Linear Algebra – MAT 202
Spring 2018, CRN: 1186 and 3100

Prerequisites: Calculus I

Instructor: Daniel Cunningham

Table of Contents

Chapter 0

Preliminaries

Linear Algebra is an important branch of mathematics which has many applications in engineering, economics, and social science. In addition, Linear Algebra is usually a students first introduction to mathematical abstraction and mathematical proof.

0.1 Proofs

How to Read a Proof

While a proof may look like a short story, it is often more challenging to read than a short story. Usually some of the computations will not seem clear, and you will have to figure out how they were done. Some of the arguments will not be immediately understandable and will require some thinking. Many of the steps will seem completely strange and may appear very mysterious. Basically, before you can understand a proof you must unravel it. First, identify the main ideas and steps of the proof. Then see how they fit together to allow one to conclude that the result is correct. One important word of advice while reading a proof. Try to remember what it is that has to be proved. Before reading the proof decide what it is exactly that must be proven. Always ask yourself, "What would I have to show in order to prove that?"

How to Write a Proof

Practice! We learn to write proofs by writing proofs. Start by just copying, nearly word for word, a proof in the text that you find interesting. Vary the wording by using your own phrases. Write out the proof using more steps and more details than you found in the original proof. Try to find a different proof of the same statement and write out your new proof. Try to change the order of the argument, if it is possible. If it is not possible, you will soon see why. All mathematicians first learned how to write proofs by going through this process of imitation.

Conjecture + Proof = Theorem

A **conjecture** is a statement that you think is plausible but whose truth has not been established. In mathematics one never accepts a conjecture as true until a mathematical **proof** of the conjecture has been given. Once a mathematical proof of the conjecture is produced we then call the conjecture a **theorem.** On the other hand, to show that a conjecture is false one must find a particular assignment of values (an example) making the statement of the conjecture false. Such an assignment is called a **counterexample** to the conjecture.

The Proof Is Completed

It is convenient to have a mark which signals the end of a proof. Mathematicians in the past, would end their proofs with letters Q.E.D., an abbreviation for the Latin expression "quod erat demonstrandum." So in English, we interpret Q.E.D. to mean "that which was to be demonstrated." In current times, mathematicians typically use the symbol $\square$ to let the reader know that the proof has been completed. In these notes we shall do the same.

Important Sets in Mathematics

Certain sets are frequently used in mathematics. The most commonly used ones are the sets of whole numbers, natural numbers, integers, rational and real numbers. These sets will be denoted by the following symbols:

1. $\mathbb{N} = \{1, 2, 3, \ldots\}$ is the set of natural numbers.
2. $\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$ is the set of integers.
3. $\mathbb{Q}$ is the set of rational numbers; that is, the set of numbers $r = \frac{a}{b}$ for integers a, b where $b \neq 0$. So, $\frac{3}{2} \in \mathbb{Q}$.
4. $\mathbb{R}$ is the set of real numbers and so, $\pi \in \mathbb{R}$.

For sets A and B we write $A \subseteq B$ to mean that the set A is a subset of the set B, that is, every element of A is also an element of B. For example, $\mathbb{N} \subseteq \mathbb{Z}$.

Example 1. Consider the set of integers $\mathbb{Z}$. We evaluate the following truth sets:

1. $\{x \in \mathbb{Z} : x \text{ is a prime number}\} = \{2, 3, 5, 7, 11, \ldots\}$.
2. $\{x \in \mathbb{Z} : x \text{ is divisible by } 3\} = \{\ldots, -12, -9, -6, -3, 0, 3, 6, 9, 12, \ldots\}$.
3. $\{z \in \mathbb{Z} : z^2 \leq 1\} = \{-1, 0, 1\}$.
4. $\{x \in \mathbb{Z} : x^2 \leq 1\} = \{-1, 0, 1\}$.

Interval Notation

In mathematics, an interval is a set consisting of all the real numbers that lie between two given real numbers a and b, where $a < b$. The numbers a and b are referred to as the endpoints of the interval. Furthermore, an interval may or may not include its endpoints.

1. The open interval (a, b) is defined to be $(a, b) = \{x \in \mathbb{R} : a < x < b\}$.
2. The closed interval $[a, b]$ is defined to be $[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}$.
3. The left-closed interval $[a, b)$ is defined to be $[a, b) = \{x \in \mathbb{R} : a \leq x < b\}$.
4. The right-closed interval $[a, b)$ is defined to be $(a, b] = \{x \in \mathbb{R} : a < x \leq b\}$.

Finally, for any real number a we now define the following rays, or half-lines.

1. The interval (a, ∞) is defined to be $(a, \infty) = \{x \in \mathbb{R} : a < x\}$.
2. The interval $[a, \infty)$ is defined to be $[a, \infty) = \{x \in \mathbb{R} : a \leq x\}$.
3. The interval $(-\infty, a)$ is defined to be $(-\infty, a) = \{x \in \mathbb{R} : x < a\}$.
4. The interval $(-\infty, a]$ is defined to be $(-\infty, a] = \{x \in \mathbb{R} : x \leq a\}$.

The symbol ∞ denotes 'infinity' and is not a number. The notation ∞ it is just a useful symbol that allows us to represent intervals that are 'without an end.' Similarly, the notation $-\infty$ is used to denote an interval 'without a beginning.'

Problem 2. Using interval notation, evaluate the following truth sets:

(1) $\{x \in \mathbb{R} : x^2 - 1 < 3\}$.
(2) $\{x \in \mathbb{R}^+ : (x-1)^2 > 1\}$.
(3) $\{x \in \mathbb{R}^- : x > \frac{1}{x}\}$.

Solution.

(1) We first solve the inequality $x^2 - 1 < 3$ for x^2 obtaining $x^2 < 4$. The solution to this latter inequality is $-2 < x < 2$. Thus, $\{x \in \mathbb{R} : x^2 - 1 < 3\} = (-2, 2)$.
(2) We are looking for all the positive real numbers x that satisfy the inequality $(x-1)^2 > 1$. We see that the solution consists of all real numbers $x > 2$. So, $\{x \in \mathbb{R}^+ : (x-1)^2 > 1\} = (2, \infty)$.
(3) We need to find all the negative real numbers x that satisfy $x > \frac{1}{x}$. We conclude $x^2 < 1$. So, we must have $-1 < x < 0$. So, $\{x \in \mathbb{R}^- : x > \frac{1}{x}\} = (-1, 0)$.

Definition. A positive rational number $\frac{m}{n}$ is in **reduced form** if $m \in \mathbb{N}$ and $n \in \mathbb{N}$ have no common factors.

Example. $\frac{4}{3}$ is in reduced form, $\frac{12}{9}$ is not in reduced form because 12 and 9 have a common factor. Clearly every rational number can be put into reduced form.

How to Prove an Equation

Equations play a critical role in modern mathematics. In this text we will establish many theorems that will require us to know how to correctly prove an equation. Because this knowledge is so important and fundamental, our first proof strategy presents two correct methods that we shall use when proving equations.

Proof Strategy 1. To prove a new equation $\varphi = \psi$ there are two approaches:

(a) Start with one side of the equation and derive the other side.

(b) Perform operations on the given equations to derive the new equation.

We now apply strategy 1(a) to prove a well known algebraic identity.

Theorem. Let a and b be arbitrary real numbers. Then $(a+b)(a-b) = a^2 - b^2$.

Proof. We[1] will start with the left hand side $(a+b)(a-b)$ and derive the right hand side as follows:

$$\begin{aligned}(a+b)(a-b) &= a(a-b) + b(a-b) && \text{by the distribution property}\\ &= a^2 - ab + ba - b^2 && \text{by the distribution property}\\ &= a^2 - b^2 && \text{by algebra.}\end{aligned}$$

Thus, we have that $(a+b)(a-b) = a^2 - b^2$. □

[1] Most mathematicians use the term "we" in their proofs. This is considered polite and is intended to include the reader in the discussion.

We now apply strategy 1(b) to prove a new equation from some given equations.

Theorem. Let m, n, i, j be integers. Suppose that $m = 2i + 5$ and $n = 3j$. Then $mn = 6ij + 15j$.

Proof. We are given that $m = 2i + 1$ and $n = 2j$. By multiplying corresponding sides of these two equation, we obtain $mn = (2i + 5)(3j)$. Thus, by algebra, we conclude that $mn = 6ij + 15j$. □

Remark 0.1.1. To prove that an equation $\varphi = \psi$ is true, it is not a correct method of proof to *assume* the equation $\varphi = \psi$ and then work on both sides of this equation to obtain an identity.

The method described in Remark 0.1.1 is a fallacious one and if applied, can produce false equations. For example, this fallacious method can be used to derive the equation $-1 = 1$. To illustrate this, let us assume the equation $-1 = 1$. Now square both sides, obtaining $(-1)^2 = 1^2$ which results in the true equation $1 = 1$. The method cited in Remark 0.1.1 would allow us to conclude that $-1 = 1$ is a true equation. This is complete nonsense. **We never want to apply a method that can produce false equations!**

Exercises 0.1

1. Let x and y be real numbers. Prove that $(x - y)(x^2 + xy + y^2) = x^3 - y^3$.

2. Let x and y be real numbers. Prove that $(x + y)(x^2 - xy + y^2) = x^3 + y^3$.

3. Let x and y be real numbers. Prove that $(x + y)^2 = x^2 + 2xy + y^2$.

4. Let x and y be real numbers. Using exercise 3, prove that $(x + y)^3 = x^3 + 3x^2y + 3xy^2 + y^3$.

5. Let φ be the positive real number satisfying the equation $\varphi^2 - \varphi - 1 = 0$. Prove that $\varphi = \frac{1}{\varphi - 1}$.

6. Let φ be the positive real number satisfying the equation $\varphi^2 - \varphi - 1 = 0$. Let $a \neq b$ be real numbers satisfying $\frac{b}{a} = \varphi$. Prove that $\frac{a}{b-a} = \varphi$.

0.2 Sets

In modern mathematics, many of the most important ideas are expressed in term of sets. A *set* is just a collection of objects. These objects are referred to as the *elements* of the set. These elements can be numbers, ordinary objects, words, other sets, functions, etc. An object a may or may not belong to a given set A. If a belongs to the set A then we say that *a is an element of A*, and we write $a \in A$. Otherwise, a is not an element of A and we write $a \notin A$. A finite set has the form $A = \{x_1, x_2, \ldots, x_n\}$ where n is a natural number and the listed elements of A are all distinct.

Basic Definitions of Set Theory

Definition 0.2.1. The following set notation is used throughout mathematics.

1. For sets A and B we write $A = B$ to mean that both sets have exactly the same elements.
2. For sets A and B we write $A \subseteq B$ to assert that the set A is a subset of the set B, that is, every element of A is also an element of B.
3. We shall say that the set A is a **proper** subset of the set B when $A \subseteq B$ and $A \neq B$.
4. We write $\varnothing$ for the empty the set, that is, the set with no members.
5. If A is a finite set, then $|A|$ represents the number of elements in A.

Venn diagrams are geometric shapes that are used to depict sets and their relationships. In Figure 1 we present a Venn diagram which illustrates the subset relation, a very important concept in set theory and mathematics.

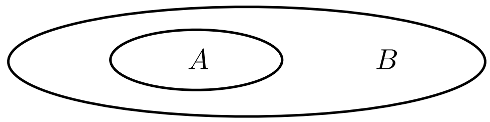


Figure 1: Venn diagram of $A \subseteq B$

Set Operations

The language of set theory is used in the definitions of nearly all of mathematics. There are three important and fundamental operations on sets that we shall now discuss: the intersection, the union and the difference of two sets. We illustrate these four set operations in Figure 2 using Venn diagrams. Shading is used to focus one's attention on the result of each set operation.

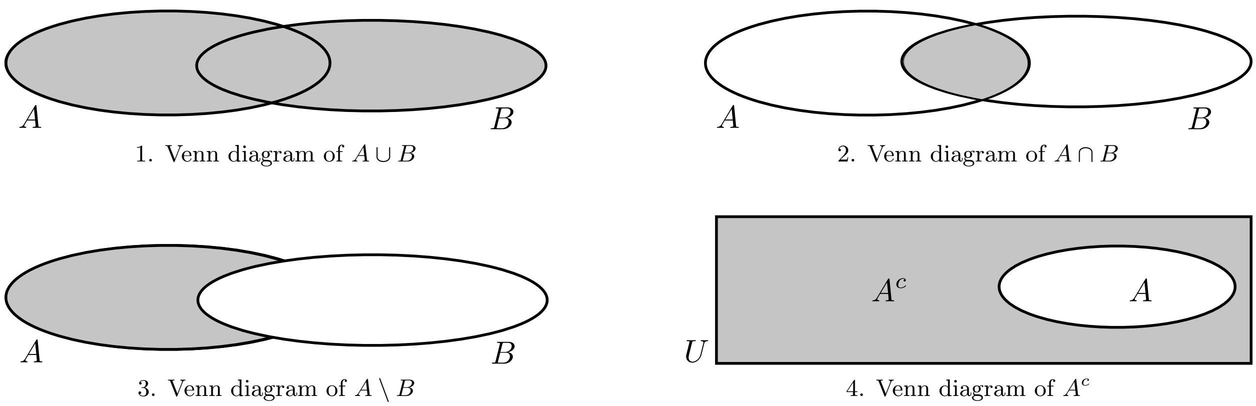


Figure 2: Set Operations

Definition 0.2.2. Given sets A and B we can build new sets using the **set operations**:

1. $A \cup B = \{x : x \in A \text{ or } x \in B\}$ is the **union** of A and B.
2. $A \cap B = \{x : x \in A \text{ and } x \in B\}$ is the **intersection** of A and B.
3. $A \setminus B = \{x : x \in A \text{ and } x \notin B\}$ is the **set difference** of A and B (also stated in English as A "minus" B).
4. Given a universe of objects U and $A \subseteq U$, the set $A^c = U \setminus A = \{x \in U : x \notin A\}$ is called the **complement** of A.

Example 1. Let $A = \{1, 2, 3, 4, 5, 6\}$ and $B = \{2, 4, 6, 8, 10, 12\}$. Then

- $A \cup B = \{1, 2, 3, 4, 5, 6, 8, 10, 12\}$.
- $A \cap B = \{2, 4, 6\}$.
- $A \setminus B = \{1, 3, 5\}$.
- $B \setminus A = \{8, 10, 12\}$.

Problem 2. Recalling the notation (see page 5) for intervals on the real line, evaluate the result of the following set operations:

1. $(-3,2) \cap (1,3)$.
2. $(-3,4) \cup (0,\infty)$.
3. $(-3,2) \setminus [1,3)$.

Solution. While reading the solution to each of these items, it may be helpful to sketch the relevant intervals on the real line.

1. Since $x \in (-3,2) \cap (1,3)$ if and only if $x \in (-3,2)$ and $x \in (1,3)$, we see that x is in this intersection precisely when x satisfies both (a) $-3 < x < 2$ and (b) $1 < x < 3$. We see that the only values for x that satisfies both (a) and (b) are those such that $1 < x < 2$. Thus, $(-3,2) \cap (1,3) = (1,2)$.

2. Since $x \in (-3,4) \cup (0,\infty)$ if and only if $x \in (-3,4)$ or $x \in (0,\infty)$, we see that x is in this union precisely when x satisfies either (a) $-3 < x < 4$ or (b) $0 < x$. We see that the only values for x that satisfies either (a) or (b) are those such that $-3 < x$. Thus, $(-3,4) \cup (0,\infty) = (-3,\infty)$.

3. Since $x \in (-3,2) \setminus [1,3)$ if and only if $x \in (-3,2)$ and $x \notin [1,3)$, we see that x is in this set difference precisely when x satisfies (a) $-3 < x < 2$ and (b) not $(1 \leq x < 3)$. We see that the only values for x that satisfies both (a) and (b) are those such that $-3 < x < 1$. Thus, $(-3,2) \setminus [1,3) = (-3,1)$.

Exercises 0.2

1. Recalling our discussion on interval notation on page 5, evaluate the following set operations:

(a) $(-2,0) \cap (-\infty,2)$.
(b) $(-2,4) \cup (-\infty,2)$.
(c) $(-\infty,0] \setminus (-\infty,2]$.
(d) $\mathbb{R} \setminus (2,\infty)$.
(e) $(\mathbb{R} \setminus (-\infty,2]) \cup (1,\infty)$.

0.3 Functions

Definition 0.3.1. We write $f\colon A \to B$ to mean that f is a **function** from the set A to the set B, that is, for every element $x \in A$ there is exactly one element $f(x)$ in B. The value $f(x)$ is called "f of x," or "the image of x under f." The set A is called the **domain** of the function f and the set B is called the **co-domain** of the function f. In addition, we shall say that $x \in A$ is an *input* for the function f and that $f(x)$ is the resulting *output*. We will also say that x gets *mapped* to $f(x)$.

Remark 0.3.2. If $f\colon A \to B$ then every $x \in A$ is assigned exactly one element $f(x)$ in B. We say that f is **single-valued**. Thus, for every $x \in A$ and $z \in A$, if $x = z$ then $f(x) = f(z)$.

Definition 0.3.3. Given a function $f\colon A \to B$ the **range** of f, denoted by $\mathcal{R}(f)$, is the set

$$\mathcal{R}(f) = \{f(a) : a \in A\} = \{b \in B : b = f(a) \text{ for some } a \in A\}.$$

The range of a function is the set of all "output" values produced by the function.

Question. Let $h : X \to Y$ be a function. What does it mean to say that $b \in \mathcal{R}(h)$? Answer: $b \in \mathcal{R}(h)$ means that $b = f(x)$ for some $x \in A$.

Example 1. Let $f : \mathbb{R} \to \mathbb{R}$ be the function in Figure 3 defined by the formula $f(x) = x^2 - x$. Then $\mathcal{R}(f) = \{f(x) : x \in \mathbb{R}\} = \{x^2 - x : x \in \mathbb{R}\} = [-\frac{1}{4}, \infty)$.

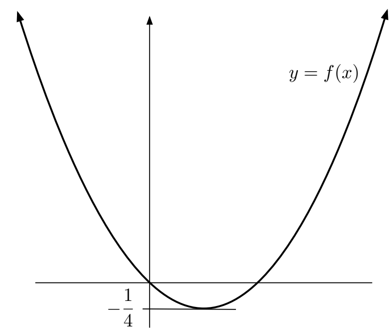


Figure 3: Graph of $f(x) = x^2 - x$

One-To-One Functions and Onto Functions

Definition. A function $f : X \to Y$ is said to be **one-to-one** (or an **injection**), if distinct elements in X get mapped to distinct elements in Y; that is,

$$\text{for all } a, b \in X, \text{ if } a \neq b \text{ then } f(a) \neq f(b),$$

or equivalently,

$$\text{for all } a, b \in X, \text{ if } f(a) = f(b) \text{ then } a = b.$$

Definition. A function $f : X \to Y$ is said to be **onto** (or a **surjection**), if for each $y \in Y$ there is an $x \in X$ such that $f(x) = y$.

Definition. A function $f : X \to Y$ is said to be **one-to-one and onto** (or a **bijection**), if f is both one-to-one and onto.

Inverse Functions

In calculus you study the inverse trigonometric functions, and you also learn that the two functions $\ln(x)$ and e^x are inverses of each other. The *inverse* of a function is another function that "reverses the action" of the original function. Not every function has an inverse. The only functions that do have an inverse are those that are one-to-one and onto.

Theorem 0.3.4. Suppose that $f : A \to B$ is one-to-one and onto. Then there is a function $f^{-1} : B \to A$ that satisfies

$$f^{-1}(b) = a \quad \text{iff} \quad f(a) = b \tag{1}$$

for all $b \in B$ and $a \in A$.

Proof. Suppose $f : A \to B$ is one-to-one and onto. We shall prove that f^{-1}, as defined by (1), is a function from B to A. To do this, we shall show that f^{-1} is single-valued. Let $b \in B$. Since $f : A \to B$ is onto, there is an $a \in A$ such that $f(a) = b$. Suppose that $a' \in A$ also satisfies $f(a') = b$.

Thus, $f(a) = f(a')$. Because f is one-to-one, it follows that $a = a'$. Therefore, for every $b \in B$ there is exactly one $a \in A$ such that $f(a) = b$. Hence, the formula $f(a) = b$ used in (1) defines a function $f^{-1}\colon B \to A$. □

Definition 0.3.5. Suppose $f\colon A \to B$ is one-to-one and onto. Then the function $f^{-1}\colon B \to A$, defined by (1), is called the **inverse function** of f.

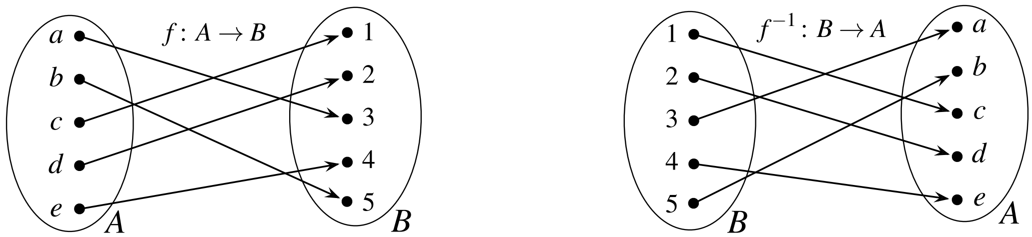


Figure 4: A function $f\colon A \to B$ and its inverse $f^{-1}\colon B \to A$

An arrow diagram of a one-to-one and onto function $f\colon A \to B$ is given in the Figure 4. The arrow diagram for the inverse function $f^{-1}\colon B \to A$ is also portrayed in Figure 4. Observe that the inverse function f^{-1} reverses the action of f and that $f(x) = y$ if and only if $f^{-1}(y) = x$, for each $x \in A$ and $y \in B$.

Composition of Functions

If the domain of a function equals the co-domain of another function, then we can use these two functions to construct a new function called the *composite function*. The composite function is defined by taking the output of one these functions and using that as the input for the other function. The formal mathematical definition appears below.

Definition 0.3.6. For functions $g\colon A \to B$ and $f\colon B \to C$, one forms the **composite function** $(f \circ g)\colon A \to C$ by defining $(f \circ g)(x) = f(g(x))$ for all $x \in A$.

For example, let $g\colon A \to B$ and $f\colon B \to C$ be the functions in Figure 5. An arrow diagram for the composite function $(f \circ g)\colon A \to C$ appears in Figure 6.

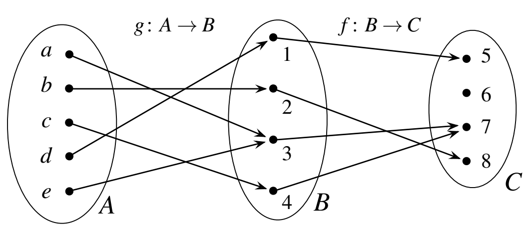


Figure 5: Two functions f and g where the domain of f equals the co-domain of g

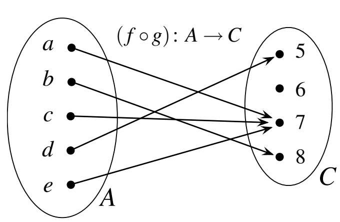


Figure 6: The resulting composite function $f \circ g$ for the functions in Figure 5

Example 2. Let $g\colon \mathbb{R} \to \mathbb{R}$ and $f\colon \mathbb{R} \to \mathbb{R}$ be the functions defined by $f(x) = \frac{1}{x^2+2}$ and $g(x) = 2x - 1$. Find formulas for $(f \circ g)(x)$ and $(g \circ f)(x)$. Is $f \circ g = g \circ f$?

Solution. Let $x \in \mathbb{R}$. We evaluate the function $(f \circ g)(x)$ as follows:

$$(f \circ g)(x) = f(g(x)) = f(2x-1) = \frac{1}{(2x-1)^2+2}.$$

Thus, $(f \circ g)(x) = \frac{1}{(2x-1)^2+2}$. We evaluate $(g \circ f)(x)$ to obtain

$$(g \circ f)(x) = g(f(x)) = g\left(\frac{1}{x^2+2}\right) = 2\left(\frac{1}{x^2+2}\right) - 1 = \frac{2}{x^2+2} - 1.$$

Hence, $(g \circ f)(x) = \frac{2}{x^2+2} - 1$. Since $(f \circ g)(0) = \frac{1}{3}$ and $(g \circ f)(0) = 0$, we conclude that $f \circ g \neq g \circ f$.

One cannot form the composition of just any two functions. When in doubt here is a simple rule to follow: *The composition $f \circ g$ is defined when the domain of the left function f is equal to the co-domain of the right function g.*

Remark 0.3.7. Given two functions $g\colon A \to E$ and $f\colon B \to C$, if $\mathcal{R}(g) \subseteq B$, then one can also define the composition $(f \circ g)\colon A \to C$. In other words, if $f(b)$ is defined for every value b of the function g, then one can define $f \circ g$.

Composing a Function with the Identity Function

The identity function just takes an input x and returns x as its output value. As a result, when one composes a function f with the identity function, the result will just be the function f.

Theorem 0.3.8. Let f be any function $f\colon A \to B$. Let $i_A\colon A \to A$ be the identity function on A and let $i_B\colon B \to B$ be the identity function on B. Then

(1) $(f \circ i_A) = f$,

(2) $(i_B \circ f) = f$.

Proof. Clearly, $(f \circ i_A)(x) = f(i_A(x)) = f(x)$ and $(i_B \circ f)(x) = i_B(f(x)) = f(x)$, for each $x \in A$. □

Composing a Function with its Inverse

Since the inverse of a function "reverses the action" of the original function, the result of composing these two functions leads to "no action."

Theorem 0.3.9. Suppose $f\colon A \to B$ is one-to-one and onto. Let $f^{-1}\colon B \to A$ be the inverse of f. Then

(1) $f^{-1}(f(a)) = a$ for all $a \in A$,

(2) $f(f^{-1}(b)) = b$ for all $b \in B$.

Proof. First we prove (1). Let $a \in A$. Since $f(a) \in B$, let $b \in B$ be such that $f(a) = b$. Theorem 0.3.4 implies $(*)$ $f^{-1}(b) = a$. After substituting $b = f(a)$ into equation $(*)$, we see that $f^{-1}(f(a)) = a$. To prove (2), let $b \in B$. Since $f^{-1}(b) \in A$, let $a \in A$ be such that $f^{-1}(b) = a$. Thus, $(\dagger)$ $f(a) = b$ by Theorem 0.3.4. Upon substituting $a = f^{-1}(b)$ into equation $(\dagger)$, we obtain $f(f^{-1}(b)) = b$. □

Corollary 0.3.10. Suppose $f\colon A \to B$ is one-to-one and onto. Let $f^{-1}\colon B \to A$ be the inverse of f. Then

(1) $(f^{-1} \circ f) = i_A$

(2) $(f \circ f^{-1}) = i_B$

where i_A is the identity function on A and i_B is the identity function on B.

Proof. Since $i_A(a) = a$ for $a \in A$ and $i_B(b) = b$ for $b \in B$, items (1) and (2) follow from the corresponding items in Theorem 0.3.9. □

Composing One-To-One Functions

Our next theorem shows that the composition of two one-to-one functions is also one-to-one.

Theorem 0.3.11. If $g\colon A \to B$ and $f\colon B \to C$ are one-to-one, then $(f \circ g)\colon A \to C$ is one-to-one.

Proof. Assume $g\colon A \to B$ and $f\colon B \to C$ are one-to-one. To prove that the function $(f \circ g)\colon A \to C$ is one-to-one, let $x \in A$ and $y \in A$. Assume $(f \circ g)(x) = (f \circ g)(y)$. Thus, (i) $f(g(x)) = f(g(y))$ by the definition of composition. Since f is one-to-one, we conclude from (i) that $g(x) = g(y)$. Because g is one-to-one, we see that $x = y$. This completes the proof. □

Composing Onto Functions

The next theorem asserts that the composition of two onto functions yields an onto function.

Theorem 0.3.12. If $g\colon A \to B$ and $f\colon B \to C$ are onto, then $(f \circ g)\colon A \to C$ is onto.

Proof. Assume $g\colon A \to B$ and $f\colon B \to C$ are onto. We shall prove that the function $(f \circ g)\colon A \to C$ is onto. Let $z \in C$. Since $f\colon B \to C$ is onto and $z \in C$, there is a $y \in B$ such that $f(y) = z$. Because $y \in B$ and $g\colon A \to B$ is onto, there is an $x \in A$ such that $g(x) = y$. We will show that $(f \circ g)(x) = z$ as follows:

$$\begin{aligned}
(f \circ g)(x) &= f(g(x)) && \text{by definition of composition} \\
&= f(y) && \text{because } g(x) = y \\
&= z && \text{because } f(y) = z.
\end{aligned}$$

Thus, $(f \circ g)(x) = z$. Therefore, $(g \circ f)\colon X \to Z$ is onto. □

Chapter 1

Matrices and Linear Systems

1.1 Introduction to Matrices and Linear Systems

A system of m linear equations in the n unknowns $x_1, \dots, x_n$ has the form

$$\begin{aligned} a_{11}x_1 + \ a_{12}x_2 + \cdots + \ a_{1n}x_n &= b_1 \\ a_{21}x_1 + \ a_{22}x_2 + \cdots + \ a_{2n}x_n &= b_2 \\ a_{31}x_1 + \ a_{32}x_2 + \cdots + \ a_{3n}x_n &= b_3 \\ \vdots \qquad\qquad & \ \ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned} \tag{1.1}$$

is called a $(m \times n)$ **system of linear equations**, where the a_{ij} and b_i are fixed constants.

Example 1. Consider the following (3×4) linear system:

$$\begin{aligned} 23x - 22y + 2w + 4z &= 26 \\ 3x + \ 2y - \ w + \ z &= 4 \\ -10x - \ 5y + 4w - 7z &= -6. \end{aligned} \tag{1.2}$$

One can check that $x = 2, y = 1, w = 3, z = -1$ is a solution to this system. Thus, we will say that $(2, 1, 3, -1)$ is a solution to the system (1.2).

Definition 1.1.1. We say that the list of numbers $(x_1, \dots, x_n)$ is a solution to the system (1.1) if it satisfies **all** the equations in the system. The set

$$\mathcal{S} = \{(x_1, \dots, x_n) : (x_1, \dots, x_n) \text{ is a solution to the system (1.1)}\}$$

is called the **solution set** of the system (1.1). The system (1.1) is **consistent** if it has a solution, that is, $\mathcal{S} \neq \varnothing$. If the system (1.1) has no solutions then it is **inconsistent**.

Question. How to find the solution set $\mathcal{S}$ to the system (1.1)?

Answer. Transform the **complicated** system (1.1), using elementary operations (algebra), to an *equivalent* system which is **easier** to solve.

1.1.1 Elementary Operations on a Linear System of Equations

Definition 1.1.2. The **Elementary Operations** that can be applied to a system of equations are the following:

(1) Multiply an equation by an non-zero number.

(2) Interchange two equations.

(3) Add a multiple of one equation to another equation.

Theorem 1.1.3. If one linear system of systems of equations is obtained from another linear system of equations by applying elementary operations, then both linear systems have exactly the same solutions.

Problem 2. Solve the system

$$\begin{aligned} x + 2y &= 8 \\ 3x - 4y &= 4. \end{aligned}$$

Solution. To be solved in class.

Problem 3. Solve the system

$$\begin{aligned} 3x + 2y + z &= 2 \\ 4x + 2y + 2z &= 8 \\ x - y + z &= 4. \end{aligned}$$

Solution. To be solved in class.

1.1.2 Matrices

Definition 1.1.4. An $m \times n$ **matrix** is a rectangular array of numbers of the form

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

and the number a_{ij} is called the ij^{th} entry of the matrix A. We shall sometimes write $A = [a_{ij}]_{m \times n}$ as shorthand. Also, we shall sometimes write $A_{m \times n}$ when we want to emphasize the "size" of the matrix A. For notational convenience we shall write

$$\mathbf{a}_i = \begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{bmatrix}$$

for the i^{th} row of the matrix A; and we shall write

$$\mathbf{A}_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{m_j} \end{bmatrix}$$

for the j^{th} column of the matrix A.

For example, the 4×5 matrix

$$A_{4\times 5} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ -2 & 0 & 20 & \pi & -4 \\ -3 & -1 & 3 & 6 & 9 \\ -4 & 4 & 5 & -4 & 8 \end{bmatrix}$$

has $a_{12} = 2$, and $a_{43} = 5$. The 2nd row of A is

$$\mathbf{a}_2 = \begin{bmatrix} -2 & 0 & 20 & \pi & -4 \end{bmatrix}$$

and the 3rd column of A is

$$\mathbf{A}_3 = \begin{bmatrix} 3 \\ 20 \\ 3 \\ 5 \end{bmatrix}.$$

1.1.3 Matrix Representation of a Linear System

In this section we will show how a system of equations can be expressed as a matrix. Matrices are helpful in rewriting a linear system in a very simple form. The algebraic properties of matrices may then be used to solve systems.

Given a system of m linear equations in the n unknowns $x_1, \ldots, x_n$

$$\begin{aligned} a_{11}x_1 + \ a_{12}x_2 + \cdots + \ a_{1n}x_n &= b_1 \\ a_{21}x_1 + \ a_{22}x_2 + \cdots + \ a_{2n}x_n &= b_2 \\ a_{31}x_1 + \ a_{32}x_2 + \cdots + \ a_{3n}x_n &= b_3 \\ \vdots \qquad\qquad & \quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned} \tag{1.3}$$

we form the two matrices

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

The matrix A is called the **coefficient** matrix and the $m \times (n+1)$ matrix

$$[A \,|\, \mathbf{b}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{bmatrix}$$

is called the **augmented** matrix. The augmented matrix provides a more compact notation for a linear system of equations.

Example 4. Given the system of equations

$$\begin{aligned} x_1 + \ \ x_2 + \ \ 2x_3 &= 9 \\ 2x_1 + 9x_2 - 10x_3 &= 39 \\ x_1 - 6x_2 + \ \ \ x_3 &= -27. \end{aligned} \tag{1.4}$$

the coefficient matrix is given by

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 9 & -10 \\ 1 & -6 & 1 \end{bmatrix}$$

and the matrix $\mathbf{b}$ is given by $\mathbf{b} = \begin{bmatrix} 9 \\ 39 \\ -27 \end{bmatrix}$. Thus, the augmented matrix of the linear system (1.4) is given by

$$[A \,|\, \mathbf{b}] = \begin{bmatrix} 1 & 1 & 2 & 9 \\ 2 & 9 & -10 & 39 \\ 1 & -6 & 1 & -27 \end{bmatrix}.$$

1.1.4 Elementary Row Operations on a Matrix

We will now show how to use the augmented matrix to solve a system of equations.

Definition 1.1.5. The **Elementary Row Operations** which can be applied to a matrix are:

(1) Multiply a row by an non-zero number

(2) Interchange two rows

(3) Add a multiple of one row to another.

Notation for Row Operations. Given a matrix A we shall let R_i represent the i-th row of A. Thus we can represent the above row operations as follows:

(1) Replace the i-th row by an non-zero multiple k of the i-th row: $kR_i \to R_i$

(2) Interchange the i-th and j-th rows: $R_i \leftrightarrow R_j$.

(3) Replace the j-th row with a multiple of the i-th row added to the j-th row: $kR_i + R_j \to R_j$, where $i \neq j$.

Definition 1.1.6. If a matrix B can be obtained from a matrix A by performing a finite sequence of row operations on A, then the matrices A and B are said to be **row equivalent**.

Theorem 1.1.7. Let $[A \,|\, \mathbf{b}]$ and $[C \,|\, \mathbf{d}]$ be the augmented matrices of two linear systems of equations each of m equations and n unknowns. If the matrices $[A \,|\, \mathbf{b}]$ and $[C \,|\, \mathbf{d}]$ are row equivalent, then both linear systems have exactly the same solutions.

Problem 5. Find the solution set of the system of equations

$$\begin{aligned} x_1 + \ \ x_2 + \ \ 2x_3 &= 9 \\ 2x_1 + 9x_2 - 10x_3 &= 39 \\ x_1 - 6x_2 + \ \ \ x_3 &= -27. \end{aligned} \tag{1.5}$$

Solution. We form the augmented matrix

$$[A\,|\,\mathbf{b}] = \begin{bmatrix} 1 & 1 & 2 & 9 \\ 2 & 9 & -10 & 39 \\ 1 & -6 & 1 & -27 \end{bmatrix}.$$

Now, using row operations, we transform the matrix $[A\,|\,\mathbf{b}]$ into the "reduced echelon" form $[C\,|\,\mathbf{d}]$ as follows:

$$[A\,|\,\mathbf{b}] = \begin{bmatrix} 1 & 1 & 2 & 9 \\ 2 & 9 & -10 & 39 \\ 1 & -6 & 1 & -27 \end{bmatrix}$$

$$-2R_1 + R_2 \to R_2$$

$$\begin{bmatrix} 1 & 1 & 2 & 9 \\ 0 & 7 & -14 & 21 \\ 1 & -6 & 1 & -27 \end{bmatrix}$$

$$-R_1 + R_3 \to R_3$$

$$\begin{bmatrix} 1 & 1 & 2 & 9 \\ 0 & 7 & -14 & 21 \\ 0 & -7 & -1 & -36 \end{bmatrix}$$

$$R_2 + R_3 \to R_3$$

$$\begin{bmatrix} 1 & 1 & 2 & 9 \\ 0 & 7 & -14 & 21 \\ 0 & 0 & -15 & -15 \end{bmatrix}$$

$$\frac{1}{7}R_2 \to R_2$$

$$\begin{bmatrix} 1 & 1 & 2 & 9 \\ 0 & 1 & -2 & 3 \\ 0 & 0 & -15 & -15 \end{bmatrix}$$

$$-\frac{1}{15}R_3 \to R_3$$

$$\begin{bmatrix} 1 & 1 & 2 & 9 \\ 0 & 1 & -2 & 3 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

$$-R_2 + R_1 \to R_1$$

$$\begin{bmatrix} 1 & 0 & 4 & 6 \\ 0 & 1 & -2 & 3 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

$$2R_3 + R_2 \to R_2$$

$$\begin{bmatrix} 1 & 0 & 4 & 6 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

$$-4R_3 + R_1 \to R_1$$

$$[C\,|\,\mathbf{d}] = \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

The matrix $[C \mid \mathbf{d}]$ is in "reduced echelon form". We will now express this new augmented matrix $[C \mid \mathbf{d}]$ as a linear system of equations and solve this new system of equations to obtain:

$$\begin{aligned} x_1 \qquad\qquad &= 2 \\ x_2 \qquad &= 5 \\ x_3 &= 1 \end{aligned}$$

This new system has the same solution set as (1.5) by Theorem 1.1.7 and thus, the solution set is $\mathcal{S} = \{(2, 5, 1)\}$.

Exercises 1.1

Pages 12 to 13 of text – #1, 3, 7, 9, 11, 13, 19, 21, 25, 27, 31, 35.

1.2 Reduced Echelon Form and Gauss-Jordon Elimination

Many of the problems we will solve in Linear Algebra require that a matrix be converted into *reduced echelon form*. Any matrix can be transformed to a matrix in reduced echelon form using elementary row operations, by a method known as Gauss-Jordan elimination. The solutions of a system of linear equations can be immediately obtained from the reduced echelon form to which the augmented matrix has been transformed.

Definition 1.2.1. A matrix is in **echelon form** if

(a) All rows consisting of zeros, if any, are on the bottom of the matrix.

(b) The first non-zero entry in each row is a 1, called the **leading one** of the row.

(c) If two rows contain leading ones, the higher row will have its leading one to the left of the leading one of the lower row.

Definition 1.2.2. A matrix is in **reduced echelon form**[1]

(a) All rows consisting of zeros, if any, are on the bottom of the matrix.

(b) The first non-zero entry in each row is a 1, called the **leading one** of the row.

(c) If two rows contain leading ones, the higher row will have its leading one to the left of the leading one of the lower row.

(d) A column containing a leading one has all other entries in the column equal to 0.

A matrix in reduced echelon form will have the general form

$$\begin{bmatrix} 1 & * & * & 0 & * & 0 & * & 0 & * & * & * \\ 0 & 0 & 0 & 1 & * & 0 & * & 0 & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 1 & * & 0 & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

where $*$ represents any real number.

[1] Some texts call this *reduced row echelon form*.

Example 1. Two matrices that are in reduced echelon form appear below:

$$M = \begin{bmatrix} 1 & 2 & 0 & 0 & 5 & -1 \\ 0 & 0 & 1 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 & 2 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad N = \begin{bmatrix} 1 & 2 & 3 & 0 & 5 & 0 \\ 0 & 0 & 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Theorem 1.2.3. Every matrix B is **row equivalent** to a matrix C in reduced echelon form.

We will illustrate the proof of Theorem 1.2.3 in our next example.

Example 2. Using elementary row operations we shall transform the matrix

$$B = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ -1 & 1 & 0 & -1 \end{bmatrix}.$$

into a matrix C in reduced echelon form as follows:

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

$$R_1 + R_3 \to R_3$$

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

$$-2R_2 + R_1 \to R_1$$

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

$$-3R_2 + R_3 \to R_3$$

$$\begin{bmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & -3 & -6 \end{bmatrix}$$

$$-\frac{1}{3}R_3 \to R_3$$

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

$$-2R_3 + R_2 \to R_2$$

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

$$R_3 + R_1 \to R_1$$

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

The matrix C is in *reduced echelon form*. This completes the example.

1.2.1 Solving a Linear System by Reduction to Echelon Form

A system of m linear equations in the n unknowns $x_1, \dots, x_n$ has the form

$$\begin{aligned} a_{11}x_1 + \ a_{12}x_2 + \cdots + \ a_{1n}x_n &= b_1 \\ a_{21}x_1 + \ a_{22}x_2 + \cdots + \ a_{2n}x_n &= b_2 \\ a_{31}x_1 + \ a_{32}x_2 + \cdots + \ a_{3n}x_n &= b_3 \\ \vdots \qquad\qquad & \ \ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned} \tag{1.6}$$

where the a_{ij} and b_i are constants.

Definition 1.2.4. We say that the list of numbers $(x_1, \dots, x_n)$ is a solution to the system (1.6) if it satisfies **all** the equations in the system. The set

$$\mathcal{S} = \{(x_1, \dots, x_n) : (x_1, \dots, x_n) \text{ is a solution to the system (1.6)}\}$$

is called the **solution set** of the system (1.6). The system (1.6) is **consistent** if it has a solution, that is, $\mathcal{S} \neq \varnothing$. If the system (1.6) has no solutions then it is **inconsistent**.

Question. How to find the solution set $\mathcal{S}$ to the system (1.6)?

Answer. Since the algorithm we use to solve the system (1.6) makes no use of the variables $(x_1, \dots, x_n)$, we consider the augmented $m \times (n+1)$ matrix

$$[A \,|\, \mathbf{b}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{bmatrix}$$

and perform *Row Operations* on $[A \,|\, \mathbf{b}]$ to transforms this matrix to a new matrix $[C \,|\, \mathbf{d}]$ which is in *reduced echelon form*, obtaining the augmented matrix $[C \,|\, \mathbf{d}]$. The system of equations associated to the augmented matrix $[C \,|\, \mathbf{d}]$ is easy to solve and has exactly the same solutions as the original system (1.6).

1.2.2 Gauss-Jordon Elimination

We now give a procedure for solving a system of equations, by transforming the augmented matrix into another augmented matrix. The system represented by the new augmented matrix is easier to solve and has the same solution set as the original system of linear equations.

Gauss-Jordan Elimination Prodedure. To solve a linear system of equations:

Step 1. Form the augmented matrix $[A \,|\, \mathbf{b}]$ of the system.

Step 2. Transform the augmented matrix $[A \,|\, \mathbf{b}]$ into *reduced echelon form* $[C \,|\, \mathbf{d}]$.

Step 3. Solve the new system of equations having the augmented matrix $[C \,|\, \mathbf{d}]$.

Problem 3. Find the solution set of the system of equations

$$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 4 \\ x_2 + 2x_3 &= 3 \\ -x_1 + \ \ x_2 \qquad\quad &= -1. \end{aligned} \tag{1.7}$$

Solution. We will (1) form the augmented matrix $[A\,|\,\mathbf{b}]$ of the system (1.7), (2) transform the augmented matrix into reduced echelon form $[C\,|\,\mathbf{d}]$, and (3) solve the new system of equations having the augmented matrix $[C\,|\,\mathbf{d}]$. The solution of the system with augmented matrix $[C\,|\,\mathbf{d}]$ will also be the solution of the system (1.7).

We apply Step 1 of the Gauss-Jordon elimination method. The augmented matrix for the system (1.7) is as follows:

$$[A\,|\,\mathbf{b}] = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ -1 & 1 & 0 & -1 \end{bmatrix}.$$

We now apply Step 2 and transform this augmented matrix into the reduced echelon form $[C\,|\,\mathbf{d}]$. This was done in Example 2 above. Thus,

$$[C\,|\,\mathbf{d}] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 2 \end{bmatrix}.$$

Apply Step 3 and solve the the new system of equations with augmented matrix $[C\,|\,\mathbf{d}]$. This new system has the same solution set as (1.7):

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

and thus the solution set is $\mathcal{S} = \{(0, -1, 2)\}$. Thus, the linear system (1.7) has only one solution.

In the following problem, we shall again use the Gauss-Jordon Elimination to solve a system of linear equations. This linear system will have an infinite number of solutions.

Problem 4. Find the solution set of the system of equations

$$\begin{aligned} x_3 + 2x_4 - x_5 &= 4 \\ x_4 - x_5 &= 3 \\ x_3 + 3x_4 - 2x_5 &= 7 \\ 2x_1 + 4x_2 + x_3 + 7x_4 \qquad &= 7 \end{aligned} \tag{1.8}$$

by Gauss-Jordan elimination.

Solution. We will (1) form the augmented matrix $[A\,|\,\mathbf{b}]$ of the system (1.8), (2) transform the augmented matrix into reduced echelon form $[C\,|\,\mathbf{d}]$, and (3) solve the new system of equations having the augmented matrix $[C\,|\,\mathbf{d}]$. The solution of the system with augmented matrix $[C\,|\,\mathbf{d}]$ will also be the solution of the system (1.8).

We now apply Step 1 of the Gauss-Jordon elimination method. We construct the the augmented matrix

$$[A\,|\,\mathbf{b}] = \begin{bmatrix} 0 & 0 & 1 & 2 & -1 & 4 \\ 0 & 0 & 0 & 1 & -1 & 3 \\ 0 & 0 & 1 & 3 & -2 & 7 \\ 2 & 4 & 1 & 7 & 0 & 7 \end{bmatrix}.$$

We now perform the following the row operations on the augmented matrix:

$$[A\,|\,\mathbf{b}] = \left[\begin{array}{ccccc|c} 0 & 0 & 1 & 2 & -1 & 4 \\ 0 & 0 & 0 & 1 & -1 & 3 \\ 0 & 0 & 1 & 3 & -2 & 7 \\ 2 & 4 & 1 & 7 & 0 & 7 \end{array}\right]$$

$$(R_1 \leftrightarrow R_4)\,\&\,(R_2 \leftrightarrow R_3)$$

$$\left[\begin{array}{ccccc|c} 2 & 4 & 1 & 7 & 0 & 7 \\ 0 & 0 & 1 & 3 & -2 & 7 \\ 0 & 0 & 0 & 1 & -1 & 3 \\ 0 & 0 & 1 & 2 & -1 & 4 \end{array}\right]$$

$$-R_2 + R_4 \to R_4$$

$$\left[\begin{array}{ccccc|c} 2 & 4 & 1 & 7 & 0 & 7 \\ 0 & 0 & 1 & 3 & -2 & 7 \\ 0 & 0 & 0 & 1 & -1 & 3 \\ 0 & 0 & 0 & -1 & 1 & -3 \end{array}\right]$$

$$R_3 + R_4 \to R_4$$

$$\left[\begin{array}{ccccc|c} 2 & 4 & 1 & 7 & 0 & 7 \\ 0 & 0 & 1 & 3 & -2 & 7 \\ 0 & 0 & 0 & 1 & -1 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

$$-R_2 + R_1 \to R_1$$

$$\left[\begin{array}{ccccc|c} 2 & 4 & 0 & 4 & 2 & 0 \\ 0 & 0 & 1 & 3 & -2 & 7 \\ 0 & 0 & 0 & 1 & -1 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

$$\frac{1}{2}R_1 \to R_1$$

$$\left[\begin{array}{ccccc|c} 1 & 2 & 0 & 2 & 1 & 0 \\ 0 & 0 & 1 & 3 & -2 & 7 \\ 0 & 0 & 0 & 1 & -1 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

$$(-3R_3 + R_2 \to R_2)\,\&\,(-2R_3 + R_1 \to R_1)$$

$$[C\,|\,\mathbf{d}] = \left[\begin{array}{ccccc|c} 1 & 2 & 0 & 0 & 3 & -6 \\ 0 & 0 & 1 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 & -1 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

The matrix $[C\,|\,\mathbf{d}]$ is in reduced echelon form, and from $[C\,|\,\mathbf{d}]$ we get the following new system of equations which has the same solution set as (1.8):

$$\begin{aligned} x_1 + 2x_2 + \qquad\quad + 3x_5 &= -6 \\ x_3 \quad + \;\, x_5 &= -2 \\ x_4 - \;\, x_5 &= 3. \end{aligned}$$

In the above system of equations, the variables x_1, x_3, x_4 are called the *leading variables* or the *dependent variables*. On the other hand, the variables x_2 and x_5 are called the *free variables* or the *independent variables*. Now, solving for the leading variables x_1, x_3, x_4 we obtain

$$\begin{aligned} x_1 &= -2x_2 - 3x_5 - 6 \\ x_3 &= -x_5 - 2 \\ x_4 &= x_5 + 3. \end{aligned}$$

Since x_2 and x_5 are *free variables,* we will replace them with r and s, respectively, to obtain the solution

$$\begin{aligned} x_1 &= -2r - 3s - 6 \\ x_2 &= r \\ x_3 &= -s - 2 \\ x_4 &= s + 3 \\ x_5 &= s \end{aligned}$$

where r and s are arbitrary, and hence, the solution set is

$$\mathcal{S} = \{(-2r - 3s - 6, r, -s - 2, s + 3, s) : r \text{ and } s \text{ are arbitrary real numbers}\}.$$

Therefore, there are an infinite number of solutions to the system (1.8).

Remark 1.2.5. When a consistent system has free variables in its "reduced echelon form system," then the original system always has an infinite number of solutions.

1.2.3 How to Recognize an Inconsistent System?

Suppose we have the linear system of equations

$$\begin{aligned} a_{11}x_1 + \ a_{12}x_2 + \cdots + \ a_{1n}x_n &= b_1 \\ a_{21}x_1 + \ a_{22}x_2 + \cdots + \ a_{2n}x_n &= b_2 \\ a_{31}x_1 + \ a_{32}x_2 + \cdots + \ a_{3n}x_n &= b_3 \\ \vdots \qquad\qquad & \quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned} \tag{1.9}$$

and we want to determine if the system has a solution. We form the augmented matrix

$$[A \,|\, \mathbf{b}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{bmatrix}$$

and perform row operations on $[A \,|\, \mathbf{b}]$ to transform this matrix to a new matrix $[C \,|\, \mathbf{d}]$ in reduced echelon form. If $[C \,|\, \mathbf{d}]$ has the form, say

$$[C \,|\, \mathbf{d}] = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} & 0 \\ c_{21} & c_{22} & \cdots & c_{2n} & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix},$$

then the system (1.9) **has no** solutions. Why? Because this last non-zero row represents the equation

$$0x_1 + 0x_2 + \cdots + 0x_n = 1.$$

Since this equation has no solutions, the original system (1.9) is **inconsistent**, that is, the system (1.9) has no solutions.

Exercises 1.2

Pages 26 to 27 of text – #11, 13, 15, 17, 19, 23, 25, 27, 29, 37, 39.

1.3 Consistent Systems of Linear Equations

Theorem 1.3.1. Let $[A\,|\,\mathbf{b}]$ be the augmented matrix for the *consistent* system in n unknowns

$$\begin{aligned} a_{11}x_1 + \ a_{12}x_2 + \cdots + \ a_{1n}x_n &= b_1 \\ a_{21}x_1 + \ a_{22}x_2 + \cdots + \ a_{2n}x_n &= b_2 \\ a_{31}x_1 + \ a_{32}x_2 + \cdots + \ a_{3n}x_n &= b_3 \\ \vdots \qquad\qquad & \ \ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m. \end{aligned} \tag{1.10}$$

Let $[C\,|\,\mathbf{d}]$ be the reduced echelon form of $[A\,|\,\mathbf{b}]$. Suppose that $[C\,|\,\mathbf{d}]$ has r many leading ones. Then $r \le n$ and there are $n - r$ free variables in the final solution to (1.10). In addition,

1. If $r = n$, then the solution to (1.10) is unique.
2. If $r < n$, then there are an infinite number of solutions to (1.10).

Proof. (Review Problem 4 on page 22.) We consider the augmented matrix

$$[A\,|\,\mathbf{b}] = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} & b_m \end{bmatrix}$$

and perform row operations on $[A\,|\,\mathbf{b}]$ to transform this matrix to the new matrix $[C\,|\,\mathbf{d}]$ in reduced echelon form. Since the system (#) is consistent, the reduced echelon matrix $[C\,|\,\mathbf{d}]$ must be of the form (with all non-zero rows on top)

$$[C\,|\,\mathbf{d}] = \begin{bmatrix} 1 & c_{12} & c_{13} & \cdots & c_{1n} & d_1 \\ 0 & c_{22} & c_{23} & \cdots & c_{2n} & d_2 \\ 0 & 0 & c_{33} & \cdots & c_{3n} & d_3 \\ 0 & 0 & 0 & \cdots & c_{4n} & d_4 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & c_{rn} & d_r \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

Let r be the number of leading ones in the matrix $[C \mid \mathbf{d}]$. Recall that n is the number of variables in the system (#) and that n is also the number of columns in the matrix C. Each non-zero row in $[C \mid \mathbf{d}]$ has exactly one leading 1. Since the system is consistent, we see that no leading 1 can occur in $\mathbf{d}$. So all of the leading 1's must occur in some (or all) of the columns in C. Therefore, $r \leq n$.

In addition, if $r = n$, then there is a unique solution, because there would be no free variables. Also, if $r < n$, then there are $n - r > 0$ many free variables and hence, there are an infinite number of solutions (we may choose **any values** we want for the free variables and then get the required values for the leading variables and thus, we can get an infinite number of solutions to the system). This completes the proof. □

Corollary 1.3.2. Let $[A \mid \mathbf{b}]$ be the augmented matrix for the system in n unknowns and m equations:

$$\begin{aligned} a_{11}x_1 + \ a_{12}x_2 + \cdots + \ a_{1n}x_n &= b_1 \\ a_{21}x_1 + \ a_{22}x_2 + \cdots + \ a_{2n}x_n &= b_2 \\ a_{31}x_1 + \ a_{32}x_2 + \cdots + \ a_{3n}x_n &= b_3 \\ \vdots \qquad\qquad & \quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m. \end{aligned} \tag{1.11}$$

If $m < n$, then either the system (1.11) is inconsistent or it has an infinite number of solutions.

Proof. Assume that $m < n$. If the system is inconsistent, then we are done. So, suppose that the system is consistent. Clearly, the number of rows in the reduced echelon form $[C \mid \mathbf{d}]$ is equal to m (since the matrix $[A \mid \mathbf{b}]$ has m rows). Hence, one can see that the number r of non-zero rows in the matrix $[C \mid \mathbf{d}]$ must satisfy $r \leq m$. Now since since $m < n$, it follows that $r < n$. Theorem 1.3.1 implies (see item 2 above) that the system must have an infinite number of solutions. □

1.3.1 Homogeneous Linear Systems of Equations

Definition 1.3.3. A system of linear equations is said to be *homogeneous* if all the b_i's are 0, that is, the system has the following form:

$$\begin{aligned} a_{11}x_1 + \ a_{12}x_2 + \cdots + \ a_{1n}x_n &= 0 \\ a_{21}x_1 + \ a_{22}x_2 + \cdots + \ a_{2n}x_n &= 0 \\ a_{31}x_1 + \ a_{32}x_2 + \cdots + \ a_{3n}x_n &= 0 \\ \vdots \qquad\qquad & \quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= 0. \end{aligned} \tag{1.12}$$

Remark 1.3.4. The homogeneous system (1.12) is *always consistent* because it has the **trivial solution** $x_1 = 0, x_2 = 0, \ldots, x_n = 0$. We are interested in homogeneous systems which have **non-trivial solutions**.

The next Theorem 1.3.5 gives a condition which, when satisfied, will guarantee that a homogeneous system has an infinite number of solutions. This theorem follows immediately from the above Corollary 1.3.2; however, we shall present a proof of Theorem 1.3.5 that "illustrates" the proof of Corollary 1.3.2.

Theorem 1.3.5. If the homogeneous system (1.12) of linear equations has more unknowns than equations (that is, if $m < n$), then the system has infinitely many solutions.

Proof. Since the argument is easy to generalize, we shall consider only the special case when $m = 3$ and $n = 5$. So, let $[A\,|\,\boldsymbol{\theta}]$ be the augmented matrix

$$[A\,|\,\boldsymbol{\theta}] = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & 0 \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & 0 \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & 0 \end{bmatrix}.$$

After performing row operations on $[A\,|\,\boldsymbol{\theta}]$, we can transform this matrix to a new matrix $[C\,|\,\boldsymbol{\theta}]$ in reduced echelon form. **[Note: the column of zeros $\boldsymbol{\theta}$ is preserved under elementary row operations.]** Now, because there are only three rows in the matrix $[A\,|\,\boldsymbol{\theta}]$, there can be at most three leading 1's in the matrix $[C\,|\,\boldsymbol{\theta}]$. (This implies that at least two of the five unknowns can be chosen arbitrarily.) Without loss of generality we can assume that $[C\,|\,\boldsymbol{\theta}]$ has the form

$$[C\,|\,\boldsymbol{\theta}] = \begin{bmatrix} 1 & c_{12} & 0 & 0 & c_{15} & 0 \\ 0 & 0 & 1 & 0 & c_{25} & 0 \\ 0 & 0 & 0 & 1 & c_{35} & 0 \end{bmatrix}$$

Since x_1, x_3, x_4 are leading variables, x_2 and x_5 are free variables. We will replace x_2 and x_5 with r and s, respectively, to obtain the system solution by solving for the leading variables x_1, x_3, x_4 (see the previous examples):

$$\begin{aligned} x_1 &= -c_{12}r - c_{15}s \\ x_2 &= r \\ x_3 &= -c_{25}s \\ x_4 &= -c_{35}s \\ x_5 &= s \end{aligned}$$

where r and s are arbitrary, and hence, there are an infinite number of solutions to the system. This completes the proof of the theorem. □

Exercises 1.3

Pages 37 to 38 of text – #1, 2, 3, 4, 21, 22, 23, 24.

1.5 Matrix Operations

Definition 1.5.1. An $m \times n$ **matrix** is a rectangular array of numbers of the form

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

and the number a_{ij} is called the ij^{th} entry of the matrix A. We shall sometimes write $A = [a_{ij}]_{m\times n}$ as shorthand. Also, we shall sometimes write $A_{m\times n}$ when we want to emphasize the "size" of the matrix A. For notational convenience we shall write

$$\mathbf{a}_i = \begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{bmatrix}$$

for the i^{th} row of the matrix A; and we shall write

$$\mathbf{A}_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}$$

for the j^{th} column of the matrix A.

For example, the 4×5 matrix

$$A_{4\times 5} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ -2 & 0 & 20 & \pi & -4 \\ -3 & -1 & 3 & 6 & 9 \\ -4 & 4 & 5 & -4 & 8 \end{bmatrix}$$

has $a_{12} = 2$, and $a_{43} = 5$. The 2nd row of A is

$$\mathbf{a}_2 = \begin{bmatrix} -2 & 0 & 20 & \pi & -4 \end{bmatrix}$$

and the 3rd column of A is

$$\mathbf{A}_3 = \begin{bmatrix} 3 \\ 20 \\ 3 \\ 5 \end{bmatrix}.$$

Definition 1.5.2. Two matrices, say $A = [a_{ij}]_{m\times n}$ and $B = [b_{ij}]_{m\times n}$, are equal if and only if they have the same size and entries, that is, $a_{ij} = b_{ij}$ for all i, j.

1.5.1 Matrix Addition and Scalar Multiplication

Definition 1.5.3. Given $A = [a_{ij}]_{m\times n}$ and $B = [b_{ij}]_{m\times n}$, we define $A + B = [c_{ij}]_{m\times n}$ where $c_{ij} = a_{ij} + b_{ij}$.

Example 1.

$$\begin{bmatrix} 1 & -1 & 2 & 3 \\ 0 & 1 & 0 & -40 \\ 2 & 0 & 3 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 & -1 & 3 \\ 1 & 0 & 1 & 41 \\ 4 & 0 & -1 & 9 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 1 & 6 \\ 1 & 1 & 1 & 1 \\ 6 & 0 & 2 & 10 \end{bmatrix}$$

Definition 1.5.4. Given $A = [a_{ij}]_{m\times n}$ and a scalar (real number) d, define **scalar multiplication** by $dA = [da_{ij}]_{m\times n}$.

Example 2.

$$2\begin{bmatrix} 1 & -1 & 2 & 3 \\ 0 & 1 & 0 & -40 \\ 2 & 0 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 2 & -2 & 4 & 6 \\ 0 & 2 & 0 & -80 \\ 4 & 0 & 6 & 2 \end{bmatrix}$$

1.5.2 Vectors in $\mathbb{R}^n$

Definition 1.5.5. $\mathbb{R}^n$ is the space (set) of all n-**vectors** $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ where each **component** x_i is a real number and is called the i-th **coordinate**.

Definition 1.5.6 (Vector Addition)**.** The **addition** of vectors $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ in $\mathbb{R}^n$ is defined by $\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}$.

Definition 1.5.7 (Scalar Multiplication)**.** The **scalar multiplication** of a vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ by a scalar c is defined by $c\mathbf{x} = \begin{bmatrix} cx_1 \\ cx_2 \\ \vdots \\ cx_n \end{bmatrix}$.

Problem 3 (Combining vector addition and scalar multiplication)**.** Consider the two vectors in $\mathbb{R}^4$ given by

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \text{ and } \mathbf{y} = \begin{bmatrix} -2 \\ -3 \\ 1 \\ 2 \end{bmatrix}.$$

Evaluate the "linear combination" $x_1\mathbf{x} + x_2\mathbf{y}$, where x_1, x_2 are scalars in $\mathbb{R}$.

Solution.

$$x_1\mathbf{x} + x_2\mathbf{y} = x_1 \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ -3 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1x_1 \\ 2x_1 \\ 3x_1 \\ 4x_1 \end{bmatrix} + \begin{bmatrix} -2x_2 \\ -3x_2 \\ 1x_2 \\ 2x_2 \end{bmatrix} = \begin{bmatrix} 1x_1 - 2x_2 \\ 2x_1 - 3x_2 \\ 3x_1 + x_2 \\ 4x_1 + 2x_2 \end{bmatrix}.$$

Problem 4 (Going Backwards)**.** Suppose that you are given the vector

$$\mathbf{z} = \begin{bmatrix} 2x_1 \\ -3x_1 + x_2 \\ -x_2 \\ x_1 - 4x_2 \end{bmatrix}$$

in $\mathbb{R}^4$, where x_1, x_2 are scalars in $\mathbb{R}$. Find vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^4$ so that $\mathbf{z} = x_1\mathbf{x} + x_2\mathbf{y}$.

Solution.

$$\mathbf{z} = \begin{bmatrix} 2x_1 \\ -3x_1 + x_2 \\ -x_2 \\ x_1 - 4x_2 \end{bmatrix} = x_1 \begin{bmatrix} 2 \\ -3 \\ 0 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ -1 \\ -4 \end{bmatrix}.$$

1.5.3 Vector Forms of General Solutions of Linear Systems

Having defined vectors, vector addition and vector scalar multiplication, we now derive the **vector form of the generalized solution** of a linear system as follows:

Problem 5. Find the vector form of the generalized solution of the system of equations

$$\begin{aligned} x_3 + 2x_4 - x_5 &= 4 \\ x_4 - x_5 &= 3 \\ x_3 + 3x_4 - 2x_5 &= 7 \\ 2x_1 + 4x_2 + x_3 + 7x_4 &= 7 \end{aligned} \tag{1.13}$$

Solution. Using Gauss-Jordan elimination on the augmented matrix

$$[A \,|\, \mathbf{b}] = \begin{bmatrix} 0 & 0 & 1 & 2 & -1 & 4 \\ 0 & 0 & 0 & 1 & -1 & 3 \\ 0 & 0 & 1 & 3 & -2 & 7 \\ 2 & 4 & 1 & 7 & 0 & 7 \end{bmatrix}$$

we obtain (see Problem 4 on page 22) the reduced echelon form matrix

$$[C \,|\, \mathbf{d}] = \begin{bmatrix} 1 & 2 & 0 & 0 & 3 & -6 \\ 0 & 0 & 1 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 & -1 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

From $[C \,|\, \mathbf{d}]$ we get the following new system of equations which has the same solution set as (1.13):

$$\begin{aligned} x_1 + 2x_2 + + 3x_5 &= -6 \\ x_3 + x_5 &= -2 \\ x_4 - x_5 &= 3. \end{aligned}$$

Note that the *independent variables* are x_2 and x_5, and the *dependent variables* are x_1, x_3, x_4. Now, solving for the leading variables x_1, x_3, x_4 we obtain

$$\begin{aligned} x_1 &= -2x_2 - 3x_5 - 6 \\ x_3 &= -x_5 - 2 \\ x_4 &= x_5 + 3. \end{aligned}$$

Since x_2 and x_5 are *free variables,* we get the solution

$$\begin{aligned} x_1 &= -2x_2 - 3x_5 - 6 \\ x_2 &= x_2 \\ x_3 &= -x_5 - 2 \\ x_4 &= x_5 + 3 \\ x_5 &= x_5. \end{aligned}$$

Putting this into vector notation, we obtain

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} -2x_2 - 3x_5 - 6 \\ x_2 \\ -x_5 - 2 \\ x_5 + 3 \\ x_5 \end{bmatrix}.$$

Thus the the vector form of the generalized solution of the system (1.13) is given by

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -3 \\ 0 \\ -1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} -6 \\ 0 \\ -2 \\ 3 \\ 0 \end{bmatrix}$$

where x_2 and x_5 are arbitrary real numbers.

1.5.4 Matrix Multiplication

Definition 1.5.8. Given $A = [a_{ij}]_{m\times p}$ and $B = [b_{ij}]_{p\times n}$, we define $AB = [a_{ij}]_{m\times p}[b_{ij}]_{p\times n} = [c_{ij}]_{m\times n}$ where

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ip}b_{pj}$$

and so,

$$AB = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ & \vdots & & \\ \hline a_{i1} & a_{i2} & \cdots & a_{ip} \\ \hline & \vdots & & \\ a_{m1} & a_{m2} & \cdots & a_{mp} \end{bmatrix} \left[\begin{array}{ccc|c|cc} b_{11} & b_{12} & \cdots & b_{1j} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2j} & \cdots & b_{2n} \\ & \vdots & \vdots & \vdots & \vdots & \\ b_{p1} & b_{m2} & \cdots & b_{pj} & \cdots & b_{pn} \end{array}\right] = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \boxed{c_{ij}} & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mn} \end{bmatrix}$$

Example 6.

$$\begin{bmatrix} 2 & -2 & 1 & 3 \\ 0 & 2 & 0 & -8 \\ 4 & 0 & 3 & 2 \end{bmatrix} \begin{bmatrix} 5 & 1 & 2 \\ 3 & -1 & -3 \\ 2 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 6 & 4 & 11 \\ 6 & -2 & -6 \\ 26 & 4 & 11 \end{bmatrix}$$

Remark 1.5.9. $AB = A_{m\times p}B_{p\times n}$ is defined if and only if the "inside dimensions" agree.

Remark 1.5.10. Even when AB and BA are both defined, it does not follow that $AB = BA$. Thus, **matrix multiplication does not satisfy the commutative law.**

Remark 1.5.11. $AB = AC$ does not necessarily imply that $B = C$. Thus, **matrix multiplication does not satisfy the cancellation law.**

1.5.5 Using Matrix Equations to Represent Linear Systems

Using matrix multiplication, a set of m linear equations in the n unknowns $x_1, \ldots, x_n$

$$\begin{aligned} a_{11}x_1 + \ a_{12}x_2 + \cdots + \ a_{1n}x_n &= b_1 \\ a_{21}x_1 + \ a_{22}x_2 + \cdots + \ a_{2n}x_n &= b_2 \\ a_{31}x_1 + \ a_{32}x_2 + \cdots + \ a_{3n}x_n &= b_3 \\ \vdots \qquad\qquad\qquad & \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned} \tag{1.14}$$

can be written as

$$\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ & \vdots & & \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

or more simply as

$$A\mathbf{x} = \mathbf{b}$$

where $A = [a_{ij}]$ is the coefficient matrix of the system (#), $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$.

Example 7. Given the system of equations

$$\begin{aligned} x_1 + \ \ x_2 + \ \ 2x_3 &= 9 \\ 2x_1 + 9x_2 - 10x_3 &= 39 \\ x_1 - 6x_2 + \quad x_3 &= -27. \end{aligned} \tag{1.15}$$

the coefficient matrix is given by

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 9 & -10 \\ 1 & -6 & 1 \end{bmatrix}$$

and the column vector $\mathbf{b}$ is given by $\mathbf{b} = \begin{bmatrix} 9 \\ 39 \\ -27 \end{bmatrix}$. The column vector of the unknowns is $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$. Thus, the linear system (1.15) can be written in matrix notation as $A\mathbf{x} = \mathbf{b}$.

Problem 8. Let A and $\mathbf{b}$ be as in the above Example 7. Solve the matrix equation $A\mathbf{x} = \mathbf{b}$ for $\mathbf{x}$.

Solution. Form the matrix $[A \,|\, \mathbf{b}]$ and put it into reduced echelon form. You will get the solution $\mathbf{x} = \begin{bmatrix} 2 \\ 5 \\ 1 \end{bmatrix}$ (see Problem 5 on page 17).

1.5.6 Other Ways to View Matrix Multiplication

Theorem 1.5.12. Let A be the $p \times m$ matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mp} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_p \end{bmatrix}$$

where $\mathbf{A}_j \in \mathbb{R}^m$ denotes the j^{th} column of A. Given a $p \times 1$ column vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}$ the matrix product $A\mathbf{x}$ is a $m \times 1$ column vector. In addition, $A\mathbf{x}$ can be written as a "linear combination" of the columns of A as follows:

$$A\mathbf{x} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mp} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} = x_1\mathbf{A}_1 + x_2\mathbf{A}_2 + \cdots + x_p\mathbf{A}_p.$$

Theorem 1.5.13. Let A be an $m \times p$ matrix and let B be the $p \times n$ matrix

$$B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ b_{p1} & b_{p2} & \cdots & b_{pn} \end{bmatrix} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 & \cdots & \mathbf{B}_n \end{bmatrix}$$

where $\mathbf{B}_j \in \mathbb{R}^p$ denotes the j^{th} column of B. The product $A\mathbf{B}_j$ is an $m \times 1$ column vector and, in addition, the matrix product AB can be expressed in another way:

$$AB = \begin{bmatrix} A\mathbf{B}_1 & A\mathbf{B}_2 & \cdots & A\mathbf{B}_n \end{bmatrix}.$$

Exercises 1.5

Pages 58 to 60 of text – #1, 7, 9, 11, 13, 15, 31, 33, 43, 45, 47, 61, 65(b).

1.6 Algebraic Properties of Matrix Operations

Definition 1.6.1. We shall write $\mathcal{O} = \mathcal{O}_{m\times n} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$ for the $m \times n$ **zero matrix**.

Definition 1.6.2. When A is a $n \times n$ matrix of the form

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

we say that A is a **square matrix** and we call the entries $a_{11}, a_{22}, \ldots, a_{nn}$ the **main diagonal**. The $n \times n$ **identity matrix** is defined to be

$$I = I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

where 1's are on the diagonal, for the $n \times n$ **identity matrix**.

Theorem 1.6.3 (Matrix Addition)**.** The follow algebraic properties hold for matrix addition:

1. $A + B = B + A$ (addition commutes)
2. $A + (B + C) = (A + B) + C$ (addition is associative)
3. $A + \mathcal{O} = \mathcal{O} + A = A$, where $\mathcal{O}$ has the same size as A (additive identity)
4. $A - A = \mathcal{O}$ (additive inverse)

Theorem 1.6.4 (Matrix Multiplication)**.** The follow algebraic properties hold for matrix multiplication:

1. $A(BC) = (AB)C$ (multiplication is associative)
2. $A(B + C) = AB + AC$ (distribution property)
3. $(A + B)C = AC + BC$ (distribution property)
4. $AI = IA = A$, where A is a square matrix (multiplicative identity)

(e) $A\mathcal{O} = \mathcal{O}A = \mathcal{O}$

Theorem 1.6.5 (Scalar Multiplication)**.** If r and s are scalars and A and B are matrices, then

1. $r(sA) = (rs)A$
2. $(r + s)A = rA + sA$
3. $r(A + B) = rA + rB$
4. $A(rB) = r(AB) = (rA)B$.

1.6.1 The Transpose of a Matrix

Given the $m \times n$ matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

we define the **transpose** of A, to be the $n \times m$ matrix

$$A^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \cdots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}$$

The first row of A becomes the first column of A^T, the second row of A becomes the second column of A^T, and so forth. Thus, the last row of A becomes the last column of A^T

Example 1. Here is a matrix and its transpose:

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{bmatrix} \quad A^T = \begin{bmatrix} 1 & 5 & 9 \\ 2 & 6 & 10 \\ 3 & 7 & 11 \\ 4 & 8 & 12 \end{bmatrix}$$

Theorem 1.6.6 (Transpose Properties)**.** If A and B are matrices, then

1. $(A+B)^T = A^T + B^T$
2. $(AB)^T = B^T A^T$
3. $(A^T)^T = A$

Example 2. Here is an example of item 2 in Theorem 1.6.6.Let

$$A = \begin{bmatrix} 1 & 3 & 2 \\ 2 & -1 & 3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 & 1 \\ 2 & 2 \\ 3 & -1 \end{bmatrix}.$$

Then

$$(AB)^T = \left(\begin{bmatrix} 1 & 3 & 2 \\ 2 & -1 & 3 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 2 & 2 \\ 3 & -1 \end{bmatrix} \right)^T = \begin{bmatrix} 12 & 7 \\ 5 & -3 \end{bmatrix}$$

and

$$B^T A^T = \begin{bmatrix} 0 & 2 & 3 \\ 1 & 2 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & -1 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 12 & 7 \\ 5 & -3 \end{bmatrix}.$$

Definition 1.6.7. Given a square $n \times n$ matrix A, we say that A is **symmetric** if $A^T = A$.

For example, the matrix

$$A = \begin{bmatrix} 1 & -3 & 4 & 8 \\ -3 & 6 & 2 & -5 \\ 4 & 2 & 7 & 3 \\ 8 & -5 & 3 & -4 \end{bmatrix}$$

is a symmetric matrix because $A = A^T$.

1.6.2 Powers of a Square Matrix

Given a square matrix A we define, for natural numbers p,

$$A^0 = I$$

$$A^1 = A$$

$$A^p = \underbrace{AA \cdots A}_{p \text{ many times}}$$

Theorem 1.6.8. If A is a square matrix, and p and q are natural numbers, then $A^pA^q = A^{p+q}$ and $(A^p)^q = A^{pq}$.

1.6.3 Scalar Products and Vector Norms

Definition 1.6.9. Let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ and let $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ be vectors in $\mathbb{R}^n$. The **scalar product** (or **dot product**), denoted by $\mathbf{x} \cdot \mathbf{y}$, is defined by

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

Definition 1.6.10. The **zero vector** $\boldsymbol{\theta}$ in $\mathbb{R}^n$ has all its components equal to 0, that is,

$$\boldsymbol{\theta} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

We now state some properties of the scalar product.

Theorem 1.6.11. If $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ are vectors in $\mathbb{R}^n$ and c is a scalar, then

1. $\mathbf{x} \cdot \boldsymbol{\theta} = 0$.
2. $\mathbf{x} \cdot \mathbf{x} \geq 0$. Furthermore, $\mathbf{x} \cdot \mathbf{x} = 0$ if and only if $\mathbf{x} = \boldsymbol{\theta}$
3. $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$
4. $(\mathbf{x} + \mathbf{y}) \cdot \mathbf{z} = \mathbf{x} \cdot \mathbf{z} + \mathbf{y} \cdot \mathbf{z}$
5. $(c\mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (c\mathbf{y}) = c(\mathbf{x} \cdot \mathbf{y})$

Proof. These properties can easily be checked. We prove item 2 by noting that

$$\mathbf{x} \cdot \mathbf{x} = x_1^2 + x_2^2 + \cdots + x_n^2 \geq 0.$$

Furthermore,

$$\begin{aligned} \mathbf{x} \cdot \mathbf{x} = 0 &\text{ if and only if } x_1^2 + x_2^2 + \cdots + x_n^2 = 0 \\ &\text{ if and only if } x_1 = x_2 = \cdots = x_n = 0 \\ &\text{ if and only if } \mathbf{x} = \boldsymbol{\theta}. \end{aligned}$$

This completes the proof of item 2. □

Definition 1.6.12. The **norm** (or length) of a vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ in $\mathbb{R}^n$ is defined by

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

Definition 1.6.13. The **distance** between the vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ and the vector $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ is defined by

$$\|\mathbf{x} - \mathbf{y}\| = \sqrt{(\mathbf{x} - \mathbf{y}) \cdot (\mathbf{x} - \mathbf{y})} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}$$

Exercises 1.6

Pages 69 to 70 of text – #1, 3, 5, 7, 13, 17, 19, 23, 41, 43, 49.

1.7 Linear Independence and Nonsingular Matrices

1.7.1 Linear Combinations

The next idea is one of the most fundamental concepts in Linear Algebra.

Definition 1.7.1. A vector $\mathbf{x}$ is a *linear combination* of the vectors $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n$ if it can be expressed in the form

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

where $x_1, x_2, \ldots, x_n$ are scalars.

Example 1. Consider the three vectors in 4-space:

$$\mathbf{A}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ -1 \end{bmatrix}, \mathbf{A}_2 = \begin{bmatrix} 1 \\ 0 \\ 2 \\ -3 \end{bmatrix}, \mathbf{A}_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ -2 \end{bmatrix}.$$

Is the vector $\mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 5 \\ -5 \end{bmatrix}$ a linear combination of the vectors $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$? In other words, are there are numbers x_1, x_2, x_3 satisfying the vector equation

$$x_1\mathbf{A}_1 + x_2\mathbf{A}_2 + x_3\mathbf{A}_3 = \mathbf{b}? \tag{1.16}$$

To answer the question, first let the vectors $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$ form the columns of a matrix A and let $\mathbf{x}$ be the column vector consisting of the unknowns x_1, x_2, x_3, that is,

$$A = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{A}_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & 1 \\ 1 & 2 & 0 \\ -1 & -3 & -2 \end{bmatrix} \text{ and } \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Now, notice that the following 5 equations are equivalent:

1. $x_1\mathbf{A}_1 + x_2\mathbf{A}_2 + x_3\mathbf{A}_3 = \mathbf{b}$
2. $x_1 \begin{bmatrix} 1 \\ 2 \\ 1 \\ -1 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 0 \\ 2 \\ -3 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 1 \\ 0 \\ -2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 5 \\ -5 \end{bmatrix}$
3. $\begin{bmatrix} 1x_1 & + & 1x_2 & + & 1x_3 \\ 2x_1 & + & 0x_2 & + & 1x_3 \\ 1x_1 & + & 2x_2 & + & 0x_3 \\ -1x_1 & - & 3x_2 & - & 2x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 5 \\ -5 \end{bmatrix}$

4. $\begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & 1 \\ 1 & 2 & 0 \\ -1 & -3 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 5 \\ -5 \end{bmatrix}$

5. $A\mathbf{x} = \mathbf{b}$

Therefore, x_1, x_2, x_3 is a solution to vector equation (1.16) if and only if $\mathbf{x}$ is a solution to the linear system $A\mathbf{x} = \mathbf{b}$. Transforming the augmented matrix $[A \,|\, \mathbf{b}]$ into reduced row echelon form we obtain

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

Therefore, the solution $x_1 = 1$, $x_2 = 2$, and $x_3 = -1$ satisfies vector equation (1.16), that is,

$$\mathbf{b} = 1\mathbf{A}_1 + 2\mathbf{A}_2 + -1\mathbf{A}_3 \text{ (verify!)}.$$

Hence, the vector $\mathbf{b}$ can be written as a linear combination of the vectors $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$.

From the above solution in Example 1, we can make the following observation.

Theorem 1.7.2. Let $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k$ be a list of vectors in $\mathbb{R}^n$; and let A be the matrix

$$A = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_k \end{bmatrix}.$$

A vector $\mathbf{b}$ in $\mathbb{R}^n$ can be written as a linear combination of the vectors $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k$ if and only if the linear system $A\mathbf{x} = \mathbf{b}$ has a solution $\mathbf{x}$.

Linear Combination Algorithm. In $\mathbb{R}^n$, to determine if a given vector $\mathbf{b}$ can be written as a linear combination of the vectors $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k$:

Step 1. Form the equation $x_1\mathbf{A}_1 + x_2\mathbf{A}_2 + \cdots + x_k\mathbf{A}_k = \mathbf{b}$.

Step 2. Rewrite the equation in Step 1 in the form $A\mathbf{x} = \mathbf{b}$ where

$$A = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_k \end{bmatrix} \text{ and } \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix}.$$

Step 3. Using Gauss-Jordon elimination, solve the linear system $A\mathbf{x} = \mathbf{b}$ for $\mathbf{x}$.

Step 4. Either you found a solution or you found that there is no such solution.

- If the system $A\mathbf{x} = \mathbf{b}$ does have a solution $x_1, x_2, \ldots x_k$, then this solution shows that $\mathbf{b}$ **can** be written as a linear combination of vectors $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k$. One should verify that $\mathbf{b} = x_1\mathbf{A}_1 + x_2\mathbf{A}_2 + \cdots + x_k\mathbf{A}_k$.
- If the system $A\mathbf{x} = \mathbf{b}$ has no solution, then $\mathbf{b}$ can **not** be written as a linear combination of vectors $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k$.

1.7.2 Linear Independence

Definition 1.7.3. A set of vectors $S = \{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k\}$ in $\mathbb{R}^n$ is **linearly dependent** if the vector equation

$$x_1\mathbf{A}_1 + x_2\mathbf{A}_2 + \cdots + x_k\mathbf{A}_k = \boldsymbol{\theta} \tag{1.17}$$

has a *non-trivial* solution $x_1, x_2, \ldots, x_k$. The set of vectors S is **linearly independent** if equation (1.17) has *only* the trivial solution $x_1 = x_2 = \cdots = x_k = 0$.

Remark 1.7.4. A set of vectors $S = \{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k\}$ is linearly **in**dependent if and only if $x_1\mathbf{A}_1 + x_2\mathbf{A}_2 + \cdots + x_k\mathbf{A}_k = \boldsymbol{\theta}$ implies $x_1 = x_2 = \cdots = x_k = 0$.

We will often say that the vectors $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k$ are linearly independent rather than say that the set of vectors $\{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k\}$ is linearly independent.

Problem 2. Consider the three vectors in 4-space ($\mathbb{R}^4$):

$$\mathbf{A}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ -1 \end{bmatrix}, \mathbf{A}_2 = \begin{bmatrix} 1 \\ 0 \\ 2 \\ -3 \end{bmatrix}, \mathbf{A}_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ -2 \end{bmatrix}.$$

Are the vectors $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$ linearly independent?

Solution. The answer is "yes," if the only solution x_1, x_2, x_3 to the vector equation

$$x_1\mathbf{A}_1 + x_2\mathbf{A}_2 + x_3\mathbf{A}_3 = \boldsymbol{\theta} \tag{1.18}$$

is the trivial solution $x_1 = x_2 = x_3 = 0$. To answer the question, we first let the vectors $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$ form the columns of a matrix A and let $\mathbf{x}$ be the column vector consisting of the unknowns x_1, x_2, x_3, that is,

$$A = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{A}_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & 1 \\ 1 & 2 & 0 \\ -1 & -3 & -2 \end{bmatrix} \text{ and } \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Once again, notice that the following 3 equations are equivalent:

1. $x_1\mathbf{A}_1 + x_2\mathbf{A}_2 + x_3\mathbf{A}_3 = \boldsymbol{\theta}$
2. $\begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & 1 \\ 1 & 2 & 0 \\ -1 & -3 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$
3. $A\mathbf{x} = \boldsymbol{\theta}$

Therefore, x_1, x_2, x_3 is a solution to vector equation (1.18) if and only if $\mathbf{x}$ is a solution to the linear system $A\mathbf{x} = \boldsymbol{\theta}$. Transforming the augmented matrix $[A \,|\, \boldsymbol{\theta}]$ into reduced row echelon form we obtain

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Therefore, the **only** solution to the vector equation (1.18) is the trivial solution $x_1 = 0$, $x_2 = 0$, and $x_3 = 0$. Hence, the vectors $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$ are linearly independent.

From the above solution for Problem 2, we can make the following observation:

Theorem 1.7.5. Let $S = \{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k\}$ be a set of vectors in $\mathbb{R}^n$; and let A be the matrix

$$A = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_k \end{bmatrix}.$$

The set of vectors S is **linearly independent** if and only if the homogeneous system $A\mathbf{x} = \boldsymbol{\theta}$ has only the *trivial* solution $\mathbf{x} = \boldsymbol{\theta}$.

Linear Dependence Algorithm. In $\mathbb{R}^n$, to determine if a given set of vectors

$$\{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k\}$$

is linearly dependent:

Step 1. Form the equation $x_1\mathbf{A}_1 + x_2\mathbf{A}_2 + \cdots + x_k\mathbf{A}_k = \boldsymbol{\theta}$.

Step 2. Rewrite the equation in Step 1 in the form $A\mathbf{x} = \boldsymbol{\theta}$ where

$$A = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_k \end{bmatrix} \text{ and } \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix}.$$

Step 3. Using Gauss-Jordon elimination, solve the linear system $A\mathbf{x} = \boldsymbol{\theta}$ for $\mathbf{x}$.

Step 4. Either you found a non-trivial solution or that the only solution is the trivial solution.

- If the system $A\mathbf{x} = \boldsymbol{\theta}$ has a *non-trivial* solution $\mathbf{x}$, then the vectors $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k$ are **linearly dependent**. You should verify that your non-trivial solution $x_1, x_2, \ldots, x_k$ satisfies the vector equation $x_1\mathbf{A}_1 + x_2\mathbf{A}_2 + \cdots + x_k\mathbf{A}_k = \boldsymbol{\theta}$.
- If the system $A\mathbf{x} = \boldsymbol{\theta}$ has *only* the *trivial* solution $\mathbf{x} = \boldsymbol{\theta}$ (that is $x_1 = x_2 = \cdots = x_k = 0$), then the vectors $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k$ are **linearly independent**.

Remark 1.7.6. The Linear Dependence Algorithm can also be implemented as follows: To determine whether or not the vectors $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k$ are linearly independent, form the matrix $A = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_k \end{bmatrix}$. Put (transform) the matrix A into reduced echelon form, obtaining the matrix B. If the matrix B has k many leading 1's, then the vectors $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k$ are linearly independent. If the matrix B has **less than** k many leading 1's, then the vectors $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k$ are linearly dependent.

Theorem 1.7.7 (Zero-One Law)**.** Let $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k$ be vectors in $\mathbb{R}^n$. Suppose that each one of these vectors has an i-th coordinate equal to 1 while all of the other vectors have 0 in their i-th coordinate. Then the vectors $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k$ are linearly independent.

We illustrate the proof of Theorem 1.7.7 with an example.

Example 3. Consider the vectors $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$ in $\mathbb{R}^4$

$$\mathbf{A}_1 = \begin{bmatrix} 0 \\ 2 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 1 \\ 3 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{A}_3 = \begin{bmatrix} 0 \\ 4 \\ 0 \\ 1 \end{bmatrix}.$$

Note the $\mathbf{A}_1$ has a 1 in the 3rd coordinate while $\mathbf{A}_2$ and $\mathbf{A}_3$ have 0 in the 3rd coordinate. The vector $\mathbf{A}_2$ has a 1 in the 1st coordinate while $\mathbf{A}_1$ and $\mathbf{A}_3$ have 0 in the 1st coordinate. Finally, we see that $\mathbf{A}_3$ has 1 in the 4th coordinate while $\mathbf{A}_1$ and $\mathbf{A}_3$ have 0 in the 4th coordinate. Thus, each one of these vectors has an i-th coordinate which is 1 while all of the other vectors have 0 in their i-th coordinate. We now show that the vectors $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$ are linearly independent. Suppose that $x_1, x_2, x_3 \in \mathbb{R}$ satisfy

$$x_1\mathbf{A}_1 + x_2\mathbf{A}_2 + x_3\mathbf{A}_3 = \boldsymbol{\theta}.$$

Hence,

$$x_1 \begin{bmatrix} 0 \\ 2 \\ 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 3 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 4 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Hence,

$$\begin{bmatrix} x_2 \\ 2x_1 + 3x_2 + 4x_3 \\ x_1 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

We conclude that $x_1 = x_2 = x_3 = 0$. Therefore, the vectors $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$ are linearly independent.

Definition 1.7.8. The **unit vectors** $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ in $\mathbb{R}^n$ are defined by

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \ldots, \quad \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

We shall refer to $\mathbf{e}_1$ as the 1st unit vector, $\mathbf{e}_2$ as the 2nd unit vector, and so on. More specifically, we shall refer to the unit vector $\mathbf{e}_i$ that contains a 1 in its i-th coordinate as the i-th unit vector.

The following should be noted:

1. It is easy to see that the set $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is linearly independent.

2. $I = [\mathbf{e}_1 \ \mathbf{e}_2 \ \cdots \ \mathbf{e}_n]$ where I is the $n \times n$ identity matrix.

Definition 1.7.9. Let A be a square $n \times n$ matrix. The matrix A is *nonsingular* if the only solution to $A\mathbf{x} = \boldsymbol{\theta}$ is $\mathbf{x} = \boldsymbol{\theta}$. We say that A *singular* if there is a non-trivial solution to $A\mathbf{x} = \boldsymbol{\theta}$, that is, a solution $\mathbf{x} \neq \boldsymbol{\theta}$.

Remark 1.7.10. For a square matrix A, A is nonsingular if and only if $A\mathbf{x} = \boldsymbol{\theta}$ implies $\mathbf{x} = \boldsymbol{\theta}$.

Theorem 1.7.11. Let $A = [\mathbf{A}_1 \ \mathbf{A}_2 \ \cdots \ \mathbf{A}_n]$ be a square $n \times n$ matrix. Then A is nonsingular if and only if the set of vectors $\{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n\}$ is linearly independent.

Proof. This follows from Theorem 1.7.5 above, where $k = n$. □

Problem 4. Let

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & 1 \\ 1 & 2 & 0 \end{bmatrix}.$$

(1) Show that A is nonsingular. (2) Show, for any vector $\mathbf{b}$, that $A\mathbf{x} = \mathbf{b}$ has a unique solution.

Solution. We solve (1) and (2) below:

(1) To show that A is nonsingular, we must show that only solution to $A\mathbf{x} = \boldsymbol{\theta}$ is $\mathbf{x} = \boldsymbol{\theta}$. Putting the augmented matrix

$$[A \,|\, \boldsymbol{\theta}] = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 2 & 0 & 1 & 0 \\ 1 & 2 & 0 & 0 \end{bmatrix}$$

into reduced echelon form, we get

$$[I \,|\, \boldsymbol{\theta}] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Thus the only solution to $A\mathbf{x} = \boldsymbol{\theta}$ is $\mathbf{x} = \boldsymbol{\theta}$. Note that A is row equivalent to the identity matrix I.

(2) We must show, for any vector $\mathbf{b}$, that $A\mathbf{x} = \mathbf{b}$ has a unique solution. By performing the same row operations as in (1), the augmented matrix

$$[A \,|\, \mathbf{b}] = \begin{bmatrix} 1 & 1 & 1 & b_1 \\ 2 & 0 & 1 & b_2 \\ 1 & 2 & 0 & b_3 \end{bmatrix}$$

can be put into reduced echelon form. We will get a matrix of the form

$$[I \,|\, \mathbf{d}] = \begin{bmatrix} 1 & 0 & 0 & d_1 \\ 0 & 1 & 0 & d_2 \\ 0 & 0 & 1 & d_3 \end{bmatrix}.$$

Thus the only solution to $A\mathbf{x} = \mathbf{b}$ is $\mathbf{x} = \mathbf{d}$.

Remark 1.7.12. Let A be a square $n \times n$ matrix. In the system $A\mathbf{x} = \mathbf{b}$ the number of variables $(x_1, \ldots, x_n)$ is the same as the number of equations. Consequently, if the system $A\mathbf{x} = \mathbf{b}$ has a unique solution then, as in the above two examples, the augmented matrix $[A \,|\, \mathbf{b}]$ can be transformed by row operations to the matrix $[I \,|\, \mathbf{d}]$ where I is the $n \times n$ identity matrix. Note that all of the variables in the reduced echelon system $I\mathbf{x} = \mathbf{d}$ are "leading variables" and hence, they are completely determined and the solution is unique.

Problem 5. Let

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & 1 \\ 3 & 1 & 2 \end{bmatrix}.$$

Show that A is singular.

Solution. Putting the augmented matrix

$$[A\,|\,\boldsymbol{\theta}] = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 2 & 0 & 1 & 0 \\ 3 & 1 & 2 & 0 \end{bmatrix}$$

into reduced echelon form, we get

$$[C\,|\,\boldsymbol{\theta}] = \begin{bmatrix} 1 & 0 & \frac{1}{2} & 0 \\ 0 & 1 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Note that $C \neq I$. Thus there is a nontrivial solution to $A\mathbf{x} = \boldsymbol{\theta}$; for example, $\mathbf{x} = \begin{bmatrix} -1/2 \\ -1/2 \\ 1 \end{bmatrix}$ is one of an infinite number of solutions.

The next theorem actually appears as Theorem 16 in section 1.9 of the text. However, we shall now state and prove this theorem.

Theorem 1.7.13. Let A be a square matrix. Then A is nonsingular if and only if A is row equivalent to the identity matrix I.

Proof. Let A be a square matrix.

($\Rightarrow$). First we prove that if A is nonsingular, then A is row equivalent to the identity matrix I. Assume A is nonsingular, that is, the unique solution to $A\mathbf{x} = \boldsymbol{\theta}$ is $\mathbf{x} = \boldsymbol{\theta}$. We prove that A is row equivalent to the identity matrix I. We are assuming that the unique solution to $A\mathbf{x} = \boldsymbol{\theta}$ is $\mathbf{x} = \boldsymbol{\theta}$. Consider the augmented matrix $[A\,|\,\boldsymbol{\theta}]$. Now transform this augmented matrix into *reduced echelon form*, say $[C\,|\,\boldsymbol{\theta}]$, by a sequence of row operations. Since the solution to the system is unique, it follows that there must be n many leading 1's. Therefore, we must have that $[C\,|\,\boldsymbol{\theta}] = [I\,|\,\boldsymbol{\theta}]$. Note that the same sequence of row operations transforms the matrix A to the identity matrix I. Hence, A is row equivalent to the identity matrix I.

($\Leftarrow$). Now we prove that if A is row equivalent to the identity matrix I, then A is nonsingular. Assume A is row equivalent to the identity matrix I. We prove that A is nonsingular, that is, we prove that the only solution to $A\mathbf{x} = \boldsymbol{\theta}$ is $\mathbf{x} = \boldsymbol{\theta}$. We now solve $A\mathbf{x} = \boldsymbol{\theta}$ by using the augmented matrix $[A\,|\,\boldsymbol{\theta}]$. Since A is row equivalent to the identity matrix I, there is a sequence of row operations applied to A that yields the identity matrix I. Apply these same row operations to the matrix $[A\,|\,\boldsymbol{\theta}]$. We will thus obtain the matrix $[I\,|\,\boldsymbol{\theta}]$ and hence, $\mathbf{x} = \boldsymbol{\theta}$ is the unique solution to $A\mathbf{x} = \boldsymbol{\theta}$. Therefore, A is nonsingular. □

Theorem 1.7.14. Let A be a square matrix. For each $\mathbf{b} \in \mathbb{R}^n$ there is a unique solution to $A\mathbf{x} = \mathbf{b}$ if and only if A is nonsingular.

Proof. Let A be a square matrix.

($\Rightarrow$). First we prove that if for each $\mathbf{b} \in \mathbb{R}^n$ there is a unique solution to $A\mathbf{x} = \mathbf{b}$, then A is nonsingular. Assume for each $\mathbf{b} \in \mathbb{R}^n$ there is a unique solution to $A\mathbf{x} = \mathbf{b}$. We prove that A is nonsingular, that is, we prove that the only solution to $A\mathbf{x} = \boldsymbol{\theta}$ is $\mathbf{x} = \boldsymbol{\theta}$. We know that $\mathbf{x} = \boldsymbol{\theta}$ is a solution to $A\mathbf{x} = \boldsymbol{\theta}$. But for $\mathbf{b} = \boldsymbol{\theta}$ our assumption implies that $\mathbf{x} = \boldsymbol{\theta}$ must be the only such solution. Therefore, A is nonsingular.

($\Leftarrow$). Now we prove that if A is nonsingular, then for each $\mathbf{b} \in \mathbb{R}^n$ there is a unique solution to $A\mathbf{x} = \mathbf{b}$. Now since A is nonsingular, we know by Theorem 1.7.13 above that A is row equivalent to the identity matrix I, that is, there is a sequence of row operations which transform A to I. So, we can solving the system $A\mathbf{x} = \mathbf{b}$ by applying these same row operations to the augmented matrix $[A \,|\, \mathbf{b}]$ and thus, obtain $[I \,|\, \mathbf{d}]$. Hence, the only solution to $A\mathbf{x} = \mathbf{b}$ must be $\mathbf{x} = \mathbf{d}$. Therefore, for each $\mathbf{b} \in \mathbb{R}^n$ there is a unique solution to $A\mathbf{x} = \mathbf{b}$. □

Theorem 1.7.15. Let A be a square $n \times n$ matrix. Assume that A is nonsingular. Then for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, if $A\mathbf{x} = A\mathbf{y}$, then $\mathbf{x} = \mathbf{y}$.

Exercises 1.7

Pages 78 to 79 of text – Odds #1–9; 16, 17, 21, 22, 29, 31, odds 35–45; 49. Prove Theorem 1.7.15.

1.9 Matrix Inverses and Their Properties

Theorem 1.9.1. Let A be square $n \times n$ matrix, and let B and C also be $n \times n$ matrices. Suppose that (1) $AB = BA = I$ and (2) $AC = CA = I$. Then $B = C$.

Proof. Let A be square $n \times n$ matrix, and let B and C also be $n \times n$ matrices. Suppose we have that (1) $AB = BA = I$ and (2) $AC = CA = I$. Then

$$\begin{aligned} AB &= BA && \text{by (1)} \\ (AB)C &= (BA)C && \text{multiplying by } C \text{ on the right} \\ (AB)C &= B(AC) && \text{by associativity} \\ IC &= BI && \text{by (1) and (2)} \\ C &= B && \text{because } I \text{ is the identity matrix.} \end{aligned}$$

Therefore, $B = C$ and this completes the proof. □

Theorem 1.9.1 allows us to make the following definition.

Definition 1.9.2. A square matrix A is **invertible**, if there is a matrix B (of the same size as A) such that $AB = BA = I$. We call this matrix B the **inverse** of A and denote it A^{-1}. If a square matrix A has no inverse, then A is called **noninvertible**.

Example 1. Let A be the matrix $A = \begin{bmatrix} 2 & 3 \\ 2 & 2 \end{bmatrix}$. One can check for the matrix $B = \begin{bmatrix} -1 & \frac{3}{2} \\ 1 & -1 \end{bmatrix}$ that $AB = BA = I_2$. Hence, A is has an inverse and $A^{-1} = B$.

1.9.1 Using Inverses to Solve Linear Systems

When a linear system has the same number of equations as the number of variables, there is a new method for solving the system if the square coefficient matrix is invertible.

Theorem 1.9.3. If A is an invertible $n \times n$ matrix, then for any $\mathbf{b} \in \mathbb{R}^n$ the system of equations $A\mathbf{x} = \mathbf{b}$ has exactly one solution.

Proof. We prove that if A is an invertible $n \times n$ matrix, then for each $\mathbf{b} \in \mathbb{R}^n$ the system of equations $A\mathbf{x} = \mathbf{b}$ has exactly one solution. So assume that A is invertible, that is, A^{-1} exists. We prove that for any $\mathbf{b} \in \mathbb{R}^n$ the system of equations $A\mathbf{x} = \mathbf{b}$ has exactly one solution. Let $\mathbf{b} \in \mathbb{R}^n$ be arbitrary. We must show that the system $A\mathbf{x} = \mathbf{b}$ has exactly one solution. That is, we want to solve for the $\mathbf{x} \in \mathbb{R}^n$ such that $A\mathbf{x} = \mathbf{b}$. By multiplying both sides of the equation $A\mathbf{x} = \mathbf{b}$ by A^{-1} we get

$$\begin{aligned} A^{-1}A\mathbf{x} &= A^{-1}\mathbf{b} \\ I\mathbf{x} &= A^{-1}\mathbf{b} \\ \mathbf{x} &= A^{-1}\mathbf{b}. \end{aligned}$$

Therefore the one and only solution is $\mathbf{x} = A^{-1}\mathbf{b}$. This completes the proof of the theorem. □

Example 2. Consider the system of equations

$$\begin{aligned} 3x_1 + 4x_2 - x_3 &= 1 \\ x_1 + 3x_3 &= 2 \\ 2x_1 + 5x_2 - 4x_3 &= -1. \end{aligned}$$

We write this system in matrix form

$$\begin{bmatrix} 3 & 4 & -1 \\ 1 & 0 & 3 \\ 2 & 5 & -4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$$

where $A = \begin{bmatrix} 3 & 4 & -1 \\ 1 & 0 & 3 \\ 2 & 5 & -4 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$. Using a method for getting A^{-1} (see section 1.9.4) one obtains

$$A^{-1} = \begin{bmatrix} \frac{3}{2} & -\frac{11}{10} & -\frac{6}{5} \\ -1 & 1 & 1 \\ -\frac{1}{2} & \frac{7}{10} & \frac{2}{5} \end{bmatrix}.$$

Therefore, the unique solution to the system is given by

$$\mathbf{x} = A^{-1}\mathbf{b} = \begin{bmatrix} \frac{3}{2} & -\frac{11}{10} & -\frac{6}{5} \\ -1 & 1 & 1 \\ -\frac{1}{2} & \frac{7}{10} & \frac{2}{5} \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ 0 \\ \frac{1}{2} \end{bmatrix}$$

1.9.2 A proof that $AB = I$ implies $BA = I$

Lemma 1.9.4. *Let A be a square $n \times n$ matrix. If A is nonsingular, then there is an $n \times n$ matrix B such that $AB = I$.*

Proof. Let A be a square $n \times n$ matrix. We must prove that if A is nonsingular, then there is an $n \times n$ matrix B such that $AB = I$. So, assume that A is nonsingular. We must find a $n \times n$ matrix B and then prove that $AB = I$. Since A is nonsingular, we know by Theorem 1.7.14 that for each $\mathbf{b} \in \mathbb{R}^n$ there is a unique solution to $A\mathbf{x} = \mathbf{b}$. Let us apply this result to each of the unit vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ in $\mathbb{R}^n$. So there is a unique solution to $A\mathbf{x} = \mathbf{e}_1$, call this solution $\mathbf{B}_1$. Let $\mathbf{B}_2$ be the unique solution to $A\mathbf{x} = \mathbf{e}_2$. Continue to get these solutions for each the unit vectors and so, finally let $\mathbf{B}_n$ be the unique solution to $A\mathbf{x} = \mathbf{e}_n$. So, we see that

$$A\mathbf{B}_1 = \mathbf{e}_1,\ A\mathbf{B}_2 = \mathbf{e}_2,\ \ldots,\ A\mathbf{B}_n = \mathbf{e}_n. \tag{1.19}$$

Let B be the $n \times n$ matrix $B = [\mathbf{B}_1\ \mathbf{B}_2\ \cdots\ \mathbf{B}_n]$. We now show that $AB = I$ as follows:

$$\begin{aligned} AB &= A[\mathbf{B}_1\ \mathbf{B}_2\ \cdots\ \mathbf{B}_n] \\ &= [A\mathbf{B}_1\ A\mathbf{B}_2\ \cdots\ A\mathbf{B}_n] && \text{by Theorem 1.5.13} \\ &= [\mathbf{e}_1\ \mathbf{e}_2\ \cdots\ \mathbf{e}_n] && \text{by equations (1.19) above} \\ &= I && \text{since } I = [\mathbf{e}_1\mathbf{e}_2\cdots\mathbf{e}_n]. \end{aligned}$$

Therefore, $AB = I$. □

Lemma 1.9.5. Let A and B be square $n \times n$ matrices. If $AB = I$, then B is nonsingular.

Proof. Let A and B be square $n \times n$ matrices. We must prove that if $AB = I$, then B is nonsingular. So, assume that $AB = I$. We now prove that B is nonsingular, that is, the only solution to $B\mathbf{x} = \boldsymbol{\theta}$ is $\mathbf{x} = \boldsymbol{\theta}$. We can now solve the equation $B\mathbf{x} = \boldsymbol{\theta}$ as follows:

$$\begin{aligned} B\mathbf{x} &= \boldsymbol{\theta} \\ AB\mathbf{x} &= A\boldsymbol{\theta} && \text{multiply on the left by } A \\ AB\mathbf{x} &= \boldsymbol{\theta} && \text{because } A\boldsymbol{\theta} = \boldsymbol{\theta} \\ I\mathbf{x} &= \boldsymbol{\theta} && \text{because } AB = I \\ \mathbf{x} &= \boldsymbol{\theta} && \text{because } I\mathbf{x} = \mathbf{x}. \end{aligned}$$

Thus, the only solution to $B\mathbf{x} = \boldsymbol{\theta}$ is $\mathbf{x} = \boldsymbol{\theta}$. Therefore, B is nonsingular. □ □

Theorem 1.9.6. Let A and B be square $n \times n$ matrices. If $AB = I$, then $BA = I$.

Proof. Let A and B be square $n \times n$ matrices. We prove that if $AB = I$, then $BA = I$. So, assume that $AB = I$. We must prove that $BA = I$. We shall use Lemmas 1.9.4 and 1.9.5 above. Since $AB = I$, we know by Lemma 1.9.5 that B is nonsingular. So, Lemma 1.9.4 implies that there is an $n \times n$ matrix C such that $BC = I$. Using some matrix algebra, we show that $C = A$ as follows:

$$\begin{aligned} AB &= I && \text{by assumption} \\ (AB)C &= IC && \text{multiply on the right by } C \\ (AB)C &= C && \text{because } IC = C \\ A(BC) &= C && \text{by associativity of matrix mult.} \\ AI &= C && \text{because } BC = I \\ A &= C && \text{because } AI = A. \end{aligned}$$

Hence, $A = C$. Therefore, $BA = BC = I$, that is, $BA = I$. □

1.9.3 Existence of Inverses

Theorem 1.9.7. Let A a square matrix. Then A has an inverse if and only if A is nonsingular.

Proof. Let A be a square matrix.

($\Rightarrow$). First we prove that if A has an inverse, then A is nonsingular. Assume A has an inverse, that is, A^{-1} exists. Thus, $A^{-1}A = I$. Lemma 1.9.5 now implies that A is nonsingular.

($\Leftarrow$). Now we prove that if A is nonsingular, then A has an inverse. Assume A is nonsingular. We prove that A has an inverse, that is, we must show that there is a matrix B such that $AB = BA = I$. Since A is nonsingular, we know by Lemma 1.9.4 that there is a matrix B such that $AB = I$. Now, Theorem 1.9.6 implies that $BA = I$. Thus, the matrix given by Lemma 1.9.4 is the inverse of A, that is, $B = A^{-1}$. Therefore, A has an inverse. □

1.9.4 Calculating the Inverse

The above results give us a method for finding the inverse of a nonsingular square matrix A. This method is based on the proof of Lemma 1.9.4. In other words to find A^{-1}, we simultaneously solve the systems

$$A\mathbf{x} = \mathbf{e}_1,\ \ A\mathbf{x} = \mathbf{e}_2,\ \ \ldots,\ \ A\mathbf{x} = \mathbf{e}_n$$

by forming the augmented matrix $[A\,|\,\mathbf{e}_1\ \ \mathbf{e}_2\ \ \cdots\ \ \mathbf{e}_n] = [A\,|\,I]$. Put this matrix into reduced echelon form, obtaining $[I\,|\,\mathbf{B}_1\ \ \mathbf{B}_2\ \ \cdots\ \ \mathbf{B}_n$. Therefore,

$$A\mathbf{B}_1 = \mathbf{e}_1,\ \ A\mathbf{B}_2 = \mathbf{e}_2,\ \ \ldots,\ \ A\mathbf{B}_n = \mathbf{e}_n \tag{1.20}$$

Let B be the $n \times n$ matrix $B = [\mathbf{B}_1\ \mathbf{B}_2\ \cdots\ \mathbf{B}_n]$. Thus, we can show that $AB = I$ as follows:

$$\begin{aligned} AB &= A[\mathbf{B}_1\ \mathbf{B}_2\ \cdots\ \mathbf{B}_n] & \\ &= [A\mathbf{B}_1\ A\mathbf{B}_2\ \cdots\ A\mathbf{B}_n] & \text{by Theorem 1.5.13} \\ &= [\mathbf{e}_1\ \mathbf{e}_2\ \cdots\ \mathbf{e}_n] & \text{by equations (1.20) above} \\ &= I & \text{since } I = [\mathbf{e}_1\mathbf{e}_2\cdots\mathbf{e}_n]. \end{aligned}$$

Therefore, $AB = I$ and thus, $A^{-1} = [\mathbf{B}_1\ \ \mathbf{B}_2\ \ \cdots\ \ \mathbf{B}_n]$.

Procedure for Finding a Inverse. To determine if the square matrix A has an inverse, do the following:

Step 1. Form the matrix $[A\,|\,I]$, where the identity matrix I has the same size as A.

Step 2. Transform the matrix $[A\,|\,I]$ into *reduced echelon form* $[M\,|\,B]$.

Step 3. Either $M = I$ or $M \neq I$.

- If $M = I$, then $A^{-1} = B$.
- If $M \neq I$, then A^{-1} does not exist.

In the next problem, we will apply the above procedure to determine whether or not the matrix

$$\begin{bmatrix} 3 & 4 & -1 \\ 1 & 0 & 3 \\ 2 & 5 & -4 \end{bmatrix}$$

has an inverse.

Problem 3. Consider the matrix $A = \begin{bmatrix} 3 & 4 & -1 \\ 1 & 0 & 3 \\ 2 & 5 & -4 \end{bmatrix}$. Find its inverse (if it exists).

Solution. Applying Step 1 we obtain the matrix $[A\,|\,I] = \begin{bmatrix} 3 & 4 & -1 & 1 & 0 & 0 \\ 1 & 0 & 3 & 0 & 1 & 0 \\ 2 & 5 & -4 & 0 & 0 & 1 \end{bmatrix}$. We now apply Step 2 and transform $[A\,|\,I]$ into reduced echelon form as follows:

$$[A\,|\,I] = \begin{bmatrix} 3 & 4 & -1 & 1 & 0 & 0 \\ 1 & 0 & 3 & 0 & 1 & 0 \\ 2 & 5 & -4 & 0 & 0 & 1 \end{bmatrix}$$

$$R_1 \leftrightarrow R_2$$

$$\begin{bmatrix} 1 & 0 & 3 & 0 & 1 & 0 \\ 3 & 4 & -1 & 1 & 0 & 0 \\ 2 & 5 & -4 & 0 & 0 & 1 \end{bmatrix}$$

$$-3R_1 + R_2 \to R_2$$

$$\begin{bmatrix} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 4 & -10 & 1 & -3 & 0 \\ 2 & 5 & -4 & 0 & 0 & 1 \end{bmatrix}$$

$$-2R_1 + R_3 \to R_3$$

$$\begin{bmatrix} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 4 & -10 & 1 & -3 & 0 \\ 0 & 5 & -10 & 0 & -2 & 1 \end{bmatrix}$$

$$R_2 \leftrightarrow R_3$$

$$\begin{bmatrix} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 5 & -10 & 0 & -2 & 1 \\ 0 & 4 & -10 & 1 & -3 & 0 \end{bmatrix}$$

$$-R_3 + R_2 \to R_2$$

$$\begin{bmatrix} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 & 1 & 1 \\ 0 & 4 & -10 & 1 & -3 & 0 \end{bmatrix}$$

$$-4R_2 + R_3 \to R_3$$

$$\begin{bmatrix} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 & 1 & 1 \\ 0 & 0 & -10 & 5 & -7 & -4 \end{bmatrix}$$

$$-\frac{1}{10}R_3 \to R_3$$

$$\begin{bmatrix} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 & 1 & 1 \\ 0 & 0 & 1 & -\frac{1}{2} & \frac{7}{10} & \frac{2}{5} \end{bmatrix}$$

$$-3R_3 + R_1 \to R_1$$

$$[M\,|\,B] = \begin{bmatrix} 1 & 0 & 0 & \frac{3}{2} & -\frac{11}{10} & -\frac{6}{5} \\ 0 & 1 & 0 & -1 & 1 & 1 \\ 0 & 0 & 1 & -\frac{1}{2} & \frac{7}{10} & \frac{2}{5} \end{bmatrix}$$

Since $M = I$, we conclude from Step 3 that

$$A^{-1} = B = \begin{bmatrix} \frac{3}{2} & -\frac{11}{10} & -\frac{6}{5} \\ -1 & 1 & 1 \\ -\frac{1}{2} & \frac{7}{10} & \frac{2}{5} \end{bmatrix}.$$

1.9.5 Inverse Properties

Theorem 1.9.8. Let A and B be square $n \times n$ matrices. Then:

1. If A has an inverse, then A^{-1} has an inverse and $(A^{-1})^{-1} = A$.
2. If A and B have inverses, then AB has an inverse and $(AB)^{-1} = B^{-1}A^{-1}$.
3. If A has an inverse and $k \in \mathbb{R}$ is non-zero, then kA has an inverse and $(kA)^{-1} = \frac{1}{k}A^{-1}$.
4. If A is has an inverse, then A^T is has an inverse and $(A^T)^{-1} = (A^{-1})^T$.

Proof. We just prove 1, 2 and 4. To prove 1, we have that $AA^{-1} = I$ and $A^{-1}A = I$. Thus, $(A^{-1})^{-1} = A$. To prove 2, by the properties of matrix operations, we obtain

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I.$$

Similarly, one can show that $(B^{-1}A^{-1})(AB) = I$. Therefore, $(AB)^{-1} = B^{-1}A^{-1}$. To prove item 4, since $AA^{-1} = I$ we conclude that $(AA^{-1})^T = I^T$. Since $I^T = I$, we see that $(AA^{-1})^T = I$. From Theorem 1.6.6(2), we deduce that $(A^{-1})^T A^T = I$. Similarly, because $A^{-1}A = I$ it follows that $A^T(A^{-1})^T = I$. Therefore, $(A^T)^{-1} = (A^{-1})^T$. □

Corollary 1.9.9. If $A_1, A_2, \ldots, A_r$ are invertible matrices having the same same size, then the matrix $A_1A_2 \cdots A_r$ is also invertible, and $(A_1A_2 \cdots A_r)^{-1} = A_r^{-1} \cdots A_2^{-1}A_1^{-1}$.

Proof. This can proved by induction on r, using Theorem 1.9.8(2). □

If A is an invertible matrix and $-p$ is a negative integer (p is a natural number), then we define

$$A^{-p} = (A^{-1})^p = \underbrace{A^{-1}A^{-1} \cdots A^{-1}}_{p\,\text{many times}}$$

We end our discussion of Chapter 1 with the statement of the following **very important theorem**. This theorem will be used extensively throughout the semester.

Theorem 1.9.10. If A is an $n \times n$ (square) matrix, then the following statements are equivalent:

1. A is nonsingular, that is, $A\mathbf{x} = \boldsymbol{\theta}$ has only the trivial solution (here, $\boldsymbol{\theta} = [0]_{n\times 1}$).
2. The column vectors of A are linearly independent.
3. For every $n \times 1$ matrix $\mathbf{b}$, the system $A\mathbf{x} = \mathbf{b}$ has a (unique) solution.
4. A has an inverse.
5. A is row equivalent to I_n.

Proof. The above theorem follows directly from Theorems 1.7.11, 1.7.14, 1.9.7 and 1.7.13. □

Exercises 1.9

Pages 102 to 103 of text – #1, 3, 6, 7, 19, 23, 25, 27, 29, 33, 35.

Chapter 3

The Vector Space $\mathbb{R}^n$

Vector spaces are the main topic of interest in linear algebra. A vector space is a mathematical structure formed by a collection of vectors that can be added together and multiplied by real numbers. In this chapter we will investigate vector spaces and the concept of dimension, which specifies the number of independent directions in the vector space.

3.1 Introduction

Recall that $\mathbb{R}^n$ is the space of all n**-vectors** $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ where each **component** x_i is a real number. Also recall the following vector operations:

- For vectors $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ in $\mathbb{R}^n$, we have that $\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}$.

- For vectors $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ in $\mathbb{R}^n$, we have that $\mathbf{x} - \mathbf{y} = \begin{bmatrix} x_1 - y_1 \\ x_2 - y_2 \\ \vdots \\ x_n - y_n \end{bmatrix}$.

- For a vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ and a scalar c, we have that $c\mathbf{x} = \begin{bmatrix} cx_1 \\ cx_2 \\ \vdots \\ cx_n \end{bmatrix}$.

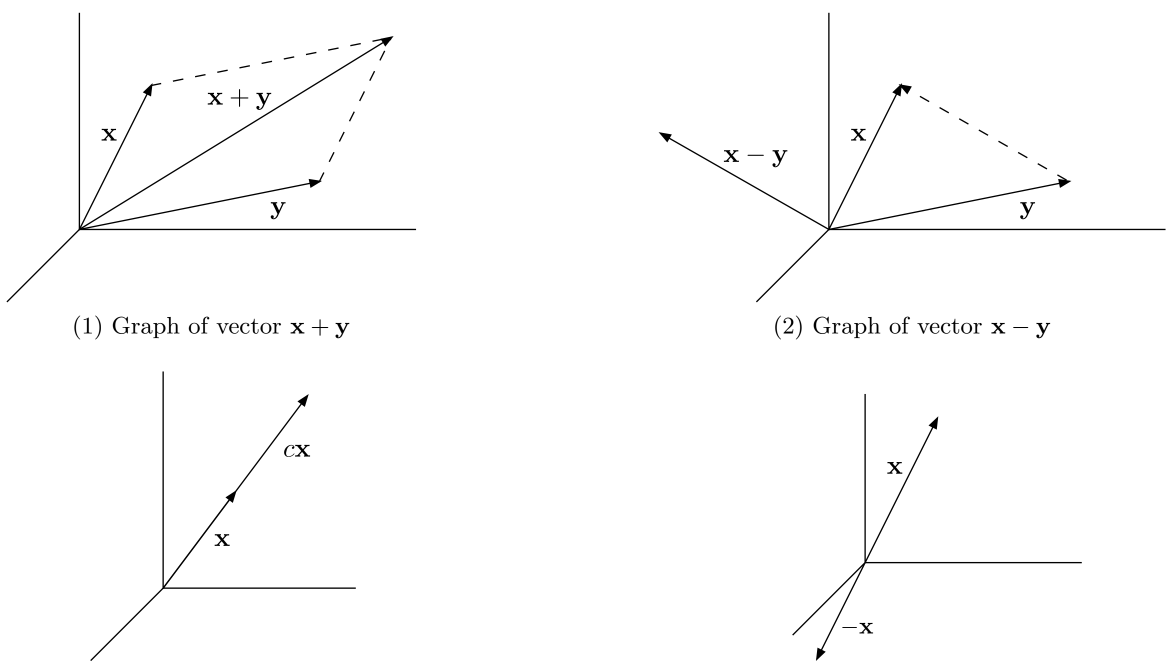
- We write $\boldsymbol{\theta} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ for the "zero" vector, and define $-\mathbf{x} = \begin{bmatrix} -x_1 \\ -x_2 \\ \vdots \\ -x_n \end{bmatrix}$.

3.1.1 Vectors in 3-space

Recall that $\mathbb{R}^3$ is the space of all **3-vectors** $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ where each **component** x_i is a real number. Also recall the following vector operations:

- For vectors $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$ in $\mathbb{R}^3$ is defined by $\mathbf{x}+\mathbf{y} = \begin{bmatrix} x_1+y_1 \\ x_2+y_2 \\ x_3+y_3 \end{bmatrix}$.
- For vectors $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$ in $\mathbb{R}^3$ is defined by $\mathbf{x}-\mathbf{y} = \begin{bmatrix} x_1-y_1 \\ x_2-y_2 \\ x_3-y_3 \end{bmatrix}$.
- For a vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ by a scalar c is defined by $c\mathbf{x} = \begin{bmatrix} cx_1 \\ cx_2 \\ cx_3 \end{bmatrix}$.
- We write $\boldsymbol{\theta} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ for the "zero" vector, and define $-\mathbf{x} = \begin{bmatrix} -x_1 \\ -x_2 \\ -x_3 \end{bmatrix}$.

The graphs in Figure 3.1 illustrate how vectors in $\mathbb{R}^3$ are (1) added, (2) subtracted, (3) scalar multiplied, and (4) made "negative."



(1) Graph of vector $\mathbf{x}+\mathbf{y}$ (2) Graph of vector $\mathbf{x}-\mathbf{y}$

(3) Graph of vector $c\mathbf{x}$ (4) Graph of vector $-\mathbf{x}$

Figure 3.1: Vector Operations

Example 1. Consider the vector space $\mathbb{R}^2$. In Figure 3.2 we give a geometric interpretation of the subset W of $\mathbb{R}^2$ defined by

$$W = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} : x_1 + x_2 = 0 \right\}.$$

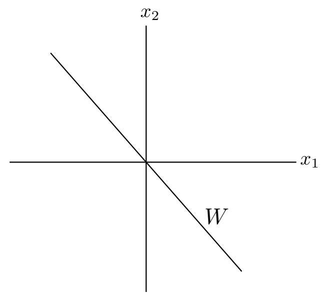


Figure 3.2: In Example 1, W is the above line in 2-Space

Example 2. Consider the vector space $\mathbb{R}^2$. In Figure 3.3 we give a geometric interpretation of the subset W of $\mathbb{R}^2$ defined by

$$W = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} : x_1 + x_2 = 1 \right\}.$$

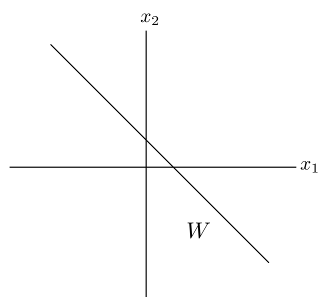


Figure 3.3: In Example 2, W is the above line in 2-Space

Example 3. Consider the vector space $\mathbb{R}^3$. In Figure 3.4 we give a geometric interpretation of the subset W of $\mathbb{R}^3$ defined by

$$W = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} : x_1 + \frac{1}{4}x_2 + x_3 = 0 \right\}.$$

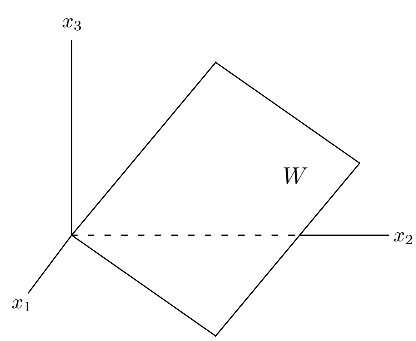


Figure 3.4: In Example 3, W is the above plane in 3-Space

Example 4. Consider the vector space $\mathbb{R}^3$. In Figure 3.5 we give a geometric interpretation of the subset W of $\mathbb{R}^3$ defined by

$$W = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} : x_1 + \frac{1}{4}x_2 + x_3 = 1 \right\}.$$

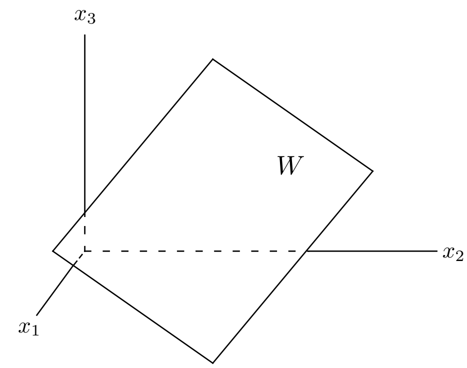


Figure 3.5: In Example 4, W is the above plane in 3-Space

3.2 Vector Space Properties

The following algebraic properties of vector addition can easily be checked.

Theorem 3.2.1. If $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ are vectors in $\mathbb{R}^n$ and if a and b are scalars, then the following properties hold:

Closure properties:

1. $\mathbf{x} + \mathbf{y}$ is a vector in $\mathbb{R}^n$, (that is, $\mathbb{R}^n$ is *closed under addition*).
2. $a\mathbf{x}$ is a vector in $\mathbb{R}^n$, (that is, $\mathbb{R}^n$ is *closed under scalar multiplication*).

Addition properties:

1. $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$
2. $\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$
3. $\mathbf{x} + \boldsymbol{\theta} = \boldsymbol{\theta} + \mathbf{x} = \mathbf{x}$
4. $\mathbf{x} + (-\mathbf{x}) = \boldsymbol{\theta}$

Scalar multiplication properties:

1. $a(b\mathbf{x}) = (ab)\mathbf{x}$
2. $a(\mathbf{x} + \mathbf{y}) = a\mathbf{x} + a\mathbf{y}$
3. $(a + b)\mathbf{x} = a\mathbf{x} + b\mathbf{x}$
4. $1\mathbf{x} = \mathbf{x}$

3.2.1 Subspaces

Definition 3.2.2. A subset W of a vector space $\mathbb{R}^n$ is called a **subspace** of $\mathbb{R}^n$, if $\boldsymbol{\theta}$ is in W and W satisfies all the properties in Theorem 3.2.1 where $\mathbb{R}^n$ is replaced by W.

Theorem 3.2.3. Let W be a subset of $\mathbb{R}^n$. Then W is a subspace of $\mathbb{R}^n$ if and only if the following conditions hold:

(s1) $\boldsymbol{\theta}$ is in W.

(s2) If $\mathbf{x}$ and $\mathbf{y}$ are vectors in W, then $\mathbf{x}+\mathbf{y}$ is in W.

(s3) If c is any scalar and $\mathbf{x}$ is a vector in W, then $c\mathbf{x}$ is in W.

Proof. We first prove that if W is a subspace, the statements (s1)-(s3) hold. Then we will prove the converse.

($\Rightarrow$). Suppose that W is a subspace of $\mathbb{R}^n$. We now show that (s1), (s2) and (s3) hold. Since $\boldsymbol{\theta} \in W$, (s1) holds. To prove (s2) and (s3), let $\mathbf{x}$ and $\mathbf{y}$ be vectors in W and let c be a scalar. Since W is a subspace, it satisfies the closure properties. Therefore, $\mathbf{x}+\mathbf{y}$ is in W and $c\mathbf{x}$ is in W.

($\Leftarrow$). Suppose that W satisfies (s1), (s2) and (s3) above. We must show that W is a subspace of $\mathbb{R}^n$. To see that W is a subspace of $\mathbb{R}^n$, one has to first show that W satisfies the addition properties and the scalar multiplication properties. Since $\mathbb{R}^n$ satisfies these properties, it follows that W also satisfies them. So the only properties that needs to be verified are the closure properties. But these properties hold by our assumption. Hence, W is a subspace of $\mathbb{R}^n$. □

Definition 3.2.4. Let W be a subset of $\mathbb{R}^n$.

1. If for all $\mathbf{x}, \mathbf{y} \in W$ we have that $\mathbf{x}+\mathbf{y} \in W$, then we say that W is *closed under addition.*
2. If for all $\mathbf{x} \in W$ and all $c \in \mathbb{R}$ we have that $c\mathbf{x} \in W$, then W is said to be *closed under scalar multiplication.*

3.2.2 Verifying That Subsets Are Subspaces

To verify that a subset W is subspace of $\mathbb{R}^n$, do the following:

Step 1. An algebraic specification for the subset W is given, and this specification serves as a test for determining whether a vector (from $\mathbb{R}^n$) is or is not in W.

Step 2. Verify that $\boldsymbol{\theta}$ satisfies the algebraic specification of W.

Step 3. Choose two arbitrary vectors $\mathbf{x}$ and $\mathbf{y}$ in W. Thus $\mathbf{x}$ and $\mathbf{y}$ are in $\mathbb{R}^n$, and both vectors $\mathbf{x}$ and $\mathbf{y}$ satisfy the algebraic specification of W.

Step 4. Verify that the sum vector $\mathbf{x}+\mathbf{y}$ meets the specification of W.

Step 5. For an arbitrary scalar c verify that the scalar multiple vector $c\mathbf{x}$ meets the specification of W.

Problem 1. Let W be the subset of $\mathbb{R}^3$ defined by

$$W = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} : x_2 = 2x_1 \text{ and } x_3 = 3x_1 \right\}.$$

Verify that W is a subspace of $\mathbb{R}^3$ and give a geometric interpretation of W.

Solution. For clarity, we explicitly number the five steps used to verify that W is a subspace of $\mathbb{R}^3$.

1. The algebraic specification for $\mathbf{x}$ to be in W is

$$x_2 = 2x_1 \quad \text{and} \quad x_3 = 3x_1 \tag{3.1}$$

2. Verify that $\boldsymbol{\theta} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ satisfies the above algebraic specification. This is clear.

3. Let $\mathbf{f}$ and $\mathbf{g}$ be two arbitrary vectors in W:

$$\mathbf{f} = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} \quad \text{and} \quad \mathbf{g} = \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix}.$$

   Because $\mathbf{f}$ and $\mathbf{g}$ are in W, each must satisfy the algebraic specification of W. That is,

$$f_2 = 2f_1 \quad \text{and} \quad f_3 = 3f_1 \tag{3.2}$$

$$g_2 = 2g_1 \quad \text{and} \quad g_3 = 3g_1. \tag{3.3}$$

4. Next, verify that the sum $\mathbf{f}+\mathbf{g}$ is in W; that is, verify equation (3.1). Now, the sum $\mathbf{f}+\mathbf{g}$ is given by

$$\mathbf{f}+\mathbf{g} = \begin{bmatrix} f_1 + g_1 \\ f_2 + g_2 \\ f_3 + g_3 \end{bmatrix}.$$

   By (3.2) and (3.3), we have

$$(f_2 + g_2) = 2(f_1 + g_1) \quad \text{and} \quad (f_3 + g_3) = 3(f_1 + g_1).$$

   Thus, $\mathbf{f}+\mathbf{g}$ is in W whenever $\mathbf{f}$ and $\mathbf{g}$ are in W (see equation (3.1)).

5. Let $\mathbf{f} \in W$ and $c \in \mathbb{R}$, where $\mathbf{f} = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix}$. As $\mathbf{f} \in W$, we have that $f_2 = 2f_1$ and $f_3 = 3f_1$. Thus,

$$cf_2 = c(2f_1) = 2(cf_1) \quad \text{and} \quad cf_3 = c(3f_1) = 3(cf_1).$$

   Therefore, the vector

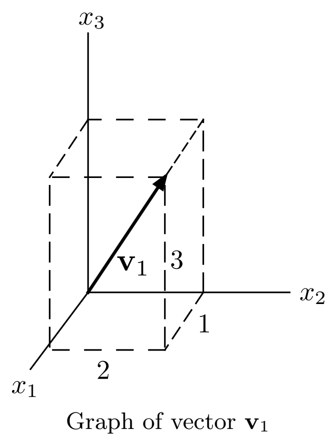
$$c\mathbf{f} = \begin{bmatrix} cf_1 \\ cf_2 \\ cf_3 \end{bmatrix}.$$

   is in W, whenever $\mathbf{f}$ is in W.

We conclude that W is a subspace of $\mathbb{R}^3$. Note that every vector $\mathbf{x}$ in W can be expressed as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ 2x_1 \\ 3x_1 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

So, W has the vector $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ as a "basis" and so, W has dimension one. Geometrically, W is an infinite line through the origin. The graph of $\mathbf{v}_1$ and W are given in Figure 3.6.



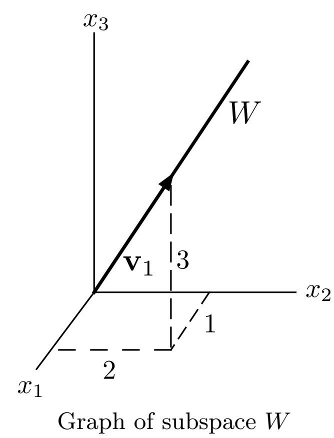


Figure 3.6: The Subspace W in Problem 1

Problem 2. Consider the vector space $\mathbb{R}^3$. Let $W = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} : x_1 = x_2 + x_3 \right\}$. Show that W is a subspace of $\mathbb{R}^3$.

Problem 3. Consider the vector space $\mathbb{R}^3$. Let $W = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} : x_1 = x_2 \cdot x_3 \right\}$. Show that W is not a subspace of $\mathbb{R}^3$.

Exercises 3.2

Pages 174 to 175 of text – #1, 2, 3, 9, 13, 15, 16, 18, 31.

3.3 Examples of Subspaces

3.3.1 The Span of a set of vectors

Theorem 3.3.1. Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ be vectors in a vector space $\mathbb{R}^n$. Let W be the set consisting of all linear combinations of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$; that is, W is the set of all vectors of the form

$$\mathbf{x} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_r\mathbf{v}_r$$

where $a_1, a_2, \ldots, a_r$ are real numbers. Then W is a subspace of $\mathbb{R}^n$.

Proof. Note that $\boldsymbol{\theta} \in W$ since

$$\boldsymbol{\theta} = 0\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_r.$$

So, to show that W is a subspace it is sufficient to show that W is closed under addition and scalar multiplication. To do this let $\mathbf{x}$ and $\mathbf{y}$ be in W and let c be a scalar. Since $\mathbf{x}$ and $\mathbf{y}$ are linear combinations of the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ there are scalars $a_1, a_2, \ldots, a_r$ and $b_1, b_2, \ldots, b_r$ such that

$$\mathbf{x} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + ar\mathbf{v}_r$$

and

$$\mathbf{y} = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_r\mathbf{v}_r.$$

But then,

$$\mathbf{x} + \mathbf{y} = (a_1 + b_1)\mathbf{v}_1 + (a_2 + b_2)\mathbf{v}_2 + \cdots + (a_r + b_r)\mathbf{v}_r$$

and

$$c\mathbf{x} = (ca_1)\mathbf{v}_1 + (ca_2)\mathbf{v}_2 + \cdots + (ca_r)\mathbf{v}_r$$

are also linear combinations of the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$. Hence, $\mathbf{x}+\mathbf{y} \in W$ and $c\mathbf{x} \in W$. Therefore, by Theorem 3.2.3, W is a subspace of $\mathbb{R}^n$. □

Definition 3.3.2. The subspace W spanned by the set of vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ will be denoted by Span(S) or Span$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$. [The book uses the notation Sp(S).]

Problem 1. Consider the three vectors in 4-space:

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ -1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 2 \\ -3 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ -2 \end{bmatrix}.$$

Is the vector $\mathbf{y} = \begin{bmatrix} 2 \\ 1 \\ 5 \\ -5 \end{bmatrix}$ in the Span$\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, that is, is $\mathbf{y}$ a linear combination of the vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$?

Solution. The answer is "yes" if there are numbers x_1, x_2, x_3 satisfying the vector equation

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 = \mathbf{y}. \tag{3.4}$$

To answer the question, first let the vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ form the columns of a matrix A and let $\mathbf{x}$ be the column vector consisting of the unknowns x_1, x_2, x_3, that is,

$$A = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & 1 \\ 1 & 2 & 0 \\ -1 & -3 & -2 \end{bmatrix} \text{ and } \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Now, notice that the following 5 equations are equivalent:

1. $x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 = \mathbf{y}$
2. $x_1 \begin{bmatrix} 1 \\ 2 \\ 1 \\ -1 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 0 \\ 2 \\ -3 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 1 \\ 0 \\ -2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 5 \\ -5 \end{bmatrix}$
3. $\begin{bmatrix} 1x_1 & + & 1x_2 & + & 1x_3 \\ 2x_1 & + & 0x_2 & + & 1x_3 \\ 1x_1 & + & 2x_2 & + & 0x_3 \\ -1x_1 & - & 3x_2 & - & 2x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 5 \\ -5 \end{bmatrix}$

4. $\begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & 1 \\ 1 & 2 & 0 \\ -1 & -3 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 5 \\ -5 \end{bmatrix}$

5. $A\mathbf{x} = \mathbf{y}$

Therefore, x_1, x_2, x_3 is a solution to vector equation (3.4) if and only if $\mathbf{x}$ is a solution to the linear system $A\mathbf{x} = \mathbf{y}$. Transforming the augmented matrix $[A\,|\,\mathbf{y}]$ into reduced row echelon form we obtain

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

Therefore, the solution $x_1 = 1$, $x_2 = 2$, and $x_3 = -1$ satisfies vector equation (3.4), that is,

$$\mathbf{y} = 1\mathbf{v}_1 + 2\mathbf{v}_2 + -1\mathbf{v}_3 \text{ (verify!)}.$$

Hence, the vector $\mathbf{y}$ can be written as a linear combination of the vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$; Thus, $\mathbf{y}$ is in the Span$\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$.

From the above solution for Problem 1, we can make the following observation:

Theorem 3.3.3. Let $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k$ be a set of vectors in $\mathbb{R}^n$ and let A be the matrix $A = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_k \end{bmatrix}$. A vector $\mathbf{y}$ in $\mathbb{R}^n$ is in the subspace $W = \text{Span}\{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k\}$ if and only if the linear system $A\mathbf{x} = \mathbf{y}$ has a solution $\mathbf{x}$.

Problem 2 (Example 1 on page 178 text)**.** Let $\mathbf{u}$ and $\mathbf{v}$ be the following vectors in $\mathbb{R}^3$:

$$\mathbf{u} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} \text{ and } \mathbf{v} = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}.$$

1. Obtain a algebraic specification for the subspace Span$\{\mathbf{u}, \mathbf{v}\}$ of $\mathbb{R}^3$.
2. Using the vectors $\mathbf{u}$ and $\mathbf{v}$, give a geometric interpretation of Span$\{\mathbf{u}, \mathbf{v}\}$.

Solution. We first obtain a algebraic specification for the subspace Span$\{\mathbf{u}, \mathbf{v}\}$. A vector $\mathbf{y} \in \mathbb{R}^3$ is in Span$\{\mathbf{u}, \mathbf{v}\}$ if and only if $A\mathbf{x} = \mathbf{y}$ has a solution $\mathbf{x}$, where $A = \begin{bmatrix} \mathbf{u} & \mathbf{v} \end{bmatrix}$. We investigate the conditions on $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$ which will imply that the system $A\mathbf{x} = \mathbf{y}$ is consistent. That is, we solve the system $A\mathbf{x} = \mathbf{y}$ by forming the augmented matrix $[A\,|\,\mathbf{y}]$ and put it into reduced echelon form. We shall do this as follows:

$$[A\,|\,\mathbf{y}] = \begin{bmatrix} 2 & 0 & y_1 \\ 1 & 1 & y_2 \\ 0 & 2 & y_3 \end{bmatrix}.$$

Now, using row operations, we transform the matrix $[A\,|\,\mathbf{y}]$ into the ‘reduced echelon form as follows:

$$[A\,|\,\mathbf{b}] = \begin{bmatrix} 2 & 0 & y_1 \\ 1 & 1 & y_2 \\ 0 & 2 & y_3 \end{bmatrix}$$

$$\frac{1}{2}R_1 \to R_1$$

$$\begin{bmatrix} 1 & 0 & \frac{y_1}{2} \\ 1 & 1 & y_2 \\ 0 & 2 & y_3 \end{bmatrix}$$

$$-R_1 + R_2 \to R_2$$

$$\begin{bmatrix} 1 & 1 & y_2 \\ 0 & 1 & y_2 - \frac{y_1}{2} \\ 0 & 2 & y_3 \end{bmatrix}$$

$$(-R_2 + R_1 \to R_1)\&(-2R_2 + R_3 \to R_3)$$

$$\begin{bmatrix} 1 & 0 & \frac{y_1}{2} \\ 0 & 1 & y_2 - \frac{y_1}{2} \\ 0 & 0 & y_1 - 2y_2 + y_3 \end{bmatrix}$$

So the system $A\mathbf{x} = \mathbf{y}$ will have a solution if and only if $y_1 - 2y_2 + y_3 = 0$; that is, when $y_1 = 2y_2 - y_3$. Hence,

$$\text{Span}\{\mathbf{u}, \mathbf{v}\} = \left\{ \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} : y_1 = 2y_2 - y_3 \right\}.$$

We give a geometric give a geometric interpretation of Span$\{\mathbf{u}, \mathbf{v}\}$ in Figure 3.7

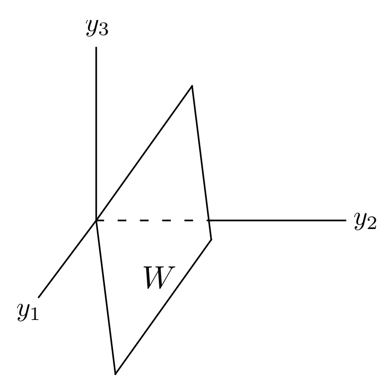


Figure 3.7: $W = \text{Span}\{\mathbf{u}, \mathbf{v}\}$

3.3.2 The Null Space of a Matrix

Definition 3.3.4. Consider the vector space $\mathbb{R}^n$. Let A be an $m \times n$ matrix. Let $\mathcal{N}(A)$ be set of all vectors $\mathbf{x}$ in $\mathbb{R}^n$ which are solutions to the homogeneous system of equations $A\mathbf{x} = \boldsymbol{\theta}$, that is,

$$\mathcal{N}(A) = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \boldsymbol{\theta}\}.$$

The set $\mathcal{N}(A)$ is called the **null space** of A or the **kernel** or A.

Theorem 3.3.5. Consider the vector space $\mathbb{R}^n$. Let A be an $m \times n$ matrix. Then $\mathcal{N}(A)$ is a subspace of $\mathbb{R}^n$.

Proof. Note that $\boldsymbol{\theta} \in \mathcal{N}(A)$ since

$$A\boldsymbol{\theta} = \boldsymbol{\theta}.$$

So, to show that $\mathcal{N}(A)$ is a subspace it is sufficient to show that $\mathcal{N}(A)$ is closed under addition and scalar multiplication. To do this let $\mathbf{x}$ and $\mathbf{y}$ be in $\mathcal{N}(A)$ and let c be a scalar. Since $\mathbf{x}$ and $\mathbf{y}$ are in $\mathcal{N}(A)$, it follows that

$$A\mathbf{x} = \boldsymbol{\theta} \text{ and } A\mathbf{y} = \boldsymbol{\theta}. \tag{3.5}$$

We must show that $\mathbf{x}+\mathbf{y}$ is in $\mathcal{N}(A)$; that is, we must show that $A(\mathbf{x}+\mathbf{y}) = \boldsymbol{\theta}$. To see this, note that

$$\begin{aligned} A(\mathbf{x}+\mathbf{y}) &= A\mathbf{x} + A\mathbf{y} && \text{by distribution of matrix mult.} \\ &= \boldsymbol{\theta} + \boldsymbol{\theta} && \text{by equations in (3.5) above} \\ &= \boldsymbol{\theta} && \text{since } \boldsymbol{\theta} + \boldsymbol{\theta} = \boldsymbol{\theta}. \end{aligned}$$

Hence, $\mathbf{x}+\mathbf{y} \in \mathcal{N}(A)$. Now we must show that $c\mathbf{x}$ is in $\mathcal{N}(A)$; that is, we must show that $A(c\mathbf{x}) = \boldsymbol{\theta}$ given that $\mathbf{x}$ is in $\mathcal{N}(A)$. To see this, note that

$$\begin{aligned} A(c\mathbf{x}) &= cA\mathbf{x} && \text{by property of matrix mult.} \\ &= c\boldsymbol{\theta} && \text{by first equation in (3.5) above} \\ &= \boldsymbol{\theta} && \text{since } c\boldsymbol{\theta} = \boldsymbol{\theta}. \end{aligned}$$

Hence, $c\mathbf{x} \in \mathcal{N}(A)$. Therefore, by Theorem 3.2.3, $\mathcal{N}(A)$ is a subspace of $\mathbb{R}^n$. □

The next problem will give a method for finding a set S of vectors in $\mathcal{N}(A)$ which span $\mathcal{N}(A)$, that is, $\text{Span}(S) = \mathcal{N}(A)$.

Problem 3. Find a set of vectors in the null space of $A = \begin{bmatrix} 1 & 1 & 2 & 9 & 18 \\ 2 & 9 & -10 & 39 & 78 \\ 1 & -6 & 1 & -27 & -54 \end{bmatrix}$ which **spans** this null space.

Solution. First notice, since A is a 3×5 matrix, that the null space of A is a subspace of $\mathbb{R}^5$. Applying Gauss-Jordon reduction to the augmented matrix $[A \,|\, \boldsymbol{\theta}]$, we obtain the augmented matrix

$$[C \,|\, \boldsymbol{\theta}] = \begin{bmatrix} 1 & 0 & 0 & 2 & 4 & 0 \\ 0 & 1 & 0 & 5 & 10 & 0 \\ 0 & 0 & 1 & 1 & 2 & 0 \end{bmatrix}.$$

Solving the equivalent system of equations $C\mathbf{x} = \boldsymbol{\theta}$, we obtain

$$\begin{aligned} x_1 &= -2x_4 - 4x_5 \\ x_2 &= -5x_4 - 10x_5 \\ x_3 &= -x_4 - 2x_5 \\ x_4 &= x_4 \\ x_5 &= x_5. \end{aligned}$$

Therefore, every solution to $A\mathbf{x} = \boldsymbol{\theta}$ can be written in the form

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = x_4 \begin{bmatrix} -2 \\ -5 \\ -1 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -4 \\ -10 \\ -2 \\ 0 \\ 1 \end{bmatrix}.$$

So the vectors $\mathbf{y}_1 = \begin{bmatrix} -2 \\ -5 \\ -1 \\ 1 \\ 0 \end{bmatrix}$ and $\mathbf{y}_2 = \begin{bmatrix} -4 \\ -10 \\ -2 \\ 0 \\ 1 \end{bmatrix}$ span the null space of A. We conclude that $\text{Span}\{\mathbf{y}_1, \mathbf{y}_2\} = \mathcal{N}(A)$.

Observation. The above vectors $\mathbf{y}_1$ and $\mathbf{y}_2$ are linearly independent (see Theorem 1.7.7).

3.3.3 The Range of a Matrix

Definition 3.3.6. Consider the vector space $\mathbb{R}^m$. Let A be an $m \times n$ matrix. Let $\mathcal{R}(A)$ be set of all vectors $\mathbf{y}$ in $\mathbb{R}^m$ such that $A\mathbf{x} = \mathbf{y}$ for some $\mathbf{x} \in \mathbb{R}^n$, that is,

$$\mathcal{R}(A) = \{\mathbf{y} \in \mathbb{R}^m : A\mathbf{x} = \mathbf{y} \text{ for some } \mathbf{x} \in \mathbb{R}^n\}.$$

Theorem 3.3.7. Consider the vector space $\mathbb{R}^m$. Let A be an $m \times n$ matrix. Then $\mathcal{R}(A)$ is a subspace of $\mathbb{R}^m$.

Proof. Recall that there is another way to interpret the product $A\mathbf{x}$. The product $A\mathbf{x}$ is a linear combination of the column vectors of the matrix $A = [\mathbf{A}_1 \ \ \mathbf{A}_2 \ \ \cdots \ \ \mathbf{A}_n]$ (where each $\mathbf{A}_i \in \mathbb{R}^m$); that is,

$$A \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{A}_1 + x_2\mathbf{A}_2 + \cdots + x_n\mathbf{A}_n.$$

It follows that the matrix equation $A\mathbf{x} = \mathbf{y}$ is equivalent to the vector equation

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

Therefore, $\mathcal{R}(A) = \text{Span}\{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n\}$. Hence, $\mathcal{R}(A)$ is a subspace of $\mathbb{R}^m$ by Theorem 3.3.1. □

Given a matrix A, we shall refer to $\mathcal{R}(A)$ as the *range space of* A.

Problem 4 (Example 4 on page 182 of text)**.** Describe, in terms of an algebraic specification, the range space of the 3×4 matrix

$$A = \begin{bmatrix} 1 & 1 & 3 & 1 \\ 2 & 1 & 5 & 4 \\ 1 & 2 & 4 & -1 \end{bmatrix}.$$

Solution. A vector $\mathbf{y} \in \mathbb{R}^3$ is in the range of A if and only if $A\mathbf{x} = \mathbf{y}$ has a solution $\mathbf{x}$. We investigate the conditions on $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$ which will imply that the system $A\mathbf{x} = \mathbf{y}$ is consistent. That is, we solve the system $A\mathbf{x} = \mathbf{y}$ by forming the augmented matrix $[A \mid \mathbf{y}]$ and put it into reduced echelon form. Now, we transform the matrix $[A \mid \mathbf{y}]$ into the reduced echelon form as follows:

$$[A \mid \mathbf{y}] = \begin{bmatrix} 1 & 1 & 3 & 1 & y_1 \\ 2 & 1 & 5 & 4 & y_2 \\ 1 & 2 & 4 & -1 & y_3 \end{bmatrix}$$

$$(-2R_1 + R_2 \to R_2)\&(-R_1 + R_3 \to R_3)$$

$$\begin{bmatrix} 1 & 1 & 3 & 1 & y_1 \\ 0 & -1 & -1 & 2 & y_2 - 2y_1 \\ 0 & 1 & 1 & -2 & y_3 - y_1 \end{bmatrix}$$

$$-R_2 \to R_2$$

$$\begin{bmatrix} 1 & 1 & 3 & 1 & y_1 \\ 0 & 1 & 1 & -2 & 2y_1 - y_2 \\ 0 & 1 & 1 & -2 & y_3 - y_1 \end{bmatrix}$$

$$(-R_2 + R_1 \to R_1)\&(-R_2 + R_3 \to R_3)$$

$$\begin{bmatrix} 1 & 0 & 2 & 3 & y_2 - y_1 \\ 0 & 1 & 1 & -2 & 2y_1 - y_2 \\ 0 & 0 & 0 & 0 & y_3 + y_2 - 3y_1 \end{bmatrix}$$

So the system $A\mathbf{x} = \mathbf{y}$ will have a solution if and only if $-3y_1 + y_2 + y_3 = 0$; that is, when $y_3 = 3y_1 - y_2$. Hence,

$$\mathcal{R}(A) = \left\{ \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} : y_3 = 3y_1 - y_2 \right\}.$$

Exercise 3.3.8. Let A be as in the above problem. Using the above algebraic specification for $\mathcal{R}(A)$, find a spanning set of vectors for $\mathcal{R}(A)$.

3.3.4 The Column Space of a Matrix

Definition 3.3.9. Given a matrix $A = [\mathbf{A}_1 \ \mathbf{A}_2 \ \cdots \ \mathbf{A}_n]$ (where each $\mathbf{A}_i \in \mathbb{R}^m$), we call

$$\text{Span}\{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n\}$$

the *column space* of A and we shall write $\text{Cspace}(A) = \text{Span}\{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n\}$.

Theorem 3.3.10. Let A be a matrix. Then the column space of A equals the range space of A; that is, $\text{Cspace}(A) = \mathcal{R}(A)$.

Proof. This follows from the proof of Theorem 3.3.7. □

3.3.5 The Row Space of a Matrix

A vector in $\mathbb{R}^n$ can be also be interpreted as a "row vector." For example, we can interpret the vector $\mathbf{A} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$ as the row vector $\mathbf{a} = [1 \ 2 \ 3 \ 4]$. Note that $\mathbf{A}^T = \mathbf{a}$.

Addition and scalar multiplication of row vectors is defined as for column vectors. For example, let $\mathbf{a} = [1\ \ 2\ \ 3\ \ 4]$ and $\mathbf{b} = [-1\ \ 2\ \ -3\ \ -5]$. Then

$$\mathbf{a} + 3\mathbf{b} = [-2\ \ 8\ \ -6\ \ -11]$$

Definition 3.3.11. Given a matrix $A = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_m \end{bmatrix}$ (where each $\mathbf{a}_i \in \mathbb{R}^n$ is a row vector), we call $\text{Span}\{\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_m\}$ the *row space* of A and we shall write $\text{Rspace}(A) = \text{Span}\{\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_m\}$.

Example 5. Let

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

be a 3×4 matrix. Then the row space of A is $\text{Span}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$, where

$$\mathbf{a}_1 = [1\ \ 2\ \ 3\ \ 4],\ \ \mathbf{a}_2 = [1\ \ 1\ \ -2\ \ -3],\ \ \mathbf{a}_3 = [0\ \ -1\ \ 2\ \ 3].$$

The following three lemmas imply that the space, spanned by a set of row vectors, is not changed by performing row operations on the set of vectors.

Lemma 3.3.12. Suppose that $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4$ are vectors in $\mathbb{R}^n$. Let c be a scalar and let

$$\mathbf{a}_3' = c\mathbf{a}_2 + \mathbf{a}_3. \tag{3.1}$$

Then $\text{Span}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4\} = \text{Span}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3', \mathbf{a}_4\}$.

Proof. Let $\mathbf{y}$ be a vector in $\text{Span}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4\}$, so

$$\mathbf{y} = c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + c_3\mathbf{a}_3 + c_4\mathbf{a}_4 \tag{3.2}$$

for some scalars c_1, c_2, c_3, c_4. We must show that $\mathbf{y}$ can also be written as a linear combination of $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3', \mathbf{a}_4\}$. Solving for $\mathbf{a}_3$ in equation (3.1) and substituting in equation (3.2) yields

$$\mathbf{y} = c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + c_3(\mathbf{a}_3' - c\mathbf{a}_2) + c_4\mathbf{a}_4.$$

Now, expanding and collecting terms gives

$$\mathbf{y} = c_1\mathbf{a}_1 + (c_2 - c_3 c)\mathbf{a}_2 + c_3\mathbf{a}_3' + c_4\mathbf{a}_4.$$

Thus, $\mathbf{y}$ can be written as a linear combination of $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3', \mathbf{a}_4\}$. Similarly, given a vector $\mathbf{x}$ in $\text{Span}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3', \mathbf{a}_4\}$, one can show that $\mathbf{x}$ is in $\text{Span}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4\}$. Therefore, $\text{Span}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4\} = \text{Span}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3', \mathbf{a}_4\}$ and this completes the proof. □

Lemma 3.3.13. Suppose that $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4$ are vectors in $\mathbb{R}^n$. Let $c \neq 0$ be a scalar and let $\mathbf{a}_3' = c\mathbf{a}_3$. Then $\text{Span}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4\} = \text{Span}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3', \mathbf{a}_4\}$.

Proof. Exercise. □

Lemma 3.3.14. Suppose that $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4$ are vectors in $\mathbb{R}^n$. Then, upon interchanging $\mathbf{a}_2$ and $\mathbf{a}_3$, we have that $\text{Span}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4\} = \text{Span}\{\mathbf{a}_1, \mathbf{a}_3, \mathbf{a}_2, \mathbf{a}_4\}$.

Theorem 3.3.15. If A and B are row equivalent $m \times n$ matrices, the row spaces of A and B are equal.

Proof. Suppose that $A = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_m \end{bmatrix}$ and $B = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_m \end{bmatrix}$. Since A and B are row equivalent, one can obtain B by applying a sequence of row operations on A. Since the row space of A is given by $\text{Span}\{\, \mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m \,\}$ and the row space of B is given by $\text{Span}\{\, \mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_m \,\}$. Lemmas 3.3.12 to 3.3.14 imply that this same sequence of row operations applied to the row vectors in the set $\{\, \mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m \,\}$ do not change the spanned space. Therefore,

$$\text{Span}\{\, \mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m \,\} = \text{Span}\{\, \mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_m \,\},$$

that is, the row spaces of A and B are equal. This completes the proof. □

Observations. Let $A = [\mathbf{A}_1 \;\; \mathbf{A}_2 \;\; \cdots \;\; \mathbf{A}_n]$ be an $m \times n$ matrix.

1. The null space $\mathcal{N}(A)$ is a subspace of $\mathbb{R}^n$.
2. The row space $\text{Rspace}(A)$ is a subspace of $\mathbb{R}^n$.
3. The range space $\mathcal{R}(A)$ is a subspace of $\mathbb{R}^m$.
4. $\text{Cspace}(A) = \text{Span}\{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n\} = \mathcal{R}(A)$
5. $\text{Cspace}(A) = \text{Rspace}(A^T)$, that is, the column space of a matrix A is the same as the row space of A^T, the transpose of A.

3.3.6 Minimal Spanning Set

Theorem 3.3.15 and item 5 of the above Observations suggests the following procedure for obtaining a "minimal" set of spanning vectors for a subspace. A **minimal spanning set** is a spanning set that has the smallest number of vectors needed to span the subspace.

Minimal Spanning Set Procedure. Given a set of vectors $S = \{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k\}$ in $\mathbb{R}^n$, to find a minimal set S' for which $\text{Span}(S') = \text{Span}(S)$, do the following:

Step 1. Form the matrix $A = [\mathbf{A}_1 \;\; \mathbf{A}_2 \;\; \cdots \;\; \mathbf{A}_k]$.

Step 2. Form the matrix $A^T = \begin{bmatrix} \mathbf{A}_1^T \\ \mathbf{A}_2^T \\ \vdots \\ \mathbf{A}_k^T \end{bmatrix}$.

Step 3. Transform the matrix A^T into reduced echelon form $B^T = \begin{bmatrix} \mathbf{B}_1^T \\ \mathbf{B}_2^T \\ \vdots \\ \mathbf{B}_k^T \end{bmatrix}$.

Step 4. Form the matrix $B = [\mathbf{B}_1 \;\; \mathbf{B}_2 \;\; \cdots \;\; \mathbf{B}_k]$.

Step 5. Since $\text{Rspace}(A^T) = \text{Rspace}(B^T)$ (by Theorem 3.3.15), it follows that $\text{Cspace}(A) = \text{Cspace}(B)$. Thus the set S' consisting of the **nonzero column vectors** of B, is a minimal set such that $\text{Span}(S') = \text{Span}(S)$.

Problem 6 (Example 5 on page 184 of text)**.** Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ where

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \mathbf{v}_2 = \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix} \mathbf{v}_3 = \begin{bmatrix} 1 \\ 4 \\ -5 \end{bmatrix} \mathbf{v}_4 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}.$$

Find a minimal set S' such that $\text{Span}(S') = \text{Span}(S)$.

Solution. We will solve this problem by applying the above procedure.

1. Form the matrix $A = [\mathbf{v}_1 \;\; \mathbf{v}_2 \;\; \mathbf{v}_3 \;\; \mathbf{v}_4] = \begin{bmatrix} 1 & 2 & 1 & 2 \\ 2 & 3 & 4 & 5 \\ 1 & 5 & -5 & -1 \end{bmatrix}$.

2. Form the matrix $A^T = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 5 \\ 1 & 4 & -5 \\ 2 & 5 & -1 \end{bmatrix}$.

3. Transform the matrix A^T into reduced echelon form $B^T = \begin{bmatrix} 1 & 0 & 7 \\ 0 & 1 & -3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

4. Form the matrix $B = [\mathbf{B}_1 \;\; \mathbf{B}_2 \;\; \mathbf{B}_3 \;\; \mathbf{B}_4] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 7 & -3 & 0 & 0 \end{bmatrix}$.

5. Thus the set $S' = \{\mathbf{B}_1, \mathbf{B}_2\}$ consisting of the **nonzero column vectors** of B, is a minimal set such that $\text{Span}(S') = \text{Span}(S)$.

Observation. The above vectors $\mathbf{B}_1$ and $\mathbf{B}_2$ are linearly independent (see Theorem 1.7.7).

3.3.7 Summary

Let A be an $m \times n$ matrix. Suppose that

$$A = [\mathbf{A}_1 \;\; \mathbf{A}_2 \;\; \cdots \;\; \mathbf{A}_n] = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_m \end{bmatrix}.$$

There are three spaces associated to the matrix A:

1. The **Null Space** of the matrix A is defined by $\mathcal{N}(A) = \{\, \mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \boldsymbol{\theta} \,\}$.

2. The **Column Space** of A is defined by $\text{Cspace}(A) = \text{Span}\{\, \mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n \,\}$.

3. The **Row Space** of A is defined by $\text{Rspace}(A) = \text{Span}\{\, \mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m \,\}$.

Exercises 3.3

Pages 186 to 188 – Odds #1–49.

3.4 Basis for Subspaces

In the previous section we were given a set of vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ in a vector space $\mathbb{R}^n$, and we showed how to build the subspace $W = \text{Span}(S)$ consisting of all linear combinations of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$, that is, all vectors of the form

$$\mathbf{x} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_r\mathbf{v}_r$$

where $a_1, a_2, \ldots a_r$ are arbitrary scalars. In this section we will "go in the opposite direction;" that is, we will be given a subspace W and will then be asked to find a set of vectors S which spans W; that is, so that $\text{Span}(S) = W$.

3.4.1 Spanning Sets

Definition 3.4.1. Let W be a subspace of $\mathbb{R}^n$. A set of vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\} \subseteq W$ is said to **span** W if $\text{Span}(S) = W$.

Consider the vector space $\mathbb{R}^n$. A set of vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\} \subseteq \mathbb{R}^n$ is said to **span** $\mathbb{R}^n$ if $\text{Span}(S) = \mathbb{R}^n$.

Example 1. Every vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ in 3-space can be expressed as a linear combination of the unit vectors $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, and $\mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, that is,

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3.$$

Thus, the set of vectors S spans $\mathbb{R}^3$; that is, $\text{Span}\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\} = \mathbb{R}^3$.

The above set S is not the only linearly independent set of vectors which spans the vector space $\mathbb{R}^3$. For example, let

$$\mathbf{v}_1 = \mathbf{e}_1, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \mathbf{e}_3.$$

Then the set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ also spans $\mathbb{R}^3$; that is, $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} = \mathbb{R}^3$.

Problem 2 (Example 1 on page 190 of text)**.** Does $\text{Span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\} = \mathbb{R}^3$ where

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} -2 \\ 3 \\ 1 \end{bmatrix} \text{ and } \mathbf{u}_3 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}?$$

Solution. A vector $\mathbf{y} \in \mathbb{R}^3$ is in $\text{Span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ if and only if there are real numbers x_1, x_2, x_3 such that

$$x_1\mathbf{u}_1 + x_2\mathbf{u}_2 + x_2\mathbf{u}_2 = \mathbf{y}, \tag{3.3}$$

and the vector equation (3.3) has a solution if and only if $A\mathbf{x} = \mathbf{y}$ has a solution $\mathbf{x}$, where $A = [\mathbf{u}_1 \ \mathbf{u}_2 \ \mathbf{u}_3]$. We investigate the conditions on $\mathbf{y} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ which will imply that the system $A\mathbf{x} = \mathbf{y}$

is consistent. That is, we solve the system $A\mathbf{x} = \mathbf{y}$ by forming the augmented matrix and put it into reduced echelon form to obtain

$$[A\,|\,\mathbf{y}] = \begin{bmatrix} 1 & -2 & 1 & a \\ -1 & 3 & 2 & b \\ 0 & 1 & 3 & c \end{bmatrix} \xrightarrow{\text{reduced echelon form}} \begin{bmatrix} 1 & 0 & 0 & 10a+9b-7c \\ 0 & 1 & 0 & 4a+4b-3c \\ 0 & 0 & 1 & -a-b+c \end{bmatrix}$$

and thus, we obtain the solution

$$\begin{aligned} x_1 &= 10a+9b-7c \\ x_2 &= -4a+4b-3c \\ x_3 &= -a-b+c. \end{aligned} \tag{3.4}$$

So the system $A\mathbf{x} = \mathbf{y}$ will have a solution for any vector $\mathbf{y} \in \mathbb{R}^3$. Hence, $\text{Span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\} = \mathbb{R}^3$. For example, $\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$ in $\text{Span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ because letting $a = 0$, $b = 1$, $c = 1$ in (3.4) we obtain

$$\begin{aligned} x_1 &= 10(0)+9(1)-7(1) = 2 \\ x_2 &= 4(0)+4(1)-3(1) = 1 \\ x_3 &= -(0)-(1)+(1) = 0, \end{aligned}$$

and these values will satisfy the vector equation

$$\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = 2\mathbf{u}_1 + 1\mathbf{u}_2 + 0\mathbf{u}_3.$$

Problem 3 (Example 2 on page 191 of text)**.** Does $\text{Span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\} = \mathbb{R}^3$ where

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} -1 \\ 0 \\ -7 \end{bmatrix} \text{ and } \mathbf{u}_3 = \begin{bmatrix} 2 \\ 7 \\ 0 \end{bmatrix}?$$

Solution. A vector $\mathbf{y} \in \mathbb{R}^3$ is in $\text{Span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ if and only if $A\mathbf{x} = \mathbf{y}$ has a solution $\mathbf{x}$, where $A = [\mathbf{u}_1\ \mathbf{u}_2\ \mathbf{u}_3]$. We investigate the conditions on $\mathbf{y} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ which will imply that the system $A\mathbf{x} = \mathbf{y}$ is consistent. That is, we solve the system $A\mathbf{x} = \mathbf{y}$ by forming the augmented matrix and put it into reduced echelon form to obtain

$$[A\,|\,\mathbf{y}] = \begin{bmatrix} 1 & -1 & 2 & a \\ 2 & 0 & 7 & b \\ 3 & -7 & 0 & c \end{bmatrix} \xrightarrow{\text{reduced echelon form}} \begin{bmatrix} 1 & 0 & \frac{7}{2} & \frac{b}{2} \\ 0 & 1 & \frac{3}{2} & -a+\frac{b}{2} \\ 0 & 0 & 0 & -7a+2b+c \end{bmatrix}.$$

So the system $A\mathbf{x} = \mathbf{y}$ will have a solution if and only if $-7a + 2b + c = 0$. Thus, in particular, $\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$ is not in $\text{Span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$. Hence, $\text{Span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\} \neq \mathbb{R}^3$.

Definition 3.4.2. A set of vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ in a vector space W is called a **basis** for W if, (1) S is linearly independent and (2) S spans W, that is, $\text{Span}(S) = W$.

Remark 3.4.3. A set of vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ in a vector space $\mathbb{R}^k$ is called a **basis** for $\mathbb{R}^k$ if, (1) S is linearly independent and (2) S spans $\mathbb{R}^k$, that is, $\text{Span}(S) = \mathbb{R}^k$.

3.4.2 Uniqueness of Representation

Theorem 3.4.4. If $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis for a vector space V, then every vector $\mathbf{y}$ in V can be expressed in the form $\mathbf{y} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$ in exactly one way.

Proof. Since S spans the vector space V, every vector $\mathbf{y}$ is a linear combination of the vectors $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$, that is, $\mathbf{y}$ has the form $\mathbf{y} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$. To show that there is only one way to write $\mathbf{y}$ as a linear combination of the vectors in S, suppose that there is another way, say, $\mathbf{y} = d_1\mathbf{v}_1 + d_2\mathbf{v}_2 + \cdots + d_n\mathbf{v}_n$. Subtracting the first equation from the second yields

$$(c_1 - d_1)\mathbf{v}_1 + (c_2 - d_2)\mathbf{v}_2 + \cdots + (c_n - d_n)\mathbf{v}_n = \boldsymbol{\theta}.$$

But the set S is linearly independent! Hence,

$$(c_1 - d_1) = 0,\ (c_2 - d_2) = 0,\ \ldots,\ (c_n - d_n) = 0$$

and thus, $c_1 = d_1, c_2 = d_2, \ldots, c_n = d_n$. □

Problem 4. Consider the vector space $\mathbb{R}^4$ and the vectors

$$\mathbf{v}_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} \mathbf{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} \mathbf{v}_4 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}.$$

Show that $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_3\}$ a basis for $\mathbb{R}^4$. Then, express the vector $\mathbf{b} = \begin{bmatrix} 3 \\ 6 \\ -3 \\ 0 \end{bmatrix}$ as a linear combination of the vectors in S.

Solution. A vector $\mathbf{y} \in \mathbb{R}^4$ is in Span$\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3,\ \mathbf{v}_4\}$ if and only if $A\mathbf{x} = \mathbf{y}$ has a solution $\mathbf{x}$, where $A = [\mathbf{v}_1\ \ \mathbf{v}_2\ \ \mathbf{v}_3\ \ \mathbf{v}_4]$. We investigate the conditions on $\mathbf{y} = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$ which will imply that the system $A\mathbf{x} = \mathbf{y}$ is consistent. That is, we solve the system $A\mathbf{x} = \mathbf{y}$ by forming the augmented matrix and put it into reduced echelon to obtain

$$[A\,|\,\mathbf{y}] = \begin{bmatrix} 0 & 1 & 1 & 1 & a \\ 1 & 0 & 1 & 1 & b \\ 1 & 1 & 0 & 1 & c \\ 1 & 1 & 1 & 0 & e \end{bmatrix} \xrightarrow{\text{reduced echelon form}} \begin{bmatrix} 1 & 0 & 0 & 0 & \left(\frac{1}{3}b - \frac{2}{3}a + \frac{1}{3}c + \frac{1}{3}d\right) \\ 0 & 1 & 0 & 0 & \left(\frac{1}{3}a - \frac{2}{3}b + \frac{1}{3}c + \frac{1}{3}d\right) \\ 0 & 0 & 1 & 0 & \left(\frac{1}{3}a + \frac{1}{3}b - \frac{2}{3}c + \frac{1}{3}d\right) \\ 0 & 0 & 0 & 1 & \left(\frac{1}{3}a + \frac{1}{3}b + \frac{1}{3}c - \frac{2}{3}d\right) \end{bmatrix}. \quad (3.5)$$

So the system $A\mathbf{x} = \mathbf{y}$ will always have a solution. Hence, Span$\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\} = \mathbb{R}^4$. Thus, in particular, $\mathbf{b} \in$ Span$\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$. We must also show that S is linearly independent. To do this, we must show that the vector equation

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 + c_4\mathbf{v}_4 = \boldsymbol{\theta}$$

has **only** the **trivial** solution; that is, $\mathbf{c} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix} = \boldsymbol{\theta}$. We apply the **Linear Dependence Algorithm**; the matrix $A = [\mathbf{v}_1 \ \mathbf{v}_2 \ \mathbf{v}_3 \ \mathbf{v}_4]$ is

$$\begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}.$$

To solve the homogeneous system $A\mathbf{c} = \boldsymbol{\theta}$ we transform the augmented matrix into reduced echelon form

$$[A \,|\, \boldsymbol{\theta}] = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \end{bmatrix} \xrightarrow{\text{reduced echelon form}} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Therefore, the only solution to the homogeneous system $A\mathbf{c} = \boldsymbol{\theta}$ is the trivial solution $\mathbf{c} = \boldsymbol{\theta}$. Hence, the set of vectors S is linearly independent.

Finally, to express the vector $\mathbf{b} = \begin{bmatrix} 3 \\ 6 \\ -3 \\ 0 \end{bmatrix}$ as a linear combination of the vectors in S, we must solve for the scalars c_1, c_2, c_3, c_4 which satisfy the vector equation

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 + c_4\mathbf{v}_4 = \mathbf{b}$$

We apply the **Linear Combination Algorithm** and solve the system $A\mathbf{c} = \mathbf{b}$ for $\mathbf{c}$. We form the augmented matrix $[A \,|\, \mathbf{b}]$ and put into reduced echelon form. We will get, from (3.5) above,

$$\begin{bmatrix} 1 & 0 & 0 & 0 & \left(\frac{1}{3}b - \frac{2}{3}a + \frac{1}{3}c + \frac{1}{3}d\right) \\ 0 & 1 & 0 & 0 & \left(\frac{1}{3}a - \frac{2}{3}b + \frac{1}{3}c + \frac{1}{3}d\right) \\ 0 & 0 & 1 & 0 & \left(\frac{1}{3}a + \frac{1}{3}b - \frac{2}{3}c + \frac{1}{3}d\right) \\ 0 & 0 & 0 & 1 & \left(\frac{1}{3}a + \frac{1}{3}b + \frac{1}{3}c - \frac{2}{3}d\right) \end{bmatrix}$$

where $a = 3, b = 6, c = -3, d = 0$. Hence, $c_1 = \frac{1}{3}(6) - \frac{2}{3}(3) + \frac{1}{3}(-3) + \frac{1}{3}(0) = -1$. Similarly, we get $c_2 = -4, c_3 = 5, c_4 = 2$. So,

$$\mathbf{b} = -1\mathbf{v}_1 + -4\mathbf{v}_2 + 5\mathbf{v}_3 + 2\mathbf{v}_4.$$

3.4.3 Finding a Basis for the Null Space of a Matrix

The next example will give a method for finding a basis for the null space of a matrix A.

Problem 5 (Example 3 on page 192 of text)**.** Find a basis for the null space of

$$A = \begin{bmatrix} 1 & 1 & 3 & 1 \\ 2 & 1 & 5 & 4 \\ 1 & 2 & 4 & -1 \end{bmatrix}.$$

Solution. First notice, since A is a 3×4 matrix, that the null space of A is a subspace of $\mathbb{R}^4$. Putting the augmented matrix $[A \,|\, \boldsymbol{\theta}]$, into reduced echelon form we get

$$[B \,|\, \boldsymbol{\theta}] = \begin{bmatrix} 1 & 0 & 2 & 3 & 0 \\ 0 & 1 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Observe that x_3 and x_4 are free variables in the system $B\mathbf{x} = \boldsymbol{\theta}$. Solving the equivalent system of equations $B\mathbf{x} = \boldsymbol{\theta}$, we obtain

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

Therefore, every solution to $A\mathbf{x} = \boldsymbol{\theta}$ can be written in the vector form

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

We conclude that the vectors $\mathbf{u}_1 = \begin{bmatrix} -2 \\ -1 \\ 1 \\ 0 \end{bmatrix}$ and $\mathbf{u}_2 = \begin{bmatrix} -3 \\ 2 \\ 0 \\ 1 \end{bmatrix}$ span the null space of A, that is, $\mathrm{Span}\{\mathbf{u}_1, \mathbf{u}_2\} = \mathcal{N}(A)$. Since the vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ are linearly independent (see Theorem 1.7.7), we see that $\{\mathbf{u}_1, \mathbf{u}_2\}$ forms a basis for $\mathcal{N}(A)$.

A basis for the null space of an $m \times n$ matrix A is obtained as follows: First solve the homogeneous system $A\mathbf{x} = \boldsymbol{\theta}$ by transforming the augmented matrix $[A \,|\, \boldsymbol{\theta}]$ into reduced echelon form to obtain $[B \,|\, \boldsymbol{\theta}]$. Now solve the equivalent system of equations $B\mathbf{x} = \boldsymbol{\theta}$ in vector form (see section 1.5.3 starting on page 30). Each free variable will yield a basis vector for the null space of A. This leads to the following theorem.

Theorem 3.4.5. Let A be an $m \times n$ matrix. Let $[B \,|\, \boldsymbol{\theta}]$ be the result of transforming the augmented matrix $[A \,|\, \boldsymbol{\theta}]$ into reduced echelon form. Then the number of free variables in the system $B\mathbf{x} = \boldsymbol{\theta}$ equals the number of vectors in a basis for the null space of A.

3.4.4 Finding a Basis for the Row Space of a Matrix

Theorem 3.4.6. Let A be a nonzero matrix and suppose that B is the result of putting A into reduced echelon form. Then the nonzero rows of B form a basis for the row space of A.

Proof. By Theorem 3.3.15 A and B have the same row space. It follows that the nonzero rows of the matrix B span the row space of A. Since the nonzero rows of a matrix in reduced echelon form are linearly independent, it follows that the nonzero rows form a basis for the row space of A. □

Problem 6. Find a basis for the row space of the matrix

$$A = \begin{bmatrix} 1 & 1 & 3 & 1 \\ 2 & 1 & 5 & 4 \\ 1 & 2 & 4 & -1 \end{bmatrix}.$$

Solution. After putting A into reduced echelon form we obtain the matrix

$$B = \begin{bmatrix} 1 & 0 & 2 & 3 \\ 0 & 1 & 1 & -2 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Let $\mathbf{b}_1 = [1\ \ 0\ \ 2\ \ 3]$ and $\mathbf{b}_2 = [0\ \ 1\ \ 1\ \ -2]$ be the nonzero rows of B. (Clearly $\mathbf{b}_1$ and $\mathbf{b}_2$ are linearly independent.) By Theorem 3.4.6 we have that $\{\mathbf{b}_1, \mathbf{b}_2\}$ is a basis for the row space of A.

The idea used in Problem 6 inspires the following algorithm for obtaining a basis for the span of any finite set of vectors. This algorithm is the minimal spanning algorithm on page 64.

Basis Algorithm. Given a set of vectors $S = \{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k\}$ in $\mathbb{R}^n$, to find a basis S' for the subspace $W = \text{Span}(S)$, do the following:

Step 1. Form the matrix $A = [\mathbf{A}_1\ \ \mathbf{A}_2\ \ \cdots\ \ \mathbf{A}_k]$.

Step 2. Form the matrix $A^T = \begin{bmatrix} \mathbf{A}_1^T \\ \mathbf{A}_2^T \\ \vdots \\ \mathbf{A}_k^T \end{bmatrix}$.

Step 3. Transform the matrix A^T into reduced echelon form $B^T = \begin{bmatrix} \mathbf{B}_1^T \\ \mathbf{B}_2^T \\ \vdots \\ \mathbf{B}_k^T \end{bmatrix}$.

Step 4. Form the matrix $B = [\mathbf{B}_1\ \ \mathbf{B}_2\ \ \cdots\ \ \mathbf{B}_k]$.

Step 5. The set S' of the nonzero column vectors of B, is a basis for $W = \text{Span}(S)$.

3.4.5 Finding a Basis for the Column Space of a Matrix

Theorem 3.4.7. Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ be vectors in $\mathbb{R}^n$. Let A be the matrix

$$A = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_k \end{bmatrix}$$

and let $\mathbf{c} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix}$ be a column matrix of scalars c_i. Then $\mathbf{c}$ is a solution to the linear system $A\mathbf{c} = \boldsymbol{\theta}$ if and only if $(c_1, c_2, \ldots, c_k)$ is a solution to the vector equation $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k = \boldsymbol{\theta}$.

Proof. See Theorem 1.7.2. □

Theorem 3.4.8. Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ be a set of vectors in $\mathbb{R}^n$. Let A be the matrix $A = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_k \end{bmatrix}$. Transforming the matrix A into reduced row echelon form we obtain the matrix $B = \begin{bmatrix} \mathbf{v}'_1 & \mathbf{v}'_2 & \cdots & \mathbf{v}'_k \end{bmatrix}$. Let $(c_1, c_2, \ldots, c_k)$ be a list scalars. Then $(c_1, c_2, \ldots, c_k)$ is a solution to the vector equation

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k = \boldsymbol{\theta}$$

if and only if $(c_1, c_2, \ldots, c_k)$ is a solution to the vector equation

$$c_1\mathbf{v}'_1 + c_2\mathbf{v}'_2 + \cdots + c_k\mathbf{v}'_k = \boldsymbol{\theta}.$$

Proof. Let $\mathbf{c} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix}$ be the column matrix of given scalars c_i. By Theorem 3.4.7, $(c_1, c_2, \ldots, c_k)$ is a solution to the vector equation

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k = \boldsymbol{\theta}$$

if and only if $\mathbf{c}$ is a solution to the homogeneous system $A\mathbf{c} = \boldsymbol{\theta}$. Since the matrices A and B are row equivalent, it follows that the augmented matrices $[A \,|\, \boldsymbol{\theta}]$ and $[B \,|\, \boldsymbol{\theta}]$ are row equivalent. Therefore, the systems $A\mathbf{c} = \boldsymbol{\theta}$ and $B\mathbf{c} = \boldsymbol{\theta}$ have exactly the same solutions (see Theorem 1.1.3). Hence, $\mathbf{c}$ is a solution to the system $A\mathbf{c} = \boldsymbol{\theta}$ if and only if $\mathbf{c}$ is a solution to the system $B\mathbf{c} = \boldsymbol{\theta}$. But Theorem 3.4.7, again, implies that $\mathbf{c}$ is a solution to the system $B\mathbf{c} = \boldsymbol{\theta}$ if and only if $(c_1, c_2, \ldots, c_k)$ is a solution to the vector equation

$$c_1\mathbf{v}'_1 + c_2\mathbf{v}'_2 + \cdots + c_k\mathbf{v}'_k = \boldsymbol{\theta}.$$

This completes the proof of Theorem 3.4.8. □

We will now apply Theorem 3.4.8 in our solution to the next problem.

Problem 7 (See Example 6 on page 197 of text)**.** Find a basis for the column space of the matrix

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 2 \\ 1 & 2 & 4 & 0 & 5 \\ 2 & 1 & -1 & 4 & 0 \\ -1 & 1 & 5 & -1 & 2 \end{bmatrix}$$

that consists of column vectors from the matrix A.

Solution. Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_5\}$ be the set of column vectors in the matrix A, that is,

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 2 \\ -1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 4 \\ -1 \\ 5 \end{bmatrix}, \quad \mathbf{v}_4 = \begin{bmatrix} 1 \\ 0 \\ 4 \\ -1 \end{bmatrix}, \quad \mathbf{v}_5 = \begin{bmatrix} 2 \\ 5 \\ 0 \\ 2 \end{bmatrix}.$$

We will find a subset of S that is a basis for the space $\text{Cspace}(A) = \text{Span}(S)$. Recall for the set of vectors S that $\text{Span}(S)$ consists of all linear combinations of the vectors in S, that is, all vectors $\mathbf{y}$ of the form

$$\mathbf{y} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 + c_4\mathbf{v}_4 + c_5\mathbf{v}_5$$

where $c_1, \ldots, c_5$ are scalars. Consider the matrix

$$A = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{v}_4 & \mathbf{v}_5 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 2 \\ 1 & 2 & 4 & 0 & 5 \\ 2 & 1 & -1 & 4 & 0 \\ -1 & 1 & 5 & -1 & 2 \end{bmatrix}.$$

Now transform this matrix into reduced row echelon form to obtain

$$B = \begin{bmatrix} \mathbf{v}_1' & \mathbf{v}_2' & \mathbf{v}_3' & \mathbf{v}_4' & \mathbf{v}_5' \end{bmatrix} = \begin{bmatrix} 1 & 0 & -2 & 0 & 1 \\ 0 & 1 & 3 & 0 & 2 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Notice that the columns $\mathbf{v}_1'$, $\mathbf{v}_2'$, $\mathbf{v}_4'$ of the matrix B contain the leading 1's. Clearly, the set of vectors $\{\mathbf{v}_1', \mathbf{v}_2', \mathbf{v}_4'\}$ is a linearly independent set. We shall now show that the set $T = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4\}$ is a basis for the space Span(S) by showing: (a) T is linearly independent and (b) T spans the space Span(S).

(a) *Show that T is linearly independent.* It follows from Theorem 3.4.8 that the set $T = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4\}$ is also linearly independent. To see why, suppose that $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_4\mathbf{v}_4 = \boldsymbol{\theta}$. Then, Theorem 3.4.8 (letting $c_3 = c_5 = 0$) implies that $c_1\mathbf{v}_1' + c_2\mathbf{v}_2' + c_4\mathbf{v}_4' = \boldsymbol{\theta}$. But this implies that $c_1 = c_2 = c_4 = 0$ is the only solution. Hence, T is linearly independent.

(b) *Show that T spans the space* Span(S). Clearly,

$$\mathbf{v}_3' = -2\mathbf{v}_1' + 3\mathbf{v}_2'$$

and so,

$$-2\mathbf{v}_1' + 3\mathbf{v}_2' + -1\mathbf{v}_3' = \boldsymbol{\theta}.$$

Theorem 3.4.8 (letting $c_4 = c_5 = 0$) implies that

$$-2\mathbf{v}_1 + 3\mathbf{v}_2 + -1\mathbf{v}_3 = \boldsymbol{\theta}$$

and so,

$$\mathbf{v}_3 = -2\mathbf{v}_1 + 3\mathbf{v}_2 \tag{3.1}$$

Similarly, one can show that

$$\mathbf{v}_5 = \mathbf{v}_1 + 2\mathbf{v}_2 + -1\mathbf{v}_4. \tag{3.2}$$

Now let $\mathbf{y}$ be a vector in Span(S), say,

$$\mathbf{y} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 + c_4\mathbf{v}_4 + c_5\mathbf{v}_5. \tag{3.3}$$

Using equations (3.1) and (3.2) to substitute for $\mathbf{v}_3$ and $\mathbf{v}_5$ in equation (3.3), and collecting "like vectors," we get

$$\mathbf{y} = (c_1 - 2c_3 + c_5)\mathbf{v}_1 + (c_2 + 3c_3 + 2c_5)\mathbf{v}_2 + (c_4 - c_5)\mathbf{v}_4.$$

Therefore, $\mathbf{y}$ can be written as a linear combination of vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4$ alone. Thus, T spans Span(S).

Therefore, $T = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4\}$ is a basis for the column space of the matrix A, and T consists only of column vectors from the matrix A.

The above solution to Problem 7 suggests the proof of the following two theorems.

Theorem 3.4.9. Let A be a nonzero matrix and suppose that B is the result of putting A into reduced echelon form. Then let $\mathbf{v}'_i, \ldots, \mathbf{v}'_j$ be the column vectors of the matrix B that contain the leading 1's. Then the corresponding column vectors $\mathbf{v}_i, \ldots, \mathbf{v}_j$ of the matrix A form a basis for the column space of A.

Theorem 3.4.10. Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ be a set of vectors in $\mathbb{R}^n$ and let $W = \text{Span}(S)$. Then there is a subset of S which forms a basis for the space W.

The idea used in Problem 7 also inspires the following algorithm for obtaining a basis for the span of any finite set S of vectors that is a subset of the set S.

Subset Basis Algorithm. Given a set of vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ in $\mathbb{R}^n$ to find a subset of S which forms a basis for the space $W = \text{Span}(S)$, do the following:

Step 1. Form the matrix $A = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_m \end{bmatrix}$.

Step 2. Transform the matrix $A = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_m \end{bmatrix}$ into reduced row echelon form, obtaining $B = \begin{bmatrix} \mathbf{v}'_1 & \mathbf{v}'_2 & \cdots & \mathbf{v}'_m \end{bmatrix}$.

Step 3. Identify all the columns of B containing leading 1's, say $\mathbf{v}'_{i_1}, \mathbf{v}'_{i_2}, \ldots, \mathbf{v}'_{i_m}$.

Step 4. The subset T of S given by $T = \{\mathbf{v}_{i_1}, \mathbf{v}_{i_2}, \ldots, \mathbf{v}_{i_m}\}$ forms a basis for $W = \text{Span}(S)$.

We now give another application of Theorem 3.4.8.

Problem 8 (Example 4 on page 193 of text)**.** Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, where

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 3 \\ 5 \\ 1 \end{bmatrix}.$$

Show that S is a linearly dependent set, and find a subset of S that is a basis for the space $W = \text{Span}(S)$.

Solution. Form the matrix

$$A = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 5 \\ 1 & 1 & 0 \end{bmatrix}.$$

Now transform this matrix into reduced row echelon form to obtain

$$B = \begin{bmatrix} \mathbf{v}'_1 & \mathbf{v}'_2 & \mathbf{v}'_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}.$$

Notice that the columns $\mathbf{v}'_1$, $\mathbf{v}'_2$ of the matrix B contain the leading 1's. Clearly, the set of vectors $\{\mathbf{v}'_1, \mathbf{v}'_2, \mathbf{v}'_3\}$ is a linearly dependent set; because

$$\mathbf{v}'_1 - 2\mathbf{v}'_2 + \mathbf{v}'_3 = \boldsymbol{\theta}.$$

So by Theorem 3.4.8,

$$\mathbf{v}_1 - 2\mathbf{v}_2 + \mathbf{v}_3 = \boldsymbol{\theta}.$$

Hence, the set S is linearly dependent. Finally, by the Subset Basis Algorithm, the set $T = \{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis for $\text{Span}(S)$. This completes the solution of Example 4 of text.

Remark 3.4.11. Let A be a matrix and let B be the result of transforming A into reduced echelon form by performing row operations on the matrix A. We make the following observations:

1. Row operations **preserve** the row space of a matrix, that is, $\text{Rspace}(A) = \text{Rspace}(B)$ (see Theorem 3.3.15).

2. Row operations **do not preserve** the column space of a matrix, that is, $\text{Cspace}(A) \neq \text{Cspace}(B)$, in general.

Example 9. Let A be the matrix $A = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$ and let B be the result of transforming A into reduced echelon form obtaining $B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$. Now observe the following:

1. Row operations **preserve** the row space of the matrix A, that is, $\text{Rspace}(A) = \mathbb{R}^2 = \text{Rspace}(B)$.

2. Row operations **do not preserve** the column space of the matrix A, that is, $\text{Cspace}(A) \neq \text{Cspace}(B)$. To see this, note that

$$\text{Cspace}(A) = \text{Span}\left\{ \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\} = \left\{ \begin{bmatrix} 0 \\ y \\ z \end{bmatrix} : y, z \in \mathbb{R} \right\}$$

and that

$$\text{Cspace}(B) = \text{Span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\} = \left\{ \begin{bmatrix} x \\ y \\ 0 \end{bmatrix} : x, y \in \mathbb{R} \right\}.$$

Thus, $\text{Cspace}(A) \neq \text{Cspace}(B)$.

Exercises 3.4

Pages 200 to 202 of text – Odds #1–15; 21, 23, 24, 25, 26, 27, 33, 35.

3.5 Dimension

3.5.1 The Definition of Dimension

The following theorem states that the number of vectors in a linearly independent set is less than or equal to the number of vectors in a spanning set.

Theorem 3.5.1. Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a spanning set for a vector space V, that is, let $\text{Span}(S) = V$. If $T = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m\}$ is any set of linearly independent vectors in V, then $m \leq n$.

Proof. Because S spans V there are scalars a_{ji} such that

$$\begin{aligned} \mathbf{y}_1 &= a_{11}\mathbf{v}_1 + a_{21}\mathbf{v}_2 + \cdots + a_{n1}\mathbf{v}_n \\ \mathbf{y}_2 &= a_{12}\mathbf{v}_1 + a_{22}\mathbf{v}_2 + \cdots + a_{n2}\mathbf{v}_n \\ &\vdots \\ \mathbf{y}_m &= a_{m1}\mathbf{v}_1 + a_{2m}\mathbf{v}_2 + \cdots + a_{nm}\mathbf{v}_n. \end{aligned} \tag{$\blacktriangle$}$$

We shall use "proof by contradiction." Suppose, for a contradiction, that $m > n$. We shall then demonstrate that this implies that $T = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m\}$ is linearly dependent (which is not the case, and therefore, we must have that $m \leq n$).

To demonstrate that the assumption "$m > n$" implies that T is linearly dependent, we must show that there are scalars $c_1, c_2, \ldots, c_m$, **not all zero**, such that

$$c_1\mathbf{y}_1 + c_2\mathbf{y}_2 + \cdots + c_m\mathbf{y}_m = \boldsymbol{\theta}.$$

To see that such scalars c_i exist, multiply the first equation in ($\blacktriangle$) by c_1, multiply the second equation in ($\blacktriangle$) by c_2, and do the same for the remaining equations. We thus obtain

$$\begin{aligned} c_1\mathbf{y}_1 &= c_1a_{11}\mathbf{v}_1 + c_1a_{21}\mathbf{v}_2 + \cdots + c_1a_{n1}\mathbf{v}_n \\ c_2\mathbf{y}_2 &= c_2a_{12}\mathbf{v}_1 + c_2a_{22}\mathbf{v}_2 + \cdots + c_2a_{n2}\mathbf{v}_n \\ &\vdots \\ c_m\mathbf{y}_m &= c_ma_{1m}\mathbf{v}_1 + c_ma_{2m}\mathbf{v}_2 + \cdots + c_ma_{nm}\mathbf{v}_n. \end{aligned}$$

Adding the left hand sides and the right hand sides together and collecting terms we get

$$\begin{aligned} c_1\mathbf{y}_1 + &c_2\mathbf{y}_2 + \cdots + c_m\mathbf{y}_m \\ &= (c_1a_{11} + c_2a_{12} + \cdots + c_ma_{1m})\mathbf{v}_1 + (c_1a_{21} + c_2a_{22} + \cdots + c_ma_{2m})\mathbf{v}_2 + \cdots \\ &\quad \cdots + (c_1a_{n1} + c_2a_{n2} + \cdots + c_ma_{nm})\mathbf{v}_n. \end{aligned}$$

Therefore, if $c_1, c_2, \ldots, c_m$ is a nontrivial solution to the homogeneous system

$$\begin{aligned} a_{11}c_1 + a_{12}c_2 + \cdots + a_{1m}c_m &= 0 \\ a_{21}c_1 + a_{22}c_2 + \cdots + a_{2m}c_m &= 0 \\ &\vdots \\ a_{n1}c_1 + a_{n2}c_2 + \cdots + a_{nm}c_m &= 0, \end{aligned} \tag{3.4}$$

then $c_1, c_2, \ldots, c_m$ is a nontrivial solution to the vector equation $c_1\mathbf{y}_1 + c_2\mathbf{y}_2 + \cdots + c_m\mathbf{y}_m = \boldsymbol{\theta}$. But the number of unknowns m in the above homogeneous system is greater than the number of equations n (because by assumption, $m > n$). Theorem 1.3.5 implies that a nontrivial solution $c_1, c_2, \ldots, c_m$ to the homogeneous system (3.4) exists. Hence, T is linearly dependent, which is a contradiction. Therefore, $m \leq n$. □

The following corollary implies that the dimension of a vector space is unambiguous.

Corollary 3.5.2. If $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ and $T = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m\}$ are basis for a vector space V, then $n = m$.

Proof. Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ and $T = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m\}$ both be a basis for a vector space V. Thus, S and T are linearly independent sets, $\mathrm{Span}(S) = V$, and $\mathrm{Span}(T) = V$. Theorem 3.5.1 thus implies that $m \leq n$ and $n \leq m$. So, $n = m$. □

Thus, the notion of the dimension of a vector space is a mathematically precise notion.

Definition 3.5.3. The **dimension** of a non-zero vector space V, denoted by $\dim(V)$, is the number of vectors in a basis for V. If $V = \{\boldsymbol{\theta}\}$, then $\dim(V) = 0$.

Problem 1. Find a basis and the dimension of the following subspace W of $\mathbf{R}^4$ defined by the follow algebraic specification

$$W = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} : x_3 = x_1 - x_2 \text{ and } x_4 = x_1 + x_2 \right\}.$$

Solution. Every vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$ in W has the form

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_1 - x_2 \\ x_1 + x_2 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ -1 \\ 1 \end{bmatrix}.$$

So the vectors $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 1 \end{bmatrix}$ span W; and one can easily check that $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly independent by Theorem 1.7.7(zero-one law). Therefore, $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis for the space W. So, W has dimension 2.

Problem 2 (See Example 2 on page 205 of text)**.** Consider the subspace $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$ of $\mathbb{R}^3$ where

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 2 \\ 4 \\ 0 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} 3 \\ 5 \\ 2 \end{bmatrix}, \quad \mathbf{u}_4 = \begin{bmatrix} 2 \\ 5 \\ -2 \end{bmatrix}.$$

Find three different basis for W by applying the following 3 techniques:

(a) Find an algebraic specification for W and find a basis for W as in the above Problem 1.
(b) Apply the Subset Basis Algorithm on page 74 of these notes.
(c) Apply the Basis Algorithm on page 71 of these notes.

Solution. We solve (a), (b) and (c), in this order, below.

(a). A vector $\mathbf{y} \in \mathbb{R}^3$ is in $\text{Span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$ if and only if $A\mathbf{x} = \mathbf{y}$ has a solution $\mathbf{x}$, where $A = [\mathbf{u}_1 \ \mathbf{u}_2 \ \mathbf{u}_3 \ \mathbf{u}_4]$. We investigate the conditions on $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$ which will imply that the system $A\mathbf{x} = \mathbf{y}$ is consistent. That is, we solve the system $A\mathbf{x} = \mathbf{y}$ by forming the augmented matrix

$$[A \,|\, \mathbf{y}] = \begin{bmatrix} 1 & 2 & 3 & 2 & y_1 \\ 1 & 4 & 5 & 5 & y_2 \\ 2 & 0 & 2 & -2 & y_3 \end{bmatrix}$$

and put it into reduced echelon form to obtain

$$\begin{bmatrix} 1 & 0 & 1 & -1 & 2y_1 - y_2 \\ 0 & 1 & 1 & \frac{3}{2} & -\frac{y_1}{2} + \frac{y_2}{2} \\ 0 & 0 & 0 & 0 & -4y_1 + 2y_2 + y_3 \end{bmatrix}$$

So the system $A\mathbf{x} = \mathbf{y}$ will have a solution if and only if $-4y_1 + 2y_2 + y_3 = 0$; that is, when $y_3 = 4y_1 - 2y_2$. Hence,

$$W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\} = \left\{ \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} : y_3 = 4y_1 - 2y_2 \right\}.$$

Thus, every vector $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$ in W has the form

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ 4y_1 - 2y_2 \end{bmatrix} = y_1 \begin{bmatrix} 1 \\ 0 \\ 4 \end{bmatrix} + y_2 \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix}.$$

So the vectors $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 4 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix}$ span W; and by Theorem 1.7.7(zero-one law) the set $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly independent. So $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis for the space W. Thus, W has dimension 2.

(b). We now apply the Subset Basis Algorithm.

Step 1. $A = [\mathbf{u}_1\ \mathbf{u}_2\ \mathbf{u}_3\ \mathbf{u}_4] = \begin{bmatrix} 1 & 2 & 3 & 2 \\ 1 & 4 & 5 & 5 \\ 2 & 0 & 2 & -2 \end{bmatrix}$.

Step 2. Transform the matrix A into reduced row echelon form, obtaining

$$B = [\mathbf{u}'_1\ \mathbf{u}'_2\ \mathbf{u}'_3\ \mathbf{u}'_4] = \begin{bmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & 1 & 3/2 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Step 3. Identify all the columns of B containing leading 1's, getting $\mathbf{u}'_1, \mathbf{u}'_2$.

Step 4. The subset T of S given by $T = \{\mathbf{u}_1, \mathbf{u}_2\}$ forms a basis for $W = \text{Span}(S)$.

(c). Now we apply the Basis Algorithm.

Step 1. Form the matrix $A = [\mathbf{u}_1\ \mathbf{u}_2\ \mathbf{u}_3\ \mathbf{u}_4] = \begin{bmatrix} 1 & 2 & 3 & 2 \\ 1 & 4 & 5 & 5 \\ 2 & 0 & 2 & -2 \end{bmatrix}$.

Step 2. Form the matrix $A^T = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 4 & 0 \\ 3 & 4 & 2 \\ 2 & 5 & -2 \end{bmatrix}$.

Step 3. Transform the matrix A^T into reduced echelon form $B^T = \begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

Step 4. Form the matrix $B = [\mathbf{B}_1\ \mathbf{B}_2\ \mathbf{B}_3\ \mathbf{B}_4] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 4 & -2 & 0 & 0 \end{bmatrix}$.

Step 5. Thus the set $\{\mathbf{B}_1, \mathbf{B}_2\}$ consisting of the nonzero column vectors of B, is a basis for W.

3.5.2 Properties of an n–Dimensional Subspace

Theorem 3.5.4. Let W be a subspace of $\mathbb{R}^k$ with dimension $\dim(W) = n$. Let T be a set of distinct vectors from W, say $T = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m\}$. Then

1. If $m > n$, then T is linearly dependent.
2. If $m < n$, then $\text{Span}(T) \neq W$.
3. If $m = n$ and T is a linearly independent set, then T is a basis for W.
4. If $m = n$ and $\text{Span}(T) = W$, then T is a basis for W.

Proof. Let W be a subspace of $\mathbb{R}^k$ with dimension $\dim(W) = n$. Let T be a set of distinct vectors from W, say $T = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m\}$. Since $\dim(W) = n$, the subspace W has a basis $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$. So, in particular, (▲) S is linearly independent and $\text{Span}(S) = W$. We now prove items 1 through 4.

1. Assume that $m > n$. Since $m > n$, Theorem 3.5.1 implies that T is linearly dependent.

2. Assume that $m < n$. We shall show that $\text{Span}(T) \neq W$. Suppose, for a contradiction, that $\text{Span}(T) = W$. Since $m < n$, Theorem 3.5.1 implies that S is linearly independent, which contradicts (▲).

3. Assume that $m = n$ and that T is a linearly independent set. We shall prove that T is a basis for W. Since T is linearly independent, we just need to show that $\text{Span}(T) = W$. Clearly, $\text{Span}(T) \subseteq W$. To show that $W \subseteq \text{Span}(T)$, let $\mathbf{u} \in W$. Suppose, for a contradiction, that $\mathbf{u} \notin \text{Span}(T)$. Consider the set of vectors $U = \{\mathbf{u}, \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m\}$. Since $\mathbf{u} \notin \text{Span}(T)$, it follows that U has $m + 1$ many vectors. Since $m + 1 = n + 1 > n$ and $\text{Span}(S) = W$, Theorem 3.5.1 implies that U is a linearly dependent set of vectors. Thus, there are scalars $a, c_1, c_2, \ldots, c_m$ (not all zero) such that

$$a\mathbf{u} + c_1\mathbf{y}_1 + c_2\mathbf{y}_2 + \cdots + c_m\mathbf{y}_m = \boldsymbol{\theta}. \tag{3.5}$$

   Because T is a linearly independent set, it follows that $a \neq 0$. Therefore, solving equation (3.5) for $\mathbf{u}$, we obtain

$$\mathbf{u} = -\frac{c_1}{a}\mathbf{y}_1 - \frac{c_2}{a}\mathbf{y}_2 - \cdots - \frac{c_m}{a}\mathbf{y}_m.$$

   Hence, $\mathbf{u} \in \text{Span}(T)$. This contradiction shows that $\text{Span}(T) = W$ and therefore, T is a basis for W.

4. Assume that $m = n$ and $\mathrm{Span}(T) = W$. We shall prove that T is a basis for W. Since $\mathrm{Span}(T) = W$, we just need to show that T is a linearly independent set. To do this, let $c_1, c_2, \ldots, c_m$ be scalars, and assume that

$$c_1\mathbf{y}_1 + c_2\mathbf{y}_2 + \cdots + c_m\mathbf{y}_m = \boldsymbol{\theta}. \tag{3.6}$$

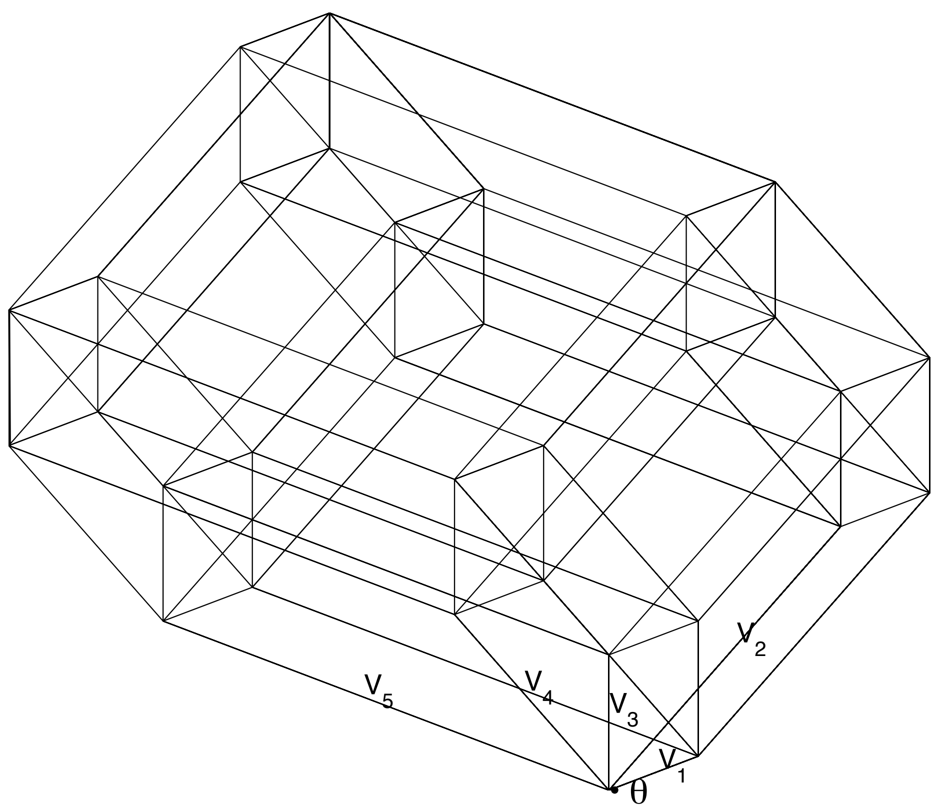
First, we will show that $c_1 = 0$. Suppose, for a contradiction, that $c_1 \neq 0$. Equation (3.6) yields

$$\mathbf{y}_1 = -\frac{c_2}{c_1}\mathbf{y}_2 - \cdots - \frac{c_m}{c_1}\mathbf{y}_m. \tag{3.7}$$

Let $T' = \{\mathbf{y}_2, \ldots, \mathbf{y}_m\}$. Thus, T' has $m-1$ many vectors, and equation (3.7) implies that $\mathrm{Span}(T) = \mathrm{Span}(T')$. Therefore, $\mathrm{Span}(T') = W$. Since $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is linearly independent and $\mathrm{Span}(T') = W$, Theorem 3.5.1 implies that $n \leq m-1$. As $n = m$, this is a contradiction. Therefore, we must have that $c_1 = 0$. A similar argument shows that $c_2 = 0$, $c_3 = 0$, ..., and $c_m = 0$. So T is a linearly independent set. Thus, T is a basis for W.

This completes the proof of the theorem. □

The vector space $\mathbb{R}^5$ has dimension 5. Using a basis $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_5\}$ for $\mathbb{R}^5$, one can construct a 5-dimensional parallelepiped (box), as illustrated in the figure below:



A 5-dimensional parallelepiped (box) that resides in the vector space $\mathbb{R}^5$

3.5.3 The Rank of a Matrix

Let A be an $m \times n$ matrix. Suppose that

$$A = [\mathbf{A}_1 \ \ \mathbf{A}_2 \ \ \cdots \ \ \mathbf{A}_n] = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_m \end{bmatrix}.$$

There are three spaces associated to the matrix A:

1. The **Null Space** of the matrix A is defined by $\mathcal{N}(A) = \{\, \mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \boldsymbol{\theta} \,\}$.

2. The **Column Space** of A is defined by $\text{Cspace}(A) = \text{Span}\{\, \mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n \,\}$.

3. The **Row Space** of A is defined by $\text{Rspace}(A) = \text{Span}\{\, \mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m \,\}$.

Definition 3.5.5. Let A be an $m \times n$ matrix. We call the dimension of the subspace $\mathcal{N}(A)$ the **nullity** of A, and we write $\text{nullity}(A) = \dim(\mathcal{N}(A))$.

Definition 3.5.6. Let A be an $m \times n$ matrix. We call the dimension of the subspace $\mathcal{R}(A)$ the **rank** of A, and we write $\text{rank}(A) = \dim(\mathcal{R}(A))$.

By the proof of Theorem 3.3.7 we know that $\text{Cspace}(A) = \mathcal{R}(A)$, the range space of A. Thus, the rank of a matrix A is also equal to the dimension of the column space of A. Hence, $\text{rank}(A) = \dim(\text{Cspace}(A))$.

Theorem 3.5.7. Let A be an $m \times n$ matrix. Then $\dim(\text{Cspace}(A)) = \dim(\text{Rspace}(A))$.

Proof. Transform the matrix A into reduced echelon form B. Theorem 3.4.6 implies that the number of leading 1's in the matrix B is the dimension of the row space of A. Theorem 3.4.9 also implies that the number of leading 1's in the matrix B is the dimension of the column space of A. Therefore, $\dim(\text{Cspace}(A)) = \dim(\text{Rspace}(A))$. ☐

Theorem 3.5.8. Let A be an $m \times n$ matrix. Then $\text{rank}(A) + \text{nullity}(A) = n$.

Proof. Transform the matrix A into reduced echelon form B. Let r be the number of leading leading 1's in the matrix B. Theorem 3.4.9 implies that r is the dimension of the column space of A. Since $\text{Cspace}(A) = \mathcal{R}(A)$, it follows that r is the rank of the matrix A. Hence, $r = \text{rank}(A)$.

Because r is the number of leading leading 1's in the matrix B, we conclude that $n - r$ is the number of free variables in the system $B\mathbf{x} = \boldsymbol{\theta}$. Theorem 3.4.5 implies that $n - r$ is the dimension of the null space of A; that is, $n - r = \text{nullity}(A)$. Therefore, $\text{rank}(A) + \text{nullity}(A) = r + (n - r) = n$. ☐

Theorem 3.5.9. Let A be a square $n \times n$ matrix. Then A is non-singular if and only if $\text{rank}(A) = n$.

Proof. This follows from Theorem 1.7.11 and the fact that $\text{rank}(A) = \dim(\text{Cspace}(A))$. ☐

Exercises 3.5

Pages 212 to 213 of text – Odds #1–27.

3.6 Orthogonal Bases for Subspaces

We first recall some definitions.

Definition. Let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ and let $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ be vectors in $\mathbb{R}^n$. The **scalar product** (or **dot product**), denoted by $\mathbf{x} \cdot \mathbf{y}$, is defined by

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

Definition. The **norm** (or length) of a vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ in $\mathbb{R}^n$ is defined by

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

Remark 3.6.1. A vector $\mathbf{x}$ has unit length if $\|\mathbf{x}\| = 1$. Note that $\|\mathbf{x}\| = 1$ if and only if $\mathbf{x} \cdot \mathbf{x} = 1$.

Definition 3.6.2. Let $\mathbf{x}$ and $\mathbf{y}$ be vectors in $\mathbb{R}^n$. We say that $\mathbf{x}$ and $\mathbf{y}$ are **orthogonal** if $\mathbf{x} \cdot \mathbf{y} = 0$.

Given a subspace W (of $\mathbb{R}^n$) with a basis $S = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$, we want to construct a new basis $T = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_k\}$ for W with the nice property: T is orthogonal, that is, $\mathbf{y}_i \cdot \mathbf{y}_j = 0$ for each pair of distinct vectors $\mathbf{y}_i$, $\mathbf{y}_j$ in T. The method used to construct T from S is called the **Gram-Schmidt Process**.

Definition 3.6.3. A set of nonzero vectors $S = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ in $\mathbb{R}^n$ is said to be **orthogonal** if any two distinct vectors $\mathbf{x}_i$, $\mathbf{x}_j$ in S are orthogonal, that is, $\mathbf{x}_i \cdot \mathbf{x}_j = 0$.

Definition 3.6.4. A set of vectors $S = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ in $\mathbb{R}^n$ is said to be **orthonormal** if S is *orthogonal* and, in addition, each vector in S has unit length, that is, $\|\mathbf{x}_i\| = 1$ for each $\mathbf{x}_i$ in S.

Definition 3.6.5. A basis $S = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ for a subspace W (of $\mathbb{R}^m$) is said to be an **orthogonal basis** if S is orthogonal.

Definition 3.6.6. A basis $S = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ for for a subspace W (of $\mathbb{R}^m$) is said to be an **orthonormal basis** if S is orthonormal.

Theorem 3.6.7. Let $S = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ be an orthogonal set of nonzero vectors in $\mathbb{R}^n$. Then S is linearly independent.

Proof. Let $S = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ be an orthogonal set of nonzero vectors in $\mathbb{R}^n$. Suppose that $c_1, c_2, \ldots, c_k$ are scalars satisfying

$$c_1 \mathbf{x}_1 + c_2 \mathbf{x}_2 + \cdots + c_k \mathbf{x}_k = \boldsymbol{\theta}.$$

We shall prove that $c_1 = c_2 = \cdots = c_k = 0$. First we prove that $c_1 = 0$. We take the dot product of $\mathbf{x}_1$ with both sides of the above equation and obtain

$$\mathbf{x}_1 \cdot (c_1 \mathbf{x}_1 + c_2 \mathbf{x}_2 + \cdots + c_k \mathbf{x}_k) = \mathbf{x}_1 \cdot \boldsymbol{\theta}.$$

Thus, by Theorem 1.6.11 we obtain

$$c_1(\mathbf{x}_1 \cdot \mathbf{x}_1) + c_2(\mathbf{x}_1 \cdot \mathbf{x}_2) + \cdots + c_k(\mathbf{x}_1 \cdot \mathbf{x}_k) = \mathbf{x}_1 \cdot \boldsymbol{\theta}.$$

Since $\mathbf{x}_1 \cdot \mathbf{x}_i = 0$ for all $i \neq 1$, we conclude that $c_1(\mathbf{x}_1 \cdot \mathbf{x}_1) = 0$ and since $\mathbf{x}_1 \cdot \mathbf{x}_1 \neq 0$ (by Theorem 1.6.11) we must have that $c_1 = 0$. A similar argument will show that $c_2 = \cdots = c_k = 0$. Therefore, S is linearly independent. □

3.6.1 Determining Coordinates

Theorem 3.6.8. Let $S = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. Then each vector $\mathbf{v}$ in W can be expressed as a linear combination of the vectors in S as follows:

$$\mathbf{v} = \left(\frac{\mathbf{v} \cdot \mathbf{x}_1}{\mathbf{x}_1 \cdot \mathbf{x}_1}\right)\mathbf{x}_1 + \left(\frac{\mathbf{v} \cdot \mathbf{x}_2}{\mathbf{x}_2 \cdot \mathbf{x}_2}\right)\mathbf{x}_2 + \cdots + \left(\frac{\mathbf{v} \cdot \mathbf{x}_k}{\mathbf{x}_k \cdot \mathbf{x}_k}\right)\mathbf{x}_k.$$

Proof. Let $S = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ be an orthogonal basis of for W. Let $\mathbf{v} \in W$. Since S is a basis for W there are scalars $c_1, c_2, \ldots, c_k$ such that

$$\mathbf{v} = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_k\mathbf{x}_k.$$

We first show that $c_1 = \frac{\mathbf{v} \cdot \mathbf{x}_1}{\mathbf{x}_1 \cdot \mathbf{x}_1}$. We take the dot product of $\mathbf{x}_1$ with both sides of the above equation and obtain

$$\mathbf{x}_1 \cdot \mathbf{v} = c_1(\mathbf{x}_1 \cdot \mathbf{x}_1) + c_2(\mathbf{x}_1 \cdot \mathbf{x}_2) + \cdots + c_k(\mathbf{x}_1 \cdot \mathbf{x}_k).$$

Since $\mathbf{x}_1 \cdot \mathbf{x}_i = 0$ for all $i \neq 1$, we conclude that $\mathbf{x}_1 \cdot \mathbf{v} = c_1(\mathbf{x}_1 \cdot \mathbf{x}_1)$. Since $\mathbf{x}_1 \cdot \mathbf{x}_1 \neq 0$, we have that $c_1 = \frac{\mathbf{v} \cdot \mathbf{x}_1}{\mathbf{x}_1 \cdot \mathbf{x}_1}$. A similar argument will show that $c_2 = \frac{\mathbf{v} \cdot \mathbf{x}_2}{\mathbf{x}_2 \cdot \mathbf{x}_2}$, …, $c_k = \frac{\mathbf{v} \cdot \mathbf{x}_k}{\mathbf{x}_k \cdot \mathbf{x}_k}$. □

3.6.2 Constructing an Orthogonal Basis

Gram-Schmidt Process. Given a basis $S = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ for a subspace W, to construct an orthogonal basis $T = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_k\}$ for W, do the following:

Step 1. Set $\mathbf{y}_1 = \mathbf{x}_1$.

Step 2. Compute the vectors $\mathbf{y}_2, \mathbf{y}_3, \ldots, \mathbf{y}_i, \ldots, \mathbf{y}_k$, in order, using the formula

$$\mathbf{y}_i = \mathbf{x}_i - \left(\frac{\mathbf{x}_i \cdot \mathbf{y}_1}{\mathbf{y}_1 \cdot \mathbf{y}_1}\right)\mathbf{y}_1 - \left(\frac{\mathbf{x}_i \cdot \mathbf{y}_2}{\mathbf{y}_2 \cdot \mathbf{y}_2}\right)\mathbf{y}_2 - \cdots - \left(\frac{\mathbf{x}_i \cdot \mathbf{y}_{i-1}}{\mathbf{y}_{i-1} \cdot \mathbf{y}_{i-1}}\right)\mathbf{y}_{i-1}.$$

The resulting set $T = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_k\}$ is an *orthogonal* basis for W.

Remark 3.6.9. If you add the next step then you will construct an orthonormal basis.

Step 3. Transform each vector $\mathbf{y}_i$, constructed in Step 2, into a unit vector; that is, let

$$\mathbf{z}_i = \frac{\mathbf{y}_i}{\|\mathbf{y}_i\|}.$$

Then $U = \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_k\}$ is an *orthonormal* basis for W.

Example 1. Use the Gram-Schmidt Process to find an orthogonal basis for the subspace W of $\mathbb{R}^4$ with basis $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$, where the vectors $\mathbf{x}_1 = [1, 1, -1, 0]$, $\mathbf{x}_2 = [0, 2, 0, 1]$, $\mathbf{x}_3 = [-1, 0, 0, 1]$ are written as row vectors.

Solution. Apply the steps in the Gram-Schmidt Process.

Step 1. Set $\mathbf{y}_1 = \mathbf{x}_1 = [1, 1, -1, 0]$.

Step 2. Compute the vectors $\mathbf{y}_2$ and $\mathbf{y}_3$, in order, using the required formula

$$\begin{aligned}
\mathbf{y}_2 &= \mathbf{x}_2 - \left(\tfrac{\mathbf{x}_2\cdot\mathbf{y}_1}{\mathbf{y}_1\cdot\mathbf{y}_1}\right)\mathbf{y}_1\\
\mathbf{y}_2 &= [0,2,0,1] - \left(\tfrac{[0,2,0,1]\cdot[1,1,-1,0]}{[1,1,-1,0]\cdot[1,1,-1,0]}\right)[1,1,-1,0]\\
\mathbf{y}_2 &= [0,2,0,1] - \left(\tfrac{2}{3}\right)[1,1,-1,0]\\
\mathbf{y}_2 &= [-\tfrac{2}{3}, \tfrac{4}{3}, \tfrac{2}{3}, 1]
\end{aligned}$$

$$\begin{aligned}
\mathbf{y}_3 &= \mathbf{x}_3 - \left(\tfrac{\mathbf{x}_3\cdot\mathbf{y}_1}{\mathbf{y}_1\cdot\mathbf{y}_1}\right)\mathbf{y}_1 - \left(\tfrac{\mathbf{x}_3\cdot\mathbf{y}_2}{\mathbf{y}_2\cdot\mathbf{y}_2}\right)\mathbf{y}_2\\
\mathbf{y}_3 &= [-1,0,0,1] - \left(\tfrac{[-1,0,0,1]\cdot[1,1,-1,0]}{[1,1,-1,0]\cdot[1,1,-1,0]}\right)[1,1,-1,0]\\
&\quad - \left(\tfrac{[-1,0,0,1]\cdot[-\frac{2}{3},\frac{4}{3},\frac{2}{3},1]}{[-\frac{2}{3},\frac{4}{3},\frac{2}{3},1]\cdot[-\frac{2}{3},\frac{4}{3},\frac{2}{3},1]}\right)[-\tfrac{2}{3}, \tfrac{4}{3}, \tfrac{2}{3}, 1]\\
\mathbf{y}_3 &= [-1,0,0,1] - \left(\tfrac{-1}{3}\right)[1,1,-1,0] - \left(\tfrac{5/3}{33/9}\right)[-\tfrac{2}{3}, \tfrac{4}{3}, \tfrac{2}{3}, 1]\\
\mathbf{y}_3 &= [-1,0,0,1] + [\tfrac{1}{3}, \tfrac{1}{3}, -\tfrac{1}{3}, 0] - [-\tfrac{10}{33}, \tfrac{20}{33}, \tfrac{10}{33}, \tfrac{5}{11}]\\
\mathbf{y}_3 &= [-\tfrac{4}{11}, -\tfrac{3}{11}, -\tfrac{7}{11}, \tfrac{6}{11}].
\end{aligned}$$

The set $T = \{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3\}$ is the desired *orthogonal* basis for W.

Exercises 3.6

Pages 224 to 225 of text – Odds #1-7, #13, 15.

3.7 Linear Transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$

We recall that a function $F\colon A \to B$ associates with each element z in the set A an element $F(z)$ in the set B. When $F\colon A \to B$ we say that F maps A into B. When we wish to indicate the geometric nature of the sets A and B, we sometimes call F a transformation rather than a function. Mathematically the concepts of "function" and "transformation" are identical. A set A may be carried into a set B by a function, while a space A is transformed into a space B by a transformation.

Our interest in this section centers on linear transformations from one vector space V into another vector space W (sometimes W and V will be the same space).

Definition 3.7.1. If $T\colon V \to W$ is a function from a vector space V to the vector space W, then T is called a linear transformation if, for all vectors $\mathbf{x}$ and $\mathbf{y}$ in V and for all scalars c, the following hold:

(a) $T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$

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

The vector space V is called the domain of T and the vector space W is called the co-domain of T.

Suppose that $T\colon V \to W$ is a linear transformation and that $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis for V. Since any vector $\mathbf{y}$ in V can be expressed as a linear combination of the vectors in S, we have that

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

for some scalars $c_1, \ldots, c_n$. By repeating properties (a) and (b) above, we obtain the equation

$$T(\mathbf{y}) = c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2) + \cdots + c_nT(\mathbf{v}_n).$$

Therefore, we can make the important observation that if you know how the linear transformation T acts on a set of basis vectors for V, then you know how T acts on all the vectors in V.

Theorem 3.7.2. Suppose that $T\colon V \to W$ is a linear transformation and suppose that $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis for V. For any vector $\mathbf{y}$ in V, the value $T(\mathbf{y})$ is completely determined by vectors in $\{T(\mathbf{v}_1), T(\mathbf{v}_2), \ldots, T(\mathbf{v}_n)\}$.

Problem 1. Suppose $T\colon \mathbb{R}^3 \to \mathbb{R}^2$ is a linear transformation and let S be the standard basis $S = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ where

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Given that

$$T(\mathbf{e}_1) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad T(\mathbf{e}_2) = \begin{bmatrix} 3 \\ 0 \end{bmatrix}, \quad T(\mathbf{e}_3) = \begin{bmatrix} 4 \\ -7 \end{bmatrix}$$

determine $T(\mathbf{x})$ when $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$.

Solution. Since

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_n$$

we see that

$$T(\mathbf{x}) = x_1T(\mathbf{e}_1) + x_2T(\mathbf{e}_2) + x_3T(\mathbf{e}_3).$$

Thus,

$$T(\mathbf{x}) = x_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 3 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 4 \\ -7 \end{bmatrix} = \begin{bmatrix} x_1 + 3x_2 + 4x_3 \\ x_1 - 7x_3 \end{bmatrix}.$$

Theorem 3.7.3. Let V be the vector space $\mathbb{R}^n$ and let A be an $m \times n$ matrix. Consider the function $T\colon \mathbb{R}^n \to \mathbb{R}^m$ defined by $T(\mathbf{x}) = A\mathbf{x}$. Then T is a linear transformation.

Proof. We show that T satisfies properties (a) and (b) of Definition 3.7.1. First we prove that (a) holds:

$$\begin{aligned} T(\mathbf{x} + \mathbf{y}) &= A(\mathbf{x} + \mathbf{y}) && \text{by definition of } T \\ &= A\mathbf{x} + A\mathbf{y} && \text{by distribution of matrix mult.} \\ &= T(\mathbf{x}) + T(\mathbf{y}) && \text{by definition of } T. \end{aligned}$$

Now we prove that (b) holds:

$$\begin{aligned} T(c\mathbf{x}) &= A(c\mathbf{x}) && \text{by definition of } T \\ &= cA\mathbf{x} && \text{by scalar mult. property} \\ &= cT(\mathbf{x}) && \text{by definition of } T. \end{aligned}$$

This completes the proof. □

Theorem 3.7.4. If $T: V \to W$ is a linear transformation, then

1. $T(\boldsymbol{\theta}) = \boldsymbol{\theta}$
2. $T(-\mathbf{x}) = -T(\mathbf{x})$
3. $T(\mathbf{x} - \mathbf{y}) = T(\mathbf{x}) - T(\mathbf{y})$

Proof. Since $0\boldsymbol{\theta} = \boldsymbol{\theta}$, it follows that $T(0\boldsymbol{\theta}) = T(\boldsymbol{\theta})$. Because T is linear, we must have that $0T(\boldsymbol{\theta}) = T(\boldsymbol{\theta})$. Clearly, $0T(\boldsymbol{\theta}) = \boldsymbol{\theta}$ and therefore, $T(\boldsymbol{\theta}) = \boldsymbol{\theta}$. For item 2, let $\mathbf{x} \in V$. We have that $-\mathbf{x} = (-1)\mathbf{x}$. Thus, $T(-\mathbf{x}) = T((-1)\mathbf{x}) = (-1)T(\mathbf{x}) = -T(\mathbf{x})$. Therefore, $T(-\mathbf{x}) = -T(\mathbf{x})$. The proof of item 3 is similar. ☐

3.7.1 Examples of Linear Transformations

Example 2. Let V be the vector space $\mathbb{R}^2$ and let A be the 2×2 matrix $A = \begin{bmatrix} 0 & 2 \\ 1 & 0 \end{bmatrix}$. Define the linear transformation $L: \mathbb{R}^2 \to \mathbb{R}^2$ by $L(\mathbf{x}) = A\mathbf{x}$. Let $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ be the unit vectors in $\mathbb{R}^2$. Evaluating $L(\mathbf{e}_1)$ and $L(\mathbf{e}_2)$ we obtain

$$L(\mathbf{e}_1) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \; L(\mathbf{e}_2) = \begin{bmatrix} 0 \\ 2 \end{bmatrix}.$$

Consequently, the unit square S is "transformed" by L to $L[S]$ as shown in Figure 3.8.

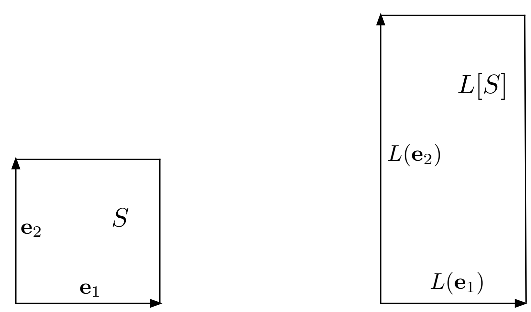


Figure 3.8: L transforms the square S into the rectangle $L[S]$

Example 3. Let V be the vector space $\mathbb{R}^2$ and let B be the 2×2 matrix $B = \frac{\sqrt{2}}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. Define the linear transformation $F: \mathbb{R}^2 \to \mathbb{R}^2$ by $F(\mathbf{x}) = B\mathbf{x}$. Let $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ be the unit vectors in $\mathbb{R}^2$. Evaluating $F(\mathbf{e}_1)$ and $F(\mathbf{e}_2)$ we obtain

$$F(\mathbf{e}_1) = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} \; F(\mathbf{e}_2) = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix}.$$

Consequently, the unit square S is "transformed" by F to $F[S]$ as shown in Figure 3.9.

Example 4. Let V be the vector space $\mathbb{R}^2$ and let C be the 2×2 matrix $C = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. Define the linear transformation $G: \mathbb{R}^2 \to \mathbb{R}^2$ by $G(\mathbf{x}) = C\mathbf{x}$. Let $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ be the unit

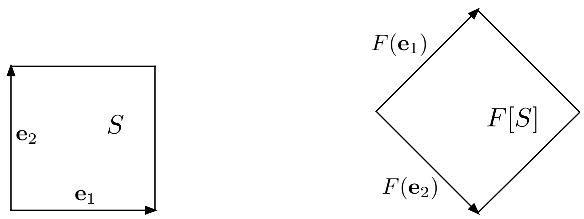


Figure 3.9: F transforms the square S into the rotated square $F[S]$

vectors in $\mathbb{R}^2$. Evaluating $G(\mathbf{e}_1)$ and $G(\mathbf{e}_2)$ we obtain

$$G(\mathbf{e}_1) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \; G(\mathbf{e}_2) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Consequently, the unit square S is "transformed" by G to $G[S]$ as shown in Figure 3.10.

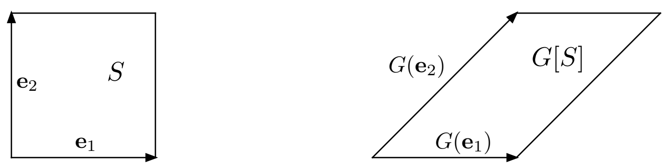


Figure 3.10: G transforms the square S into the parallelogram $G[S]$

Example 5. Let V be the vector space $\mathbb{R}^3$ and let A be the 3×3 matrix

$$A = \begin{bmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix}.$$

Let $T\colon \mathbb{R}^3 \to \mathbb{R}^3$ be defined by $T(\mathbf{x}) = A\mathbf{x}$. Letting $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ and using matrix multiplication we can get a "formula" for T:

$$T(\mathbf{x}) = A\mathbf{x} = \begin{bmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -2x_3 \\ x_1 + 2x_2 + x_3 \\ x_1 + 3x_3 \end{bmatrix}.$$

Thus we obtain the formula

$$T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \right) = \begin{bmatrix} -2x_3 \\ x_1 + 2x_2 + x_3 \\ x_1 + 3x_3 \end{bmatrix}.$$

Conversely, given a formula (as in the above example) for a linear transformation, we can obtain a matrix representation (namely A) such that $T(\mathbf{x}) = A\mathbf{x}$.

Problem 6. Let $T\colon \mathbb{R}^3 \to \mathbb{R}^4$ be the linear transformation defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \right) = \begin{bmatrix} -2x_3 \\ x_1 + x_2 \\ x_1 + 2x_2 + x_3 \\ x_1 + 3x_3 \end{bmatrix}.$$

Find a matrix A such that $T(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^3$.

Solution. We shall find the desired matrix as follows:

$$T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} -2x_3 \\ x_1 + x_2 \\ x_1 + 2x_2 + x_3 \\ x_1 + 3x_3 \end{bmatrix} = x_1 \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ 2 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -2 \\ 0 \\ 1 \\ 3 \end{bmatrix}$$

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

Thus, $A = \begin{bmatrix} 0 & 0 & -2 \\ 1 & 1 & 0 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix}$ is the desired matrix.

3.7.2 Finding the Matrix of a Linear Transformation

Theorem 3.7.5. Let $T\colon \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and let $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ be the unit vectors in $\mathbb{R}^n$. Define the matrix $A = [T(\mathbf{e}_1)\ \ T(\mathbf{e}_2)\ \ \cdots\ \ T(\mathbf{e}_n)]$. Then $T(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$.

Proof. Let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$. We will show that $T(\mathbf{x}) = A\mathbf{x}$. Observe that $\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n$. Since T is a linear transformation, we obtain

$$T(\mathbf{x}) = x_1T(\mathbf{e}_1) + x_2T(\mathbf{e}_2) + \cdots + x_nT(\mathbf{e}_n) = [T(\mathbf{e}_1)\ \ T(\mathbf{e}_2)\ \ \cdots\ \ T(\mathbf{e}_n)] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = A\mathbf{x}$$

by Theorem 1.5.12, where $A = [T(\mathbf{e}_1)\ \ T(\mathbf{e}_2)\ \ \cdots\ \ T(\mathbf{e}_n)]$. □

Problem 7 (see problem 6 on page 87)**.** Let $T\colon \mathbb{R}^3 \to \mathbb{R}^4$ be the linear transformation defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} -2x_3 \\ x_1 + x_2 \\ x_1 + 2x_2 + x_3 \\ x_1 + 3x_3 \end{bmatrix}.$$

Find a matrix A such that $T(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^3$.

Solution. Let $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ be the unit vectors in $\mathbb{R}^3$. Thus, by Theorem 3.7.5 the matrix A is

$$A = [T(\mathbf{e}_1)\ \ T(\mathbf{e}_2)\ \ T(\mathbf{e}_3)] = \begin{bmatrix} 0 & 0 & -2 \\ 1 & 1 & 0 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix}.$$

3.7.3 Coordinates with Respect to a Basis

Let V be a vector space and let $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis for V. By Theorem 3.4.4, we have that every vector $\mathbf{y}$ in V there is a unique list of scalars $c_1, c_2, \ldots, c_n$ such that

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

Thus, using the given order of the vectors in B, we can call the scalars $c_1, c_2, \ldots, c_n$ **the coordinates of y with respect to the basis** B, and call

$$[\mathbf{y}]_B = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$$

the **coordinate vector of y with respect to the basis** B. Using this coordinate system for V, we can represent any linear transformation $T\colon V \to V$ in terms of a matrix.

Theorem 3.7.6. Let $T\colon V \to V$ be a linear transformation and $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis for V. Define the matrix $A = [\,[T(\mathbf{v}_1)]_B\ [T(\mathbf{v}_2)]_B\ \cdots\ [T(\mathbf{v}_n)]_B\,]$. Then $[T(\mathbf{x})]_B = A[\mathbf{x}]_B$ for all $\mathbf{x} \in V$.

There is a slight generalization of Theorem 3.7.6.

Theorem 3.7.7. Let $T\colon V \to W$ be a linear transformation. Let $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis for V and $C = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$ be a basis for W. If $A = [\,[T(\mathbf{v}_1)]_C\ [T(\mathbf{v}_2)]_C\ \cdots\ [T(\mathbf{v}_n)]_C\,]$, then $[T(\mathbf{x})]_C = A[\mathbf{x}]_B$ for all $\mathbf{x} \in V$.

3.7.4 Null Space and Range Space of a Linear Transformation

Definition 3.7.8. Let $T\colon V \to W$ be a linear transformation.

- The null space of T (or the kernel of T), denoted by $\mathcal{N}(T)$, is the set of all vectors $\mathbf{x}$ in V such that $T(\mathbf{x}) = \boldsymbol{\theta}$, that is,

  $$\mathcal{N}(T) = \{\mathbf{x} \in V : T(\mathbf{x}) = \boldsymbol{\theta}\}.$$

- The range space of T, denoted by $\mathcal{R}(T)$, is set of all vectors $\mathbf{y}$ in W such that $T(\mathbf{x}) = \mathbf{y}$ for some $\mathbf{x} \in V$, that is,

  $$\mathcal{R}(T) = \{\mathbf{y} \in W : T(\mathbf{x}) = \mathbf{y} \text{ for some } \mathbf{x} \in V\}.$$

Theorem 3.7.9. Let V be the vector space $\mathbb{R}^n$ and let A be an $m \times n$ matrix. Consider the linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^m$ defined by $T(\mathbf{x}) = A\mathbf{x}$. Then $\mathcal{N}(T) = \mathcal{N}(A)$ and $\mathcal{R}(T) = \mathcal{R}(A)$.

Definition 3.7.10. Let $T\colon V \to W$ be a linear transformation. The dimension of the subspace $\mathcal{N}(T)$ is called the **nullity** of T. The dimension of the subspace $\mathcal{R}(T)$ is called the **rank** of T.

Problem 8. Consider the linear transformation $T\colon \mathbb{R}^3 \to \mathbb{R}^4$ defined by $T(\mathbf{x}) = A\mathbf{x}$ where

$$A = \begin{bmatrix} 2 & 0 & -2 \\ 1 & 1 & 0 \\ 1 & 2 & 1 \\ 1 & 0 & -1 \end{bmatrix}.$$

Find a basis for the null space and range space of T. Then determine the nullity and rank of T.

Solution. By Theorem 3.7.9 we just need to find the null space of the matrix A and find the range of A. First we find the null space of A. Applying Gauss-Jordon reduction to the augmented matrix $[A \,|\, \boldsymbol{\theta}]$, we obtain the augmented matrix

$$[C \,|\, \boldsymbol{\theta}] = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Solving the equivalent system of equations $C\mathbf{x} = \boldsymbol{\theta}$, we obtain

$$\begin{aligned} x_1 &= x_3 \\ x_2 &= -x_3 \\ x_3 &= x_3. \end{aligned}$$

Therefore, every solution to $A\mathbf{x} = \boldsymbol{\theta}$ can be written in the form

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

So the vector $\mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$ forms the basis of the null space of A. We conclude that $\mathbf{v}_1$ forms a basis for the null space of T and that the nullity of T is 1.

Now we find a basis for the range of A. Since the column space of A equals the range space of A, we just need to find a basis for the column space of A. To do this, we apply Theorem 3.4.9. Transforming the matrix A into reduced echelon form we obtain

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Theorem 3.4.9 implies that the vectors $\mathbf{y}_1 = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ and $\mathbf{y}_2 = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 0 \end{bmatrix}$ form a basis for the range of A. We conclude that $\mathbf{y}_1$, $\mathbf{y}_2$ form a basis for the range space of T and that the rank of T is 2.

3.7.5 One-to-One Linear Transformations

Definition 3.7.11. A linear transformation $T\colon V \to W$ is said to be one-to-one if the following holds: For any vectors $\mathbf{u}$ and $\mathbf{v}$ in V, if $T(\mathbf{u}) = T(\mathbf{v})$ then $\mathbf{u} = \mathbf{v}$.

Theorem 3.7.12. A linear transformation $T\colon V \to W$ is one-to-one if and only if $\mathcal{N}(T) = \{\boldsymbol{\theta}\}$.

Proof. Let $T\colon V \to W$ be a linear transformation. Assume that T is one-to-one. We will prove that $\mathcal{N}(T) = \{\boldsymbol{\theta}\}$. To do this, let $\mathbf{x} \in \mathcal{N}(T)$. Thus, $T(\mathbf{x}) = \boldsymbol{\theta}$. Since $T(\boldsymbol{\theta}) = \boldsymbol{\theta}$ by Theorem 3.7.4, we conclude that $T(\mathbf{x}) = T(\boldsymbol{\theta})$. Because T is one-to-one, we infer that $\mathbf{x} = \boldsymbol{\theta}$. Therefore, $\mathcal{N}(T) = \{\boldsymbol{\theta}\}$. For the converse, assume that $\mathcal{N}(T) = \{\boldsymbol{\theta}\}$. We shall prove that T is one-to-one. Let $\mathbf{u}$ and $\mathbf{v}$ be in V and assume that $T(\mathbf{u}) = T(\mathbf{v})$. Thus, $T(\mathbf{u}) - T(\mathbf{v}) = \boldsymbol{\theta}$. Because T is linear, we have that $T(\mathbf{u} - \mathbf{v}) = \boldsymbol{\theta}$. Thus, $\mathbf{u} - \mathbf{v}$ is in $\mathcal{N}(T)$. Since $\mathcal{N}(T) = \{\boldsymbol{\theta}\}$, this implies that $\mathbf{u} - \mathbf{v} = \boldsymbol{\theta}$. Therefore, $\mathbf{u} = \mathbf{v}$ and thus, T is one-to-one. □

Corollary 3.7.13. Let $T\colon V \to W$ be a linear transformation. Then T is one-to-one if and only if for all $\mathbf{x} \in V$, if $T(\mathbf{x}) = \boldsymbol{\theta}$, then $\mathbf{x} = \boldsymbol{\theta}$.

Lemma 3.7.14. Let $T\colon V \to W$ be a one-to-one linear transformation. If $\mathbf{x}_1, \mathbf{x}_2, \dots \mathbf{x}_k$ are linear independent vectors in V, then $T(\mathbf{x}_1), T(\mathbf{x}_2), \dots T(\mathbf{x}_k)$ are linear independent vectors in W.

Proof. Let $T\colon V \to W$ be a one-to-one linear transformation. Assume that $\mathbf{x}_1, \mathbf{x}_2, \dots \mathbf{x}_k$ are linear independent vectors in V. Suppose that

$$c_1 T(\mathbf{x}_1) + c_2 T(\mathbf{x}_2) + \cdots c_k T(\mathbf{x}_k) = \boldsymbol{\theta}.$$

As T is a linear transformation, we conclude that

$$T(c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots c_k\mathbf{x}_k) = \boldsymbol{\theta}.$$

Since T is one-to-one, Corollary 3.7.13 implies that

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots c_k\mathbf{x}_k = \boldsymbol{\theta}.$$

Because $\mathbf{x}_1, \mathbf{x}_2, \dots \mathbf{x}_k$ are linearly independent, we infer that $c_1 = c_2 = \cdots c_k = 0$. Therefore, the vectors $T(\mathbf{x}_1), T(\mathbf{x}_2), \dots T(\mathbf{x}_k)$ are linear independent. □

Theorem 3.7.15. Let $T\colon V \to W$ be a linear transformation such that $\dim(V) = \dim(W)$. If T is one-to-one, then T is onto W.

Proof. Let $T\colon V \to W$ be a linear transformation such that $\dim(V) = \dim(W) = n$. Assume that T is one-to-one. Let $\{x_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ be a basis for V. By Lemma 3.7.14, we conclude that $T(\mathbf{x}_1), T(\mathbf{x}_2), \dots T(\mathbf{x}_n)$ are linear independent vectors in W. Theorem 3.5.4(3) implies that $\{T(\mathbf{x}_1), T(\mathbf{x}_2), \dots T(\mathbf{x}_n)\}$ is a basis for W. To show that T is onto W, let $\mathbf{y} \in W$. Since $\{T(\mathbf{x}_1), T(\mathbf{x}_2), \dots T(\mathbf{x}_n)\}$ is a basis for W, there are scalars $c_1, c_2, \dots, c_n$ such that

$$c_1 T(\mathbf{x}_1) + c_2 T(\mathbf{x}_2) + \cdots c_n T(\mathbf{x}_n) = \mathbf{y}.$$

Because T is a linear transformation, we see that

$$T(c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots c_n\mathbf{x}_n) = \mathbf{y}.$$

Since $c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots c_n\mathbf{x}_n$ is in V, we conclude that T is onto W. □

Exercises 3.7

Pages 239 to 241 of text – # 2, 3, 4, 8, 11, 14, 19, 22, 24, 25, 26, 28.

Chapter 4

The Eigenvalue Problem

A scalar λ is said to be an eigenvalue for square matrix A if there is a nonzero vector $\mathbf{v}$ such that $A\mathbf{v} = \lambda\mathbf{v}$. Such a vector $\mathbf{v}$ is called an eigenvector. The key method for finding eigenvalues is the use of determinants. The determinate of a square matrix A is a real number, denoted by $\det(A)$, that is obtained by an involved computation using the matrix A. Throughout this chapter we will be working with square matrices.

4.2 Determinants

Definition 4.2.1. Let

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

Then $\det(A) = a_{11}a_{22} - a_{21}a_{12}$.

Problem 1. Find the determinate of the 2×2 matrix

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

Solution. The determinant is given by $\det(A) = 1 \cdot (-1) - 3 \cdot 4 = -13$.

4.2.1 Minors and Cofactors

In order to define the determinate of $n \times n$ matrices when $n > 2$, we shall introduce a procedure for deleting a particular row and a particular column from a matrix.

Definition 4.2.2. Let

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

be an $n \times n$ matrix. Let **minor matrix** M_{ij} is the $(n-1) \times (n-1)$ matrix obtained by deleting the i-th row and j-th column of A.

Problem 2. Consider the 4×4 matrix

$$A = \begin{bmatrix} 1 & 2 & 1 & 3 \\ 2 & 1 & 2 & 1 \\ 4 & 2 & 0 & -1 \\ -2 & 3 & 1 & 1 \end{bmatrix}.$$

Find the minor matrices M_{21}, M_{22}, M_{23} and M_{24}.

Solution. M_{21} is obtained from A be deleting the 2nd row and the 1st column of A. So,

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

Similarly, M_{22} is obtained from A be deleting the 2nd row and the 2nd column of A. M_{23} is obtained from A be deleting the 2nd row and the 3rd column of A. M_{24} is obtained from A be deleting the 2nd row and the 4th column of A. Thus,

$$M_{22} = \begin{bmatrix} 1 & 1 & 3 \\ 4 & 0 & -1 \\ -2 & 1 & 1 \end{bmatrix}, \quad M_{23} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 2 & -1 \\ -2 & 3 & 1 \end{bmatrix}, \quad M_{24} = \begin{bmatrix} 1 & 2 & 1 \\ 4 & 2 & 0 \\ -2 & 3 & 1 \end{bmatrix}.$$

Definition 4.2.3. Let

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

be an $n \times n$ matrix. Let M_{ij} be the $(n-1) \times (n-1)$ minor matrix obtained by deleting the i-th row and j-th column of A. The **cofactor** A_{ij} is defined to be

$$A_{ij} = (-1)^{i+j} \det(M_{ij}).$$

Problem 3. Let $A = \begin{bmatrix} 2 & 1 & 1 \\ -1 & 1 & 2 \\ 1 & 2 & 3 \end{bmatrix}$. Evaluate A_{23}.

Solution. Since $A_{23} = (-1)^{2+3} \det(M_{23})$, we first obtain the matrix M_{23} by deleting the 2nd row and the 3rd column of A to obtain

$$M_{23} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

We see that $\det(M_{23}) = 2 \cdot 2 - 1 \cdot 1 = 3$. Therefore, $A_{23} = (-1)^{2+3} \det(M_{23}) = -3$.

Definition 4.2.4. Let

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

be an $n \times n$ matrix. For any row $R_i = \begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{bmatrix}$, the determinant of A is the number $\det(A)$ defined by

$$\det(A) = a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in}.$$

Remark 4.2.5. No matter which row you select in Definition 4.2.4, you will always get the same value for $\det(A)$. The above definition of the $\det(A)$ is referred to as the i-th row expansion. The text defines $\det(A)$ using the method called the j-th column expansion (see Theorem 4.2.6 below).

Problem 4. Let $A = \begin{bmatrix} 2 & 1 & 1 \\ -1 & 1 & 2 \\ 1 & 2 & 3 \end{bmatrix}$. Select a row and use Definition 4.2.4 to compute $\det(A)$.

Solution. We shall use the first row of A. Thus,

$$\begin{aligned} \det(A) &= a_{11}A_{11} + a_{12}A_{12} + a_{13}A_{13} \\ &= 2 \cdot A_{11} + 1 \cdot A_{12} + 1 \cdot A_{13} \qquad (\star) \end{aligned}$$

where the cofactors A_{11}, A_{12}, A_{13} are computed as follows:

$$\begin{aligned} A_{11} &= (-1)^{1+1} \det\left(\begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}\right) = (1)[1 \cdot 3 - 2 \cdot 2] = -1 \\ A_{12} &= (-1)^{1+2} \det\left(\begin{bmatrix} -1 & 2 \\ 1 & 3 \end{bmatrix}\right) = (-1)[(-1) \cdot 3 - 2 \cdot 1] = 5 \\ A_{13} &= (-1)^{1+3} \det\left(\begin{bmatrix} -1 & 1 \\ 1 & 2 \end{bmatrix}\right) = (1)[(-1) \cdot 2 - 1 \cdot 1] = -3. \end{aligned}$$

Therefore, by $(\star)$, $\det(A) = 2 \cdot (-1) + 1 \cdot 5 + 1 \cdot (-3) = 0$.

Problem 5. Let $A = \begin{bmatrix} 1 & 2 & -1 & 1 \\ -1 & 0 & 2 & -2 \\ 3 & -1 & 1 & 1 \\ 2 & 0 & -1 & 2 \end{bmatrix}$. Select a row and compute $\det(A)$.

Solution. We shall use the second row of A because it has a 0. Thus,

$$\begin{aligned} \det(A) &= a_{21}A_{21} + a_{22}A_{22} + a_{23}A_{23} + a_{24}A_{24} \\ &= -1 \cdot A_{21} + 0 \cdot A_{22} + 2 \cdot A_{23} - 2 \cdot A_{24} \\ &= -1 \cdot A_{21} + 2 \cdot A_{23} - 2 \cdot A_{24} \\ &= -1 \cdot (-1)^{2+1} \det(M_{21}) + 2 \cdot (-1)^{2+3} \det(M_{23}) - 2 \cdot (-1)^{2+4} \det(M_{24}) \\ &= \det(M_{21}) - 2\det(M_{23}) - 2\det(M_{24}) \\ &= 5 - 2(-8) - 2(9) = 3 \end{aligned}$$

where the values $\det(M_{21}) = 5$, $\det(M_{23}) = -8$ and $\det(M_{24}) = 9$ are computed as follows: To compute $\det(M_{21})$ we first identify M_{21} and obtain

$$M_{21} = \begin{bmatrix} 2 & -1 & 1 \\ -1 & 1 & 1 \\ 0 & -1 & 2 \end{bmatrix}.$$

We now evaluate $\det(M_{21})$, using the 3rd row of M_{21}, and get

$$\begin{aligned} \det(M_{21}) &= (-1)(-1)^{3+2} \det\left(\begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix}\right) + (2)(-1)^{3+3} \det\left(\begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}\right) \\ &= 1(2 \cdot 1 - (-1)(1)) + 2(2 \cdot 1 - (-1)(-1)) = 3 + 2 = 5. \end{aligned}$$

Because

$$M_{23} = \begin{bmatrix} 1 & 2 & 1 \\ 3 & -1 & 1 \\ 2 & 0 & 2 \end{bmatrix},$$

using the 3rd row in M_{23}, we obtain

$$\begin{aligned}\det(M_{23}) &= (2)(-1)^{3+1}\det\left(\begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix}\right) + (2)(-1)^{3+3}\det\left(\begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix}\right) \\ &= 2(2+1) + 2(-1-6)) = -8.\end{aligned}$$

Finally, since

$$M_{24} = \begin{bmatrix} 1 & 2 & -1 \\ 3 & -1 & 1 \\ 2 & 0 & -1 \end{bmatrix}$$

we use the 3rd row in M_{24} to evaluate

$$\begin{aligned}\det(M_{24}) &= (2)(-1)^{3+1}\det\left(\begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}\right) + (-1)(-1)^{3+3}\det\left(\begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix}\right) \\ &= 2(2-1) + (-1)(-1-6)) = 9.\end{aligned}$$

Theorem 4.2.6. Let

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

be an $n \times n$ matrix. Then

$$\begin{aligned}\det(A) &= a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in} && i\text{-th row expansion} \\ &= a_{1j}A_{1j} + a_{2j}A_{2j} + \cdots + a_{nj}A_{nj} && j\text{-th column expansion.}\end{aligned}$$

Thus, one can compute the determinate of a square matrix by selecting a row, or a column, with the most zeros for the evaluation of this determinate. Recall the following definition of a singular matrix.

Definition 4.2.7. Let A be a square $n \times n$ matrix. The matrix A is **nonsingular** if the only solution to $A\mathbf{x} = \boldsymbol{\theta}$ is $\mathbf{x} = \boldsymbol{\theta}$. We say that A **singular** if there is a non-trivial solution to $A\mathbf{x} = \boldsymbol{\theta}$, that is, a solution $\mathbf{x} \neq \boldsymbol{\theta}$.

The next two results are very important in linear algebra.

Theorem 4.2.8. Let A and B be $n \times n$ matrices. Then $\det(AB) = \det(A)\det(B)$.

Theorem 4.2.9. Let A be an $n \times n$ matrix. Then A is singular if and only if $\det(A) = 0$.

4.2.2 Triangular Matrices

Definition 4.2.10. A square $n \times n$ matrix A is **upper triangular** if all of the entries below the main diagonal are 0; that is, if A has the form

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}.$$

A square matrix A is **lower triangular** if all of the entries above the main diagonal are 0. A matrix is said to be **triangular** if it is either upper or lower triangular.

Theorem 4.2.11. Let A be a triangular $n \times n$ matrix. Then $\det(A)$ is just the product of the diagonal entries of the matrix A; that is, $\det(A) = a_{11}a_{22}\cdots a_{nn}$.

Problem 6. Evaluate the determinate of each of the following triangular matrices:

$$A = \begin{bmatrix} 2 & 1 & -3 & 5 \\ 0 & -4 & 9 & 25 \\ 0 & 0 & -2 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 4 & 2 & 0 & 0 \\ -5 & 11 & -2 & 0 \\ 10 & 7 & 9 & 3 \end{bmatrix}$$

Solution. $\det(A) = 2 \cdot (-4) \cdot (-2) \cdot 1 = 16$ and $\det(B) = 1 \cdot 2 \cdot (-2) \cdot 3 = -12$.

Problem 7. Let $A = \begin{bmatrix} 2 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 6 & 2 & 0 \\ 0 & 1 & 0 & 3 \end{bmatrix}$. Select a row R_i and use Definition 4.2.4 to evaluate $\det(A)$.

Solution. We shall use the second row because it has a lot of 0's. Thus,

$$\det(A) = a_{21}A_{21} + a_{22}A_{22} + a_{23}A_{23} + a_{24}A_{24} = 1 \cdot A_{22}.$$

So we just need to evaluate the cofactor A_{22}, obtaining

$$A_{22} = (-1)^{2+2} \det\left(\begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}\right) = (1)[12] = 12.$$

Hence, $\det(A) = 12$.

Exercises 4.2

Pages 288 to 289 of text – Odds #9-17.

4.4 & 4.5 Eigenvalues, Eigenvectors, and Eigenspaces

Definition 4.4.12. If A is a square $n \times n$ matrix, then a nonzero vector $\mathbf{x}$ in $\mathbb{R}^n$ is called an eigenvector of A if $A\mathbf{x}$ is a scalar multiple of $\mathbf{x}$, that is, $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ. The scalar λ is called an *eigenvalue* of A, and the nonzero vector $\mathbf{x}$ is said to be an *eigenvector* corresponding to the eigenvalue λ.

Remark 4.4.13. We make some observations about the definition of an eigenvalue and eigenvector.

1. Notice that $\mathbf{x} = \boldsymbol{\theta}$ is not allowed to be an eigenvector, even though $A\boldsymbol{\theta} = \boldsymbol{\theta}$.
2. Notice that if $\mathbf{x}$ is a nonzero vector in the null space of A, that is $A\mathbf{x} = \boldsymbol{\theta}$, then $\mathbf{x}$ is an eigenvector with eigenvalue (the number) 0, because $A\mathbf{x} = \boldsymbol{\theta} = 0\mathbf{x}$.
3. Not all matrices have an eigenvector.

Definition 4.4.14. Let A is a square $n \times n$ matrix. Let λ be a fixed eigenvalue of A. The *eigenspace* corresponding to λ is defined to be the set $E_\lambda = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \lambda\mathbf{x}\}$.

The reader will observe that if $\mathbf{x}$ is an eigenvector corresponding to λ, then any scalar multiple $c\mathbf{x}$ is also an eigenvector. This is a special case of the next result.

Theorem 4.4.15. Let A is a square $n \times n$ matrix. Let λ be a fixed eigenvalue of A. Then E_λ is a subspace of $\mathbb{R}^n$.

Proof. Let A is a square $n \times n$ matrix. Let λ be a fixed eigenvalue of A. Note that $\boldsymbol{\theta} \in E_\lambda$ since $A\boldsymbol{\theta} = \boldsymbol{\theta}$ and thus, $A\boldsymbol{\theta} = \lambda\boldsymbol{\theta}$. So, to show that E_λ is a subspace it is sufficient to show that E_λ is closed under addition and scalar multiplication. To do this let $\mathbf{x}$ and $\mathbf{y}$ be in E_λ and let c be a scalar. Since $\mathbf{x}$ and $\mathbf{y}$ are in E_λ, it follows that

$$A\mathbf{x} = \lambda\mathbf{x} \text{ and } A\mathbf{y} = \lambda\mathbf{y}. \tag{4.1}$$

We must show that $\mathbf{x} + \mathbf{y}$ is in E_λ; that is, we must show that $A(\mathbf{x} + \mathbf{y}) = \lambda(\mathbf{x} + \mathbf{y})$. To see this, note that

$$\begin{aligned} A(\mathbf{x} + \mathbf{y}) &= A\mathbf{x} + A\mathbf{y} && \text{by distribution of matrix mult.} \\ &= \lambda\mathbf{x} + \lambda\mathbf{y} && \text{by equations in (4.1) above} \\ &= \lambda(\mathbf{x} + \mathbf{y}) && \text{by distribution of scalar mult.} \end{aligned}$$

Hence, $\mathbf{x} + \mathbf{y} \in E_\lambda$. Now we must show that $c\mathbf{x}$ is in E_λ; that is, we must show that $A(c\mathbf{x}) = \lambda(c\mathbf{x})$ given that $\mathbf{x}$ is in E_λ. To see this, note that

$$\begin{aligned} A(c\mathbf{x}) &= cA\mathbf{x} && \text{by property of matrix mult.} \\ &= c\lambda\mathbf{x} && \text{by first equation in (4.1) above} \\ &= \lambda c\mathbf{x} && \text{since } c\lambda = \lambda c. \end{aligned}$$

Hence, $c\mathbf{x} \in E_\lambda$. Therefore, by Theorem 3.2.3, E_λ is a subspace of $\mathbb{R}^n$. □

Theorem 4.4.16. Let A is a square $n \times n$ matrix and let λ be a fixed eigenvalue of A. Then $E_\lambda = \mathcal{N}(A - \lambda I)$; that is, the eigenspace E_λ is the null space of the matrix $A - \lambda I$.

Proof. We are given that λ is an eigenvalue of the square $n \times n$ matrix. Let $\mathbf{x} \in \mathbb{R}^n$. Suppose that $\mathbf{x}$ is in eigenspace E_λ. Then $A\mathbf{x} = \lambda\mathbf{x}$ and so,

$$\begin{aligned} A\mathbf{x} &= \lambda\mathbf{x} \\ A\mathbf{x} &= \lambda I\mathbf{x} && \text{because } I\mathbf{x} = \mathbf{x} \text{ where } I \text{ is the identity matrix} \\ A\mathbf{x} - \lambda I\mathbf{x} &= \boldsymbol{\theta} && \text{subtracting } \lambda I\mathbf{x} \text{ from both sides} \\ (A - \lambda I)\mathbf{x} &= \boldsymbol{\theta} && \text{distribution prop. of matrix mult.} \end{aligned}$$

We conclude that $\mathbf{x}$ is in the null space of the matrix $A - \lambda I$. Similarly, if $\mathbf{x}$ is in the null space of the matrix $A - \lambda I$, then $A\mathbf{x} = \lambda\mathbf{x}$. Thus, $\mathbf{x}$ is in the eigenspace E_λ. □

4.4.3 Finding Eigenvalues and Eigenvectors

From the proof of Theorem 4.4.16, we see that the equation $A\mathbf{x} = \lambda\mathbf{x}$ is equivalent to the equation $(A - \lambda I)\mathbf{x} = \boldsymbol{\theta}$. Hence, we now have the following theorem.

Theorem 4.4.17. Let A is a square $n \times n$ matrix and let λ be a scalar. The following are equivalent:

(a) There is a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$.

(b) The matrix $(A - \lambda I)$ is singular.

(c) The scalar λ is a root to the equation $\det(A - \lambda I) = 0$.

Proof. We show that (a) implies (b), (b) implies (c), and (c) implies (a).

$(a) \Rightarrow (b)$: Suppose that a nonzero vector $\mathbf{x}$ satisfies $A\mathbf{x} = \lambda\mathbf{x}$. Thus, $(A - \lambda I)\mathbf{x} = \boldsymbol{\theta}$. Since $\mathbf{x}$ is nonzero, we conclude that matrix $(A - \lambda I)$ is singular.

$(b) \Rightarrow (c)$: Suppose that the matrix $(A-\lambda I)$ is singular. Theorem 4.2.9 implies that $\det(A-\lambda I) = 0$.

$(c) \Rightarrow (a)$: Suppose that $\det(A-\lambda I) = 0$. Theorem 4.2.9 implies that $A-\lambda I$ is singular. Therefore, there is a nonzero vector $\mathbf{x}$ satisfying $(A - \lambda I)\mathbf{x} = \boldsymbol{\theta}$. Therefore, $A\mathbf{x} = \lambda\mathbf{x}$. □

Our next corollary shows that to find an eigenvalue of a square matrix A, one just needs to find the roots of an equation.

Corollary 4.4.18. Let A is a square $n \times n$ matrix and let λ be a scalar. Then λ is an eigenvalue for A if and only if $\det(A - \lambda I) = 0$.

Definition 4.4.19. Let A is a square $n \times n$ matrix.

1. The polynomial $p(\lambda) = \det(A - \lambda I)$ is called the *characteristic polynomial* of A.
2. The polynomial equation $\det(A - \lambda I) = 0$ is called the *characteristic equation* of A.
3. Suppose that r is a root of the characteristic equation $\det(A - \lambda I) = 0$. The *algebraic multiplicity* of r is defined to be the the number of times that r occurs as root.

Problem 1. Find the characteristic polynomial $p(\lambda)$ of the matrix

$$A = \begin{bmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix}.$$

Find all the eigenvalues of A and their algebraic multiplicity.

Solution. To find the eigenvalues of A, by Corollary 4.4.18, we must find the roots of the characteristic equation $\det(A-\lambda I)=0$. To do this, we first determine the matrix $A-\lambda I$ as follows:

$$A-\lambda I=\begin{bmatrix}0 & 0 & -2\\ 1 & 2 & 1\\ 1 & 0 & 3\end{bmatrix}-\begin{bmatrix}\lambda & 0 & 0\\ 0 & \lambda & 0\\ 0 & 0 & \lambda\end{bmatrix}=\begin{bmatrix}-\lambda & 0 & -2\\ 1 & 2-\lambda & 1\\ 1 & 0 & 3-\lambda\end{bmatrix}.$$

To find the roots of the characteristic equation of A we now compute (using the first row) the

$$\begin{aligned}\det(A-\lambda I) &= (-\lambda)(2-\lambda)(3-\lambda)+(-2)(-(2-\lambda)(1))\\ &= (-\lambda)(2-\lambda)(3-\lambda)+2(2-\lambda)\\ &= (2-\lambda)[(-\lambda)(3-\lambda)+2]\\ &= (2-\lambda)[(\lambda^2-3\lambda+2)]\\ &= (2-\lambda)(\lambda-2)(\lambda-1)\end{aligned}$$

Thus, the characteristic polynomial of A is $p(\lambda)=(2-\lambda)(\lambda-2)(\lambda-1)$. The eigenvalues of A are the roots of the characteristic equation $p(\lambda)=0$ which are $\lambda=2$ and $\lambda=1$. The algebraic multiplicity of $\lambda=2$ is two and algebraic multiplicity of $\lambda=1$ is one.

Definition 4.4.20. Let A is a square $n\times n$ matrix.

1. Suppose that λ is an eigenvalue of A. The *geometric multiplicity* of λ is defined to be $\dim(E_\lambda)$.
2. If the algebraic multiplicity of any eigenvalue λ is different than the geometric multiplicity of λ, then the matrix A is said to be *defective.*

Problem 2. Determine the dimension and a basis for all of the eigenspaces of the matrix

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

given in Problem 1. Afterwards, determine whether or not the matrix A is defective.

Solution. The eigenvalues of A were derived in our solution of Problem 1, where we obtained $\lambda=2$ and $\lambda=1$. Theorem 4.4.16 implies that the eigenspace E_λ of A corresponding to the eigenvalue $\lambda=2$, is the null space of the matrix $A-\lambda I=A-2I$. So we must determine the dimension and a basis for the null space of the matrix

$$A-2I=\begin{bmatrix}-2 & 0 & -2\\ 1 & 0 & 1\\ 1 & 0 & 1\end{bmatrix}.$$

Putting this matrix into reduced echelon form we get the matrix

$$B=\begin{bmatrix}1 & 0 & 1\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}.$$

We thus obtain the solution to the system $(A-2I)\mathbf{x}=\boldsymbol{\theta}$ to be

$$\begin{aligned}x_1&=-x_3\\ x_2&=x_2\\ x_3&=x_3.\end{aligned}$$

Therefore, every solution to $(A - 2I)\mathbf{x} = \boldsymbol{\theta}$ can be written in the form

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}.$$

Thus, the vectors $\mathbf{y}_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ and $\mathbf{y}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ form a basis for the eigenspace E_2. Hence, the dimension of the subspace E_2 is two. Therefore, the geometric multiplicity of $\lambda = 2$ is two and is the same as the algebraic multiplicity of $\lambda = 2$ (see Problem 1).

A basis and the dimension for the eigenspace corresponding to $\lambda = 1$ can be determined as above. The reader should complete this problem by showing that the dimension of the eigenspace corresponding to $\lambda = 1$ equals one, and that the vector $\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$ forms a basis for this eigenspace. Therefore, the geometric multiplicity of $\lambda = 1$ equals one and is the same as the algebraic multiplicity of $\lambda = 1$. Since every eigenvalue has its geometric multiplicity equal to its algebraic multiplicity, we conclude that the matrix A is **not defective**.

4.4.4 Distinct Eigenvalues

Theorem 4.4.21. Let A be an $n \times n$ matrix with eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ and corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$. Suppose that these eigenvalues are all distinct, that is, $\lambda_i \neq \lambda_j$ when $i \neq j$ and $1 \leq i, j \leq k$. Then the vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ are linearly independent.

Proof. Let A be an $n \times n$ matrix. Suppose, for a contradiction, that the theorem is false for the matrix A. Thus, there must be a smallest k for which the theorem is false. We will now work with this k. So the matrix A possesses eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$, with corresponding distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$, such that $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ are linearly **de**pendent. First we observe that the eigenvector $\mathbf{x}_1$ is linearly independent. To see this, suppose that $c\mathbf{x}_1 = \boldsymbol{\theta}$. Since $\mathbf{x}_1 \neq \boldsymbol{\theta}$ (because it is an eigenvector), it follows that $c = 0$. Thus, $k > 1$. Since k is the smallest for which the theorem is false, it follows that the eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{k-1}$ are linearly **in**dependent.

Since the vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{k-1}, \mathbf{x}_k$ are linearly **de**pendent, there are scalars $c_1, c_2, \ldots, c_{k-1}, c_k$, which are **not all** 0, such that

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_{k-1}\mathbf{x}_{k-1} + c_k\mathbf{x}_k = \boldsymbol{\theta}. \tag{4.2}$$

We will now show that $c_k \neq 0$. If $c_k = 0$ then we would obtain, from (4.2), the equation

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_{k-1}\mathbf{x}_{k-1} = \boldsymbol{\theta}. \tag{4.3}$$

Because $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{k-1}$ are linearly **in**dependent, we would be able to conclude from (4.3) that all of the scalars $c_1, c_2, \ldots, c_{k-1}, c_k$ must be 0, which is not the case. Thus, we must have that $c_k \neq 0$.

Since $c_k \neq 0$ and $\mathbf{x}_k \neq \boldsymbol{\theta}$, equation (4.2) implies that there is a least one scalar c_ℓ such that $c_\ell \neq 0$ where $1 \leq \ell \leq k-1$. Multiplying both sides of (4.2) on the left by the matrix A, we get

$$c_1A\mathbf{x}_1 + c_2A\mathbf{x}_2 + \cdots + c_{k-1}A\mathbf{x}_{k-1} + c_kA\mathbf{x}_k = A\boldsymbol{\theta}.$$

Since $A\mathbf{x}_i = \lambda\mathbf{x}_i$ for each $i = 1, 2, \ldots k$, we conclude that

$$c_1\lambda_1\mathbf{x}_1 + c_2\lambda_2\mathbf{x}_2 + \cdots + c_{k-1}\lambda_{k-1}\mathbf{x}_{k-1} + c_k\lambda_k\mathbf{x}_k = \boldsymbol{\theta}. \tag{4.4}$$

Multiplying both sides of (4.2) by the scalar λ_k, we get

$$c_1\lambda_k\mathbf{x}_1 + c_2\lambda_k\mathbf{x}_2 + \cdots + c_{k-1}\lambda_k\mathbf{x}_{k-1} + c_k\lambda_k\mathbf{x}_k = \boldsymbol{\theta}. \tag{4.5}$$

Subtracting equation (4.4) from equation (4.5), we see that

$$c_1(\lambda_k - \lambda_1)\mathbf{x}_1 + c_2(\lambda_k - \lambda_2)\mathbf{x}_2 + \cdots + c_{k-1}(\lambda_k - \lambda_{k-1})\mathbf{x}_{k-1} = \boldsymbol{\theta}.$$

Because $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{k-1}$ are linearly independent, we have that

$$c_1(\lambda_k - \lambda_1) = 0, \;\; c_2(\lambda_k - \lambda_2) = 0, \;\; \ldots, \;\; c_{k-1}(\lambda_k - \lambda_{k-1}) = 0.$$

Hence, in particular, $c_\ell(\lambda_k - \lambda_\ell) = 0$. We noted above that $c_\ell \neq 0$. Thus $\lambda_k - \lambda_\ell = 0$ and so, $\lambda_k = \lambda_\ell$. We conclude that not all of the given eigenvalues are distinct. This contradiction shows that the theorem is true. □

Corollary 4.4.22. Let A be an $n \times n$ matrix. If A has n distinct eigenvalues, then A has n linearly independent eigenvectors.

Corollary 4.4.23. Let A be an $n \times n$ matrix with eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ and corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$. If these eigenvalues are all distinct, then $\mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_k \neq \boldsymbol{\theta}$.

Exercises 4.4 & 4.5

Page 305 of text – Odds #1-11. Page 314 of text – Odds #1-17.

4.7 Diagonalization

Definition 4.7.1. A square $n \times n$ matrix D is a **diagonal matrix** if all of the entries off the main diagonal are 0; that is, if D has the form

$$D = \begin{bmatrix} d_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ 0 & 0 & d_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & d_n \end{bmatrix}.$$

Example 3. The following square matrices are diagonal matrices:

$$A = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 7 & 0 & 0 \\ 0 & 0 & -5 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

Theorem 4.7.2. Let D be the diagonal matrix

$$D = \begin{bmatrix} d_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ 0 & 0 & d_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & d_n \end{bmatrix}.$$

Then

$$D^k = \begin{bmatrix} d_1^k & 0 & 0 & \cdots & 0 \\ 0 & d_2^k & 0 & \cdots & 0 \\ 0 & 0 & d_3^k & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & d_n^k \end{bmatrix}$$

for all $k \geq 1$.

Theorem 4.7.3. Let A be an $n \times n$ matrix. Suppose that $A = SDS^{-1}$ were D is an $n \times n$ diagonal matrix D and S is a nonsingular $n \times n$ matrix. Then $A^k = SD^kS^{-1}$ for all $k \geq 1$.

Proof. Exercise, using mathematical induction. □

Example 4. Whenever $A = [\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_n]$ is an $n \times n$ matrix with columns $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$, then $A\mathbf{e}_i = \mathbf{x}_i$ whenever $\mathbf{e}_i$ is the i-th unit vector in $\mathbb{R}^n$. For example, let A be the 3×3 matrix

$$A = \begin{bmatrix} 3 & 1 & -2 \\ -1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \mathbf{x}_3]$$

and let $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, and $\mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ be the unit vectors in $\mathbb{R}^3$. Then

$$A\mathbf{e}_1 = \begin{bmatrix} 3 & 1 & -2 \\ -1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix} = \mathbf{x}_1$$

$$A\mathbf{e}_2 = \begin{bmatrix} 3 & 1 & -2 \\ -1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \mathbf{x}_2$$

$$A\mathbf{e}_3 = \begin{bmatrix} 3 & 1 & -2 \\ -1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \\ 3 \end{bmatrix} = \mathbf{x}_3.$$

Similarly, whenever $A = [\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_n]$ is an $n \times n$ matrix with columns $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$, then $A(\lambda\mathbf{e}_i) = \lambda\mathbf{x}_i$ whenever λ is a scalar and $\mathbf{e}_i$ is the i-th unit vector in $\mathbb{R}^n$.

Definition 4.7.4. A square $n \times n$ matrix A is said to be **diagonalizable** if there is an $n \times n$ diagonal matrix D such that $S^{-1}AS = D$ for some nonsingular $n \times n$ matrix S.

Theorem 4.7.5. A square $n \times n$ matrix A is diagonalizable if and only if the matrix A possesses n linearly independent eigenvectors.

Proof. Let A be a square $n \times n$ matrix.

($\Rightarrow$). We must prove that if A is diagonalizable, then A possesses n linearly independent eigenvectors. So, assume that A is diagonalizable. Thus, there is an $n \times n$ diagonal matrix D such that $S^{-1}AS = D$ for some nonsingular $n \times n$ matrix S. Let $S = [\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_n]$ and suppose that D

has the form

$$D = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

Since the matrix S is nonsingular, Theorem 1.7.11 implies that the vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are linearly independent. We shall prove that $A\mathbf{x}_i = \lambda_i \mathbf{x}_i$ for each $i = 1, 2, \ldots, n$. So, let i be such a natural number. Since $S^{-1}AS = D$, it follows that $AS = SD$. Thus, $(AS)\mathbf{e}_i = (SD)\mathbf{e}_i$ where $\mathbf{e}_i$ is the i-th unit vector in $\mathbb{R}^n$. Thus, $A(Se_i) = S(D\mathbf{e}_i)$. Since $S\mathbf{e}_i = \mathbf{x}_i$ and $D\mathbf{e}_i = \lambda_i \mathbf{e}_i$, we conclude that

$$A\mathbf{x}_i = S\lambda_i \mathbf{e}_i = \lambda_i S\mathbf{e}_i = \lambda_i \mathbf{x}_i.$$

Therefore, $A\mathbf{x}_i = \lambda_i \mathbf{x}_i$ and thus, each $\mathbf{x}_i$ is an eigenvector for A. We can now conclude that A possesses n linearly independent eigenvectors.

($\Leftarrow$). Now suppose that A possesses n linearly independent eigenvectors. We shall prove that A is diagonalizable. Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ be n linearly independent eigenvectors for the matrix A and let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be their corresponding eigenvalues. Hence $A\mathbf{x}_i = \lambda_i \mathbf{x}_i$ for $i = 1, 2, \ldots, n$. Let

$$D = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix} \tag{4.6}$$

and let $S = [\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_n]$. Since $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are linearly independent, Theorem 1.7.11 implies that the matrix S is nonsingular. Thus, S^{-1} exists by Theorem 1.9.7. We shall now prove that $AS = SD$ as follows:

$$\begin{aligned} AS = A[\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_n] &= [A\mathbf{x}_1 \ A\mathbf{x}_2 \ \cdots \ A\mathbf{x}_n] \\ &= [\lambda_1\mathbf{x}_1 \ \lambda_2\mathbf{x}_2 \ \cdots \ \lambda_n\mathbf{x}_n] \\ &= [\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_n] \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix} = SD. \end{aligned}$$

We conclude that $AS = SD$ and hence, $S^{-1}AS = D$. Therefore, A is diagonalizable. □

Remark 4.7.6. Suppose that A is an $n \times n$ diagonalizable matrix. The proof of Theorem 4.7.5 gives a procedure for diagonalizing A. First find n linearly independent eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ for the matrix A, with corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Let $S = [\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_n]$. Then $S^{-1}AS = D$ where D is as (4.6) above.

Problem 5. Consider the 3×3 matrix

$$A = \begin{bmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix}.$$

Show that A is diagonalizable and find a 3×3 matrix S so that $S^{-1}AS = D$ were D is a 3×3 diagonal matrix.

Solution. The matrix A was used in Problem 2 on page 99. In our solution to this Problem 2, we obtain the following three eigenvectors for the matrix A:

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

with corresponding eigenvalues $\lambda_1 = 2, \lambda_2 = 2, \lambda_3 = 1$. By applying the Linear Dependence Algorithm on page 40, one can show that the vectors $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ are linearly independent. Thus, Theorem 4.7.5 implies that A is diagonalizable. Let

$$S = [\mathbf{x}_1 \ \mathbf{x}_2 \ \mathbf{x}_3] = \begin{bmatrix} 0 & -1 & -2 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \text{ and } D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Thus, by the proof of Theorem 4.7.5, we have that $S^{-1}AS = D$.

Given a finite set B of vectors, we let $|B|$ denote the number of vectors in the set B.

Theorem 4.7.7. Let A be a square $n \times n$ matrix. Suppose that A has k distinct eigenvalues $\lambda_1, \lambda_2 \dots, \lambda_k$ where $k < n$. For each natural number $i \le k$, let B_{λ_i} be a basis for E_{λ_i}. Then the set of vectors $U = B_{\lambda_1} \cup B_{\lambda_2} \cup \dots \cup B_{\lambda_k}$ is a linearly independent set of vectors. Hence, if $|U| = n$, then the matrix A is diagonalizable.

Proof. Let A be a square $n \times n$ matrix. Suppose that A has k distinct eigenvalues $\lambda_1, \lambda_2 \dots, \lambda_k$ where $k < n$. For each $i \le k$, let $B_{\lambda_i} = \{\mathbf{x}_{i,1}, \dots, \mathbf{x}_{i,j_i}\}$ be a basis for the eigenspace E_{λ_i}. We note that ($\blacktriangle$) a nonzero linear combination of the vectors in B_{λ_i} is also an eigenvector with eigenvalue λ_i. Moreover, since the eigenvalues are distinct, it follows that $B_{\lambda_i} \cap B_{\lambda_j} = \varnothing$ whenever $i \ne j$. Let $U = B_{\lambda_1} \cup B_{\lambda_2} \cup \dots \cup B_{\lambda_k}$. To show that U is a linearly independent set of vectors, suppose that

$$(c_{1,1}\mathbf{x}_{1,1} + \dots + c_{1,j_1}\mathbf{x}_{1,j_1}) + (c_{2,1}\mathbf{x}_{2,1} + \dots + c_{2,j_2}\mathbf{x}_{2,j_2}) + \dots + (c_{k,1}\mathbf{x}_{k,1} + \dots + c_{k,j_k}\mathbf{x}_{k,j_k}) = \boldsymbol{\theta}.$$

Corollary 4.4.23 and ($\blacktriangle$) imply that each of the above parenthetical linear combinations must be equal to $\boldsymbol{\theta}$. Since each B_{λ_i} is a linearly independent set of vectors, we conclude that

$$c_{1,1} = \dots = c_{1,j_1} = c_{2,1} = \dots = c_{2,j_2} = \dots \ \dots = c_{k,1} = \dots = c_{k,j_k} = 0.$$

Thus, U is a linearly independent set of vectors. Hence, if $|U| = n$, then Theorem 4.7.5 implies that the matrix A is diagonalizable. □

Corollary 4.7.8. Let A be an $n \times n$ matrix. If the sum of the dimensions of all the eigenspaces of A is equal to n, then A is diagonalizable.

Exercises 4.7

Pages 336 to 337 of text – Odds #1-11.