^2$ by $T(A) = A\begin{bmatrix} 2 \\ 1 \end{bmatrix}$. Because of some properties of a matrix times a vector, T is linear.

   (a) Find $T\left(\begin{bmatrix} -3 & 1 \\ 2 & 4 \end{bmatrix}\right)$.

   (b) Demonstrate that $\ker(T) \neq [\,0\,]$ by giving a matrix $B \neq [\,0\,]$ for which $T(B) = \mathbf{0}$. (Here $[\,0\,]$ denotes the 2×2 matrix with all zero entries.)

   (c) Give the kernel of T using set builder notation. Then give a basis for the kernel, denoting it by $\mathcal{B}_{\ker(T)}$, and give the nullity of T.

   (d) What is the rank of T? Give a matrix A for which $T(A) = \begin{bmatrix} e \\ f \end{bmatrix}$. What does this demonstrate about T?

   (e) T is not invertible - why not?

2. Consider the transformation of the previous exercise and the bases

$$\mathcal{B} = \left\{\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\right\} \qquad \text{and} \qquad \mathcal{C} = \left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\}$$

   for M_{22} and $\mathbb{R}^2$.

   (a) For $A = \begin{bmatrix} -3 & 1 \\ 2 & 4 \end{bmatrix}$, give the coordinate vector $[A]_{\mathcal{B}}$.

   (b) Give the coordinate vector $[T(A)]_{\mathcal{C}}$. (Your answer to 1(a) should be useful here...)

   (c) Find the matrix $[T]_{\mathcal{B},\mathcal{C}}$ of the transformation.

   (d) Without doing any work, what should $[T]_{\mathcal{B},\mathcal{C}}[A]_{\mathcal{B}}$ be? Now, using your answers to (c) and (a), find $[T]_{\mathcal{B},\mathcal{C}}[A]_{\mathcal{B}}$.

3 Orthogonality and Inner Product Spaces

Outcome/Performance Criteria:

3. Understand inner products and inner product spaces.

   (a) Determine whether an operation is an inner product on a vector space.

   (b) In an inner product space, find the inner product of two vectors, determine whether two vectors are orthogonal.

   (c) In an inner product space, find the norm of a vector.

   (d) Determine whether a basis for an inner product space is orthogonal; determine whether a basis is orthonormal.

   (e) In an inner product space, find the representation of a vector with respect to an orthogonal or orthonormal basis without solving a system of equations.

   (f) Given a basis for a subspace W of $\mathbb{R}^n$ and a $\mathbf{v}$ in $\mathbb{R}^n$, find the unique orthogonal decomposition $\mathbf{v} = \mathbf{w} + \mathbf{w}^\perp$ with $\mathbf{w}$ in W and $\mathbf{w}^\perp$ in $W^\perp$.

   (g) Apply the Gram-Schmidt process to a set of vectors in an inner product space to obtain an orthogonal basis; normalize a vector or set of vectors in an inner product space.

   (h) Find the norm of a vector, or the distance between two vectors, using a given norm.

3.1 Inner Products and Orthogonality

Performance Criteria:

3. (a) Determine whether an operation is an inner product on a vector space.
   (b) In an inner product space, find the inner product of two vectors, determine whether two vectors are orthogonal.
   (c) In an inner product space, find the norm of a single vector, find the distance between two vectors.

In a previous linear algebra course, and in vector calculus, you saw and used the **dot product** of two vectors. This is what we call a *binary operation* that "takes in" two vectors and "gives out" a real number that is the dot product of the two vectors. The dot product has certain properties that also hold for some binary operations in abstract vector spaces. A binary operation on such a vector space that has those properties is called an inner product, which is then a generalization of the dot product to these spaces.

Definition 3.1: Let V be a vector space. A binary operation $\langle *, * \rangle$ that assigns a real number $\langle \mathbf{u}, \mathbf{v} \rangle$ to every pair of vectors $\mathbf{u}, \mathbf{v} \in V$ is called an **inner product** if all of the following hold for every $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and any scalar c:

1) $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$

2) $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$

3) $\langle c\mathbf{u}, \mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$

4) $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ with equality if and only if $\mathbf{u} = \mathbf{0}$

A vector space along with an inner product is called an **inner product space**.

◇ **Example 3.1(a):** Let $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ be vectors in $\mathbb{R}^n$, and define $\langle \mathbf{u}, \mathbf{v} \rangle$ to be the dot product

$$\langle \mathbf{u}, \mathbf{v} \rangle = u_1v_1 + u_2v_2 + \cdots + u_nv_n.$$

This binary operation is an inner product on $\mathbb{R}^n$.

◇ **Example 3.1(b):** Let $\mathbf{u}$ and $\mathbf{v}$ be as in the previous example, and let $w_1, w_2, ..., w_n$ be *fixed* non-negative constants. Then the binary operation

$$\langle \mathbf{u}, \mathbf{v} \rangle = w_1u_1v_1 + w_2u_2v_2 + \cdots + w_nu_nv_n$$

is an inner product on $\mathbb{R}^n$. This type of inner product is called a **weighted inner product** with weight vector $\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$.

◇ **Example 3.1(c):** Let $p_1(x) = a_1x^2 + b_1x + c_1$ and $p_2(x) = a_2x^2 + b_2x + c_2$ be elements of $\mathscr{P}_2$. the binary operation defined by

$$\langle p_1, p_2 \rangle = a_1a_2 + b_1b_2 + c_1c_2$$

is an inner product on $\mathscr{P}_2$.

◇ **Example 3.1(d):** Let $\mathscr{C}[a, b]$ denote the functions that are continuous on the interval $[a, b]$, and define

$$\langle f, g \rangle = \int_a^b f(x)\, g(x)\, dx$$

for any $f, g \in \mathscr{C}[a, b]$. This is an inner product.

◇ **Example 3.1(e):** Let $\mathscr{C}[a, b]$ be as in the previous example, and let $w(x)$ be a *fixed* continuous function that is non-negative. For any $f, g \in \mathscr{C}[a, b]$, define

$$\langle f, g \rangle = \int_a^b w(x)\, f(x)\, g(x)\, dx$$

This is a weighted inner product on $\mathscr{C}[a, b]$, with weight function $w(x)$.

A student of mathematics should at some point note the fact that sums and integrals are analogous. This should be no surprise, as an integral is defined to be a limit of sums. We can see the analogy more clearly if we write the inner product from (a) above and compare with the inner product from (d):

$$\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{j=1}^{n} u_j v_j \qquad \text{and} \qquad \langle f, g \rangle = \int_a^b f(x)\, g(x)\, dx$$

The product $u_j v_j$ can be thought of as the product of $\mathbf{u}$ and $\mathbf{v}$ at some j with $1 \le j \le n$, and $f(x)\, g(x)$ is the product of the functions f and g at some value of x with $a \le x \le b$. The dx in the integral is really a weight, with the corresponding weight for the sum being one. And, finally, summing from zero to n is analogous to integrating from a to b.

The following are easily derived from the properties of an inner product:

> **Theorem 3.2:** Let V be an inner product space with inner product $\langle *, * \rangle$. Then for every $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and any scalar c,
>
> 1) $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$
>
> 2) $\langle \mathbf{u}, c\mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$
>
> 3) $\langle \mathbf{u}, \mathbf{0} \rangle = \langle \mathbf{0}, \mathbf{u} \rangle = 0$

In $\mathbb{R}^2$ or $\mathbb{R}^3$, two vectors are perpendicular if and only if their dot product is zero. The following definition is the analog for general inner product spaces.

> **Definition 3.3:** Let V be an inner product space with inner product $\langle *, * \rangle$. Two vectors $\mathbf{u}$ and $\mathbf{v}$ in V are **orthogonal** if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.

◇ **Example 3.1(f):** The vectors $\mathbf{u} = \begin{bmatrix} -3 \\ 5 \\ 2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 4 \\ 4 \\ -4 \end{bmatrix}$ in $\mathbb{R}^3$ are orthogonal with the inner product $\langle \mathbf{u}, \mathbf{v} \rangle = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$.

◇ **Example 3.1(g):** The polynomials $p_(x) = -3x^2 + 5x + 2$ and $q(x) = 4x^2 + 4x - 4$ in $\mathscr{P}_2$ are orthogonal with the inner product defined by

$$\langle p_1, p_2 \rangle = a_1 a_2 + b_1 b_2 + c_1 c_2 \qquad \text{for} \qquad p_1(x) = a_1 x^2 + b_1 x + c_1 \text{ and } p_2(x) = a_2 x^2 + b_2 x + c_2.$$

Note that the vectors in Example 3.1(f) are the coordinate vectors of these polynomials with respect to the standard basis $\mathcal{B} = \{x^2, x, 1\}$.

◇ **Example 3.1(h):** Let $\mathbf{w} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. Then the vectors $\mathbf{u} = \begin{bmatrix} 6 \\ -3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ in $\mathbb{R}^2$ are orthogonal with the weighted inner product $\langle \mathbf{u}, \mathbf{v} \rangle = w_1 u_1 v_1 + w_2 u_2 v_2 + \cdots + w_n u_n v_n$ defined in Example 3.1(b). Note that these same vectors *are not* orthogonal by the standard inner product $\langle \mathbf{u}, \mathbf{v} \rangle = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$.

◇ **Example 3.1(i):** Letting $f(x) = x$ and $g(x) = x^2$, we see that

$$\int_{-1}^{1} f(x)\, g(x)\, dx = \int_{-1}^{1} x^3\, dx = \left. \frac{x^4}{4} \right|_{-1}^{1} = \frac{1}{4} - \frac{1}{4} = 0.$$

Therefore f and g are orthogonal in $\mathscr{C}[-1,1]$ with the inner product from Example 3.1(d).

Recall that if $\mathbf{x} \in \mathbb{R}^n$ has components x_1, x_2, ..., x_n, the *magnitude* of $\mathbf{x}$ is given by

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}.$$

This is the motivation for the following definition:

> **Definition 3.4:** Let V be an inner product space with inner product $\langle *, * \rangle$. Then for every $\mathbf{v} \in V$ we define the **norm** of $\mathbf{v}$, denoted by $\|\mathbf{v}\|$, to be
>
> $$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}.$$
>
> We sometimes use either of the terms **magnitude** or **length** synonymously with norm.

Later we will see other norms that do not come from an inner product. When a norm is obtained from an inner product in the manner described above, we say the norm is *induced by* the inner product.

◇ **Example 3.1(j):** Consider $p(x) = 3x^2 - 6x + 1$ with the inner product of Example 3.1(c) and 3.1(g) (see above). Using that inner product,

$$\|p\| = \sqrt{\langle p, p \rangle} = \sqrt{(3)(3) + (-6)(-6) + (1)(1)} = \sqrt{46}.$$

◇ **Example 3.1(k):** Consider $f(x) = x^2$ on the interval $[-1, 1]$ with the inner product

$$\langle f, g \rangle = \int_a^b f(x)\, g(x)\, dx.$$

We have

$$\|f\| = \sqrt{\langle f, f \rangle} = \left(\int_{-1}^{1} f(x)\, f(x)\, dx \right)^{\frac{1}{2}} = \left(\int_{-1}^{1} [f(x)]^2\, dx \right)^{\frac{1}{2}} = \left(\int_{-1}^{1} x^4\, dx \right)^{\frac{1}{2}} = \left(\frac{x^5}{5} \bigg|_{-1}^{1} \right)^{\frac{1}{2}} = \sqrt{\frac{2}{5}}$$

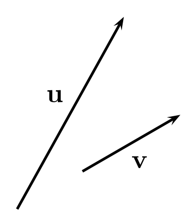
The expression $\left(\int_{-1}^{1} [f(x)]^2\, dx \right)^{\frac{1}{2}}$ is seen in applications as the **root mean square** (RMS) norm. Often we divide the integral by the length of the interval over which we are integrating. In the above case we would then have

$$\|f\| = \left(\frac{1}{2} \int_{-1}^{1} [f(x)]^2\, dx \right)^{\frac{1}{2}}$$

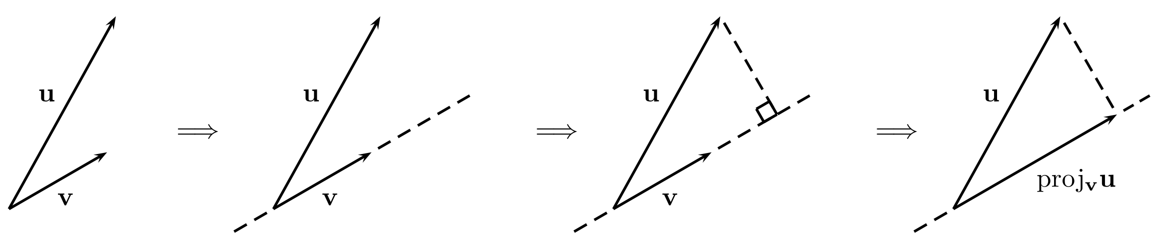
The "root" of root mean square should be obvious, and "mean" refers to the fact that $\frac{1}{2} \int_{-1}^{1} [f(x)]^2\, dx$ is the mean, or average, of the "square" of f over the interval $[-1, 1]$, which has length two (hence the $\frac{1}{2}$ in front of the integral). *In this course we will only divide by the length of the interval if doing so is included in the definition of the inner product.*

We now turn to the idea of the projection of one vector onto another, which you should be familiar with from other courses. Given two vectors $\mathbf{u}$ and $\mathbf{v}$, we can create a new vector called the **projection of $\mathbf{u}$ onto $\mathbf{v}$**, denoted by $\text{proj}_{\mathbf{v}}\mathbf{u}$. This is a very useful idea, in many ways. Geometrically, we can find a projection for two vectors in $\mathbb{R}^2$ in the way demonstrated in this example:

◇ **Example 3.1(l):** For the vectors $\mathbf{u}$ and $\mathbf{v}$ shown to the right, find the projection $\text{proj}_{\mathbf{v}}\mathbf{u}$.



We put the two vectors tail-to-tail, extend $\mathbf{v}$ to form a "screen" and then finish as shown:



Algebraically we find the projection of $\mathbf{u}$ onto $\mathbf{v}$ with the formula

$$\text{proj}_{\mathbf{v}}\mathbf{u} = \frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{v} \cdot \mathbf{v}}\, \mathbf{v}$$

where $\mathbf{u} \cdot \mathbf{v}$ is standard dot product. Note that since both $\mathbf{u} \cdot \mathbf{v}$ and $\mathbf{v} \cdot \mathbf{v}$ are scalars, so is $\frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{v} \cdot \mathbf{v}}$. That scalar is then multiplied times $\mathbf{v}$, resulting in a vector parallel $\mathbf{v}$. If the scalar is positive the projection is in the direction of $\mathbf{v}$, as shown in the example above; when the scalar is negative the projection is in the direction opposite the vector being projected onto. The projection of $\mathbf{u}$ onto $\mathbf{v}$ is probably best thought of as *the part of* $\mathbf{u}$ *that is parallel to* $\mathbf{v}$.

The concept of the projection of one vector onto another holds in any inner product space:

> **Definition 3.5:** Let V be an inner product space. If $\mathbf{u}, \mathbf{v} \in V$, the projection of $\mathbf{u}$ onto $\mathbf{v}$ is the vector given by
>
> $$\text{proj}_{\mathbf{v}}\mathbf{u} = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle}\,\mathbf{v}$$

We should note that the projection of $\mathbf{u}$ onto $\mathbf{v}$ is itself a vector, parallel to $\mathbf{v}$. The quantity $\dfrac{\langle \mathbf{u}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle}$ is a scalar that determines the length of the projection.

◇ **Example 3.1(m):** Again using the inner product from Example 3.2(b), find the projection of $p(x) = x^2 + 3$ onto $q(x) = 5x + 2$.

First we find

$$\langle p, q \rangle = \int_{-1}^{1} (x^2 + 3)(5x + 2)\,dx = \frac{40}{3} \qquad \text{and} \qquad \langle q, q \rangle = \int_{-1}^{1} (5x + 2)^2\,dx = \frac{74}{3}$$

It then follows that

$$\text{proj}_q p = \frac{\langle p, q \rangle}{\langle q, q \rangle}\,q(x) = \frac{40/3}{74/3}(5x + 2) = \tfrac{40}{74}(5x + 2) = \tfrac{100}{37}x + \tfrac{40}{37}$$

Section 3.1 Exercises

1. Let $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ represent vectors in $\mathbb{R}^2$. For each of the following, determine whether the given binary operation $\langle \mathbf{u}, \mathbf{v} \rangle$ is in fact an inner product. If it is, say so. If it is not give a specific counterexample showing that it violates one of the conditions of an inner product.

   (a) $\langle \mathbf{u}, \mathbf{v} \rangle = u_1^2 v_1^2 + u_2^2 v_2^2$ (b) $\langle \mathbf{u}, \mathbf{v} \rangle = u_1 v_1$ (c) $\langle \mathbf{u}, \mathbf{v} \rangle = u_1 + v_2$

   (d) $\langle \mathbf{u}, \mathbf{v} \rangle = 4u_1 v_1 + u_2 v_2$ (e) $\langle \mathbf{u}, \mathbf{v} \rangle = 4(u_1 v_1 + u_2 v_2)$ (f) $\langle \mathbf{u}, \mathbf{v} \rangle = -4(u_1 v_1 + u_2 v_2)$

2. Let $p(x) = x^2$ and $q(x) = 1$. Note that $p, q \in \mathscr{P}_2$.

   (a) Find $\langle p, q \rangle$ for the inner product defined in Example 3.1(g).

   (b) Find $\langle p, q \rangle$ for the inner product defined in Example 3.1(i).

   (c) Are p and q orthogonal?

   (d) Find $\|p\|$ and $\|q\|$ for the norm induced by the inner product of Example 3.1(g).

   (e) Find $\|p\|$ and $\|q\|$ for the norm induced by the inner product of Example 3.1(i).

3. Consider the vector space $\mathscr{C}[0, 2]$ with the inner product defined by $\langle f, g \rangle = \displaystyle\int_0^2 f(x)\,g(x)\,dx$ for all $f, g \in \mathscr{C}[0, 2]$.

   (a) Find $\langle x, e^x \rangle$. You may use your calculator to compute this - round to the hundredth's place if you give a decimal answer.

(b) Find $\|3x+2\|$ using the norm induced by the inner product.

(c) Find $\langle \sin \pi x, \cos \pi x\rangle$. What does this tell us about the functions $\sin \pi x$ and $\cos \pi x$?

4. (a) Give a polynomial $q \in \mathscr{P}_1$ that is orthogonal to $p(x) = 3x - 1$ with the inner product of Example 3.1(c) adapted to $\mathscr{P}_1$.

   (b) Give a polynomial $q \in \mathscr{P}_1$ that is orthogonal to $p(x) = 3x - 1$ with the inner product

$$\langle p, q\rangle = \int_0^1 p(x)\, q(x)\, dx.$$

5. Let $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ represent vectors in $\mathbb{R}^2$, and let $A = \begin{bmatrix} 5 & 2 \\ 2 & 1 \end{bmatrix}$. The product $\mathbf{u}^T A\mathbf{v}$ is

$$\mathbf{u}^T A\mathbf{v} = [u_1 \;\; u_2] \begin{bmatrix} 5 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix},$$

and defines an inner product $\langle \mathbf{u}, \mathbf{v}\rangle$. Use this inner product for the following.

   (a) Find the inner product of $\mathbf{u} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$.

   (b) Find $\|\mathbf{v}\|$ for the vector $\mathbf{v}$ of part (a) and the norm induced by the inner product described above.

   (c) Find a vector $\mathbf{w}$ that is orthogonal to the $\mathbf{u}$ of part (a).

6. Give a specific counterexample showing that binary operation $\langle \mathbf{u}, \mathbf{v}\rangle = \mathbf{u}^T A\mathbf{v}$ is *NOT* an inner product when A is the matrix $\begin{bmatrix} 2 & 2 \\ 3 & 4 \end{bmatrix}$.

7. Give a specific counterexample showing that binary operation $\langle \mathbf{u}, \mathbf{v}\rangle = \mathbf{u}^T A\mathbf{v}$ is *NOT* an inner product when A is the matrix $\begin{bmatrix} 3 & 2 \\ 2 & 1 \end{bmatrix}$.

8. Find the projection of $p(x) = x^2 - 5x + 3$ onto $q(x) = 2x^2 + x - 1$ using the inner product of Example 3.1(c).

9. (a) Find the projection of $f(x) = \sin x$ onto $g(x) = x$ using the integral inner product of Example 3.1(d) with $a = -1$ and $b = 1$. Compute the needed inner products as decimals to the ten-thousandths place (show the values you obtain), then round your final coefficient to the hundredth's place.

   (b) Find the projection of $f(x) = 2x + 3$ onto $g(x) = x$ again using the integral inner product of Example 3.1(d) with $a = -1$ and $b = 1$. Use exact values throughout.

10. Find the projection of $\mathbf{u} = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$ onto $\mathbf{v} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$, using the inner product of Exercise 5 above. Compute all inner products and ensuing values in exact form.

3.2 Orthogonal Bases

Performance Criteria:

3. (d) Determine whether a basis for an inner product space is orthogonal; determine whether a basis is orthonormal.

   (e) In an inner product space, find the representation of a vector with respect to an orthogonal or orthonormal basis without solving a system of equations.

Definition 3.6: Let V be an inner product space. A basis $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ for V is called an **orthogonal basis** if $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ whenever $i \neq j$.

◇ **Example 3.2(a):** Consider $\mathbb{R}^3$ with the standard dot product as the inner product. The set

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \right\}$$

can be shown to be a linearly independent set in $\mathbb{R}^3$ so, by Theorem 1.15(c), it is a basis for $\mathbb{R}^3$. It is also easy to see that the dot product of any two of them is zero, so they are an orthogonal basis.

◇ **Example 3.2(b):** Consider the vector space $\mathscr{P}_2[-1,1]$, the polynomials of degree two or less on the interval $[-1,1]$, with the inner product $\langle p_1, p_2 \rangle = \int_{-1}^{1} p_1(x)\, p_2(x)\, dx$. Because $\mathscr{P}_2$ has dimension three and the set $\{1, x, x^2 - \frac{1}{3}\}$ is linearly independent, it is a basis for $\mathscr{P}_2[-1,1]$. Moreover, we see that

$$\langle 1, x \rangle = \int_{-1}^{1} x\; dx = \frac{x^2}{2}\bigg|_{-1}^{1} = 0,$$

$$\langle 1, x^2 - \tfrac{1}{3} \rangle = \int_{-1}^{1} (x^2 - \tfrac{1}{3})\; dx = \left[\tfrac{1}{3}(x^3 - x)\right]_{-1}^{1} = \tfrac{1}{3}\left\{[1^3 - 1] - [(-1)^3 - (-1)]\right\} = 0,$$

$$\langle x, x^2 - \tfrac{1}{3} \rangle = \int_{-1}^{1} (x^3 - \tfrac{1}{3}x)\; dx = \left[\tfrac{1}{4}x^4 - \tfrac{1}{6}x^2\right]_{-1}^{1} = [(\tfrac{1}{4} - \tfrac{1}{6}) - (\tfrac{1}{4} - \tfrac{1}{6})] = 0.$$

Therefore the set $\{1, x, x^2 - \frac{1}{3}\}$ is an orthogonal basis for $\mathscr{P}_2[-1,1]$ with the given inner product.

Orthogonal bases have a very important feature that we'll illustrate here. Suppose that V is a vector space with an orthogonal basis $\mathcal{B}$ are as in the above definition, and that $\mathbf{u} \in V$ has the following representation with respect to the basis $\mathcal{B}$:

$$\mathbf{u} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_j\mathbf{v}_j + \cdots + c_n\mathbf{v}_n. \tag{1}$$

If we take the inner product of each side of the above with the basis vector $\mathbf{v}_1$ and apply properties of the inner product given in the previous section, the result is as follows:

$$\begin{aligned}
\langle \mathbf{u}, \mathbf{v}_j \rangle &= \langle c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_j\mathbf{v}_j + \cdots + c_n\mathbf{v}_n, \mathbf{v}_j \rangle \\
&= \langle c_1\mathbf{v}_1, \mathbf{v}_j \rangle + \langle c_2\mathbf{v}_2, \mathbf{v}_j \rangle + \cdots + \langle c_j\mathbf{v}_j, \mathbf{v}_j \rangle + \cdots + \langle c_n\mathbf{v}_n, \mathbf{v}_j \rangle \\
&= c_1\langle \mathbf{v}_1, \mathbf{v}_j \rangle + c_2\langle \mathbf{v}_2, \mathbf{v}_j \rangle + \cdots + c_j\langle \mathbf{v}_j, \mathbf{v}_j \rangle + \cdots + c_n\langle \mathbf{v}_n, \mathbf{v}_j \rangle
\end{aligned}$$

Because of the fact that $\mathcal{B}$ is an orthogonal basis, all of the inner products on the right side are zero, with the exception of $\langle \mathbf{v}_j, \mathbf{v}_j \rangle$, giving us $\langle \mathbf{u}, \mathbf{v}_j \rangle = c_j \langle \mathbf{v}_j, \mathbf{v}_j \rangle$. Dividing both sides by $\langle \mathbf{v}_j, \mathbf{v}_j \rangle$ results in the following theorem:

Theorem 3.7: Let V be an inner product space with orthogonal basis $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$. Every $\mathbf{u} \in V$ has the representation

$$\mathbf{u} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_j\mathbf{v}_j + \cdots + c_n\mathbf{v}_n, \tag{1}$$

where each c_j is given by $c_j = \dfrac{\langle \mathbf{u}, \mathbf{v}_j \rangle}{\langle \mathbf{v}_j, \mathbf{v}_j \rangle}$.

The above theorem frees us from the need to solve systems of equations to find the scalars c_j in (1). Instead we can simply compute some inner products. Perhaps most importantly, it gives us a method for finding the scalars for a linear combination in an infinite dimensional vector space.

◇ **Example 3.2(c):** Let $\mathbf{u} = \begin{bmatrix} 3 \\ -6 \\ 2 \end{bmatrix}$ and $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \right\}$. Find the representation of $\mathbf{u}$ with respect to the given basis for $\mathbb{R}^3$.

The set $\mathcal{B}$ is an orthogonal basis for $\mathbb{R}^3$; let's denote its elements by $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$. Then

$$c_1 = \frac{\langle \mathbf{u}, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} = \frac{3+2}{2} = \frac{5}{2}, \qquad c_2 = \frac{\langle \mathbf{u}, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle} = \frac{-6}{1} = -6, \qquad c_3 = \frac{\langle \mathbf{u}, \mathbf{v}_3 \rangle}{\langle \mathbf{v}_3, \mathbf{v}_3 \rangle} = \frac{3-2}{2} = \frac{1}{2},$$

giving the representation

$$\mathbf{u} = \begin{bmatrix} 3 \\ -6 \\ 2 \end{bmatrix} = \frac{5}{2}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - 6\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$

Before continuing we should make a very important observation. If we take the representation (1) for the vector $\mathbf{u}$ and substitute in the expression for each c_j we get

$$\mathbf{u} = \frac{\langle \mathbf{u}, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 + \frac{\langle \mathbf{u}, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\mathbf{v}_2 + \cdots + \frac{\langle \mathbf{u}, \mathbf{v}_j \rangle}{\langle \mathbf{v}_j, \mathbf{v}_j \rangle}\mathbf{v}_j + \cdots + \frac{\langle \mathbf{u}, \mathbf{v}_n \rangle}{\langle \mathbf{v}_n, \mathbf{v}_n \rangle}\mathbf{v}_n \tag{1}$$

Note that each part $\dfrac{\langle \mathbf{u}, \mathbf{v}_j \rangle}{\langle \mathbf{v}_j, \mathbf{v}_j \rangle}\mathbf{v}_j$ of the above sum is the projection of $\mathbf{u}$ onto the basis vector $\mathbf{v}_j$!

We can carry the idea of an orthogonal basis a bit farther:

Definition 3.8: Let V be an inner product space. A basis $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ for V is called an **orthonormal basis** if it is an orthogonal basis and the norm of each $\mathbf{v}_j$ is one.

If each $\mathbf{v}_j$ has norm one, then for each we have $\langle \mathbf{v}_j, \mathbf{v}_j \rangle = 1$ and Theorem 3.6 simplifies to the following.

Theorem 3.9: Let V be an inner product space with orthonormal basis $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$. Every $\mathbf{u} \in V$ has the representation

$$\mathbf{u} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_j\mathbf{v}_j + \cdots + c_n\mathbf{v}_n,$$

where each c_j is given by $c_j = \langle \mathbf{u}, \mathbf{v}_j \rangle$.

If we have an orthogonal basis, it is easy to obtain an orthonormal basis. For any vector $\mathbf{v}$ the vector $\frac{1}{\|\mathbf{v}\|}\mathbf{v}$ has the same "direction" as $\mathbf{v}$, but has norm one. The process of doing this with a vector is often referred to as **normalizing the vector**.

◇ **Example 3.2(d):** Normalize each of the polynomials (using the inner product from Example 3.2(b)) in the orthogonal basis $\{1, x, x^2 - \frac{1}{3}\}$ for $\mathscr{P}_2[-1, 1]$ to obtain an orthonormal basis.

Let's begin by finding the inner product of each polynomial in the basis with itself

$$\langle 1, 1\rangle = \int_{-1}^{1} 1\,dx = x\Big|_{-1}^{1} = 2,$$

$$\langle x, x\rangle = \int_{-1}^{1} x^2\,dx = \tfrac{1}{3}x^3\Big|_{-1}^{1} = \tfrac{1}{3}\left[1^3 - (-1)^3\right] = \frac{2}{3},$$

$$\langle x^2 - \tfrac{1}{3}, x^2 - \tfrac{1}{3}\rangle = \int_{-1}^{1} (x^2 - \tfrac{1}{3})^2\,dx = \int_{-1}^{1} (x^4 - \tfrac{2}{3}x^2 + \frac{1}{9})\,dx = \left[\tfrac{1}{5}x^5 - \tfrac{2}{9}x^3 + \tfrac{1}{9}x\right]_{-1}^{1} = \tfrac{8}{45}.$$

The norms of $1, x, x^2 - \frac{1}{3}$ are then obtained by taking the square root of each of the above values to get the respective norms $\sqrt{2}, \sqrt{\frac{2}{3}}, \frac{2\sqrt{2}}{3\sqrt{5}}$. We then divide each of $1, x, x^2 - \frac{1}{3}$ by the above norms of each of the above values (which is of course equivalent to multiplying by their reciprocals) to get the orthonormal basis

$$\left\{\frac{1}{\sqrt{2}},\ \sqrt{\frac{3}{2}}x,\ \sqrt{\frac{45}{8}}\left(x^2 - \frac{1}{3}\right)\right\}.$$

Section 3.2 Exercises

1. Consider the vectors $\mathbf{u} = \begin{bmatrix} 7 \\ -1 \end{bmatrix}$, $\mathbf{v}_1 = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ in $\mathbb{R}^2$. Hopefully you can see quite easily that $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal.

   (a) Find a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$ that equals $\mathbf{u}$ by setting up an equation representing that and solving the resulting system of equations. Conclude by giving $\mathbf{u}$ as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$.

   (b) Show clearly how to use Theorem 3.6 to find the scalars c_1 and c_2 for which $\mathbf{u} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$. Do your c_1 and c_2 equal those you found in (a)? If not, find your error!

2. The vectors $\mathbf{v}_1 = \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix}$ in $\mathbb{R}^3$ are orthogonal.

   (a) Find a third vector $\mathbf{v}_3$ that is orthogonal to both of those. You could do this one of two ways: (1) By being observant about the two that are given and making an educated guess for the third and checking it or (2) by finding the cross product of those two, which is a third vector that is orthogonal to both of them.

   (b) "Normalize" the three vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ to get an orthonormal basis $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$.

   (c) Use Theorem 3.8 to obtain a linear combination of your orthonormal basis elements from (b) that equals the vector $\mathbf{w} = \begin{bmatrix} 3 \\ -5 \\ 1 \end{bmatrix}$.

3. The vectors $\begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$ are orthogonal. Use Theorem 3.6 to find values of c_1, c_2 and c_3 such that

$$\begin{bmatrix} 7 \\ -4 \\ 6 \end{bmatrix} = c_1 \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} + c_3 \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

4. Example 3.2(b) showed that the set $\{1, x, x^2 - \frac{1}{3}\}$ is an orthogonal basis for $\mathscr{P}_2[-1,1]$ with the inner product $\langle f, g\rangle = \displaystyle\int_{-1}^{1} f(x)\, g(x)\, dx$. Let $\mathscr{C}[-1,1]$ be the set of continuous functions on the interval $[-1,1]$. This is an infinite dimensional vector space, so we can't represent functions with finite linear combinations of basis vectors, *but we CAN approximate functions* by forming infinite linear combinations of the form

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 + \cdots$$

and just using the first few terms. Use Theorem 3.6 to find a three term approximation of e^x with respect to the basis $\{1, x, x^2 - \frac{1}{3}\}$. Use your calculator to compute coefficients and round each coefficient to two places past the decimal. When you have your approximation, simplify it, then graph it and e^x together over the interval $[-2,2]$ to see how good the approximation is over $[-1,1]$.

5. Let $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ represent vectors in $\mathbb{R}^2$. For the matrix $A = \begin{bmatrix} 5 & 2 \\ 2 & 1 \end{bmatrix}$, define

$$\langle \mathbf{u}, \mathbf{v}\rangle = \mathbf{u}^T A\mathbf{v} = \begin{bmatrix} u_1 & u_2 \end{bmatrix} \begin{bmatrix} 5 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

for any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$. Let $\mathbf{u} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$.

   (a) Find the norm of each vector with respect to the given inner product. *Give your answers in exact form.*

   (b) Find $\langle \mathbf{u}, \mathbf{v}\rangle$, $\langle \mathbf{u}, \mathbf{w}\rangle$ and $\langle \mathbf{v}, \mathbf{w}\rangle$.

   (c) Two of the vectors are orthogonal. Which are they?

   (d) Normalize the orthogonal vectors by multiplying each by the reciprocal of its norm to obtain an orthonormal basis. Give it, labelling it $\mathcal{B}$.

   (e) Use Theorem 3.8 to find the representation of $\begin{bmatrix} 2 \\ -3 \end{bmatrix}$ with respect to $\mathcal{B}$. Write your answer as a linear combination.

6. Again, the vectors $\begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$ are orthogonal.

   (a) Normalize each of the three vectors, giving your answers in decimal form, rounded to the nearest thousandth.

   (b) Create a matrix Q whose columns are the three normalized vectors. Use some technology to find Q^{-1}. Realizing that neither Q nor Q^{-1} are exact, what do you think the relationship between Q and Q^{-1} is?

   (c) Make up a vector $\mathbf{v}$ in $\mathbb{R}^3$ and find its norm, $\|\mathbf{v}\|$, to the thousandths place. Then calculate $Q\mathbf{v}$ and $\|Q\mathbf{v}\|$. What do you notice?

   (d) Use MATLAB or your calculator to find $\det(Q)$, the determinant of Q. What is it?

3.3 Orthogonal Subspaces

Performance Criteria:

3. (f) Given a basis for a subspace W of $\mathbb{R}^n$ and a $\mathbf{v}$ in $\mathbb{R}^n$, find the unique orthogonal decomposition $\mathbf{v} = \mathbf{w} + \mathbf{w}^\perp$ with $\mathbf{w}$ in W and $\mathbf{w}^\perp$ in $W^\perp$.

Definition 3.10: Let V be an inner product space, and let W be a subspace of V. The **orthogonal complement** of W is the set

$$W^\perp = \{\mathbf{u} \in V \mid \langle \mathbf{u}, \mathbf{w} \rangle = 0 \text{ for all } \mathbf{w} \in W\}.$$

Finding a good description of the orthogonal complement of a given subspace can be a bit complicated. In the next section we'll see a method for finding a basis for the orthogonal subspace, which of course is as good of a description of a space as we can provide. In this section we'll stick with a few relatively simple examples, and we'll introduce a concept or two that will be needed in the next section. To help us there are a number of facts that hold for $W^\perp$ and W, which we summarize in the next theorem.

Theorem 3.11: Let V be an inner product space, with W a subspace of V.

a) $W^\perp$ is a subspace of V.

b) $(W^\perp)^\perp = W$

c) $W \cap W^\perp = \mathbf{0}$

d) $W \cup W^\perp = V$

e) $\dim(W) + \dim(W^\perp) = \dim(V)$

◇ **Example 3.3(a):** Let W be the span of $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ and $\mathbf{u}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ in $\mathbb{R}^3$, with the standard inner product (dot product). The span of a set is always a subspace, so this is a subspace of $\mathbb{R}^3$, and it has dimension two because the two vectors are independent. The vector $\mathbf{u}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ can easily be seen to be orthogonal to both u_1 and u_2. We know from part (e) of the above theorem that $W^\perp$ has dimension one, so $\text{span}(\mathbf{u}_3)$ is the orthogonal complement of W.

The orthogonal complement of a subspace is a very geometric idea. With a bit of thought we realize that the subspace W of the previous example is the xy-plane in $\mathbb{R}^3$, and the orthogonal complement is the z-axis.

Here's an example where we have to get a bit fancier to find the orthogonal complement:

◇ **Example 3.3(b):** Let W be the span of $\mathbf{u}_1 = \begin{bmatrix} 1 \\ -4 \\ 2 \end{bmatrix}$ and $\mathbf{u}_2 = \begin{bmatrix} 3 \\ 0 \\ -5 \end{bmatrix}$ in $\mathbb{R}^3$, with the standard inner product (dot product). W is a subspace of $\mathbb{R}^3$; geometrically, it is a plane through the origin. In vector calculus you learned that the cross product of two vectors is another vector that is perpendicular to each of the original vectors, so it is perpendicular to the plane containing them. Therefore $\mathbf{w} = \mathbf{u} \times \mathbf{v} = \begin{bmatrix} -20 \\ 1 \\ 12 \end{bmatrix}$ is orthogonal to W, as is any scalar multiple of $\mathbf{u} \times \mathbf{v}$. Therefore $W^\perp$ is the span of $\mathbf{w}$.

◇ **Example 3.3(c):** Let $\mathcal{B} = \{\mathbf{e}_1, \mathbf{e}_2, ...\mathbf{e}_5\}$ be the standard basis for $\mathbb{R}_5$, with the standard dot product as the inner product. Note that this is an orthogonal basis, so $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = 0$ if $i \neq j$. If we let $W = \text{span}(\{\mathbf{e}_1, \mathbf{e}_2\})$, then the remaining basis vectors for $\mathbb{R}^5$ are all orthogonal to all vectors in W, so $W^\perp = \text{span}(\{\mathbf{e}_3, \mathbf{e}_4, \mathbf{e}_5\})$.

Our primary interest in orthogonal subspaces is that the will allow us to "decompose" a given vector in an inner product space V into two parts that are orthogonal to each other, one part in some subspace W of V, and the other part in the orthogonal complement of W. To see how to do that, we first need the idea of projecting a vector not onto another vector, but onto a subspace.

Definition 3.12: Let V be an inner product space, and let W be a subspace of V with orthogonal basis $\mathcal{B} = \{\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_n\}$. For any $\mathbf{u} \in V$, the **orthogonal projection of u on** W is

$$\text{proj}_W \mathbf{u} = \frac{\langle \mathbf{u}, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle} \mathbf{w}_1 + \frac{\langle \mathbf{u}, \mathbf{w}_2 \rangle}{\langle \mathbf{w}_2, \mathbf{w}_2 \rangle} \mathbf{w}_2 + \cdots + \frac{\langle \mathbf{u}, \mathbf{w}_n \rangle}{\langle \mathbf{w}_n, \mathbf{w}_n \rangle} \mathbf{w}_n$$

and the **component of u orthogonal to** W is

$$\text{perp}_W \mathbf{u} = \mathbf{u} - \text{proj}_W \mathbf{u}.$$

Theorem 3.13: Let V be an inner product space, and let W be a subspace of V. For any $\mathbf{u} \in V$ there exist unique vectors $\mathbf{w} \in W$ and $\mathbf{w}^\perp \in W^\perp$ such that

$$\mathbf{u} = \mathbf{w} + \mathbf{w}^\perp.$$

In particular, $\mathbf{w} = \text{proj}_W \mathbf{u}$ and $\mathbf{w}^\perp = \text{perp}_W \mathbf{u}$.

◇ **Example 3.3(d):** Let W be the span of the orthogonal vectors $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$ and $\mathbf{u}_2 = \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix}$ in $\mathbb{R}^3$, with the standard inner product (dot product). Note that $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal. For the vector $\mathbf{v} = \begin{bmatrix} -4 \\ 5 \\ 2 \end{bmatrix}$, find vectors $\mathbf{w} \in W$ and $\mathbf{w}^\perp \in W^\perp$ such that $\mathbf{v} = \mathbf{w} + \mathbf{w}^\perp$.

First we project $\mathbf{v}$ onto W to find $\mathbf{w}$:

$$\mathbf{w} = \text{proj}_W \mathbf{v} = \frac{\langle \mathbf{v}, \mathbf{u}_1 \rangle}{\langle \mathbf{u}_1, \mathbf{u}_1 \rangle} \mathbf{u}_1 + \frac{\langle \mathbf{v}, \mathbf{u}_2 \rangle}{\langle \mathbf{u}_2, \mathbf{u}_2 \rangle} \mathbf{u}_2 = \frac{5}{6} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} + \frac{-10}{5} \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix} = \frac{1}{6} \begin{bmatrix} -19 \\ 5 \\ 22 \end{bmatrix}.$$

We then obtain

$$\mathbf{w}^\perp = \text{perp}_W \mathbf{v} = \mathbf{v} - \text{proj}_W \mathbf{v} = \begin{bmatrix} -4 \\ 5 \\ 2 \end{bmatrix} - \frac{1}{6} \begin{bmatrix} -19 \\ 5 \\ 22 \end{bmatrix} = \frac{5}{6} \begin{bmatrix} -1 \\ 5 \\ -2 \end{bmatrix}$$

It is clear that $\mathbf{v} = \mathbf{w} + \mathbf{w}^\perp$. Let's verify that $\mathbf{w}$ and $\mathbf{w}^\perp$ are indeed orthogonal to each other:

$$\mathbf{w}^T A \mathbf{w}^\perp = \begin{bmatrix} -4 & 5 & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ \frac{5}{3} \end{bmatrix} = \begin{bmatrix} -\frac{52}{3} & \frac{56}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ \frac{5}{3} \end{bmatrix} = 0.$$

Therefore $\mathbf{w}$ and $\mathbf{w}^\perp$ are the orthogonal components of $\mathbf{v}$ in W and $W^\perp$, respectively.

Section 3.3 Exercises

When working in $\mathbb{R}^n$, assume the inner product is the standard dot product unless told other wise.

1. In $\mathbb{R}^2$, let $W = \text{span}\left(\begin{bmatrix} 3 \\ 2 \end{bmatrix}\right)$.

   (a) Describe the orthogonal complement $W^\perp$ as the span of a vector or set of vectors. Then draw a picture that shows both W and $W^\perp$.

   (b) Find the vectors $\mathbf{w}$ and $\mathbf{w}^\perp$ of Theorem 3.13 for the vector $\mathbf{u} = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$. Give your answers in exact form, then with decimal components rounded to the nearest tenth.

   (c) Verify that your vectors from part (b) satisfy the condition that $\mathbf{w} + \mathbf{w}^\perp = \mathbf{u}$.

   (d) Add $\mathbf{w}$ and $\mathbf{w}^\perp$ to your graph from part (a).

2. In $\mathbb{R}^3$, let $W = \text{span}\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}\right)$.

   (a) Give a vector whose span is the orthogonal complement $W^\perp$ of W. You should be able to do this by inspection.

   (b) Note that the two given vectors whose span is W are orthogonal to each other, so we can use Definition 3.12 to find the vectors $\mathbf{w}$ and $\mathbf{w}^\perp$ of Theorem 3.13. Do so.

3. In $\mathbb{R}^3$, let $W = \text{span}\left(\begin{bmatrix} 3 \\ -2 \\ 5 \end{bmatrix}\right)$. Describe the orthogonal complement $W^\perp$ in the same way. (**Hint:** Any two vectors that are both orthogonal to the given vector and independent will form a basis for $W^\perp$.)

4. Again let $W = \text{span}\left(\begin{bmatrix} 3 \\ -2 \\ 5 \end{bmatrix}\right)$ in $\mathbb{R}^3$, but with the weighted inner product $\langle \mathbf{u}, \mathbf{v} \rangle = 3u_1v_1 + u_2v_2 + 2u_3v_3$. Give the orthogonal complement $W^\perp$ as the span of a set of vectors.

5. The set $\mathcal{B} = \{x,\ 3x^2 - 1,\ 5x^3 - 3x\}$ is an orthogonal set in $\mathscr{C}[-1, 1]$ with the inner product $\langle f, g \rangle = \int_{-1}^{1} f(x)\, g(x)\, dx$, so $W = \text{span}(\mathcal{B})$ is a subspace with orthogonal basis $\mathcal{B}$. Find the projection of $f(x) = e^{-3x}$ onto W. You should use your calculator to compute integrals, and **find all coefficients as decimal approximations, rounded to four places past the decimal.** Combine like terms in the resulting polynomial.

6. For the matrix $A = \begin{bmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{bmatrix}$ and any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$, define $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T A \mathbf{v}$, as in Example 3.3(d). For this exercise we "are in" the inner product space of $\mathbb{R}^3$ *with this inner product* so you must always use this inner product when finding norms or inner products.

   (a) Find the norm of $\mathbf{u} = \begin{bmatrix} -5 \\ 2 \\ 4 \end{bmatrix}$.

   (b) Find the projection of the $\mathbf{u}$ from part (a) onto $\mathbf{w} = \begin{bmatrix} 3 \\ 1 \\ -2 \end{bmatrix}$.

   (c) Verify that $\mathbf{v} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} 3 \\ 1 \\ -2 \end{bmatrix}$ are orthogonal.

   (d) $W = \text{span}(\{\mathbf{v}, \mathbf{w}\})$ is a subspace of $\mathbb{R}^3$ and those two vectors form an orthogonal basis. Find the projection of $\mathbf{u}$ from part (a) onto W.

7. Consider the inner product given in Example 3.1(c), and let $W = \text{span}(\{x^2 + 5x - 2, 3x + 1\})$.

   (a) What is the dimension of W? What is the dimension of $W_\perp$?

   (b) Find a basis for $W_\perp$. (**Hint:** One of the examples in the section is useful.)

8. Let $\mathcal{S} = \left\{ \begin{bmatrix} 0 \\ -5 \\ 1 \\ 0 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ 2 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ -2 \\ 0 \end{bmatrix} \right\}$ in $\mathbb{R}^6$ with the standard inner product.

   (a) The span of a set of vectors is *always* a subspace, so $W = \text{span}(\mathcal{S})$ is a subspace of $\mathbb{R}^6$. Is $\mathcal{S}$ a basis for that subspace? Justify your answer with a brief explanation.

   (b) Give the simplest basis for W that you can think of.

   (c) Give a basis for $W_\perp$.

3.4 Gram-Schmidt Orthogonalization

Performance Criteria:

3. (g) Apply the Gram-Schmidt process to a set of vectors in an inner product space to obtain an orthogonal basis; normalize a vector or set of vectors in an inner product space.

In this section we develop the Gram-Schmidt process, which uses a basis for a vector space to create an orthogonal basis for the space. The fundamental idea is that if we have a subspace W with an orthogonal basis and some vector $\mathbf{u}$ not in W, the vector $\text{perp}_W\mathbf{u} = \mathbf{u} - \text{proj}_W\mathbf{u}$ is orthogonal to W. So let us begin with a subspace W of a vector space V, and suppose that $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k\}$ is a basis for W. We will construct from this basis an orthogonal basis $\mathcal{C} = \{\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_k\}$ for W.

- First we let $\mathbf{w}_1 = \mathbf{v}_1$, and we define the subspace $W_1 = \text{span}(\mathbf{v}_1) = \text{span}(\mathbf{w}_1)$.
- Because $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k\}$ is a basis, $\mathbf{v}_2$ is linearly independent of $\mathbf{v}_1$. Thus the vector defined by

$$\mathbf{w}_2 = \text{perp}_{W_1}\mathbf{v}_2 = \mathbf{v}_2 - \text{proj}_{W_1}\mathbf{v}_2 = \mathbf{v}_2 - \frac{\langle \mathbf{v}_2, \mathbf{w}_1\rangle}{\langle \mathbf{w}_1, \mathbf{w}_1\rangle}\,\mathbf{w}_1$$

  is a nonzero vector that is orthogonal to $W_1 = \text{span}(\mathbf{w}_1)$. We then define $W_2 = \text{span}(\mathbf{v}_1, \mathbf{v}_2) = \text{span}(\mathbf{w}_1, \mathbf{w}_2)$.
- Next we find a third vector $\mathbf{w}_3$ that is orthogonal to both $\mathbf{w}_2$ and $\mathbf{w}_1$:

$$\begin{aligned}\mathbf{w}_3 = \text{perp}_{W_2}\mathbf{v}_3 &= \mathbf{v}_3 - \text{proj}_{W_2}\mathbf{v}_3\\ &= \mathbf{v}_3 - \frac{\langle \mathbf{v}_3, \mathbf{w}_1\rangle}{\langle \mathbf{w}_1, \mathbf{w}_1\rangle}\,\mathbf{w}_1 - \frac{\langle \mathbf{v}_3, \mathbf{w}_2\rangle}{\langle \mathbf{w}_2, \mathbf{w}_2\rangle}\,\mathbf{w}_2\end{aligned}$$

  Note what has happened here: We have taken the vector $\mathbf{v}_3$ and removed the components of it in the directions of $\mathbf{w}_1$ and $\mathbf{w}_2$. Now we let $W_3 = \text{span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) = \text{span}(\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3)$.
- We continue on in this manner. For each new $\mathbf{v}_j$ we subtract off its projections onto each of the previous $\mathbf{w}_i$ to obtain $\mathbf{w}_j$. The final W_k is equal to W.

Theorem 3.14: Let V be an inner product space and let W be a subspace of V with basis $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k\}$. The set $\mathcal{C} = \{\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_k\}$ created in the manner above is an orthogonal basis for W.

◇ **Example 3.4(a):** The vectors $\mathbf{u}_1 = \begin{bmatrix}1\\1\\0\end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix}1\\0\\1\end{bmatrix}$, $\mathbf{u}_3 = \begin{bmatrix}0\\1\\1\end{bmatrix}$ are linearly independent, so they form a basis for $\mathbb{R}^3$. Perform the Gram-Schmidt process on this set to obtain an orthogonal basis, using the standard dot product for $\mathbb{R}^3$.

We take our first basis vector $\mathbf{w}_1$ to be the vector $\mathbf{u}_1$. Next we obtain

$$\mathbf{w}_2 = \mathbf{u}_2 - \text{proj}_{\mathbf{u}_1}\mathbf{u}_2 = \mathbf{u}_2 - \frac{\langle \mathbf{u}_2, \mathbf{w}_1\rangle}{\langle \mathbf{w}_1, \mathbf{w}_1\rangle}\,\mathbf{w}_1 = \begin{bmatrix}1\\0\\1\end{bmatrix} - \tfrac{1}{2}\begin{bmatrix}1\\1\\0\end{bmatrix} = \begin{bmatrix}\frac{1}{2}\\-\frac{1}{2}\\1\end{bmatrix}$$

and

$$\mathbf{w}_3 = \mathbf{u}_3 - \frac{\langle \mathbf{u}_3, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\,\mathbf{w}_1 - \frac{\langle \mathbf{u}_3, \mathbf{w}_2 \rangle}{\langle \mathbf{w}_2, \mathbf{w}_2 \rangle}\,\mathbf{w}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} - \tfrac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - \frac{\frac{1}{2}}{\frac{3}{2}}\begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{bmatrix} = \begin{bmatrix} -\frac{2}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix}$$

$\mathcal{C} = \{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ is an orthogonal basis for $\mathbb{R}^3$.

◇ **Example 3.4(b):** Apply the Gram-Schmidt process to the basis $\mathcal{B} = \{1, x, x^2, x^3\}$ for the subspace $\mathscr{P}_3[-1,1]$ of $\mathscr{C}[-1,1]$, with the inner product $\langle f, g\rangle = \displaystyle\int_{-1}^{1} f(x)\,g(x)\,dx$.

We take our first basis vector to be $\mathbf{w}_1 = 1$, and then subtract the projection of x onto 1 from x to get our second basis element:

$$\mathbf{w}_2 = x - \text{proj}_1 x = x - \frac{\langle x, 1\rangle}{\langle x, x\rangle}\,(1) = x - \frac{0}{\frac{2}{3}} = x.$$

Next we have

$$\mathbf{w}_2 = x^2 - \text{proj}_1 x^2 - \text{proj}_x x^2 = x^2 - \frac{\langle x^2, 1\rangle}{\langle 1, 1\rangle}\,(1) - \frac{\langle x^2, x\rangle}{\langle x, x\rangle}\,x = x^2 - \frac{\frac{2}{3}}{2}\,(1) - \frac{0}{\frac{2}{3}}\,x = x^2 - \tfrac{1}{3}$$

And for our final vector we have

$$\begin{aligned}
\mathbf{w}_3 &= x^3 - \text{proj}_1 x^3 - \text{proj}_x x^3 - \text{proj}_{(x^2-1/3)} x^3 \\
&= x^3 - \frac{\langle x^3, 1\rangle}{\langle 1, 1\rangle}\,(1) - \frac{\langle x^3, x\rangle}{\langle x, x\rangle}\,x - \frac{\langle x^3, x^2 - \frac{1}{3}\rangle}{\langle x^2 - \frac{1}{3}, x^2 - \frac{1}{3}\rangle}\,(x^2 - \tfrac{1}{3}) \\
&= x^3 - \frac{0}{2}\,(1) - \frac{\frac{2}{5}}{\frac{2}{3}}\,x - \frac{0}{\frac{8}{45}}\,(x^2 - \tfrac{1}{3}) \\
&= x^3 - \tfrac{3}{5}x
\end{aligned}$$

The set $\mathcal{C} = \{1, x, x^2 - \frac{1}{3}, x^3 - \frac{3}{5}x\}$ is an orthogonal basis for $\mathscr{P}_3[-1,1]$ with the given inner product.

The polynomials found in the last example are called the **Legendre polynomials**.

Section 3.4 Exercises

1. The set of vectors $\mathbf{v}_1 = \begin{bmatrix} -3 \\ 5 \\ 1 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 4 \\ 0 \\ 2 \end{bmatrix}$ and $\mathbf{v}_3 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$ are linearly independent, so they form a basis for $\mathbb{R}^3$. In this exercise you will

   (1) perform the Gram-Schmidt process on the set of vectors to get an orthogonal basis $\mathcal{B}_1 = \{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ for $\mathbb{R}^3$, then

   (2) normalize each of those vectors to obtain an orthonormal basis $\mathcal{B}_2 = \{\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3\}$ for $\mathbb{R}^3$.

   NOTE: The inner product to be used in this exercise is the standard dot product for $\mathbb{R}^3$.

   (a) The orthogonal basis we will create will consist of the vectors $\mathbf{w}_1$, $\mathbf{w}_2$ and $\mathbf{w}_3$, which is the same notation as used in the description of the process at the start of the section. Then $\mathbf{w}_1$ is just $\mathbf{v}_1$.

   (b) $\mathbf{w}_2$ will be $\mathbf{v}_2$ with its part in the direction of $\mathbf{w}_1$ removed; that is, $\mathbf{w}_2 = \mathbf{v}_2 - \text{proj}_{\mathbf{w}_1}(\mathbf{v}_2)$. Find $\mathbf{w}_2$ in decimal form, with each component rounded to the thousandth's place.

(c) $\mathbf{w}_3$ will be $\mathbf{v}_3$ with its parts in the directions of $\mathbf{w}_1$ and $\mathbf{w}_2$ removed; that is, $\mathbf{w}_3 = \mathbf{v}_3 - \text{proj}_{\mathbf{w}_1}(\mathbf{v}_3) - \text{proj}_{\mathbf{w}_2}(\mathbf{v}_3)$. Find $\mathbf{w}_3$ in decimal form, with each component rounded to the thousandth's place.

(d) Normalize each of the three vectors $\mathbf{w}_1$, $\mathbf{w}_2$ and $\mathbf{w}_3$ to get an orthonormal basis for $\mathbb{R}^3$ consisting of three unit vectors $\mathbf{q}_1$, $\mathbf{q}_2$ and $\mathbf{q}_3$. Again, round to the thousandth's place.

(e) You can check your answer as follows: Let your vectors be the columns of a matrix Q. Determine the transpose Q^T of Q and then find the product Q^TQ. You should get something close to the identity matrix.

2. The polynomials $p_1(x) = x^2 + x + 2$, $p_2(x) = 3x + 1$, $p_3(x) = x^2 - 2x$ are linearly independent in $\mathscr{P}_2$. Using the inner product given in Example 3.1(c), perform the Gram-Schmidt process to these polynomials to obtain an orthogonal basis.

3. Consider again the set of vectors $\mathbf{v}_1 = \begin{bmatrix} -3 \\ 5 \\ 1 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 4 \\ 0 \\ 2 \end{bmatrix}$ and $\mathbf{v}_3 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$ which are a basis for $\mathbb{R}^3$. Perform the Gram-Schmidt process on the vectors with the inner product $\langle \mathbf{u}, \mathbf{v} \rangle = 3u_1v_1 + u_2v_2 + 2u_3v_3$ in order to form a basis that is orthogonal with respect to that inner product.

4. For the matrix $A = \begin{bmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{bmatrix}$ and any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$, define $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T A \mathbf{v}$. For this exercise we "are in" the inner product space of $\mathbb{R}^3$ *with this inner product* so you must always use this inner product when finding norms or inner products. The three vectors $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ are linearly independent, so they form a basis for $\mathbb{R}^3$. Apply the Gram-Schmidt process to obtain an orthogonal basis.

3.5 Distance and Approximation

Performance Criteria:

3. (h) Find the norm of a vector, or the distance between two vectors, using a given norm.

Norms

In Section 3.1 we used an inner product to define the norm of a vector in an inner product space, with the norm being a function that assigns to every vector in the space a real number that is its "size," in some sense. That definition is always valid when we have an inner product, but what about vector spaces without inner products? In such spaces we can still have a norm, as long as it exhibits to the following properties.

Definition 3.15: A **norm** on a vector space V is a mapping that associates with each vector $\mathbf{v}$ a real number $\|\mathbf{v}\|$, called the **norm** of $\mathbf{v}$, such that the following properties are satisfied for all vectors $\mathbf{u}$ and $\mathbf{v}$ and all scalars c.

1) $\|\mathbf{v}\| \geq 0$, and $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = \mathbf{0}$

2) $\|c\mathbf{v}\| = |c|\|\mathbf{v}\|$

3) $\|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$

A vector space with a norm is called a **normed linear space**.

The next comment is important enough that I think I'll put a box around it!

Just as the same vector space can have different inner products defined on it, *the same vector space can have several different norms defined on it*; thus the same vector space can become more than one normed linear space.

The following definitions illustrate the above comment and give us three important examples of norms.

Definition 3.16: Let $\mathbf{v} = [v_1, v_2, ..., v_n]$ be a vector in $\mathbb{R}^n$.

a) The **2-norm**, or **Euclidean norm**, of $\mathbf{v}$ is defined by

$$\|\mathbf{v}\|_2 = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$$

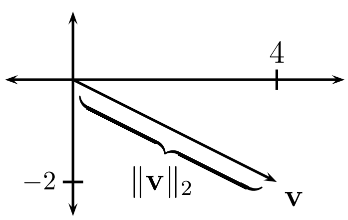
b) The **1-norm**, or **sum norm**, of $\mathbf{v}$ is defined by

$$\|\mathbf{v}\|_1 = |v_1| + |v_2| + \cdots + |v_n|$$

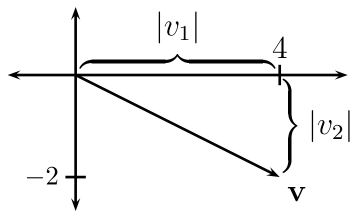
c) The **∞-norm**, or **max norm**, of $\mathbf{v}$ is defined by

$$\|\mathbf{v}\|_\infty = \max\{|v_1|, |v_2|, ..., |v_n|\}$$

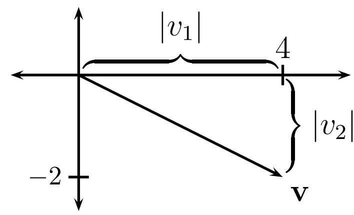
Let's see what these norms look like in $\mathbb{R}^2$. The pictures below show the vector $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 4 \\ -2 \end{bmatrix}$ and the graphical interpretation of each of the three norms from Definition 3.16.



$$\|\mathbf{v}\|_2 = \sqrt{v_1^2 + v_2^2} = \sqrt{4^2 + (-2)^2} = \sqrt{20}$$



$$\|\mathbf{v}\|_1 = |v_1| + |v_2| = |4| + |-2| = 6$$



$$\|\mathbf{v}\|_\infty = \max\{|v_1|, |v_2|\} = \max\{|4|, |-2|\} = 4$$

◇ **Example 3.5(a):** For the vector $\mathbf{u} = \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix}$ in $\mathbb{R}^3$, find the 1-norm, 2-norm and ∞-norm.

$$\|\mathbf{u}\|_1 = |2| + |-3| + |1| = 6, \qquad \|\mathbf{u}\|_2 = \sqrt{2^2 + (-3)^2 + 1^2} = \sqrt{14}, \qquad \|\mathbf{u}\|_\infty = \max\{|2|, |-3|, |1|\} = 3$$

The names 1-norm and 2-norm can be explained as follows. For any $p \geq 1$ the expression

$$\|\mathbf{v}\|_p = \sqrt[p]{|v_1|^p + |v_2|^p + \cdots + |v_n|^p} \tag{1}$$

defines a norm. In the cases where $p = 1$ and $p = 2$ this gives us the 1- and 2-norms. You will see in the exercises why the ∞-norm is named that.

Norms of Functions

For functions we define similar norms. Let f be a function in an "appropriate" vector space of functions that are defined on some interval $[a, b]$, and let p be a real number greater than or equal to one. Then we define

$$\|f\|_p = \left(\frac{1}{b-a} \int_a^b |f(x)|^p \, dx \right)^{\frac{1}{p}} \tag{2}$$

Note the similarity between the definitions (1) and (2). The cases $p = 1$ and $p = 2$ give us the one and two norms for functions:

$$\|f\|_1 = \frac{1}{b-a} \int_a^b |f(x)| \, dx \qquad\qquad \|f\|_2 = \left(\frac{1}{b-a} \int_a^b [f(x)]^2 \, dx \right)^{\frac{1}{2}}$$

The 2-norm is induced by the inner product defined by $\langle f, g \rangle = \dfrac{1}{b-a} \displaystyle\int_a^b f(x)\, g(x) \, dx$ and is commonly referred to by mathematicians as the L_2-norm ("ell-two norm"). In applications it is the **root mean square** (RMS) average of the function f. Sometimes the factor of $1/(b-a)$ is omitted.

◇ **Example 3.5(b):** Find the 1- and 2-norms of $f(x) = \sin x$ in the space $\mathscr{C}[0, 2\pi]$.

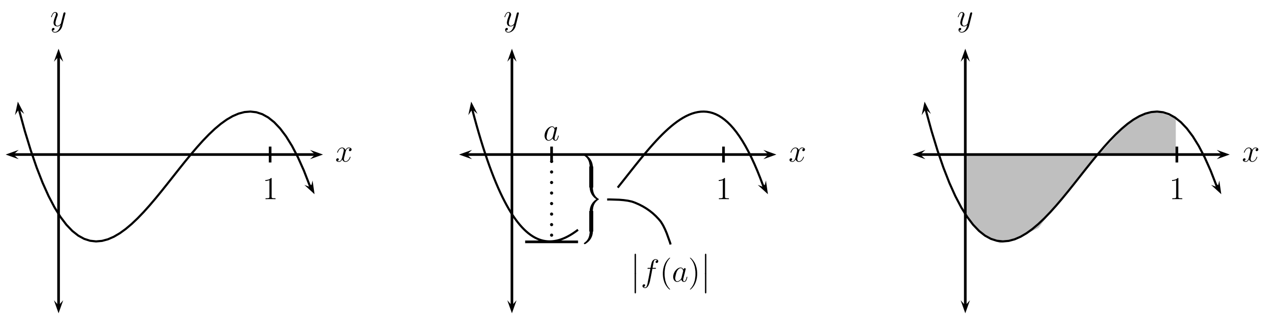
$$\|f\|_1 = \frac{1}{2\pi} \int_0^{2\pi} |\sin x| \, dx = \frac{1}{\pi} \int_0^{\pi} \sin x \, dx = \frac{2}{\pi}, \qquad \|f\|_2 = \left(\frac{1}{2\pi} \int_0^{2\pi} \sin^2 x \, dx \right)^{\frac{1}{2}} = \frac{1}{\sqrt{2}}$$

The ∞-norm of a function $f \in \mathscr{C}[a,b]$ is defined to be $\|f\|_\infty = \max\{|f(x)| : x \in [a,b]\}$. This is sometimes written instead as $\|f\|_\infty = \max\limits_{x\in[a,b]}\{|f(x)|\}$.

◇ **Example 3.5(c):** Find $\|\sin x\|_\infty$ on the interval $[0, 2\pi]$.

Because the maximum and minimum values of $\sin x$ on the interval $[0, 2\pi]$ are one and negative one, the maximum value of $|\sin x|$ is one. Therefore $\|\sin x\|_\infty = 1$.

It's possible to interpret the 1- and ∞-norms of functions graphically. Consider the function whose graph whose shown below and to the left, on the interval $[0,1]$. In the center below see see that the largest value of $|f(x)|$ occurs at $x = a$, where the function has a minimum value. The ∞-norm of f is then $|f(a)|$. The graph below and to the right has the area between the function and the x-axis shaded, and that total area is the 1-norm of f.



$\|f\|_\infty = \max\limits_{x\in[0,1]}\{|f(x)|\} = |f(a)|$ $\qquad$ $\|f\|_1 =$ the total shaded area

Distance

In many applications we approximate quantities or functions, and it is generally desirable to measure the distance between an approximation and whatever it is that is being approximated. To do this we need some formal method for measuring distances between objects, and that method relies on the concept of a distance function.

Definition 3.17: A **distance function** on a vector space V is a binary operation that assigns to each pair of vectors $(\mathbf{u}, \mathbf{v})$ a real number $\mathrm{d}(\mathbf{u}, \mathbf{v})$, called the **distance** between $\mathbf{u}$ and $\mathbf{v}$, such that the following properties are satisfied for all vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in V.

1) $\mathrm{d}(\mathbf{u}, \mathbf{v}) \geq 0$, and $\mathrm{d}(\mathbf{u}, \mathbf{v}) = 0$ if and only if $\mathbf{u} = \mathbf{v}$

2) $\mathrm{d}(\mathbf{u}, \mathbf{v}) = \mathrm{d}(\mathbf{v}, \mathbf{u})$

3) $\mathrm{d}(\mathbf{u}, \mathbf{w}) \leq \mathrm{d}(\mathbf{u}, \mathbf{v}) + \mathrm{d}(\mathbf{v}, \mathbf{w})$

A distance function is sometimes referred to as a **metric**; any set of objects for which a metric is defined is called a **metric space**. *There are metric spaces that are not vector spaces.*

The following theorem gives one of the most common ways for obtaining a distance function on a vector space. Using the same language as before, a norm *induces* a distance, or metric.

Theorem 3.18: In any vector space with a norm $\|\ \|$, the function defined by $\mathrm{d}(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$ is a distance function.

Here is a common scenario: We have an infinite dimensional inner product space V of functions, and there is some useful orthogonal basis $\mathcal{B} = \{f_1, f_2, f_3, ...\}$ for that space. Let W_n denote the subspace of V spanned by the first n elements of $\mathcal{B}$. Given some $g \in V$, the function $g_n = \text{proj}_{W_n} g$ is an approximation for g. We can use the distance function induced by the inner product (well, the distance function is induced by the norm, which in turn is induced by the inner product) to find out how close the approximation g_n is to g.

◇ **Example 3.5(d):** The set $\mathcal{B} = \{1, x, x^2 - \frac{1}{3}, x^3 - \frac{3}{5}x, ...\}$ is an orthogonal basis for $\mathscr{C}[-1, 1]$. The projections of $g(x) = \sin x$ onto W_2 and W_4 are

$$g_2(x) = \text{proj}_{W_2} \sin x = 0.9035x \qquad \text{and} \qquad g_4(x) = \text{proj}_{W_4} \sin x = 0.9981x - 0.1576x^3$$

Use the 2-norm induced by the standard integral inner product $\langle f, g\rangle = \dfrac{1}{b-a}\displaystyle\int_a^b f(x)\, g(x)\, dx$ to determine "how close" each of these is to $g(x) = \sin x$.

The distance of the first approximation from g is

$$\|g_2 - g\|_2 = \left(\frac{1}{2}\int_{-1}^{1}\big[g_2(x) - g(x)\big]^2\, dx\right)^{\frac{1}{2}} = \left(\frac{1}{2}\int_{-1}^{1}(0.9035x - \sin x)^2\, dx\right)^{\frac{1}{2}} = .02383$$

$$\|g_4 - g\|_2 = \left(\frac{1}{2}\int_{-1}^{1}\big[g_4(x) - g(x)\big]^2\, dx\right)^{\frac{1}{2}} = \left(\frac{1}{2}\int_{-1}^{1}(0.9981x - 0.1576x^3 - \sin x)^2\, dx\right)^{\frac{1}{2}} = .0003076$$

As we might expect, the second approximation is significantly better than the first.

Section 3.5 Exercises

1. For the vector $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$, find $\|\mathbf{v}\|_1$, $\|\mathbf{v}\|_2$ and $\|\mathbf{v}\|_\infty$. (Label each, of course!)

2. Let $f(x) = x^3 + x^2 - \frac{1}{2}$, $x \in [-1, 1]$.

   (a) Find $\|f\|_1$ and $\|f\|_2$, giving some indication of how you do it. Use your calculator to calculate integrals, and round your answers to the hundredth's place.

   (b) Find $\|f\|_\infty$ by examining the graph of the function.

3. In this exercise you will be reminded of how to use differential calculus to find $\|f\|_\infty$ for $f(x) = x^3 + x^2 - \frac{1}{2}$ on the interval $[-1, 1]$.

   (a) Find all critical values of x by setting the derivative equal to zero and solving. Then determine the values of the function at those values of x to obtain relative minima or maxima.

   (b) Find the values of the function at the ends of the interval $[-1, 1]$.

   (c) The ∞-norm is the maximum of the absolute values of all function values obtained in parts (a) and (b). Give it.

4. Use calculus and algebra, showing clearly how you do it, to determine $\|xe^{-x}\|_\infty$ on the interval $[0, 2]$.

5. The set of points (x, y) in $\mathbb{R}^2$ that have norm one when considered as vectors is called the **unit circle**, where the word circle now means all points equidistant from the origin, not the familiar geometric shape! Sketch the unit circle for each of the following norms on $\mathbb{R}^2$.

   (a) 2-norm (b) 1-norm (c) ∞-norm

6. For $1 \leq p < \infty$, the p-norm of any vector $\mathbf{v} = \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2$ is $\|\mathbf{v}\|_p = \sqrt[p]{|x|^p + |y|^p}$. In this exercise you will see why we call the max norm the ∞-norm.

   (a) Give $\|\mathbf{v}\|_p$ using an exponent for the root.

   (b) When finding the unit circle for the p-norm in $\mathbb{R}^2$, we need to set your expression from part (a) equal to one. Open the online graphing utility *Desmos* and select "Start Graphing." Enter the expression for the p-norm (with the letter p) and set it equal to one. Click where it says to add a slider for p. Select the values one and two for p to check your first two answers to the previous exercise.

   (c) Sketch the unit "circles" for $p = 4$ and $p = 8$ and label each to tell them apart.

   (d) As $p \to \infty$, What seems to happen to the unit circle?

7. In Exercise 4 of Section 3.2 you found that e^x can be approximated on $[-1, 1]$ by

$$p(x) = 0.54x^2 + 1.10x + 1.$$

   The **error** of this approximation can be measured by the distance $\mathrm{d}(e^x, p(x)) = \|e^x - p(x)\|$ induced by any of the norms discussed in this section.

   (a) Find the error using the 1 norm and the 2-norm.

   (b) Find the distance using the ∞-norm by finding the value of x where the graphs of the two functions appear to be farthest apart, then calculating the absolute value of the difference in the values of the two functions at that point.

3.6 Least Squares Solutions to Inconsistent Systems

Performance Criterion:

3. (i) Find the least-squares approximation to the solution of an inconsistent system of equations. Solve a problem using least-squares approximation.

   (j) Give the least squares error and least squares error vector for a least squares approximation to a solution to a system of equations.

Recall that an inconsistent system is one for which there is no solution. Often we wish to solve inconsistent systems and it is just not acceptable to have no solution. In those cases we can find some vector (whose components are the values we are trying to find when attempting to solve the system) that is "closer to being a solution" than all other vectors. The theory behind this process is part of the second term of this course, but we now have enough knowledge to find such a vector in a "cookbook" manner.

Suppose that we have a system of equations $A\mathbf{x} = \mathbf{b}$. Pause for a moment to reflect on what we know and what we are trying to find when solving such a system: We have a system of linear equations, and the entries of A are the coefficients of all the equations. The vector $\mathbf{b}$ is the vector whose components are the right sides of all the equations, and the vector $\mathbf{x}$ is the vector whose components are the unknown values of the variables we are trying to find. So we know A and $\mathbf{b}$ and we are trying to find $\mathbf{x}$. If A is invertible, the solution vector $\mathbf{x}$ is given by $\mathbf{x} = A^{-1}\mathbf{b}$. If A is not invertible there will be no solution vector $\mathbf{x}$, but we can usually find a vector $\bar{\mathbf{x}}$ (usually spoken as "ex-bar") that comes "closest" to being a solution. Here is the formula telling us how to find that $\bar{\mathbf{x}}$:

Theorem 3.19: THE LEAST SQUARES THEOREM: Let A be an $m \times n$ matrix and let $\mathbf{b}$ be in $\mathbb{R}^m$. If $A\mathbf{x} = \mathbf{b}$ has a least squares solution $\bar{\mathbf{x}}$, it is the solution to

$$A^T A\bar{\mathbf{x}} = A^T\mathbf{b}$$

◇ **Example 3.6(a):** Find the least squares solution to
$$\begin{aligned} 1.3x_1 + 0.6x_2 &= 3.3 \\ 4.7x_1 + 1.5x_2 &= 13.5 \\ 3.1x_1 + 5.2x_2 &= -0.1 \end{aligned}.$$

First we note that if we try to solve by row reduction we get no solution; this is an overdetermined system because there are more equations than unknowns. The matrix A and vector $\mathbf{b}$ are

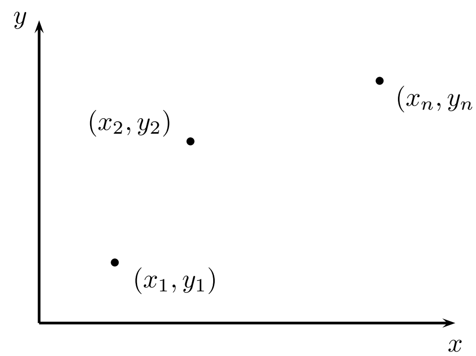
$$A = \begin{bmatrix} 1.3 & 0.6 \\ 4.7 & 1.5 \\ 3.1 & 5.2 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 3.3 \\ 13.5 \\ -0.1 \end{bmatrix}$$

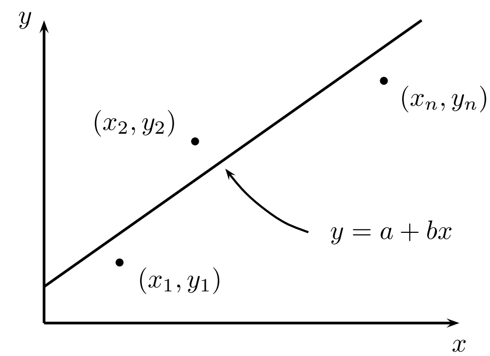
From this

$$A^T A = \begin{bmatrix} 33.39 & 23.95 \\ 23.9529.65 & \end{bmatrix} \qquad \text{and} \qquad A^T\mathbf{b} = \begin{bmatrix} 67.43 \\ 21.71 \end{bmatrix}$$

Finally, solving $A^T A\bar{\mathbf{x}} = A^T\mathbf{b}$ gives $\bar{\mathbf{x}} = \begin{bmatrix} 3.5526 \\ -2.1374 \end{bmatrix}$

A classic example of when we want to do something like this is when we have a bunch of (x, y) data pairs from some experiment, and when we graph all the pairs they describe a trend. We then want to find a simple function $y = f(x)$ that best models that data. In some cases that function might be a line, in other cases maybe it is a parabola, and in yet other cases it might be an exponential function. Let's try to make the connection between this and linear algebra. Suppose that we have the data points (x_1, y_1), (x_2, y_2), ..., (x_n, y_n), and when we graph these points they arrange themselves in roughly a line, as shown below and to the left. We then want to find an equation of the form $a + bx = y$ (note that this is just the familiar $y = mx + b$ rearranged and with different letters for the slope and y-intercept) such that $a + bx_i \approx y_i$ for $i = 1, 2, ..., n$, as shown below and to the right.



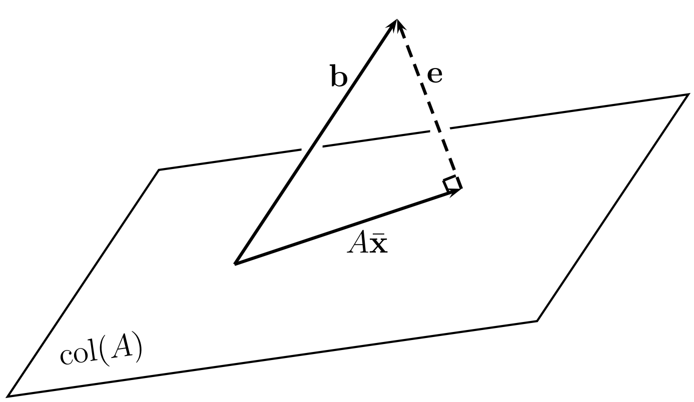


If we substitute each data pair into $a + bx = y$ we get a system of equations which can be thought of in several different ways. Remember that all the x_i and y_i are known values - the unknowns are a and b.

$$\begin{array}{c} a + x_1 b = y_1 \\ a + x_2 b = y_2 \\ \vdots \\ a + x_n b = y_n \end{array} \quad \Longleftrightarrow \quad \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_n \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \Longleftrightarrow \quad A\mathbf{x} = \mathbf{b}$$

Above we first see the system that results from putting each of the (x_i, y_i) pairs into the equation $a + bx = y$. After that we see the $A\mathbf{x} = \mathbf{b}$ form of the system. We must be careful of the notation here. A is the matrix whose columns are a vector in $\mathbb{R}^n$ consisting of all ones and a vector whose components are the x_i values. It would be logical to call this last vector $\mathbf{x}$, but instead $\mathbf{x}$ is the vector $\begin{bmatrix} a \\ b \end{bmatrix}$. $\mathbf{b}$ is the column vector whose components are the y_i values. Our task, as described by this interpretation, is to *find a vector* $\mathbf{x}$ *in* $\mathbb{R}^2$ *that* A *transforms into the vector* $\mathbf{b}$ *in* $\mathbb{R}^n$. Even if such a vector did exist, it couldn't be given as $\mathbf{x} = A^{-1}\mathbf{b}$ because A is not square, so can't be invertible. However, it is likely no such vector exists, but we *CAN* find the least-squares solution $\bar{\mathbf{x}} = \begin{bmatrix} a \\ b \end{bmatrix}$. When we do, its components a and b are the intercept and slope of our line.

Theoretically, here is what is happening. Least squares is generally used in situations that are **over determined**. This means that there is too much information and it is bound to "disagree" with itself somehow. In terms of systems of equations, we are talking about cases where there are more equations than unknowns. Now the fact that the system $A\mathbf{x} = \mathbf{b}$ has no solution means that $\mathbf{b}$ is not in the column space of A. The least squares solution to $A\mathbf{x} = \mathbf{b}$ is simply the vector $\bar{\mathbf{x}}$ for which $A\bar{\mathbf{x}}$ is the projection of $\mathbf{b}$ onto the column space of A. This is shown simplistically below, for the situation where the column space is a plane in $\mathbb{R}^3$.



To recap a bit, suppose we have a system of equations $A\mathbf{x} = \mathbf{b}$ where there is no vector $\mathbf{x}$ for which $A\mathbf{x}$ equals $\mathbf{b}$. What the least squares approximation allows us to do is to find a vector $\bar{\mathbf{x}}$ for which $A\bar{\mathbf{x}}$ is as "close" to $\mathbf{b}$ as possible. We generally determine "closeness" of two objects by finding the difference between them. Because both $A\bar{\mathbf{x}}$ and $\mathbf{b}$ are both vectors of the same length, we can subtract them to get a vector $\mathbf{e}$ that we will call the **error vector**, shown above. The **least squares error** is then the magnitude of this vector:

> **Definition 3.20:** If $\bar{\mathbf{x}}$ is the least-squares solution to the system $A\mathbf{x} = \mathbf{b}$, the **least squares error vector** is
> $$\mathbf{e} = \mathbf{b} - A\bar{\mathbf{x}}$$
> and the **least squares error** is $\|\mathbf{e}\|$, the magnitude of $\mathbf{e}$.

◇ **Example 3.6(b):** Find the least squares error vector and least squares error for the solution obtained in Example 3.6(a).

The least squares error vector is

$$\mathbf{e} = \mathbf{b} - A\bar{\mathbf{x}} = \begin{bmatrix} 3.3 \\ 13.5 \\ -0.1 \end{bmatrix} - \begin{bmatrix} 1.3 & 0.6 \\ 4.7 & 1.5 \\ 3.1 & 5.2 \end{bmatrix} \begin{bmatrix} 3.5526 \\ -2.1374 \end{bmatrix} = \begin{bmatrix} -0.0359 \\ 0.0089 \\ 0.0016 \end{bmatrix}$$

The least squares error is $\|e\| = 0.0370$.

Section 3.6 Exercises

1. Consider the three points $P_1(1,0)$, $P_2(2,1)$ and $P_3(3,5)$.

   (a) Using graph paper, plot the three points. Use something straight (ruler, edge of a CD case, etc.) to verify that there is not a line through all three points. In this assignment and the next we will be searching for the line that "best fits" the three points.

   (b) On the same graph, draw the line $y = \frac{3}{2}x - 1$ and put dots at the three points on the line with x-coordinates 1, 2 and 3. Find the distance between the point P_1 and the point on the line having the same x-coordinate as P_1; we will call that value ε_1. Repeat for P_2 and P_3.

   (c) The vector $\mathbf{e} = [\varepsilon_1, \varepsilon_2, \varepsilon_3]$ is called the error vector, and its (Euclidean) norm is the total error when the line $y = \frac{3}{2}x - 1$ is used to approximate the three points P_1, P_2 and P_3. Find that error.

2. For the same three points, repeat Exercise 1 (new graph and all) for the line $y = 2x - 2$.

3. Of the two lines, which gave the smaller error? In the next assignment you will find the **least-squares regression line**, which has the smallest such error possible.

4 Matrix Factorizations

Outcome/Performance Criteria:

4. Compute and apply matrix factorizations. Find change of basis matrices.

   (a) Use LU-factorization to solve a system of equations, given the LU-factorization of its coefficient matrix.

   (b) Determine a matrix that performs a given row operation.

   (c) Find the LU-factorization of a matrix by hand.

   (d) Find the QR factorization of a matrix.

   (e) Use the QR factorization to compute the least squares solution to a system of equations.

   (f) Diagonalize a matrix when possible; know the forms of the matrices P and D from $P^{-1}AP = D$.

   (g) Find the orthogonal diagonalization and outer product spectral decomposition of a symmetric matrix.

   (h) Find the singular values of a matrix; find the singular value decomposition of a matrix.

   (i) Give the outer product form of the singular value decomposition of a matrix.

4.1 Solving a System With An LU-Factorization

Performance Criterion:

4. (a) Use LU-factorization to solve a system of equations, given the LU-factorization of its coefficient matrix.
   (b) Determine a matrix that performs a given row operation.
   (c) Find the LU-factorization of a matrix by hand.

In many cases an $n \times n$ square matrix A can be "factored" into a product of a lower triangular matrix and an upper triangular matrix, in that order. That is, $A = LU$ where L is lower triangular and U is upper triangular. In that case, for a system $A\mathbf{x} = \mathbf{b}$ that we are trying to solve for $\mathbf{x}$ we have

$$A\mathbf{x} = \mathbf{b} \quad \Rightarrow \quad (LU)\mathbf{x} = \mathbf{b} \quad \Rightarrow \quad L(U\mathbf{x}) = \mathbf{b}$$

Note that $U\mathbf{x}$ is simply a vector; let's call it $\mathbf{y}$. We then have two systems, $L\mathbf{y} = \mathbf{b}$ and $U\mathbf{x} = \mathbf{y}$. To solve the system $A\mathbf{x} = \mathbf{b}$ we first solve $L\mathbf{y} = \mathbf{b}$ for $\mathbf{y}$. Once we know $\mathbf{y}$ we can then solve $U\mathbf{x} = \mathbf{y}$ for $\mathbf{x}$, which was our original goal.

To take full advantage of this method, the process of using L to find $\mathbf{y}$ from $\mathbf{b}$ is carried out by a process called **forward substitution**. Looking at $L\mathbf{y} = \mathbf{b}$ as a system of equations, the first equation contains only y_1, so it is easily solved for. The second equation contains y_1 and y_2, so y_2 can be found easily using the already known value of y_1. This process continues until the values of all of the components of $\mathbf{y}$ are found. Then, when finding the components of $\mathbf{x}$, we consider the system of equations represented by $U\mathbf{x} = \mathbf{y}$. The last equation contains only x_n, so we solve for it. The next to last equation contains only x_{n-1} and x_n, and x_{n-1} is easily found using the just found value of x_n. We then proceed "backward" through the equations, finding each component in turn, from last to first. This process is called **back-substitution**. The reader should carefully note the carrying out of these processes in the following example.

◇ **Example 4.1(a):** Solve the system of equations $\begin{array}{rcr} 7x_1 - 2x_2 + x_3 & = & 12 \\ 14x_1 - 7x_2 - 3x_3 & = & 17 \\ -7x_1 + 11x_2 + 18x_3 & = & 5 \end{array}$, given that the coefficient matrix factors as

$$\begin{bmatrix} 7 & -2 & 1 \\ 14 & -7 & -3 \\ -7 & 11 & 18 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & -3 & 1 \end{bmatrix} \begin{bmatrix} 7 & -2 & 1 \\ 0 & -3 & -5 \\ 0 & 0 & 4 \end{bmatrix}$$

Because of the above factorization we can write the system in matrix form as follows:

$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & -3 & 1 \end{bmatrix} \begin{bmatrix} 7 & -2 & 1 \\ 0 & -3 & -5 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 12 \\ 17 \\ 5 \end{bmatrix}$$

We now let $\begin{bmatrix} 7 & -2 & 1 \\ 0 & -3 & -5 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \quad (*)$ and the above system becomes

$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & -3 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 12 \\ 17 \\ 5 \end{bmatrix} \quad (**)$$

The system $(**)$ is easily solved for the vector $\mathbf{y} = [y_1, y_2, y_3]$ by **forward-substitution**. From the first row we see that $y_1 = 12$; from that it follows that $y_2 = 17 - 2y_1 = 17 - 24 = -7$. Finally, $y_3 = 5 + y_1 + 3y_2 = -4$.

Now that we know $\mathbf{y}$, the system $(*)$ becomes

$$\begin{bmatrix} 7 & -2 & 1 \\ 0 & -3 & -5 \\ 0 & 0 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 12 \\ -7 \\ -4 \end{bmatrix}$$

This is now solved by back-substitution. We can see that $x_3 = -1$, so

$$-3x_2 - 5x_3 = -7 \quad \Longrightarrow \quad -3x_2 + 5 = -7 \quad \Longrightarrow \quad x_2 = 4$$

Finally,

$$7x_1 - 2x_2 + x_3 = 12 \quad \Longrightarrow \quad 7x_1 - 9 = 12 \quad \Longrightarrow \quad x_1 = 3$$

The solution to the original system of equations is $(3, 4, -1)$.

The obvious question to ask at this point is how we factor a matrix A into L and U. The idea is this:

- We find an **elementary matrix** E_1 that, when multiplied times A, accomplished the first step of row-reduction. After that we find a second elementary matrix E_2 that accomplishes the second step of row-reduction in the same way. Continuing, we end up with a sequence E_1, E_2, ..., E_j that, when performed in sequence, perform enough row operations to reduce A to an upper triangular matrix U, which is the U of the LU-factorization.
- We now have $E_jE_{j-1}\cdots E_3E_2E_1A = U$. It turns out that each elementary matrix is invertible, so multiplying on the left by each inverse in turn results in $A = E_1^{-1}E_2^{-1}\cdots E_{j-1}^{-1}E_j^{-1}U$.
- The product $E_1^{-1}E_2^{-1}\cdots E_{j-1}^{-1}E_j^{-1}$ turns out to be a lower triangular matrix - it is L!

Each elementary matrix E_i has ones on the diagonal and every other entry is zero except for one, which turns out to be the multiplier needed to "zero out" an entry of A. The inverse of an elementary matrix is easily obtained by changing the sign of the nonzero term not on the diagonal. Finally, it turns out that the product of the inverses of the elementary matrices (in the correct order!) is simply a lower triangular matrix with ones on the diagonal and all of the nonzero entries of the E_i^{-1} in their places below the diagonal.

The process for doing the above will be carried out in the exercises.

Section 4.1 Exercises

1. In this exercise you will be working with the system $\begin{aligned} x_1 + 3x_2 - 2x_3 &= -4 \\ 3x_1 + 7x_2 + x_3 &= 4 \\ -2x_1 + x_2 + 7x_3 &= 7 \end{aligned}$.

   For the purposes of the exercise we will let

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ -2 & -\frac{7}{2} & 1 \end{bmatrix}, \quad U = \begin{bmatrix} 1 & 3 & -2 \\ 0 & -2 & 7 \\ 0 & 0 & \frac{55}{2} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -4 \\ 4 \\ 7 \end{bmatrix}$$

   (a) Write the system $L\mathbf{y} = \mathbf{b}$ as a system of three equations in the three unknowns y_1, y_2, y_3. Then solve the system by hand, showing clearly how it is done. In the end, give the vector $\mathbf{y}$.

   (b) Write the system $U\mathbf{x} = \mathbf{y}$ as a system of three equations in the three unknowns x_1, x_2, x_3. Then solve the system by hand, showing clearly how it is done. In the end, give the vector $\mathbf{x}$.

   (c) Use the linear combination of vectors interpretation of the system to show that the x_1, x_2, x_3 you found in part (b) is a solution to the system of equations. Show the scalar multiplication and vector addition as two separate steps.

(d) Multiply L times U, in that order. What do you notice about the result? If you don't see something, you may have gone astray somewhere!

2. Let A be the coefficient matrix for the system from the previous exercise.

(a) Give the matrix E_1 be the matrix for which E_1A is the result of the first row operation used to reduce A to U. Give the matrix E_1A.

(b) Give the matrix E_2 such that $E_2(E_1A)$ is the result after the second row operation used to reduce A to U. Give the matrix E_2E_1A.

(c) Give the matrix E_3 such that $E_3(E_2E_1A)$ is U.

(d) Find the matrix $B = E_3E_2E_1$, then use your calculator to find B^{-1}. What is it? If you don't recognize it, you are asleep or you did something wrong!

3. (a) Fill in the blanks of the second matrix below with the entries from E_1. Then, without using your calculator, fill in the blanks in the first matrix so that the product of the first two matrices is the 3×3 identity, as shown.

$$\begin{bmatrix} \underline{\quad} & \underline{\quad} & \underline{\quad} \\ \underline{\quad} & \underline{\quad} & \underline{\quad} \\ \underline{\quad} & \underline{\quad} & \underline{\quad} \end{bmatrix} \begin{bmatrix} \underline{\quad} & \underline{\quad} & \underline{\quad} \\ \underline{\quad} & \underline{\quad} & \underline{\quad} \\ \underline{\quad} & \underline{\quad} & \underline{\quad} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Call the matrix you found F_1. Do the same thing with E_2 and E_3 to find matrices F_2 and F_3.

(b) Find the product $F_1F_2F_3$, in that order. Again, you should recognize the result.

4.2 QR Factorization and Least Squares

Performance Criteria:

4. (d) Find the QR factorization of a matrix.

   (e) Use the QR factorization to compute the least squares solution to a system of equations.

Recall that a set of vectors is an **orthonormal set** if every pair of vectors in the set (other than a vector paired with itself) is orthogonal and each vector in the set has norm one.

Definition 4.1: An **orthogonal matrix** is a square matrix whose columns form an orthonormal set.

Theorem 4.2: If Q is an orthogonal matrix, then $Q^TQ = I$.

Suppose we have a non-square matrix Q whose columns form an orthonormal set. Then Q must have more rows than columns (do you know why?), and it is still true that $Q^TQ = I$. This is because the product Q^TQ amounts to taking the dot products of all of the columns with each other, which results in a one when a column is dotted with itself and a zero when it is dotted with any other column.

Let's now think back to the Gram-Schmidt process:

- We begin with a linearly independent set $\mathbf{v}_1$, $\mathbf{v}_2$,...,$\mathbf{v}_n$, and we start forming an orthogonal set of vectors by letting $\mathbf{w}_1 = \mathbf{v}_1$.
- We use $\mathbf{v}_2$ to form a new vector $\mathbf{w}_2$ that is orthogonal to $\mathbf{w}_1$ and such that $\text{span}(\mathbf{w}_1, \mathbf{w}_2) = \text{span}(\mathbf{v}_1, \mathbf{v}_2)$. We then use $\mathbf{v}_3$ to form a new vector $\mathbf{w}_3$ that is orthogonal to $\mathbf{w}_1$ and $\mathbf{w}_2$, and such that $\text{span}(\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3) = \text{span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$.
- This process continues until we have created an orthogonal set $\{\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_n\}$ that has the same span as $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$.
- We can then normalize each of the $\mathbf{w}$ vectors to obtain an orthonormal set $\{\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_n\}$. For this set we have $\text{span}(\mathbf{q}_1) = \text{span}(\mathbf{v}_1)$, $\text{span}(\mathbf{q}_1, \mathbf{q}_2) = \text{span}(\mathbf{v}_1, \mathbf{v}_2), ...,$ $\text{span}(\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_j) = \text{span}(\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_j)$.

From the last bullet above, each $\mathbf{v}_j$ equals a linear combination of $\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_j$:

$$\mathbf{v}_j = r_{1j}\mathbf{q}_1 + r_{2j}\mathbf{q}_2 + \cdots + r_{jj}\mathbf{q}_j,$$

which we can then extend to a linear combination of all the $\mathbf{q}_k$:

$$\mathbf{v}_j = r_{1j}\mathbf{q}_1 + r_{2j}\mathbf{q}_2 + \cdots + r_{jj}\mathbf{q}_j + 0\mathbf{q}_{j+1} + \cdots + 0\mathbf{q}_n.$$

One might recognize this as the linear combination form of a matrix times a vector (see page 13) where the matrix is the one whose columns are the vectors $\mathbf{q}_k$, in order. This product then forms the kth column of a matrix A whose columns are the vectors $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$. Let $\mathbf{r}_j$ be the vector whose components are $r_{1j}, r_{2j}, ..., r_{jj}, 0, ..., 0$. Then $\mathbf{v}_j = Q\mathbf{r}_j$ and

$$A = \begin{bmatrix} \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \\ \\ \end{bmatrix} = \begin{bmatrix} \\ Q\mathbf{r}_1 & Q\mathbf{r}_2 & \cdots & Q\mathbf{r}_n \\ \\ \end{bmatrix} = Q \begin{bmatrix} \\ \mathbf{r}_1 & \mathbf{r}_2 & \cdots & \mathbf{r}_n \\ \\ \end{bmatrix} = QR$$

The third equality comes from the fact that a product AB can be computed by taking A times the individual columns of B, with the results making up the columns of the product.

Now the components of each $\mathbf{r}_j$ are zero after the jth component r_{jj}, so R has the form

$$R = \begin{bmatrix} & & & \\ \mathbf{r}_1 & \mathbf{r}_2 & \cdots & \mathbf{r}_n \\ & & & \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & \cdots & r_{1n} \\ 0 & r_{22} & r_{23} & \cdots & r_{2n} \\ 0 & 0 & r_{33} & \cdots & r_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & r_{nn} \end{bmatrix}$$

We have proved the following:

Theorem 4.3: Let A be an $m \times n$ matrix with linearly independent columns. Then there exist an $m \times n$ matrix Q with orthonormal columns and an upper triangular square matrix R such that $A = QR$. The product QR is called the QR**-factorization** of A.

A QR-factorization can be carried out as described below. Doing so has some numerical advantages over the method described in Section 3.6.

Theorem 4.4: Let A be an $m \times n$ matrix with linearly independent columns and let $\mathbf{b}$ be in $\mathbb{R}^m$. If $A = QR$ is a QR-factorization of A, then the unique least squares solution $\bar{\mathbf{x}}$ of $A\mathbf{x} = \mathbf{b}$ is

$$\bar{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$$

4.3 A Review of Eigenvectors and Eigenvalues

The next factorization we will look at is what is called the diagonalization of a square matrix. In order to proceed with an investigation of diagonalizing matrices, we need to review eigenvalues and eigenvectors.

> **Definition 4.5:** A scalar λ is called an **eigenvalue** of a matrix A if there is a *nonzero* vector $\mathbf{x}$ such that
> $$A\mathbf{x} = \lambda\mathbf{x}.$$
> The vector $\mathbf{x}$ is an **eigenvector** corresponding to the eigenvalue λ.

One comment is in order at this point. Suppose that $\mathbf{x}$ has n components. Then $\lambda\mathbf{x}$ does as well, so A must have n rows. However, for the multiplication $A\mathbf{x}$ to be possible, A must also have n columns. For this reason, *only square matrices have eigenvalues and eigenvectors.* We now see how to determine whether a vector is an eigenvector of a matrix.

◇ **Example 4.3(a):** Determine whether either of the vectors $\mathbf{w}_1 = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$ and $\mathbf{w}_2 = \begin{bmatrix} -3 \\ 3 \end{bmatrix}$ are eigenvectors for the matrix $A = \begin{bmatrix} -4 & -6 \\ 3 & 5 \end{bmatrix}$ of Example 11.1(a). If either is, give the corresponding eigenvalue.

We see that

$$A\mathbf{w}_1 = \begin{bmatrix} -4 & -6 \\ 3 & 5 \end{bmatrix}\begin{bmatrix} 4 \\ -1 \end{bmatrix} = \begin{bmatrix} -10 \\ 7 \end{bmatrix} \quad \text{and} \quad A\mathbf{w}_2 = \begin{bmatrix} -4 & -6 \\ 3 & 5 \end{bmatrix}\begin{bmatrix} -3 \\ 3 \end{bmatrix} = \begin{bmatrix} -6 \\ 6 \end{bmatrix}$$

$\mathbf{w}_1$ is not an eigenvector of A because there is not scalar λ such that $A\mathbf{w}_1$ is equal to $\lambda\mathbf{w}_1$. $\mathbf{w}_2$ *IS* an eigenvector, with corresponding eigenvalue 2, because $A\mathbf{w}_2 = 2\mathbf{w}_2$.

Note that for the 2×2 matrix A of Examples 11.1(a) and (b) we have seen two eigenvalues now. It turns out that those are the only two eigenvalues, which illustrates the following:

> **Theorem 4.6:** The number of eigenvalues of an $n \times n$ matrix is at most n.

Do not let the use of the Greek letter lambda intimidate you - it is simply some scalar! It is tradition to use λ to represent eigenvalues. Now suppose that $\mathbf{x}$ is an eigenvector of an $n \times n$ matrix A, with corresponding eigenvalue λ, and let c be any scalar. Then for the vector $c\mathbf{x}$ we have

$$A(c\mathbf{x}) = c(A\mathbf{x}) = c(\lambda\mathbf{x}) = (c\lambda)\mathbf{x} = \lambda(c\mathbf{x})$$

This shows that any scalar multiple of $\mathbf{x}$ is also an eigenvector of A with the same eigenvalue λ. The set of all scalar multiples of $\mathbf{x}$ is of course a subspace of $\mathbb{R}^n$, and we call it the **eigenspace** corresponding to λ. $\mathbf{x}$, or any scalar multiple of it, is a basis for the eigenspace. The two eigenspaces you have seen so far have dimension one, but an eigenspace can have a higher dimension.

> **Definition 4.7:** For a given eigenvalue λ_j of an $n \times n$ matrix A, the **eigenspace** E_j corresponding to λ is the set of all eigenvectors corresponding to λ_j. It is a subspace of $\mathbb{R}^n$.

So where are we now? We know what eigenvectors, eigenvalues and eigenspaces are, and we know how to determine whether a vector is an eigenvector of a matrix. There are two big questions at this point:

- Why do we care about eigenvalues and eigenvectors?
- If we are just given a square matrix A, how do we find its eigenvalues and eigenvectors?

We will not see the answer to the first question for a bit, but we'll now tackle answering the second question. We begin by rearranging the eigenvalue/eigenvector equation $A\mathbf{x} = \lambda\mathbf{x}$ a bit. First, we can subtract $\lambda\mathbf{x}$ from both sides to get

$$A\mathbf{x} - \lambda\mathbf{x} = \mathbf{0}.$$

Note that the right side of this equation must be the zero *vector*, because both $A\mathbf{x}$ and $\lambda\mathbf{x}$ are vectors. At this point we want to factor $\mathbf{x}$ out of the left side, but if we do so carelessly we will get a factor of $A - \lambda$, which makes no sense because A is a matrix and λ is a scalar! Note, however, that multiplying a vector by the scalar λ is the same as multiplying by a diagonal vector with all diagonal entries being λ, and that matrix is just λI. Therefore we can replace $\lambda\mathbf{x}$ with $(\lambda I)\mathbf{x}$, allowing us to factor $\mathbf{x}$ out:

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

Now $A - \lambda I$ is just a matrix - let's call it B for now. Any *nonzero* (by definition) vector $\mathbf{x}$ that is a solution to $B\mathbf{x} = \mathbf{0}$ is an eigenvector for A. Clearly the zero vector is a solution to $B\mathbf{x} = \mathbf{0}$, and if B is invertible that will be the only solution. But since eigenvectors are nonzero vectors, A will have eigenvectors only if B is not invertible. Recall that one test for invertibility of a matrix is whether its determinant is nonzero. For B to not be invertible, then, its determinant must be zero. But B is $A - \lambda I$, so we want to find values of λ for which $\det(A - \lambda I) = 0$. (Note that the determinant of a matrix is a scalar, so the zero here is just the scalar zero.) We introduce a bit of special language that we use to discuss what is happening here:

Definition 4.8: Taking λ to be an unknown, $\det(A - \lambda I)$ is a polynomial called the **characteristic polynomial** of A. The equation $\det(A - \lambda I) = 0$ is called the **characteristic equation** for A, and its solutions are the eigenvalues of A.

Before looking at a specific example, you would probably find it useful to go back and look at Examples 7.4(a),(b) and (c), and to recall the following.

Determinant of a 2×2 Matrix

The determinant of the matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is $\det(A) = ad - bc$.

Determinant of a 3×3 Matrix

To find the determinant of a 3×3 matrix,

- Augment the matrix with its first two columns.
- Find the product down each of the three complete "downward diagonals" of the augmented matrix, and the product up each of the three "upward diagonals."
- Add the products from the downward diagonals and subtract each of the products from the upward diagonals. The result is the determinant.

Now we're ready to look at a specific example of how to find the eigenvalues of a matrix.

◇ **Example 4.3(b):** Find the eigenvalues of the matrix $A = \begin{bmatrix} -4 & -6 \\ 3 & 5 \end{bmatrix}$.

We need to find the characteristic polynomial $\det(A - \lambda I)$, then set it equal to zero and solve.

$$A - \lambda I = \begin{bmatrix} -4 & -6 \\ 3 & 5 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -4 & -6 \\ 3 & 5 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} -4-\lambda & -6 \\ 3 & 5-\lambda \end{bmatrix}$$

$$\det(A - \lambda I) = (-4-\lambda)(5-\lambda) - (3)(-6)) = (-20 - \lambda + \lambda^2) + 18 = \lambda^2 - \lambda - 2$$

We now factor this and set it equal to zero to find the eigenvalues:

$$\lambda^2 - \lambda - 2 = (\lambda - 2)(\lambda + 1) = 0 \qquad \Longrightarrow \qquad \lambda = 2, -1$$

We use subscripts to distinguish the different eigenvalues: $\lambda_1 = 2$, $\lambda_2 = -1$.

We now need to find the eigenvectors or, more generally, the eigenspaces, corresponding to each eigenvalue. We defined eigenspaces in the previous section, but here we will give a slightly different (but equivalent) definition.

Definition 4.9: For a given eigenvalue λ_j of an $n \times n$ matrix A, the **eigenspace** E_j corresponding to λ_j is the set of all solutions to the equation

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

It is a subspace of $\mathbb{R}^n$.

Note that we indicate the correspondence of an eigenspace with an eigenvalue by subscripting them with the same number.

◇ **Example 4.3(c):** Find the eigenspace E_1 of the matrix $A = \begin{bmatrix} -4 & -6 \\ 3 & 5 \end{bmatrix}$ corresponding to the eigenvalue $\lambda_1 = 2$.

For $\lambda_1 = 2$ we have $A - \lambda I = \begin{bmatrix} -4 & -6 \\ 3 & 5 \end{bmatrix} - \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} -6 & -6 \\ 3 & 3 \end{bmatrix}$

The augmented matrix of the system $(A - \lambda I)\mathbf{x} = \mathbf{0}$ is then $\begin{bmatrix} -6 & -6 & 0 \\ 3 & 3 & 0 \end{bmatrix}$, which reduces to $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

The solution to this system is all vectors of the form $t \begin{bmatrix} -1 \\ 1 \end{bmatrix}$. We can the describe the eigenspace E_1 corresponding to $\lambda_1 = 2$ by either

$$E_1 = \left\{ t \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\} \qquad \text{or} \qquad E_1 = \text{span}\left(\begin{bmatrix} -1 \\ 1 \end{bmatrix} \right)$$

It would be beneficial for the reader to repeat the above process for the second eigenvalue $\lambda_2 = -1$ and check the answer against what was seen in Example 4.3(a).

When first seen, the whole process for finding eigenvalues and eigenvectors can be a bit bewildering! Here is a summary of the process:

Finding Eigenvalues and Bases for Eigenspaces

The following procedure will give the eigenvalues and corresponding eigenspaces for a square matrix A.

1) Find $\det(A - \lambda I)$. This is the characteristic polynomial of A.

2) Set the characteristic polynomial equal to zero and solve to get the eigenvalues.

3) For a given eigenvalue λ_i, solve the system $(A - \lambda_i I)\mathbf{x} = \mathbf{0}$. The set of solutions is the eigenspace corresponding to λ_i. The vector or vectors whose linear combinations make up the eigenspace are a basis for the eigenspace.

4.4 Diagonalization of Matrices

Performance Criterion:

4. (f) Diagonalize a matrix when possible; know the forms of the matrices P and D from $P^{-1}AP = D$.

We begin with an example involving the matrix A from Examples 4.3(a) and (b).

◇ **Example 4.4(a):** For $A = \begin{bmatrix} -4 & -6 \\ 3 & 5 \end{bmatrix}$ and $P = \begin{bmatrix} -1 & -2 \\ 1 & 1 \end{bmatrix}$, find the product $P^{-1}AP$.

First we obtain $P^{-1} = \dfrac{1}{(-1)(1)-(1)(-2)} \begin{bmatrix} 1 & 2 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ -1 & -1 \end{bmatrix}$. Then

$$P^{-1}AP = \begin{bmatrix} 1 & 2 \\ -1 & -1 \end{bmatrix}\begin{bmatrix} -4 & -6 \\ 3 & 5 \end{bmatrix}\begin{bmatrix} -1 & -2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix}$$

We want to make note of a few things here:

- The columns of the matrix P are eigenvectors for A.
- The matrix $D = P^{-1}AP$ is a diagonal matrix.
- The diagonal entries of D are the eigenvalues of A, in the order of the corresponding eigenvectors in P.

For a square matrix A, the process of creating such a matrix D in this manner is called **diagonalization** of A. This cannot always be done, but often it can. (We will fret about exactly when it can be done later.)

◇ **Example 4.4(b):** Diagonalize the matrix $A = \begin{bmatrix} 3 & 12 & -21 \\ -1 & -6 & 13 \\ 0 & -2 & 6 \end{bmatrix}$.

First we find the eigenvalues by solving $\det(A - \lambda I) = 0$:

$$\begin{aligned}
\det \begin{bmatrix} 3-\lambda & 12 & -21 \\ -1 & -6-\lambda & 13 \\ 0 & -2 & 6-\lambda \end{bmatrix} &= (3-\lambda)(-6-\lambda)(6-\lambda) - 42 + 26(3-\lambda) + 12(6-\lambda) \\
&= (-18 + 3\lambda + \lambda^2)(6-\lambda) - 42 + 78 - 26\lambda + 72 - 12\lambda \\
&= -108 + 18\lambda + 18\lambda - 3\lambda^2 + 6\lambda^2 - \lambda^3 + 108 - 38\lambda \\
&= -\lambda^3 + 3\lambda^2 - 2\lambda \\
&= -\lambda(\lambda^2 - 3\lambda + 2) \\
&= -\lambda(\lambda-2)(\lambda-1)
\end{aligned}$$

The eigenvalues of A are then $\lambda = 0, 1, 2$. We now find an eigenvector corresponding to $\lambda = 0$ by solving the system $(A - \lambda I)\mathbf{x} = 0$. The augmented matrix and its row-reduced form are shown below:

$$\begin{bmatrix} 3 & 12 & -21 & 0 \\ -1 & -6 & 13 & 0 \\ 0 & -2 & 6 & 0 \end{bmatrix} \Longrightarrow \begin{bmatrix} 1 & 4 & -\frac{21}{3} & 0 \\ 0 & 1 & -3 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \Longrightarrow \begin{array}{l} \text{Let } x_3 = 1. \\ \text{Then } x_2 = 3 \\ \text{and } x_1 = -5 \end{array}$$

The eigenspace corresponding to the eigenvalue $\lambda = 0$ is then the span of the vector $\mathbf{v}_1 = [-5, 3, 1]$. For $\lambda = 1$ we have

$$\begin{bmatrix} 2 & 12 & -21 & 0 \\ -1 & -7 & 13 & 0 \\ 0 & -2 & 5 & 0 \end{bmatrix} \implies \begin{bmatrix} 1 & 6 & -\frac{21}{2} & 0 \\ 0 & 1 & -\frac{5}{2} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \implies \begin{array}{l} \text{Let } x_3 = 1. \\ \text{Then } x_2 = \frac{5}{2} \\ \text{and } x_1 = -\frac{9}{2} \end{array}$$

The eigenspace corresponding to the eigenvalue $\lambda = 1$ is then the span of the vector $\mathbf{v}_2 = [-9, 5, 2]$ (obtained by multiplying the solution vector by two in order to get a vector with integer components). Finally, for $\lambda = 2$ we have

$$\begin{bmatrix} 1 & 12 & -21 & 0 \\ -1 & -8 & 13 & 0 \\ 0 & -2 & 4 & 0 \end{bmatrix} \implies \begin{bmatrix} 1 & 12 & -21 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \implies \begin{array}{l} \text{Let } x_3 = 1. \\ \text{Then } x_2 = 2 \\ \text{and } x_1 = -3 \end{array}$$

so the eigenspace corresponding to the eigenvalue $\lambda = 2$ is then the span of the vector $\mathbf{v}_3 = [-3, 2, 1]$. The diagonalization of A is then $D = P^{-1}AP$, where

$$D = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \qquad \text{and} \qquad P = \begin{bmatrix} -5 & -9 & -3 \\ 3 & 5 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

4.5 Diagonalization of Symmetric Matrices

Performance Criterion:

4. (g) Find the orthogonal diagonalization and outer product spectral decomposition of a symmetric matrix.

Recall the following definition and theorem:

Definition: An **orthogonal matrix** is a square matrix whose columns form an orthonormal set.

Theorem: If Q is an orthogonal matrix, then $Q^TQ = I$.

In the previous section we saw that many, but not all, square matrices can be diagonalized, meaning that they can be written in the form $A = PDP^{-1}$. Moreover, in some cases the columns of P and diagonal entries of D might be complex numbers. But in the case that A is symmetric, we can do better:

Theorem 4.10: If A is symmetric, then there exist a diagonal matrix D and an orthogonal matrix Q such that $A = QDQ^T$.

Here are a few comments about Q and D:

- The diagonal entries of D are the eigenvalues of A, which *are always real.* Any repeated eigenvalues are listed multiple times according to their multiplicities.
- If all of the eigenvalues are distinct, then the columns of Q are the normalized eigenvectors, in the order that their eigenvectors appear on the diagonal of D.
- The eigenvectors corresponding to any repeated eigenvalues will be linearly independent but not necessarily orthogonal. In that case the Gram-Schmidt process and normalization are applied to the eigenvectors to obtain an orthonormal set of vectors, and those vectors are the columns of Q.

◇ **Example 4.5(a):** Find the orthogonal diagonalization of $A = \begin{bmatrix} 7 & 2 \\ 2 & 4 \end{bmatrix}$.

Theorem 4.11: Suppose that the symmetric matrix A is orthogonally diagonalized as QDQ^T, where Q has columns $\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_n$ and the diagonal entries of D are $\lambda_1, \lambda_2, ..., \lambda_n$. Then A can also be written as

$$A = \lambda_1\mathbf{q}_1\mathbf{q}_1^T + \lambda_2\mathbf{q}_2\mathbf{q}_2^T + \cdots + \lambda_n\mathbf{q}_n\mathbf{q}_n^T \qquad (1)$$

The sum (1) is called the **outer product spectral decomposition** of A.

◇ **Example 4.5(b):** Find the outer product spectral decomposition of $A = \begin{bmatrix} 7 & 2 \\ 2 & 4 \end{bmatrix}$.

◇ **Example 4.5(c):** Find the orthogonal diagonalization and outer product spectral decomposition of

$$B = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}.$$

4.6 Singular Value Decomposition

Performance Criterion:

4. (h) Find the singular values and singular vectors of a matrix; find the singular value decomposition of a matrix.
   (i) Give the outer product form of the singular value decomposition of a matrix.

In Section 4.4 we saw that in many cases a square matrix A can be diagonalized. That is, there exists a diagonal matrix D and an invertible matrix P for which $A = PDP^{-1}$. The matrix D has the eigenvalues of A along its diagonal, and the columns of P are the eigenvalues of A in the same order that the eigenvalues appear in D. Then we saw that *every* symmetric matrix A is *orthogonally diagonalizable*. This means that $A = QDQ^T$, where D is again diagonal with the eigenvalues of A along the diagonal, but now Q is an orthogonal matrix. In this case the columns of Q are the normalized eigenvectors of A.

The process of diagonalizing is sometimes referred to as decomposing the matrix A into the product of three other matrices. Symmetric matrices can always be decomposed, but we might not be able to decompose a non-symmetric square matrix A as PDP^{-1}. Remarkably, though, *any* matrix of *any* dimensions can be decomposed in a manner quite similar to QDQ^T. This decomposition is called the **singular value decomposition** (SVD) of the matrix. Before describing it fully we need to make a couple of definitions.

Given a matrix A of any dimensions, the matrix A^TA is always symmetric, with real, nonnegative eigenvalues. This alows us to make the following definition.

Definition 4.12: The **singular values** of a matrix A are the square roots of the eigenvalues of A^TA. Each singular value is denoted by a subscripted σ, and it is customary to subscript them so that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$.

◇ **Example 4.6(a):** Find the singular values of $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -2 & 2 \end{bmatrix}$.

Let D represent a diagonal matrix (so necessarily square), and let O represent a matrix of any dimensions with all entries being zero. By a matrix

$$\left[\begin{array}{c:c} D & O \\ \hdashline O & O \end{array}\right] \tag{1}$$

we mean a diagonal matrix D with perhaps additional rows or columns of zeros added at the bottom or on the right.

We can now state the main result of this section:

Theorem 4.13: Given an $m \times n$ matrix A, there exist an $m \times m$ orthogonal matrix U, an $n \times n$ orthogonal matrix V, and a matrix Σ of the form (1) above such that $A = U\Sigma V^T$. The product $U\Sigma V^T$ is called the **singular value decomposition** of A, sometimes just referred to as the SVD.

So how do we find the singular value decomposition of a matrix? The process is described below.

Obtaining a Singular Value Decomposition

To find the singular value decomposition of a a matrix A:

1. Find the eigenvalues for A^TA; their square roots are the singular values. Create a diagonal matrix D with the *nonzero* singular values along the diagonal, in order from largest to smallest. Included repeated singular values.

2. Find the eigenvectors for A^TA. If none of the eigenvalues were repeated, normalize the eigenvectors and put them into a matrix V *in the same order as the singular values corresponding to the their eigenvalues appear in* D. If any of the eigenvalues were repeated, put the eigenvectors in order, perform Gram-Schmidt orthogonalization on them, then normalize them and put them into V. The columns of V are the **right singular vectors**.

3. Find the eigenvectors for AA^T; they are the left singular vectors. Do with them the same thing as described in (b), calling the resulting matrix U. It's columns are the **left singular vectors**.

4. Create a new matrix Σ by adding rows and/or columns of zeros at the bottom and/or right side of D in such a way that the products $U\Sigma$ and ΣV^T are both defined.

5. The product $U\Sigma V^T$ is the singular value decomposition of A.

◇ **Example 4.6(b):** Find the singular value decomposition of $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -2 & 2 \end{bmatrix}$.

Theorem 4.14: Suppose that A has the singular value decomposition $U\Sigma V^T$ with nonzero singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$. The sum

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T,$$

where $\mathbf{u}_1, ..., \mathbf{u}_m$ and $\mathbf{v}_1, ..., \mathbf{v}_n$ are the columns of U and V, respectively, is called the **outer product form of the singular value decomposition** of A.

◇ **Example 4.6(c):** Give the outer product form of the singular value decomposition of $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -2 & 2 \end{bmatrix}$.