

# Lectures in Linear Algebra

David M. McClendon

Department of Mathematics  
Ferris State University

2024 edition  
©2024 David M. McClendon

---

# Contents

---

<b>Contents</b>	<b>2</b>
<b>1 Vector spaces</b>	<b>5</b>
1.1 Some motivating problems . . . . .	5
1.2 Why do we study “linear” things? . . . . .	11
1.3 Introducing vector spaces . . . . .	18
1.4 Examples of vector spaces . . . . .	25
1.5 Summary of Chapter 1 . . . . .	28
1.6 Homework exercises for Chapter 1 . . . . .	29
<b>2 Matrices</b>	<b>33</b>
2.1 Vocabulary . . . . .	33
2.2 Matrix operations . . . . .	37
2.3 Matrix multiplication . . . . .	39
2.4 <i>Mathematica</i> and calculator commands for matrix operations . . . . .	44
2.5 Homework exercises for Chapter 2 . . . . .	47
<b>3 Subspaces</b>	<b>49</b>
3.1 Introducing subspaces . . . . .	49
3.2 Span . . . . .	56
3.3 Distinguishing between subspaces . . . . .	59
3.4 Affine subspaces . . . . .	62
3.5 Linear independence . . . . .	68
3.6 Basis and dimension . . . . .	74
3.7 Summary of Chapter 3 . . . . .	84
3.8 Homework exercises for Chapter 3 . . . . .	86
<b>4 Dot products and orthogonality</b>	<b>91</b>
4.1 Definitions and properties of dot products . . . . .	91

4.2	Norms, distances and length . . . . .	95
4.3	Orthogonality . . . . .	99
4.4	Orthogonal complements and projections . . . . .	101
4.5	Orthonormal bases and the Gram-Schmidt procedure . . . . .	111
4.6	Angles and the Cauchy-Schwarz inequality . . . . .	115
4.7	Hyperplanes . . . . .	117
4.8	Summary of Chapter 4 . . . . .	123
4.9	Homework exercises for Chapter 4 . . . . .	126
<b>5</b>	<b>Linear transformations</b>	<b>130</b>
5.1	Introduction . . . . .	130
5.2	The standard matrix of a linear transformation . . . . .	132
5.3	Proving whether a transformation is linear . . . . .	144
5.4	Prototypical examples of linear transformations . . . . .	148
5.5	Kernels and images . . . . .	151
5.6	Injectivity, surjectivity, bijectivity and inverses . . . . .	156
5.7	Fundamental Theorem of Linear Algebra . . . . .	162
5.8	More on invertibility . . . . .	166
5.9	Summary of Chapter 5 . . . . .	170
5.10	Homework exercises for Chapter 5 . . . . .	172
<b>6</b>	<b>Systems of linear equations</b>	<b>182</b>
6.1	Definitions and theory . . . . .	182
6.2	Rank and dimension issues . . . . .	191
6.3	Gaussian elimination . . . . .	195
6.4	Summary so far . . . . .	210
6.5	Computing the inverse of a matrix . . . . .	211
6.6	Least-squares approximations . . . . .	213
6.7	Solving systems of equations using technology . . . . .	219
6.8	Homework exercises for Chapter 6 . . . . .	222
<b>7</b>	<b>Determinants</b>	<b>228</b>
7.1	Definition and properties . . . . .	228
7.2	Computing determinants . . . . .	232
7.3	Determinants and cross products . . . . .	237
7.4	Homework exercises for Chapter 7 . . . . .	239
<b>8</b>	<b>Eigentheory</b>	<b>241</b>
8.1	Motivation: a differential equation . . . . .	241
8.2	Eigenvalues and eigenvectors . . . . .	244
8.3	Eigentheory of matrices . . . . .	250
8.4	Computing eigenvalues and eigenvectors . . . . .	254
8.5	Matrix exponentials . . . . .	262

8.6 Homework exercises for Chapter 8 . . . . .	267
<b>A All about sets</b>	<b>270</b>
A.1 Sets . . . . .	270
A.2 Venn diagrams . . . . .	271
A.3 Relationships between, and operations on, sets . . . . .	272
A.4 Sums and differences of sets . . . . .	275
<b>B <i>Mathematica</i> information</b>	<b>276</b>
B.1 What is <i>Mathematica</i> ? . . . . .	276
B.2 Important general concepts re: <i>Mathematica</i> syntax . . . . .	278
B.3 <i>Mathematica</i> quick reference guides . . . . .	280
B.4 Linear algebra in <i>Mathematica</i> . . . . .	284
<b>Index</b>	<b>287</b>

## Chapter 1

---

# Vector spaces

---

### 1.1 Some motivating problems

#### For the economist: the Leontief input-output model

##### EXAMPLE 1

An economy is divided into three sectors: coal, steel and electricity. Each industry depends on the others for raw materials as follows:

- To make \$1 worth of coal, it takes no coal, \$.10 of steel and \$.10 of electricity.
- To make \$1 worth of steel, it takes \$.20 of coal, \$.10 of steel and \$.20 of electricity.
- To make \$1 worth of electricity, it takes \$.40 of coal, \$.20 of steel and \$.10 of electricity.

If we want the economy to output \$1 billion worth of coal, \$.7 billion worth of steel and \$2.9 billion worth of electricity, how much coal, steel and electricity will the economy actually need to produce?

*Setup of Solution:* Let  $c$ ,  $s$  and  $e$  be the amounts of coal, steel and electricity produced (in billions).

	Amount of coal used	Amount of steel used	Amount of electricity used
To produce $c$ coal:	$0c$	$.1c$	$.1c$
To produce $s$ steel:	$.2s$	$.1s$	$.2s$
To produce $e$ electricity:	$.4e$	$.2e$	$.1e$
Total:	$.2s + .4e$	$.1c + .1s + .2e$	$.1c + .2s + .1e$

So to output 1 billion worth of coal, we need

$$(\text{amount of coal produced}) - (\text{amount of coal used}) = 1$$

The steel and electricity work similarly, so we end up with these equations:

$$\begin{cases} c - .2s - .4e = 1 \\ -.1c + .9s - .2e = .7 \\ -.1c - .2s + .9e = 2.9 \end{cases}$$

This is called a “system of equations” which we’d like to solve for  $c$ ,  $s$  and  $e$ .

## For the actuarial scientist: bonus-malus systems

### EXAMPLE 2

You work for an insurance company which charges customers two different rates for automobile policies: rate  $r$  if the policyholder has gone two years without an accident, and rate  $s$  if the policyholder has had an accident in the last two years ( $r < s$ ). Assume that the probability that a policyholder has an accident in any given year is  $p$  ( $p$  is between 0 and 1).

You are given  $p$ , and need to determine the long-term probability that you will be charging a customer  $r$  (as opposed to  $s$ ). This tells your company what values they need to choose for  $r$  and  $s$  to ensure a profit.

*Setup of solution:* Divide the policyholders into three categories:

- A “good”, or “bonus” policyholder has gone two years without an accident.
- A “medium” policyholder has gone one, but not two, years without an accident.
- A “bad”, or “malus” policyholder has had an accident in the past year.

Let  $x$ ,  $y$  and  $z$  be the long-term probability that a policyholder is good, medium or bad (so that  $x + y + z = 1$ ). Observe that in one year:

- A good person has probability  $p$  of becoming bad, and probability  $1 - p$  of staying good;

- A medium person has probability  $p$  of becoming bad, and probability  $1 - p$  of becoming good;
- A bad person has probability  $p$  of staying bad, and probability  $1 - p$  of becoming medium.

So in the long run, the good people in a year (i.e.  $x$ ) will be  $1 - p$  of the current good people  $x$  plus  $1 - p$  of the current medium people  $y$ , i.e.

$$x = (1 - p)x + (1 - p)y \quad \Rightarrow$$

Similarly, the medium people (i.e.  $y$ ) will be  $1 - p$  of the bad people  $z$ , i.e.

$$y = (1 - p)z \quad \Rightarrow \quad y + (p - 1)z = 0$$

Last, the bad people (i.e.  $z$ ) will be  $p$  of the good plus  $p$  of the medium plus  $p$  of the bad, i.e.

$$z = px + py + pz \quad \Rightarrow \quad px + py + (p - 1)z = 0$$

Altogether, we have

$$\begin{cases} px + (p - 1)y = 0 \\ y + (p - 1)z = 0 \\ px + py + (p - 1)z = 0 \\ x + y + z = 1 \end{cases}$$

where  $p$  is a constant. This is another system of equations, that you can solve for  $x$ ,  $y$  and  $z$  (hopefully).

What you really care about is  $x$ , because that's the portion of customers who get the bonus rate.

## For the surveyor: error correction

### EXAMPLE 3

You are a surveyor, studying a plot of land. You take several measurements with an instrument that measures the elevation difference between points.

You find the following:

- from point A to point D, there is an elevation drop of 30 feet
- from point A to point B, there is an elevation drop of 20 feet

- from point B to point D, there is an elevation drop of 8 feet
- from point B to point C, there is an elevation drop of 5 feet
- from point C to point D, there is an elevation drop of 2 feet
- from point C to point A, there is an elevation gain of 26 feet

What is the best estimate of the elevation differences between these points?

*Setup of solution:* Let  $a, b, c$  and  $d$  be elevations A, B, C and D. We have:

$$\begin{cases} a - d = 30 \\ a - b = 20 \\ b - d = 8 \\ b - c = 5 \\ c - d = 2 \\ c - a = -26 \end{cases}$$

The goal would be to solve for  $a, b, c$  and  $d$ , but there is no solution.

So instead, we seek the best *approximation* of a solution of this system.

This approximation is called a “*least-squares solution*” (even though it isn’t actually a solution).

## For the engineer: differential equations

### EXAMPLE 4

Originally, the liquid is  $215^\circ\text{F}$  (pretty hot). To cool the liquid, you put it in a refrigerator which has temperature  $45^\circ\text{F}$ . Describe the function  $f(t)$  which gives the temperature of the liquid at time  $t$ .

*Setup of solution:* Apply Newton’s Law of Heating and Cooling (MATH 330, maybe), which says that the rate of change of the temperature of an object is proportional to the difference between the temperature of the object and the ambient temperature. In this case, that means

i.e. we get the “initial value problem”

$$\begin{cases} f'(t) - kf(t) = -45k \\ f(0) = 215 \end{cases}$$

where  $k$  is some constant.



EXAMPLE 5

An RLC electrical circuit consists of a voltage source,  $6\Omega$  resistor,  $1H$  inductor and  $.2F$  capacitor in series. If the voltage supplied to the circuit at time  $t$  is  $2 \sin t$  V, the initial charge in the circuit is  $.1$  and the initial current is  $1$  ohm, find the charge  $q(t)$  across the circuit at time  $t$ .

*Setup of solution:* Apply Kirchoff's voltage law, Ohm's Law and Faraday's Law to obtain the differential equation

$$Lq''(t) + Rq'(t) + \frac{1}{C}q(t) = E_S(t)$$

which in this example is

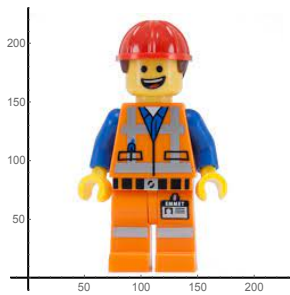
$$q''(t) + 6q'(t) + \frac{1}{.2}q(t) = 2 \sin t.$$

Again, we get a differential equation (we also know  $q(0) = .1$  and  $q'(0) = 1$ ).

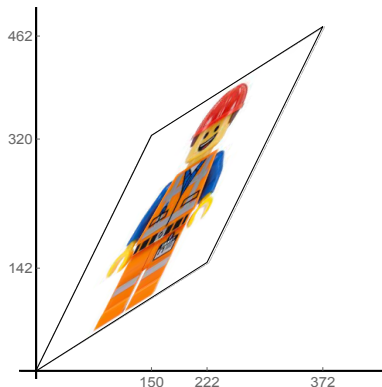
**For the graphic designer: warping pictures**

EXAMPLE 6

Here is a picture of Emmet from the LEGO movie, which is 225 pixels by 225 pixels:



Suppose you need to warp the image of Emmet so that it fits in a parallelogram like this:



1. If a pixel was at position  $(x, y)$  in the old picture, what will its location be in the original picture? In other words, what “formula” is needed to distort the image so that it fits in the right space?
2. If a pixel ends up at position  $(x, y)$  in the new picture, what was its location in the original picture?

## For the pure mathematician: an indefinite integral

### EXAMPLE 7

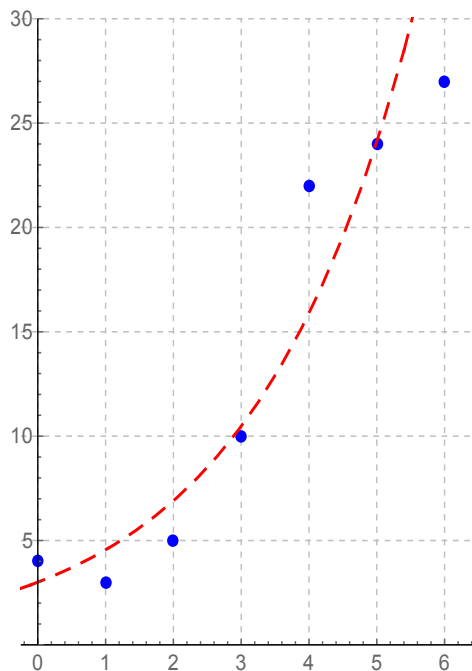
$$\int \cos x \, dx$$

Definite integrals are used to compute areas, displacements of objects, arc lengths, average values, probabilities and expected values, moments, centers of mass, volumes, surface areas, magnetic flux, work, etc., and to compute definite integrals, you need indefinite integrals.

## For the statistician: fit a curve to some data

### EXAMPLE 8

Find the equation of the form  $y = Ae^{bt}$  which best fits the data points  $(0, 4)$ ,  $(1, 3)$ ,  $(2, 5)$ ,  $(3, 10)$ ,  $(4, 22)$ ,  $(5, 24)$  and  $(6, 27)$ .



**What I hope to convince you of in MATH 322:**

All these examples are really the same problem.

The commonality between these problems is that they are all “linear”  
(whatever that means).

## 1.2 Why do we study “linear” things?

### **REASON 1:**

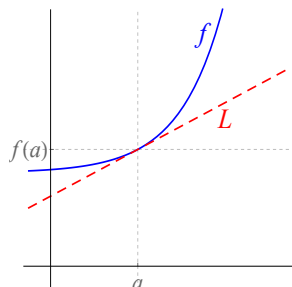
#### **Linear things approximate nonlinear things**

In calculus, we learn about **derivatives**. Computing the derivative of a function is useful because something associated to a derivative solves lots of problems, including

- slope of line tangent to function
- instantaneous rate of change of function
- instantaneous velocity (if the original function gives position at time  $t$ )
- optimization problems (find maximum and/or minimum of function)
- L'Hôpital's rule (for evaluating limits of form  $\frac{0}{0}$  or  $\frac{\infty}{\infty}$ )
- related rates problems
- Newton's method
- etc.

The concept behind all these applications is that if you have any function  $f$  and you want to approximate the values of  $f$  for values of  $x$  near  $a$ , you can *approximate  $f$  by a linear function* (namely, the tangent line at  $a$ ):

$$\begin{aligned} f(x) &\approx L(x) \\ &= f(a) + f'(a)(x - a). \end{aligned}$$



Doing this makes sense because linear equations are easier to work with than arbitrary equations (see reason # 2 below for more on this).

**Long-term goal (MATH 320, perhaps):** So far, you might only have heard about functions  $f : \mathbb{R} \rightarrow \mathbb{R}$  (that is, functions  $y = f(x)$  or  $x \mapsto y$ ). But functions may have much more general domains and ranges; for example, a set of parametric equations describing planar motion can be thought of as a single function

$$f : t \mapsto (x, y)$$

which is a function  $f : \mathbb{R} \rightarrow \mathbb{R}^2$ .

A function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}^3$  might be something like

$$f(u, v) = (u^3v, \sin(uv + 1), e^{2u} - 3).$$

A natural question is to ask what the derivative of such a function is. It should have something to do with the best linear approximation of the function (just as the derivative of a function  $y = f(x)$  gives the slope of the best linear approximation to  $f$  at  $x$ ). But what does linear mean in that context?

**The first major goal of linear algebra is to describe what “linear” means, in a general sense.**

**REASON 2 (TO STUDY LINEAR THINGS):****Linear things are (relatively) easy to study**

Suppose you have an equation in one variable (say  $x$ ). Every such equation can be written as  $f(x) = b$ , where  $f$  is some function from  $\mathbb{R}$  to  $\mathbb{R}$  and  $b$  is some constant. For example,

$$\sin x + x^2 - 6 = 3(x - 2) + x - 5$$

can be rewritten as

To think about such an equation conceptually/theoretically, we need to ask the following questions:

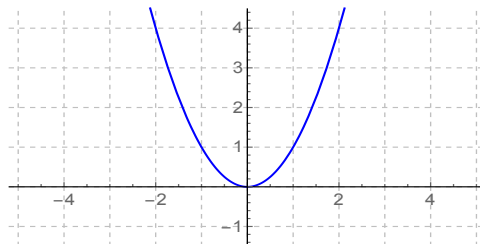
- 1.
- 2.
- 3.

**Question 1 above is easy to answer theoretically:**

There is at least one solution to  $f(x) = b \iff$

$\iff$

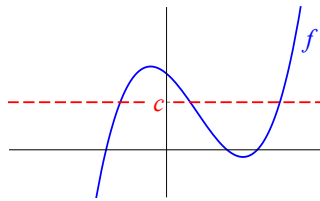
**Example:**  $x^2 = b$  has at least one solution if and only if  $b \geq 0$ .



Unfortunately, in practice, Question 1 might be hard to approach, because we might not have access to a graph of  $y = f(x)$ .

**Question 2 is also easy to answer theoretically:**

The number of solutions to  $f(x) = b$  is equal to



But in practice, Question 2 is also hard because we might not have access to a graph of  $y = f(x)$ , and even if we have a graph, there may be intersection points we’re not aware of (that lie off the piece of the graph we see).

**Question 3, in general, is impossible:**

Consider the equation  $x = \cos x$ , i.e.  $\cos x - x = 0$ . We know there is a solution (since the graph of  $\cos x - x$  crosses the  $x$ -axis), but good luck finding its exact value.

**However:** there is one class of equations in one variable for which we can completely (and easily) answer questions 1,2 and 3. Suppose  $f$  is a function whose graph is a line, i.e.

In this case the equation  $f(x) = c$  becomes

## 1.2. Why do we study “linear” things?

On the previous page we saw that every linear equation in one variable can be rewritten as  $Ax = b$ , where  $A$  and  $b$  are constants.

Let’s look at Questions 1 2 and 3 in this situation:

Equation $Ax = b$	Does the equation have a solution?	How many solutions does the equation have?	What is/are the solution(s)?
Case 1: $A \neq 0$			
Case 2: $A = 0, b = 0$			
Case 3: $A = 0, b \neq 0$			

Summarizing, we have the first theorem (of many) of our course:

**Theorem 1.1 (Linear Equations in One Variable)** *Every linear equation in one variable  $x$  can be rewritten as  $Ax = b$  for constants  $A$  and  $b$ .*

1. *If  $A \neq 0$ , then an equation has exactly one solution, namely  $x = \frac{b}{A} = A^{-1}b$ .*
2. *If  $A = 0$  and  $b = 0$ , then the equation has infinitely many solutions (every  $x \in \mathbb{R}$  is a solution).*
3. *If  $A = 0$  and  $b \neq 0$ , then the equation has no solution.*

**Note:** Once you get much beyond linear equations, Questions 1-3 no longer have such nice answers. Studying these types of questions is the field of mathematical research called *algebraic geometry*.

## 1.2. Why do we study “linear” things?

---

In MATH 322, we want to think about linear systems with more than one equation and/or more than one variable (in part so that we could eventually describe the derivative of a function of several variables, but also because such equations arise naturally). For example, suppose we want to solve for  $x$  and  $y$  if

$$\begin{cases} a_{11}x + a_{12}y = b_1 \\ a_{21}x + a_{22}y = b_2 \end{cases}$$

where the  $a$ s and  $b$ s are constants. This is a system of two equations in two variables. There are a variety of ways one might solve this system:

### EXAMPLE 9

---

Solve the system

$$\begin{cases} 2x - 3y = 13 \\ x - 2y = 7 \end{cases}$$

Method # 1: Addition / elimination

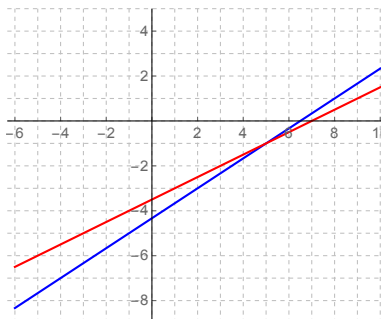
$$\begin{cases} 2x - 3y = 13 \\ x - 2y = 7 \end{cases}$$

Method # 2: Substitution

$$\begin{cases} 2x - 3y = 13 \\ x - 2y = 7 \end{cases}$$

Method # 3: Graphical (imprecise in general)

$$\begin{cases} 2x - 3y = 13 \\ x - 2y = 7 \end{cases}$$





## 1.2. Why do we study “linear” things?

These methods are fine if you have two equations in two variables, but what if you have  $m$  equations in  $n$  variables? An example is something like

$$\begin{cases} 3w + 2x - y + 5z = 8 \\ 2w - x + 4y - z = -3 \\ w + 5x + 2y + 6z = 17 \end{cases}$$

This is a system of three equations in four variables. How do you solve this for  $w, x, y, z$ ? Does this system even have a solution? If so, how many solutions does it have?

### **Big picture conceptual approach:**

- think of the left hand side of this system *as a single function* called  $T$ ;
- think of the four variables  $(w, x, y, z)$  *as a single variable*  $\mathbf{x}$  (a.k.a.  $\vec{x}$ );
- group the right-hand side of the equation  $(8, -3, 17)$  *into a single object* called  $\mathbf{b}$  (a.k.a.  $\vec{b}$ ).
- Then the system above *becomes a single equation*

$$T(\mathbf{x}) = \mathbf{b}.$$

### **What’s hard:**

$\mathbf{x}$  and  $\mathbf{b}$  aren’t numbers—they are objects called *vectors*.

Similarly,  $T$  isn’t a function like any you’ve seen before.

It turns out that  $T$  is something called a *linear transformation* (and is represented by an object called a *matrix*).

**Where we're going:**

Just like what happened with our earlier situation

$$Ax = b,$$

whether or not the **system** (now written as a single equation)

$$T(\mathbf{x}) = \mathbf{b} \quad (\text{a.k.a. } Ax = \mathbf{b} \text{ where } A \text{ is a matrix})$$

**has a solution** (and what that solution is) boils down to whether or not  **$\mathbf{b}$  is in the range of  $T$**  (and this probably depends on some properties of  $T$  and/or  $\mathbf{b}$ ).

So to study these systems in general, we need to figure out what these properties are, and study the nature of linear transformations (and vectors and matrices) in general. This gets us back to the first goal of MATH 322, which we already stated earlier:

**The first major goal of linear algebra is to describe what “linear” means, in a general sense.**

Once we do this, we will then know how to solve systems of any number of equations in any number of variables. Turns out, we will also discover lots of other useful stuff along the way.

## 1.3 Introducing vector spaces

### RECALL

Our goal is to study what “linear” means, in a general sense. To get started, remember that what we know about linear equations from high school is that they all (other than vertical lines) have equation

$$y = mx + b$$

What operations are required to describe this equation?

- 1.
- 2.

Based on this, it is reasonable to expect that if we are going to define “linear” in a general sense, we probably need to assume that there is some notion of each of the two operations above.

A general setting in which we can add objects and multiply objects by real numbers is called a *vector space*:

**Definition 1.2** A (real) vector space  $V$  is a set, together with two operations:

- **addition:**  $+$  :  $V \times V \rightarrow V$  (i.e.  $(\mathbf{u}, \mathbf{v}) \mapsto \mathbf{u} + \mathbf{v}$ )
- **scalar multiplication:**  $\mathbb{R} \times V \rightarrow V$  (i.e.  $(c, \mathbf{v}) \mapsto c\mathbf{v}$ )

such that the following rules (collectively called the **Vector Space Laws**) are satisfied:

**Laws of Vector Addition** For all  $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ ,

1. Addition is closed: This means  $\mathbf{u} + \mathbf{v} \in V$ .
2. Addition is commutative: This means  $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ .
3. Addition is associative: This means  $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$ .
4. Additive identity element: This means there exists a special element of  $V$ , called the **zero vector (of  $V$ )** and denoted  $\mathbf{0}$ , such that  $\mathbf{v} + \mathbf{0} = \mathbf{v}$ .
5. Additive inverses exist: For all  $\mathbf{v} \in V$ , there exists a special element  $-\mathbf{v} \in V$  such that  $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$ .

**Laws of Scalar Multiplication** For all  $c, d \in \mathbb{R}$  and all  $\mathbf{v} \in V$ ,

1. Scalar multiplication is closed: This means  $c\mathbf{v} \in V$ .
2. Scalar multiplication is associative: This means  $(cd)\mathbf{v} = c(d\mathbf{v})$ .
3. Identity element for scalar multiplication: This means  $1\mathbf{v} = \mathbf{v}$ .

**Distributive Laws** For all  $c, d \in \mathbb{R}$  and all  $\mathbf{u}, \mathbf{v} \in V$ ,

$$c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v} \quad \text{and} \quad (c + d)\mathbf{v} = c\mathbf{v} + d\mathbf{v}.$$

In linear algebra, real numbers are called **scalars** and elements of the vector space are called **vectors**.

**Comment:** There are other types of vector spaces (other than “real” vector spaces). These vector spaces have different types of scalars (the scalars can be complex numbers, rational numbers, or more exotic things). In MATH 322, we will not deal with these.

**Notation:** Vectors are usually referred to by boldface letters (like  $\mathbf{v}$ ) when typed, and as letters with arrows over them (like  $\vec{v}$ ) when hand-written. However, sometimes we get lazy and just refer to a vector with a letter (like  $v$ ). The zero scalar is denoted  $0$ ; the zero vector is denoted  $\mathbf{0}$  or  $\vec{0}$ .

**WARNING:** Vectors from two different vector spaces cannot be added to one another; for example, if  $V$  and  $W$  are two unrelated vector spaces with  $\mathbf{u}, \mathbf{v} \in V$  but  $\mathbf{w} \in W$ , then  $\mathbf{u} + \mathbf{v}$  makes sense but  $\mathbf{u} + \mathbf{w}$  is nonsense.

### Our first example of a vector space: $\mathbb{R}^2$

**Theorem 1.3** *The set  $\mathbb{R}^2 = \{(x, y) : x, y \in \mathbb{R}\}$  is a real vector space, where the addition and scalar multiplication are defined coordinate-wise, i.e.*

$$(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2) \quad \text{and} \quad c(x_1, y_1) = (cx_1, cy_1).$$

**PROOF** Let  $\mathbf{u}, \mathbf{v}$ , and  $\mathbf{w}$  be arbitrary elements of  $\mathbb{R}^2$ , and let  $c, d \in \mathbb{R}$ . By definition of  $\mathbb{R}^2$ , we have  $\mathbf{u} = (u_1, u_2)$ ;  $\mathbf{v} = (v_1, v_2)$  and  $\mathbf{w} = (w_1, w_2)$ . To verify  $\mathbb{R}^2$  is a vector space, we have to check all the vector space laws one by one:

1. *Addition is closed:* This is obvious by the definition of vector addition.
2. *Addition is commutative:* We need to check  $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ :

$$\begin{aligned} \mathbf{u} + \mathbf{v} &= (u_1, u_2) + (v_1, v_2) \\ &= (u_1 + v_1, u_2 + v_2) \quad (\text{by def'n of } + \text{ in } \mathbb{R}^2) \\ &= (v_1 + u_1, v_2 + u_2) \quad (\text{since } + \text{ is commutative in } \mathbb{R}) \\ &= (v_1, v_2) + (u_1, u_2) \quad (\text{by def'n of } + \text{ in } \mathbb{R}^2) \\ &= \mathbf{v} + \mathbf{u}. \end{aligned}$$

3. *Addition is associative:*

$$\begin{aligned} \mathbf{u} + (\mathbf{v} + \mathbf{w}) &= (u_1, u_2) + ((v_1, v_2) + (w_1, w_2)) \\ &= (u_1, u_2) + (v_1 + w_1, v_2 + w_2) \quad (\text{by def'n of } + \text{ in } \mathbb{R}^2) \\ &= (u_1 + (v_1 + w_1), u_2 + (v_2 + w_2)) \quad (\text{by def'n of } + \text{ in } \mathbb{R}^2) \\ &= ((u_1 + v_1) + w_1, (u_2 + v_2) + w_2) \quad (\text{since } + \text{ is associative in } \mathbb{R}) \\ &= (u_1 + v_1, u_2 + v_2) + (w_1, w_2) \quad (\text{by def'n of } + \text{ in } \mathbb{R}^2) \\ &= ((u_1, u_2) + (v_1, v_2)) + (w_1, w_2) \quad (\text{by def'n of } + \text{ in } \mathbb{R}^2) \\ &= (\mathbf{u} + \mathbf{v}) + \mathbf{w}. \end{aligned}$$

4. *Additive identity element:* Let  $\mathbf{0} = (0, 0)$ . Then

$$\begin{aligned}\mathbf{v} + \mathbf{0} &= (v_1, v_2) + (0, 0) \\ &= (v_1 + 0, v_2 + 0) \quad (\text{by def'n of } + \text{ in } \mathbb{R}^2) \\ &= (v_1, v_2) \quad (\text{since } 0 \text{ is additive identity in } \mathbb{R}) \\ &= \mathbf{v}.\end{aligned}$$

5. *Additive inverses exist:* Given  $\mathbf{v} = (v_1, v_2)$ , let  $-\mathbf{v} = (-v_1, -v_2)$ . Then

$$\begin{aligned}\mathbf{v} + (-\mathbf{v}) &= (v_1, v_2) + (-v_1, -v_2) \\ &= (v_1 + (-v_1), v_2 + (-v_2)) \quad (\text{by def'n of } + \text{ in } \mathbb{R}^2) \\ &= (0, 0) \quad (\text{by properties of additive inverses in } \mathbb{R}) \\ &= \mathbf{0} \quad (\text{by def'n of } \mathbf{0}).\end{aligned}$$

6. *Scalar multiplication is closed:* This is obvious by the definition of scalar multiplication.

7. *Scalar multiplication is associative:*

$$\begin{aligned}(cd)\mathbf{v} &= (cd)(v_1, v_2) \\ &= ((cd)v_1, (cd)v_2) \quad (\text{by def'n of scalar multiplication}) \\ &= (c(dv_1), c(dv_2)) \quad (\text{by associativity of } \cdot \text{ in } \mathbb{R}) \\ &= c(dv_1, dv_2) \quad (\text{by def'n of scalar multiplication}) \\ &= c(d(v_1, v_2)) \quad (\text{by def'n of scalar multiplication}) \\ &= c(d\mathbf{v}).\end{aligned}$$

8. *Identity element for scalar multiplication:*

$$\begin{aligned}1\mathbf{v} &= 1(v_1, v_2) \\ &= (1v_1, 1v_2) \quad (\text{by def'n of scalar multiplication}) \\ &= (v_1, v_2) \quad (\text{since } 1 \text{ is mult. identity in } \mathbb{R}) \\ &= \mathbf{v}.\end{aligned}$$

9. *Distributive Law # 1:*

$$\begin{aligned}c(\mathbf{u} + \mathbf{v}) &= c((u_1, u_2) + (v_1, v_2)) \\ &= c(u_1 + v_1, u_2 + v_2) \quad (\text{by def'n of } + \text{ in } \mathbb{R}^2) \\ &= (c(u_1 + v_1), c(u_2 + v_2)) \quad (\text{by def'n of scalar multiplication}) \\ &= (cu_1 + cv_1, cu_2 + cv_2) \quad (\text{by distributivity of } \mathbb{R}) \\ &= (cu_1, cu_2) + (cv_1, cv_2) \quad (\text{by def'n of } + \text{ in } \mathbb{R}^2) \\ &= c(u_1, u_2) + c(v_1, v_2) \quad (\text{by def'n of scalar multiplication}) \\ &= c\mathbf{u} + c\mathbf{v}.\end{aligned}$$

10. *Distributive Law # 2:*

$$\begin{aligned}
 (c + d)\mathbf{v} &= (c + d)(v_1, v_2) \\
 &= ((c + d)v_1, (c + d)v_2) \quad (\text{by def'n of scalar multiplication}) \\
 &= (cv_1 + dv_1, cv_2 + dv_2) \quad (\text{by distributivity of } \mathbb{R}) \\
 &= (cv_1, cv_2) + (dv_1, dv_2) \quad (\text{by def'n of } + \text{ in } \mathbb{R}^2) \\
 &= c(v_1, v_2) + d(v_1, v_2) \quad (\text{by def'n of scalar multiplication}) \\
 &= c\mathbf{v} + d\mathbf{v}.
 \end{aligned}$$

Since all the laws hold,  $\mathbb{R}^2$  is indeed a vector space over  $\mathbb{R}$ .  $\square$

**Note:** Actually checking the vector space laws is extremely tedious; for the most part, we don't do this in undergraduate mathematics.

**Logical technicality:** To verify the vector space laws, you need to describe exactly what the additive identity element (a.k.a. the zero vector) is what exactly is meant by an additive inverse.

The work in Theorem 1.3 generalizes:

**Theorem 1.4** For any  $n \geq 1$ , define  $\mathbb{R}^n$  to be the set of  $n$ -tuples of real numbers with addition and scalar multiplication defined coordinate-wise. For each  $n$ ,  $\mathbb{R}^n$  is a vector space (and these are the most important examples of vector spaces).

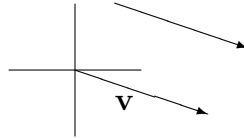
EXAMPLE 10

Let  $\mathbf{v} = (2, -1, 3, 0)$  and  $\mathbf{w} = (-3, -2, 0, 2)$ .

1. To what vector space should one assume  $\mathbf{v}$  and  $\mathbf{w}$  belong?
2. Compute  $4\mathbf{v}$ .
3. Compute  $\mathbf{v} + \mathbf{w}$ .
4. Compute  $2\mathbf{v} - 3\mathbf{w}$ .

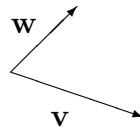
## Pictorial representations of vectors

To get a picture of elements of  $\mathbb{R}^n$  (or any vector space, really), we think of them as “arrows”:

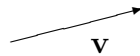


In reality, vectors in  $\mathbb{R}^n$  are more like “floating” arrows, in that two vectors pointing the same direction with the same length are really the same vector (even if they are drawn in different spots).

Using the idea of vectors as “arrows”, vector addition then corresponds to “head-to-tail” or “parallelogram” addition:

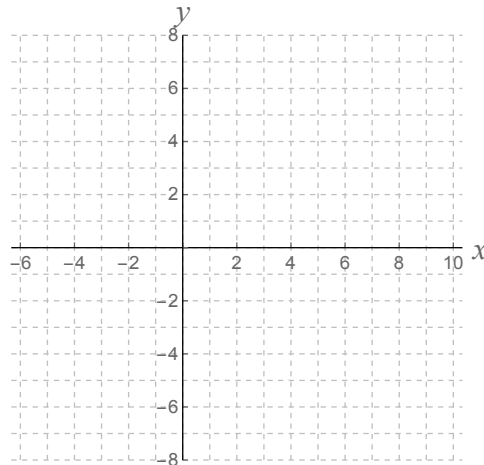


Scalar multiplication corresponds to “stretching”:

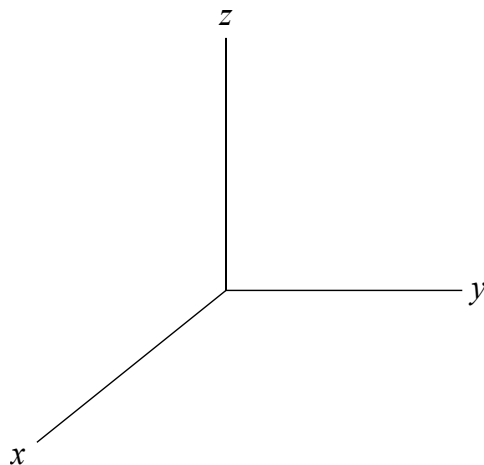


### EXAMPLE 11

1. Let  $\mathbf{v} = (3, -1)$  and  $\mathbf{u} = (2, 1)$ .

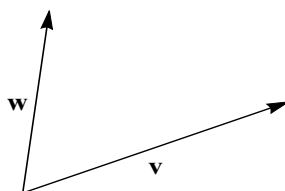


2. Let  $\mathbf{v} = (1, 2, 4)$  and  $\mathbf{w} = (-6, -2, 5)$ .



3. Let  $\mathbf{v} = (0, -2, 3, 1)$  and  $\mathbf{w} = (2, 8, -1, 4)$ .

4. Let  $\mathbf{v}$  and  $\mathbf{w}$  be as shown here:





In the definition of a vector space, we assume the ten so-called “vector space laws”. From those ten laws, it is possible to show that the following standard rules also hold in any real vector space:

**Theorem 1.5 (Additional laws of real vector spaces)** *Let  $V$  be a real vector space. Then:*

1. Uniqueness of additive identity: *There is only one zero vector  $\mathbf{0}$  in  $V$ .*
2. Uniqueness of additive inverses: *For every  $\mathbf{v} \in V$ , there is only one vector  $-\mathbf{v}$  which is the additive inverse of  $\mathbf{v}$ .*
3. Rule for additive inverses: *For every  $\mathbf{v} \in V$ ,  $(-1)\mathbf{v} = -\mathbf{v}$ .*
4. Zero property # 1: *For every  $\mathbf{v} \in V$ ,  $0\mathbf{v} = \mathbf{0}$ .*
5. Zero property # 2: *For every  $c \in \mathbb{R}$ ,  $c\mathbf{0} = \mathbf{0}$ .*

## 1.4 Examples of vector spaces

Here is a fairly exhaustive list of vector spaces; you should familiarize yourself with this list and know the symbol(s) used to represent each space:

1. “Traditional” vector spaces: For any  $n \in \mathbb{N}$ ,  $\mathbb{R}^n = \{(x_1, \dots, x_n) : x_j \in \mathbb{R} \forall j\}$  is a real vector space, where the addition and scalar multiplication are defined coordinate-wise. (P.S.  $\forall$  means “for all”;  $\exists$  means “there exists”)
2. Real numbers:  $\mathbb{R}$  is a real vector space (where the addition and scalar multiplication are the usual numerical operations). In particular,  $\mathbb{R}^1 = \mathbb{R}$ .
3. Zero vector space:  $\{\mathbf{0}\}$  (the set consisting only of a zero vector) is a vector space. In particular,  $\mathbb{R}^0 = \{\mathbf{0}\}$ .
4. Function spaces: Each of the following sets is a real vector space; for all these sets of functions, the addition is described by  $(f+g)(x) = f(x) + g(x)$  and the scalar multiplication is  $(cf)(x) = c \cdot f(x)$ . (In all cases, the additive identity element is the constant function  $f(x) = 0$  and the additive inverse of  $f$  is the function  $-f$ .) Specific function spaces include:
  - a) *Polynomials*: The set  $\mathbb{R}[x]$  of polynomials whose coefficients are in  $\mathbb{R}$  is a real vector space.

- b) *Polynomials of bounded degree*: The set  $\mathcal{P}_n$  of polynomials with real coefficients whose degree is at most  $n$  is a real vector space.
- c) *Continuous functions*: The set  $C(\mathbb{R}, \mathbb{R})$  of continuous functions from  $\mathbb{R}$  to  $\mathbb{R}$  is a real vector space.
- d) *Differentiable functions*: The set of differentiable functions from  $\mathbb{R}$  to  $\mathbb{R}$  is a real vector space.
- e) *Analytic functions*: The set  $C^\omega$  of analytic functions from  $\mathbb{R}$  to  $\mathbb{R}$  is a real vector space (a function is *analytic* if it can be written as a power series which converges everywhere).
5. *Sequence spaces*: In these examples, the addition and scalar multiplication are defined term-by-term, i.e.

$$(x_1, x_2, x_3, \dots) + (y_1, y_2, y_3, \dots) = (x_1 + y_1, x_2 + y_2, x_3 + y_3, \dots)$$

and

$$c(x_1, x_2, \dots) = (cx_1, cx_2, \dots).$$

- a) The set  $\mathbb{R}^{\mathbb{N}}$  of infinite sequences of elements of  $\mathbb{R}$  is a real vector space (where the addition and scalar multiplication are done term-by-term).
- b) The set  $\mathbb{R}^\infty$  of infinite sequences where all but finitely many elements of the sequence are 0 also forms a real vector space. Here, some care needs to be taken to verify that addition is closed.
- c) The set of convergent sequences of real numbers forms a vector space over  $\mathbb{R}$ .
6. *Matrix spaces*: The set of  $m \times n$  matrices (this means  $m$  rows and  $n$  columns) with elements in  $\mathbb{R}$ , denoted  $M_{mn}(\mathbb{R})$ , is a real vector space where the addition and scalar multiplication are defined entry-wise. (Notation: the set of square  $n \times n$  matrices with entries in  $\mathbb{R}$  is denoted  $M_n(\mathbb{R})$  rather than  $M_{nn}(\mathbb{R})$ .) Matrix spaces are discussed in more detail in Chapter 2.

**Key commonalities:** With all these examples of vector spaces,

1. You can add two things in the set, and the sum is always something in the set.
2. You can multiply things in the set by a scalar, and the result is always something in the set.

If either (1) or (2) fail, then the set is not a vector space.

## EXAMPLE 12

Suppose  $V$  and  $W$  are (different) real vector spaces; suppose  $\mathbf{v}_1, \mathbf{v}_2$  and  $\mathbf{v}_3$  are vectors in  $V$ ; suppose  $\mathbf{w}_1$  and  $\mathbf{w}_2$  are vectors in  $W$ ; suppose all other letters represent scalars. Determine whether each of the following expressions results in a **scalar**, a **vector**, or if the expression is **nonsense**.

1.  $\mathbf{v}_1 + 5c\mathbf{v}_3$

2.  $c(\mathbf{v}_1 - \mathbf{v}_1)$

3.  $c(d\mathbf{v}_2)$

4.  $\mathbf{v}_1(\mathbf{v}_2 + c\mathbf{v}_3)$

5.  $\mathbf{v}_2(c + d)$

6.  $\mathbf{v}_1 + \mathbf{w}_1$

7.  $\frac{1}{c}\mathbf{v}_1$

8.  $0\mathbf{w}_2$

9.  $c(d_1 + d_2)$

## 1.5 Summary of Chapter 1

1. A **(real) vector space** is a set  $V$  of objects called **vectors** which can be added to one another and multiplied by **scalars** (a scalar is a real number) in such a way that the addition and scalar multiplication obey a bunch of algebraic rules (listed earlier in this packet).
2. A vector space is the most general setting in which one can define addition and multiplication. Since linear functions are made up of addition and multiplication, this means vector spaces are the most general settings for “linear” objects.
3. Examples of vector spaces include:
  - “traditional vector spaces”  $\mathbb{R}^n$ , where the vectors are ordered  $n$ -tuples of real numbers;
  - “function spaces”, where the vectors are functions;
  - “sequence spaces”, where the vectors are sequences;
  - “matrix spaces”, where the vectors are matrices.

Each of these vector spaces has their own notion of addition and scalar multiplication. You cannot add vectors from two different vector spaces.

4. A good pictorial representation of a vector is a “floating arrow”. In this setting, scalar multiplication corresponds to stretching and addition corresponds to “head-to-tail” addition.
5. The whole point of linear algebra is to define what **linear** means in a general sense. We are interested in doing this because linear things are relatively easy to work with, and because you can approximate any real-world problem with a linear problem.
6. Eventually we will get around to studying systems of linear equations. So far, we have looked only at systems of one linear equation in one variable. Such an equation can always be rewritten as

$$Ax = b.$$

- If  $A \neq 0$ , there is exactly one solution of this equation, namely  $x = A^{-1}b$ .
- If  $A = 0$  and  $b = 0$ , then every  $x$  is a solution of this equation (there are infinitely many solutions).
- If  $A = 0$  and  $b \neq 0$ , the equation has no solution.

## 1.6 Homework exercises for Chapter 1

**Note:** Exercises 1-12 are unrelated to material we cover during lecture; these questions are based on the material in Appendix A (“All about sets”) in the back of these notes.

In Problems 1-3, consider the following sets:

$$E = \{1, 3, 5, 7\}$$

$$F = \{0, 2, 4, 6\}$$

$$G = \{1, 2, 4, 5, 7\}$$

$$H = \{x : 0 \leq x \leq 7 \text{ and } x \text{ is an integer}\}$$

1. Classify the following statements as true or false:

- |                     |                           |
|---------------------|---------------------------|
| a) $2 \in G$        | e) $E \subseteq H$        |
| b) $2 \notin E$     | f) $H \subseteq G$        |
| c) $2 \in F \cap H$ | g) $E \cap F = \emptyset$ |
| d) $E \in H$        | h) $E \cup F = H$         |

2. Describe the set by giving a list of its elements:

- |               |                        |
|---------------|------------------------|
| a) $E \cup F$ | c) $(E \cap F) \cap G$ |
| b) $E \cap F$ | d) $(E \cup F) \cap G$ |

3. Which two of the sets  $E, F, G$  and  $H$  are disjoint?

4. a) Suppose  $A$  and  $B$  are two arbitrary sets with  $A \subseteq B$ .

- i. In this situation, what is  $A \cap B$ ?
- ii. What is  $A \cup B$ ?

b) Let  $S$  be any set.

- |                          |                                   |
|--------------------------|-----------------------------------|
| i. What is $S \cup S$ ?  | iii. What is $S \cap \emptyset$ ? |
| ii. What is $S \cap S$ ? | iv. What is $S \cup \emptyset$ ?  |

5. Let  $E, F$  and  $G$  be any three sets. Classify each statement as true or false:

- |  |   |
|--|---|
| a) $E \cap F = F \cap E$                   | d) $E \cup (F \cup G) = (E \cup F) \cup G$          |
| b) $E \cup F = F \cup E$                   | e) $E \cap (F \cup G) = (E \cap F) \cup G$          |
| c) $E \cap (F \cap G) = (E \cap F) \cap G$ | f) $E \cap (F \cup G) = (E \cap F) \cup (E \cap G)$ |

*Hint:* For each question, draw a Venn diagram and figure out which region(s) of the Venn diagram go with the left- and right-hand sides of the equation. If the same regions go with both sides, the statement is true; otherwise, the statement is false.

6. In this problem, you are given various subsets of  $\mathbb{R}^2$ . For each subset, sketch a picture of  $\mathbb{R}^2$  and indicate (by shading, or color-coding, etc.) which points belong to the set. Please draw a different picture for each problem:
- $A = \{(x, y) : y = 2x\}$ ;
  - $B = \{(x, y) : x \geq 1\}$ ;
  - $A \cap B$  (where  $A$  and  $B$  are as in the previous two parts);
  - $C = \{(x, y) : x + y = 3\}$ .
7. Same directions as the previous problem:
- $A \cup C$  (where  $A$  and  $C$  are as in the previous problem);
  - $A \cap C$  (where  $A$  and  $C$  are as in the previous problem);
  - $D = \{(x, y) : x^2 = y^2\}$ ;
  - $E = \{(x, y) : y < x^2\}$ .
8. In this problem, let  $E$  be the set of all even integers, let  $D$  be the set of all odd integers, let  $P$  be the set of positive integers, and let  $N$  be the set of all negative integers.
- Is the statement  $E + E = 2E$  true or false? Explain.
  - Describe each of the following sets (i.e. they are one of  $E, D, P, N$  or  $\mathbb{Z}$ ):

$$E + D \quad D + D \quad D + P \quad P + P \quad -P \quad E + 3$$

9. In this problem, consider the following intervals in  $\mathbb{R}$ :

$$A = [0, 3) \quad (\text{i.e. } A = \{x : 0 \leq x < 3\})$$

$$B = [-5, -1]$$

$$C = [2, \infty)$$

Write each of the sets below in interval notation:

- |            |                |
|------------|----------------|
| a) $4A$    | e) $C + C + C$ |
| b) $-3C$   | f) $A - A$     |
| c) $A + B$ | g) $C - 3$     |
| d) $A + C$ | h) $B + 4$     |

10. In this problem, consider the following subsets of  $\mathbb{R}^2$ :

$$A = \{(x, y) : y = x\}$$

$$B = \{(x, y) : y = 3x + 6\}$$

$$C = \{(x, y) : |x| \leq 1\}$$

$$D = \{(x, y) : y \geq 3\}$$

$$E = \{(x, y) : x \geq 2\}$$

For each given set below, sketch a picture of  $\mathbb{R}^2$  and indicate (by shading, etc.) which points belong to the set. Please draw a different picture for each problem:

a)  $3A$

c)  $-C$

b)  $2B$

d)  $D + E$

11. Same directions and sets as the previous problem:

a)  $A + E$

c)  $A + (0, 1)$

b)  $A + B$

d)  $A + (-3, 4)$

12. Consider the following subsets of  $\mathbb{R}^2$ :

$$A = \{(x, y) \in \mathbb{R}^2 : x \geq 2\}$$

$$B = \{(x, y) \in \mathbb{R}^2 : y = \frac{1}{2}x\}$$

$$C = \{(x, y) \in \mathbb{R}^2 : y = x^2\}$$

$$D = \{(x, y) \in \mathbb{R}^2 : x = 0\}$$

$$E = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$$

For each subset of  $\mathbb{R}^2$  given below, sketch a picture of  $\mathbb{R}^2$  and indicate (by shading, etc.) which points belong to the set. Please draw a different picture for each problem.

a)  $B + (2, 3)$

f)  $A + B$

b)  $C - (0, 1)$

g)  $B + D$

c)  $C + (0, 5)$

h)  $-E$

d)  $D + (3, 2)$

i)  $4E$

e)  $D + (4, 0)$

j)  $D + E$

13. a) Let  $W$  be a subset of a vector space  $V$  and let  $\mathbf{v} \in V$ . Based on your answers to parts (a)-(e) of the preceding question, describe in your own words what you think the difference is between the sets  $W$  and  $W + \mathbf{v}$ .  
*Hint:* One or more of the words “rotate”, “reflect”, “shift”, “translate”, “stretch”, “shrink” may be useful in describing what is going on.
- b) Let  $E$  be the pentagon in  $\mathbb{R}^2$  whose vertices are  $(-1, 0)$ ,  $(0, 1)$ ,  $(2, 1)$ ,  $(2, -1)$  and  $(0, -1)$ . Sketch a picture of  $E + (-5, -2)$ , indicating the coordinates of the vertices of this object.
14. Classify the following statements as true or false (no explanation is required). The sets  $A, \dots, E$  in this problem are the same ones as in Problem 12.
- |                |                                  |
|----------------|----------------------------------|
| a) $A = A + A$ | e) $D = D - D$                   |
| b) $B = 2B$    | f) $E = 2E$                      |
| c) $C = 2C$    | g) $cD = D$ for any constant $c$ |
| d) $B = B + B$ | h) $0A = D$                      |
15. Which one or ones of the sets  $A, B, C, D, E$  from Problem 12 contain the zero vector?
16. Which one or ones of the sets  $A, B, C, D, E$  from Problem 12 are “closed under addition”? (A set is called **closed under addition** if the sum of any two elements from that set is also in that set.)
17. Which one or ones of the sets  $A, B, C, D, E$  from Problem 12 are “closed under scalar multiplication”? (A set is called **closed under scalar multiplication** if the product of any scalar times any element of the set must also be in that set.)
18. Compute the following quantities, if they make sense. If they don’t make any sense, just write “nonsense”.
- |                               |                                       |
|-------------------------------|---------------------------------------|
| a) $(2, -3, 1) + (-7, 5, 2)$  | c) $2(1, 4, -2, 5) + (2, 3, 0, 1, 0)$ |
| b) $3(-1, 0, 4) - 2(1, 3, 0)$ | d) $4(1, 2) + 3(0, -1) + (5, 6)$      |



## Chapter 2

---

# Matrices

---

### 2.1 Vocabulary

In the last chapter, we discussed several examples of vector spaces (sets of ordered  $n$ -tuples, the zero vector space, spaces of functions, etc.). Now, we introduce one more class of a vector spaces, whose importance will be in helping us represent the “ $T$ ”s in the  $T(\mathbf{x}) = \mathbf{b}$  we want to study.

**Definition 2.1** Given positive integers  $m$  and  $n$ , an  $m \times n$  **matrix** with entries in  $\mathbb{R}$  is an array of numbers  $a_{ij} \in \mathbb{R}$  where  $1 \leq i \leq m$  and  $1 \leq j \leq n$ . We denote matrices by capital letters (usually); a matrix with entries  $a_{ij}$  is usually denoted  $A$ . We arrange the entries of the matrix in a rectangle as follows:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & \cdots & a_{mn} \end{pmatrix}$$

The set of matrices of size  $m \times n$  with entries in  $\mathbb{R}$  is denoted  $M_{mn}(\mathbb{R})$ .

Two matrices are **equal** if they are the same size and if all their entries coincide, i.e.  $A = B$  if they are both  $m \times n$  and if  $a_{ij} = b_{ij}$  for all  $i, j$ .

In particular,  $a_{ij}$  is the entry of  $A$  in the  $i^{\text{th}}$  row and  $j^{\text{th}}$  column.  $m$  is the number of rows of  $A$ ;  $n$  is the number of columns of  $A$ .

**Definition 2.2** A matrix is called **square** if it has the same number of rows as columns. The set of square  $n \times n$  matrices with entries in  $\mathbb{R}$  is denoted  $M_n(\mathbb{R})$  (as opposed to  $M_{nn}(\mathbb{R})$ ).

EXAMPLES:  $A = \begin{pmatrix} 1 & 6 \\ -4 & 5 \end{pmatrix}$   $B = (1 \ 5 \ 7)$

**Definition 2.3** Given a matrix  $A$ , the **diagonal entries** of  $A$  are the numbers  $a_{11}, a_{22}, a_{33}, \dots, a_{nn}$  (this list may stop at  $a_{mm}$  depending on when you run out of entries).

EXAMPLE:  $A = \begin{pmatrix} 2 & 7 \\ 5 & 1 \\ 4 & -2 \end{pmatrix}$

**Definition 2.4** A matrix  $A$  is called **diagonal** if it is square and all of its nondiagonal entries are zero.

EXAMPLE:  $A = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}$

**Definition 2.5** The **trace** of a matrix, denoted  $\text{tr}(A)$ , is the sum of the diagonal entries of that matrix.

EXAMPLE:  $A = \begin{pmatrix} 7 & -2 & 4 \\ 1 & -2 & 5 \\ 0 & 1 & -1 \end{pmatrix}$   $\text{tr}(A) =$

**Definition 2.6** A square matrix is called **upper triangular** (abbreviated **upper  $\Delta$** ) if all its entries below its diagonal are zero. A matrix is called **lower triangular** (abbreviated **lower  $\Delta$** ) if all the entries above its diagonal are zero. A matrix is called **triangular** if it is either lower triangular or upper triangular.

EXAMPLES:  $A = \begin{pmatrix} 2 & 3 \\ 0 & -4 \end{pmatrix}$  is upper  $\Delta$ .  $B = \begin{pmatrix} -7 & 0 & 0 \\ 3 & -2 & 0 \\ 1 & -3 & 4 \end{pmatrix}$  is lower  $\Delta$ .

**Note:** diagonal matrices are both upper triangular and lower triangular.

**Definition 2.7** The **transpose** of an  $m \times n$  matrix  $A$ , denoted  $A^T$  or  $A^t$ , is the  $n \times m$  matrix satisfying  $(a^T)_{ij} = a_{ji}$  for all  $i, j$ .

EXAMPLE:  $A = \begin{pmatrix} 1 & -4 \\ 2 & -1 \\ 3 & 0 \end{pmatrix}$   $A^T =$

**Definition 2.8** The  $n \times n$  **identity matrix**, denoted  $I$  or  $I_n$ , is the diagonal  $n \times n$  matrix with all diagonal entries equal to 1.

$$I = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

**Definition 2.9** Given an  $m \times n$  matrix  $A$  as above, the vectors

$$\begin{aligned} &(a_{11}, a_{12}, a_{13}, \dots, a_{1n}), \\ &(a_{21}, a_{22}, a_{23}, \dots, a_{2n}), \\ &\quad \dots, \\ &(a_{m1}, a_{m2}, \dots, a_{mn}) \end{aligned}$$

are called the **rows** of  $A$ ; note that each row of  $A$  is an element of  $\mathbb{R}^n$ . The vectors

$$\begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}, \dots, \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}$$

are called the **columns** of  $A$ ; note that each column of  $A$  is an element of  $\mathbb{R}^m$ .

### EXAMPLE 1

For each given matrix:

- Give the size of the matrix.
- Write down the  $(3, 2)$ -entry of the matrix (if it exists).
- Write down the diagonal entries of the matrix.
- Write down the second row of the matrix.
- What is the trace of the matrix?
- Is the matrix square? diagonal? upper triangular? lower triangular?
- Write down the transpose of the matrix.

$$1. A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

$$2. B = \begin{pmatrix} -4 & 2 & 5 & 6 \\ 7 & 0 & -3 & 2 \\ 4 & 2 & 1 & -2 \end{pmatrix}$$

- a)  $B \in$
- b)  $b_{32} =$
- c) diagonal entries are
- d) the second row of  $B$  is
- e)  $\text{tr}(B) =$
- f)  $B$  is not square and hence not diagonal or triangular

$$g) B^T = \begin{pmatrix} -4 & 7 & 4 \\ 2 & 0 & 2 \\ 5 & -3 & 1 \\ 6 & 2 & 2 \end{pmatrix}.$$

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

- a)  $C \in M_{25}(\mathbb{R})$ , i.e.  $C$  is  $2 \times 5$
- b)  $c_{32}$  DNE
- c) diagonal entries are  $-1$  and  $3$
- d) the second row of  $C$  is  $(0, 3, 1, 4, -6)$  (this is an element of  $\mathbb{R}^5$ )
- e)  $\text{tr}(C) = -1 + 3 = 2$
- f)  $C$  is not square and hence not diagonal or triangular

$$g) C^T = \begin{pmatrix} -1 & 0 \\ 2 & 3 \\ 0 & 1 \\ -3 & 4 \\ 0 & -6 \end{pmatrix}$$

$$4. D = \begin{pmatrix} 3 & 0 \\ 1 & -5 \end{pmatrix}$$

- a)  $D \in M_2(\mathbb{R})$
- b)  $d_{32}$  DNE
- c) diagonal entries are  $3$  and  $-5$
- d) the second row of  $D$  is  $(1, -5)$  (this is an element of  $\mathbb{R}^2$ )
- e)  $\text{tr}(D) = 3 + (-5) = -2$
- f)  $D$  is square and lower triangular, but not diagonal

$$g) D^T = \begin{pmatrix} 3 & 1 \\ 0 & -5 \end{pmatrix}$$

## 2.2 Matrix operations

The first thing we want to do is show that for each  $m$  and  $n$ , the set  $M_{mn}(\mathbb{R})$  of  $m \times n$  matrices forms a vector space. To do this, we need to define what it means to add two matrices, and what it means to multiply a matrix by a scalar. Essentially, *addition and scalar multiplication of matrices are performed entry-by-entry*. More precisely,

**Definition 2.10** Given two matrices  $A, B \in M_{mn}(\mathbb{R})$  and a scalar  $r \in \mathbb{R}$ , we define the matrix  $A + B \in M_{mn}(\mathbb{R})$  by

$$A + B = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & a_{13} + b_{13} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & a_{23} + b_{23} & \cdots & a_{2n} + b_{2n} \\ a_{31} + b_{31} & a_{32} + b_{32} & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & \cdots & a_{mn} + b_{mn} \end{pmatrix}$$

(equivalently we set  $(a + b)_{ij} = a_{ij} + b_{ij}$  for all  $i, j$ ); and we define the matrix  $rA \in M_{mn}(\mathbb{R})$  by

$$rA = \begin{pmatrix} ra_{11} & ra_{12} & ra_{13} & \cdots & ra_{1n} \\ ra_{21} & ra_{22} & ra_{23} & \cdots & ra_{2n} \\ ra_{31} & ra_{32} & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ ra_{m1} & ra_{m2} & \cdots & \cdots & ra_{mn} \end{pmatrix}$$

(equivalently we set  $(ra)_{ij} = r(a_{ij})$  for all  $i, j$ ).

**Note:** You can only add two matrices of the same size.

### EXAMPLE 2

Let  $A = \begin{pmatrix} 4 & 1 & -1 \\ 2 & 0 & -3 \end{pmatrix}$ ,  $B = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & -1 \end{pmatrix}$  and  $C = \begin{pmatrix} 2 & 1 \\ 1 & 5 \end{pmatrix}$ .

1.  $A + B =$

2.  $3A - 2B = \begin{pmatrix} 3(4) - 2(1) & 3(1) - 2(1) & 3(-1) - 2(2) \\ 3(2) - 2(0) & 3(0) - 2(1) & 3(-3) - 2(-1) \end{pmatrix} = \boxed{\begin{pmatrix} 10 & 1 & -5 \\ 6 & -2 & -7 \end{pmatrix}}$ .

3.  $4C = \begin{pmatrix} 4(2) & 4(1) \\ 4(1) & 4(5) \end{pmatrix} = \boxed{\begin{pmatrix} 8 & 4 \\ 4 & 20 \end{pmatrix}}$ .

4.  $A + C$  **DNE** because  $A$  and  $C$  are not the same size.

**Theorem 2.11** *The operations defined above make each  $M_{mn}(\mathbb{R})$  into a real vector space; the additive identity element of  $M_{mn}(\mathbb{R})$  is the  $m \times n$  **zero matrix***

$$\mathbf{0} = \mathbf{0}_{m \times n} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & 0 \end{pmatrix}.$$

### Associating vectors in $\mathbb{R}^n$ to column vectors

We associate vectors in  $\mathbb{R}^n$  to  $n \times 1$  matrices as follows:

$$\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n \quad \longleftrightarrow \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in M_{n1}(\mathbb{R}).$$

In particular, an  $n \times 1$  matrix is also called a **column vector**. A column vector with  $n$  entries is the same thing as a vector in  $\mathbb{R}^n$ . Given a vector  $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ , if we want to think of that vector as a row vector, we take the transpose of  $\mathbf{x}$ :

$$\mathbf{x}^T = (x_1 \ x_2 \ x_3 \ \cdots \ x_n) \in M_{1n}(\mathbb{R}).$$

**Theorem 2.12 (Properties of elementary matrix operations)** *Let  $A$  and  $B$  be matrices of the same size, and let  $r \in \mathbb{R}$ . Then:*

**Transpose is linear:** *This means  $(A + B)^T = A^T + B^T$  and  $(rA)^T = r(A^T)$*

**Transpose undoes itself:**  $(A^T)^T = A$ .

**Transpose preserves trace:**  $tr(A^T) = tr(A)$ .

**Trace preserves addition:**  $tr(A + B) = tr(A) + tr(B)$

## 2.3 Matrix multiplication

There is another operation one can perform on matrices, which doesn't directly have anything to do with thinking of matrices as vectors. The importance of this operation will be seen later in the course; for now we simply define it:

**Definition 2.13** Given matrices  $A \in M_{mn}(\mathbb{R})$  and  $B \in M_{pq}(\mathbb{R})$ , if  $n = p$  then we can define the **product**  $AB$ , which is an  $m \times q$  matrix  $AB$  defined entrywise by setting

$$(ab)_{ij} = \sum_{k=1}^{n(=p)} a_{ik}b_{kj}.$$

(If  $n \neq p$ ,  $AB$  is undefined.)

### REMARKS

- If  $A$  is a square matrix, we write  $A^2$  for  $AA$ ,  $A^3$  for  $AAA$ , etc. But if  $A$  isn't square, then  $A^2$  is undefined.
- In general **matrix multiplication is not commutative**:  $AB \neq BA$  most of the time, even if both products are defined.

### How to recognize matrix multiplication:

Matrix multiplication is ubiquitous in mathematics. Whenever you encounter one of these type of formulas in a math problem:

$$(\text{answer})_{\square\Delta} = \sum_{\star} \text{something}_{\square\star} \text{something else}_{\star\Delta}$$

$$\text{answer}(\square, \Delta) = \sum_{\star} \text{something}(\square, \star) \text{something else}(\star, \Delta)$$

you should instinctively think

“Hey! The answer is the matrix product of “something” times “something else”.

If you are familiar with dot products, you can see that each entry of  $AB$  is a dot product:

$$(i, j)\text{-entry of } AB = (i^{\text{th}} \text{ row of } A) \cdot (j^{\text{th}} \text{ column of } B)$$

If you aren't familiar with dot products, don't worry... we'll discuss them later.

EXAMPLE 3

---

Let  $A = \begin{pmatrix} 1 & -1 & 0 \\ 2 & 1 & -3 \end{pmatrix}$ ,  $B = \begin{pmatrix} -2 & 1 \\ 1 & 3 \end{pmatrix}$  and  $C = \begin{pmatrix} 5 & -1 \\ 1 & 2 \\ 0 & 1 \end{pmatrix}$ . Compute each quantity. If the quantity is undefined, say so (with justification).

1.  $AB$

2.  $BA$

3.  $A^2$

4.  $B^2$



5. (Recall that  $A = \begin{pmatrix} 1 & -1 & 0 \\ 2 & 1 & -3 \end{pmatrix}$ ,  $B = \begin{pmatrix} -2 & 1 \\ 1 & 3 \end{pmatrix}$  and  $C = \begin{pmatrix} 5 & -1 \\ 1 & 2 \\ 0 & 1 \end{pmatrix}$ .)

$CB^2A$

**Theorem 2.14 (Properties of matrix multiplication)** Let  $A, B, C$  be matrices with entries in  $\mathbb{R}$ , let  $I$  be the identity matrix of the appropriate size and let  $r \in \mathbb{R}$ . Then, **so long as everything is defined**, we have:

**Matrix multiplication is associative:** This means

$$A(BC) = (AB)C \quad \text{and} \quad r(AB) = (rA)B = A(rB)$$

**Matrix multiplication distributes over addition:** This means

$$A(B + C) = AB + AC \quad \text{and} \quad (A + B)C = AC + BC$$

**$I$  is a multiplicative identity:** This means  $IA = A$  and  $AI = A$ .

**Transpose of a prod. is the backward prod. of transposes:**  $(AB)^T = B^T A^T$ .

**Trace of a product doesn't care about the order:**  $tr(AB) = tr(BA)$ .

#### EXAMPLE 4

Suppose  $A \in M_{mn}(\mathbb{R})$ ,  $\mathbf{b} \in \mathbb{R}^n$ ,  $C \in M_{nn}(\mathbb{R})$ ,  $D \in M_{nm}(\mathbb{R})$  (assume in this problem that neither  $m$  nor  $n$  are 1, and assume  $m \neq n$ ). For each of the following expressions, determine if the expression is a **scalar**, a **vector**, a **matrix** (in which you should give its size), or **nonsense**.

1.  $A\mathbf{b}$

2.  $AC^2$

3.  $CAB$

4.  $\mathbf{b}CI$

5.  $\mathbf{b}\mathbf{b}^T$

6.  $\mathbf{b}^T\mathbf{b}$

7.  $\mathbf{b}^TIC$

8.  $\mathbf{b}^TC\mathbf{b}$

9.  $(\mathbf{b}^T\mathbf{b})D$

10.  $DACD$

*Solution:*  $D_{n \times m} A_{m \times n} C_{n \times n} D_{n \times m} = (DACD)_{n \times m}$  which is an  $n \times m$  matrix.

11.  $DAC^2\mathbf{b}$

*Solution:*  $D_{n \times m} A_{m \times n} C_{n \times n} C_{n \times n} \mathbf{b}_{n \times 1} = (DAC^2\mathbf{b})_{n \times 1}$  which is a vector in  $\mathbb{R}^n$ .

12.  $\mathbf{b}^TDA$

13.  $(AC)^T$

## 2.4 *Mathematica* and calculator commands for matrix operations

Matrix operations for large-sized matrices can and should be done electronically using either a computer software package or calculator. In this section I give directions for matrix operations using the computer software package *Mathematica* and I give directions for matrix operations on a TI-83/84 series graphics calculator. For other calculators, consult your user manual or the internet, or ask your instructor or classmate.

### Matrix operations on *Mathematica*

**Defining a matrix:** To store a matrix as a variable, there are two methods.

1. Type the matrix in using braces very carefully. For example, to save the matrix  $\begin{pmatrix} 2 & 4 & 7 \\ -5 & 3 & 1 \end{pmatrix}$  as  $A$ , execute

$$A = \{\{2,4,7\}, \{-5,3,1\}\}$$

Note that the entries are separated by commas, every row of the matrix needs braces around it, and the entire matrix needs braces around it.

To type in a column vector, you need only one set of braces, so if you execute  $b = \{1,2,3\}$ , this defines the column vector  $(1, 2, 3) = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ .

2. Use the Basic Math Assistant Palette. Click Palettes and Basic Math Assistant, then on the Basic Math Assistant click the fourth tab under Basic Commands that looks like a matrix. In the *Mathematica* notebook, type  $A=$ , then click the large button that looks like a matrix, then click AddRow or Add Column until the matrix is the appropriate size. Click in each box of the matrix and type in the appropriate numbers. For example, your command to define the  $A$  above would look like

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

**Matrix multiplication:** To multiply two matrices in *Mathematica*, you need a period between the matrices. For example, after defining matrices  $A$  and  $B$ , you can compute the matrix product by

$$A.B$$

The output you will get won't look like a matrix; to make it look like a matrix you can type

`A.B //MatrixForm`

For matrix powers, you will need to type `A.A` rather than `A^2`. For a larger matrix power (say  $A^{100}$ ), run the following:

`MatrixPower[A,100] //MatrixForm`

**Other matrix operations:** Once you have saved a matrix as a letter or string, you can perform standard operations on it as follows (add `//MatrixForm` to the end of the command to make the output look like a matrix):

1. For the transpose of  $A$ , execute `Transpose[A]`.
2. For the trace of  $A$ , execute `Tr[A]`.
3. To multiply  $A$  by a scalar (say 5), execute `A`.
4. To add two matrices (say  $A$  and  $B$ ), execute `A + B`.
5. To get the  $i, j$  entry of a matrix, use double braces: execute `A[[2,3]]` (to get the 2, 3–entry).
6. To call the  $n \times n$  identity matrix, use a command like `IdentityMatrix[4]` (this generates the  $4 \times 4$  identity matrix).

### Matrix operations on TI-83/84-type calculators

The main button you need to find is the `[MATRX]` button, which may be above the `[SIN]` key or the second function on the `[x-1]`, button depending on your particular model.

**Defining a matrix:** Suppose you want to enter the matrix

$$\begin{pmatrix} 2 & 4 & 7 \\ -5 & 3 & 1 \end{pmatrix}$$

into your calculator. There are two ways to do this:

1. On your normal screen, type `[[2,4,7][-5,3,1]]`, then hit `[STO→]`, `[MATRX]` and choose a name for the matrix from the `NAME` menu. Note that there are no commas between the rows.
2. Alternatively, you can hit `[MATRX]`. You will see a menu where across the top of the screen the calculator says `NAMES`, `MATH`, `EDIT`. Use the right arrow to highlight `EDIT`, and then if necessary move the cursor up or down until you get to the letter you want to save your matrix as. Type in the size of the matrix (in this case it is  $2 \times 3$ ), and then type the entries in one by one. Once you get done, hit `[QUIT]` to return to the home screen.

**Matrix multiplication:** To multiply two matrices (say  $A$  times  $B$ ) with a TI-83/84 calculator, first save the two matrices and then return to the home screen. Now, hit [MATRX], highlight NAMES, and go down to [A] and hit [ENTER]. Now hit [MATRX], highlight names, and go down to [B] and hit [ENTER].

Your home screen should now look like [A][B]; now hit [ENTER] and the matrix product will be displayed.

**Other matrix operations:** Assuming you have saved the matrices you want to work with:

1. For the transpose of  $A$ , type [MATRX], highlight NAMES, go down to [A] and hit [ENTER]. Then hit [MATRX], highlight MATH and choose  $T$ . Hit [ENTER]; the home screen will look like  $[A]^{nT}$ . Hit [ENTER] and the transpose will be displayed.
2. To multiply  $A$  by a scalar (say 5), hit 5 and then hit [MATRX], highlight NAMES, and go down to [A] and hit [ENTER]. Your home screen will look like 5[A]; now hit [ENTER] and the product will be displayed.
3. To add two matrices (say  $A$  and  $B$ ), use the same commands as for matrix multiplication, but put a + in between the matrices.

## 2.5 Homework exercises for Chapter 2

In problems 1 through 6, let

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \end{pmatrix}; \quad B = \begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 3 & 2 \end{pmatrix}; \quad C = \begin{pmatrix} 3 & -1 & 3 \\ 4 & 1 & 5 \\ 2 & 1 & 3 \end{pmatrix};$$

$$D = \begin{pmatrix} 3 & -2 \\ 2 & 4 \end{pmatrix}; \quad E = \begin{pmatrix} 2 & -4 & 5 \\ 0 & 1 & 4 \\ 3 & 2 & 1 \end{pmatrix}; \quad F = \begin{pmatrix} -4 & 5 \\ 2 & 3 \end{pmatrix};$$

$$\mathbf{x} = (2, 1, -1); \quad \mathbf{y} = (1, -3).$$

- Compute each of the following quantities, if they make sense. If they don't make any sense, just write "nonsense".
  - the size of  $A$
  - the size of  $C$
  - the size of  $B^T$
  - the diagonal entries of  $E$
  - the trace of  $C$
  - $\text{tr}(F)$
- Compute each of the following quantities, if they make sense. If they don't make any sense, just write "nonsense".
  - $b_{1,2}$
  - $b_{2,1}$
  - $2c_{1,3} - c_{2,3}$
  - the second row of  $A$
  - the third column of  $C$
  - the third row of  $D$
- Compute each of the following quantities, if they make sense. If they don't make any sense, just write "nonsense".
  - $A + B$
  - $C + E$
  - $2F - 3D$
  - $A^T$
  - $(A^T)^T$
  - $(2 + 3)B$
- Compute each of the following quantities, if they make sense. If they don't make any sense, just write "nonsense".
  - $2(3B)$
  - $D - D^T$
  - $(2A)^T$
  - $(B + A^T)^T$
  - $(B + F)^T$
  - $C + I$  (you should know what  $I$  is)

5. Compute each of the following quantities, if they make sense. If they don't make any sense, just write "nonsense".

a)  $AB$

d)  $FB$

b)  $(3A)B$

e)  $A(C + E)$

c)  $BF$

f)  $F^3$

6. Compute each of the following quantities, if they make sense. If they don't make any sense, just write "nonsense".

a)  $ABDF$

d)  $Ax$

b)  $DB^T$

e)  $xA$

c)  $DA^T$

f)  $y^T Fy$

7. a) Give an example of two  $2 \times 2$  matrices  $A$  and  $B$  such that  $AB \neq BA$  (verify that these products are unequal).

b) Give an example of two  $2 \times 2$  matrices  $A$  and  $B$  such that:

- neither  $A$  nor  $B$  are diagonal matrices;
- $B \neq cA$  for any scalar  $c$ ; and
- $AB = BA$  (verify that the products are equal).



## Chapter 3

---

# Subspaces

---

### 3.1 Introducing subspaces

#### RECALL

We have seen several examples of real vector spaces so far:

- The zero vector space  $\{\mathbf{0}\}$ ;
- $\mathbb{R}^n$ , the space of “traditional” vectors;
- spaces of functions like  $C(\mathbb{R}, \mathbb{R})$ ,  $\mathbb{R}[x]$ ,  $\mathcal{P}_n$ , etc.;
- spaces of sequences; and
- spaces  $M_{mn}(\mathbb{R})$  of  $m \times n$  matrices.

Some of these vector spaces are subsets of one another. For example, we know from calculus that every polynomial is a continuous function. Restated, this says that the set  $\mathbb{R}[x]$  of polynomials is a subset of the space  $C(\mathbb{R}, \mathbb{R})$  of all continuous functions from  $\mathbb{R}$  to  $\mathbb{R}$ . In set language, we would write

When one vector space is a subset of another, we say that the first space is a *subspace* of the second. More precisely:

**Definition 3.1** *Let  $V$  be a real vector space, and let  $W \subseteq V$  be a subset of  $V$ .  $W$  is called a **subspace** of  $V$  if  $W$  is itself a real vector space under the operations of addition and scalar multiplication that define  $V$ .*

The only way  $W$  would not be a real vector space is if it is not closed under addition or scalar multiplication, so this definition can be restated as follows:

**Definition 3.2 (equivalent to Definition 3.1)** Let  $V$  be a real vector space, and let  $W \subseteq V$  be a subset of  $V$ .  $W$  is called a **subspace** of  $V$  if:

1.  $W$  is **nonempty**, meaning there exists some  $\mathbf{w} \in W$ ;
2.  $W$  is **closed under addition**, meaning that if  $\mathbf{w}_1 \in W$  and  $\mathbf{w}_2 \in W$ , then  $\mathbf{w}_1 + \mathbf{w}_2 \in W$ .
3.  $W$  is **closed under scalar multiplication**, meaning that if  $\mathbf{w}_1 \in W$  and  $r \in \mathbb{R}$ , then  $r\mathbf{w}_1 \in W$ .

**Theorem 3.3** Let  $V$  be a real vector space, and let  $W \subseteq V$  be a subspace. Then the zero vector  $\mathbf{0}$  must be in  $W$ .

PROOF Let  $W$  be a subspace of  $V$ . Then  $W \neq \emptyset$ , so there is some  $\mathbf{w} \in W$ . Since  $W$  is closed under scalar multiplication,

$$0\mathbf{w} = \mathbf{0} \in W$$

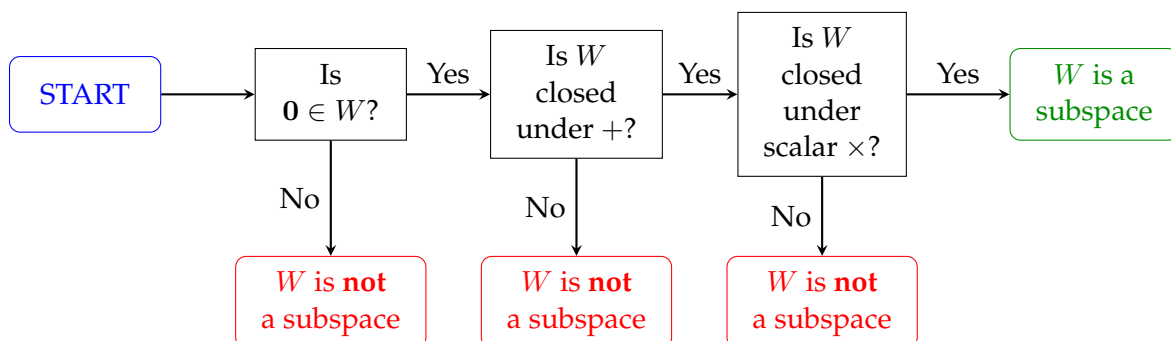
as desired.  $\square$

**A preview:** It turns out that a deep understanding of subspaces is key to solving problems like those we encountered in Chapter 1.

Therefore, we need to obtain a good understanding of what is and what isn't a subspace. Eventually, we'll develop intuition for this, but for now, we need to learn how to *prove* whether or not subsets of a vector space are subspaces.

### How to prove if a subset of a vector space is a subspace

Definition 3.2 and Theorem 3.3 give us a standard way to decide whether a subset of a vector space is a subspace. If you are given vector space  $V$  and subset  $W \subseteq V$ , to decide whether  $W$  is a subspace, ask the following questions (this is called the *brute-force method* for determining whether or not  $W$  is a subspace:



EXAMPLE 1

In these examples, you are given a real vector space  $V$  and a subset  $W$  of  $V$ . Determine, with justification, whether or not  $W$  is a subspace of  $V$ .

1.  $V =$  any vector space.  $W = \{0\}$ .

INTUITION	RIGOROUS PROOF

2.  $V = M_2(\mathbb{R})$ .  $W =$  the set of diagonal  $2 \times 2$  matrices.

INTUITION	RIGOROUS PROOF

3.  $V = \mathbb{R}^2$ .  $W =$  the set of vectors  $(x, y)$  such that  $y = 3$ .

INTUITION	RIGOROUS PROOF
<ul style="list-style-type: none"> <li>• Is <math>\mathbf{0} \in W</math>?</li> </ul>	

4.  $V = \mathbb{R}^2$ .  $W =$  the  $x$ -axis.

INTUITION	RIGOROUS PROOF
<ul style="list-style-type: none"> <li>• Is <math>\mathbf{0} \in W</math>?</li> </ul>	
<ul style="list-style-type: none"> <li>• <math>W</math> closed under <math>+</math>?</li> </ul>	
<ul style="list-style-type: none"> <li>• <math>W</math> closed under <math>\cdot</math> ?</li> </ul>	

5.  $V = \mathbb{R}^2$ .  $W =$  the upper half-plane =  $\{(x, y) : y \geq 0\}$

INTUITION	RIGOROUS PROOF
<ul style="list-style-type: none"> <li>• Is <math>\mathbf{0} \in W</math>?</li> </ul>	
<ul style="list-style-type: none"> <li>• <math>W</math> closed under <math>+</math>?</li> </ul>	
<ul style="list-style-type: none"> <li>• <math>W</math> closed under <math>\cdot</math> ?</li> </ul>	

6.  $V = C(\mathbb{R}, \mathbb{R})$ .  $W =$  the set of differentiable functions from  $\mathbb{R}$  to  $\mathbb{R}$ .

INTUITION	RIGOROUS PROOF
<ul style="list-style-type: none"> <li>• Is <math>\mathbf{0} \in W</math>?</li> </ul>	
<ul style="list-style-type: none"> <li>• <math>W</math> closed under <math>+</math>?</li> </ul>	
<ul style="list-style-type: none"> <li>• <math>W</math> closed under <math>\cdot</math> ?</li> </ul>	

7.  $V = C(\mathbb{R}, \mathbb{R})$ .  $W =$  the set of functions passing through  $(5, 0)$ .

INTUITION	RIGOROUS PROOF
<ul style="list-style-type: none"> <li>• Is <math>\mathbf{0} \in W</math>?</li> </ul>	
<ul style="list-style-type: none"> <li>• <math>W</math> closed under <math>+</math>?</li> </ul>	
<ul style="list-style-type: none"> <li>• <math>W</math> closed under <math>\cdot</math> ?</li> </ul>	

8.  $V = C(\mathbb{R}, \mathbb{R})$ .  $W =$  the set of functions passing through  $(5, 2)$ .

INTUITION	RIGOROUS PROOF
<ul style="list-style-type: none"> <li>• Is <math>\mathbf{0} \in W</math>?</li> </ul>	

9.  $V =$  any vector space.  $W = V$ .

INTUITION	RIGOROUS PROOF
<ul style="list-style-type: none"> <li>• Is <math>\mathbf{0} \in W</math>?</li> </ul>	<p>Let <math>\mathbf{w}_1, \mathbf{w}_2 \in W</math> and <math>r \in \mathbb{R}</math>.</p> <p>Then <math>\mathbf{w}_1 \in V</math> and <math>\mathbf{w}_2 \in V</math>.</p>
<ul style="list-style-type: none"> <li>• <math>W</math> closed under <math>+</math>?</li> </ul>	<p>Therefore <math>\mathbf{w}_1 + \mathbf{w}_2 \in V</math>, since <math>V</math> is a vector space. Since <math>V = W</math>, that means <math>\mathbf{w}_1 + \mathbf{w}_2 \in W</math>, so <math>W</math> is closed under <math>+</math>.</p>
<ul style="list-style-type: none"> <li>• <math>W</math> closed under <math>\cdot</math>?</li> </ul>	<p>Also, <math>r\mathbf{w}_1 \in V</math>, since <math>V</math> is a vector space. Since <math>V = W</math>, that means <math>r\mathbf{w}_1 \in W</math>, so <math>W</math> is closed under scalar multiplication.</p>
	<p>Therefore <math>W</math> is a subspace of <math>V</math>. <math>\square</math></p>

**To write a proof that a subset  $W$  is a subspace of  $V$ , you need to do both of these things:**

1. Take two generic elements of  $W$  (see worksheet) and add them. Verify that the sum belongs to  $W$ .
2. Take a generic element of  $W$  and multiply it by a generic constant (like  $r$ ). Verify that this product belongs to  $W$ .

**To write a proof that a subset  $W$  is NOT a subspace of  $V$ , you need to do one of these three things:**

1. Explain why  $0$  does not belong to  $W$ .
2. Alternatively, write down two **specific** elements of  $W$  (i.e. **with numbers**) whose sum is not in  $W$ .
3. Alternatively, write down a **specific** element of  $W$  (i.e. **with numbers**) and a **specific** scalar (i.e. **a number**) such that when you multiply that scalar by that element, the product is not in  $W$ .

Performing either (2) or (3) in this context is called *finding an explicit counterexample*.

## Intuition regarding subspaces

How do you know which of the two above procedures you are supposed to be carrying out?

In order for a subset  $W$  to be a subspace of vector space  $V$ ,  $W$  should ...

- **contain the zero vector**,
- **be unbounded**, i.e. extend “forever” in any direction it goes (otherwise it won’t be closed under scalar multiplication);
- **be flat/straight** (otherwise it won’t be closed under scalar multiplication);
- **and be convex**, i.e. for any two points in  $W$ , the line segment connecting the points should stay entirely within  $W$  (otherwise, it won’t be closed under addition).

If  $W$  is these four things, it is probably a subspace. That said, these intuitive concepts are **NOT** substitutes for a proof.

## 3.2 Span

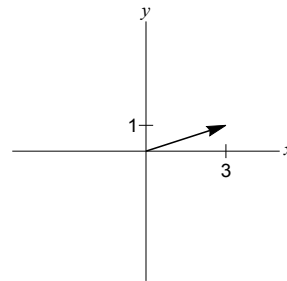
Here is an important example of a class of subspaces:

**Definition 3.4** Let  $V$  be a real vector space and let  $\mathbf{v} \in V$  be any vector. Define  $W = \{c\mathbf{v} : c \in \mathbb{R}\}$  to be the set of scalar multiples of  $\mathbf{v}$ .  $W$  is called the **span** of  $\mathbf{v}$  and denoted  $\text{Span}(\mathbf{v})$  or  $\langle \mathbf{v} \rangle$ .

**General pictorial representation** of the span of a single vector:

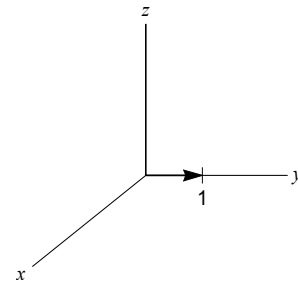
### EXAMPLE 2

$$V = \mathbb{R}^2; \mathbf{v} = (3, 1).$$



### EXAMPLE 3

$$V = \mathbb{R}^3; \mathbf{v} = (0, 1, 0).$$



### EXAMPLE 4

$$V = C(\mathbb{R}, \mathbb{R}), f(x) = \sin x.$$



**Theorem 3.5** *Let  $V$  be a real vector space. Then the span of any vector is a subspace of  $V$ .*

PROOF Suppose  $\mathbf{w}_1, \mathbf{w}_2 \in \text{Span}(\mathbf{v})$  and  $r \in \mathbb{R}$ .

That means  $\mathbf{w}_1 = c_1\mathbf{v}$ ,  $\mathbf{w}_2 = c_2\mathbf{v}$ .

So  $\mathbf{w}_1 + \mathbf{w}_2 = c_1\mathbf{v} + c_2\mathbf{v} = (c_1 + c_2)\mathbf{v} \in \text{Span}(\mathbf{v})$ , so  $\text{Span}(\mathbf{v})$  is closed under  $+$ .

Also,  $r\mathbf{w}_1 = r(c_1\mathbf{v}) = (rc_1)\mathbf{v} \in \text{Span}(\mathbf{v})$ , so  $\text{Span}(\mathbf{v})$  is closed under scalar multiplication.

Therefore  $\text{Span}(\mathbf{v})$  is a subspace of  $V$ .

**Definition 3.6** *Let  $V$  be a real vector space and let  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n \in V$  be any collection of vectors. Define the **span** of these vectors to be the set  $W \subseteq V$ , denoted  $\text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_n)$  or  $\langle \mathbf{v}_1, \dots, \mathbf{v}_n \rangle$ , by*

$$W = \text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_n) = \{c_1\mathbf{v}_1 + \dots + c_n\mathbf{v}_n : c_j \in \mathbb{R} \forall j\}$$

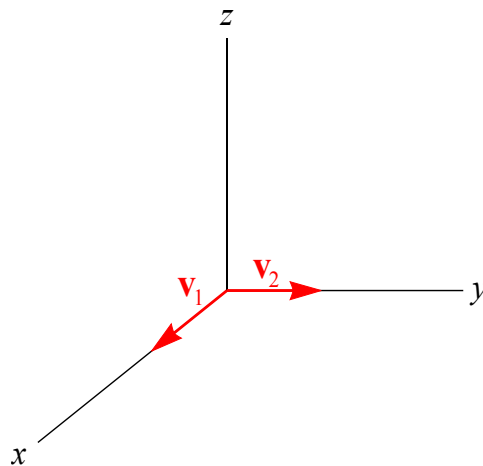
*(this the set of linear combinations of the  $\mathbf{v}_j$ ).*

The word “span” can also be used as a verb:

**Definition 3.7** *Let  $V$  be a real vector space (or subspace) and let  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n \in V$  be a collection of vectors. We say the set  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$  **spans**  $V$  if  $V = \text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_n)$ .*

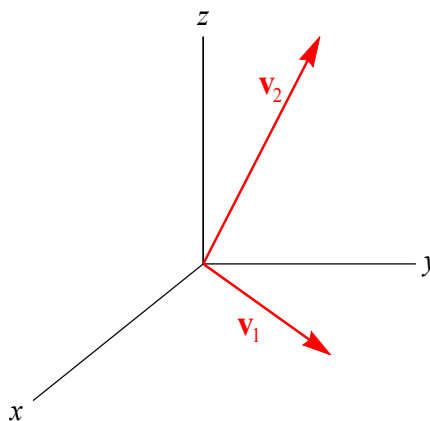
#### EXAMPLE 5

$V = \mathbb{R}^3$ ;  $\mathbf{v}_1 = (1, 0, 0)$ ,  $\mathbf{v}_2 = (0, 1, 0)$



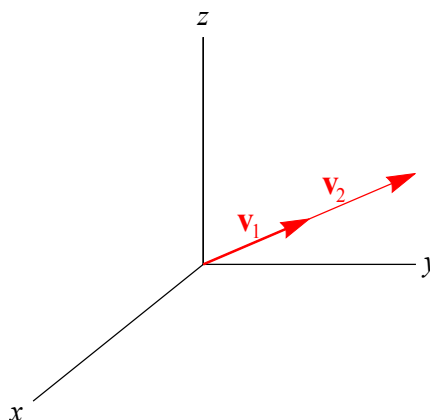
## EXAMPLE 6

$$V = \mathbb{R}^3; \mathbf{v}_1 = (1, 2, -1), \mathbf{v}_2 = (-2, 1, 3)$$



## EXAMPLE 7

$$V = \mathbb{R}^3; \mathbf{v}_1 = (1, 3, 1), \mathbf{v}_2 = (2, 6, 2)$$



**Theorem 3.8 (Spans are subspaces)** Let  $V$  be a real vector space. Then the span of any collection of vectors in  $V$  is a subspace of  $V$ .

PROOF HW (this is similar to the proof of Theorem 3.5).

*Hint:* If  $w$  is a generic element of  $\text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_n)$ , then  $w$  can be written down as

$$\sum_{j=1}^n c_j \mathbf{v}_j, \text{ where } c_j \in \mathbb{R} \text{ for all } j.$$

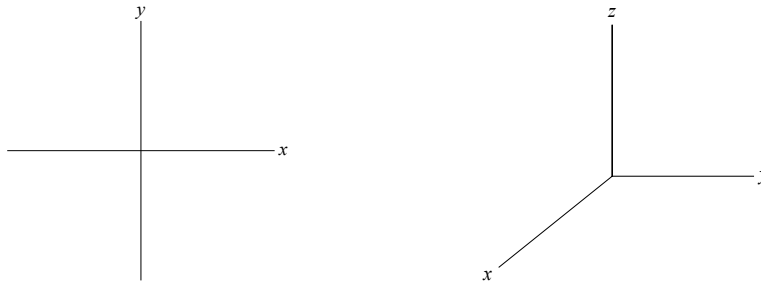
### 3.3 Distinguishing between subspaces

#### QUESTIONS

1. In a general sense, what kinds of sets are subspaces?
2. How do you distinguish between vector spaces?
3. How do you distinguish between different subspaces of a vector space?

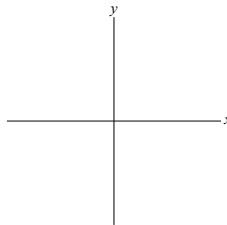
#### FIRST EXAMPLE

What is the difference between  $\mathbb{R}^2$  and  $\mathbb{R}^3$ ?



#### Some questions related to these observations:

1. Is it possible to span  $\mathbb{R}^2$  with only 1 vector?
2. Is it possible to span  $\mathbb{R}^3$  with less than 3 vectors?
3.  $\mathbb{R}^2$  is also  $\text{Span}((1, 0), (0, 1), (1, 1))$ :

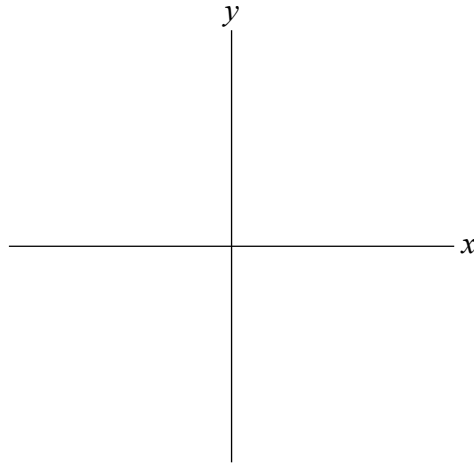


If  $W = \text{Span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \dots, \mathbf{v}_n)$ , which  $\mathbf{v}_j$  are needed? Which are not?

4.  $\mathbb{R}^2$  is also spanned by lots of other sets of two vectors. Is  $\mathbb{R}^2$  spanned by any set of two vectors in  $\mathbb{R}^2$ ?

## SECOND EXAMPLE

Let  $V = \mathbb{R}^2$  and let  $W_1 = \text{Span}(1, 0)$ ;  $W_2 = \text{Span}(1, 1)$ ;  $W_3 = \text{Span}(-4, 0)$ .



But what does “going in the same direction” mean, in general?

**Definition 3.9** Let  $V$  be a real vector space. Two vectors  $\mathbf{v}$  and  $\mathbf{w}$  in  $V$  are called **parallel** (denoted  $\mathbf{v} \parallel \mathbf{w}$ ) if one is a scalar multiple of another, i.e.

$$\mathbf{v} \parallel \mathbf{w} \iff (\exists c \in \mathbb{R} \text{ s.t. } c\mathbf{v} = \mathbf{w}) \text{ or } (\exists c \in \mathbb{R} \text{ s.t. } c\mathbf{w} = \mathbf{v}).$$

Two vectors  $\mathbf{v}$  and  $\mathbf{w}$  are **in the same direction** if the  $c$  in the above definition can be taken to be greater than or equal to zero.

**Some mathematical shorthand:**

- $\iff$  means “if and only if”
- $\Rightarrow$  means “therefore”
- “s.t.” means “such that”
- $\exists$  is short for “there exists”
- $\forall$  is short for “for all”

## EXAMPLE 8

$$V = \mathbb{R}^2; \mathbf{v} = (2, -3); \mathbf{w} = (-4, 6).$$

## EXAMPLE 9

$$V = \mathbb{R}^2; \mathbf{v} = (0, 0); \mathbf{w} = (1, 2).$$

## EXAMPLE 10

$$V = M_2(\mathbb{R}); A = \begin{pmatrix} 2 & -3 \\ 0 & 1 \end{pmatrix}; B = \begin{pmatrix} 8 & -12 \\ 0 & 4 \end{pmatrix}.$$

The following characterization of parallelism is important because it will generalize to collections of more than two vectors.

**Theorem 3.10** *Let  $\mathbf{v}, \mathbf{w} \in V$ , where  $V$  is a real vector space. Then*

$$\mathbf{v} \parallel \mathbf{w} \iff \exists c_1, c_2 \in \mathbb{R} \text{ with } c_1, c_2 \text{ not both zero,} \\ \text{so that } c_1\mathbf{v} + c_2\mathbf{w} = \mathbf{0}.$$

PROOF ( $\Rightarrow$ ) Assume  $\mathbf{v} \parallel \mathbf{w}$ . Then

$$\begin{aligned} \mathbf{v} &= c\mathbf{w} \text{ for some scalar } c \\ \Rightarrow \mathbf{v} - c\mathbf{w} &= \mathbf{0} \\ \Rightarrow 1\mathbf{v} + (-c)\mathbf{w} &= \mathbf{0} \\ \Rightarrow c_1\mathbf{v} + c_2\mathbf{w} &= \mathbf{0} \text{ (where } c_1 = 1, c_2 = -c). \end{aligned}$$

( $\Leftarrow$ ) Suppose  $c_1\mathbf{v} + c_2\mathbf{w} = \mathbf{0}$  where either  $c_1$  or  $c_2$  is nonzero.

If  $c_1 \neq 0$ , divide through by  $c_1$  to get

$$\mathbf{v} + \frac{c_2}{c_1}\mathbf{w} = \mathbf{0} \Rightarrow \mathbf{v} = \frac{-c_2}{c_1}\mathbf{w} \Rightarrow \mathbf{v} \parallel \mathbf{w}.$$

If  $c_2 \neq 0$ , divide through by  $c_2$  to get

$$\frac{c_1}{c_2}\mathbf{v} + \mathbf{w} = \mathbf{0} \Rightarrow \mathbf{w} = \frac{-c_1}{c_2}\mathbf{v} \Rightarrow \mathbf{v} \parallel \mathbf{w}. \quad \square$$

## 3.4 Affine subspaces

**Definition 3.11** Let  $V$  be a real vector space. An **affine subspace**  $A$  of  $V$  is a translate of a subspace, i.e.  $A$  is an affine subspace of  $V$  if there exists a vector  $\mathbf{p} \in V$  and a subspace  $W \subseteq V$  such that

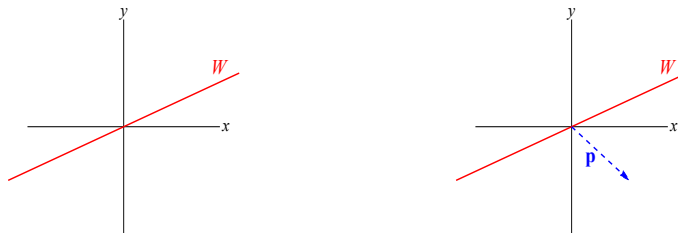
$$A = \mathbf{p} + W = \{\mathbf{p} + \mathbf{w} : \mathbf{w} \in W\}.$$

In this setting, the vector  $\mathbf{p}$  is called a **translation vector** for  $A$ , and the subspace  $W$  is called the **associated subspace** of  $A$ .

### FIRST EXAMPLES

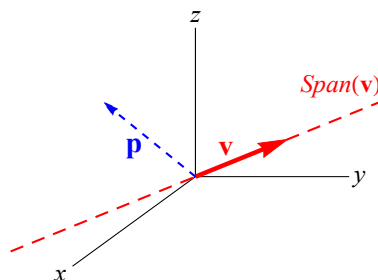
- For any vector space  $V$ ,  $V$  is a subspace of itself (hence an affine subspace of itself).
- For any vector space  $V$ , any set  $\{\mathbf{v}\}$  consisting of exactly one vector  $\mathbf{v} \in V$  constitutes an affine subspace called a **point** (since any such set is the translation of the subspace  $\{\mathbf{0}\}$  by the translation vector  $\mathbf{v}$ ).

**Note:** “Affine” is not an adjective which describes some subspaces. In fact, subspaces are special kinds of affine subspaces, not the other way around:



### Lines

**Definition 3.12** Let  $V$  be a real vector space. A **line**  $\Lambda$  in  $V$  is an affine subspace whose associated subspace is the span of a single nonzero vector  $\mathbf{v} \in V$ .  $\mathbf{v}$  is called a **direction vector** for the line.



**Definition 3.13** Let  $\Lambda$  be a line in vector space  $V$ . A set of **parametric equations** for  $\Lambda$  is a coordinate-wise version of the equation  $\mathbf{x} = \mathbf{p} + t\mathbf{v}$ , where  $\mathbf{p} \in \Lambda$  and  $\mathbf{v}$  is a direction vector of  $\Lambda$ .

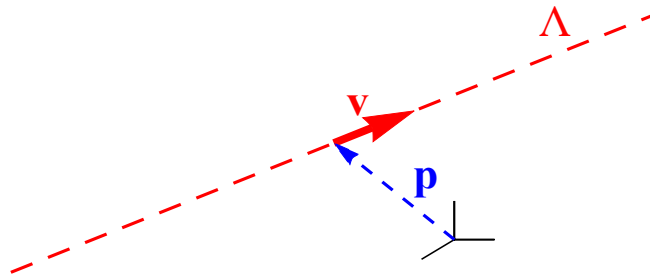
**Note:** parametric equations for a line (or other affine subspace) are **NEVER** unique, because they depend on a choice of  $\mathbf{p}$  and  $\mathbf{v}$ .

EXAMPLE 11

Find parametric equations for the line which passes through  $(4, -1, 3)$  and  $(2, 5, -1)$  in  $\mathbb{R}^3$ .

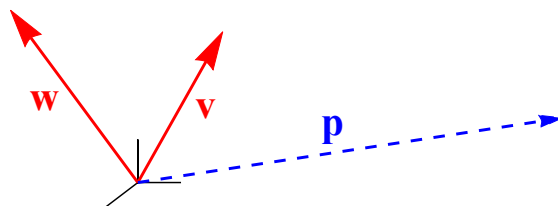
Parametric equations for a line “lay a  $t$ -axis” on that line, i.e. **coordinatize** the line where

$t = 0 \leftrightarrow$  choice of translation vector  $\mathbf{p}$   
 one unit of  $t \leftrightarrow$  choice of direction vector  $\mathbf{v}$



## Planes

**Definition 3.14** Let  $V$  be a real vector space. A **plane**  $\Pi$  in  $V$  is an affine subspace whose associated subspace is the span of two nonparallel vectors  $\mathbf{v}, \mathbf{w} \in V$ .



**Definition 3.15** Let  $\Pi$  be a plane in vector space  $V$ . A set of **parametric equations** for  $\Pi$  is a coordinate-wise version of the equation  $\mathbf{x} = \mathbf{p} + s\mathbf{v} + t\mathbf{w}$ , where  $\mathbf{p} \in \Pi$  and  $\mathbf{v}$  and  $\mathbf{w}$  span the subspace associated to  $\Pi$ .

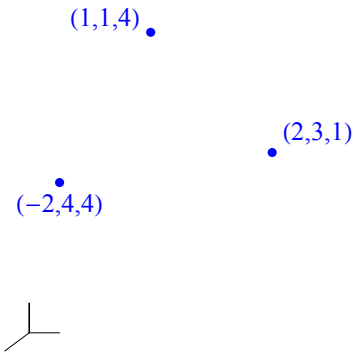
### EXAMPLE 12

Find parametric equations of the plane in  $\mathbb{R}^3$  containing the point  $(3, 2, -7)$  and whose associated subspace is spanned by  $(1, 5, -1)$  and  $(2, 0, -3)$ .



## EXAMPLE 13

Find parametric equations of the plane in  $\mathbb{R}^3$  passing through  $(1, 1, 4)$ ,  $(2, 3, 1)$  and  $(-2, 4, 4)$ .



Parametric equations for a plane  $\Pi$  “lay an  $s, t$ -plane” on  $\Pi$  where

$(s, t)$  origin  $\leftrightarrow$  choice of translation vector  $\mathbf{p}$

$s$ -axis  $\leftrightarrow$  choice of  $\mathbf{v}$

$t$ -axis  $\leftrightarrow$  choice of  $\mathbf{w}$

The  $s$ - and  $t$ -axes need not be  $\perp$ :



## Affine subspaces of $\mathbb{R}$ , $\mathbb{R}^2$ and $\mathbb{R}^3$

### QUESTION

Can we completely list all the subspaces and affine subspaces of a given vector space  $V$ ?

**Answer:** In most cases, yes; for now, we do this when  $V$  is  $\mathbb{R}$ ,  $\mathbb{R}^2$  or  $\mathbb{R}^3$ .

**Theorem 3.16** *The only subspaces of  $\mathbb{R}$  are the zero subspace  $\{0\}$  and the entire real number line  $\mathbb{R}$ .*

**PROOF** First, both  $\{0\}$  and  $\mathbb{R}$  are obviously subspaces of  $\mathbb{R}$ .

So we need to show that there aren't any other subspaces of  $\mathbb{R}$ .

To do this, let  $W \subseteq \mathbb{R}$  be a subspace. There are two cases:

*Case 1:*  $W = \{0\}$ .

*Case 2:*  $W \neq \{0\}$ . In this situation, there is a  $w \neq 0$  s.t.  $w \in W$ .

Then  $rw \in W$  for all  $r \in \mathbb{R}$ , so all real numbers are in  $W$ . Thus  $W = \mathbb{R}$ .  $\square$

**Corollary 3.17** *The only affine subspaces of  $\mathbb{R}$  are points and the entire real number line  $\mathbb{R}$ .*

**PROOF** Let  $A$  be an affine subspace of  $\mathbb{R}$ .

Then  $A = p + W$  where  $W$  is a subspace of  $\mathbb{R}$  and  $p \in \mathbb{R}$ .

By the previous theorem, there are two cases:

*Case 1:*  $W = \{0\}$ . Then  $A = \{p\}$ , a point.

*Case 2:*  $W = \mathbb{R}$ . Then  $A = p + \mathbb{R} = \mathbb{R}$ .  $\square$

**Theorem 3.18** *The only subspaces of  $\mathbb{R}^2$  are the zero subspace  $\{0\}$ , lines passing through the origin, and all of  $\mathbb{R}^2$ . Consequently, the only affine subspaces of  $\mathbb{R}^2$  are points, lines, and  $\mathbb{R}^2$ .*

**PROOF** It is clear that points, lines and all of  $\mathbb{R}^2$  are affine subspaces.

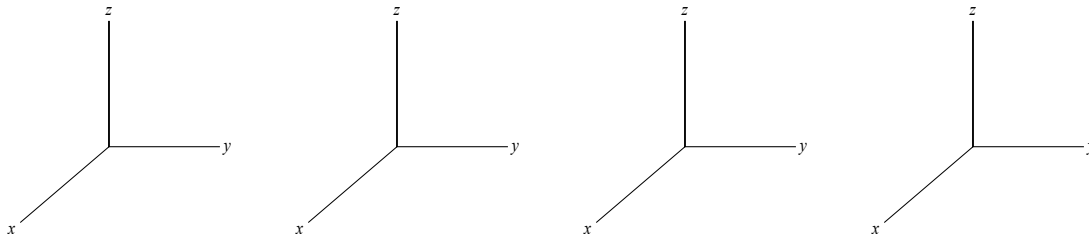
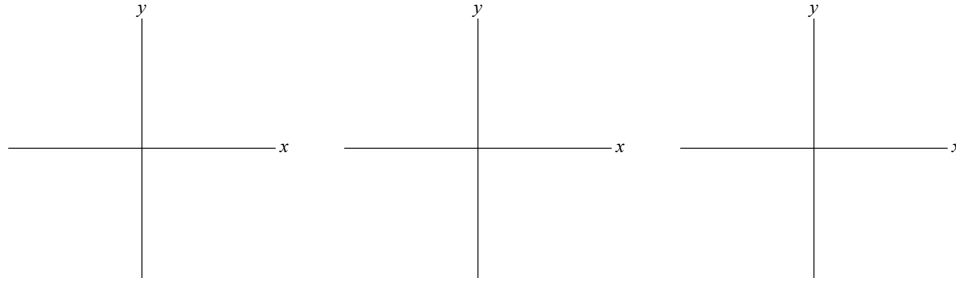
We have to prove there aren't any other affine subspaces; we'll do this later.  $\square$

**Corollary 3.19** *The only subspaces of  $\mathbb{R}^3$  are the zero subspace  $\{0\}$ , lines passing through the origin, planes passing through the origin, and all of  $\mathbb{R}^3$ . Consequently, the only affine subspaces of  $\mathbb{R}^3$  are points, lines, planes and  $\mathbb{R}^3$  itself.*

**PROOF** It is clear that points, lines, planes, and all of  $\mathbb{R}^3$  are affine subspaces.

We have to prove there aren't any other affine subspaces; we'll do this later.  $\square$

**Generic pictures of subspaces:**



These pictures reflect the idea that subspaces contain the zero vector and are flat, convex, and unbounded.

Generally speaking, when you are told “ $W$  is a subspace of  $V$ ”, think of  $W$  as being something like a line or plane passing through the origin.

## 3.5 Linear independence

### MOTIVATING EXAMPLE

Let  $V = \mathbb{R}^n$ ; take two vectors  $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$  and let  $W = \text{Span}(\mathbf{v}, \mathbf{w})$ . What kind of object is  $W$ ?

Classifying these cases depends on whether or not  $\mathbf{v}$  and  $\mathbf{w}$  are parallel.

## Linear dependence

### NEW QUESTION

What is the right notion of “parallelism” for a family of  $n > 2$  vectors  $\mathbf{v}_1, \dots, \mathbf{v}_n$ ?

**Recall:** Theorem 3.10 from earlier gives an equivalent characterization of parallelism of two vectors:

$$\mathbf{v}_1 \parallel \mathbf{v}_2 \iff \exists c_1, c_2 \in \mathbb{R} \text{ (not both zero) s.t. } c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 = \mathbf{0}.$$

EXAMPLE  $(2, 1) \parallel (4, 2)$  since

The equation  $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = \mathbf{0}$  from Theorem 3.10 generalizes naturally to sets of more than two vectors:

**Definition 3.20** *Let  $V$  be a real vector space. A collection of vectors  $\mathbf{v}_1, \dots, \mathbf{v}_n \in V$  is called **linearly dependent (lin. dep.)** if  $\exists c_1, \dots, c_n \in \mathbb{R}$ , with not all  $c_j = 0$ , such that  $c_1\mathbf{v}_1 + \dots + c_n\mathbf{v}_n = \mathbf{0}$ .*

Here is an equivalent formulation of the same idea:

**Definition 3.21** *Let  $V$  be a real vector space, and let  $\mathbf{v}_1, \dots, \mathbf{v}_n \in V$ . Any expression of the form  $c_1\mathbf{v}_1 + \dots + c_n\mathbf{v}_n$  is called a **linear combination** of the  $\mathbf{v}_j$ . The expression  $0\mathbf{v}_1 + \dots + 0\mathbf{v}_n$  is called the **trivial combination** (this combination always equals  $\mathbf{0}$ ). Any other combination (other than the trivial one) is called a **nontrivial combination** of the  $\mathbf{v}_j$  otherwise (i.e. if not all the  $c_j$  are zero). A set of vectors is **linearly dependent** if and only if they have a nontrivial combination which equals  $\mathbf{0}$ .*

The idea behind this definition is that a lin. dep. set of vectors “repeats the same direction unnecessarily”, in the same way that two parallel vectors do. This idea is made precise in the upcoming Theorems 3.22 and 3.23.

EXAMPLE 14

In  $\mathbb{R}^3$ , set  $\mathbf{v}_1 = (1, 0, 0)$ ;  $\mathbf{v}_2 = (0, 1, 0)$ ;  $\mathbf{v}_3 = (1, 2, 0)$ .

**Theorem 3.22** *Let  $V$  be a real vector space.*

A collection of vectors  $\mathbf{v}_1, \dots, \mathbf{v}_n \in V$  is linearly dependent  $\iff$  for some  $k$ , we can write  $\mathbf{v}_k$  as a linear combination of the previous vectors in the list, i.e.

$$\mathbf{v}_k = \sum_{j=1}^{k-1} d_j \mathbf{v}_j = d_1 \mathbf{v}_1 + d_2 \mathbf{v}_2 + \dots + d_{k-1} \mathbf{v}_{k-1}.$$

PROOF ( $\Rightarrow$ ) Assume the  $\mathbf{v}_j$  are lin. dep. Then

$$c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \dots + c_n \mathbf{v}_n = \mathbf{0}$$

where  $c_1, \dots, c_n \in \mathbb{R}$  with not all the  $c_j$  zero.

Let  $k$  be the largest subscript so that  $c_k \neq 0$ . Then

$$\begin{aligned} c_1 \mathbf{v}_1 + \dots + c_{k-1} \mathbf{v}_{k-1} + c_k \mathbf{v}_k &= \mathbf{0} \\ \Rightarrow c_k \mathbf{v}_k &= -c_1 \mathbf{v}_1 - \dots - c_{k-1} \mathbf{v}_{k-1} \\ \Rightarrow \mathbf{v}_k &= \frac{-c_1}{c_k} \mathbf{v}_1 - \dots - \frac{c_{k-1}}{c_k} \mathbf{v}_{k-1} \\ \Rightarrow \mathbf{v}_k &= d_1 \mathbf{v}_1 + \dots + d_{k-1} \mathbf{v}_{k-1}. \end{aligned}$$

( $\Leftarrow$ ) Suppose  $\mathbf{v}_k = d_1 \mathbf{v}_1 + \dots + d_{k-1} \mathbf{v}_{k-1}$ . Then

$$d_1 \mathbf{v}_1 + \dots + d_{k-1} \mathbf{v}_{k-1} - 1 \mathbf{v}_k + 0 \mathbf{v}_{k+1} + \dots + 0 \mathbf{v}_n = \mathbf{0}$$

so  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  lin. dep. by definition.  $\square$

Why do we care about whether or not a set of vectors is linearly dependent? Because linear dependence has a lot to do with what the span of these vectors is. Consider the following theorem:

**Theorem 3.23 (Removing lin. dep. vectors doesn't change span)** *Let  $V$  be a real vector space. Let  $\mathbf{v}_1, \dots, \mathbf{v}_n$  be a collection of linearly dependent vectors such that*

$$\mathbf{v}_k = d_1 \mathbf{v}_1 + \dots + d_{k-1} \mathbf{v}_{k-1}.$$

*Then  $\text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_n) = \text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n)$ .*

*In other words,  $\mathbf{v}_k$  can be removed from the list without changing the span of the vectors.*

PROOF Let  $W_1 = \text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_n)$  and let  $W_2 = \text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n)$ .

We will show  $W_1 = W_2$  by showing each set is a subset of the other:

( $\supseteq$ ): Suppose  $\mathbf{w} \in W_2$ . Then by definition of span,

$$\mathbf{w} = \sum_{j=1, j \neq k}^n c_j \mathbf{v}_j = \sum_{j=1}^n c_j \mathbf{v}_j \text{ (by setting } c_k = 0)$$

So by definition,  $\mathbf{w} \in \text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_n)$  so  $\mathbf{w} \in W_1$ .

( $\subseteq$ ): Suppose  $\mathbf{w} \in W_1$ . Then by definition of span,

$$\begin{aligned} \mathbf{w} &= \sum_{j=1}^n c_j \mathbf{v}_j = \sum_{j=1, j \neq k}^n c_j \mathbf{v}_j + c_k \mathbf{v}_k \\ &= \sum_{j=1, j \neq k}^n c_j \mathbf{v}_j + c_k \sum_{j=1}^{k-1} d_j \mathbf{v}_j \\ &= \sum_{j=1}^{k-1} (c_j + c_k d_j) \mathbf{v}_j + \sum_{j=k+1}^n c_j \mathbf{v}_j. \end{aligned}$$

Thus  $\mathbf{w} \in W_2$  by definition of span.

Since  $W_1 \subseteq W_2$  and  $W_2 \subseteq W_1$ ,  $W_1 = W_2$ .  $\square$

## Linear independence

If a collection of vectors are not linearly dependent, then they are called *linearly independent*:

**Definition 3.24** Let  $V$  be a real vector space. A collection of vectors  $\mathbf{v}_1, \dots, \mathbf{v}_n \in V$  is called **linearly independent (lin. ind.)** if the collection is not linearly dependent.

Equivalently, the vectors are lin. ind. if the only combination of them that makes  $\mathbf{0}$  is the trivial one.

Equivalently, the vectors are lin. ind. if  $c_1 \mathbf{v}_1 + \dots + c_n \mathbf{v}_n = \mathbf{0}$  implies  $c_1 = c_2 = \dots = c_n = 0$ .

**Theorem 3.25 (Elementary properties of lin. dep. and lin. ind. sets)** *Let  $V$  be a real vector space. Then:*

1. *The zero vector  $\mathbf{0}$  cannot be part of a linearly independent set of vectors.*
2. *A collection of one vector  $\mathbf{v}$  is linearly independent if and only if  $\mathbf{v} \neq \mathbf{0}$ .*
3. *A collection of two vectors  $\mathbf{v}, \mathbf{w}$  is linearly dependent if and only if  $\mathbf{v} \parallel \mathbf{w}$ .*
4. *Any subset of a collection of linearly independent vectors is linearly independent.*
5. *Any collection of vectors containing a linearly dependent subcollection is itself linearly dependent.*

EXAMPLE 15

Let  $\mathbf{v}_1 = (1, 0, 0)$ ;  $\mathbf{v}_2 = (0, 1, 0)$ ;  $\mathbf{v}_3 = (1, 1, 4)$ . Is the set  $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$  of vectors linearly independent, or linearly dependent? Explain.

CONCEPTUAL IDEA	ALGEBRAIC JUSTIFICATION



**EXAMPLE 16**

Let  $\mathbf{v}_1 = (1, -1, 2, 5)$ ;  $\mathbf{v}_2 = (0, 2, -1, 4)$ ;  $\mathbf{v}_3 = (-3, 0, -1, 2)$ . Is the set  $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$  of vectors linearly independent, or linearly dependent? Explain.

## 3.6 Basis and dimension

**Definition 3.26** Let  $V$  be a real vector space (or a subspace of some other real vector space). A collection of vectors  $\mathbf{v}_1, \dots, \mathbf{v}_n \in V$  is called a **basis** of  $V$  if

1.  $\mathbf{v}_1, \dots, \mathbf{v}_n$  are *linearly independent*; and
2.  $V = \text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_n)$ .

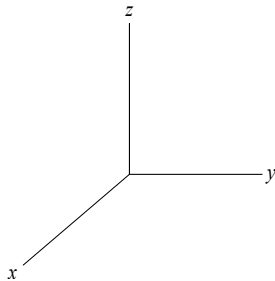
**Note:** Regarding part (2) of this definition, it is always true that  $\text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_n) \subseteq V$ . The question is whether every  $\mathbf{v} \in V$  is a linear combination of the  $\mathbf{v}_j$ .

**Heuristic:** Statement 1 says that no “directions” in  $V$  are repeated by the  $\mathbf{v}_j$  (i.e. that the set of the  $\mathbf{v}_j$ 's isn't unnecessarily big). Statement 2 says that no “directions” are “missed” by the  $\mathbf{v}_j$  (i.e. that the set of the  $\mathbf{v}_j$ 's isn't too small to capture all of  $V$ ).

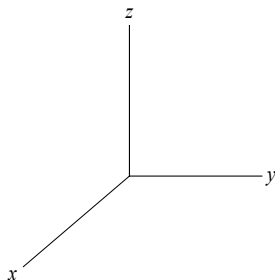
## EXAMPLE 17

Determine whether each given set of vectors forms a basis of  $\mathbb{R}^3$ .

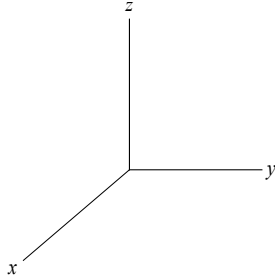
1.  $\{\mathbf{v}_1, \mathbf{v}_2\}$ , where  $\mathbf{v}_1 = (1, 0, 0)$ ;  $\mathbf{v}_2 = (0, 1, 0)$ .



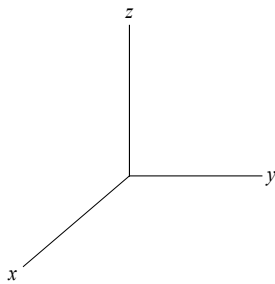
2.  $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ , where  $\mathbf{v}_1 = (1, 0, 0)$ ;  $\mathbf{v}_2 = (0, 1, 0)$ ;  $\mathbf{v}_3 = (0, 0, 1)$ ;  $\mathbf{v}_4 = (1, -1, 2)$ .



3.  $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ , where  $\mathbf{v}_1 = (1, 0, 0)$ ;  $\mathbf{v}_2 = (0, 1, 0)$ ;  $\mathbf{v}_3 = (1, 2, 0)$ .



4.  $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ , where  $\mathbf{v}_1 = (1, 0, 0)$ ;  $\mathbf{v}_2 = (0, 1, 0)$ ;  $\mathbf{v}_3 = (0, 0, 1)$ .



A vector space (or subspace) is completely described by giving a basis of that space, because of the following theorem:

**Theorem 3.27 (Unique Representation Theorem)** *Let  $V$  be a real vector space and let  $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  be a basis of  $V$ . Then, for every  $\mathbf{v} \in V$  there is **exactly one choice** of  $c_1, \dots, c_n \in \mathbb{R}$  such that*

$$\mathbf{v} = \sum_{j=1}^n c_j \mathbf{v}_j.$$

*Equivalently, there is exactly one linear combination of the basis vectors that makes each  $\mathbf{v} \in V$ .*

PROOF Since  $\mathcal{B}$  is a basis,  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  spans  $V$ . By definition of span, every  $\mathbf{v} \in V$  can be written as

$$\mathbf{v} = \sum_{j=1}^n c_j \mathbf{v}_j = c_1 \mathbf{v}_1 + \dots + c_n \mathbf{v}_n$$

in *at least one* way.

Now suppose there are two combinations of the vectors that make  $\mathbf{v}$ , i.e.

$$\mathbf{v} = \sum_{j=1}^n c_j \mathbf{v}_j = \sum_{j=1}^n d_j \mathbf{v}_j.$$

Subtracting the right-hand side from both sides, we get

$$\sum_{j=1}^n c_j \mathbf{v}_j - \sum_{j=1}^n d_j \mathbf{v}_j = \mathbf{0}$$

so

$$\sum_{j=1}^n (c_j - d_j) \mathbf{v}_j = \mathbf{0}.$$

But since  $\mathcal{B}$  is a basis,  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  are linearly independent!

So the left-hand side of the above equation has to be the trivial combination.

This means  $c_j - d_j = 0$  for all  $j$ , i.e.  $c_j = d_j$  for all  $j$ .

Therefore, the two combinations we started with **weren't actually different**.

This means  $\mathbf{v}$  can't be written as a linear combination of the  $\mathbf{v}_j$  in more than one way.  $\square$

#### EXAMPLE 18

One basis of  $\mathbb{R}^3$  is  $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ . Take the vector  $\mathbf{v} = (-2, 4, 3)$ .

## The Exchange Lemma

We now come to the most important theoretical principle of linear algebra, upon which the rest of the subject (including all applications) rests:

**Theorem 3.28 (Exchange Lemma)** *Let  $V$  be a real vector space. If  $\{\mathbf{v}_1, \dots, \mathbf{v}_m\}$  spans  $V$ , and if  $\{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  is linearly independent in  $V$ , then  $m \geq n$ .*

**In English:** Spanning sets must contain at least as many vectors as linearly independent sets.

PROOF We are given that  $V = \text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_m)$ .

Since  $\mathbf{w}_1 \in V = \text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_m)$ . Therefore the set

$$\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m, \mathbf{w}_1\}$$

is linearly dependent (since the last vector is in the span of the others).

Writing these vectors in a different order doesn't change the fact they are lin. dep., so

$$\{\mathbf{w}_1, \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$$

is also lin. dep.

Now discard the first  $\mathbf{v}_j$  in this list which depends on the previous vectors.

Discarding this  $\mathbf{v}_j$  doesn't change the span of the vectors in this list, so we have

$$V = \text{Span}(\mathbf{w}_1, \mathbf{v}_1, \dots, \mathbf{v}_m) = \text{Span}(\mathbf{w}_1, \mathbf{v}_1, \dots, \mathbf{v}_{j-1}, \mathbf{v}_{j+1}, \dots, \mathbf{v}_m).$$

Essentially, what we have done above is "exchange" one of the  $\mathbf{v}_j$ s for  $\mathbf{w}_1$  without changing the span of the set.

This technique is why this theorem is called the Exchange Lemma.

Now we “exchange” again.  $\mathbf{w}_2 \in V$ , so  $\mathbf{w}_2 \in \text{Span}(\mathbf{w}_1, \mathbf{v}_1, \dots, \mathbf{v}_{j-1}, \mathbf{v}_{j+1}, \dots, \mathbf{v}_m)$  so  $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{v}_1, \dots, \mathbf{v}_{j-1}, \mathbf{v}_{j+1}, \dots, \mathbf{v}_m\}$  is lin. dep.

Remove the first dependent vector  $\mathbf{v}_k$  in this list to obtain

$$\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{v}_1, \dots, \mathbf{v}_{j-1}, \mathbf{v}_{j+1}, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_m\}$$

which still spans  $V$ .

Keep exchanging in this fashion.

**If  $n > m$ ,** then we can exchange all the  $\mathbf{v}$ s with  $\mathbf{w}$ s to obtain a set

$$\{\mathbf{w}_1, \dots, \mathbf{w}_m\}$$

which spans  $V$ , and still have  $\mathbf{w}_{m+1}$  leftover. **But**

$$\mathbf{w}_{m+1} \in V = \text{Span}(\mathbf{w}_1, \dots, \mathbf{w}_m),$$

so  $\{\mathbf{w}_1, \dots, \mathbf{w}_{m+1}\}$  is lin. dep. (since the last vector is in the span of the others).

**This contradicts the hypothesis** that  $\{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  is lin. ind.

Thus  $n \leq m$ , proving the Exchange Lemma.  $\square$

## Dimension

**Corollary 3.29 (Dimension Theorem)** *Let  $V$  be a real vector space which has a basis consisting of  $n$  vectors. Then any other basis of  $V$  must also consist of  $n$  vectors.*

PROOF Suppose  $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  and  $\mathcal{B}' = \{\mathbf{w}_1, \dots, \mathbf{w}_m\}$  are two bases of  $V$ .

$\mathcal{B}$  is spanning set;  $\mathcal{B}'$  is lin. indep  $\Rightarrow n \geq m$  by Exchange Lemma.

$\mathcal{B}$  is lin. indep.;  $\mathcal{B}'$  is spanning set  $\Rightarrow m \geq n$  by Exchange Lemma.

Thus  $m = n$ .  $\square$

**Definition 3.30** *Let  $V$  be a real vector space. If  $V$  is spanned by a finite set of vectors, we say  $V$  is **finite dimensional** and write  $\dim V < \infty$ .*

*In this case, the **dimension** of  $V$  is the number of elements in any basis of  $V$ .*

*(We define  $\dim(\{\mathbf{0}\}) = 0$  even though  $\{\mathbf{0}\}$  does not have a basis.)*

*If  $V$  is not spanned by any finite set of vectors, we say  $V$  is **infinite dimensional** and write  $\dim V = \infty$ .*

**Definition 3.31** Let  $V$  be a real vector space, and  $A \subseteq V$  be an affine subspace of  $V$ .

The **dimension** of  $A$  is the dimension of the subspace  $W = A - \mathbf{p}$ , where  $\mathbf{p}$  is any translation vector for  $A$ .

In particular, a **point** is an affine subspace of dimension zero; a **line** is an affine subspace of dimension one; and a **plane** is an affine subspace of dimension two.

A **hyperplane** is an affine subspace whose dimension is one less than the dimension of  $V$ .

EXAMPLE 19

Find a basis for, and the the dimension of, of each of the following vector spaces:

1.  $\mathbb{R}^n$

2.  $M_n(\mathbb{R})$ , the set of  $n \times n$  matrices with entries in  $\mathbb{R}$ .

3.  $W =$  the set of  $3 \times 3$  diagonal matrices with entries in  $\mathbb{R}$ .

4. The set  $\mathbb{R}[x]$  of polynomials with real coefficients.



## Theoretical consequences of the Exchange Lemma

**Theorem 3.32 (Spanning Set Theorem)** *If  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  spans vector space  $V$ , then some subset of  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  forms a basis of  $V$  (so  $n \geq \dim V$ ).*

**Consequence:** No two vectors span  $\mathbb{R}^3$ ; no  $m$  vectors span  $\mathbb{R}^n$  if  $n > m$ , etc.

PROOF Start with spanning set  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ .

If the vectors are lin. indep., they form a basis.

Otherwise, they are lin. dep., so there is a first vector in the list which depends on the preceding vectors. Discard this vector from the list.

If what is left is lin. indep., then it forms a basis; otherwise, keep discarding vectors which depend on the others until what is left is lin. indep.

Since discarding dependent vectors doesn't change the span, what's left will span  $V$  and be linearly independent, hence form a basis.  $\square$

**Theorem 3.33 (Linearly Independent Set Theorem)** *If  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  is a linearly independent set of vectors in vector space  $V$ , then  $n \leq \dim V$ .*

**Consequence:** No four vectors in  $\mathbb{R}^3$  are linearly independent; no  $m$  vectors in  $\mathbb{R}^n$  are linearly independent if  $m > n$ , etc.

PROOF Follows immediately from the Exchange Lemma, since there are  $\dim V$  vectors in a basis of  $V$  (which spans  $V$ ).  $\square$

**Theorem 3.34 (Basis Extension Theorem)** *Let  $V$  be a finite dimensional real vector space. If  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  is a linearly independent set of vectors in  $V$ , then  $V$  has a basis of the form*

$$\{\mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{w}_1, \dots, \mathbf{w}_m\}$$

(where  $\dim(V) = m + n$ ).

**Note:**  $m$  could equal 0 in this theorem (if the  $\mathbf{v}_j$ s already form a basis).

PROOF Let  $W_0 = \text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_n)$ .

If  $W_0 = V$ , then we are done ( $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  forms a basis of  $V$ ).

Otherwise, there is  $\mathbf{w}_1 \notin W_0$ .

Thus  $\{\mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{w}_1\}$  is lin. ind. (since the first  $n$  vectors in this list are lin. ind. and the last vector is not a linear combination of the first  $n$ ).

Let  $W_1 = \text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{w}_1)$ .

If  $W_1 = V$ , then we are done ( $\{\mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{w}_1\}$  forms a basis of  $V$ ).

Otherwise, there is  $\mathbf{w}_2 \notin W_1$ .

Thus  $\{\mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{w}_1, \mathbf{w}_2\}$  is lin. ind.

Let  $W_2 = \text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{w}_1, \mathbf{w}_2)$ .

If  $W_2 = V$ , then we are done ( $\{\mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{w}_1\}$  forms a basis of  $V$ ).

Otherwise, continue this process.

Eventually, if  $\dim V = m+n$ , we get a list of lin. ind. vectors  $\{\mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{w}_1, \dots, \mathbf{w}_m\}$ .

Let  $W_m$  be the span of these vectors.

If  $W_m \neq V$ , then there is  $\mathbf{w}_{m+1} \notin W_m$  that could be added to this collection to produce a list of  $n+m+1 > \dim V$  lin. ind. vectors, contradicting the Linearly Independent Set Theorem.  $\square$

**Theorem 3.35 (Basis Theorem)** *Let  $V$  be a vector space with  $n = \dim V < \infty$ . Then:*

1.  $V$  has a basis (so long as  $V \neq \mathbf{0}$ ).
2. Any set of  $n$  linearly independent vectors in  $V$  forms a basis of  $V$ .
3. Any set of  $n$  vectors which span  $V$  forms a basis of  $V$ .

**PROOF** For (1), if  $V \neq \mathbf{0}$  choose a nonzero  $\mathbf{v} \in V$ .  $\{\mathbf{v}\}$  is linearly independent; by the Basis Extension Theorem it can be extended to a basis of  $V$ .

For (2), if the set of linearly independent vectors  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  isn't a basis, then these vectors don't span  $V$ , so there is  $\mathbf{w} \in V$  which is not in their span. Thus  $\{\mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{w}\}$  is a set of  $n+1 > \dim V$  lin. ind. vectors in  $V$ , contradicting the Linearly Independent Set Theorem.

For (3), if the spanning set  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  isn't a basis, then these vectors are lin. dep. Remove all the dependent vectors from this set; this leaves a set of  $< n$  vectors which still span  $V$  (since removing dependent vectors doesn't change the span of a set). This contradicts the Spanning Set Theorem.  $\square$

**Theorem 3.36** *If  $W$  is a subspace of real vector space  $V$ , then  $\dim W \leq \dim V$ .*

*Also, if  $W$  is a subspace of  $V$  and  $\dim W = \dim V < \infty$ , then  $W = V$ .*

PROOF First, let's prove the first statement. If  $\dim V = \infty$ , the result is trivial.

Otherwise, let  $n = \dim W$  and let  $\{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  be a basis of  $W$ . This basis can be extended to a basis of  $V$ , which must necessarily have  $\geq n$  elements.

Therefore  $\dim V \geq n = \dim W$  as desired.

For the second statement, if  $n = \dim V = \dim W < \infty$ , let  $\{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  be a basis of  $W$ . Thus  $\{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  are a set of  $n$  lin. ind. vectors in  $V$ , hence form a basis of  $V$  by the Basis Theorem. Therefore  $W = \text{Span}(\mathbf{w}_1, \dots, \mathbf{w}_n) = V$ .  $\square$

We can now revisit some ideas we encountered earlier:

**Theorem 3.37 (Classification of subspaces of  $\mathbb{R}$ ,  $\mathbb{R}^2$  and  $\mathbb{R}^3$ )**

- *The only subspaces of  $\mathbb{R}$  are  $\{\mathbf{0}\}$  and  $\mathbb{R}$ .*
- *The only subspaces of  $\mathbb{R}^2$  are  $\{\mathbf{0}\}$ , lines passing through the origin, and  $\mathbb{R}^2$ .*
- *The only subspaces of  $\mathbb{R}^3$  are  $\{\mathbf{0}\}$ , lines passing through the origin, planes passing through the origin, and  $\mathbb{R}^3$ .*

PROOF For statement (3), notice  $\dim \mathbb{R}^3 = 3$ . If  $W \subseteq \mathbb{R}^3$  is a subspace, then  $\dim W \leq 3$ . Since  $\dim W$  is a nonnegative integer, we have four possibilities:

- $\dim W = 0$ . Then  $W$  is a point by definition.
- $\dim W = 1$ . Then  $W$  is a line by definition.
- $\dim W = 2$ . Then  $W$  is a plane by definition.
- $\dim W = 3$ . Then  $W = \mathbb{R}^3$  by Corollary 3.36.

The proofs of the first two statements are similar.  $\square$

**Corollary 3.38 (Classification of affine subspaces of  $\mathbb{R}$ ,  $\mathbb{R}^2$  and  $\mathbb{R}^3$ )**

1. *The only affine subspaces of  $\mathbb{R}$  are points, and all of  $\mathbb{R}$ .*
2. *The only affine subspaces of  $\mathbb{R}^2$  are points, lines and all of  $\mathbb{R}^2$ .*
3. *The only affine subspaces of  $\mathbb{R}^3$  are points, lines, planes and all of  $\mathbb{R}^3$ .*

PROOF This is obvious, given the preceding theorem.  $\square$

## 3.7 Summary of Chapter 3

### Fundamental definitions:

- A **subspace**  $W$  of vector space  $V$  is a subset of  $V$  which is itself a vector space. That means that  $W$  must contain  $\mathbf{0}$ , be closed under addition, and closed under scalar multiplication.
- An **affine subspace**  $A$  of vector space  $V$  is a translate of a subspace of  $V$ .
- The **span** of a set of vectors is the set of linear combinations of those vectors; the span of any set of vectors is always a subspace. A set of vectors **spans** a vector space (or subspace)  $V$  if every vector in  $V$  is a linear combination of vectors in that set.
- A set of vectors  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  is called **linearly independent** if the only way to write  $\mathbf{0} = c_1\mathbf{v}_1 + \dots + c_n\mathbf{v}_n$  is to set all the  $c_j = 0$ . A set of vectors is called **linearly dependent** if it is not linearly independent.
- A set of vectors is called a **basis** of vector space  $V$  if the set spans  $V$  and is linearly independent.
- Any two bases of a vector space  $V$  must have the same number of elements; the count of the number of vectors in any basis of  $V$  is called the **dimension** of  $V$ .

### Fundamental theoretical concepts relating these definitions

- A set of vectors is lin. dep. if and only if there is some vector in the list which is a linear combination of the previous vectors in the list.
- $\mathbf{0}$  is never part of a lin. ind. set of vectors.
- A set of one vector is lin. ind. if and only if the vector is nonzero.
- A set of two vectors is lin. ind. if and only if the vectors are nonparallel.
- A subset of a linearly independent set must also be lin. ind.
- If a set of vectors has a lin. dep. subset, then it is also lin. dep.
- Any set of more than  $\dim V$  vectors in  $V$  must be lin. dep.
- No set of less than  $\dim V$  vectors in  $V$  can span  $V$ .
- Lin. dep. vectors can be removed from a list without changing the span of that list.

- Every spanning set for  $V$  can be reduced to a basis of  $V$  by removing the lin. dep. vectors.
- Any set of  $\dim V$  vectors which span  $V$  must also be a basis of  $V$  (i.e. must also be lin. ind.).
- Any lin. ind. set of vectors in  $V$  can be extended to a basis of  $V$ .
- Any set of  $\dim V$  vectors which are lin. ind. must also be a basis of  $V$  (i.e. must also span  $V$ ).
- Every vector space has a basis.
- Given a basis of  $V$ , every vector in  $V$  can be written as a linear combination of those basis vectors in exactly one way.
- If  $W$  is a subspace of  $V$ , then  $\dim W \leq \dim V$ .
- If  $W$  is a subspace (or affine subspace) of  $V$  and  $\dim W = \dim V < \infty$ , then  $V = W$ .

### Classification of subspaces and affine subspaces

These objects are classified according to their dimension:

- The only subspaces of  $\mathbb{R}$  are  $\{0\}$  (dimension 0) and  $\mathbb{R}$  (dimension 1).  
The only affine subspaces of  $\mathbb{R}$  are points (dimension 0) and  $\mathbb{R}$  (dimension 1).
- The only subspaces of  $\mathbb{R}^2$  are  $\{0\}$  (dimension 0), lines passing through the origin (dimension 1), and all of  $\mathbb{R}^2$  (dimension 2).  
The only affine subspaces of  $\mathbb{R}^2$  are points (dimension 0), lines (dimension 1) and all of  $\mathbb{R}^2$  (dimension 2).
- The only subspaces of  $\mathbb{R}^3$  are  $\{0\}$  (dimension 0), lines passing through the origin (dimension 1), planes passing through the origin (dimension 2), and all of  $\mathbb{R}^3$  (dimension 3).  
The only affine subspaces of  $\mathbb{R}^3$  are points (dimension 0), lines (dimension 1), planes (dimension 2), and all of  $\mathbb{R}^3$  (dimension 3).

In general, you should think of an affine subspace as an object akin to a point, line or plane (but perhaps larger-dimensional), and you should think of a subspace as an affine subspace which passes through  $0$ .

A **hyperplane** in  $V$  is an affine subspace whose dimension is 1 less than the dimension of  $V$ .

## 3.8 Homework exercises for Chapter 3

1. Let  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  be vectors in some vector space  $V$ . Define the **span** of this collection of vectors to be the set of linear combinations of those vectors, i.e.

$$\text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_n) = \{c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_n\mathbf{v}_n : c_1, c_2, \dots, c_n \in \mathbb{R}\}.$$

Prove that the span of any collection of vectors is a subspace.

*Hint:* a generic element of the span is a vector  $\mathbf{w}$  where  $\mathbf{w} = c_1\mathbf{v}_1 + \dots + c_n\mathbf{v}_n$ .

2. Let  $V$  be a vector space and let  $W_1$  and  $W_2$  be two subspaces of  $V$ . Prove that  $W_1 + W_2$  is also a subspace of  $V$ .

*Hint:* Let  $W = W_1 + W_2$ . A generic element of  $W$  is a vector  $\mathbf{w}$  such that  $\mathbf{w} = \mathbf{w}_1 + \mathbf{w}_2$  where  $\mathbf{w}_1 \in W_1$  and  $\mathbf{w}_2 \in W_2$ .

3. Sketch a picture of each of the following subsets of  $\mathbb{R}^2$  (please draw a different picture for each set):

a)  $\text{Span}((2, 3))$

d)  $\text{Span}((4, 2), (2, 1))$

b)  $\text{Span}((-4, -1))$

c)  $\text{Span}((1, 0), (1, 1))$

e)  $\text{Span}((0, 0))$

4. Sketch a picture of each of the following subsets of  $\mathbb{R}^2$  (please draw a different picture for each set):

a)  $(2, -1) + \text{Span}((1, 1))$

c)  $\text{Span}((2, -2)) + \text{Span}((-3, 3))$

b)  $(5, 3) + \text{Span}((-2, 0))$

d)  $\text{Span}((3, 0)) + \text{Span}((1, -1))$

5. Give an example of a subset of  $\mathbb{R}^2$  which contains the zero vector and is closed under scalar multiplication, but is **not** a subspace of  $\mathbb{R}^2$ .

In each part of Problems 6-9, you are given a vector space  $V$  and a subset  $W$  of  $V$ . Determine whether or not  $W$  is a subspace of  $V$ , and write a rigorous proof to support your assertion.

6. a)  $V = \mathbb{R}^3; W = \{(x, y, z) : x + y + z = 0\}$

b)  $V = \mathbb{R}^3; W = \{(x, y, z) : x + y + z = 1\}$

c)  $V = \mathbb{R}^2; W = \{(x, y) : y = \frac{1}{2}x\}$

7. a)  $V = \mathbb{R}^2; W = \text{the } x\text{-axis}$

b)  $V = \mathbb{R}^2; W = \{(x, y) : x + y \geq 1\}$

c)  $V = \mathbb{R}^4; W = \{(w, x, y, z) : x = 2w \text{ and } y = -3w\}$

8. a)  $V = C(\mathbb{R}, \mathbb{R})$ ;  $W =$  the set of functions which are equal to their derivative  
 b)  $V = C(\mathbb{R}, \mathbb{R})$ ;  $W =$  the set of functions  $f$  which are of the form  $a \sin x + b \cos x$  for constants  $a, b \in \mathbb{R}$   
 c)  $V = M_{23}(\mathbb{R})$ ;  $W =$  the set of  $2 \times 3$  matrices such that the entries in the first row sum to zero
9. a)  $V = M_2(\mathbb{R})$ ;  $W =$  the set of  $2 \times 2$  matrices which equal to  $(-1)$  times their transpose  
 b)  $V = M_2(\mathbb{R})$ ;  $W =$  the set of  $2 \times 2$  matrices which have at least one entry which is zero  
 c)  $V = M_2(\mathbb{R})$ ;  $W = \{A \in M_{22}(\mathbb{R}) : \text{tr}(A) = 0\}$
10. Write parametric equations for each of the given lines:  
 a) The line in  $\mathbb{R}^2$  that passes through the points  $(-3, 5)$  and  $(2, -6)$ .  
 b) The  $y$ -axis in  $\mathbb{R}^2$ .  
 c) The line in  $\mathbb{R}^2$  whose Cartesian equation is  $y = 5x + 4$ .  
*Hint:* find any two points on this line and then proceed as in part (a).
11. Write parametric equations for each of the given lines:  
 a) The line in  $\mathbb{R}^3$  which has direction vector  $(2, -3, 0)$  and passes through the point  $(1, -3, 4)$ .  
 b) The line in  $\mathbb{R}^3$  passing through the points  $(0, 1, -5)$  and  $(4, -2, -1)$ .  
 c) The line in  $\mathbb{R}^6$  passing through the points  $(-1, 1, 2, 0, 1, 4)$  and  $(2, -1, 0, 3, 1, -2)$ .
12. Write parametric equations for each of the given planes:  
 a) The plane in  $\mathbb{R}^3$  that passes through the points  $(1, 2, 3)$ ,  $(-1, 4, 2)$  and  $(3, 2, 2)$ .  
 b) The  $xy$ -plane in  $\mathbb{R}^3$  (this is the set of points whose  $z$ -coordinate is zero).  
*Hint:* first find any three noncollinear points in this plane and then proceed as in part (a).  
 c) The plane in  $\mathbb{R}^3$  that passes through the points  $(0, 5, -2)$ ,  $(-5, 3, 2)$  and  $(4, 1, -1)$ .
13. Here are the parametric equations for two lines in  $\mathbb{R}^3$ :

$$\begin{cases} x = 2 + 3t \\ y = 1 - t \\ z = 4 + 7t \end{cases} ; \begin{cases} x = 3 - 7t \\ y = -2 + 5t \\ z = -6 - 4t \end{cases}$$

Show that these two lines intersect in a point. Find the coordinates of this point.

*Hint:* If two lines intersect, they must meet at the same  $(x, y, z)$ , but the  $t$  doesn't have to be the same for both lines (think about why this is).

14. Here are the parametric equations for two lines in  $\mathbb{R}^3$ :

$$\begin{cases} x = 3t \\ y = 2 - t \\ z = -1 + t \end{cases} ; \begin{cases} x = 1 + 4t \\ y = -2 + t \\ z = -3 - 3t \end{cases}$$

Show that these two lines do not intersect.

15. Two airplanes fly along straight lines. At time  $t$ , plane 1 is at  $(75, 50, 25) + t(5, 10, 1)$  and plane 2 is at  $(60, 80, 34) + t(10, 5, -1)$ .
- Do the flight paths of these planes intersect? Explain.
  - Do the planes crash into one another? Explain.
16. Given each of the following vector spaces  $V$  and lists  $\mathcal{S}$  of vectors, determine whether or not  $\mathcal{S}$  is a linearly independent set. Give a quick reason for your answer.
- $V = \mathbb{R}^2$ ;  $\mathcal{S} = \{(1, 3)\}$ .
  - $V = \mathbb{R}^2$ ;  $\mathcal{S} = \{(2, 5), (-4, -10)\}$ .
  - $V = \mathbb{R}^2$ ;  $\mathcal{S} = \{(-3, 7), (4, -10)\}$ .
  - $V = \mathbb{R}^4$ ;  $\mathcal{S} = \{(1, 1, 1, 1), (2, 2, 2, 2), (1, 3, 5, 8), (0, -2, 5, 7)\}$ .
  - $V = \mathbb{R}^3$ ;  $\mathcal{S} = \{(1, 2, 7), (-1, 7, 5), (10, -3, 6), (8, 4, -1)\}$ .
17. Same directions as the previous problem:
- $V = \mathbb{R}^7$ ;  $\mathcal{S} = \{(2, 3, 8, -5, 0, 0, 0)\}$ .
  - $V = \mathbb{R}^4$ ;  $\mathcal{S} = \{(1, 2, 3, 4), (5, 6, 7, 8), (0, 0, 0, 0)\}$
  - $V = \mathbb{R}^3$ ;  $\mathcal{S} = \{(0, 1, 4), (0, 0, 1), (0, 7, -3)\}$ .
  - $V = \mathbb{R}^3$ ;  $\mathcal{S} = \{(0, 1, 2), (0, -3, -6), (2, 5, -8)\}$ .
18. Give an example of three vectors  $\mathbf{u}$ ,  $\mathbf{v}$  and  $\mathbf{w}$ , all belonging to the same vector space, such that  $\{\mathbf{u}, \mathbf{v}\}$ ,  $\{\mathbf{u}, \mathbf{w}\}$  and  $\{\mathbf{v}, \mathbf{w}\}$  are all linearly independent sets, but the set  $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$  is linearly dependent.
19. Let  $V = C^\infty(\mathbb{R}, \mathbb{R})$  be the vector space whose elements are functions from  $\mathbb{R}$  to  $\mathbb{R}$  which are differentiable infinitely many times (this is a real vector space since it is closed under addition and scalar multiplication and contains zero).



- a) In this vector space, is the set of functions  $\{\cos^2 x, \sin^2 x, 1\}$  linearly independent? Why or why not?

*Hint:* think of an identity relating  $\cos^2 x$  and  $\sin^2 x$ .

- b) In this vector space, is the set of functions  $\{\sin 2x, \sin x \cos x\}$  linearly independent? Why or why not?

*Hint:* do a Google search to find an identity relating these functions.

- c) In this vector space, is the set of functions  $\{\cos^2 x, \cos 2x, 1\}$  linearly independent? Why or why not?

*Hint:* do a Google search to find an identity relating some of these functions.

20. In this problem we will determine whether or not the set of functions

$$\{e^x, e^{-x}, 1\}$$

linearly independent.

*Note:* These functions are linearly independent if and only if no identity of the form  $c_1 e^x + c_2 e^{-x} = 1$  exists. Now, you're probably not aware of such an identity, but just because you aren't aware of such an identity doesn't mean that one doesn't exist. To prove that no such identity exists (i.e. to prove that the functions are linearly independent), we will carry out the following steps:

- a) Suppose we write 0 (the constant function 0) as a linear combination of  $\{e^x, e^{-x}, 1\}$ :

$$c_1 e^x + c_2 e^{-x} + c_3 1 = 0 \tag{3.1}$$

Show that it must be the case that  $c_1 + c_2 + c_3 = 0$ .

*Hint:* if the above equation holds as an equality between functions, then it is supposed to hold for all particular values of  $x$ . Plug in  $x = 0$  and see what happens.

- b) Differentiate both sides of equation (3.1) and subsequently (by plugging in  $x = 0$ ) show that  $c_1 - c_2 = 0$ .
- c) Differentiate both sides of the equation again, and show that  $c_1 + c_2 = 0$ .
- d) Solve the equations obtained in parts (a), (b) and (c) to explain why  $c_1 = c_2 = c_3 = 0$  (and thus why the three functions  $\{e^x, e^{-x}, 1\}$  are linearly independent).

**Note:** In Math 330, you learn a more general technique to prove that a collection of functions is linearly independent: you compute the Wronskian of those functions, and if there is any  $x$  where the Wronskian is nonzero, then the functions are linearly independent. Essentially, what this problem is doing is verifying that the Wronskian of  $\{e^x, e^{-x}, 1\}$  is nonzero when  $x = 0$ .

21. Suppose  $f$  and  $g$  are functions such that  $f(17) = 2$ ,  $f'(17) = 1$ ,  $g(17) = 3$  and  $g'(17) = 2$ . Are  $f$  and  $g$  linearly independent? Why or why not?
22. Given each of the vector spaces (or subspaces)  $W$ , find a basis of  $W$ , and the dimension of  $W$ . (You may assume without proof that these are all subspaces.)
- $W = \mathbb{R}^4$
  - $W = M_{42}(\mathbb{R})$  ( $4 \times 2$  matrices with real entries)
  - $W = \mathbb{P}_3$  (this is notation for the set of polynomials whose degree is  $\leq 3$ ).
23. Given each of the vector spaces (or subspaces)  $W$ , find a basis of  $W$ , and the dimension of  $W$ . (You may assume without proof that these are all subspaces.)
- $W = \text{Span}((2, 3), (12, 18), (-2, -3))$
  - $W = \{(x, y, z) \in \mathbb{R}^3 : 2x + y - 3z = 0\}$ .
  - $W =$  the line in  $\mathbb{R}^4$  with parametric equations  $x_1 = 4t$ ,  $x_2 = -3t$ ,  $x_3 = 2t$ ,  $x_4 = 0$ .
24. Given each of the vector spaces (or subspaces)  $W$ , find a basis of  $W$ , and the dimension of  $W$ . (You may assume without proof that these are all subspaces.)
- $V = C(\mathbb{R}, \mathbb{R})$ ;  $W$  is the subspace of  $V$  consisting of functions  $f$  satisfying the differential equation  $f'(x) = f(x)$ .
  - $W = \left\{ \begin{pmatrix} 2s - 5t \\ s \\ 4t \end{pmatrix} : s, t \in \mathbb{R} \right\}$
  - $W =$  the set of  $2 \times 2$  matrices which are equal to their transpose

## Chapter 4

---

# Dot products and orthogonality

---

### 4.1 Definitions and properties of dot products

#### GOALS OF THIS CHAPTER

Generalize geometric ideas like *length, distance, angle, direction, perpendicularity*, etc. to vectors.

**Definition 4.1** The dot product on  $\mathbb{R}^n$  is the function  $\mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$  (where the output associated to inputs  $\mathbf{v} = (v_1, \dots, v_n)$  and  $\mathbf{w} = (w_1, \dots, w_n)$  is written  $\mathbf{v} \cdot \mathbf{w}$  or  $\mathbf{v}^T \mathbf{w}$ ) defined by

$$\mathbf{v} \cdot \mathbf{w} = \sum_{j=1}^n v_j w_j = v_1 w_1 + v_2 w_2 + \dots + v_n w_n.$$

QUICK EXAMPLE:  $(3, -2, 5) \cdot (1, 0, -2) =$

**NOTE:**  $\mathbf{v} \cdot \mathbf{w}$  is a **scalar**, not a vector.

**Also note:** Dot product is connected with matrix multiplication. Writing  $\mathbf{v}$  and  $\mathbf{w}$  as column vectors, we see

This explains why we also denote the dot product  $\mathbf{v} \cdot \mathbf{w}$  by  $\mathbf{v}^T \mathbf{w}$ .

**Another note:** Recall that matrix multiplication is defined by setting the  $(i, j)$ -entry of  $AB$  equal to

$$(AB)_{ij} = \sum_{k=1}^n a_{ik} b_{kj}$$

So entries of the product matrix  $AB$  are the dot products of rows of  $A$  with columns of  $B$ .

## Dot product spaces

Spaces of functions also have a notion of dot product, which plays the same role as the above dot product does for traditional vectors:

**Definition 4.2** Let  $[a, b]$  be a closed, bounded interval in  $\mathbb{R}$ . The **dot product** on  $C([a, b], \mathbb{R})$  is the function  $C([a, b], \mathbb{R}) \times C([a, b], \mathbb{R}) \rightarrow \mathbb{R}$  (where the output associated to  $f$  and  $g$  is written  $f \cdot g$ ) defined by

$$f \cdot g = \int_a^b f(x)g(x) dx.$$

### QUICK EXAMPLE

Let  $f(x) = x$ ,  $g(x) = x^2$ . Then  $f, g \in C([0, 1], \mathbb{R})$ , and in this vector space,

$$f \cdot g =$$

**Definition 4.3** The **dot product** on  $M_n(\mathbb{R})$  (the space of  $n \times n$  matrices) is the function  $M_n(\mathbb{R}) \times M_n(\mathbb{R}) \rightarrow \mathbb{R}$  (where the output associated to  $A$  and  $B$  is written  $A \cdot B$ ) defined by  $A \cdot B = \text{tr}(AB^T)$ .

### Similarities between the various dot product formulas

Vector space	Dot product formula
$\mathbb{R}^n$ (traditional vectors)	$\mathbf{v} \cdot \mathbf{w} = \sum_{j=1}^n v_j w_j$ multiply coordinate-wise, then add
$C([a, b], \mathbb{R})$ (vectors are functions)	$f \cdot g = \int_a^b f(x)g(x) dx$ multiply $x$ -by- $x$ , then integrate
$M_n(\mathbb{R})$ (vectors are matrices)	$A \cdot B = \text{tr}(AB^T)$ matrix multiply, then add down the diagonal

Henceforth, we will call any vector space, together with its appropriate choice of dot product, a **dot product space**. (There are other examples of such spaces.) This is not standard mathematics notation; usually one would call such a thing an **inner product space**, but this requires a discussion of generalizations of dot product that will not be discussed in our course.

### Properties of dot products

**Theorem 4.4 (Properties of dot products)** *Let  $V$  be a dot product space. Then, for any vectors  $\mathbf{v}, \mathbf{v}_1, \mathbf{v}_2, \mathbf{w}, \mathbf{w}_1, \mathbf{w}_2 \in V$  and any scalar  $r \in \mathbb{R}$ , we have:*

1. Dot product is **symmetric**, meaning  $\mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v}$ .
2. Dot product is **bilinear**, meaning

$$\begin{aligned}(\mathbf{v}_1 + \mathbf{v}_2) \cdot \mathbf{w} &= \mathbf{v}_1 \cdot \mathbf{w} + \mathbf{v}_2 \cdot \mathbf{w}; \\ \mathbf{v} \cdot (\mathbf{w}_1 + \mathbf{w}_2) &= \mathbf{v} \cdot \mathbf{w}_1 + \mathbf{v} \cdot \mathbf{w}_2; \\ r(\mathbf{v} \cdot \mathbf{w}) &= (r\mathbf{v}) \cdot \mathbf{w} = \mathbf{v} \cdot (r\mathbf{w}).\end{aligned}$$

3. Dot product is **positive**, meaning  $\mathbf{v} \cdot \mathbf{v} \geq 0$ .
4. Dot product is **definite**, meaning  $\mathbf{v} \cdot \mathbf{v} = 0$  only when  $\mathbf{v} = \mathbf{0}$ .

## 4.1. Definitions and properties of dot products

---

**Note:** The bilinearity property extends to the following fact: if  $\mathbf{v}_i, \mathbf{w}_j \in V$  and  $r_i, s_j \in \mathbb{R}$ , then

$$\left( \sum_{i=1}^m r_i \mathbf{v}_i \right) \cdot \left( \sum_{j=1}^n s_j \mathbf{w}_j \right) = \sum_{i=1}^m \sum_{j=1}^n r_i s_j (\mathbf{v}_i \cdot \mathbf{w}_j).$$

**PROOF** We will prove the properties of dot product in the situation where  $V = \mathbb{R}^n$  (the other dot product spaces require their own proofs, which are similar).

Throughout this proof, let  $\mathbf{v} = (v_1, \dots, v_n)$ , let  $\mathbf{v}_1 = (v_{1,1}, \dots, v_{1,n})$ , and use similar notation for any other vectors.

First, let's prove statement 1 (symmetry): this is straightforward, as

$$\mathbf{v} \cdot \mathbf{w} = \sum_{j=1}^n v_j w_j = \sum_{j=1}^n w_j v_j = \mathbf{w} \cdot \mathbf{v}.$$

Next, let's verify the statements in 2 (bilinearity):

$$\begin{aligned} (\mathbf{v}_1 + \mathbf{v}_2) \cdot \mathbf{w} &= \sum_{j=1}^n (v_{1,j} + v_{2,j}) w_j = \sum_{j=1}^n v_{1,j} w_j + \sum_{j=1}^n v_{2,j} w_j = \mathbf{v}_1 \cdot \mathbf{w} + \mathbf{v}_2 \cdot \mathbf{w}; \\ \mathbf{v} \cdot (\mathbf{w}_1 + \mathbf{w}_2) &= \sum_{j=1}^n v_j (w_{1,j} + w_{2,j}) = \sum_{j=1}^n v_j w_{1,j} + \sum_{j=1}^n v_j w_{2,j} = \mathbf{v} \cdot \mathbf{w}_1 + \mathbf{v} \cdot \mathbf{w}_2; \\ r(\mathbf{v} \cdot \mathbf{w}) &= r \sum_{j=1}^n v_j w_j = \sum_{j=1}^n (r v_j) w_j = (r\mathbf{v}) \cdot \mathbf{w} \\ &= \sum_{j=1}^n v_j (r w_j) = \mathbf{v} \cdot (r\mathbf{w}). \end{aligned}$$

Now, let's prove statement 3 (positivity):

$$\mathbf{v} \cdot \mathbf{v} = v_1 v_1 + \dots + v_n v_n = \sum_{j=1}^n v_j^2 \geq 0.$$

Finally, we prove statement 4: if  $0 = \mathbf{v} \cdot \mathbf{v} = v_1 v_1 + \dots + v_n v_n = \sum_{j=1}^n v_j^2$ , it must be

that  $v_j^2 = 0$  for all  $j$  (otherwise this sum would be strictly positive).

That means  $v_j = 0$  for all  $j$ , i.e.  $\mathbf{v} = (0, 0, \dots, 0) = \mathbf{0}$ .  $\square$

## Dual relations

This last elementary result is very important. It will be used later when analyzing matrices and studying systems of linear equations:

**Theorem 4.5 (Dual relations)** *Let  $A \in M_{mn}(\mathbb{R})$ . Then, for any  $\mathbf{x} \in \mathbb{R}^n$  and any  $\mathbf{y} \in \mathbb{R}^m$ , we have*

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

*(Similarly, for any  $B \in M_{nm}(\mathbb{R})$ ,  $\mathbf{x} \cdot B\mathbf{y} = B^T\mathbf{x} \cdot \mathbf{y}$ .)*

PROOF This is a direct calculation:

$$\begin{aligned} A\mathbf{x} \cdot \mathbf{y} &= (A\mathbf{x})^T\mathbf{y} \\ &= \mathbf{x}^T A^T\mathbf{y} \quad (\text{transpose of product} = \text{backwards product of transposes}) \\ &= \mathbf{x}^T(A^T\mathbf{y}) \\ &= \mathbf{x} \cdot (A^T\mathbf{y}). \quad \square \end{aligned}$$

## 4.2 Norms, distances and length

RECALL

The **absolute value** of a real number is its distance from 0:

$$|x| = \sqrt{x^2} = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases}$$

We want to generalize the idea of absolute value to vectors:

**Definition 4.6** *Let  $V$  be a dot product space. For any vector  $\mathbf{v} \in V$ , we define the **norm** (a.k.a. **length** a.k.a. **magnitude** a.k.a. **absolute value**) of  $\mathbf{v}$  to be*

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}.$$

**Note:** Since dot product is positive, we'll never have to take the square root of a negative number here.

**Remark:** It is also useful to keep in mind that since  $\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$ ,

$$\|\mathbf{v}\|^2 = \mathbf{v} \cdot \mathbf{v}.$$

This makes  $\|\mathbf{v}\|^2$  easier to work with than  $\|\mathbf{v}\|$  when writing arguments.

#### EXAMPLE 1

Compute  $\|\mathbf{v}\|$ , if  $\mathbf{v} = (-4, 2, 5, 1, 0)$ .

*Solution:*  $\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{(-4)^2 + 2^2 + 5^2 + 1^2 + 0^2} = \boxed{\sqrt{46}}.$

#### EXAMPLE 2

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

*Solution:*

$$\|A\| = \sqrt{A \cdot A} = \sqrt{\text{tr}(AA^T)} = \sqrt{\text{tr} \left[ \begin{pmatrix} 2 & -3 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ -3 & 4 \end{pmatrix} \right]} = \sqrt{\text{tr} \begin{pmatrix} 13 & * \\ * & 17 \end{pmatrix}} = \boxed{\sqrt{30}}.$$

**Theorem 4.7 (Properties of norms)** Let  $V$  be a dot product space, let  $\mathbf{v}, \mathbf{w} \in V$  be vectors and let  $r \in \mathbb{R}$  be a scalar. Then:

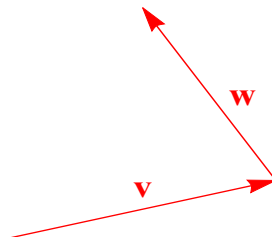
1. Norms are **positive**, meaning  $\|\mathbf{v}\| \geq 0$ .
2. Norms are **definite**, meaning  $\|\mathbf{v}\| = 0$  if and only if  $\mathbf{v} = \mathbf{0}$ .
3. Norms are **multiplicative**, meaning  $\|r\mathbf{v}\| = |r| \|\mathbf{v}\|$ .
4. The **Triangle Inequality** holds, meaning  $\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$ .

PROOF Statements (1) and (2) follow from statements (3) and (4) of Theorem 4.4 (properties of dot products). For property (3), observe

$$\|r\mathbf{v}\| = \sqrt{r\mathbf{v} \cdot r\mathbf{v}} = \sqrt{r^2(\mathbf{v} \cdot \mathbf{v})} = \sqrt{r^2} \sqrt{\mathbf{v} \cdot \mathbf{v}} = |r| \|\mathbf{v}\|.$$

We'll prove property (4) in Section 4.6.  $\square$

**Why is  $\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$  called the "Triangle Inequality"?**



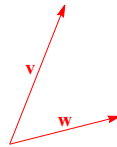


## Distance

The norm of a vector should be thought of as its distance from  $\mathbf{0}$ . More generally:

**Definition 4.8** Let  $V$  be a dot product space and let  $\mathbf{v}, \mathbf{w} \in V$ . Then the **distance** from  $\mathbf{v}$  to  $\mathbf{w}$  is

$$\text{dist}(\mathbf{v}, \mathbf{w}) = \|\mathbf{v} - \mathbf{w}\|.$$



**Observation:** By the multiplicative property of norms,  $\|\mathbf{v} - \mathbf{w}\| = \|\mathbf{w} - \mathbf{v}\|$  since

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

Therefore **distance is symmetric**:  $\text{dist}(\mathbf{v}, \mathbf{w}) = \text{dist}(\mathbf{w}, \mathbf{v})$ .

**Theorem 4.9 (Properties of distance)** Let  $V$  be a dot product space and let  $\mathbf{u}, \mathbf{v}, \mathbf{w}$  be vectors in  $V$  and let  $r \in \mathbb{R}$  be a scalar. Then:

1. Distances are **positive**, meaning  $\text{dist}(\mathbf{v}, \mathbf{w}) \geq 0$ .
2. Distances are **definite**, meaning  $\text{dist}(\mathbf{v}, \mathbf{w}) = 0$  if and only if  $\mathbf{v} = \mathbf{w}$ .
3. Distances are **symmetric**, meaning  $\text{dist}(\mathbf{v}, \mathbf{w}) = \text{dist}(\mathbf{w}, \mathbf{v})$ .
4. Distances are **multiplicative**, meaning  $\text{dist}(r\mathbf{v}, r\mathbf{w}) = |r| \text{dist}(\mathbf{v}, \mathbf{w})$ .
5. The **Triangle Inequality** holds, meaning  $\text{dist}(\mathbf{v}, \mathbf{w}) \leq \text{dist}(\mathbf{v}, \mathbf{u}) + \text{dist}(\mathbf{u}, \mathbf{w})$ .

PROOF For statement 1, notice  $\text{dist}(\mathbf{v}, \mathbf{w}) = \|\mathbf{v} - \mathbf{w}\| \geq 0$ .

For statement 2, suppose  $\text{dist}(\mathbf{v}, \mathbf{w}) = 0$ .

Then  $\|\mathbf{v} - \mathbf{w}\| = 0$  so  $\mathbf{v} - \mathbf{w} = \mathbf{0}$  so  $\mathbf{v} = \mathbf{w}$ .

Statement 3 was proven in the blue box before Theorem 4.9.

Statement 4 follows from the fact norms are multiplicative.

Statement 5 follows from the earlier triangle inequality in Theorem 4.7 (admittedly, that version hasn't been proven yet):

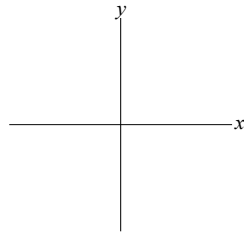
$$\begin{aligned} \text{dist}(\mathbf{v}, \mathbf{w}) &= \|\mathbf{v} - \mathbf{w}\| = \|\mathbf{v} - \mathbf{u} + \mathbf{u} - \mathbf{w}\| \\ &= \|(\mathbf{v} - \mathbf{u}) + (\mathbf{u} - \mathbf{w})\| \\ &\leq \|\mathbf{v} - \mathbf{u}\| + \|\mathbf{u} - \mathbf{w}\| \\ &= \text{dist}(\mathbf{v}, \mathbf{u}) + \text{dist}(\mathbf{u}, \mathbf{w}). \quad \square \end{aligned}$$

## Unit vectors and normalization

**Definition 4.10** Let  $V$  be a dot product space. A vector  $\mathbf{v} \in V$  is called a **unit vector** if  $\|\mathbf{v}\| = 1$ .

### EXAMPLE 3

Describe all the unit vectors in  $\mathbb{R}^2$ .



**Theorem 4.11** Let  $V$  be a dot product space. Given any nonzero vector  $\mathbf{v} \in V$ , there is a unit vector in the same direction as  $\mathbf{v}$ . This unit vector is called a **normalized version** of  $\mathbf{v}$ .

PROOF Given  $\mathbf{v} \neq \mathbf{0}$ , define  $\mathbf{u} = \frac{1}{\|\mathbf{v}\|} \mathbf{v}$ . This vector is clearly in the same direction as  $\mathbf{v}$ ;

$$\|\mathbf{u}\| = \left\| \frac{1}{\|\mathbf{v}\|} \mathbf{v} \right\| = \frac{1}{\|\mathbf{v}\|} \|\mathbf{v}\| = 1$$

so  $\mathbf{u}$  is a unit vector.  $\square$

### EXAMPLE 4

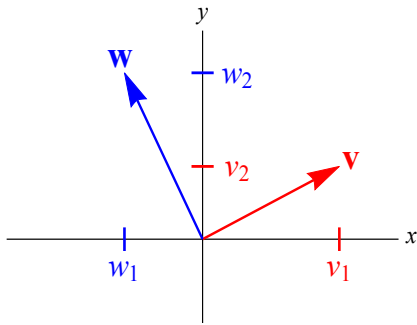
Find a unit vector in the same direction as  $(2, -3, 5, 1)$ .

## 4.3 Orthogonality

### GOAL

Extend the notion of *perpendicularity* in  $\mathbb{R}^2$  to arbitrary vector spaces.

To get started, think about what makes  $\mathbf{v} = (v_1, v_2)$  and  $\mathbf{w} = (w_1, w_2) \perp$  in  $\mathbb{R}^2$ ?



**Definition 4.12** Let  $V$  be a dot product space. Two vectors  $\mathbf{v}, \mathbf{w} \in V$  are called **orthogonal** (a.k.a. **perpendicular**) if  $\mathbf{v} \cdot \mathbf{w} = 0$ , in which case we write  $\mathbf{v} \perp \mathbf{w}$ .

**Note:** The zero vector  $\mathbf{0}$  is orthogonal to every vector (since  $\mathbf{0} \cdot \mathbf{v} = 0$  for any  $\mathbf{v} \in V$ ).

### EXAMPLE 5

Determine whether or not the vectors  $\mathbf{v} = (2, -3, 1, -4)$  and  $\mathbf{w} = (0, 1, 7, 1)$  in  $\mathbb{R}^4$  are orthogonal.

### EXAMPLE 6

Determine whether or not the functions  $f(x) = \sin x$  and  $g(x) = \cos x$  are orthogonal in  $C([0, \pi], \mathbb{R})$ .

**Theorem 4.13 (Pythagorean Theorem)** Let  $V$  be a dot product space and let  $\mathbf{v}, \mathbf{w} \in V$ . Then

$$\|\mathbf{v} + \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 \iff \mathbf{v} \perp \mathbf{w}.$$

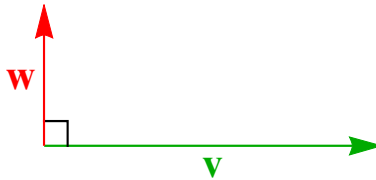
PROOF ( $\Leftarrow$ ) Assume  $\mathbf{v} \perp \mathbf{w}$ . Then

$$\begin{aligned} \|\mathbf{v} + \mathbf{w}\|^2 &= (\mathbf{v} + \mathbf{w}) \cdot (\mathbf{v} + \mathbf{w}) \\ &= \mathbf{v} \cdot \mathbf{v} + \mathbf{w} \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{w} + \mathbf{w} \cdot \mathbf{w} \\ &= \mathbf{v} \cdot \mathbf{v} + 0 + 0 + \mathbf{w} \cdot \mathbf{w} \\ &= \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2. \end{aligned}$$

( $\Rightarrow$ ) Assume  $\|\mathbf{v} + \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2$ . That means

$$\begin{aligned} (\mathbf{v} + \mathbf{w}) \cdot (\mathbf{v} + \mathbf{w}) &= \mathbf{v} \cdot \mathbf{v} + \mathbf{w} \cdot \mathbf{w} \\ \mathbf{v} \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{w} + \mathbf{w} \cdot \mathbf{v} + \mathbf{w} \cdot \mathbf{w} &= \mathbf{v} \cdot \mathbf{v} + \mathbf{w} \cdot \mathbf{w} \\ 2\mathbf{v} \cdot \mathbf{w} &= 0 \\ \mathbf{v} \cdot \mathbf{w} &= 0 \\ \Rightarrow \mathbf{v} \perp \mathbf{w}. \quad \square \end{aligned}$$

Why is  $\|\mathbf{v} + \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2$  called the “Pythagorean Theorem”?



## 4.4 Orthogonal complements and projections

**Definition 4.14** Let  $V$  be a dot product space and let  $W$  be a subspace of  $V$ . Define the **orthogonal complement of  $W$** , denoted  $W^\perp$  and pronounced “ $W$  perp”, to be the set of vectors orthogonal to every  $\mathbf{w} \in W$ ; that is,

$$W^\perp = \{\mathbf{v} \in V : \mathbf{v} \cdot \mathbf{w} = 0 \forall \mathbf{w} \in W\}.$$

**Theorem 4.15** Let  $V$  be a dot product space and let  $W$  be a subspace of  $V$  spanned by  $\mathbf{w}_1, \dots, \mathbf{w}_n$ . Let  $\mathbf{v} \in V$ . Then

$$\mathbf{v} \in W^\perp \iff \mathbf{v} \cdot \mathbf{w}_j = 0 \forall j = 1, 2, \dots, n.$$

PROOF ( $\Rightarrow$ ) follows from the definition of  $W^\perp$ , since  $\mathbf{w}_j \in W$  for all  $j$ .

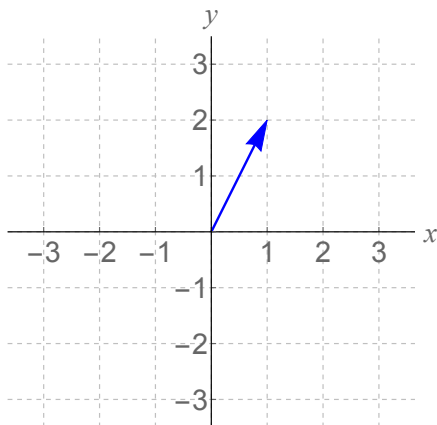
( $\Leftarrow$ ) Let  $\mathbf{w} \in W$ . Since the  $\mathbf{w}_j$  span  $W$ ,  $\mathbf{w} = \sum_j c_j \mathbf{w}_j$  so

$$\mathbf{v} \cdot \mathbf{w} = \mathbf{v} \cdot \left( \sum_j c_j \mathbf{w}_j \right) = \sum_j c_j (\mathbf{v} \cdot \mathbf{w}_j) = \sum_j c_j 0 = 0.$$

This means  $\mathbf{v} \perp \mathbf{w}$  for all  $\mathbf{w} \in W$ , so  $\mathbf{v} \in W^\perp$  as desired.  $\square$

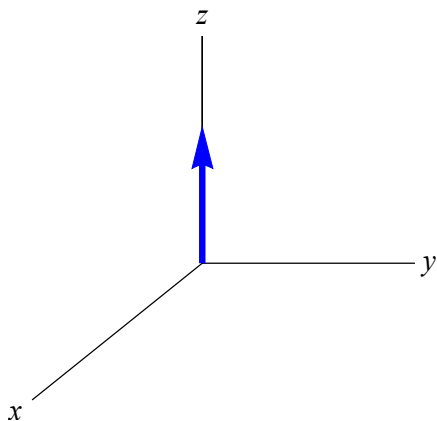
## EXAMPLE 7

Let  $V = \mathbb{R}^2$  and let  $W = \text{Span}((1, 2))$ . What is  $W^\perp$ ?



## EXAMPLE 8

Let  $V = \mathbb{R}^3$  and let  $W = \text{Span}((0, 0, 1))$ . What is  $W^\perp$ ?



## OBSERVATIONS IN EXAMPLES 7 AND 8

- A.
- B.
- C.
- D. **Orthogonal Decomposition:** every  $\mathbf{v} \in V$  can be written as

$$\mathbf{v} = \pi_W(\mathbf{v}) + \pi_{W^\perp}(\mathbf{v}),$$

where  $\pi_W(\mathbf{v}) \in W$  and  $\pi_{W^\perp}(\mathbf{v}) \in W^\perp$ :

- E. **Basis Union Property:** if  $\mathcal{B}$  is a basis of  $W$  and  $\mathcal{B}^\perp$  is a basis of  $W^\perp$ , then  $\mathcal{B} \cup \mathcal{B}^\perp$  forms a basis of  $V$ .

Our next goal is to establish that Observations A-E hold in general (for any subspace  $W$  of any vector space  $V$ ).

**Theorem 4.16 (Orthogonal complements are subspaces)** *Let  $V$  be a dot product space and let  $W$  be a subspace of  $V$ . Then  $W^\perp$  is also a subspace of  $V$ .*

PROOF We verify the three essential characteristics of a subspace.

Keep in mind that a generic element of  $W^\perp$  is a vector  $\mathbf{v}$  such that  $\mathbf{v} \cdot \mathbf{w} = 0$  for all  $\mathbf{w} \in W$ .

- $\mathbf{0} \in W^\perp$  since  $\mathbf{0} \cdot \mathbf{w} = 0$  for any  $\mathbf{w} \in V$ , hence any  $\mathbf{w} \in W$ .

- Let  $\mathbf{v}_1, \mathbf{v}_2 \in W^\perp$ .

Then  $\mathbf{v}_1 \cdot \mathbf{w} = 0$  and  $\mathbf{v}_2 \cdot \mathbf{w} = 0$  for any  $\mathbf{w} \in W$ .

Therefore  $(\mathbf{v}_1 + \mathbf{v}_2) \cdot \mathbf{w} = \mathbf{v}_1 \cdot \mathbf{w} + \mathbf{v}_2 \cdot \mathbf{w} = 0 + 0 = 0$  for any  $\mathbf{w} \in W$ .

This means  $\mathbf{v}_1 + \mathbf{v}_2 \in W^\perp$ .

- Let  $\mathbf{v} \in W^\perp$  and  $r \in \mathbb{R}$ .

Then  $\mathbf{v} \cdot \mathbf{w} = 0$  for any  $\mathbf{w} \in W$ .

Therefore  $(r\mathbf{v}) \cdot \mathbf{w} = r(\mathbf{v} \cdot \mathbf{w}) = r(0) = 0$  for any  $\mathbf{w} \in W$ .

This means  $r\mathbf{v} \in W^\perp$ .

Therefore  $W^\perp$  is a subspace of  $V$ .  $\square$

**Theorem 4.17** *Let  $V$  be a dot product space and let  $W$  be a subspace of  $V$ . Then  $W \cap W^\perp = \{\mathbf{0}\}$ .*

PROOF Clearly  $\mathbf{0} \in W \cap W^\perp$ , since  $\mathbf{0}$  is in every subspace.

Now suppose  $\mathbf{w} \in W \cap W^\perp$ . Then  $\mathbf{w} \perp \mathbf{w}$ , so  $\mathbf{w} \cdot \mathbf{w} = 0$ .

By definiteness, this implies  $\mathbf{w} = \mathbf{0}$ .  $\square$

## Projections

**Theorem 4.18 (Orthogonal Decomposition Theorem (dimension 1))** *Let  $V$  be a dot product space; let  $\mathbf{w} \neq \mathbf{0}$  be a vector in  $V$  and let  $W = \text{Span}(\mathbf{w})$  (so that  $W$  is a one-dimensional subspace of  $V$ ). Then for any  $\mathbf{v} \in V$ , we can write*

$$\mathbf{v} = \pi_{\mathbf{w}}(\mathbf{v}) + \pi_{\mathbf{w}^\perp}(\mathbf{v})$$

where  $\pi_{\mathbf{w}}(\mathbf{v}) \parallel \mathbf{w}$  (i.e.  $\pi_{\mathbf{w}}(\mathbf{v}) \in W$ ) and  $\pi_{\mathbf{w}^\perp}(\mathbf{v}) \perp \mathbf{w}$  (i.e.  $\pi_{\mathbf{w}^\perp}(\mathbf{v}) \in W^\perp$ ).

#### 4.4. Orthogonal complements and projections

PROOF Given  $\mathbf{v}, \mathbf{w}$ , set

$$\pi_{\mathbf{w}}(\mathbf{v}) = \frac{\mathbf{v} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}} \mathbf{w} \quad \text{and} \quad \pi_{\mathbf{w}^\perp}(\mathbf{v}) = \mathbf{v} - \pi_{\mathbf{w}}(\mathbf{v}).$$

Clearly  $\pi_{\mathbf{w}}(\mathbf{v}) \parallel \mathbf{w}$ , since it is a multiple of  $\mathbf{w}$ .

It is also clear that  $\pi_{\mathbf{w}}(\mathbf{v}) + \pi_{\mathbf{w}^\perp}(\mathbf{v}) = \mathbf{v}$ .

So, we are left to verify  $\pi_{\mathbf{w}^\perp}(\mathbf{v}) \perp \mathbf{w}$ :

$$\begin{aligned} \pi_{\mathbf{w}^\perp}(\mathbf{v}) \cdot \mathbf{w} &= (\mathbf{v} - \pi_{\mathbf{w}}(\mathbf{v})) \cdot \mathbf{w} = \left( \mathbf{v} - \frac{\mathbf{v} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}} \mathbf{w} \right) \cdot \mathbf{w} \\ &= \mathbf{v} \cdot \mathbf{w} - \frac{\mathbf{v} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}} \mathbf{w} \cdot \mathbf{w} \\ &= \mathbf{v} \cdot \mathbf{w} - \mathbf{v} \cdot \mathbf{w} = 0. \end{aligned}$$

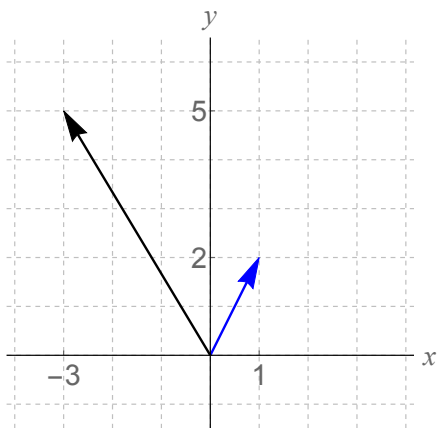
Therefore  $\pi_{\mathbf{w}^\perp}(\mathbf{v}) \perp \mathbf{w}$  as desired.  $\square$

**Definition 4.19** Let  $V$  be a dot product space, and let  $\mathbf{v}, \mathbf{w} \in V$  with  $\mathbf{w} \neq \mathbf{0}$ . The **projection of  $\mathbf{v}$  onto  $\mathbf{w}$** , denoted  $\pi_{\mathbf{w}}\mathbf{v}$  or  $\text{proj}_{\mathbf{w}}\mathbf{v}$  (and sometimes  $\mathbf{v}^{\mathbf{w}}$ ), is the vector

$$\pi_{\mathbf{w}}(\mathbf{v}) = \pi_{\mathbf{w}}\mathbf{v} = \text{proj}_{\mathbf{w}}\mathbf{v} = \frac{\mathbf{v} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}} \mathbf{w}.$$

EXAMPLE 9

Compute the projection of  $(-3, 5)$  onto  $(1, 2)$ .





#### 4.4. Orthogonal complements and projections

---

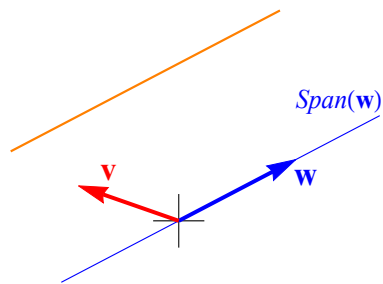
##### EXAMPLE 10

Compute the projection of  $f(x) = x^2$  onto  $g(x) = x$  in  $C([0, 1], \mathbb{R})$ .

---

##### GENERAL PICTURE

---



#### 4.4. Orthogonal complements and projections

**Theorem 4.20 (Properties of projections)** Let  $V$  be a dot product space, and let  $\mathbf{v}, \mathbf{w} \in V$  with  $\mathbf{w} \neq \mathbf{0}$ . Then:

1. Alternate formulas:  $\pi_{\mathbf{w}}\mathbf{v} = \frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{w}\|^2}\mathbf{w} = \frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{w}\|}\mathbf{u} = (\mathbf{v} \cdot \mathbf{u})\mathbf{u}$ , where  $\mathbf{u}$  is a normalized version of  $\mathbf{w}$ .
2. Projection onto orthogonal vector is zero:  $\mathbf{v} \perp \mathbf{w}$  if and only if  $\pi_{\mathbf{w}}\mathbf{v} = \mathbf{0}$ .
3. Projection onto parallel vector is itself:  $\mathbf{v} \parallel \mathbf{w}$  if and only if  $\pi_{\mathbf{w}}\mathbf{v} = \mathbf{v}$ .
4. Projection  $\perp$  what's leftover:  $\pi_{\mathbf{w}}\mathbf{v} \perp (\mathbf{v} - \pi_{\mathbf{w}}\mathbf{v})$ .
5. Projecting twice is redundant:  $\pi_{\mathbf{w}}(\pi_{\mathbf{w}}\mathbf{v}) = \pi_{\mathbf{w}}\mathbf{v}$ .

PROOF These are all calculations using the definitions.  $\square$

**Theorem 4.21 (Orthogonal Decomposition Theorem (general case))** Let  $V$  be a dot product space and let  $W$  be a finite-dimensional subspace of  $V$ . Then for any  $\mathbf{v} \in V$ , we can write

$$\mathbf{v} = \pi_W(\mathbf{v}) + \pi_{W^\perp}(\mathbf{v}),$$

where  $\pi_W(\mathbf{v}) \in W$  and  $\pi_{W^\perp}(\mathbf{v}) \in W^\perp$ . The  $\pi_W(\mathbf{v})$  in this theorem is called the **projection of  $\mathbf{v}$  onto  $W$**  and the  $\pi_{W^\perp}(\mathbf{v})$  is the **component of  $\mathbf{v}$  orthogonal to  $W$** .

PROOF First, let  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$  be a basis of  $W$  with two properties:

1.  $\|\mathbf{x}_j\| = 1$  for all  $j$  (i.e.  $\mathbf{x}_j \cdot \mathbf{x}_j = 1$  for all  $j$ ); and
2.  $\mathbf{x}_j \perp \mathbf{x}_k$  for all  $j \neq k$  (i.e.  $\mathbf{x}_j \cdot \mathbf{x}_k = 0$  for all  $j \neq k$ ).

**Note:** It is not clear that such a basis exists. We will prove in the next section that such a basis (called an *orthonormal basis*) always exists.

Given such a basis, and given  $\mathbf{v}$ , define

$$\pi_W(\mathbf{v}) = \sum_{j=1}^n (\mathbf{v} \cdot \mathbf{x}_j)\mathbf{x}_j = (\mathbf{v} \cdot \mathbf{x}_1)\mathbf{x}_1 + (\mathbf{v} \cdot \mathbf{x}_2)\mathbf{x}_2 + \dots + (\mathbf{v} \cdot \mathbf{x}_n)\mathbf{x}_n.$$

Since  $\pi_W(\mathbf{v})$  is a combination of the  $\mathbf{x}_j$ , it is in  $W = \text{Span}(\mathbf{x}_1, \dots, \mathbf{x}_n)$  as desired.

Next, define

$$\pi_{W^\perp}(\mathbf{v}) = \mathbf{v} - \pi_W(\mathbf{v}).$$

This ensures that  $\mathbf{v} = \pi_W(\mathbf{v}) + \pi_{W^\perp}(\mathbf{v})$ .

All that is left to prove is that  $\pi_{W^\perp}(\mathbf{v}) \in W^\perp$ .

To verify this, we will check that  $\pi_{W^\perp}(\mathbf{v})$  is orthogonal to all the  $\mathbf{x}_k$  (which is sufficient since the  $\mathbf{x}_k$  span  $W$ ):

$$\begin{aligned}
 \pi_{W^\perp}(\mathbf{v}) \cdot \mathbf{x}_j &= (\mathbf{v} - \pi_W(\mathbf{v})) \cdot \mathbf{x}_k \\
 &= \left( \mathbf{v} - \sum_{j=1}^n (\mathbf{v} \cdot \mathbf{x}_j) \mathbf{x}_j \right) \cdot \mathbf{x}_k \\
 &= \mathbf{v} \cdot \mathbf{x}_k - \sum_{j=1}^n (\mathbf{v} \cdot \mathbf{x}_j) (\mathbf{x}_j \cdot \mathbf{x}_k) \\
 &= \mathbf{v} \cdot \mathbf{x}_k - [(\mathbf{v} \cdot \mathbf{x}_1)0 + \dots + (\mathbf{v} \cdot \mathbf{x}_{k-1})0 + (\mathbf{v} \cdot \mathbf{x}_k)1 \\
 &\quad + (\mathbf{v} \cdot \mathbf{x}_{k+1})0 + \dots + (\mathbf{v} \cdot \mathbf{x}_n)0] \\
 &= \mathbf{v} \cdot \mathbf{x}_k - \mathbf{v} \cdot \mathbf{x}_k = 0.
 \end{aligned}$$

Therefore  $\pi_{W^\perp}(\mathbf{v})$  is orthogonal to all the  $\mathbf{x}_j$ , hence is in  $W^\perp$  as desired.  $\square$

**Theorem 4.22 (Uniqueness of orthogonal decompositions)** *The  $\pi_W(\mathbf{v})$  and the  $\pi_{W^\perp}(\mathbf{v})$  described in Theorem 4.21 are unique (in other words, you don't get to choose between more than one possible  $\pi_W(\mathbf{v})$ s or more than one possible  $\pi_{W^\perp}(\mathbf{v})$ s).*

PROOF Suppose there are two vectors, say  $\pi_W(\mathbf{v})$  and  $\hat{\pi}_W(\mathbf{v})$  which are “both” the projection of  $\mathbf{v}$  onto subspace  $W$ . That means:

$$\pi_W(\mathbf{v}) \in W \quad \hat{\pi}_W(\mathbf{v}) \in W \quad \mathbf{v} - \pi_W(\mathbf{v}) \in W^\perp \quad \mathbf{v} - \hat{\pi}_W(\mathbf{v}) \in W^\perp$$

Since  $W$  is a subspace,  $\pi_W(\mathbf{v}) - \hat{\pi}_W(\mathbf{v}) \in W$ , and since  $W^\perp$  is a subspace,

$$(\mathbf{v} - \pi_W(\mathbf{v})) - (\mathbf{v} - \hat{\pi}_W(\mathbf{v})) = \pi_W(\mathbf{v}) - \hat{\pi}_W(\mathbf{v}) \in W^\perp.$$

Combining these statements, we see  $\pi_W(\mathbf{v}) - \hat{\pi}_W(\mathbf{v}) \in W \cap W^\perp$ .

But by Observation B (Theorem 4.17), that means  $\pi_W(\mathbf{v}) - \hat{\pi}_W(\mathbf{v}) = \mathbf{0}$ .

In other words,  $\pi_W(\mathbf{v}) = \hat{\pi}_W(\mathbf{v})$ . This means we didn't really have two different projections of  $\mathbf{v}$  onto  $W$ ; we only had one.  $\square$

**Corollary 4.23** *Let  $V$  be a finite-dimensional inner product space and let  $W$  be a subspace of  $V$ . Given any basis  $\mathcal{B}$  of  $W$  and any basis  $\mathcal{B}^\perp$  of  $W^\perp$ ,  $\mathcal{B} \cup \mathcal{B}^\perp$  is a basis of  $V$ .*

PROOF Write  $\mathcal{B} = \{\mathbf{w}_1, \dots, \mathbf{w}_m\}$  and write  $\mathcal{B}^\perp = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ .

*Claim 1:*  $\mathcal{B} \cup \mathcal{B}^\perp$  spans  $V$ .

*Proof of Claim 1:* Let  $\mathbf{v} \in V$ . Then by the Orthogonal Decomposition Theorem,

$$\mathbf{v} = \pi_W(\mathbf{v}) + \pi_{W^\perp}(\mathbf{v}) = \sum_{i=1}^m c_i \mathbf{w}_i + \sum_{j=1}^n b_j \mathbf{v}_j.$$

Thus  $\mathbf{v}$  can be written as a linear combination of vectors in  $\mathcal{B} \cup \mathcal{B}^\perp$ , so  $\mathcal{B} \cup \mathcal{B}^\perp$  is a spanning set for  $V$ .

*Claim 2:*  $\mathcal{B} \cup \mathcal{B}^\perp$  is lin. ind.

*Proof of Claim 2:* Suppose  $\sum_{i=1}^m c_i \mathbf{w}_i + \sum_{j=1}^n b_j \mathbf{v}_j = \mathbf{0}$ .

$$\text{Therefore } \sum_{i=1}^m c_i \mathbf{w}_i = - \sum_{j=1}^n b_j \mathbf{v}_j.$$

Since the **left-hand side** of this equation is in  $W$  and the **right-hand side** is in  $W^\perp$ , both sides must belong to  $W \cap W^\perp$ , hence both sides are  $\mathbf{0}$ .

But since  $\mathcal{B}$  and  $\mathcal{B}^\perp$  are bases, they are lin. ind., so all the  $b_j$  and  $c_i$  must be zero, hence  $\mathcal{B} \cup \mathcal{B}^\perp$  is lin. ind. set.

By Claims 1 and 2,  $\mathcal{B} \cup \mathcal{B}^\perp$  is a basis of  $V$ .  $\square$

**Corollary 4.24 (Dimension Formula)** *Let  $V$  be a finite-dimensional inner product space and let  $W$  be a subspace of  $V$ . Then*

$$\dim V = \dim W + \dim(W^\perp).$$

PROOF Let  $\mathcal{B}$  be a basis of  $W$  and  $\mathcal{B}^\perp$  be a basis of  $W^\perp$ . By the previous theorem,  $\mathcal{B} \cup \mathcal{B}^\perp$  (which must have  $\dim W + \dim W^\perp$  vectors) is a basis of  $V$ , so

$$\begin{aligned} \dim V &= \#(\text{vectors in } \mathcal{B} \cup \mathcal{B}^\perp) \\ &= \#(\text{vectors in } \mathcal{B}) + \#(\text{vectors in } \mathcal{B}^\perp) \\ &= \dim W + \dim W^\perp. \end{aligned}$$

as wanted.  $\square$

**Corollary 4.25** *Let  $V$  be a finite-dimensional inner product space and let  $W$  be a subspace of  $V$ . Then  $(W^\perp)^\perp = W$ .*

PROOF First,  $\dim(W^\perp)^\perp + \dim W^\perp = \dim V$ , so

$$\dim(W^\perp)^\perp = \dim V - \dim W^\perp = \dim W.$$

Now let  $\mathbf{w} \in W$ ;  $\mathbf{w} \perp \mathbf{v}$  for all  $\mathbf{v} \in W^\perp$ , so  $\mathbf{w} \in (W^\perp)^\perp$ . Thus  $W \subseteq (W^\perp)^\perp$ .

But since these subspaces have the same dimension, they must be equal by an earlier theorem (Theorem 3.36).  $\square$

**WARNING:** If  $\dim V = \infty$ , the preceding corollary is false! It turns out that in this situation,  $(W^\perp)^\perp$  can contain more vectors than  $W$ .

## Distance to a subspace

**Definition 4.26** *Let  $V$  be a dot product space and let  $W$  be a subspace of  $V$ . Given any  $\mathbf{v} \in V$ , let the **distance from  $\mathbf{v}$  to  $W$** , denoted by  $\text{dist}(\mathbf{v}, W)$ , be the minimum distance from  $\mathbf{v}$  to any vector in  $W$ .*

**Theorem 4.27 (Distance to a subspace)** Let  $V$  be a dot product space and let  $W$  be a finite-dimensional subspace of  $V$ . Given any  $\mathbf{v} \in V$ ,  $\text{dist}(\mathbf{v}, W) = \|\pi_{W^\perp}(\mathbf{v})\|$ .

PROOF Let  $\mathbf{w} \in W$ . Then  $\pi_W(\mathbf{v}) - \mathbf{w} \in W$ , since it is the difference of two vectors in  $W$  and  $W$  is a subspace. Thus  $(\pi_W(\mathbf{v}) - \mathbf{w}) \perp \pi_{W^\perp}(\mathbf{v})$  since  $\pi_{W^\perp}(\mathbf{v}) \in W^\perp$ .

*Claim 1:*  $\|\pi_{W^\perp}(\mathbf{v})\| \leq \text{dist}(\mathbf{v}, W)$ .

*Proof of Claim 1:* So by the Pythagorean Theorem,

$$\|\pi_W(\mathbf{v}) - \mathbf{w}\|^2 + \|\pi_{W^\perp}(\mathbf{v})\|^2 = \|(\pi_W(\mathbf{v}) - \mathbf{w}) + \pi_{W^\perp}(\mathbf{v})\|^2,$$

i.e.

$$\|\pi_W(\mathbf{v}) - \mathbf{w}\|^2 + \|\pi_{W^\perp}(\mathbf{v})\|^2 = \|\mathbf{v} - \mathbf{w}\|^2,$$

so

$$\|\pi_{W^\perp}(\mathbf{v})\|^2 \leq \|\mathbf{v} - \mathbf{w}\|^2.$$

Take the square root of both sides of this inequality to get

$$\|\pi_{W^\perp}(\mathbf{v})\| \leq \text{dist}(\mathbf{v}, \mathbf{w}).$$

Since this holds for any  $\mathbf{w} \in W$ , we have  $\|\pi_{W^\perp}(\mathbf{v})\| \leq \text{dist}(\mathbf{v}, W)$ .

*Claim 2:*  $\|\pi_{W^\perp}(\mathbf{v})\| \geq \text{dist}(\mathbf{v}, W)$ .

*Proof of Claim 2:* Observe  $\|\pi_{W^\perp}(\mathbf{v})\| = \|\mathbf{v} - \pi_W(\mathbf{v})\| = \text{dist}(\mathbf{v}, \pi_W(\mathbf{v}))$ .

Since  $\pi_W(\mathbf{v}) \in W$ , this must be at least  $\text{dist}(\mathbf{v}, W)$ .

So  $\|\pi_{W^\perp}(\mathbf{v})\| \geq \text{dist}(\mathbf{v}, W)$ .

The two claims put together imply the result.  $\square$

## 4.5 Orthonormal bases and the Gram-Schmidt procedure

## EXAMPLE 11

Let  $V = \mathbb{R}^4$  and let  $W = \text{Span}((1, 2, 1, -1), (0, 2, 1, -2))$ . Let  $\mathbf{v} = (2, -1, 0, 0)$ . Compute  $\pi_W(\mathbf{v})$ .

*Theoretical solution:* We know (from the proof of the Orthogonal Decomposition Theorem) that if we had a basis  $\{\mathbf{x}_1, \mathbf{x}_2\}$  of  $W$  with the properties that  $\|\mathbf{x}_j\| = 1$  for all  $j$  and  $\mathbf{x}_j \perp \mathbf{x}_k$  for  $j \neq k$ , then

$$\pi_W(\mathbf{v}) = (\mathbf{v} \cdot \mathbf{x}_1)\mathbf{x}_1 + (\mathbf{v} \cdot \mathbf{x}_2)\mathbf{x}_2.$$

**Question:** How do you find  $\{\mathbf{x}_1, \mathbf{x}_2\}$ ?

**Definition 4.28** Let  $V$  be a dot product space. A set  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  of vectors in  $V$  is called **(pairwise) orthogonal** if  $\mathbf{x}_j \perp \mathbf{x}_k$  for all  $j \neq k$ . The set is called **orthonormal** if it is orthogonal and if  $\|\mathbf{x}_j\| = 1$  for all  $j$ .

## EXAMPLES

- $\{(2, 3), (-3, 2)\}$  is an orthogonal set of vectors in  $\mathbb{R}^2$  which is not orthonormal.
- $\left\{\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \left(\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)\right\}$  is an orthonormal set of vectors in  $\mathbb{R}^2$ .
- $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$  is an orthonormal set of vectors in  $\mathbb{R}^n$  where

$$\mathbf{e}_j = (0, 0, \dots, 0, 1, 0, 0, \dots, 0)$$

(the 1 is in the  $j^{\text{th}}$  position).

We are especially interested in **bases** that are orthonormal. This is because we can easily write a vector as a linear combination of vectors coming from an orthonormal basis. For example, consider  $\mathbb{R}^3$  which has orthonormal basis  $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$  where  $\mathbf{e}_1 = (1, 0, 0)$ ;  $\mathbf{e}_2 = (0, 1, 0)$ ;  $\mathbf{e}_3 = (0, 0, 1)$ .

Now, let  $\mathbf{v} = (2, -3, 6)$ . Then  $\mathbf{v} =$

## 4.5. Orthonormal bases and the Gram-Schmidt procedure

---

This generalizes:

**Theorem 4.29** *Let  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  be an orthonormal basis of dot product space  $V$ . Then for every  $\mathbf{v} \in V$ , we have*

$$\mathbf{v} = \sum_{j=1}^n (\mathbf{v} \cdot \mathbf{x}_j) \mathbf{x}_j.$$

**PROOF** By the unique decomposition theorem, we know that we can write

$$\mathbf{v} = \sum_{k=1}^n c_k \mathbf{x}_k$$

uniquely. Now for each  $j$ ,

$$\mathbf{v} \cdot \mathbf{x}_j = \sum_{k=1}^n c_k \mathbf{x}_k \cdot \mathbf{x}_j = \sum_{k=1}^n c_k (\mathbf{x}_k \cdot \mathbf{x}_j).$$

Since the  $\mathbf{x}_j$  are orthonormal, we have  $\mathbf{x}_k \cdot \mathbf{x}_j = 0$  if  $\mathbf{x}_j \neq \mathbf{x}_k$  and  $\mathbf{x}_j \cdot \mathbf{x}_j = \|\mathbf{x}_j\|^2 = 1^2 = 1$ , so the above expression reduces to

$$\mathbf{v} \cdot \mathbf{x}_j = c_j.$$

Thus  $\mathbf{v} = \sum_{k=1}^n c_k \mathbf{x}_k = \sum_{k=1}^n (\mathbf{v} \cdot \mathbf{x}_k) \mathbf{x}_k$  as desired.  $\square$

From the proof of the Orthogonal Decomposition Theorem, we get the following formula:

**Corollary 4.30 (Projection Formula)** *Let  $V$  be an inner product space and let  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$  be an orthonormal basis of a finite-dimensional subspace  $W$ . Then for every  $\mathbf{v} \in V$ , we have*

$$\pi_W(\mathbf{v}) = \sum_{j=1}^n (\mathbf{v} \cdot \mathbf{x}_j) \mathbf{x}_j.$$

So we now have a theoretical procedure to compute projections of vectors onto subspaces:

1. Find an orthonormal basis of the subspace; then
2. apply the Projection Formula.

**Problem:** How do we find orthonormal bases?

*Solution:* Use a procedure called *Gram-Schmidt*.



**Theorem 4.31 (Gram-Schmidt Theorem)** *Given a basis  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  of a dot product space  $V$  ( $V$  could be a subspace of some other dot product space), there is a basis  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  of  $V$  such that*

1.  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  is an orthonormal basis, and
2.  $\text{Span}(\mathbf{x}_1, \dots, \mathbf{x}_k) = \text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$  for all  $k \leq n$ .

PROOF The orthonormal basis  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  is constructed by the Gram-Schmidt procedure described below.  $\square$

The following corollary fills in the gap in the proof of the Orthogonal Decomposition Theorem:

**Corollary 4.32** *Every finite-dimensional dot product space has an orthonormal basis.*

### Gram-Schmidt procedure

This series of steps constructs an orthonormal basis  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  from a given basis  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ :

Step 1: Define

$$\begin{aligned} \mathbf{w}_1 &= \mathbf{v}_1 \\ \mathbf{w}_2 &= \mathbf{v}_2 - \pi_{\mathbf{w}_1} \mathbf{v}_2 \\ \mathbf{w}_3 &= \mathbf{v}_3 - \pi_{\mathbf{w}_1} \mathbf{v}_3 - \pi_{\mathbf{w}_2} \mathbf{v}_3 \\ &\vdots \\ \mathbf{w}_n &= \mathbf{v}_n - \pi_{\mathbf{w}_1} \mathbf{v}_n - \pi_{\mathbf{w}_2} \mathbf{v}_n - \dots - \pi_{\mathbf{w}_{n-1}} \mathbf{v}_n \end{aligned}$$

This produces an orthogonal set  $\{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  which is a basis.

Step 2: convert the orthogonal set to an orthonormal set by normalizing each vector: for each  $j$ , set  $\mathbf{x}_j = \frac{\mathbf{w}_j}{\|\mathbf{w}_j\|}$ .

#### BACK TO EXAMPLE 11

Let  $V = \mathbb{R}^4$  and let  $W = \text{Span}((1, 2, 1, -1), (0, 2, 1, -2))$ . Let  $\mathbf{v} = (2, -1, 0, 0)$ . How can we write  $\mathbf{v}$  as  $\pi_W(\mathbf{v}) + \pi_{W^\perp}(\mathbf{v})$  where  $\pi_W(\mathbf{v}) \in W$ ;  $\pi_{W^\perp}(\mathbf{v}) \in W^\perp$ ?

## 4.5. Orthonormal bases and the Gram-Schmidt procedure

---

*Step 1:* Use Gram-Schmidt to find orthonormal basis of  $W$ .

Recall  $W = \text{Span}(\mathbf{v}_1, \mathbf{v}_2)$  where  $\begin{cases} \mathbf{v}_1 = (1, 2, 1, -1) \\ \mathbf{v}_2 = (0, 2, 1, -2) \end{cases}$ .

“ $\mathbf{v}$ 's  $\mapsto$   $\mathbf{w}$ 's”:

“ $\mathbf{w}$ 's  $\mapsto$   $\mathbf{x}$ 's”:

*Step 2:* Use projection formula to find  $\pi_W(\mathbf{v})$  (recall  $\mathbf{v} = (2, -1, 0, 0)$ ).

From Step 1, here's our orthonormal basis of  $W$ :

$$\{\mathbf{x}_1, \mathbf{x}_2\} = \left\{ \left( \frac{1}{\sqrt{7}}, \frac{2}{\sqrt{7}}, \frac{1}{\sqrt{7}}, \frac{-1}{\sqrt{7}} \right), \left( \frac{-1}{\sqrt{2}}, 0, 0, \frac{-1}{\sqrt{2}} \right) \right\}$$

*Step 3:* Find  $\pi_{W^\perp}(\mathbf{v})$  by setting  $\pi_{W^\perp}(\mathbf{v}) = \mathbf{v} - \pi_W(\mathbf{v})$ .

This gives us a (not very nice) method to compute projections of vectors onto subspaces. A less messy way to carry out (Step 1 of) the Gram-Schmidt procedure is to use the following command in *Mathematica*:

```
Orthogonalize[{{1,2,1,-1}, {0,2,1,-2}}]
```

Later in the course we will see another method of computing projections by hand that is less terrible.

## 4.6 Angles and the Cauchy-Schwarz inequality

**Theorem 4.33 (Cauchy-Schwarz Inequality)** *Let  $V$  be a dot product space. Then for any  $\mathbf{v}, \mathbf{w} \in V$ ,  $|\mathbf{v} \cdot \mathbf{w}| \leq \|\mathbf{v}\| \|\mathbf{w}\|$ .*

PROOF If  $\mathbf{w} = \mathbf{0}$ , then both sides of the C-S inequality are zero, so it is trivially true. Henceforth, assume  $\mathbf{w} \neq \mathbf{0}$ . Then

$$\begin{aligned}
 0 \leq \|\mathbf{v} - \pi_{\mathbf{w}}\mathbf{v}\|^2 &= (\mathbf{v} - \pi_{\mathbf{w}}\mathbf{v}) \cdot (\mathbf{v} - \pi_{\mathbf{w}}\mathbf{v}) \\
 &= \mathbf{v} \cdot \mathbf{v} - 2\pi_{\mathbf{w}}\mathbf{v} \cdot \mathbf{v} + \pi_{\mathbf{w}}\mathbf{v} \cdot \pi_{\mathbf{w}}\mathbf{v} \\
 &= \|\mathbf{v}\|^2 - 2\left(\frac{\mathbf{v} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}}\mathbf{w}\right) \cdot \mathbf{v} + \left(\frac{\mathbf{v} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}}\mathbf{w}\right) \cdot \left(\frac{\mathbf{v} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}}\mathbf{w}\right) \\
 &= \|\mathbf{v}\|^2 - 2\frac{(\mathbf{v} \cdot \mathbf{w})^2}{\mathbf{w} \cdot \mathbf{w}} + \frac{(\mathbf{v} \cdot \mathbf{w})^2}{(\mathbf{w} \cdot \mathbf{w})^2}\mathbf{w} \cdot \mathbf{w} \\
 &= \|\mathbf{v}\|^2 - 2\frac{(\mathbf{v} \cdot \mathbf{w})^2}{\mathbf{w} \cdot \mathbf{w}} + \frac{(\mathbf{v} \cdot \mathbf{w})^2}{\mathbf{w} \cdot \mathbf{w}} \\
 &= \|\mathbf{v}\|^2 - \frac{(\mathbf{v} \cdot \mathbf{w})^2}{\mathbf{w} \cdot \mathbf{w}} \\
 &= \|\mathbf{v}\|^2 - \frac{(\mathbf{v} \cdot \mathbf{w})^2}{\|\mathbf{w}\|^2}.
 \end{aligned}$$

Therefore

$$\frac{(\mathbf{v} \cdot \mathbf{w})^2}{\|\mathbf{w}\|^2} \leq \|\mathbf{v}\|^2 \Rightarrow (\mathbf{v} \cdot \mathbf{w})^2 \leq \|\mathbf{v}\|^2 \|\mathbf{w}\|^2 \Rightarrow |\mathbf{v} \cdot \mathbf{w}| \leq \|\mathbf{v}\| \|\mathbf{w}\|. \quad \square$$

## 4.6. Angles and the Cauchy-Schwarz inequality

Using the C-S Inequality, we are now able to prove the following result, which was mentioned earlier but not proven:

**Theorem 4.34 (Triangle Inequality)** *Let  $V$  be a dot product space. Then for all  $\mathbf{v}, \mathbf{w} \in V$ ,  $\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$ .*

PROOF

$$\begin{aligned} \mathbf{v} \cdot \mathbf{w} &\leq \|\mathbf{v}\| \|\mathbf{w}\| \quad (\text{C-S Inequality}) \\ \Rightarrow 2\mathbf{v} \cdot \mathbf{w} &\leq 2\|\mathbf{v}\| \|\mathbf{w}\| \\ \Rightarrow \mathbf{v} \cdot \mathbf{v} + 2\mathbf{v} \cdot \mathbf{w} + \mathbf{w} \cdot \mathbf{w} &\leq \|\mathbf{v}\|^2 + 2\|\mathbf{v}\| \|\mathbf{w}\| + \|\mathbf{w}\|^2 \\ \Rightarrow (\mathbf{v} + \mathbf{w}) \cdot (\mathbf{v} + \mathbf{w}) &\leq (\|\mathbf{v}\| + \|\mathbf{w}\|)^2 \\ \Rightarrow \|\mathbf{v} + \mathbf{w}\|^2 &\leq (\|\mathbf{v}\| + \|\mathbf{w}\|)^2. \end{aligned}$$

Take the square root of both sides to get the desired result.  $\square$

**Corollary 4.35 (Generalized Triangle Inequality)** *Let  $V$  be a dot product space. Then for all  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n \in V$ ,*

$$\left\| \sum_{j=1}^n \mathbf{v}_j \right\| \leq \sum_{j=1}^n \|\mathbf{v}_j\|.$$

**Definition 4.36** *Let  $V$  be a dot product space. Given nonzero vectors  $\mathbf{v}, \mathbf{w} \in V$ , the angle between  $\mathbf{v}$  and  $\mathbf{w}$  is*

$$\theta = \arccos \left( \frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{v}\| \|\mathbf{w}\|} \right).$$

**Note:**  $\frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{v}\| \|\mathbf{w}\|}$  is guaranteed to be between  $-1$  and  $1$  by the C-S Inequality.

Rewritten, this definition says  $\cos \theta = \frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{v}\| \|\mathbf{w}\|}$ .

Rewritten further, this says  $\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\| \|\mathbf{w}\| \cos \theta$ . (This formula should look familiar if you have already taken Calculus III or have seen vectors in physics.)

## EXAMPLE 12

Compute the angle between the vectors  $(0, 1)$  and  $(1, 1)$  in  $\mathbb{R}^2$ .

## 4.7 Hyperplanes

**Definition 4.37** Let  $V$  be a real vector space of dimension  $n < \infty$ . A **hyperplane** is an affine subspace of  $V$  whose dimension is  $n - 1$ .

## EXAMPLES

- In  $\mathbb{R}$  (dimension 1), hyperplanes
- In  $\mathbb{R}^2$  (dimension 2), hyperplanes
- In  $\mathbb{R}^3$  (dimension 3), hyperplanes

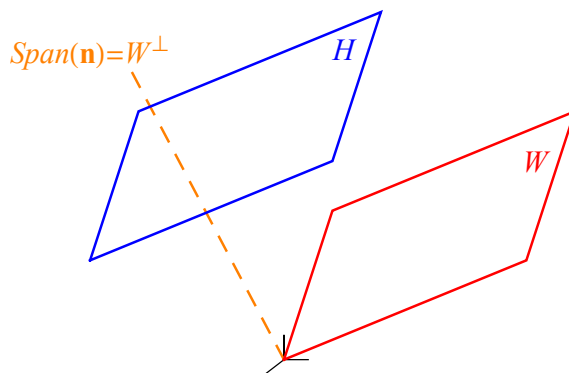
**Theorem 4.38** Let  $V$  be a finite-dimensional dot product space and let  $H$  be a hyperplane. Then there is a nonzero vector  $\mathbf{n} \in V$ , called a **normal vector** to the hyperplane, and a scalar  $d \in \mathbb{R}$  such that

$$\mathbf{x} \in H \iff \mathbf{n} \cdot \mathbf{x} = d.$$

The equation  $\mathbf{n} \cdot \mathbf{x} = d$  is called a **normal equation** or **standard equation** of the hyperplane.

PROOF Let  $W$  be the subspace of  $V$  associated to  $H$ .

Then  $\dim W^\perp = \dim V - \dim W = n - (n - 1) = 1$ , so  $W^\perp$  is spanned by a single nonzero vector. Call such a vector  $\mathbf{n}$ .



Next, select  $\mathbf{x}_0 \in H$  (so that  $H = W + \mathbf{x}_0$ ) and let  $d = \mathbf{n} \cdot \mathbf{x}_0$ .

We claim that for this choice of  $\mathbf{n}$  and  $d$ ,  $\mathbf{x} \in H \iff \mathbf{n} \cdot \mathbf{x} = d$ .

( $\Rightarrow$ ) Suppose  $\mathbf{x} \in H$ .

Then  $\mathbf{x} - \mathbf{x}_0 \in W$  so  $(\mathbf{x} - \mathbf{x}_0) \perp \mathbf{n}$ .

Thus  $(\mathbf{x} - \mathbf{x}_0) \cdot \mathbf{n} = 0$  so  $\mathbf{x} \cdot \mathbf{n} = \mathbf{x}_0 \cdot \mathbf{n} = d$ .

( $\Leftarrow$ ) Suppose  $\mathbf{n} \cdot \mathbf{x} = d$ .

Then  $\mathbf{n} \cdot (\mathbf{x} - \mathbf{x}_0) = d - d = 0$  so  $\mathbf{x} - \mathbf{x}_0 \in W$ .

Thus  $\mathbf{x} \in W + \mathbf{x}_0 = H$ .  $\square$

#### EXAMPLE 13

Let  $V = \mathbb{R}^2$  and let  $H$  be the hyperplane  $y = -2x + 4$ .

## EXAMPLE 14

Let  $V = \mathbb{R}^3$  and let  $H$  be the hyperplane passing through  $(0, 1, 4)$ ,  $(-2, -3, 8)$  and  $(2, 2, -1)$ . Write a normal equation of  $H$ .

## EXAMPLE 15

Find a normal equation of the plane passing through  $(3, -1, 2)$  and containing the line whose parametric equations are  $x = 2t + 1$ ,  $y = t - 5$ ,  $z = -7t + 4$ .

**EXAMPLE 16**

The two planes with equations  $x + y + z = 3$  and  $2x - 3y + 4z = 9$  intersect in a line. Find parametric equations of this line.



## Cross products

**Definition 4.39** Let  $\mathbf{a} = (a_1, a_2, a_3)$  and  $\mathbf{b} = (b_1, b_2, b_3)$  be two vectors in  $\mathbb{R}^3$ . The cross product of  $\mathbf{a}$  and  $\mathbf{b}$ , denoted  $\mathbf{a} \times \mathbf{b}$ , is the vector

$$\mathbf{a} \times \mathbf{b} = (a_2b_3 - a_3b_2, a_3b_1 - a_1b_3, a_1b_2 - a_2b_1).$$

**EXAMPLE:**  $(1, 4, -3) \times (2, -1, 3) =$

Some of you may know an alternate formula for cross product using determinants. You may use that formula if you know it (we'll get to it later in this course).

**Theorem 4.40 (Properties of cross product)** Let  $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^3$ . Then:

1.  $(\mathbf{a} \times \mathbf{b}) \perp \mathbf{a}$  and  $(\mathbf{a} \times \mathbf{b}) \perp \mathbf{b}$ ;
2.  $\mathbf{a} \parallel \mathbf{b}$  if and only if  $\mathbf{a} \times \mathbf{b} = \mathbf{0}$ ;
3.  $\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}$ ;
4.  $(\mathbf{a} + \mathbf{b}) \times \mathbf{c} = (\mathbf{a} \times \mathbf{c}) + (\mathbf{b} \times \mathbf{c})$  and  $\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}$ ;
5.  $r\mathbf{a} \times \mathbf{b} = r(\mathbf{a} \times \mathbf{b}) = \mathbf{a} \times r\mathbf{b}$  for any  $r \in \mathbb{R}$ ;
6. If  $\mathbf{a}$  and  $\mathbf{b}$  are not parallel, then  $(\text{Span}(\mathbf{a}, \mathbf{b}))^\perp = \text{Span}(\mathbf{a} \times \mathbf{b})$ ;
7.  $\|\mathbf{a} \times \mathbf{b}\| = \|\mathbf{a}\| \|\mathbf{b}\| \sin \theta$ , where  $\theta$  is the angle between  $\mathbf{a}$  and  $\mathbf{b}$ .

**PROOF** These are all direct calculations using the definition; the important fact in linear algebra is statement (1), so we will prove the first part of that statement here:

$$\begin{aligned} (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{a} &= (a_2b_3 - a_3b_2, a_3b_1 - a_1b_3, a_1b_2 - a_2b_1) \cdot (a_1, a_2, a_3) \\ &= a_1(a_2b_3 - a_3b_2) + a_2(a_3b_1 - a_1b_3) + a_3(a_1b_2 - a_2b_1) \\ &= a_1a_2b_3 - a_1b_2a_3 + b_1a_2a_3 - a_1a_2b_3 + a_1b_2a_3 - b_1a_2a_3 \\ &= 0. \quad \square \end{aligned}$$

## EXAMPLE 14, REVISITED

Let  $V = \mathbb{R}^3$  and let  $H$  be the hyperplane passing through  $(0, 1, 4)$ ,  $(-2, -3, 8)$  and  $(2, 2, -1)$ . Write a normal equation of  $H$ .

## 4.8 Summary of Chapter 4

1. The dot product of two vectors is a scalar. Each vector space has its own formula for dot product; examples of dot product include:

- if  $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ ,  $\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^T \mathbf{w} = \sum_{j=1}^n v_j w_j$ .
- if  $f, g \in C([a, b], \mathbb{R})$ ,  $f \cdot g = \int_a^b f(x)g(x) dx$ .
- if  $A, B \in M_n(\mathbb{R})$ ,  $A \cdot B = \text{tr}(AB^T)$ .

There are vector spaces with no dot product, but if  $V$  has a dot product formula, then we call  $V$  a **dot product space**.

2. Dot products are useful because they lead to geometric descriptions of vectors. In particular, if  $V$  is a dot product space we can define:

- a) The **norm** (length) of a vector  $\mathbf{v} \in V$  is the scalar  $\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$ .  
(In other words,  $\mathbf{v} \cdot \mathbf{v} = \|\mathbf{v}\|^2$ .)
- b) The **distance** between two vectors  $\mathbf{v}, \mathbf{w} \in V$  is the scalar  $\text{dist}(\mathbf{v}, \mathbf{w}) = \|\mathbf{v} - \mathbf{w}\|$ .
- c) A **unit vector** is a vector whose norm is 1.  
Given any nonzero  $\mathbf{v}$ , there is always a unit vector  $\mathbf{u}$  in the same direction as  $\mathbf{v}$ , called a **normalized version of  $\mathbf{v}$** , which is computed by setting  $\mathbf{u} = \frac{1}{\|\mathbf{v}\|} \mathbf{v}$ .
- d) Two vectors  $\mathbf{v}$  and  $\mathbf{w}$  are called **orthogonal** (written  $\mathbf{v} \perp \mathbf{w}$ ) if  $\mathbf{v} \cdot \mathbf{w} = 0$ . Orthogonality generalizes the idea of “perpendicularity” from planar geometry.
- e) The **angle**  $\theta$  between nonzero vectors  $\mathbf{v}$  and  $\mathbf{w}$  is  $\theta = \arccos\left(\frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{v}\| \|\mathbf{w}\|}\right)$ .  
(In other words,  $\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\| \|\mathbf{w}\| \cos \theta$ .)

The reason we can define angles this way is because of the Cauchy-Schwarz Inequality, which says that for any  $\mathbf{v}, \mathbf{w} \in V$ ,

$$|\mathbf{v} \cdot \mathbf{w}| \leq \|\mathbf{v}\| \|\mathbf{w}\|.$$

3. Dot product in  $\mathbb{R}^m$  and  $\mathbb{R}^n$  satisfies the important **dual relation**: if  $A \in M_{mn}(\mathbb{R})$ ,  $\mathbf{x} \in \mathbb{R}^n$  and  $\mathbf{y} \in \mathbb{R}^m$ , then

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

4. A basis  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  of dot product space (or subspace)  $V$  is called **orthonormal** if the vectors in the basis are pairwise orthogonal, and if every vector in the basis is a unit vector.

a) If  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  is an orthonormal basis of  $V$ , then every  $\mathbf{v} \in V$  satisfies

$$\mathbf{v} = \sum_{j=1}^n (\mathbf{v} \cdot \mathbf{x}_j) \mathbf{x}_j.$$

b) To find an orthonormal basis of  $V$ , start with any basis  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  of  $V$  and apply the **Gram-Schmidt procedure** to convert this basis into an orthonormal basis.

5. Given a dot product space  $V$  and a subspace  $W$  of  $V$ , define the **orthogonal complement**  $W^\perp$  to be the set of vectors which are orthogonal to every  $\mathbf{w} \in W$ .

- $W^\perp$  is a subspace of  $V$ , and  $\dim V = \dim W + \dim W^\perp$ .
- $W \cap W^\perp = \{\mathbf{0}\}$ .
- $(W^\perp)^\perp = W$  if  $V$  is finite dimensional.
- $\mathbf{v} \in W^\perp$  if and only if  $\mathbf{v}$  is orthogonal to every vector in a spanning set (or basis) of  $W$ .
- Given any  $\mathbf{v} \in V$ , we can write  $\mathbf{v} = \pi_W(\mathbf{v}) + \pi_{W^\perp}(\mathbf{v})$  where  $\pi_{W^\perp}(\mathbf{v}) \in W$  and  $\pi_W(\mathbf{v}) \in W^\perp$ .

The  $\pi_W(\mathbf{v})$  so obtained is called the **projection of  $\mathbf{v}$  onto  $W$** .

a) If  $\dim W = 1$ , then  $W = \text{Span}(\mathbf{w})$  for some nonzero  $\mathbf{w}$ ; in this case

$$\pi_W(\mathbf{v}) = \pi_{\mathbf{w}}\mathbf{v} = \frac{\mathbf{v} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}} \mathbf{w}.$$

b) In general,  $\pi_W(\mathbf{v})$  can be computed by taking a basis  $\{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  of  $W$ , converting that basis into an orthonormal basis  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  via Gram-Schmidt, and then using the **Projection Formula**:

$$\pi_W(\mathbf{v}) = \sum_{j=1}^n (\mathbf{v} \cdot \mathbf{x}_j) \mathbf{x}_j.$$

c) To compute  $\pi_{W^\perp}(\mathbf{v})$ , first compute  $\pi_W(\mathbf{v})$  and then set

$$\pi_{W^\perp}(\mathbf{v}) = \mathbf{v} - \pi_W(\mathbf{v}).$$

d)  $\text{dist}(\mathbf{v}, W)$ , the minimum distance from vector  $\mathbf{v}$  to subspace  $W$ , satisfies  $\text{dist}(\mathbf{v}, W) = \|\pi_{W^\perp}(\mathbf{v})\|$ .

6. A **hyperplane** in vector space  $V$  is an affine subspace whose dimension is 1 less than the dimension of  $V$ .

a) If  $V$  is a dot product space, then every hyperplane has a **normal equation**  $\mathbf{n} \cdot \mathbf{x} = d$ .

$\mathbf{n}$  is called a **normal vector** to the hyperplane.

b) If  $V = \mathbb{R}^2$ , hyperplanes are lines, and the normal equation of a line is  $ax + by = d$ . (Here, the normal vector is  $\mathbf{n} = (a, b)$ .)

c) If  $V = \mathbb{R}^3$ , hyperplanes are planes, and the normal equation of a plane is

$$ax + by + cz = d$$

(here, the normal vector is  $\mathbf{n} = (a, b, c)$ ).

Given two nonparallel vectors  $\mathbf{a}$  and  $\mathbf{b}$  in  $\mathbb{R}^3$ , the normal vector  $\mathbf{n}$  to the plane containing  $\mathbf{a}$  and  $\mathbf{b}$  is  $\mathbf{n} = \mathbf{a} \times \mathbf{b}$ .

## 4.9 Homework exercises for Chapter 4

1. Let  $\mathbf{x}$  and  $\mathbf{y}$  be vectors in  $\mathbb{R}^4$ ; let  $\mathbf{v}$  and  $\mathbf{w}$  be vectors in  $\mathbb{R}^3$ , and let  $A$  be a  $4 \times 3$  matrix. Determine whether each of the following quantities are a **scalar**, a **vector** (in which case you should say if it belongs to  $\mathbb{R}^3$  or  $\mathbb{R}^4$ ), a **matrix** (in which case you should give its size), or **nonsense**.

*Note:* For the purposes of this problem, matrices with a single column should be called vectors.

- |   |  |
|---|--|
| a) $\mathbf{x} \cdot \mathbf{x}$                    | e) $(\mathbf{x} + 2\mathbf{y}) \cdot \mathbf{y}$ |
| b) $\mathbf{x} \cdot \mathbf{x} \cdot \mathbf{x}$   | f) $\mathbf{v} \cdot A\mathbf{x}$                |
| c) $(\mathbf{x} \cdot \mathbf{x}) \cdot \mathbf{x}$ | g) $A\mathbf{v} \cdot A\mathbf{w}$               |
| d) $(\mathbf{x} \cdot \mathbf{x})\mathbf{x}$        | h) $\mathbf{x} \cdot A^T\mathbf{y}$              |

2. Same directions as the previous problem (where  $\mathbf{v}$ ,  $\mathbf{w}$ ,  $A$ ,  $\mathbf{x}$  and  $\mathbf{y}$  are the same types of objects as in the previous problem):

- |   |   |
|---|---|
| a) $(\mathbf{v} \cdot \mathbf{w})A\mathbf{x}$                     | e) $\ \mathbf{y}\ \mathbf{y}$                                       |
| b) $A((\mathbf{x} \cdot \mathbf{y})A\mathbf{w} \cdot \mathbf{y})$ | f) $\ \mathbf{w}\ \mathbf{x}$                                       |
| c) $\text{dist}(\mathbf{x}, \mathbf{x})$                          | g) $\text{dist}(\mathbf{w}, 2\mathbf{v})A$                          |
| d) $\ 2\mathbf{y} + \mathbf{x}\ $                                 | h) $\text{dist}(\mathbf{x}, \mathbf{w})\mathbf{y} \cdot \mathbf{x}$ |

3. Let  $\mathbf{v} = (-2, 2, -1)$ , let  $\mathbf{w} = (0, 1, 2)$  and let  $\mathbf{x} = (3, -1, -1)$ . Compute the following:

- |   |   |
|---|---|
| a) $\mathbf{v} \cdot \mathbf{w}$                | e) $\ \mathbf{v}\ $                                     |
| b) $\mathbf{x} \cdot 2\mathbf{w}$               | f) $\ \mathbf{v} + 2\mathbf{w} - \mathbf{x}\ $          |
| c) $\mathbf{x} \cdot (\mathbf{v} - \mathbf{x})$ | g) $\frac{\mathbf{v}}{\ \mathbf{v}\ }$                  |
| d) $\text{dist}(\mathbf{v}, \mathbf{w})$        | h) $\text{dist}(3\mathbf{v}, 2\mathbf{x} - \mathbf{w})$ |

4. Suppose  $\mathbf{v}$ ,  $\mathbf{w}$ ,  $\mathbf{x}$  are vectors in some dot product space  $V$  such that  $\mathbf{v} \cdot \mathbf{w} = 4$ ,  $\|\mathbf{v}\| = 5$ ,  $\mathbf{v} \cdot \mathbf{x} = -3$  and  $\mathbf{x} \cdot \mathbf{x} = 20$ . Compute the following:

- |  |  |
|--|--|
| a) $-\mathbf{x} \cdot 2\mathbf{v}$                         | e) $\ \mathbf{x}\ $  |
| b) $\mathbf{v} \cdot (4\mathbf{x} - 3\mathbf{w})$          | f) $(\mathbf{v} + \mathbf{x}) \cdot (\mathbf{v} + \mathbf{x})$   |
| c) $\mathbf{v} \cdot \mathbf{v}$                           | g) $2\mathbf{x} \cdot \mathbf{w} - \mathbf{x} \cdot 2\mathbf{w}$ |
| d) $\mathbf{w} \cdot \frac{1}{\ \mathbf{w}\ ^2}\mathbf{w}$ | h) $\mathbf{v} \cdot \ \mathbf{v}\ \mathbf{v}$                   |

5. Consider the dot product space  $C([-1, 1], \mathbb{R})$ , and let  $f(x) = x$  and  $g(x) = x^2$ . Compute the following:

- |                |                        |
|----------------|------------------------|
| a) $f \cdot g$ | d) $\ f + g\ $         |
| b) $f \cdot f$ | e) $\text{dist}(f, g)$ |
| c) $\ f\ $     | f) $\text{dist}(f, 2)$ |

6. Find all real numbers  $k$  such that the vectors  $(1, 2, k)$  and  $(2, 4, 3)$  are distance 3 apart.
7. Let  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$ . Suppose that for *some* nonzero  $\mathbf{v} \in \mathbb{R}^3$ ,  $\mathbf{v} \cdot \mathbf{a} = \mathbf{v} \cdot \mathbf{b}$ . Is it necessarily the case that  $\mathbf{a} = \mathbf{b}$ ? Justify your answer (this means prove the result if it is true, and provide a counterexample if it is false).
8. Let  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$ . Suppose that for *every*  $\mathbf{v} \in \mathbb{R}^3$ ,  $\mathbf{v} \cdot \mathbf{a} = \mathbf{v} \cdot \mathbf{b}$ . Prove that  $\mathbf{a} = \mathbf{b}$ .  
*Hint:* Write  $\mathbf{a} = (a_1, a_2, a_3)$  and  $\mathbf{b} = (b_1, b_2, b_3)$ . Make some clever choices of  $\mathbf{v}$  and use the hypothesis of this problem to show that  $a_1 = b_1$ ,  $a_2 = b_2$  and  $a_3 = b_3$  (which means  $\mathbf{a} = \mathbf{b}$ ).
9. Let  $\mathbf{v} = (-2, 0, 1, 3)$  and  $\mathbf{w} = (1, 5, -2, 3)$ .
- Find  $\pi_{\mathbf{w}}(\mathbf{v})$ , the projection of  $\mathbf{v}$  onto  $\mathbf{w}$ .
  - Find the projection of  $\mathbf{w}$  onto  $\mathbf{v}$ .
  - Find  $\pi_{\mathbf{w}^\perp}(\mathbf{v})$ , the component of  $\mathbf{v}$  orthogonal to  $\mathbf{w}$ .
  - Find the cosine of the angle between  $\mathbf{v}$  and  $\mathbf{w}$ .
  - Find a unit vector in the same direction as  $\mathbf{w}$ .
  - Find a vector of length 8 in the same direction as  $\mathbf{w}$ .
- Hint:* Start with the answer to part (e).
- Determine whether the vectors  $\mathbf{v}$  and  $(1, 1, 5, -1)$  are orthogonal.
  - If  $(3, 1, 4x, 2x) \perp \mathbf{v}$ , find the value of  $x$ .
10. Without using any methods from calculus, find the point on the line  $y = 3x$  which is closest to the point  $(8, 3)$ .  
*Hint:* This has something to do with projections. Draw a picture and identify some vectors to help you figure out what to compute.
11. Use the Gram-Schmidt procedure to convert each of the following sets of vectors into an orthonormal basis of the space they span:
- $\{(3, -4, 5), (-3, 14, -7)\}$
  - $\{(1, -4, 0, 1), (7, -7, -4, 1), (6, 3, 6, -3)\}$
  - $\{(1, -1, 0, 1, 1), (3, -3, 2, 5, 5), (5, 1, 3, 2, 8)\}$
12. Let  $\mathbf{v} = (1, 3, 5)$  and let  $W = \text{Span}((1, 3, -2), (5, 1, 4))$ .

- a) Compute the projection of  $\mathbf{v}$  onto  $W$ .
- b) Compute the component of  $\mathbf{v}$  orthogonal to  $W$ .
- c) Find the distance between  $\mathbf{v}$  and  $W$ .
- d) Find the element of  $W$  which is closest to  $\mathbf{v}$ .
13. Let  $V = C([0, 1], \mathbb{R})$ . Consider the subspace  $W$  of  $V$  consisting of functions of the form  $f(x) = mx + b$ , where  $m$  and  $b$  are constants.
- a) Find a basis of  $W$ .
- b) Use the Gram-Schmidt procedure to convert your answer to part (a) into an orthonormal basis of  $W$ .
- c) Let  $g(x) = x^2$ . Find the projection of  $g$  onto  $W$ .  
*Note:* The answer to part (c) should be thought of as the linear function which best approximates the function  $g(x) = x^2$  on the interval  $[0, 1]$ .
14. Find the linear function which best approximates the function  $f(x) = x^3 - x^2$  on the interval  $[0, 1]$ .
15. Find the linear function which best approximates the function  $f(x) = x^3 - 4x^2 + 4x$  on the interval  $[0, 3]$ .
16. Suppose an object sits on the plane in  $\mathbb{R}^3$  that is described as the set of points satisfying the equation  $8x - 5y + z = 0$ . If a force is applied to the object in the direction  $\mathbf{f} = (2, -1, -3)$ , in what direction will the object move? (Ignore the effect of friction, and assume that the force is sufficiently large to move the object.)  
*Hint:* This has something to do with projections.
17. In each part of this problem, you are given a vector space  $V$  and a subspace  $W$  of  $V$ . Describe the orthogonal complement  $W^\perp$  (by “describe”, I mean give the dimension of  $W^\perp$  and write down a basis of  $W^\perp$ ).
- a)  $V = \mathbb{R}^2; W = \text{Span}((3, 5))$
- b)  $V = \mathbb{R}^2; W = \text{the } y\text{-axis}$
- c)  $V = \mathbb{R}^2; W = \{(x, y) : 3x - 7y = 0\}$
- d)  $V = \mathbb{R}^3; W = \text{Span}((2, 1, -3), (0, 2, -5))$
- e)  $V = \mathbb{R}^3; W = \text{Span}((2, 1, -3))$
- f)  $V = \mathbb{R}^3; W = \{(x, y, z) : 2x + 3y + z = 0\}$   
*Hint:* First, find a basis of  $W$ .

18. In each part of this problem, you are given a vector space  $V$  and a subspace  $W$  of  $V$ . Describe the orthogonal complement  $W^\perp$ .



- a)  $V = \mathbb{R}^4; W = \text{Span}((1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 0, 1))$
- b)  $V = \mathbb{R}^4; W = \text{Span}((0, 3, 0, 0))$
- c)  $V = \mathbb{R}^4; W = \text{Span}((1, -2, 5, 2))$
- d)  $V = \mathbb{R}^4; W = \text{Span}((1, 0, -1, 0), (-4, -1, 0, 1))$
19. Prove (using vectors and dot products) that the diagonals of a rhombus are perpendicular.
- Hint:* Draw a rhombus and think of the sides as vectors; give them names like  $\mathbf{v}$  and  $\mathbf{w}$ . (You should only need use two letters for the sides, even though there are four sides.) Since the shape you drew is a rhombus, what is assumed to be true about  $\mathbf{v}$  and  $\mathbf{w}$ ? Draw the diagonals of the rhombus, figure out what they are in terms of  $\mathbf{v}$  and  $\mathbf{w}$ , and then show they are orthogonal.
20. Prove (using vectors and dot products) that if the diagonals of a parallelogram have the same length, then the parallelogram is a rectangle.
21. Let  $S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  be a set of nonzero vectors in an inner product space such that for all  $i \neq j$ ,  $\mathbf{v}_i \perp \mathbf{v}_j$ . Prove that  $S$  is a linearly independent set of vectors.
- Hint:* Suppose  $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_n\mathbf{v}_n = \mathbf{0}$ . Now take the dot product of both sides of this equation with  $\mathbf{v}_1$ , and use the hypothesis of orthogonality (and some other stuff) to explain why  $c_1 = 0$ . Proceed from there.
22. a) Write the normal equation of the plane in  $\mathbb{R}^3$  passing through the points  $(2, 1, -3)$ ,  $(4, -3, 2)$  and  $(1, -5, 1)$ .
- b) The line in  $\mathbb{R}^3$  passing through  $(1, 0, 2)$  and  $(3, -1, 5)$  intersects the plane described in part (a) in one point. Find the coordinates of this point.

## Chapter 5

---

# Linear transformations

---

### 5.1 Introduction

#### RECALL

In Chapter 1 we discussed the problem of solving a system of several equations in several variables. Here is an example:

$$\begin{cases} 3w + 2x - y + 5z = 8 \\ 2w - x + 4y - z = -3 \\ w + 5x + 2y + 6z = 17 \end{cases}$$

We are interested in the following questions about this system:

- 1.
- 2.
- 3.

#### BIG PICTURE

Think of the **left hand side** of this system **as a single function** called  $T$  and we think of the four **variables**  $(w, x, y, z)$  **as a single variable**  $\mathbf{x}$ . Last, we group the **right-hand side** of the equation  $(8, -3, 17)$  into a **single vector** called  $\mathbf{b}$ . Then the system above **becomes a single equation**

$$T(\mathbf{x}) = \mathbf{b}.$$

At this point in the course, we know quite a bit about  $\mathbf{x}$  and  $\mathbf{b}$ :

- $\mathbf{x} = (w, x, y, z)$  is a vector in  $\mathbb{R}^4$ ;
- $\mathbf{b} = (8, -3, 17)$  is a vector in  $\mathbb{R}^3$ .

In Chapters 3 and 4 we studied vector spaces (and subspaces, and dimension, and orthogonality, etc.) in detail, so we have a good idea of how to think about the  $\mathbf{x}$  and the  $\mathbf{b}$ . **What we don't know much about yet is the  $T$ .** All we know about  $T$  is that

It turns out that this  $T$  is something called a *linear transformation*. We will see that:

- every linear transformation has four subspaces associated to it which determine the answers to the three questions above (thus the need to fully understand subspaces), and
- these four subspaces come in two pairs of orthogonal complements (thus the need to understand dot products and orthogonality).

Developing an understanding of linear transformations and these associated subspaces is the content of this chapter. For now, we start with a definition:

**Definition 5.1** Let  $V_1$  and  $V_2$  be real vector spaces. A function  $T : V_1 \rightarrow V_2$  is called a **linear transformation** if, for every  $\mathbf{v}, \mathbf{w} \in V_1$  and every  $r \in \mathbb{R}$ ,

1.  $T$  **preserves addition**, meaning  $T(\mathbf{v} + \mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w})$ , and
2.  $T$  **preserves scalar multiplication**, meaning  $T(r\mathbf{v}) = rT(\mathbf{v})$ .

**Theorem 5.2 (Linear transformations preserve zero)** If  $T : V_1 \rightarrow V_2$  is a linear transformation, then  $T(\mathbf{0}) = \mathbf{0}$ .

PROOF Let  $\mathbf{v} \in V_1$  be arbitrary. Then  $T(\mathbf{0}) = T(0\mathbf{v}) = 0T(\mathbf{v}) = \mathbf{0}$ .  $\square$

## 5.2 The standard matrix of a linear transformation

EXAMPLE 1

---

Suppose  $T : \mathbb{R} \rightarrow \mathbb{R}$  is a linear transformation. What must be true about  $T$ ?

**Theorem 5.3** *Every linear transformation from  $\mathbb{R}$  to  $\mathbb{R}$  is of the form  $T(x) = ax$ , where  $a \in \mathbb{R}$  is a constant.*

EXAMPLE 2

---

What are all the linear transformations from  $\mathbb{R}^2$  to  $\mathbb{R}^3$ ?

*Solution:* Proceeding as in Example 1, suppose  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^3$  is linear.

Let  $(a, c, e) = T(1, 0)$  and let  $(b, d, f) = T(0, 1)$ .

Then for any  $(x, y) \in \mathbb{R}^2$ ,

$$\begin{aligned} T(x, y) &= T(x, 0) + T(0, y) \\ &= xT(1, 0) + yT(0, 1) \\ &= x(a, c, e) + y(b, d, f) \\ &= (ax, cx, ex) + (by, dy, fy) \\ &= (ax + by, cx + dy, ex + fy). \end{aligned}$$

This shows that every linear transformation  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^3$  is of the form

$$T(x, y) = (ax + by, cx + dy, ex + fy)$$

for suitable constants  $a, b, c, d, e$  and  $f$ .

There is another way to think about this solution that is useful:

The work we did above generalizes:

## 5.2. The standard matrix of a linear transformation

**Definition 5.4** The **standard basis of  $\mathbb{R}^n$**  is the basis  $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$  (written in that order) of  $\mathbb{R}^n$  where

$$\mathbf{e}_1 = (1, 0, 0, \dots, 0), \quad \mathbf{e}_2 = (0, 1, 0, \dots, 0), \quad \dots \quad \mathbf{e}_n = (0, 0, \dots, 0, 1)$$

In general,  $\mathbf{e}_j$  is a vector that has a 1 in the  $j^{\text{th}}$  position and 0s in all other positions.

**Note:** the standard basis  $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$  is an orthonormal basis of  $\mathbb{R}^n$ .

**Theorem 5.5** Let  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  be a linear transformation. Then there is a matrix  $A \in M_{mn}(\mathbb{R})$ , called the **standard matrix of  $T$**  such that  $T(\mathbf{x}) = A\mathbf{x}$  for all  $\mathbf{x} \in \mathbb{R}^n$ . Furthermore, the columns of  $A$  are, in order,  $T(\mathbf{e}_1)$ ,  $T(\mathbf{e}_2)$ , ..., and  $T(\mathbf{e}_n)$ .

PROOF Let  $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$  be the standard basis of  $\mathbb{R}^n$ . Let

$$A = \left( \begin{array}{cccc} \uparrow & \uparrow & & \uparrow \\ T(\mathbf{e}_1) & T(\mathbf{e}_2) & \cdots & T(\mathbf{e}_n) \\ \downarrow & \downarrow & & \downarrow \end{array} \right).$$

Now let  $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ ; we claim  $T(\mathbf{x}) = A\mathbf{x}$ . To check this, just compute:

$$\begin{aligned} A\mathbf{x} &= x_1T(\mathbf{e}_1) + x_2T(\mathbf{e}_2) + \dots + x_nT(\mathbf{e}_n) \quad (\text{by def'n of matrix multiplication}) \\ &= T(x_1\mathbf{e}_1) + T(x_2\mathbf{e}_2) + \dots + T(x_n\mathbf{e}_n) \quad (\text{since } T \text{ preserves scalar } \cdot) \\ &= T(x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \dots + x_n\mathbf{e}_n) \quad (\text{since } T \text{ preserves } +) \\ &= T(\mathbf{x}). \quad \square \end{aligned}$$

**Important Consequence:** Every linear transformation from  $\mathbb{R}^n$  to  $\mathbb{R}^m$  is really **matrix multiplication**, once you write down the standard matrix (which will be  $m \times n$ ).

## EXAMPLE 3

In each part of this example, you are given a linear transformation (take my word that these are all actually linear transformations). Find its standard matrix  $A$ , and then compute  $T(2, -1)$ .

- a)  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  rotates the plane by angle  $\theta$  radians counterclockwise about the origin.

- b)  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  defined by  $T(\mathbf{x}) = \mathbf{x}$  (the *identity transformation*).

*Solution:*  $T(\mathbf{e}_1) = T(1, 0) = (1, 0)$  and  $T(\mathbf{e}_2) = T(0, 1) = (0, 1)$ , so the standard matrix of  $T$  is

$$A = \begin{pmatrix} \begin{matrix} \uparrow \\ T(\mathbf{e}_1) \\ \downarrow \end{matrix} & \begin{matrix} \uparrow \\ T(\mathbf{e}_2) \\ \downarrow \end{matrix} \end{pmatrix} = \boxed{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}.$$

$$\text{Therefore } T(2, -1) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ -1 \end{pmatrix} = \boxed{\begin{pmatrix} 2 \\ -1 \end{pmatrix}} = \boxed{(2, -1)}.$$

## 5.2. The standard matrix of a linear transformation

---

c)  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  which projects vectors onto the subspace spanned by  $(4, 1)$ .

*Solution:*

Similarly,  $T(\mathbf{e}_2) = T(0, 1) = \pi_{(4,1)}(0, 1) = \frac{(0,1) \cdot (4,1)}{(4,1) \cdot (4,1)}(4, 1) = \frac{1}{17}(4, 1) = \left(\frac{4}{17}, \frac{1}{17}\right)$ .

Therefore the standard matrix of  $T$  is

$$A = \begin{pmatrix} \frac{16}{17} & \frac{4}{17} \\ \frac{4}{17} & \frac{1}{17} \end{pmatrix}$$

$$\text{and } T(2, -1) = \begin{pmatrix} \frac{16}{17} & \frac{4}{17} \\ \frac{4}{17} & \frac{1}{17} \end{pmatrix} \begin{pmatrix} 2 \\ -1 \end{pmatrix} = \boxed{\begin{pmatrix} \frac{28}{17} \\ \frac{7}{17} \end{pmatrix}} = \boxed{\left(\frac{28}{17}, \frac{7}{17}\right)}.$$

d)  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^4$  defined by  $T(x, y) = (3x + y, 5x - 2y, -y, -4x + 7y)$ .

*Solution:*  $T(\mathbf{e}_1) = T(1, 0) = (3, 5, 0, -4)$  and  $T(\mathbf{e}_2) = T(0, 1) = (1, -2, -1, 7)$ , so the standard matrix of  $T$  is

$$A = \boxed{\begin{pmatrix} 3 & 1 \\ 5 & -2 \\ 0 & -1 \\ -4 & 7 \end{pmatrix}}.$$

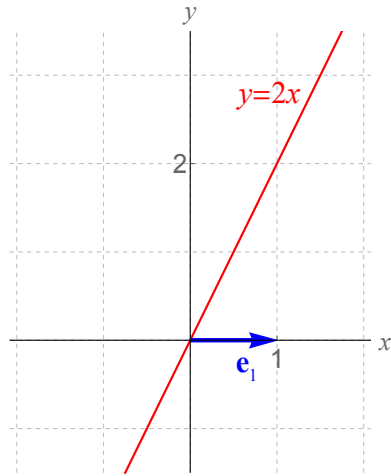
$$\text{Therefore } T(2, -1) = \begin{pmatrix} 3 & 1 \\ 5 & -2 \\ 0 & -1 \\ -4 & 7 \end{pmatrix} \begin{pmatrix} 2 \\ -1 \end{pmatrix} = \boxed{\begin{pmatrix} 5 \\ 12 \\ 1 \\ -15 \end{pmatrix}} = \boxed{(5, 12, 1, -15)}.$$



## 5.2. The standard matrix of a linear transformation

---

e)  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  where  $T$  reflects points through the line  $y = 2x$ .



**More examples with standard matrices****EXAMPLE 4**

Suppose  $T : \mathbb{R}^3 \rightarrow \mathbb{R}^2$  is a linear transformation such that  $T(1, 0, 0) = (2, -3)$ ,  $T(0, 1, 0) = (-2, 5)$  and  $T(0, 0, 1) = (1, 4)$ . Compute  $T(2, -1, 4)$ .

**EXAMPLE 5**

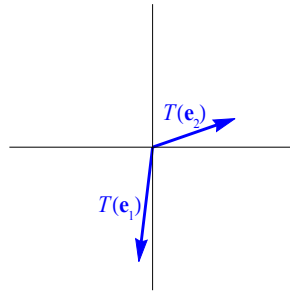
Suppose  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  is a linear transformation such that  $T(1, 5) = (3, -7)$  and  $T(2, -1) = (-5, -3)$ . Compute  $T(3, 2)$ .

Now that we know  $T(\mathbf{x}) = A\mathbf{x} = \begin{pmatrix} -2 & 1 \\ -2 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$ , we see that

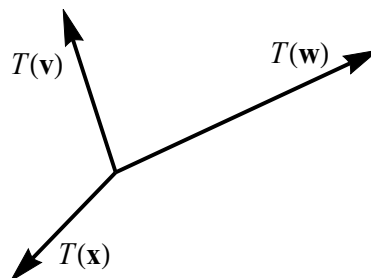
$$T(3, 2) = A \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} -2 & 1 \\ -2 & -1 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \boxed{\begin{pmatrix} -4 \\ -8 \end{pmatrix}} = \boxed{(-4, -8)}.$$

**Linear transformations and pictures****EXAMPLE 6**

Below, you are given sketches of  $T(\mathbf{e}_1)$  and  $T(\mathbf{e}_2)$  for some linear transformation  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ . On that picture, sketch the relative position of the vectors  $T(2, 0)$ ,  $T(\frac{1}{2}, 3)$  and  $T(-2, -1)$ .

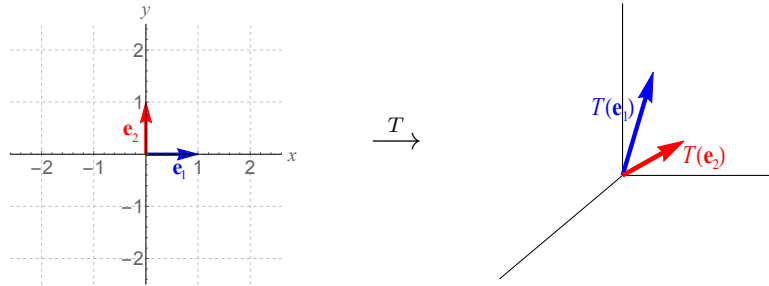
**EXAMPLE 7**

Below, you are given sketches of  $T(\mathbf{v})$ ,  $T(\mathbf{w})$  and  $T(\mathbf{x})$  for some linear transformation  $T$ . On that picture, sketch the relative position of the vectors  $T(\frac{3}{2}\mathbf{x})$ ,  $T(\mathbf{v} + \mathbf{x})$ ,  $T(\mathbf{w} - 2\mathbf{v})$  and  $T(2\mathbf{v} + \mathbf{w} + \mathbf{x})$ .

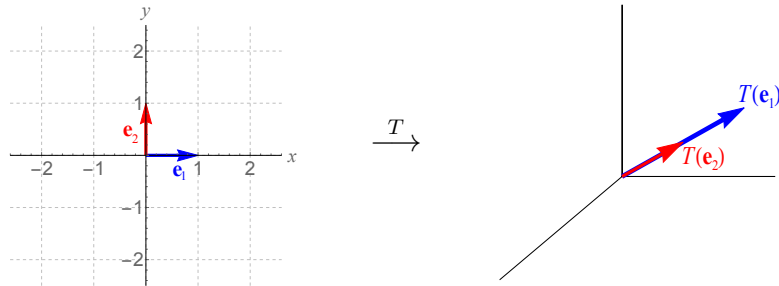


## 5.2. The standard matrix of a linear transformation

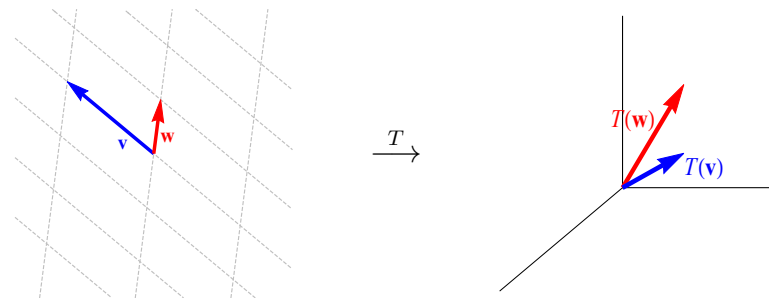
Examples 6 and 7 demonstrate how **linear transformations preserve grids**. For example, if you have a linear transformation  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^3$  with  $T(e_1)$  and  $T(e_2)$  as indicated, then the rest of the plane gets mapped as follows:



Sometimes, however, you get a picture like this:



The starting grid doesn't have to be orthonormal:

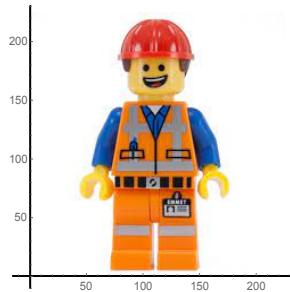


These ideas enable us to revisit one of the motivating problems from Chapter 1:

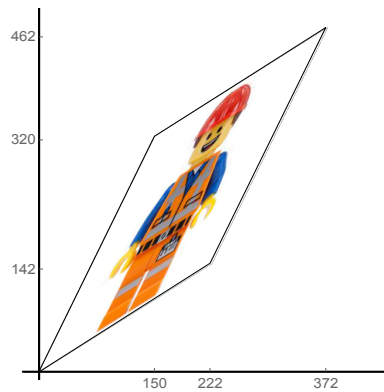
## 5.2. The standard matrix of a linear transformation

EXAMPLE 8 (THIS WAS EXAMPLE 6 FROM SECTION 1.1)

Here is a picture of Emmet from the LEGO movie, which is 225 pixels by 225 pixels:



Suppose you warp the image of Emmet so that it fits in this parallelogram:



- If a pixel was at position  $(x, y)$  in the old picture, what will its location be in the original picture? In other words, what “formula” is needed to distort the image so that it fits in the right space?
- Suppose that the left corner of Emmet’s mouth is the pixel that was originally at position  $(110, 170)$ . At what position will the left corner of Emmet’s mouth end up in the warped image?

## Compositions of linear transformations

### RECALL

To *compose* two functions means to do one, then the other:

**Theorem 5.6 (Compositions of linear transformations are linear)** *Let  $T_1 : V_1 \rightarrow V_2$  and  $T_2 : V_2 \rightarrow V_3$  be linear transformations. Then the transformation  $T_2 \circ T_1 : V_1 \rightarrow V_3$  is linear as well.*

PROOF First, we show  $T_2 \circ T_1$  preserves addition: let  $\mathbf{v}, \mathbf{w} \in V_1$ ; then

$$\begin{aligned} (T_2 \circ T_1)(\mathbf{v} + \mathbf{w}) &= T_2(T_1(\mathbf{v} + \mathbf{w})) \\ &= T_2(T_1(\mathbf{v}) + T_1(\mathbf{w})) \quad (\text{since } T_1 \text{ preserves } +) \\ &= T_2(T_1(\mathbf{v})) + T_2(T_1(\mathbf{w})) \quad (\text{since } T_2 \text{ preserves } +) \\ &= (T_2 \circ T_1)(\mathbf{v}) + (T_2 \circ T_1)(\mathbf{w}). \end{aligned}$$

Next, we show  $T_2 \circ T_1$  preserves scalar  $\cdot$ : let  $\mathbf{v} \in V_1$  and  $r \in \mathbb{R}$ ; then

$$\begin{aligned} (T_2 \circ T_1)(r\mathbf{v}) &= T_2(T_1(r\mathbf{v})) \\ &= T_2(rT_1(\mathbf{v})) \quad (\text{since } T_1 \text{ preserves scalar } \cdot) \\ &= rT_2(T_1(\mathbf{v})) \\ &= r(T_2 \circ T_1)(\mathbf{v}). \end{aligned}$$

Thus  $T_2 \circ T_1$  is linear, since it preserves  $+$  and scalar  $\cdot$ .  $\square$

**Theorem 5.7 (Standard matrix of composition is matrix product)** *Let  $T_1 : \mathbb{R}^n \rightarrow \mathbb{R}^m$  and  $T_2 : \mathbb{R}^m \rightarrow \mathbb{R}^p$  be linear transformations with standard matrices  $A_1 \in M_{mn}(\mathbb{R})$  and  $A_2 \in M_{pm}(\mathbb{R})$ , respectively. Then the standard matrix of  $T_2 \circ T_1 : \mathbb{R}^n \rightarrow \mathbb{R}^p$  is  $A_2A_1 \in M_{pn}(\mathbb{R})$ .*

PROOF Let  $\{\mathbf{e}_j\}$  represent standard basis elements of  $\mathbb{R}^n, \mathbb{R}^m, \mathbb{R}^p$ , etc.

Let the entries of  $A_1$  be denoted by  $a$ s and the entries of  $A_2$  be denoted by  $b$ s.

Then the  $(i, j)$ -entry of  $A_2A_1$ , by the definition of matrix multiplication, is

$$\sum_{l=1}^m b_{il}a_{lj}.$$

Now, the  $i^{\text{th}}$  entry of the vector  $(T_2 \circ T_1)(\mathbf{e}_j)$  is

$$\begin{aligned} [(T_2 \circ T_1)(\mathbf{e}_j)]_i &= [T_2(T_1(\mathbf{e}_j))]_i = [T_2(j^{\text{th}} \text{ column of } A_1)]_i \\ &= [T_2(a_{1j}, a_{2j}, \dots, a_{mj})]_i \\ &= \left[ T_2 \left( \sum_{l=1}^m a_{lj} \mathbf{e}_l \right) \right]_i \end{aligned}$$

From the previous page,

$$\begin{aligned}
 [(T_2 \circ T_1)(\mathbf{e}_j)]_i &= \left[ T_2 \left( \sum_{l=1}^m a_{lj} \mathbf{e}_l \right) \right]_i \\
 &= \left[ \sum_{l=1}^m a_{lj} T_2(\mathbf{e}_l) \right]_i \quad (\text{since } T_2 \text{ is linear}) \\
 &= \left[ \sum_{l=1}^m a_{lj} T_2(\mathbf{e}_l) \right]_i \\
 &= \sum_{l=1}^m a_{lj} \left[ l^{\text{th}} \text{ column of } A_2 \right]_i \\
 &= \sum_{l=1}^m a_{lj} [(i, l) \text{ - entry of } A_2] \\
 &= \sum_{l=1}^m b_{il} a_{lj}. \\
 &= (i, j) \text{ - entry of matrix } A_2 A_1.
 \end{aligned}$$

This proves that for all  $i$  and  $j$ ,  $[(T_2 \circ T_1)(\mathbf{e}_j)]_i$  is the  $(i, j)$ -entry of  $A_2 A_1$ .

Therefore the  $j^{\text{th}}$  column of the standard matrix of  $T_2 \circ T_1$  is the  $j^{\text{th}}$  column of  $A_2 A_1$ , i.e. the standard matrix of  $T_2 \circ T_1$  is  $A_2 A_1$ .  $\square$

**Importance:** This theorem further explains why matrix multiplication has the complicated definition that it does.

#### EXAMPLE 9

Find the standard matrix of the linear transformation  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  that first multiplies vectors by  $\begin{pmatrix} 2 & -1 \\ 3 & 0 \end{pmatrix}$  and then rotates the plane by  $\frac{\pi}{4}$  clockwise.

## 5.3 Proving whether a transformation is linear

### RECALL

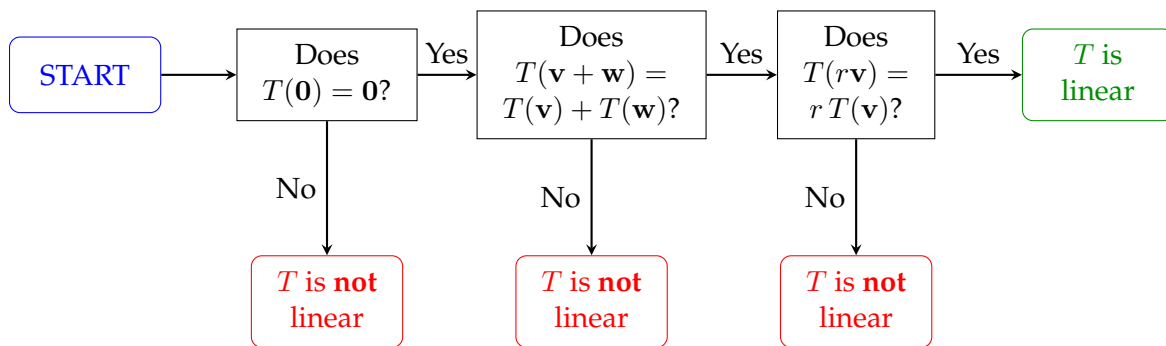
By definition,  $T : V_1 \rightarrow V_2$  is a linear transformation if

- $T$  preserves addition, meaning
- $T$  preserves scalar multiplication, meaning

We also proved a theorem in Section 5.1 which says that for any linear  $T$ ,

- $T$  preserves  $\mathbf{0}$  (i.e.  $T(\mathbf{0}) = \mathbf{0}$ ).

These three facts give you a good mechanism to decide whether a given function is a linear transformation. If you are given a function  $T : V_1 \rightarrow V_2$ , ask the following questions (this is called the **brute-force** method for determining whether or not  $T$  is a linear transformation):



This should remind you of the procedure we used in Chapter 4 to prove whether or not a subset of a vector space is a subspace.

**To prove that a transformation  $T : V_1 \rightarrow V_2$  is linear**, you need to do **both** of these things:

1. Take two generic elements of  $V_1$ , say  $\mathbf{v}$  and  $\mathbf{w}$ . Work out  $T(\mathbf{v} + \mathbf{w})$  and  $T(\mathbf{v}) + T(\mathbf{w})$  separately, and show they come out to the same thing.

**Be careful not to write  $T(\mathbf{v} + \mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w})$  until you know that is a true statement. (Don't assume what you want to prove!)**

2. Take a generic element of  $V_1$ , say  $\mathbf{v}$  and a generic scalar  $r$ . Work out  $T(r\mathbf{v})$  and  $rT(\mathbf{v})$  separately, and show they come out to the same thing.

**Be careful not to write  $T(r\mathbf{v}) = rT(\mathbf{v})$  until you know that is true.**



**To prove that a transformation  $T : V_1 \rightarrow V_2$  is **NOT** linear**, you need to do **one** of these three things:

1. Explain why  $T(\mathbf{0}) \neq \mathbf{0}$ .
2. Alternatively, write down two **specific** elements of  $V_1$  (i.e. **with numbers**), say  $\mathbf{v}$  and  $\mathbf{w}$ , and explain why  $T(\mathbf{v} + \mathbf{w}) \neq T(\mathbf{v}) + T(\mathbf{w})$ .
3. Alternatively, write down a **specific** element of  $V_1$  (i.e. **with numbers**), say  $\mathbf{v}$ , and a **specific** scalar (i.e. **a number**), say  $r$ , and explain why  $T(r\mathbf{v}) \neq rT(\mathbf{v})$ .

EXAMPLE 10

In each part, you are given a function  $T : V_1 \rightarrow V_2$ , where  $V_1$  and  $V_2$  are real vector spaces. Determine, with justification, whether or not  $T$  is a linear transformation.

a)  $V_1 = V_2 = \mathbb{R}^3$ .  $T(x_1, x_2, x_3) = (3x_1 + x_2 - x_3, 0, x_2)$ .

b)  $V_1 = V_2 = \mathbb{R}^2$ .  $T(x, y) = (5x - 1, y + x)$ .

c)  $V_1 = V_2 = M_3(\mathbb{R}), T(A) = A^T.$

*Solution:*

- $T(\mathbf{0}) = T(\mathbf{0}_{3 \times 3}) = (\mathbf{0}_{3 \times 3})^T = \mathbf{0}_{3 \times 3} = \mathbf{0}$
- $T(A + B) = (A + B)^T$   
 $T(A) + T(B) = A^T + B^T$
- $T(rA) = (rA)^T$
- $rT(A) = rA^T$

Therefore  $\boxed{T \text{ is linear}}$ .

d)  $V_1 = C(\mathbb{R}, \mathbb{R}); V_2 = \mathbb{R}^2. T(f) = (f(2), 3f(0)).$

e)  $V_1 = V_2 = \mathbb{R}; T(x) = |x|.$

*Solution:*

- $T(0) = |0| = 0.$
- $T(3 + (-2)) = T(1) = |1| = 1, \text{ but}$   
 $T(3) + T(-2) = |3| + |-2| = 3 + 2 = 5.$

Therefore  $\boxed{T \text{ is not linear}}$ .

f)  $V_1 = M_4(\mathbb{R}); V_2 = \mathbb{R}^4$ .  $T(A)$  = the 3rd column of  $A$ .

*Solution:* Let  $A \in M_4(\mathbb{R})$ . We can write  $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ \vdots & \vdots & \vdots & \vdots \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}$ . That makes  $T(A) = (a_{13}, a_{23}, a_{33}, a_{43})$ , and if we use similar notation for  $B \in M_4(\mathbb{R})$ , then  $T(B) = (b_{13}, b_{23}, b_{33}, b_{43})$ . Therefore:

g)  $V_1 = V_2 = \mathbb{R}^2$ .  $T(\mathbf{x})$  is the reflection of point  $\mathbf{x}$  through the line  $x = 3$ .

h)  $V_1 =$  the space of differentiable functions from  $\mathbb{R}$  to  $\mathbb{R}$ ;  $V_2 = \mathbb{R}$ .  $T(f) = f'(0)$ .

*Solution:*

- $T(\mathbf{0}) = T(\text{the constant function } f(x) = 0) = (0)'(0) = 0$ .
- $T(f + g) = (f + g)'(0)$   
 $T(f) + T(g) = f'(0) + g'(0)$
- $T(rf) = (rf)'(0)$   
 $rT(f) = r f'(0)$

Therefore  $T$  is linear.

## 5.4 Prototypical examples of linear transformations

### 1. Matrix multiplication (the most important example)

Every matrix  $A \in M_{mn}(\mathbb{R})$  defines a linear transformation  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  by  $T(\mathbf{x}) = A\mathbf{x}$ .

REASON:

### 2. Projection

Let  $V$  be any vector space and let  $\mathbf{w} \in V$  be nonzero. Then the map  $T : V \rightarrow V$  defined by  $T(\mathbf{v}) = \pi_{\mathbf{w}}\mathbf{v}$  is linear. (Recall that  $T(\mathbf{v}) = \pi_{\mathbf{w}}(\mathbf{v}) = \frac{\mathbf{v} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}}\mathbf{w}$ .)

REASON: •  $T(\mathbf{0}) = \pi_{\mathbf{w}}\mathbf{0} = \mathbf{0}$   
 •  $T(\mathbf{v}_1 + \mathbf{v}_2) = \pi_{\mathbf{w}}(\mathbf{v}_1 + \mathbf{v}_2) = \pi_{\mathbf{w}}(\mathbf{v}_1) + \pi_{\mathbf{w}}(\mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2)$   
 •  $T(r\mathbf{v}_1) = \pi_{\mathbf{w}}(r\mathbf{v}_1) = r\pi_{\mathbf{w}}(\mathbf{v}_1) = rT(\mathbf{v}_1)$ .

Projection onto a subspace (i.e. the map  $\pi_W : \mathbf{v} \mapsto \pi_W(\mathbf{v})$  for  $W \subseteq V$ ) is also linear.

### 3. Dot product by a fixed vector

Fix a vector  $\mathbf{v} \in \mathbb{R}^n$ . Then the transformations  $T(\mathbf{x}) = \mathbf{v} \cdot \mathbf{x}$  and  $T(\mathbf{x}) = \mathbf{x} \cdot \mathbf{v}$  are both linear transformations from  $\mathbb{R}^n$  to  $\mathbb{R}$  (the same kind of thing works for the other dot product spaces).

### 4. Differentiation

Let  $V = C^\infty(\mathbb{R}, \mathbb{R})$ , the space of functions which are infinitely differentiable. Then  $T : V \rightarrow V$  defined by  $T(f) = f'$  is linear.

REASON: •  $T(\mathbf{0}) = 0' = 0$   
 •  $T(f + g) = (f + g)' = f' + g' = T(f) + T(g)$   
 •  $T(rf) = (rf)' = r f' = rT(f)$ .

## 5. Definite integration

Let  $V = C([a, b], \mathbb{R})$ , the space of continuous functions from  $[a, b]$  to  $\mathbb{R}$ . Then  $T : V \rightarrow \mathbb{R}$ , defined by  $T(f) = \int_a^b f(x) dx$  is linear.

REASON: •  $T(\mathbf{0}) = \int_a^b 0 dx = 0$

$$\bullet T(f + g) = \int_a^b [f(x) + g(x)] dx = \int_a^b f(x) dx + \int_a^b g(x) dx = T(f) + T(g)$$

$$\bullet T(rf) = \int_a^b rf(x) dx = r \int_a^b f(x) dx = rT(f).$$

## 6. Evaluation

Let  $V$  be some vector space of functions and fix  $a \in \mathbb{R}$ . Then  $T : V \rightarrow \mathbb{R}$ , defined by  $T(f) = f(a)$ , is linear.

REASON: •  $T(\mathbf{0}) = 0(a) = 0$

$$\bullet T(f + g) = (f + g)(a) = f(a) + g(a) = T(f) + T(g)$$

$$\bullet T(rf) = (rf)(a) = r f(a) = rT(f).$$

EXAMPLE: If  $a = 0$ , then  $T(f) = f(0)$ .

That means  $T(e^x) = e^0 = 1$ ;  $T(x^3) = 0$ ;  $T(x^2 + 5) = 5$ ; etc.

## 7. Certain geometric transformations

like rotations about the origin, reflections, etc.

## 8. Compositions of linear transformations

If  $T_1 : V_1 \rightarrow V_2$  and  $T_2 : V_2 \rightarrow V_3$  are linear, then  $T_2 \circ T_1 : V_1 \rightarrow V_3$  is linear as well (this was proven earlier).

## 9. Sums of other linear transformations

**Theorem 5.8** Let  $T_1 : V_1 \rightarrow V_2$  and  $T_2 : V_1 \rightarrow V_2$  be linear transformations. Then the transformation  $T_1 + T_2 : V_1 \rightarrow V_2$ , defined by setting

$$(T_1 + T_2)(\mathbf{v}) = T_1(\mathbf{v}) + T_2(\mathbf{v})$$

for all  $\mathbf{v} \in V_1$ , is linear.

PROOF First, we show  $T_1 + T_2$  preserves addition: let  $\mathbf{v}, \mathbf{w} \in V_1$ ; then

$$\begin{aligned} (T_1 + T_2)(\mathbf{v} + \mathbf{w}) &= T_1(\mathbf{v} + \mathbf{w}) + T_2(\mathbf{v} + \mathbf{w}) \quad (\text{by definition of } T_1 + T_2) \\ &= T_1(\mathbf{v}) + T_1(\mathbf{w}) + T_2(\mathbf{v}) + T_2(\mathbf{w}) \quad (\text{since } T_1 \text{ and } T_2 \text{ are linear}) \\ &= (T_1 + T_2)(\mathbf{v}) + (T_1 + T_2)(\mathbf{w}) \quad (\text{by definition of } T_1 + T_2). \end{aligned}$$

Next, we show  $T_1 + T_2$  preserves scalar multiplication: let  $\mathbf{v} \in V_1$  and  $r \in \mathbb{R}$ ; then

$$\begin{aligned} (T_1 + T_2)(r\mathbf{v}) &= T_1(r\mathbf{v}) + T_2(r\mathbf{v}) \quad (\text{by definition of } T_1 + T_2) \\ &= rT_1(\mathbf{v}) + rT_2(\mathbf{v}) \quad (\text{since } T_1 \text{ and } T_2 \text{ are linear}) \\ &= r [T_1(\mathbf{v}) + T_2(\mathbf{v})] \\ &= r(T_1 + T_2)(\mathbf{v}) \quad (\text{by definition of } T_1 + T_2). \end{aligned}$$

## 10. Scalar multiples of other linear transformations

**Theorem 5.9** Let  $T : V_1 \rightarrow V_2$  be a linear transformation and let  $c \in \mathbb{R}$ . Then the transformation  $cT : V_1 \rightarrow V_2$ , defined by setting

$$(cT)(\mathbf{v}) = cT(\mathbf{v})$$

for all  $\mathbf{v} \in V_1$ , is linear.

PROOF HW (as a hint, this is similar to the proof of Theorem 5.8... you need to show  $cT$  preserves addition and scalar multiplication).

**Remark:** Theorems 5.8 and ?? prove that for fixed vector spaces  $V_1$  and  $V_2$ , the set of all linear transformations from  $V_1$  to  $V_2$  is itself a vector space! This vector space is denoted  $L(V_1, V_2)$ .

**Associated Notation:** Therefore, to say " $T \in L(V_1, V_2)$ " means that  $T$  is a linear transformation from  $V_1$  to  $V_2$ .

## 5.5 Kernels and images

### CONCEPT

A linear transformation  $T : V_1 \rightarrow V_2$  determines some subspaces of  $V_1$  and  $V_2$ . Understanding these subspaces tells us a lot about the linear transformation, and will tell us something about the solution of equations  $T(\mathbf{x}) = \mathbf{b}$ .

### Kernels

**Definition 5.10** Let  $T : V_1 \rightarrow V_2$  be a linear transformation. The **kernel** of  $T$ , denoted  $\ker(T)$ , is the subset of  $V_1$  defined by

$$\ker(T) = \{\mathbf{v} \in V_1 : T(\mathbf{v}) = \mathbf{0}\}.$$

What this means in English: the kernel of a linear transformation is the set of vectors in the domain which get mapped to  $\mathbf{0}$  under the transformation.

**Theorem 5.11 (Kernels are subspaces)** Let  $T : V_1 \rightarrow V_2$  be a linear transformation. Then  $\ker(T)$  is a subspace of  $V_1$ .

PROOF We show that  $\ker(T)$  has the essential characteristics of subspaces:

- Let  $\mathbf{v}, \mathbf{w} \in \ker(T)$ . Thus  $T(\mathbf{v}) = \mathbf{0}$  and  $T(\mathbf{w}) = \mathbf{0}$ .  
Then  $T(\mathbf{v} + \mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w}) = \mathbf{0} + \mathbf{0} = \mathbf{0}$  so  $\mathbf{v} + \mathbf{w} \in \ker(T)$ .  
Therefore  $\ker(T)$  is closed under  $+$ .
- Let  $\mathbf{v} \in \ker(T)$  and let  $r \in \mathbb{R}$ .  
Then  $T(r\mathbf{v}) = rT(\mathbf{v}) = r\mathbf{0} = \mathbf{0}$  so  $r\mathbf{v} \in \ker(T)$ .  
Therefore  $\ker(T)$  is closed under scalar  $\cdot$ .

Thus  $\ker(T)$  is a subspace of  $V_1$ .  $\square$

### Images

**Definition 5.12** Let  $T : V_1 \rightarrow V_2$  be a linear transformation. The **image** of  $T$ , denoted  $\text{im}(T)$  or  $\text{Im}(T)$ , is the subset of  $V_2$  defined by

$$\text{im}(T) = \{\mathbf{w} \in V_2 : \exists \mathbf{v} \in V_1 \text{ such that } T(\mathbf{v}) = \mathbf{w}\}.$$

What this means in English: the image of a linear transformation is its range, i.e. is the set of outputs of the linear transformation.

**Theorem 5.13 (Images are subspaces)** Let  $T : V_1 \rightarrow V_2$  be a linear transformation. Then  $\text{im}(T)$  is a subspace of  $V_2$ .

PROOF We show that  $\text{im}(T)$  has the essential characteristics of subspaces:

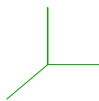
- Let  $\mathbf{w}_1, \mathbf{w}_2 \in \text{im}(T)$ .  
Thus there are vectors  $\mathbf{v}_1, \mathbf{v}_2 \in V_1$  such that  $T(\mathbf{v}_1) = \mathbf{w}_1$  and  $T(\mathbf{v}_2) = \mathbf{w}_2$ .  
Then  $T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2) = \mathbf{w}_1 + \mathbf{w}_2$  so  $\mathbf{w}_1 + \mathbf{w}_2 \in \text{im}(T)$ .  
Therefore  $\text{im}(T)$  is closed under  $+$ .
- Let  $\mathbf{w} \in \text{im}(T)$  and let  $r \in \mathbb{R}$ .  
Then there is a vector  $\mathbf{v} \in V_1$  such that  $T(\mathbf{v}) = \mathbf{w}$ .  
Now, notice that  $T(r\mathbf{v}) = rT(\mathbf{v}) = r\mathbf{w}$ , so  $r\mathbf{w} \in \text{im}(T)$ .  
Therefore,  $\text{im}(T)$  is closed under scalar  $\cdot$ .  
Thus  $\text{im}(T)$  is a subspace of  $V_2$ .  $\square$

**Definition 5.14** Let  $T : V_1 \rightarrow V_2$  be a linear transformation. The **rank** of  $T$ , denoted  $r$  or  $r(T)$ , is the dimension of the image of  $T$ :

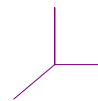
$$r(T) = \dim(\text{im}(T)).$$

### The key picture associated to linear transformations

A linear transformation  $T : V_1 \rightarrow V_2$  always suggests the following picture:



$V_1$   
 $\dim V_1 = n$



$V_2$   
 $\dim V_2 = m$



## Why do we care about kernels and images?

Suppose you have an equation of the form

$$T(\mathbf{x}) = \mathbf{b}$$

where  $T : V_1 \rightarrow V_2$  is linear (the goal is to find  $\mathbf{x} \in V_1$ , given  $T$  and  $\mathbf{b} \in V_2$ ).

What's the relevance of the image  $im(T)$ ?

What's the relevance of the kernel  $ker(T)$ ?

**Theorem 5.15 (Solution of  $T(\mathbf{x}) = \mathbf{b}$ )** Suppose  $T : V_1 \rightarrow V_2$  is a linear transformation. Then the equation  $T(\mathbf{x}) = \mathbf{b}$ ...

... has *no solution* if  $\mathbf{b} \notin im(T)$ .

... has *exactly one solution* if  $\mathbf{b} \in im(T)$  and  $ker(T) = \{\mathbf{0}\}$ .

... has *infinitely many solutions* if  $\mathbf{b} \in im(T)$  and  $ker(T) \neq \{\mathbf{0}\}$ .

Furthermore, if  $\mathbf{b} \in im(T)$ , then the *solution set* of the equation  $T(\mathbf{x}) = \mathbf{b}$  is the *affine subspace*  $\mathbf{x}_p + ker(T)$ , where  $\mathbf{x}_p$  is any **particular solution** of the equation.

PROOF First, if  $\mathbf{b} \notin im(T)$ , then  $T(\mathbf{x}) = \mathbf{b}$  has no solution (by def'n of image).

Henceforth, we assume  $\mathbf{b} \in im(T)$ .

By definition of image, there is  $\mathbf{x}_p \in V_1$  such that  $T(\mathbf{x}_p) = \mathbf{b}$ .

Let  $S$  be the set of solutions of  $T(\mathbf{x}) = \mathbf{b}$ .

We claim  $S = \mathbf{x}_p + ker(T)$  and will prove this statement by showing each set is a subset of the other:

To show  $S \subseteq \mathbf{x}_p + \ker(T)$ :

Let  $\mathbf{x} \in S$ . Then  $T(\mathbf{x}) = \mathbf{b}$ . Thus

$$T(\mathbf{x} - \mathbf{x}_p) = T(\mathbf{x}) - T(\mathbf{x}_p) = \mathbf{b} - \mathbf{b} = \mathbf{0}$$

so  $\mathbf{x} - \mathbf{x}_p \in \ker(T)$ . Thus  $\mathbf{x} = \mathbf{x}_p + (\mathbf{x} - \mathbf{x}_p) \in \mathbf{x}_p + \ker(T)$ .

To show  $\mathbf{x}_p + \ker(T) \subseteq S$ :

Let  $\mathbf{x} \in \mathbf{x}_p + \ker(T)$ . Thus  $\mathbf{x} = \mathbf{x}_p + \mathbf{k}$  where  $T(\mathbf{k}) = \mathbf{0}$ .

Therefore

$$T(\mathbf{x}) = T(\mathbf{x}_p + \mathbf{k}) = T(\mathbf{x}_p) + T(\mathbf{k}) = \mathbf{b} + \mathbf{0} = \mathbf{b}$$

so  $\mathbf{x} \in S$ .

At this point, we've shown that if  $\mathbf{b} \in \text{im}(T)$ , the solution set  $S$  of  $T(\mathbf{x}) = \mathbf{b}$  is the affine subspace  $\mathbf{x}_p + \ker(T)$ .

If  $\ker(T) = \{\mathbf{0}\}$ , then this affine subspace is a point, so there is one solution.

Otherwise, this affine subspace has dimension  $\geq 1$ , so it is infinite, so there are infinitely many solutions.  $\square$

**Theorem 5.16 (Image-Kernel Theorem)** Let  $V_1$  and  $V_2$  be vector spaces where  $V_1$  is finite-dimensional. If  $T : V_1 \rightarrow V_2$  is any linear transformation, then

$$\dim(\text{im}(T)) + \dim(\ker(T)) = \dim(V_1).$$

**PROOF** Let  $T : V_1 \rightarrow V_2$  be linear and suppose  $\dim(\ker(T)) = k$  and  $\dim(V_1) = n$ .

Since  $\ker(T)$  is a subspace of  $V_1$ ,  $k \leq n$ .

*Case 1:  $k = n$ .* In this case, since  $\dim(\ker(T)) = \dim(V_1)$  and  $\ker(T) \subseteq V_1$ , we can conclude that  $\ker(T) = V_1$ .

This means that  $T(\mathbf{x}) = \mathbf{0}$  for all  $\mathbf{x} \in V_1$ .

Thus  $\text{im}(T) = \{\mathbf{0}\}$  and  $\dim(\text{im}(T)) = 0$ , so

$$\dim(\text{im}(T)) + \dim(\ker(T)) = 0 + n = n = \dim(V_1), \text{ as desired.}$$

*Case 2:  $k < n$ .* In this case,  $\dim((\ker(T))^\perp) = n - k > 0$ .

Write down a basis of  $(\ker(T))^\perp$  and call this basis  $\{\mathbf{v}_1, \dots, \mathbf{v}_{n-k}\}$ .

Let  $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$  be a basis of  $\ker(T)$ , so that  $\{\mathbf{v}_1, \dots, \mathbf{v}_{n-k}, \mathbf{x}_1, \dots, \mathbf{x}_k\}$  forms a basis of  $V_1$ .

We claim that  $\{T(\mathbf{v}_1), \dots, T(\mathbf{v}_{n-k})\}$  is a basis of  $\text{im}(T)$ .

- To show  $\{T(\mathbf{v}_1), \dots, T(\mathbf{v}_{n-k})\}$  spans  $im(T)$ :

Let  $\mathbf{w} \in im(T)$ . Then there exists  $\mathbf{v} \in V_1$  such that  $T(\mathbf{v}) = \mathbf{w}$ .

$$\begin{aligned} \mathbf{v} &= c_1\mathbf{v}_1 + \dots + c_{n-k}\mathbf{v}_{n-k} + d_1\mathbf{x}_1 + \dots + d_k\mathbf{x}_k \\ \Rightarrow \mathbf{w} = T(\mathbf{v}) &= c_1T(\mathbf{v}_1) + \dots + c_{n-k}T(\mathbf{v}_{n-k}) + d_1T(\mathbf{x}_1) + \dots + d_kT(\mathbf{x}_k) \\ &= c_1T(\mathbf{v}_1) + \dots + c_{n-k}T(\mathbf{v}_{n-k}) + \mathbf{0} + \dots + \mathbf{0} \\ &\quad \text{(since } \mathbf{x}_j \in \ker(T)\text{)} \\ &= c_1T(\mathbf{v}_1) + \dots + c_{n-k}T(\mathbf{v}_{n-k}). \end{aligned}$$

- To show  $\{T(\mathbf{v}_1), \dots, T(\mathbf{v}_{n-k})\}$  is linearly independent:

Suppose  $c_1T(\mathbf{v}_1) + \dots + c_{n-k}T(\mathbf{v}_{n-k}) = \mathbf{0}$ . Then

$$T(c_1\mathbf{v}_1 + \dots + c_{n-k}\mathbf{v}_{n-k}) = \mathbf{0}$$

so

$$c_1\mathbf{v}_1 + \dots + c_{n-k}\mathbf{v}_{n-k} \in \ker(T) \cap (\ker(T))^\perp.$$

Therefore  $c_1\mathbf{v}_1 + \dots + c_{n-k}\mathbf{v}_{n-k} = \mathbf{0}$ .

Since the  $\mathbf{v}_j$  are lin. ind., all the  $c_j$  are 0.

Therefore  $\{T(\mathbf{v}_1), \dots, T(\mathbf{v}_{n-k})\}$  is lin. ind.

Since  $\{T(\mathbf{v}_1), \dots, T(\mathbf{v}_{n-k})\}$  spans  $im(T)$  and is linearly independent, this list forms a basis for  $im(T)$  so  $\dim(im(T)) = n - k$ , the number of vectors in this basis.

To conclude, we have

$$\dim(im(T)) + \dim(\ker(T)) = (n - k) + k = n = \dim(V_1)$$

as desired.  $\square$

**Corollary 5.17 (Rank-Nullity Theorem, version 1)** Suppose  $T : V_1 \rightarrow V_2$  is a linear transformation with  $\dim V_1 = n < \infty$  and  $\dim V_2 = m < \infty$ . Suppose also that  $T$  has rank  $r(T)$ . Then

$$\boxed{\dim(im(T)) = r(T)} \quad \text{and} \quad \boxed{\dim(\ker(T)) = n - r(T)}.$$

## 5.6 Injectivity, surjectivity, bijectivity and inverses

**Definition 5.18** Let  $X$  and  $Y$  be sets and let  $f : X \rightarrow Y$  be a function.

1.  $f$  is called **injective** (or 1 – 1 or **one-to-one**) if whenever  $f(x) = f(y)$ , it must be the case that  $x = y$ .
2.  $f$  is called **surjective** (or **onto**) if for every  $y \in Y$ , there is an  $x \in X$  such that  $f(x) = y$ .
3.  $f$  is called **bijective** if it both injective and surjective.
4.  $f$  is called **invertible** if there is a function  $f^{-1} : Y \rightarrow X$  such that  $f^{-1}(f(x)) = x$  for all  $x \in X$  and  $f(f^{-1}(y)) = y$  for all  $y \in Y$ . In this setting the function  $f^{-1}$  is called an **inverse** of  $f$ .

This vocabulary can be restated in the language of solving equations:

1. To say  $f$  is **injective** means that **for every  $y \in Y$ , the equation  $f(x) = y$  has at most one solution  $x$ .**
2. To say  $f$  is **surjective** means that **for every  $y \in Y$ , the equation  $f(x) = y$  has at least one solution  $x$ .**
3. Thus  $f$  is **bijective** if and only if **for every  $y \in Y$ , the equation  $f(x) = y$  has exactly one solution  $x$ .**

**Theorem 5.19** A function is invertible if and only if it is bijective.

PROOF ( $\Leftarrow$ ) Suppose  $f$  is bijective. Then for every  $y \in Y$ , the equation  $f(x) = y$  has exactly one solution  $x$ ; define  $f^{-1} : Y \rightarrow X$  by setting  $f^{-1}(y)$  to be the solution  $x$  of  $f(x) = y$ . It is easy to see that  $f^{-1}(f(x)) = x$  and  $f(f^{-1}(y)) = y$ , so  $f$  is invertible with inverse  $f^{-1}$ .

( $\Rightarrow$ ) Now suppose  $f$  is invertible. Then the equation  $f(x) = y$  has exactly one solution, namely  $x = f^{-1}(y)$ , so  $f$  is bijective.  $\square$

Putting all this together, we have the following results:

**Theorem 5.20 (Equivalent properties to injectivity)** Let  $T : V_1 \rightarrow V_2$  be a linear transformation. Then the following are equivalent (if any one of these is true, they are all true; if any one is false, then they are all false):

1.  $T$  is injective.
2. The equation  $T(\mathbf{x}) = \mathbf{b}$  always has at most one solution, no matter what  $\mathbf{b} \in V_2$  is.
3.  $\ker(T) = \{\mathbf{0}\}$ .

Moreover, it is possible for  $T$  to be injective only if  $\dim V_1 \leq \dim V_2$ .

PROOF (1)  $\Leftrightarrow$  (2) based on the remark on the previous page.

(2)  $\Leftrightarrow$  (3) by Theorem 5.15 (which describes the solutions of  $T(\mathbf{x}) = \mathbf{b}$ ).

To prove the last statement, by the Image-Kernel Theorem,

$$\begin{aligned} \dim(\ker(T)) + \dim(\operatorname{im}(T)) &= \dim(V_1) \\ 0 + \dim(\operatorname{im}(T)) &= \dim(V_1) \quad (\text{since } T \text{ is injective}) \end{aligned}$$

Therefore  $\dim(\operatorname{im}(T)) = \dim(V_1)$ . Since  $\operatorname{im}(T)$  is a subspace of  $V_2$ , we have

$$\dim(V_2) \geq \dim(\operatorname{im}(T)) = \dim(V_1). \quad \square$$

**Theorem 5.21 (Equivalent properties to surjectivity)** Let  $T : V_1 \rightarrow V_2$  be a linear transformation. Then the following are equivalent:

1.  $T$  is surjective.
2. The equation  $T(\mathbf{x}) = \mathbf{b}$  always has at least one solution, no matter what  $\mathbf{b} \in V_2$  is.
3.  $\operatorname{im}(T) = V_2$ .

Moreover, it is possible for  $T$  to be surjective only if  $\dim V_1 \geq \dim V_2$ .

PROOF (1)  $\Leftrightarrow$  (2) based on the remark on the previous page.

(2)  $\Leftrightarrow$  (3) by Theorem 5.15.

To prove the last statement, by the Image-Kernel Theorem,

$$\begin{aligned} \dim(\ker(T)) + \dim(\operatorname{im}(T)) &= \dim(V_1) \\ \dim(\ker(T)) + \dim(V_2) &= \dim(V_1) \quad (\text{since } T \text{ is surjective}) \end{aligned}$$

Therefore  $\dim V_2 \leq \dim V_1$ .  $\square$

If  $T$  is bijective (both injective and surjective), we can say quite a bit more:

**Theorem 5.22 (Equivalent properties to bijectivity)** *Let  $T : V_1 \rightarrow V_2$  be a linear transformation. Then the following are equivalent:*

1.  $T$  is bijective.
2.  $T$  is invertible.
3. The equation  $T(\mathbf{x}) = \mathbf{b}$  always has exactly one solution, no matter what  $\mathbf{b} \in V_2$  is.
4.  $\text{im}(T) = V_2$  and  $\ker(T) = \{\mathbf{0}\}$ .

*Moreover, this situation is only possible if  $\dim V_1 = \dim V_2$ . In this situation, the inverse  $T^{-1} : V_2 \rightarrow V_1$  is always a linear transformation, and the one and only solution of equation  $T(\mathbf{x}) = \mathbf{b}$  is  $\mathbf{x} = T^{-1}(\mathbf{b})$ .*

PROOF (1)  $\Leftrightarrow$  (2) from Theorem 5.19; (2)  $\Leftrightarrow$  (3) by the remark two pages ago; and (3)  $\Leftrightarrow$  (4) by Theorem 5.15; so statements (1)-(4) are equivalent.

Now, suppose  $T$  is bijective.

Since  $T$  is injective,  $\dim V_1 \leq \dim V_2$  by Theorem 5.20.

Since  $T$  is surjective,  $\dim V_1 \geq \dim V_2$  by Theorem 5.21.

Therefore  $\dim V_1 = \dim V_2$ .

Next, we show that if  $T$  is invertible, then  $T^{-1}$  is linear.

We show this by brute-force:

- Let  $\mathbf{w}_1, \mathbf{w}_2 \in V_2$  be such that  $T^{-1}(\mathbf{w}_1) = \mathbf{v}_1$  and  $T^{-1}(\mathbf{w}_2) = \mathbf{v}_2$ .

Then  $T(\mathbf{v}_1) = \mathbf{w}_1$  and  $T(\mathbf{v}_2) = \mathbf{w}_2$ .

Since  $T$  is linear,  $T(\mathbf{v}_1 + \mathbf{v}_2) = \mathbf{w}_1 + \mathbf{w}_2$  so

$$T^{-1}(\mathbf{w}_1 + \mathbf{w}_2) = \mathbf{v}_1 + \mathbf{v}_2 = T^{-1}(\mathbf{w}_1) + T^{-1}(\mathbf{w}_2).$$

Therefore  $T^{-1}$  preserves addition.

- Let  $\mathbf{w} \in V_2$  be such that  $T^{-1}(\mathbf{w}) = \mathbf{v}$  and let  $r \in \mathbb{R}$ .

Since  $T$  is linear,  $T(r\mathbf{v}) = r\mathbf{w}$  so  $T^{-1}(r\mathbf{w}) = r\mathbf{v} = rT^{-1}(\mathbf{w})$ .

Therefore  $T^{-1}$  preserves scalar multiplication.

By the brute-force method,  $T^{-1}$  is linear.

Last, suppose  $T$  is bijective and consider the equation  $T(\mathbf{x}) = \mathbf{b}$ .

Apply the transformation  $T^{-1}$  to both sides to get

$$\begin{aligned} T^{-1}(T(\mathbf{x})) &= T^{-1}(\mathbf{b}) \\ \mathbf{x} &= T^{-1}(\mathbf{b}). \quad \square \end{aligned}$$

EXAMPLE 11

For each of these linear transformations  $T : V_1 \rightarrow V_2$  (you don't need to prove they are linear), your directions are:

1. Find the dimension of, and a basis for,  $\ker(T)$ .
2. Find the dimension of, and a basis for,  $\text{im}(T)$ .
3. Sketch the key picture associated to the transformation  $T$ .
4. Determine whether or not  $T$  is injective, surjective, bijective or neither.
5. Determine the possible number of solutions to  $T(\mathbf{x}) = \mathbf{b}$  for various  $\mathbf{b} \in V_2$ .

a) Let  $T : \mathbb{R}^3 \rightarrow \mathbb{R}^2$  be the transformation defined by

$$T(x_1, x_2, x_3) = (x_1, x_3).$$

## 5.6. Injectivity, surjectivity, bijectivity and inverses

---

b) Let  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be the transformation defined by  $T(\mathbf{x}) = A\mathbf{x}$  where

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



## 5.6. Injectivity, surjectivity, bijectivity and inverses

---

- c) Let  $V = \mathcal{P}^3$  be the space of polynomials with degree  $\leq 3$  and let  $T : \mathcal{P}^3 \rightarrow \mathcal{P}^3$  be defined by  $T(f) = f'$ .

## 5.7 Fundamental Theorem of Linear Algebra

## EXAMPLE 11

Given a  $m \times n$  matrix  $A$ , the transformation  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  defined by  $T(\mathbf{x}) = A\mathbf{x}$  is an important example of a linear transformation. In this setting the kernel and image of  $T$  have synonyms:

**Definition 5.23** Let  $A \in M_{mn}(\mathbb{R})$  (i.e.  $A$  is an  $m \times n$  matrix of real numbers).

1. The **null space** of  $A$ , denoted  $N(A)$ , is the set of vectors  $\mathbf{x} \in \mathbb{R}^n$  such that  $A\mathbf{x} = \mathbf{0}$ . (In other words,  $N(A) = \ker(T)$  where  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is defined by  $T(\mathbf{x}) = A\mathbf{x}$ .)
2. The **column space** of  $A$ , denoted  $C(A)$ , is the span of the columns of  $A$ . (In other words,  $C(A) = \text{im}(T)$  where  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is defined by  $T(\mathbf{x}) = A\mathbf{x}$ .)
3. The **row space** of  $A$ , denoted  $R(A)$ , is the span of the rows of  $A$ .
4. The **left null space** of  $A$  is the null space of  $A^T$ , i.e. is the set of vectors  $\mathbf{y} \in \mathbb{R}^m$  such that  $A^T\mathbf{y} = \mathbf{0}$ . (In other words,  $N(A^T) = \ker(T^*)$  where  $T^* : \mathbb{R}^m \rightarrow \mathbb{R}^n$  is defined by  $T^*(\mathbf{y}) = A^T\mathbf{y}$ .)

Collectively, these four spaces are called the **four fundamental subspaces** associated to the matrix  $A$ .

## EXAMPLE 12

Describe the four fundamental subspaces associated to  $A = \begin{pmatrix} 2 & 4 & -3 \\ 1 & 2 & 1 \end{pmatrix}$ .

OBSERVE

Given an  $m \times n$  matrix  $A$ :

- $R(A) = C(A^T)$
- $C(A) = R(A^T)$
- If  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is defined by  $T(\mathbf{x}) = A\mathbf{x}$ , then  $N(A) = \ker(T)$ .
- If  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is defined by  $T(\mathbf{x}) = A\mathbf{x}$ , then  $C(A) = \text{im}(T)$ .
- $N(A)$  and  $R(A)$  are subspaces of  $\mathbb{R}^n$ .
- $N(A^T)$  and  $C(A)$  are subspaces of  $\mathbb{R}^m$ .

**Why do we know these sets are subspaces?**  $R(A)$  and  $C(A)$  are defined as **spans** (hence automatically subspaces), and  $N(A)$  and  $N(A^T)$  are **kernels** of linear transformations (hence automatically subspaces).

It turns out that these subspaces are two pairs of orthogonal complements:

**Theorem 5.24 (Fundamental Theorem of Linear Algebra)** *Let  $A \in M_{mn}(\mathbb{R})$  (i.e.  $A$  is an  $m \times n$  matrix of real numbers). Then*

1.  $[R(A)]^\perp = N(A)$ ; and
2.  $[C(A)]^\perp = N(A^T)$ .

PROOF To prove statement (1), notice

$$\begin{aligned} \mathbf{x} \in N(A) &\iff A\mathbf{x} = \mathbf{0} \\ &\iff \mathbf{x} \perp (\text{any row of } A) \\ &\iff \mathbf{x} \in [R(A)]^\perp. \end{aligned}$$

To prove statement (2), apply the same argument to matrix  $A^T$ :

$$[C(A)]^\perp = [R(A^T)]^\perp = N(A^T). \quad \square$$

**Corollary 5.25 (Rank Theorem)** Let  $A \in M_{mn}(\mathbb{R})$ . Then the following numbers are all equal to the same number, called the **rank** of  $A$  and denoted  $r$  or  $r(A)$ :

1.  $\dim(C(A))$
2.  $\dim(R(A))$
3. the number of linearly independent columns of  $A$
4. the number of linearly independent rows of  $A$
5. the rank of  $T$  where  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is defined by  $T(\mathbf{x}) = A\mathbf{x}$ .

PROOF It is clear that (1) and (3) are equal, and that (2) and (4) are equal.

Since  $C(A) = \text{im}(T)$ , (1) and (5) are equal.

Last, we will show (1) and (2) are equal.

Define  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  by  $T(\mathbf{x}) = A\mathbf{x}$ . By the Image-Kernel Theorem,

$$\begin{aligned}\dim(\text{im}(T)) + \dim(\ker(T)) &= n \\ \dim(C(A)) + \dim(N(A)) &= n \\ \dim(C(A)) &= n - \dim(N(A)).\end{aligned}$$

But at the same time, by the FTLA  $[N(A)]^\perp = R(A)$  so

$$\dim(R(A)) = n - \dim(N(A))$$

Therefore  $\dim(R(A)) = \dim(C(A))$ , proving (1) and (2) are equal.  $\square$

**Corollary 5.26 (Dimensions of fundamental subspaces)** Let  $A \in M_{mn}(\mathbb{R})$  have rank  $r = r(A)$ . Then:

1.  $C(A)$  is a  $r$ -dimensional subspace of  $\mathbb{R}^m$ .
2.  $R(A)$  is a  $r$ -dimensional subspace of  $\mathbb{R}^n$ .
3.  $N(A)$  is an  $(n - r)$ -dimensional subspace of  $\mathbb{R}^n$ .
4.  $N(A^T)$  is an  $(m - r)$ -dimensional subspace of  $\mathbb{R}^m$ .

PROOF (1) and (2) follows from the Rank Theorem; (3) and (4) follow from the Rank-Nullty Theorem.  $\square$

EXAMPLE 13

---

Describe the four fundamental subspaces associated to  $A = \begin{pmatrix} 2 & -1 & 3 & 1 \\ 6 & -3 & 9 & 3 \end{pmatrix}$ .

EXAMPLE 14

---

Describe the four fundamental subspaces associated to  $A = \begin{pmatrix} 1 & 3 & -1 & 0 \\ 2 & -5 & 4 & 0 \\ 0 & 1 & 4 & -2 \end{pmatrix}$ .

## 5.8 More on invertibility

RECALL

$$T \text{ is invertible} \iff T \text{ is bijective} \iff \begin{cases} \ker(T) = \{\mathbf{0}\} \text{ and} \\ \text{im}(T) = V_2 \end{cases}$$

Later, we'll be able to say more about when these conditions hold. As a preview, suppose  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  has standard matrix  $A$ . Then, on the one hand,

$$T \text{ is invertible} \iff \begin{cases} N(A) = \{\mathbf{0}\} \iff r(A) = n \\ C(A) = \mathbb{R}^m \iff r(A) = m \end{cases}$$

and this can only hold if  $r(A) = m = n$  (i.e.  $A$  is a square matrix). But at the same time, by definition,

$$T \text{ is invertible} \iff T^{-1} : \mathbb{R}^m \rightarrow \mathbb{R}^n \text{ exists}$$

and  $T^{-1} \circ T(\mathbf{x}) = \mathbf{x}$  for all  $\mathbf{x}$ ;  $T \circ T^{-1}(\mathbf{x}) = \mathbf{x}$  for all  $\mathbf{x}$ .

Let  $B$  be the standard matrix of  $T^{-1}$ . Then, since  $T \circ T^{-1}$  and  $T^{-1} \circ T$  are both the identity transformation, the standard matrix of both  $T \circ T^{-1}$  and  $T^{-1} \circ T$  is  $I$ . Therefore  $B$  is some matrix such that

Such a matrix  $B$  has another name. It is called the *inverse* of  $A$  and is denoted by  $A^{-1}$  (pronounced "A inverse").

**Definition 5.27** A matrix  $A \in M_{mn}(\mathbb{R})$  is called **invertible** if there is another matrix  $A^{-1} \in M_{nm}(\mathbb{R})$  and called an **inverse (matrix)** of  $A$ , such that

$$AA^{-1} = I_m \text{ and } A^{-1}A = I_n.$$

**Theorem 5.28** A linear transformation  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is invertible if and only if its standard matrix is invertible. (This can only happen if  $n = m$ , based on previous results.)

As a consequence, if a matrix is invertible, then it must be square and have full rank (i.e.  $r(A) = m = n$ ). Conversely, any square matrix with full rank must be invertible.

PROOF ( $\Rightarrow$ ) was proven above.

( $\Leftarrow$ ) If  $T$  has standard matrix  $A$ , and  $A$  is invertible with inverse  $A^{-1}$ , define  $T^{-1} : \mathbb{R}^m \rightarrow \mathbb{R}^n$  by  $T^{-1}(\mathbf{x}) = A^{-1}\mathbf{x}$ .

It's easy to check that  $T$  and  $T^{-1}$  are inverses:

**Theorem 5.29 (Properties of inverses)** Let  $A, B \in M_n(\mathbb{R})$  be invertible. Then

1.  $A$  has only one inverse.
2.  $(A^{-1})^{-1} = A$ .
3.  $(AB)^{-1} = B^{-1}A^{-1}$ .
4.  $A^T$  is invertible and  $(A^T)^{-1} = (A^{-1})^T$ .

PROOF To prove (1), suppose  $B$  and  $C$  are both inverses of  $A$ . Then

$$AB = I \text{ and } AC = I, \text{ so}$$

$$\begin{aligned} AB &= AC && \text{(from above)} \\ BAB &= BAC && \text{(by multiplying both sides on the left by } B\text{)} \\ IB &= IC && \text{(since } B \text{ is an inverse of } A\text{)} \\ B &= C. \end{aligned}$$

So  $A$  cannot have two different inverses.

(2) follows from the definition of inverse.

To prove (3), notice

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

and

$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}IB = B^{-1}B = I.$$

To prove (4), start with  $AA^{-1} = I$  and take the transpose of both sides to get

$$[AA^{-1}]^T = I^T = I, \text{ i.e. } (A^{-1})^T A^T = I.$$

By definition of inverse,  $(A^{-1})^T = (A^T)^{-1}$ .  $\square$

## Inverses of $1 \times 1$ and $2 \times 2$ matrices

### QUESTION

---

When is a  $1 \times 1$  matrix invertible? What is the inverse of a  $1 \times 1$  matrix?

**Theorem 5.30 (1 × 1 inverses)** Let  $A = \begin{pmatrix} a \end{pmatrix}$ . Then  $A$  is invertible  $\iff a \neq 0$ , in which case

$$A^{-1} = \begin{pmatrix} \frac{1}{a} \end{pmatrix}.$$

QUESTION

When is a  $2 \times 2$  matrix invertible? What is the inverse of a  $2 \times 2$  matrix?

*Answer:* Let  $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$  and suppose  $A$  is invertible. Write  $A^{-1} = \begin{pmatrix} w & x \\ y & z \end{pmatrix}$ .

**Theorem 5.31 (2 × 2 inverses)** Let  $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ . Then  $A$  is invertible  $\iff ad - bc \neq 0$ , in which case

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

EXAMPLE 15

Compute the inverse of  $A = \begin{pmatrix} 3 & 2 \\ 7 & 8 \end{pmatrix}$ , if the inverse exists. If the inverse does not exist, say so.



QUESTION

---

When is a  $3 \times 3$  or larger matrix invertible?

*Theoretical answer:*

an  $n \times n$  matrix  $A$   
is invertible  $\iff A$  has full rank  $\iff N(A) = \{\mathbf{0}\} \iff C(A) = \mathbb{R}^n$ .

*Practical answer:* coming in Chapter 6.

## 5.9 Summary of Chapter 5

1. A transformation  $T : V_1 \rightarrow V_2$  is called a **linear transformation** if  $T$  preserves addition and scalar multiplication. Any such  $T$  must also preserve the zero vector.

Typical examples of linear transformations include matrix multiplication, differentiation, integration, evaluation, projection, geometric transformations like rotations and reflections, and dot product by a fixed vector.

Sums, scalar multiples and compositions of linear transformations are also linear transformations.

2. The **kernel** of a linear transformation is the set of vectors in  $V_1$  which are mapped to  $\mathbf{0}$  under the transformation.  $\ker(T)$  is always a subspace of  $V_1$ .
3. The **image** of a linear transformation is the set of vectors in  $V_2$  which are actual outputs of the transformation.
4. If  $T$  is a linear transformation, then the equation  $T(\mathbf{x}) = \mathbf{b}$ ...
  - ... has **no solution** if  $\mathbf{b} \notin \text{im}(T)$ ;
  - ... has **exactly one solution** if  $\mathbf{b} \in \text{im}(T)$  and  $\ker(T) = \{\mathbf{0}\}$ ;
  - ... has **infinitely many solutions** if  $\mathbf{b} \in \text{im}(T)$  and  $\ker(T) \neq \{\mathbf{0}\}$ .
 When  $\mathbf{b} \in \text{im}(T)$ , the solution set of  $T(\mathbf{x}) = \mathbf{b}$  is the affine subspace  $\mathbf{x}_p + \ker(T)$  where  $\mathbf{x}_p$  is any particular solution of  $T(\mathbf{x}) = \mathbf{b}$ .

5.  $T$  is called **injective** if any of these equivalent conditions hold:

- $\ker(T) = \{\mathbf{0}\}$ ;
- $T$  takes different inputs to different outputs;
- the equation  $T(\mathbf{x}) = \mathbf{b}$  has at most one solution for every  $\mathbf{b} \in V_2$ .

It is possible for  $T$  to be injective only if  $\dim V_1 \leq \dim V_2$ .

6.  $T$  is called **surjective** if any of these equivalent conditions hold:

- $\text{im}(T) = V_2$ ;
- the equation  $T(\mathbf{x}) = \mathbf{b}$  has at least one solution for every  $\mathbf{b} \in V_2$ .

It is possible for  $T$  to be surjective only if  $\dim V_1 \geq \dim V_2$ .

7. A transformation  $T$  is called **bijective** if it is both injective and surjective. The following are equivalent:
  - $T$  is bijective;
  - $T$  is invertible (in which case  $T^{-1}$  is linear);
  - the equation  $T(\mathbf{x}) = \mathbf{b}$  always has exactly one solution which is  $\mathbf{x} = T^{-1}(\mathbf{b})$ .

It is possible for  $T$  to be bijective only if  $\dim V_1 = \dim V_2$ .

8. The dimensions of the kernel and image of a linear transformation always add to the dimension of the domain of the transformation:

$$\dim(\ker(T)) + \dim(\text{im}(T)) = \dim V_1.$$

9. Every  $m \times n$  matrix  $A$  has four subspaces associated to it: the **null space**, **column space**, **row space** and **left nullspace**. There is a number  $r = r(A)$  called the **rank** of the matrix such that

- the column space  $C(A)$  is a  $r$ -dimensional subspace of  $\mathbb{R}^m$ ;
- the row space  $R(A)$  is a  $r$ -dimensional subspace of  $\mathbb{R}^n$ ;
- the null space  $N(A)$  is an  $(n - r)$ -dimensional subspace of  $\mathbb{R}^n$ ;
- the left nullspace  $N(A^T)$  is an  $(m - r)$ -dimensional subspace of  $\mathbb{R}^m$ .
- $[R(A)]^\perp = N(A)$  and  $[C(A)]^\perp = N(A^T)$  (this is the Fundamental Theorem of Linear Algebra)

10. Every linear transformation  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is of the form  $T(\mathbf{x}) = A\mathbf{x}$  where  $A$  is an  $m \times n$  matrix called the **standard matrix** of  $T$ . The columns of  $A$  are, from left to right,  $T(\mathbf{e}_1), T(\mathbf{e}_2), \dots, T(\mathbf{e}_n)$ . In this setting,

- $\text{im}(T)$  is the same thing as  $C(A)$ , the span of the columns of  $A$ .
- $\ker(T)$  is the same thing as  $N(A)$ , the null space of  $A$ .

11. A matrix  $A$  is called **invertible** if there is another matrix  $A^{-1}$  (called the inverse of  $A$ ) such that  $AA^{-1} = I$  and  $A^{-1}A = I$ . The following are equivalent:

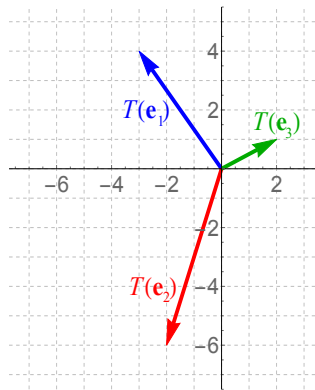
- $A$  is invertible.
- the linear transformation  $T(\mathbf{x}) = A\mathbf{x}$  is invertible.
- $A$  is square and has full rank (i.e.  $r = m = n$ ).

A  $1 \times 1$  matrix  $A = (a)$  is invertible if and only if  $a \neq 0$ , in which case  $A^{-1} = \left(\frac{1}{a}\right)$ .

A  $2 \times 2$  matrix  $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$  is invertible if and only if  $ad - bc \neq 0$ , in which case  $A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ .

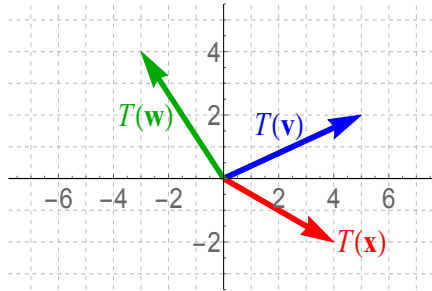
## 5.10 Homework exercises for Chapter 5

1.
  - a) Suppose  $A$  is a  $7 \times 3$  matrix whose entries are real numbers. Let  $T$  be the linear transformation defined by  $T(\mathbf{x}) = A\mathbf{x}$ . What vector space is the domain of  $T$ ? To what vector space do the outputs of  $T$  belong?
  - b) How many rows and columns must a matrix  $A$  have in order to define a linear transformation from  $\mathbb{R}^4$  into  $\mathbb{R}^5$  by the formula  $T(\mathbf{x}) = A\mathbf{x}$ ?
2. Suppose  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^3$  is a linear transformation such that  $T(1, 0) = (-1, 3, 7)$  and  $T(0, 1) = (0, -2, -2)$ .
  - a) What is  $T(x, y)$  for an arbitrary vector  $(x, y) \in \mathbb{R}^2$ ?
  - b) Find  $T(2, 3)$  and  $T(-4, 1)$ .
3. Find the standard matrix of each of the following linear transformations.
  - a)  $T : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  where  $T$  is projection onto the vector  $(2, 1, -2)$ ;
  - b)  $T : \mathbb{R}^4 \rightarrow \mathbb{R}^2$  where  $T(x_1, x_2, x_3, x_4) = (x_1 - 5x_3 + x_4, 0)$ .
4. Find the standard matrix of each of the following linear transformations.
  - a)  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  where  $T$  rotates vectors by an angle of  $2\pi/3$  radians clockwise;
  - b)  $T : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  described by  $T(1, 0, 0) = (1, 2, 3)$ ;  $T(0, 2, 0) = (4, 4, 8)$ ;  $T(0, 0, 1) = (0, 0, 1)$ ;
  - c)  $T : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  where  $T$  reflects points in  $\mathbb{R}^3$  through the  $xy$ -plane.
5. Suppose  $T : \mathbb{R}^3 \rightarrow \mathbb{R}^2$  is a linear transformation where  $T(\mathbf{e}_1)$ ,  $T(\mathbf{e}_2)$  and  $T(\mathbf{e}_3)$  are as indicated below:

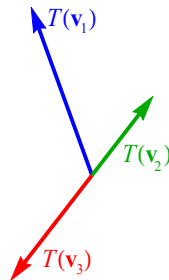


- a) Compute  $T(5, 3, -1)$ .
- b) Compute  $T(-2, 0, 5)$ .

6. Suppose  $\mathbf{v}$ ,  $\mathbf{w}$  and  $\mathbf{x}$  are linearly independent vectors in a vector space  $V_1$ . Suppose also that  $T : V_1 \rightarrow \mathbb{R}^2$  is a linear transformation, and that  $T(\mathbf{v})$ ,  $T(\mathbf{w})$  and  $T(\mathbf{x})$  are as shown below:

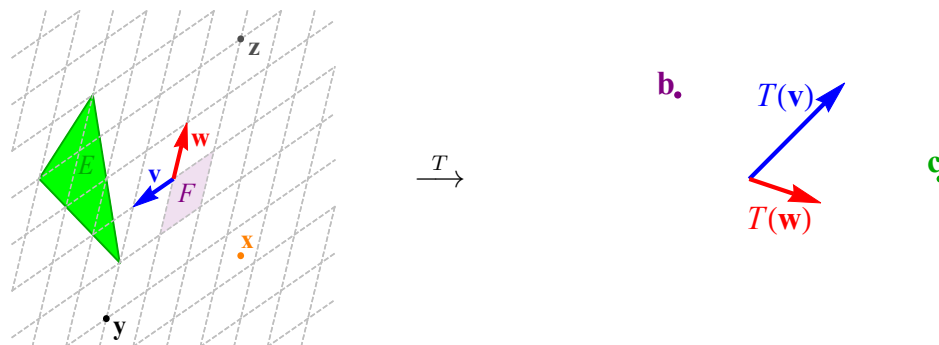


- Compute  $T(-3\mathbf{v})$ .
  - Compute  $T(\frac{1}{2}\mathbf{w} - 3\mathbf{x})$ .
  - Compute  $T(2\mathbf{v} - 5\mathbf{w} + 4\mathbf{x})$ .
  - Find a vector  $\mathbf{a} \in V_1$  so that  $T(\mathbf{a}) = (3, 6)$ .  
(Your formula for  $\mathbf{a}$  should be in terms of  $\mathbf{v}$ ,  $\mathbf{w}$  and/or  $\mathbf{x}$ .)
  - Find a nonzero vector  $\mathbf{b} \in V_1$  so that  $T(\mathbf{b}) = \mathbf{0}$ .  
(Your formula for  $\mathbf{b}$  should be in terms of  $\mathbf{v}$ ,  $\mathbf{w}$  and/or  $\mathbf{x}$ .)
7. Suppose  $T : V_1 \rightarrow V_2$  is a linear transformation, and let  $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4 \in V_1$ . Suppose that  $T(\mathbf{v}_1)$ ,  $T(\mathbf{v}_2)$  and  $T(\mathbf{v}_3)$  are as indicated in the abstract picture below, and that  $T(\mathbf{v}_4) = \mathbf{0}$ . Copy this picture and sketch each of the vectors indicated in parts (a)-(f) on the picture, labelling which is which.



- $T(-\frac{1}{2}\mathbf{v}_1)$
- $T(2\mathbf{v}_3 - 3\mathbf{v}_2)$
- $T(\mathbf{v}_2 + \mathbf{v}_1)$
- $T(3\mathbf{v}_4 + 2\mathbf{v}_2)$
- $T(-2\mathbf{v}_3 - \mathbf{v}_1)$
- $T(\frac{4}{3}\mathbf{v}_3 + \frac{1}{5}\mathbf{v}_4)$

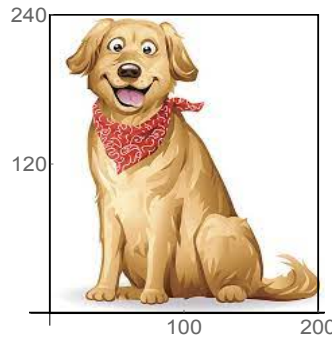
8. Suppose  $T : V_1 \rightarrow V_2$  is a linear transformation. Let  $\mathbf{v}, \mathbf{w} \in V_1$  be as indicated in the picture below at left, and suppose  $T(\mathbf{v})$  and  $T(\mathbf{w})$  are as indicated in the picture at right. Copy these pictures and indicate your answers to each of these questions on the relevant picture.



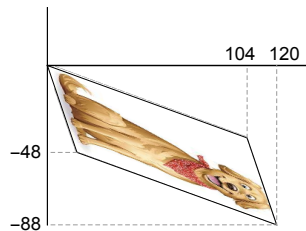
- Sketch the approximate location of  $T(\mathbf{x})$  on the right-hand picture.
  - Sketch the approximate location of  $T(\mathbf{y})$  on the right-hand picture.
  - Sketch the approximate location of  $T(\mathbf{z})$  on the right-hand picture.
  - Sketch the approximate location of some vector  $\mathbf{v}_1$  satisfying  $T(\mathbf{v}_1) = \mathbf{b}$  on the left-hand picture.
  - Sketch the approximate location of some vector  $\mathbf{v}_2$  satisfying  $T(\mathbf{v}_2) = \mathbf{c}$  on the left-hand picture.
  - Let  $F$  be the parallelogram shown in the right-hand picture. Sketch the set  $T(F) = \{T(\mathbf{x}) : \mathbf{x} \in F\}$  on the right-hand picture.
  - Let  $E$  be the triangular region shown in the left-hand picture. Sketch the set  $T(E) = \{T(\mathbf{x}) : \mathbf{x} \in E\}$  on the right-hand picture.
9. A transformation  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is called an **orthogonal transformation** if  $T$  “preserves dot product”, i.e.  $T(\mathbf{x}) \cdot T(\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$  for every  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ .
- Prove that if  $A \in M_n(\mathbb{R})$  is such that  $A^T A = I$ , then the linear transformation defined by  $T(\mathbf{x}) = A\mathbf{x}$  is orthogonal. *Hint:* The dual relations from Chapter 4 may be helpful.
  - Prove that if  $A$  is the standard matrix of orthogonal transformation  $T$ , then  $A^T A = I$ .  
*Note:* This means that one can define a square matrix  $A$  to be **orthogonal** if  $A^T A = I$ .
  - Prove that a rotation in  $\mathbb{R}^2$  (by any angle  $\theta$ ) is an orthogonal transformation.
10. a) Prove that if  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is orthogonal, and if  $\mathbf{x} \perp \mathbf{y}$ , then  $T(\mathbf{x}) \perp T(\mathbf{y})$ .

5.10. Homework exercises for Chapter 5

- b) Prove that “orthogonal transformation preserve norms”, i.e. if  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is orthogonal, then  $\|T(\mathbf{x})\| = \|\mathbf{x}\|$  for all  $\mathbf{x} \in \mathbb{R}^n$ .
- c) Prove that “orthogonal transformation preserve angles”, i.e. if  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is orthogonal, then the angle between  $\mathbf{x}$  and  $\mathbf{y}$  is the same as the angle between  $T(\mathbf{x})$  and  $T(\mathbf{y})$ .
11. Suppose  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  is a linear transformation such that  $T(-1, 2) = (2, 1)$  and  $T(1, 3) = (-2, 4)$ .
- a) Give a formula for  $T(x, y)$ , for any  $(x, y) \in \mathbb{R}^2$ .
- b) Compute  $T(2, 5)$ .
12. Consider the picture below, which occupies the region  $[0, 200] \times [0, 240]$  in  $\mathbb{R}^2$ :



- a) Suppose you stretch the picture by a factor of 3 vertically, and then rotate the picture  $145^\circ$  counterclockwise about the origin (the lower-left corner). At what position  $(x, y)$  will the center of the dog’s nose end up at, if it is currently at  $(78, 196)$ ? (I’m looking for a decimal approximation here.)
- b) Suppose you took the original picture and rotated it  $110^\circ$  clockwise about the center of the picture. Now, at what position will the center of the dog’s nose end up at? (I’m looking for a decimal approximation here.)
- c) Suppose you needed to distort this picture so that it fit in the parallelogram shown below:



At what position will the center of the dog’s nose end up in this picture?

13. Suppose  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  is the linear transformation that stretches vectors by a factor of 2, then rotates the plane  $\pi/3$  radians counterclockwise about the origin. Compute the standard matrix of  $T$ .
14. Let  $T : \mathbb{R}^3 \rightarrow \mathbb{R}^2$  be a linear transformation such that  $T(1, 0, 0) = (1, -1)$ ,  $T(0, 1, 0) = (2, 0)$  and  $T(0, 0, 1) = (0, 1)$ . Let  $S : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be projection onto the vector  $(3, 4)$ , and let  $R : \mathbb{R}^2 \rightarrow \mathbb{R}^3$  be the linear transformation  $R(x, y) = (x + 2y, y, -x - y)$ . Find the standard matrix of the linear transformation  $RST$ .
15. Suppose  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  is a linear transformation that first reflects points through the  $x$ -axis, then reflects points through the  $y$ -axis.
- Find the standard matrix of  $T$ .
  - Show that  $T$  is a rotation, and find the angle  $T$  rotates points by.
16. Let  $V$  be a vector space and define  $T : V \rightarrow V$  by  $T(\mathbf{v}) = 3\mathbf{v}$ . Determine, with proof, whether or not  $T$  is a linear transformation.
17. Let  $T : \mathbb{R}^2 \rightarrow \mathbb{R}$  be defined by  $T(\mathbf{x}) = \|\mathbf{x}\|$ . Determine, with proof, whether or not  $T$  is a linear transformation.
18. Let  $T : C(\mathbb{R}, \mathbb{R}) \rightarrow C(\mathbb{R}, \mathbb{R})$  be the function which sends the function  $f(x)$  to the function  $xf(x)$ . For example,  $T(\sin x) = x \sin x$ ;  $T(x^4) = x^5$ ; etc. Determine, with proof, whether or not this  $T$  is a linear transformation.
19. Let  $T : C([0, 1], \mathbb{R}) \rightarrow \mathbb{R}$  be the function which sends the function  $f$  to its maximum value on  $[0, 1]$ . Determine, with proof, whether or not this  $T$  is a linear transformation.
20. Prove Theorem 5.9 from the notes, which says that if  $T : V_1 \rightarrow V_2$  is a linear transformation and  $c \in \mathbb{R}$ , then the transformation  $cT : V_1 \rightarrow V_2$  defined by  $(cT)(\mathbf{v}) = cT(\mathbf{v})$  is linear.
21. For each of the following linear transformations  $T : V_1 \rightarrow V_2$ :
- Find the dimension of the kernel of  $T$ , and give a basis of  $\ker(T)$ ;
  - Find the dimension of the image of  $T$ , and give a basis of  $\text{im}(T)$ .

You do not need to prove that these transformations  $T$  are linear.

- $V_1 = V_2 = \mathbb{R}^3$ ;  $T(x, y, z) = (x - y, 0, 0)$ .
- $V_1 = \mathbb{R}^3$ ;  $V_2 = \mathbb{R}^2$ ;  $T(x, y, z) = (x + y + 2z, y - 2z)$ .
- $V_1 = V_2 = \mathbb{R}^2$ ;  $T$  reflects points through the line  $y = x$ .
- $V_1 = \mathbb{R}^2$ ;  $V_2 = \mathbb{R}^4$ ;  $T(x, y) = (x + 3y, -2x - 6y, 6x + 18y, 0)$ .



22. Same directions as the previous question:

- a)  $V_1 = V_2 = \mathbb{P}^3$  (the space of polynomials of degree  $\leq 3$ );  $T(f) = f''$ .
- b)  $V_1 = \mathbb{P}^3$ ;  $V_2 = \mathbb{R}$ ;  $T(f) = f(0) + f(1)$ .
- c)  $V_1 = V_2 = M_2(\mathbb{R})$ ;  $T$  is defined by

$$T \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a+b & 0 \\ 0 & c+d \end{pmatrix}.$$

- d)  $V_1 = M_{32}(\mathbb{R})$ ;  $V_2 = \mathbb{R}^2$ ;  $T(A)$  is the first row of  $A$ .

23. Let  $T : V_1 \rightarrow V_2$  be a linear transformation.

Prove that if  $T$  is injective, then for any linearly independent set  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  of vectors in  $V_1$ , the set  $\{T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)\}$  is a linearly independent set of vectors in  $V_2$ .

*Hint:* Start by assuming that there are scalars  $c_1, \dots, c_n$  such that

$$c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2) + \dots + c_nT(\mathbf{v}_n) = \mathbf{0}.$$

Show that the scalars must all be equal to zero. Use the fact that  $T$  is linear, and use the fact that  $T$  is injective (make it clear in your proof where you use these facts).

24. Let  $T : V_1 \rightarrow V_2$  be a linear transformation.

Prove that if  $T$  is surjective, then for any set  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  of vectors which span  $V_1$ , the set  $\{T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)\}$  spans  $V_2$ .

*Hint:* Let  $\mathbf{w} \in V_2$  (the goal is to write  $\mathbf{w}$  as a linear combination of the vectors  $T(\mathbf{v}_j)$ ). First, use the fact that  $T$  is surjective to find a  $\mathbf{v} \in V_1$  such that  $T(\mathbf{v}) = \mathbf{w}$ . Then, use the fact that the vectors  $\mathbf{v}_j$  span  $V_1$  to do something with  $\mathbf{v}$ . Proceed from there.

25. Suppose  $A$  is a  $5 \times 8$  matrix.

- a) Which vector space is  $C(A)$  a subspace of? (I'm thinking of the answer as being  $\mathbb{R}^n$  for some  $n$ .)
- b) Which vector space is  $R(A)$  a subspace of?
- c) Which vector space is  $N(A)$  a subspace of?
- d) Which vector space is  $N(A^T)$  a subspace of?

26. a) Suppose  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is a linear transformation. Is it possible that for every  $\mathbf{b}$ , the equation  $T(\mathbf{x}) = \mathbf{b}$  has no solution? Give a specific example if this is possible; if this is impossible, explain why not.

- b) Suppose  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is a linear transformation. Is it possible that for every  $\mathbf{b}$ , the equation  $T(\mathbf{x}) = \mathbf{b}$  has exactly one solution  $\mathbf{x}$ ? Give a specific example if this is possible; if this is impossible, explain why not.
- c) Suppose  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is a linear transformation. Is it possible that for every  $\mathbf{b}$ , the equation  $T(\mathbf{x}) = \mathbf{b}$  has exactly two solutions? Give a specific example if this is possible; if this is impossible, explain why not.
- d) Suppose  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is a linear transformation. Is it possible that for every  $\mathbf{b}$ , the equation  $T(\mathbf{x}) = \mathbf{b}$  has infinitely many solutions? Give a specific example if this is possible; if this is impossible, explain why not.
27. For each given matrix, find the dimension of, and bases for, each of the four fundamental subspaces of the matrix.

a)  $\begin{pmatrix} 2 & -3 & 6 \\ 3 & 0 & 9 \end{pmatrix}$

c)  $\begin{pmatrix} 2 & 0 & 5 & 3 & -1 \end{pmatrix}$

b)  $\begin{pmatrix} 1 & -2 & 5 \\ -1 & 2 & -5 \\ 2 & -4 & 10 \\ 1 & -2 & 5 \end{pmatrix}$

d)  $\begin{pmatrix} 2 & -1 & 3 & 0 \\ 0 & 0 & 0 & 0 \\ 4 & -2 & 6 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix}$

28. Let  $T(\mathbf{x}) = A\mathbf{x}$ , where  $A$  is a  $9 \times 9$  matrix with 6 linearly independent columns.  
*Note:* Filling out the answer sheet on the next page is a convenient way to answer Exercises 28-33.

- What vector space is the domain of  $T$ ?
  - To what vector space do the outputs of  $T$  belong?
  - $\ker(T)$  is a subspace of what vector space?
  - $\text{im}(T)$  is a subspace of what vector space?
  - Find the dimension of  $\ker(T)$ .
  - Find the dimension of  $\text{im}(T)$ .
  - Is there a vector  $\mathbf{b}$  such that the equation  $T(\mathbf{x}) = \mathbf{b}$  has no solution?
  - Is there a vector  $\mathbf{b}$  such that the equation  $T(\mathbf{x}) = \mathbf{b}$  has exactly one solution?
  - Is there a vector  $\mathbf{b}$  such that the equation  $T(\mathbf{x}) = \mathbf{b}$  has infinitely many solutions?
  - Is  $T$  injective? Explain.
  - Is  $T$  surjective? Explain.
  - Is  $T$  bijective? Explain.
  - Is  $T$  invertible? Explain.
  - Does  $A^{-1}$  exist? Explain.
29. Let  $T(\mathbf{x}) = A\mathbf{x}$ , where  $A$  is a  $9 \times 9$  matrix with 9 linearly independent columns. Answer the same questions (a)-(n) that were asked in Exercise 28.

30. Let  $T(\mathbf{x}) = A\mathbf{x}$ , where  $A$  is a  $4 \times 8$  matrix with 4 linearly independent columns. Answer the same questions (a)-(n) that were asked in Exercise 28.
31. Let  $T(\mathbf{x}) = A\mathbf{x}$ , where  $A$  is a  $4 \times 8$  matrix with 2 linearly independent columns. Answer the same questions (a)-(n) that were asked in Exercise 28.
32. Let  $T(\mathbf{x}) = A\mathbf{x}$ , where  $A$  is a  $10 \times 7$  matrix with 5 linearly independent columns. Answer the same questions (a)-(n) that were asked in Exercise 28.
33. Let  $T(\mathbf{x}) = A\mathbf{x}$ , where  $A$  is a  $10 \times 7$  matrix with 7 linearly independent columns. Answer the same questions (a)-(n) that were asked in Exercise 28.

5.10. Homework exercises for Chapter 5

	# 28	# 29	# 30	# 31	# 32	# 33
size of $A$	$9 \times 9$	$9 \times 9$	$4 \times 8$	$4 \times 8$	$10 \times 7$	$10 \times 7$
# lin. indep. cols	6	9	4	2	5	7
(a) $Dom(T)$						
(b) outputs of $T$ are elements of...						
(c) $\ker(T) \subseteq ?$						
(d) $im(T) \subseteq ?$						
(e) $\dim \ker(T)$						
(f) $\dim im(T)$						
(g) $\exists \mathbf{b}$ such that $T(\mathbf{x}) = \mathbf{b}$ has no solution?						
(h) $\exists \mathbf{b}$ such that $T(\mathbf{x}) = \mathbf{b}$ has one solution?						
(i) $\exists \mathbf{b}$ such that $T(\mathbf{x}) = \mathbf{b}$ has infinitely many solutions?						
(j) Is $T$ injective?						
(k) Is $T$ surjective?						
(l) $T$ bijective?						
(m) Is $T$ invertible?						
(n) Does $A^{-1}$ exist?						

34. In this exercise, we prove the following result, which will be used in Chapter 8:

**Theorem:** Let  $V_1, V_2$  and  $V_3$  be vector spaces; let  $S : V_1 \rightarrow V_2$  and  $T : V_2 \rightarrow V_3$  be linear transformations. If  $\ker(T) \subseteq \text{im}(S)$ ,  $\dim(\ker(T)) < \infty$  and  $\dim(\ker(S)) < \infty$ , we have

$$\dim(\ker(T \circ S)) = \dim \ker(T) + \dim \ker(S).$$

To prove this theorem, let  $m = \dim \ker(T)$  and  $n = \dim \ker(S)$ . Let  $\{\mathbf{x}_1, \dots, \mathbf{x}_m\}$  be a basis of  $\ker(T)$ , and let  $\{\mathbf{y}_1, \dots, \mathbf{y}_n\}$  be a basis of  $\ker(S)$ . Since each  $\mathbf{x}_k \in \ker(T) \subseteq \text{im}(S)$ , we know that there is a vector  $\mathbf{v}_k \in V_1$  such that  $S(\mathbf{v}_k) = \mathbf{x}_k$ . The theorem follows (by counting basis elements) once you prove that  $\{\mathbf{v}_1, \dots, \mathbf{v}_m, \mathbf{y}_1, \dots, \mathbf{y}_n\}$  is a basis of  $\ker(T \circ S)$ , which requires two things you have to do:

- a) show  $\{\mathbf{v}_1, \dots, \mathbf{v}_m, \mathbf{y}_1, \dots, \mathbf{y}_n\}$  span  $\ker(T \circ S)$ ; and
  - b) show  $\{\mathbf{v}_1, \dots, \mathbf{v}_m, \mathbf{y}_1, \dots, \mathbf{y}_n\}$  are linearly independent.
35. Compute the inverse of each of the following matrices (if the matrix is not invertible, say so and explain why):
- a) (4)
  - b) (0)
  - c)  $\begin{pmatrix} 2 & 3 \\ -5 & -1 \end{pmatrix}$
  - d)  $\begin{pmatrix} 1 & 5 \\ 2 & 0 \end{pmatrix}$
36. Compute the inverse of each of the following matrices (if the matrix is not invertible, say so and explain why):
- a)  $\begin{pmatrix} 10 & 5 \\ -6 & -3 \end{pmatrix}$
  - b)  $\begin{pmatrix} 3 & 0 & 0 \\ 0 & -5 & 0 \\ 0 & 0 & -1 \end{pmatrix}$
  - c)  $\begin{pmatrix} 3 & 0 & 0 \\ 0 & -5 & 0 \\ 0 & 0 & 0 \end{pmatrix}$
37. Consider the picture of the dog given in Exercise 12. If it is distorted as described in part (c) of Exercise 12 and a pixel ended up at position  $(70, -55)$ , what was the original position of the pixel?

## Chapter 6

---

# Systems of linear equations

---

### 6.1 Definitions and theory

#### RECALL

Several of the motivating examples we discussed in Chapter 1 reduced to *systems of linear equations*. An example of such a system would be

$$\begin{cases} 3w - 2x + 5y - z = 12 \\ 5w + 3x - y + 7z = -19 \\ 8w - 11x + y + 4z = 6 \end{cases}$$

where the goal would be to solve for  $(w, x, y, z)$ .

In Section 1.1, we saw these examples that led to systems of linear equations:

- the Leontief input-output model (for the economist);
- bonus-malus systems (for the actuarial scientist);
- error correction (for the surveyor)

#### OTHER EXAMPLES OF PROBLEMS LEADING TO SYSTEMS OF LINEAR EQUATIONS

---

##### 1. Determining the orthogonal complement of a subspace:

*Example:* Let  $V = \mathbb{R}^4$ ; let  $W = \text{Span}((2, 4, 0, -1), (1, -1, 2, 3))$ . What is  $W^\perp$ ?

**2. Determining whether a vector lies in the span of some other vectors:**

*Example:* Let  $V = \mathbb{R}^3$  and let  $W = \text{Span}((1, -1, 2), (2, 0, -3))$ . Is the vector  $(-2, -8, 17)$  in  $W$ ?

**3. Word problems:**

*Example 1:* Dan has some 32-cent stamps, some 29-cent stamps, and some 3-cent stamps. The number of 29-cent stamps is 10 less than the number of 32-cent stamps, while the number of 3-cent stamps is 5 less than the number of 29-cent stamps. The total value of the stamps is \$9.45. How many of each stamp does Dan have?

*Solution:* Let  $x =$  the number of 32-cent stamps;  $y =$  the number of 29-cent stamps;  $z =$  the number of 3-cent stamps. Then:

*Example 2:* A cashier has 25 coins, all of which are nickels, dimes and quarters. If the total value of her coins is \$4.90, how many of each type of coin might she have?

*Solution:* Let  $n =$  the number of nickels;  $d =$  the number of dimes;  $q =$  the number of quarters. Then we have:

$$\begin{cases} n + d + q = 25 \\ .05n + .1d + .25q = 4.90 \end{cases}$$

**4. Fitting data to a model:**

*Example:* Find the equation of the plane which best fits the following six points:

$$\begin{array}{ccc} (1, 2, 5) & (2, 3, 4) & (2, 2, 6) \\ (1, 3, 4) & (1, 4, 5) & (3, 3, 3) \end{array}$$

*Solution:* We know from Chapter 4 that every plane in  $\mathbb{R}^3$  has the normal equation

$$ax + by + cz = d.$$

Assuming  $c \neq 0$ , we can solve the normal equation for  $z$  to get

$$z = \alpha + \beta x + \gamma y$$

where  $\alpha, \beta, \gamma$  are constants.

Now, we can plug in the given data points to the equation  $\alpha + \beta x + \gamma y = z$ :



## Definition of a linear system

**Definition 6.1** A linear equation in  $n$  variables is an equation which can be rewritten as

$$a_1x_1 + a_2x_2 + \dots + a_nx_n = b$$

where  $a_1, a_2, \dots, a_n, b \in \mathbb{R}$ .

**Definition 6.2** A linear system of  $m$  equations in  $n$  variables (a.k.a.  $m \times n$  linear system) is a set of  $m$  linear equations in the same  $n$  variables:

$$(*) \begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 \\ \vdots \qquad \qquad \qquad \ddots \qquad \qquad \qquad \vdots = \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m \end{cases}$$

A **solution** of system  $(*)$  is a vector  $\mathbf{x} = (x_1, x_2, \dots, x_n) = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$  which satisfies all the equations in  $(*)$ .

A system is called **consistent** if it has a solution, and is called **inconsistent** otherwise.

The **solution set** of a system is the set of all solutions; this set is a subset of  $\mathbb{R}^n$ .

Two  $m \times n$  linear systems are called **equivalent** if they have the same solution set.

### EXAMPLE

$4x = 8$  is equivalent to  $25x = 50$ , since both equations have solution set  $\{2\}$ .

### EXAMPLE

The systems  $\begin{cases} 2x + 2y = 9 \\ 4x + 4y = 14 \end{cases}$  and  $\begin{cases} x - y = 3 \\ 2x - 2y = -4 \end{cases}$  are equivalent, because both of their solution sets are  $\emptyset$  (neither system has a solution).

## Fundamental questions associated to linear systems:

Given an  $m \times n$  linear system,

1. Is the system consistent? (In other words, is there a solution?)
2. If the system is consistent, how many solutions does it have?
3. If the system is consistent, what is the solution set?

To approach these questions, we need to view a linear system in several different ways:

**Ways to think about a linear system****1. As a system of linear equations:**

This means the way the system (\*) is written in the definition.

**2. As a matrix equation:**

Given (\*) on the previous page,

**3. As a vector equation:**

Given (\*) on the previous page, write it as

$$\begin{aligned} (*) \text{ is consistent} &\iff \mathbf{b} \in \text{Span}(\text{columns of } A) \iff \mathbf{b} \in C(A) \\ (*) \text{ is inconsistent} &\iff \mathbf{b} \notin C(A) \end{aligned}$$

**4. As a functional equation involving a linear transformation:**

Define  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  by  $T(\mathbf{x}) = A\mathbf{x}$ , where  $A$  is as above.

That means (\*) becomes

## EXAMPLE 1

---

Consider the system

$$(*) \quad \begin{cases} 3w & -2x & +y & -4z & = & 8 \\ w & +x & +7y & & = & 2 \\ -w & & +y & -3z & = & -3 \end{cases}$$

Describe this system as a matrix equation, as a vector equation, and as a functional equation.

Summarizing some of what we discussed when talking about linear transformations and fundamental subspaces, we have the following theorem, which gives a complete theoretical description of the solution set of a linear system:

**Theorem 6.3 (Solution set of a linear system)** Let  $(*)$  be an  $m \times n$  linear system which is  $A\mathbf{x} = \mathbf{b}$  as a matrix equation, and which is  $T(\mathbf{x}) = \mathbf{b}$  as a functional equation (meaning  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is the linear transformation  $T(\mathbf{x}) = A\mathbf{x}$ ). Then:

1.  $(*)$  is consistent  $\iff \mathbf{b} \in C(A) \iff \mathbf{b} \in \text{im}(T)$ .  
 $(*)$  is inconsistent  $\iff \mathbf{b} \notin C(A) \iff \mathbf{b} \notin \text{im}(T)$ .
2. If  $(*)$  is consistent, then the solution set of the system is the affine subspace  $\mathbf{x}_p + \ker(T) = \mathbf{x}_p + N(A)$  where  $\mathbf{x}_p$  is any one "particular" solution of the system.
3. If  $(*)$  is consistent, then

$(*)$  has  
 exactly one solution  $\iff N(A) = \ker(T) = \{\mathbf{0}\} \iff T$  is injective  $\iff r = n$ .

and

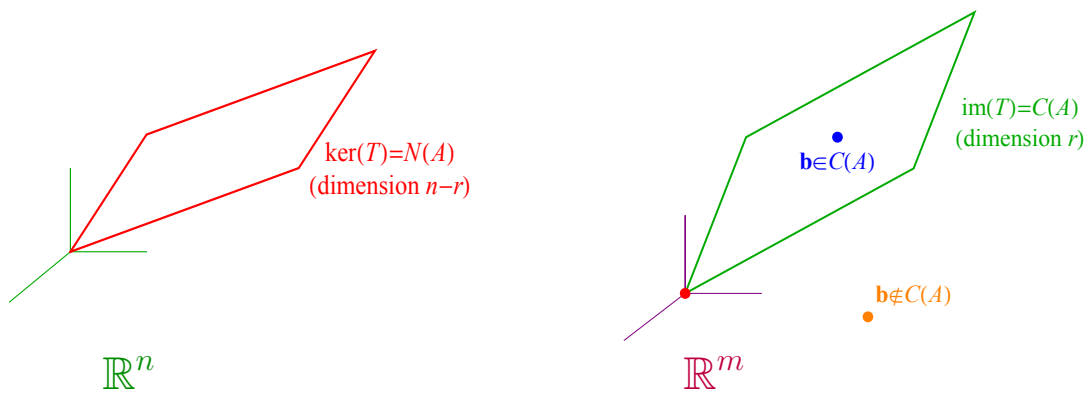
$(*)$  has  
 infinitely many solutions  $\iff N(A) = \ker(T) \neq \{\mathbf{0}\} \iff T$  is not injective  $\iff r \neq n$ .

**The essential content of this theorem is in the following four facts:**

1. A system  $A\mathbf{x} = \mathbf{b}$  has **no solution** if and only if  $\mathbf{b} \notin C(A)$ .
2. A system  $A\mathbf{x} = \mathbf{b}$  has **exactly one solution** if and only if  $\mathbf{b} \in C(A)$  and  $N(A) = \{\mathbf{0}\}$ .
3. A system  $A\mathbf{x} = \mathbf{b}$  has **infinitely many solutions** if and only if  $\mathbf{b} \in C(A)$  and  $N(A) \neq \{\mathbf{0}\}$ .
4. If the system  $A\mathbf{x} = \mathbf{b}$  is consistent, then its **solution set is** the affine subspace  $\mathbf{x}_p + N(A)$ .

**THE KEY PICTURE**

$$T(\mathbf{x}) = A\mathbf{x} \text{ (rank } r\text{)}$$



Let's see how this works in a very simple setting. Suppose we have one equation in one variable, i.e. the entire linear system is just

$$Ax = b$$

where  $A$  and  $b$  are constants.

Compare this with Theorem 1.1, all the way back in Section 1.2.

Now let's try this with two equations in two variables. Suppose we have a system

$$(*) \begin{cases} a_{11}x + a_{12}y = b_1 \\ a_{21}x + a_{22}y = b_2 \end{cases} \quad \text{i.e.} \quad \mathbf{Ax} = \mathbf{b}, \text{ where} \\ A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}; \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}; \mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}.$$

## 6.2 Rank and dimension issues

RECALL FROM THE PREVIOUS SECTION

If  $A \in M_{mn}(\mathbb{R})$  and  $T(\mathbf{x}) = A\mathbf{x}$ , then...

1. the system  $A\mathbf{x} = \mathbf{b}$  has **no solution** if and only if  $\mathbf{b} \notin C(A)$ .
2. the system  $A\mathbf{x} = \mathbf{b}$  has **exactly one solution** if and only if  $\mathbf{b} \in C(A)$  and  $N(A) = \{\mathbf{0}\}$ .  
(This is equivalent to  $\mathbf{b} \in im(T)$  and  $T$  being **injective**.)
3. the system  $A\mathbf{x} = \mathbf{b}$  has **infinitely many solutions** if and only if  $\mathbf{b} \in C(A)$  and  $N(A) \neq \{\mathbf{0}\}$ .  
(This is equivalent to  $\mathbf{b} \in im(T)$  and  $T$  being **not injective**.)
4. if the system  $A\mathbf{x} = \mathbf{b}$  is consistent, then its **solution set is** the affine subspace  $\mathbf{x}_p + N(A)$ .

Adding to this, suppose  $T$  is surjective ( $r = m$ ). Then  $im(T) = C(A) = \mathbb{R}^m$ , so every  $\mathbf{b} \in \mathbb{R}^m$  lies in  $im(T)$ . Then

Now suppose  $T$  is injective ( $r = n$ ). Then  $\ker(T) = N(A) = \{\mathbf{0}\}$ , and then

Even further, suppose  $T$  is bijective ( $r = m = n$ ). Then  $T$  is both surjective and injective, so

To summarize:

**Theorem 6.4 (Equivalent properties to surjectivity)** Let  $A \in M_{mn}(\mathbb{R})$  and suppose  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is defined by  $T(\mathbf{x}) = A\mathbf{x}$ . Then the following are equivalent (and none of them are possible unless  $m \leq n$ ):

1.  $r(A) = m$ ;
2.  $im(T) = \mathbb{R}^m$ ;
3.  $T$  is surjective;
4.  $C(A) = \mathbb{R}^m$ ;
5.  $N(A^T) = \{\mathbf{0}\}$ ;
6. the rows of  $A$  are linearly independent;
7.  $T$  “maps spanning sets to spanning sets”, i.e. given any set  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  which spans  $\mathbb{R}^n$ , the set  $\{T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)\}$  spans  $\mathbb{R}^m$ ;
8.  $A\mathbf{x} = \mathbf{b}$  has at least one solution for every  $\mathbf{b} \in \mathbb{R}^m$ .

PROOF (1)  $\Leftrightarrow$  (4) : By definition,  $r(A) = \dim C(A)$ .

Since  $C(A)$  is a subspace of  $\mathbb{R}^m$ ,  $C(A)$  equals all of  $\mathbb{R}^m$  if and only if  $\dim C(A) = m$ .

(2)  $\Leftrightarrow$  (4) because  $C(A) = im(T)$ .

(2)  $\Leftrightarrow$  (3)  $\Leftrightarrow$  (8) by definition of surjectivity.

(4)  $\Leftrightarrow$  (5) by the FTLA (which says  $N(A^T) = [C(A)]^\perp$ ).

(1)  $\Leftrightarrow$  (6) by the Rank Theorem.

(3)  $\Rightarrow$  (7) was Homework Exercise 23 in Chapter 5.

(7)  $\Rightarrow$  (3): Let  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  be a basis of  $\mathbb{R}^n$ . Let  $\mathbf{w} \in \mathbb{R}^m$ .

By (7),  $\{T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)\}$  span  $\mathbb{R}^m$ , so we can write

$$\begin{aligned}\mathbf{w} &= c_1T(\mathbf{v}_1) + \dots + c_nT(\mathbf{v}_n) \\ &= T(c_1\mathbf{v}_1 + \dots + c_n\mathbf{v}_n).\end{aligned}$$

Therefore  $\mathbf{w} = T(\mathbf{v})$  where  $\mathbf{v} = c_1\mathbf{v}_1 + \dots + c_n\mathbf{v}_n$ , meaning  $T$  is surjective.  $\square$

**Corollary 6.5** Let  $A \in M_{mn}(\mathbb{R})$  where  $m > n$  (i.e.  $A$  has more rows than columns, i.e. the system has *more equations than variables*). Then *there must be some  $\mathbf{b} \in \mathbb{R}^m$  such that the equation  $A\mathbf{x} = \mathbf{b}$  has no solution.*



**Theorem 6.6 (Equivalent properties to injectivity)** Let  $A \in M_{mn}(\mathbb{R})$  and suppose  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is defined by  $T(\mathbf{x}) = A\mathbf{x}$ . Then the following are equivalent (and none of them are possible unless  $m \geq n$ ):

1.  $r(A) = n$ ;
2.  $R(A) = \mathbb{R}^n$ ;
3.  $N(A) = \{\mathbf{0}\}$ ;
4.  $T$  is injective;
5.  $\ker(T) = \{\mathbf{0}\}$ ;
6. the columns of  $T$  are linearly independent;
7.  $T$  “maps linearly independent sets to linearly independent sets”, i.e. given any lin. ind. set  $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$  in  $\mathbb{R}^n$ ,  $\{T(\mathbf{v}_1), \dots, T(\mathbf{v}_k)\}$  is a lin. ind. set in  $\mathbb{R}^m$ ;
8.  $A\mathbf{x} = \mathbf{b}$  has at most one solution for every  $\mathbf{b} \in \mathbb{R}^m$ .

PROOF (1)  $\Leftrightarrow$  (2): By the Rank Theorem,  $r(A) = \dim R(A)$ .

Since  $R(A)$  is a subspace of  $\mathbb{R}^n$ ,  $R(A)$  equals all of  $\mathbb{R}^n$  if and only if  $\dim R(A) = n$ .

(2)  $\Leftrightarrow$  (3) by the FTLA (which says  $N(A) = [R(A)]^\perp$ ).

(3)  $\Leftrightarrow$  (5) because  $N(A) = \ker(T)$ .

(4)  $\Leftrightarrow$  (5)  $\Leftrightarrow$  (8) by definition of injectivity.

(1)  $\Leftrightarrow$  (6) by the definition of rank.

(4)  $\Rightarrow$  (7) was Homework Exercise 22 in Chapter 5.

(7)  $\Rightarrow$  (4): Suppose  $T(\mathbf{x}) = \mathbf{0}$ .

If  $\mathbf{x} \neq \mathbf{0}$ , then  $\{\mathbf{x}\}$  is a linearly independent set of one vector in  $\mathbb{R}^n$ .

By hypothesis,  $\{T(\mathbf{x})\}$  is linearly independent in  $\mathbb{R}^m$ , so  $T(\mathbf{x}) \neq \mathbf{0}$ .

But this is a contradiction! Therefore  $\mathbf{x} = \mathbf{0}$ .

Therefore  $\ker(T) = \{\mathbf{0}\}$ , proving that  $T$  is injective.  $\square$

**Corollary 6.7** 1. Let  $A \in M_{mn}(\mathbb{R})$  where  $m < n$  (i.e.  $A$  has more columns than rows). Then the equation  $A\mathbf{x} = \mathbf{b}$  never has exactly one solution.

2. A linear system with *more variables than equations* never has exactly one solution.

**Theorem 6.8 (Equivalent properties to bijectivity)** Let  $A \in M_n(\mathbb{R})$  be an  $n \times n$  **square** matrix (i.e.  $m = n$ ) and suppose  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is defined by  $T(\mathbf{x}) = A\mathbf{x}$ . Then Properties 1-28 below are all equivalent:

**Properties of the entire matrix  $A$ :**

1.  $A$  is invertible;
2. There is a matrix  $B \in M_n(\mathbb{R})$  such that  $AB = I$ ;
3.  $A$  is row equivalent to the identity matrix  $I$ ; ← DISCUSSED LATER
4.  $A$  has  $n$  pivots; ← DISCUSSED LATER
5.  $\text{rref}(A) = I$ . ← DISCUSSED LATER

**Properties of the rows of  $A$ :**

6.  $\text{rank}(A) = n$ ;
7.  $R(A) = \mathbb{R}^n$ ;
8. the rows of  $A$  form a basis of  $\mathbb{R}^n$ ;
9. the rows of  $A$  are linearly independent;
10. the rows of  $A$  span  $\mathbb{R}^n$ .

**Properties of the columns of  $A$ :**

11.  $C(A) = \mathbb{R}^n$ ;
12.  $N(A) = \{\mathbf{0}\}$ ;
13. the columns of  $A$  form a basis of  $\mathbb{R}^n$ ;
14. the columns of  $A$  are linearly independent;
15. the columns of  $A$  span  $\mathbb{R}^n$ .

**Properties of the system  $A\mathbf{x} = \mathbf{b}$ :**

16.  $A\mathbf{x} = \mathbf{0}$  has only the solution  $\mathbf{x} = \mathbf{0}$ ;
17. there is a **single**  $\mathbf{b} \in \mathbb{R}^n$  such that the equation  $A\mathbf{x} = \mathbf{b}$  has exactly one solution;
18. for **every**  $\mathbf{b} \in \mathbb{R}^n$ , the equation  $A\mathbf{x} = \mathbf{b}$  has exactly one solution (namely  $\mathbf{x} = A^{-1}\mathbf{b}$ ).

**Properties of the transformation  $T$ :**

19.  $T$  is bijective;
20.  $T$  is invertible (i.e. the function  $T^{-1} : \mathbb{R}^n \rightarrow \mathbb{R}^n$  exists);
21.  $T$  is injective;
22.  $\ker(T) = \{\mathbf{0}\}$ ;
23.  $T$  is surjective;
24.  $\text{im}(T) = \mathbb{R}^n$ ;
25.  $T$  maps linearly independent sets to linearly independent sets;
26.  $T$  maps spanning sets to spanning sets;
27.  $T$  maps bases to bases, meaning that for **any** basis  $\mathcal{B}$  of  $\mathbb{R}^n$ ,  $T(\mathcal{B})$  is also a basis of  $\mathbb{R}^n$ ;
28. given any **one** single basis  $\mathcal{B}$  of  $\mathbb{R}^n$ ,  $T(\mathcal{B})$  is a basis of  $\mathbb{R}^n$ .

## 6.3 Gaussian elimination

### GOAL

We know what the solution to a system of linear equations looks like *theoretically*. What we want now is to develop a *practical* method that will find the solution of a system of linear equations. Ideally, this method will

- 1.
- 2.
- 3.

### EXAMPLE 2

Suppose  $N(A) = \text{Span}((2, -1, 3), (1, 0, 5))$  and that  $A \begin{pmatrix} 1 \\ 1 \\ 4 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 3 \end{pmatrix}$ . What are all the solutions  $\mathbf{x}$  of the system  $A\mathbf{x} = \begin{pmatrix} 2 \\ 0 \\ 3 \end{pmatrix}$ ?

### BASIC STRATEGY

Start with some system and replace it with an equivalent system (i.e. one that has the same solution set as the original system) that is easier to solve.

*Philosophy:* We can

- 1.
- 2.
- 3.

without changing the solution set of the system.

*More philosophy:* the “names” of the variables don’t matter, i.e.

$$\begin{cases} 7x - 2y = 5 \\ 4x + 7y = 11 \end{cases} \text{ is the same as } \begin{cases} 7s - 2t = 5 \\ 4s + 7t = 11 \end{cases} .$$

**Definition 6.9** Given a  $m \times n$  system of linear equations, the **augmented matrix** of the system is the  $m \times (n + 1)$  matrix

$$(A|\mathbf{b}) = \left( \begin{array}{ccc|c} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} & b_m \end{array} \right).$$

EXAMPLE 3

Write the augmented matrix of the following system:

$$\begin{cases} 3x_1 + 2x_2 = -7 \\ -x_1 - 3x_2 = 4 \end{cases}$$

EXAMPLE 4

Write the corresponding system of equations for each augmented matrix. Solve the system if it is “easy”:

a)  $(A|\mathbf{b}) = \left( \begin{array}{cccc|c} 4 & -1 & 2 & 6 & 0 \\ 1 & 0 & 2 & -1 & 5 \end{array} \right)$

b)  $(A|\mathbf{b}) = \left( \begin{array}{cccc|c} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 2 & 0 & -1 \\ 0 & 0 & 0 & 1 & 2 \end{array} \right)$

$$\text{c) } (A|\mathbf{b}) = \left( \begin{array}{ccc|c} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 7 \\ 0 & 0 & 1 & -3 \end{array} \right)$$

$$\text{d) } (A|\mathbf{b}) = \left( \begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 1 \end{array} \right)$$

---

#### HOW TO IMPLEMENT THE BASIC STRATEGY

---

Given a system of linear equations,

1. Write the augmented matrix  $(A|\mathbf{b})$  of the system.
2. Do some “operations” on  $(A|\mathbf{b})$  to produce an augmented matrix of an “easy” system, equivalent to the original one.

*What “operations” are allowed here?*

- i.
  - ii.
  - iii.
3. Solve the “easy” system.

#### EXAMPLE 4

---

In each example, you are given a matrix and an indicated row operation (or operations). Perform the indicated operations:

- a) Swap the first and third row, then multiply the third row by 3.

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

- b) Multiply the third row by  $-2$ , then add 4 times the first row to the third row:

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

**Definition 6.10** Two matrices are called **row equivalent** if one can be produced from the other by a sequence of operations of types (i), (ii) and/or (iii) above.

**Theorem 6.11** If two systems of equations have row equivalent augmented matrices, then the systems are equivalent (they have the same solution set).

**Note:** The allowable row operations are all reversible:

- To “undo” the swapping of two rows, swap the rows again.
- To “undo” the multiplying of a row by nonzero constant  $c$ , multiply the same row by  $\frac{1}{c}$ .
- To “undo” the adding of  $c$  times Row  $i$  to Row  $j$ , add  $-c$  times Row  $i$  to Row  $j$ .

Thus the definition of “row equivalent” is symmetric: if  $A$  is row equivalent to  $B$ , then  $B$  is row equivalent to  $A$ .

#### QUESTION

Looking back at Example 4, why were some of these systems easier to study than others?

**Definition 6.12** A matrix is in **(row-)echelon form** (*ref*) if

1. all rows of zeros are at the bottom of the matrix, and
2. defining the **pivot** (a.k.a. **leading entry**) of each row to be the left-most nonzero entry in that row, the pivot of every row is to the right of the pivot of any above row. (This implies, among other things, that all entries directly below any pivot are zero.)

**Definition 6.13** A matrix is in **reduced (row-)echelon form** (*rref*) if

1. it is in row echelon form, and
2. all pivots are equal to 1, and
3. each pivot is the only nonzero entry in its column.

**FACT:** A matrix may have several row echelon forms, but it has one and only one *rref*. So we can talk about **the** *rref* of  $A$  or  $(A|\mathbf{b})$ , rather than **an** *rref* of  $A$  or  $(A|\mathbf{b})$ .

**Definition 6.14** Given a matrix  $A$ , the **pivot columns** of  $A$  are the columns which have a pivot in the rref of  $A$ . The **free columns** of  $A$  are the columns which are not pivot columns.

**Note:** If the matrix is augmented  $(A|\mathbf{b})$ , then the last column (the  $\mathbf{b}$ ) is not a pivot nor a free column. It “doesn’t count”.

Now for the key idea which ties the theory to the practice:

**Theorem 6.15** If  $A$  and  $B$  are row equivalent matrices, then  $R(A) = R(B)$ .

PROOF If  $A$  and  $B$  are row equivalent, then you can convert one to the other by a sequence of row operations. But none of the three types of elementary row operations changes the row space of a matrix.  $\square$

**Consequence:**

$$\begin{aligned} \dim(R(A)) &= \dim(R(\text{ref}(A))) \\ &= \# \text{ of nonzero rows in } \text{ref}(A) \\ &= \# \text{ of pivots of } A \\ &= \text{rank of } A \\ &= r. \end{aligned}$$

This proves:

**Theorem 6.16 (Rank Theorem II)** Let  $A \in M_{mn}(\mathbb{R})$ . Then

$$r(A) = \# \text{ of pivots of } A = \# \text{ of nonzero rows in } \text{ref}(A).$$

## EXAMPLE 5

For each matrix, determine whether or not it is in row echelon form. Determine whether or not it is in reduced row echelon form. Write down all the pivots, the pivot columns and the free columns. Find the rank of  $A$ , and the dimension of the four fundamental subspaces associated to  $A$ .

$$\text{a) } A = \left( \begin{array}{ccccc|c} 1 & 5 & -1 & 6 & 0 & 1 \\ 0 & 1 & 4 & 2 & 1 & -3 \\ 0 & 0 & 0 & 5 & 4 & 5 \\ 0 & 0 & 0 & 0 & -1 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right)$$

pivots:

rank  $r(A) =$

row echelon form?

$\dim C(A) =$

 $rref?$ 

$\dim R(A) =$

pivot columns:

$\dim N(A) =$

free columns:

$\dim N(A^T) =$

$$\text{b) } A = \left( \begin{array}{ccccc} 1 & 2 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 1 \end{array} \right)$$

pivots:

rank  $r(A) =$

row echelon form?

$\dim C(A) =$

 $rref?$ 

$\dim R(A) =$

pivot columns:

$\dim N(A) =$

free columns:

$\dim N(A^T) =$



### Algorithm for putting a matrix in ref or rref form by hand

Suppose you are given a matrix (it can be an augmented matrix  $(A | \mathbf{b})$ , or just a coefficient matrix  $A$ , or any other matrix) and you want to put it into echelon form. Here's a formulaic way called **Gaussian elimination** or **row reduction** which always works (this is not the only method, however).

**Note:** No one does this by hand if they have a computer or graphics calculator available. See Section 6.7.

#### Part 1 (Downward reduction): Transforming a matrix into row-echelon form

1. Place an imaginary cursor at the upper-left entry of the matrix.
2. If the cursor entry and all entries below the cursor entry are zero, move the cursor one column to the right. Repeat this step if necessary.
3. The cursor should now be in a position such that there is a nonzero entry either in the cursor position or directly below the cursor. If necessary, switch the cursor row with a row *beneath* the cursor row to put a nonzero entry in the cursor position.
4. Add multiples of the cursor row to each of the rows beneath the cursor row to create zeros in all entries beneath the cursor.
5. Move the cursor one row down and one column to the right. If the cursor is not on the bottom row, return to step 2. If you reach the last nonzero row, you are done and the matrix is in row-echelon form.

These steps are called *downward reduction* because the **cursor is always moving down** (from the upper-left corner downward and to the right).

#### Part 2 (Upward reduction): Transforming a row-echelon form into a rref

1. If necessary, carry out downward reduction as in Part 1 to place the matrix in row-echelon form.
2. Multiply each pivot row by the reciprocal of its pivot, so that all the pivots become 1.
3. Place the cursor on the right-most pivot position.
4. If necessary, add multiples of the cursor row to every row above the cursor row so that all entries above the cursor become zero.
5. Once all entries above the cursor are zero, move the cursor up one row and to the left until you reach a pivot position.
6. If the cursor is on the first row, you are done. The matrix is in rref form. Otherwise, return to step 4.

These steps are called *upward reduction* because the **cursor is always moving up** (from the bottom pivot row up and to the left).

## EXAMPLE 6

Let  $A = \begin{pmatrix} 0 & 6 & 4 & -12 \\ 3 & 3 & 0 & 9 \\ 2 & 0 & -3 & 10 \end{pmatrix}$ . Find  $rref(A)$ .

## Downward reduction

Matrix	Current Step	Thought Process	Subsequent Row Operation
$\begin{pmatrix} \underline{0} & 6 & 4 & -12 \\ 3 & 3 & 0 & 9 \\ 2 & 0 & -3 & 10 \end{pmatrix}$	1,2	Cursor column contains nonzero entries below cursor. Proceed to step 3.	
$\begin{pmatrix} \underline{0} & 6 & 4 & -12 \\ 3 & 3 & 0 & 9 \\ 2 & 0 & -3 & 10 \end{pmatrix}$	3	I need a nonzero entry at the cursor position.	Swap rows 1 and 2
$\begin{pmatrix} \underline{3} & 3 & 0 & 9 \\ 0 & 6 & 4 & -12 \\ 2 & 0 & -3 & 10 \end{pmatrix}$	4	I need to make all entries at below the cursor zero.	Add $-\left(\frac{2}{3}\right)$ times row 1 to row 3
$\begin{pmatrix} \underline{3} & 3 & 0 & 9 \\ 0 & 6 & 4 & -12 \\ 0 & -2 & -3 & 4 \end{pmatrix}$	5	Move the cursor down and right. Return to step 2.	
$\begin{pmatrix} 3 & 3 & 0 & 9 \\ 0 & \underline{6} & 4 & -12 \\ 0 & -2 & -3 & 4 \end{pmatrix}$	2	Cursor column contains nonzero entries below cursor; cursor entry is nonzero. Proceed to step 4.	
$\begin{pmatrix} 3 & 3 & 0 & 9 \\ 0 & \underline{6} & 4 & -12 \\ 0 & -2 & -3 & 4 \end{pmatrix}$	4	I need to make all entries at below the cursor zero.	Add $\frac{1}{3}$ times row 2 to row 3
$\begin{pmatrix} 3 & 3 & 0 & 9 \\ 0 & \underline{6} & 4 & -12 \\ 0 & 0 & -\frac{5}{3} & 0 \end{pmatrix}$	5	Move the cursor down and right. It'll be on the bottom row, so I'm done.	

This matrix is in row-echelon form. What you would typically write down on your sheet of paper is something like this:

$$\begin{pmatrix} 0 & 6 & 4 & -12 \\ 3 & 3 & 0 & 9 \\ 2 & 0 & -3 & 10 \end{pmatrix} \xrightarrow{R_1 \leftrightarrow R_2} \begin{pmatrix} 3 & 3 & 0 & 9 \\ 0 & 6 & 4 & -12 \\ 2 & 0 & -3 & 10 \end{pmatrix} \\
 \xrightarrow{-\left(\frac{2}{3}\right)R_1 + R_2} \begin{pmatrix} 3 & 3 & 0 & 9 \\ 0 & 6 & 4 & -12 \\ 0 & -2 & -3 & 4 \end{pmatrix} \\
 \xrightarrow{\frac{1}{3}R_2 + R_3} \begin{pmatrix} 3 & 3 & 0 & 9 \\ 0 & 6 & 4 & -12 \\ 0 & 0 & -\frac{5}{3} & 0 \end{pmatrix}.$$

## Upward reduction

Matrix	Current Step	Thought Process	Subsequent Row Operation
$\begin{pmatrix} 3 & 3 & 0 & 9 \\ 0 & 6 & 4 & -12 \\ 0 & 0 & -\frac{5}{3} & 0 \end{pmatrix}$	1,2	Matrix is in ref form. Need to make pivots 1.	Multiply row 1 by $\frac{1}{3}$ ; multiply row 2 by $\frac{1}{6}$ ; multiply row 3 by $-\left(\frac{3}{5}\right)$ .
$\begin{pmatrix} 1 & 1 & 0 & 3 \\ 0 & 1 & \frac{2}{3} & -2 \\ 0 & 0 & 1 & 0 \end{pmatrix}$	3	Place cursor on right-most pivot.	
$\begin{pmatrix} 1 & 1 & 0 & 3 \\ 0 & 1 & \frac{2}{3} & -2 \\ 0 & 0 & \underline{1} & 0 \end{pmatrix}$	4	I need to make all entries at above the cursor zero.	Add $-\left(\frac{2}{3}\right)$ times row 3 to row 2
$\begin{pmatrix} 1 & 1 & 0 & 3 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & \underline{1} & 0 \end{pmatrix}$	5	Move the cursor up and left. Return to step 4.	
$\begin{pmatrix} 1 & 1 & 0 & 3 \\ 0 & \underline{1} & 0 & -2 \\ 0 & 0 & 1 & 0 \end{pmatrix}$	4	I need to make all entries at above the cursor zero.	Add $-1$ times row 2 to row 1
$\begin{pmatrix} 1 & 0 & 0 & 5 \\ 0 & \underline{1} & 0 & -2 \\ 0 & 0 & 1 & 0 \end{pmatrix}$	5	Move the cursor up and left. It'll be on the first row, so I'm done.	

Notice that this matrix is in its rref form. What you would typically write down on your sheet of paper for the upward steps is something like this:

$$\begin{aligned}
 \begin{pmatrix} 3 & 3 & 0 & 9 \\ 0 & 6 & 4 & -12 \\ 0 & 0 & -\frac{5}{3} & 0 \end{pmatrix} &\xrightarrow{\frac{1}{3}\cdot R_1, \frac{1}{6}\cdot R_2, -\frac{3}{5}\cdot R_3} \begin{pmatrix} 1 & 1 & 0 & 3 \\ 0 & 1 & \frac{2}{3} & -2 \\ 0 & 0 & 1 & 0 \end{pmatrix} \\
 &\xrightarrow{-\left(\frac{2}{3}\right)R_3+R_2} \begin{pmatrix} 1 & 1 & 0 & 3 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 0 \end{pmatrix} \\
 &\xrightarrow{-1\cdot R_2+R_1} \begin{pmatrix} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 0 \end{pmatrix}.
 \end{aligned}$$

**P.S.** The answer to the question that was asked (compute  $rref(A)$ ) is

$$rref(A) = \boxed{\begin{pmatrix} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 0 \end{pmatrix}}.$$

**Putting all this together**EXAMPLE 7

---

$$\text{Solve the system } \begin{cases} x + 2y & = 4 \\ x & + z = -3 \\ x - y & z = -4 \end{cases} .$$

## EXAMPLE 8

---

Solve the system  $\begin{cases} 3x_1 + x_2 - 6x_3 = -10 \\ 2x_1 + x_2 - 5x_3 = -8 \end{cases}$ .

EXAMPLE 9

---

Solve the system  $\begin{cases} x_1 & & +x_3 & +x_4 & = & 5 \\ x_1 & +x_2 & +x_3 & +x_4 & = & 6 \\ x_1 & -x_2 & +x_3 & +x_4 & = & 5 \end{cases} .$

## EXAMPLE 11

Find  $C(A)$  and find a basis for  $C(A)$ , if  $A = \begin{pmatrix} 2 & -1 & 1 & -3 \\ 3 & 2 & 5 & -8 \\ 1 & 6 & 7 & -8 \end{pmatrix}$ .

*“Old” solution:*

*New solution:* Perform Gaussian elimination on  $A$ .

$$A = \begin{pmatrix} 2 & -1 & 1 & -3 \\ 3 & 2 & 5 & -8 \\ 1 & 6 & 7 & -8 \end{pmatrix} \xrightarrow{R_1 \leftrightarrow R_3} \begin{pmatrix} 1 & 6 & 7 & -8 \\ 3 & 2 & 5 & -8 \\ 2 & -1 & 1 & -3 \end{pmatrix}$$

$$\xrightarrow{\substack{-3 \times R_1 + R_2 \\ -2 \times R_1 + R_3}} \begin{pmatrix} 1 & 6 & 7 & -8 \\ 0 & -16 & -16 & 16 \\ 0 & -13 & -13 & 13 \end{pmatrix}$$

$$\xrightarrow{\frac{-13}{16} \times R_2 + R_3} \begin{pmatrix} 1 & 6 & 7 & -8 \\ 0 & -16 & -16 & 16 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\xrightarrow{\frac{-1}{16} \times R_2} \begin{pmatrix} 1 & 6 & 7 & -8 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

**Notice:**

Therefore in any *ref* form of a matrix, the **linearly independent columns are the pivot columns**, and the **linearly dependent columns are the free columns**.

**WARNING:** In general,  $C(A) \neq C(\text{ref}(A))$ . These subspaces only have the same dimension.

**What about  $R(A)$ ?** From an earlier result,  $R(A) = R(\text{ref}(A)) = R(\text{rref}(A))$  since  $A$ ,  $\text{ref}(A)$  and  $\text{rref}(A)$  are row equivalent.

To summarize:

**Theorem 6.17 (Bases for column space and row space)** Let  $A \in M_{mn}(\mathbb{R})$ . Then:

1. A basis for  $C(A)$  consists of the pivot columns of  $A$  (not  $\text{ref}(A)$  or  $\text{rref}(A)$ ).
2. A basis for  $R(A)$  consists of the nonzero rows in any echelon form of  $A$ .

### Computing a basis of null space of a matrix

Let  $A \in M_{mn}(\mathbb{R})$ . Remember that  $\mathbf{x} = (x_1, \dots, x_n) \in N(A) \iff A\mathbf{x} = \mathbf{0}$ .

$$(A|\mathbf{0}) \xrightarrow{\text{row ops}} \text{rref}(A|\mathbf{0}) = \left( \begin{array}{cccccc|c} 1 & & * & * & & * & 0 \\ & 1 & * & * & & * & 0 \\ & & 1 & * & * & * & 0 \\ & & & & 1 & * & \vdots \\ & & & & & 1 & * & \vdots \\ & & & & & & 1 & \vdots \\ & & & & & & & \ddots \\ & & & & & & & \ddots \\ & & & & & & & 0 \end{array} \right)$$

Now write as a system of equations, and solve each equation for the left-most variable. Then, write

$$N(A) = \{(x_1, \dots, x_n)\}$$



and substitute in for each pivot variable, separate and write as a span. In this example, we would get

Remember that in general,  $\dim N(A) = n - r = \#$  of free columns.

EXAMPLE 12

Find a basis for  $N(A)$  if  $A = \begin{pmatrix} 1 & 1 & -2 & -5 \\ 2 & 0 & 2 & 2 \\ 3 & 0 & 3 & 3 \\ 4 & 0 & 4 & 4 \end{pmatrix}$ .

*Solution:* Perform Gaussian elimination on  $(A|\mathbf{0})$ :

$$\begin{aligned}
 A &= \left( \begin{array}{cccc|c} 1 & 1 & -2 & -5 & 0 \\ 2 & 0 & 2 & 2 & 0 \\ 3 & 0 & 3 & 3 & 0 \\ 4 & 0 & 4 & 4 & 0 \end{array} \right) \xrightarrow{-2R_1+R_2, -3R_1+R_3, -4R_1+R_4} \left( \begin{array}{cccc|c} 1 & 1 & -2 & -5 & 0 \\ 0 & -2 & 6 & 12 & 0 \\ 0 & -3 & 9 & 18 & 0 \\ 0 & -4 & 12 & 24 & 0 \end{array} \right) \\
 &\xrightarrow{-\frac{1}{2}\cdot R_2, -\frac{1}{3}\cdot R_3, -\frac{1}{4}\cdot R_4} \left( \begin{array}{cccc|c} 1 & 1 & -2 & -5 & 0 \\ 0 & 1 & -3 & -6 & 0 \\ 0 & 1 & -3 & -6 & 0 \\ 0 & 1 & -3 & -6 & 0 \end{array} \right) \\
 &\xrightarrow{-R_2+R_3, -R_2+R_4} \left( \begin{array}{cccc|c} 1 & 1 & -2 & -5 & 0 \\ 0 & 1 & -3 & -6 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right) \\
 &\xrightarrow{-R_2+R_3, -R_2+R_4} \left( \begin{array}{cccc|c} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & -3 & -6 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right) = rref(A|\mathbf{0})
 \end{aligned}$$

## 6.4 Summary so far

Let  $A \in M_{mn}(\mathbb{R})$  have rank  $r$  ( $r$  is the number of pivots in any row echelon form of  $A$ ). Then:

1. The system  $Ax = \mathbf{0}$  always has at least one solution (namely  $\mathbf{x} = \mathbf{0}$ ). There are two possible situations, (a) or (b):

a)  $Ax = \mathbf{0}$  has more than one solution. This is equivalent to all of the following:

- $N(A) \neq \{\mathbf{0}\}$
- $\dim N(A) \geq 1$
- $r < n$  (i.e. there is at least one free column).

In this case, for any  $\mathbf{b} \in \mathbb{R}^n$ :

- $Ax = \mathbf{b}$  has no solution if and only if  $\mathbf{b} \notin C(A)$ ;
- $Ax = \mathbf{b}$  has infinitely many solutions if and only if  $\mathbf{b} \in C(A)$  (in this case the solution set of  $Ax = \mathbf{b}$  is  $\mathbf{x}_p + N(A)$  where  $\mathbf{x}_p$  is any particular solution of the system) (this case is assured if  $r = m < n$ );
- $Ax = \mathbf{b}$  never has exactly one solution.

b)  $Ax = \mathbf{0}$  has exactly one solution (only  $\mathbf{x} = \mathbf{0}$ ). This is equivalent to all of the following:

- $N(A) = \{\mathbf{0}\}$
- $\dim N(A) = 0$
- $r = n$  (i.e. all the columns are pivot columns).

In this case, for any  $\mathbf{b} \in \mathbb{R}^n$ :

- $Ax = \mathbf{b}$  has no solution if and only if  $\mathbf{b} \notin C(A)$ ;
- $Ax = \mathbf{b}$  has exactly one solution if and only if  $\mathbf{b} \in C(A)$  (this is assured if  $r = m = n$ ; see below);
- $Ax = \mathbf{b}$  never has infinitely many solutions.

Notice that if  $m < n$  (that is, there are fewer equations than variables), case (b) above is impossible because  $r \leq m$  (so  $r$  cannot be equal to  $n$ ).

2. In the special case where  $r = m = n$  (i.e.  $A$  is square and has **full rank**), then  $C(A) = R(A) = \mathbb{R}^n$  and  $N(A) = N(A^T) = \{\mathbf{0}\}$ . Furthermore, for every  $\mathbf{b} \in \mathbb{R}^n$ , the system  $Ax = \mathbf{b}$  has exactly one solution.

This situation is equivalent to many different things, which are described in Theorem 6.8.

3. The system  $Ax = \mathbf{b}$  has no solution  $\Leftrightarrow \mathbf{b} \notin C(A) \Leftrightarrow$  an echelon form of the augmented matrix  $(A | \mathbf{b})$  contains a **false row** of the form  $(0 \ 0 \ \cdots \ 0 \ | \ z)$  where  $z \neq 0$ .

## 6.5 Computing the inverse of a matrix

## QUESTION

Suppose  $A \in M_n(\mathbb{R})$  is an invertible matrix. How do you compute its inverse  $A^{-1}$ ?

Suppose  $A \in M_n(\mathbb{R})$  is invertible. Write  $A = (a_{ij})$  and write  $A^{-1} = (b_{ij})$ . (The object is to solve for the  $b_{ij}$ s.) Since  $AA^{-1} = I$ , we see by the way that matrix multiplication works that

$$A \rightarrow \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ \vdots & \vdots & & 1 & & \vdots \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix} \leftarrow I$$

Now look at the  $j^{\text{th}}$  column of this product. This produces the equation

$$A \begin{pmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{pmatrix} = \mathbf{e}_j,$$

which we could solve by Gaussian elimination to produce the vector  $\begin{pmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{pmatrix}$ ,

which is the  $j^{\text{th}}$  column of  $A^{-1}$ .

**Theorem 6.18** Suppose  $A \in M_n(\mathbb{R})$  is invertible. If  $\mathbf{b}_j$  is the  $j^{\text{th}}$  column of  $A^{-1}$ , then  $\mathbf{b}_j$  is the solution of  $A\mathbf{b}_j = \mathbf{e}_j$ .

So to get all the columns of  $A^{-1}$ , we solve the equations  $Ax = \mathbf{e}_j$  all at once. This is called the *Gauss-Jordan method* of computing the inverse of a square matrix  $A$ .

**Gauss-Jordan method** to compute the inverse of a matrix  $A$ :

1. Write the augmented matrix  $(A|I)$ .
2. Perform Gaussian elimination to put  $A$  in rref form.
  - If you get a row of zeros in the  $A$  part of the matrix at any point,  $A$  is not invertible.
  - Otherwise, you will end up with a matrix of the form  $(I|A^{-1})$ .

**Note:** In practice, one uses a computer or calculator to compute the inverse of a matrix that is  $3 \times 3$  or larger (see Section 6.7).

EXAMPLE 13

Compute the inverse of  $A = \begin{pmatrix} 1 & 2 & 1 \\ -1 & 0 & 2 \\ 0 & 1 & 4 \end{pmatrix}$ .

$$\text{Solution: } (A|I) = \left( \begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ -1 & 0 & 2 & 0 & 1 & 0 \\ 0 & 1 & 4 & 0 & 0 & 1 \end{array} \right)$$

## 6.6 Least-squares approximations

EXAMPLE 14

Find the equation of the plane in  $\mathbb{R}^3$  passing through the following points:

$$(1, 1, 1) \quad (2, 2, 2) \quad (2, 1, 1) \quad (1, 3, 4)$$

*First attempt at a solution:* We know every plane in  $\mathbb{R}^3$  has the normal equation

$$\mathbf{n} \cdot \mathbf{x} = d, \quad \text{a.k.a.} \quad ax + by + cz = d$$

so by solving for  $z$ , we see that every plane (with  $c \neq 0$ ) in  $\mathbb{R}^3$  has equation

$$z =$$

To determine this equation, plug each of the given points in for  $x, y$  and  $z$ :

$$\left\{ \begin{array}{l} \\ \\ \\ \end{array} \right.$$

This translates into the matrix equation  $A\mathbf{x} = \mathbf{b}$  where  $\mathbf{x} = (\alpha, \beta, \gamma)$ ;

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \end{pmatrix}; \quad \mathbf{b} = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 4 \end{pmatrix}.$$

Notice:

first column of  $A \leftrightarrow$  a column of all ones  
 last column of  $A \leftrightarrow$   
 second column of  $A \leftrightarrow$   
 the column vector  $\mathbf{b} \leftrightarrow$

We'd solve such a system by Gaussian elimination/row reductions:

$$(A | \mathbf{b}) = \left( \begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 3 & 4 \end{array} \right) \xrightarrow{\text{row ops}} \left( \begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{array} \right) = \text{ref}(A | \mathbf{b})$$

Therefore

EXAMPLE 14, RESTATED

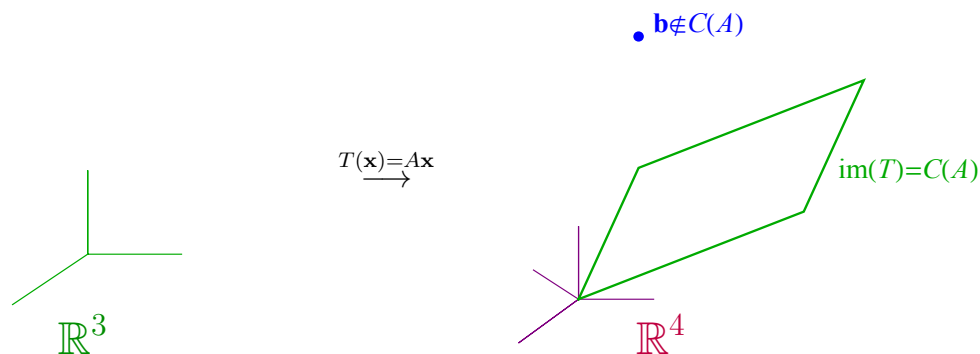
Find the equation of the plane in  $\mathbb{R}^3$  which *best fits* the following points:

$$(1, 1, 1) \quad (2, 2, 2) \quad (2, 1, 1) \quad (1, 3, 4)$$

**Solution:** Start by repeating the work on the previous page to obtain the matrix equation  $A\mathbf{x} = \mathbf{b}$  where

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \end{pmatrix}_{4 \times 3}; \quad \mathbf{b} = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 4 \end{pmatrix}$$

From the last page, we know there is no solution  $\mathbf{x}$  to this system, which means  $\mathbf{b} \notin C(A)$ . This means we can think of a picture like this:



The plane which best fits the data will be given by the  $\mathbf{x} = (\alpha, \beta, \gamma)$  which makes  $A\mathbf{x}$  as close to possible to  $\mathbf{b}$ . Thus

$A\mathbf{x}$  should be equal to

QUESTION: How do you compute  $\hat{\mathbf{b}}$ ?

## QUESTION FROM THE PREVIOUS PAGE

How do you compute  $\hat{\mathbf{b}}$ , the projection of  $\mathbf{b}$  onto  $C(A)$ ?

*Old method (from Chapter 4):*  $C(A)$  is spanned by the columns of  $A$ .

So use Gram-Schmidt to convert the columns of  $A$  into an orthonormal basis  $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$  of  $C(A)$ , and then use the projection formula.

This gives

$$\hat{\mathbf{b}} = (\mathbf{b} \cdot \mathbf{x}_1)\mathbf{x}_1 + (\mathbf{b} \cdot \mathbf{x}_2)\mathbf{x}_2 + (\mathbf{b} \cdot \mathbf{x}_3)\mathbf{x}_3.$$

While this method always works, there are two big drawbacks: it is very computationally intensive, and doesn't tell you an  $\hat{\mathbf{x}}$  that satisfies  $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$ .

*New method:* Since  $\hat{\mathbf{b}} \in C(A)$ , the system  $A\mathbf{x} = \hat{\mathbf{b}}$  has at least one solution.

Call this solution  $\hat{\mathbf{x}}$ . Therefore  $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$ .

However, since  $\hat{\mathbf{b}}$  is the projection of  $\mathbf{b}$  onto  $C(A)$ ,

$$\mathbf{b} - \hat{\mathbf{b}} \in [C(A)]^\perp = N(A^T).$$

That means  $A^T(\mathbf{b} - \hat{\mathbf{b}}) = \mathbf{0}$ . Now, substituting in  $A\hat{\mathbf{x}}$  for  $\hat{\mathbf{b}}$ , we see

$$\mathbf{0} = A^T(\mathbf{b} - A\hat{\mathbf{x}})$$

$$\mathbf{0} = A^T\mathbf{b} - A^T A\hat{\mathbf{x}}$$

$$A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$$

$$(A^T A)^{-1} A^T A\hat{\mathbf{x}} = (A^T A)^{-1} A^T\mathbf{b}$$

Therefore

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T\mathbf{b}$$

and since  $\hat{\mathbf{b}} = A\hat{\mathbf{x}}$ ,

$$\hat{\mathbf{b}} = A(A^T A)^{-1} A^T\mathbf{b}.$$

$\hat{\mathbf{x}}$  is called the **least-squares** approximation to the original equation  $A\mathbf{x} = \mathbf{b}$ .

## BACK TO EXAMPLE 14

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \end{pmatrix}; \quad \mathbf{b} = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 4 \end{pmatrix}$$

The least-squares approximation is

$$\begin{aligned}\hat{\mathbf{x}} &= (A^T A)^{-1} A^T \mathbf{b} \\ &= \left[ \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 1 \\ 1 & 2 & 1 & 3 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \end{pmatrix} \right]^{-1} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 1 \\ 1 & 2 & 1 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 1 \\ 4 \end{pmatrix} \\ &= \begin{pmatrix} 4 & 6 & 7 \\ 6 & 10 & 10 \\ 7 & 10 & 15 \end{pmatrix}^{-1} \begin{pmatrix} 8 \\ 11 \\ 18 \end{pmatrix} \\ &= \begin{pmatrix} 5 & -2 & -1 \\ -2 & \frac{11}{10} & \frac{1}{5} \\ -1 & \frac{1}{5} & \frac{1}{5} \end{pmatrix} \begin{pmatrix} 8 \\ 11 \\ 18 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{-3}{10} \\ \frac{7}{5} \end{pmatrix}.\end{aligned}$$

Since  $\hat{\mathbf{x}} = (\alpha, \beta, \gamma) = (0, \frac{-3}{10}, \frac{7}{5})$ , the plane best fitting the data points

$$(1, 1, 1) \quad (2, 2, 2) \quad (2, 1, 1) \quad (1, 3, 4)$$

is

If you have access to *Mathematica*, implementing this entire process is easy: type in the matrices  $A$  and  $\mathbf{b}$  and then run the command

$$\text{LeastSquares}[A, \mathbf{b}] // \text{MatrixForm}$$

This will compute and display  $\hat{\mathbf{x}}$ .

**Potential problem:** The only drawback to the least-square method is that it requires the matrix  $A^T A$  to be invertible to work. Fortunately, this is assured in most situations because of the following fact:

**FACT:** Let  $A \in M_{mn}(\mathbb{R})$ . If the columns of  $A$  are linearly independent, then the  $n \times n$  matrix  $A^T A$  is invertible.

In the context of Example 14, this is assured if the same  $x, y$  coordinates from the original list of points are never repeated.



To summarize:

**Theorem 6.19 (Least-squares)** Let  $A \in M_{mn}(\mathbb{R})$  and let  $\mathbf{b} \in \mathbb{R}^m$ . Then, assuming the columns of  $A$  are linearly independent,

1. The projection of  $\mathbf{b}$  onto the column space of  $A$  is

$$\hat{\mathbf{b}} = \pi_{C(A)}\mathbf{b} = A(A^T A)^{-1}A^T\mathbf{b}$$

2. The **least-squares solution** of the system  $A\mathbf{x} = \mathbf{b}$  is the vector  $\hat{\mathbf{x}} \in \mathbb{R}^n$  given by

$$\hat{\mathbf{x}} = (A^T A)^{-1}A^T\mathbf{b}$$

$\hat{\mathbf{x}}$  solves the equation  $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$ , and is the closest approximation to any “solution” of  $A\mathbf{x} = \mathbf{b}$ . It isn’t generally a solution of  $A\mathbf{x} = \mathbf{b}$ , however.

**Theorem 6.20** Let  $A \in M_n(\mathbb{R})$  be an invertible square matrix and let  $\mathbf{b} \in \mathbb{R}^n$ . Then the least-squares solution of the system  $A\mathbf{x} = \mathbf{b}$  is  $\hat{\mathbf{x}} = A^{-1}\mathbf{b}$  and is the same as the actual solution of the original system.

PROOF If  $A$  is invertible, so is  $A^T$  and  $(A^T A)^{-1} = A^{-1}(A^T)^{-1}$ . Therefore

$$\hat{\mathbf{x}} = (A^T A)^{-1}A^T\mathbf{b} = A^{-1}(A^T)^{-1}A^T\mathbf{b} = A^{-1}I\mathbf{b} = A^{-1}\mathbf{b}$$

and  $A\hat{\mathbf{x}} = A(A^{-1}\mathbf{b}) = \mathbf{b}$ .  $\square$

We wrap this material up by restating the already-proven facts in terms of projections:

**Definition 6.21** Let  $A \in M_{mn}(\mathbb{R})$  be a matrix whose columns are linearly independent. Then define the  $m \times m$  matrix  $P$  to be

$$P = A(A^T A)^{-1}A^T.$$

$P$  is called the **projection matrix** for  $A$ .

**Theorem 6.22** Let  $A \in M_{mn}(\mathbb{R})$  be a matrix whose columns are linearly independent, and let  $P$  be its projection matrix. Then for any  $\mathbf{y} \in \mathbb{R}^m$ ,  $P\mathbf{y} = \pi_{C(A)}\mathbf{y}$ , the projection of  $\mathbf{y}$  onto the column space of  $A$ .

## EXAMPLE 15

Compute the projection of  $(2, -5, 1, 7)$  onto the subspace  $W$  spanned by  $(1, 3, -2, 1)$  and  $(4, -1, 1, 3)$ .

$$\begin{aligned} &= \frac{1}{401} \begin{pmatrix} 1 & 4 \\ 3 & -1 \\ -2 & 1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} -286 \\ 541 \end{pmatrix} \\ &= \frac{1}{401} \begin{pmatrix} 1878 \\ -1399 \\ 1113 \\ 1337 \end{pmatrix} \\ &= \boxed{\left( \frac{1878}{401}, -\frac{1399}{401}, \frac{1113}{401}, \frac{1337}{401} \right)}. \end{aligned}$$

## 6.7 Solving systems of equations using technology

### Linear systems on *Mathematica*

*Mathematica* does each of the computations you need to be able to do with one command line:

1. To solve a system of linear equations  $Ax = b$  using *Mathematica*:
  - a) Type in the matrix  $A$  (using the syntax from Section 2.4).
  - b) Type in the vector  $b$  (also using the syntax from Section 2.4).
  - c) Execute `LinearSolve[A,b]`. This will produce a particular solution  $x_p$  (if possible) and will tell you if there is no solution.
  - d) Execute `NullSpace[A]`. This will spit out a list of vectors which form a basis of  $N(A)$ .
  - e) The solution, as the theory tells us, is  $x_p + N(A)$ . Make sure you write it correctly.

As an example, suppose we were trying to solve

$$\begin{cases} x + 2y - 3z = 4 \\ 2x + y - 5z = 1 \end{cases}$$

In *Mathematica*, execute the following commands:

*In:* `A = {{1,2,-3},{2,1,-5}}`

*In:* `b = {4,1}`

*In:* `LinearSolve[A,b]`

*Out:* `{{-2/3},{7/3},{0}}` ← This is a particular solution  $x_p$ .

*In:* `NullSpace[A]`

*Out:* `{{7,1,3}}` ← This is a basis for  $N(A)$

The solution is therefore

$$x_p + N(A) = \left( \frac{-2}{3}, \frac{7}{3}, 0 \right) + t(7, 1, 3) = \boxed{\left( \frac{-2}{3} + 7t, \frac{7}{3} + t, 3t \right)}.$$

2. Other things *Mathematica* can do:

- a) To find the *rref* form of a matrix, type in the matrix (saving it as, say  $A$ ) and then execute `RowReduce[A] //MatrixForm`.
- b) To find the rank of a matrix, type in the matrix and execute `MatrixRank[A]`.
- c) To find a basis for the null space of a matrix, type in the matrix and execute `NullSpace[A]`.
- d) To find the inverse of a square matrix, type in the matrix and execute `Inverse[A]`.
- e) To find the least-squares solution of a system, type in  $A$  and  $b$  and execute `LeastSquares[A,b] //MatrixForm`; the output is  $\hat{x}$ .

### Linear systems on TI-83/84 type calculators

1. To solve a system of linear equations  $Ax = b$  using such a calculator:
  - a) Type in the augmented matrix  $(A|b)$  and save it (as a single matrix) using the directions in Section 2.4.
  - b) Then hit `[MATRX]`, go to the right to find `MATH`, then go down to `rref`. Hit `[ENTER]`, then pull up the matrix you have saved by hitting `[MATRX]` and scrolling down to the matrix you want. Then execute this command; your calculator will spit out the reduced row-echelon form of the augmented matrix.
  - c) Solve the system of equations corresponding to the *rref* using the procedures described in Section 6.3.

As an example, if we were trying to solve

$$\begin{cases} x + 2y - 3z = 4 \\ 2x + y - 5z = 1 \end{cases}$$

we would first save the matrix

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

as `[A]`, then ask the calculator to compute `rref([A])`. You will get (after converting the answer to fractions)

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

This translates back to the system

$$\begin{cases} x + \frac{1}{2}y - \frac{5}{2}z = \frac{1}{2} \\ y - \frac{1}{3}z = \frac{7}{3} \end{cases}$$

which you solve by hand (as in Section 6.3) to get the solution.

2. Other things the graphics calculator will do:
  - a) To find the *rref* form of a matrix  $A$ , first type in the matrix and save it using the directions of Chapter 2. Then hit [MATRX], go to the right to find MATH, then go down to *rref*(. Hit [ENTER], then pull up the matrix you have saved by hitting [MATRX] and scrolling down to the matrix you want. Then execute this command; your calculator will spit out the reduced row-echelon form of  $A$ .
  - b) To find the inverse of a matrix, first type in the matrix and save it using the directions of Chapter 2. Then hit [MATRX] and scroll down to the matrix you want to take the inverse of; hit [ENTER]. Then hit [ $x^{-1}$ ] and execute this command; your calculator will spit out the inverse of  $A$  and will give you an error if  $A$  is not invertible.

### Using an online equation solver

1. To solve a system of linear equations  $Ax = b$  using a free online tool:
  - a) Open a web browser to <http://www.quickmath.com>
  - b) Type the equations in the box, separated by commas. Click "Solve".
  - c) The solution will appear below (including cases of no solution or infinitely many solutions).
2. There are probably lots of other online tools which can be used to compute bases of a null space or perform row reductions or compute inverses of matrices; I'll leave it to you to use Google to find them. (I don't suggest using WolframAlpha, because the output you get is a bit clunky.)

## 6.8 Homework exercises for Chapter 6

1. Assume that the following chart shows the number of grams of nutrients per ounce of food indicated:

	BEEF	POTATO	CABBAGE
PROTEIN	20	5	1
FAT	4	7	12
CARBOHYDRATES	15	20	5

Suppose the army desires to use these three delectable foods to feed new recruits a dinner providing 305 grams of protein, 365 grams of fat, and 575 grams of carbohydrates. Write a system of equations which, when solved, will figure out how much of each food should be prepared for each recruit. (Be sure to define any variables which are used in your problem.)

2. Consider the following system of linear equations:

$$\begin{cases} x - 2y + 3z - \frac{w}{2} - 2 = 5 - 3x + z \\ -\sqrt{3}x + y + 5z + \pi^2 w = -4 \\ 2(x - 3z) = 6 \end{cases}$$

- Write the system as a matrix equation  $Ax = b$  (i.e. what is  $A$ ? What is  $x$ ? What is  $b$ ?)
  - Write the system as a vector equation.
  - Essentially, this system is asking whether some vector (say  $v$ ) is in the span of some other vectors (say  $w_1, w_2, \dots$ ). In this context, what is the  $v$ ? What are the  $w_1, w_2, \dots$ ?
3. Suppose  $A$  is some matrix with  $N(A) = \text{Span}((1, 3, 0, -2), (1, 0, 0, -2))$ . Suppose also that  $A(1 - 3, 0, 2) = (2, 0, 3, 1, 4)$ .
- How many equations, and how many variables are in the system  $Ax = (2, 0, 3, 1, 4)$ ?
  - What are the dimensions of the four fundamental subspaces of  $A$ ?
  - Describe all solutions to the equation  $Ax = (2, 0, 3, 1, 4)$ .
4. Suppose  $A$  is some matrix with  $N(A) = \{0\}$ . Suppose also that  $A(1 - 3, 0, 2) = (8, 1, -2, 5)$ .
- How many equations, and how many variables are in the system  $Ax = (8, 1, -2, 5)$ ?
  - What are the dimensions of the four fundamental subspaces of  $A$ ?

- c) Describe all solutions to the equation  $A\mathbf{x} = (8, 1, -2, 5)$ .
5. Suppose  $A$  is some matrix with  $N(A) = \{0\}$ . Suppose also that  $A(2, 5) = (1, 1, 1, 1, 1)$ .
- How many equations, and how many variables are in the system  $A\mathbf{x} = (1, 1, 1, 1, 1)$ ?
  - What are the dimensions of the four fundamental subspaces of  $A$ ?
  - Describe all solutions to the equation  $A\mathbf{x} = (1, 1, 1, 1, 1)$ .
6. Use Gaussian elimination to row reduce the following matrix to a row-echelon form, then to a reduced row-echelon form. In this problem, and only in this problem, you are required to perform Gaussian elimination by hand and show all your steps.

$$A = \begin{pmatrix} 1 & 2 & -3 & 0 \\ 2 & 4 & -2 & 2 \\ 3 & 6 & -4 & 3 \end{pmatrix}$$

7. Find the reduced row-echelon form of the matrix below:

*Note:* Henceforth, to perform Gaussian elimination I recommend that you use a calculator, or *Mathematica*, or an online row reducer like the ones at

<http://www.math.purdue.edu/~dvb/matrix.html> or  
<http://www.math.odu.edu/~bogacki/cgi-bin/lat.cgi?c=roc>

$$\begin{pmatrix} 2 & 2 & -1 & 6 & 4 \\ 4 & 4 & 1 & 10 & 13 \\ 6 & 6 & 0 & 20 & 19 \end{pmatrix}$$

8. Solve this system:

$$\begin{cases} x - 2y + z = 7 \\ 2x - y + 4z = 17 \\ 3x - 2y + 2z = 14 \end{cases}$$

9. Solve this system:

$$\begin{cases} x + 2y - z = 3 \\ x + 3y + z = 5 \\ 3x + 8y + 4z = 17 \end{cases}$$

10. Solve this system:

$$\begin{cases} x_1 + 2x_2 - 5x_3 = 1 \\ x_1 - 3x_2 + 3x_3 - x_4 = -4 \\ x_2 = -5 \\ -2x_1 + 2x_2 + 2x_3 + 3x_4 = -2 \end{cases}$$

11. Solve this system:

$$\begin{cases} x_1 - 2x_2 + x_3 + x_4 = 7 \\ 2x_2 - x_3 - 7x_4 = 6 \\ -3x_3 + 2x_4 = -6 \end{cases}$$

12. Solve this system:

$$\begin{cases} x + 2y + 3z = 1 \\ 2x + 3y + 4z = 0 \\ 3x + 4y + 5z = 1 \end{cases}$$

13. Solve this system:

$$\begin{cases} x_1 - 2x_2 - 3x_3 + 5x_4 - 2x_5 = 4 \\ 2x_3 - 6x_4 + 3x_5 = 2 \\ 5x_5 = 10 \end{cases}$$

14. Solve three of these four problems:

- Determine if the following three planes in  $\mathbb{R}^3$  have at least one common point of intersection:  $x + 2y + z = 4$ ,  $y - z = 1$  and  $x + 3y = 0$ .
- Find the unique polynomial of degree at most 3 which goes through the points  $(1, 1)$ ,  $(2, 3)$ ,  $(3, 6)$  and  $(4, 10)$ .  
*Hint:* write the polynomial as  $y = ax^3 + bx^2 + cx + d$ ; substitute in the given points to get a system of four equations in the four variables  $(a, b, c, d)$ .
- Express  $(1, -2, 5)$  as a linear combination of  $(1, -3, 2)$ ,  $(2, -4, -1)$  and  $(1, -5, 7)$ , if it can be done; if not, explain why this problem is impossible.
- Let  $W$  be the subspace of  $\mathbb{R}^4$  spanned by  $(1, 2, -5, 2)$  and  $(0, 1, 3, -1)$ . Find a basis of  $W^\perp$ .

15. Assume the matrices  $A$  and  $B$  given below are row equivalent. Find a basis for  $C(A)$ , a basis for  $R(A)$ , and a basis for  $N(A)$ .

$$A = \begin{pmatrix} 1 & -4 & 9 & -7 \\ -1 & 2 & -4 & 1 \\ 5 & -6 & 10 & 7 \end{pmatrix}; B = \begin{pmatrix} 1 & 0 & -1 & 5 \\ 0 & -2 & 5 & -6 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

16. Assume the matrices  $A$  and  $B$  given below are row equivalent. Find a basis for  $C(A)$ , a basis for  $R(A)$ , and a basis for  $N(A)$ .

$$A = \begin{pmatrix} 1 & 1 & -3 & 7 & 9 & -9 \\ 1 & 2 & -4 & 10 & 13 & -12 \\ 1 & -1 & -1 & 1 & 1 & -3 \\ 1 & -3 & 1 & -5 & -7 & 3 \\ 1 & -2 & 0 & 0 & -5 & -4 \end{pmatrix}; B = \begin{pmatrix} 1 & 1 & -3 & 7 & 9 & -9 \\ 0 & 1 & -1 & 3 & 4 & -3 \\ 0 & 0 & 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$



17. In each of the following, determine (a) if the equation  $Ax = 0$  has a nontrivial solution (“nontrivial” means a solution other than  $\mathbf{x} = \mathbf{0}$ ) and (b) if the equation  $Ax = \mathbf{b}$  has at least one solution for every choice of  $\mathbf{b}$ .

- $A$  is a  $3 \times 3$  matrix with 3 pivot columns.
- $A$  is a  $2 \times 4$  matrix with rank 2.
- $A$  is an  $8 \times 6$  matrix with 1-dimensional null space.
- $A$  is a  $5 \times 3$  matrix with 3 pivot columns.
- $A$  is a  $4 \times 4$  matrix with rank 3.

18. Consider the system of equations  $Ax = \mathbf{b}$  where

$$A = \begin{pmatrix} 1 & -2 & 4 & 0 \\ -1 & -1 & -1 & 0 \\ 0 & 4 & -4 & 3 \\ 3 & 2 & 4 & -2 \\ 1 & 0 & 2 & 2 \end{pmatrix} \text{ and } \mathbf{b} = \begin{pmatrix} 4 \\ 2 \\ -11 \\ -2 \\ -2 \end{pmatrix}.$$

For each  $j$ , let  $\mathbf{a}_j$  denote the  $j^{\text{th}}$  column of  $A$ .

- Find the solution set of this system.
  - Find bases for the column space of  $A$ , the row space of  $A$ , and the null space of  $A$ .
  - Find the dimensions of  $R(A)$ ,  $C(A)$ ,  $N(A)$  and  $N(A^T)$ .
  - Find the rank of  $A$ .
  - Let  $T : \mathbb{R}^a \rightarrow \mathbb{R}^b$  be the linear transformation defined by  $T(\mathbf{x}) = A\mathbf{x}$ . What are  $a$  and  $b$ ? Find bases for the kernel and image of  $T$ , and the dimensions of the kernel and image of  $T$ .
  - Is  $\mathbf{b}$  in the span of the columns of  $A$ ? If so, write  $\mathbf{b}$  as a linear combination of the columns of  $A$ . If not, explain why not.
  - Find a vector which is not in the column space of  $A$ .
  - Is there any vector  $\mathbf{v} \in \mathbb{R}^5$  for which  $A\mathbf{x} = \mathbf{v}$  has exactly one solution? If so, find such a vector  $\mathbf{v}$ . If not, explain why not.
  - Is the set  $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4\}$  linearly independent? Why or why not?
  - Is the set  $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_4\}$  linearly independent? Why or why not?
  - Is the set  $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4, \mathbf{b}\}$  linearly independent? Why or why not?
19. Use the Gauss-Jordan method to find the inverse of this matrix (if the inverse exists). In this problem, and only in this problem, you are required to perform this method by hand and show all your steps.

$$\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

20. Use appropriate technology to find the inverse of this matrix (if the inverse exists).

$$\begin{pmatrix} 3 & 10 & 3 & 8 \\ 3 & -2 & 8 & 7 \\ 2 & 1 & 4 & -5 \\ 5 & 11 & 7 & 3 \end{pmatrix}$$

21. Use appropriate technology to find the inverse of this matrix (if the inverse exists).

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

22. Find the least-squares solution (a decimal approximation is okay; keep at least four decimal places) of the following system of equations:

$$\begin{cases} w - 2x + y - 2z = 4 \\ 2w + x + 4z = 5 \\ -3x - 2y + 5z = -2 \\ -w - x - y + 3z = 0 \\ w + 4y + z = 1 \\ 2x - y + 6z = -3 \end{cases}$$

23. Suppose you obtain a bunch of data points  $(x, y, z)$  which are supposed to fit a model of the form

$$z = \alpha + \beta x^2 + \gamma y^3$$

where  $\alpha, \beta, \gamma$  are constants. Suppose that the eight data points you obtain are:

$$\begin{aligned} (1, 2, -1) & \quad (0, 2, 0) & \quad (-2, 3, -1) & \quad (1, 5, 2) \\ (-1, -1, -3) & \quad (-4, 0, -6) & \quad (0, -2, 7) & \quad (2, 2, 7) \end{aligned}$$

- Write down the matrix equation  $A\mathbf{x} = \mathbf{b}$  you would solve to find  $\alpha, \beta, \gamma$  (i.e. what is  $A$ ? What is  $\mathbf{x}$ ? What is  $\mathbf{b}$ ?)
- Write down a formula in terms of  $A$  and/or  $\mathbf{b}$  which gives the least-squares solution  $\hat{\mathbf{x}}$ .
- Write down the formula in terms of  $A$ ,  $\mathbf{b}$  and/or  $\hat{\mathbf{x}}$  which gives the corresponding  $\hat{\mathbf{b}}$ .
- Compute the least-squares solution  $\hat{\mathbf{x}}$  to the matrix equation of part (a) (a decimal approximation is okay; keep at least four decimal places).
- Compute the corresponding  $\hat{\mathbf{b}}$ .
- Compute  $\|\mathbf{b} - \hat{\mathbf{b}}\|$ .

- g) What is the significance of the quantity you computed in part (f)? In other words, if that quantity is small, what does that say about the least-squares solution? If it is big, what does that say about the least-squares solution?
- h) What is the model that best fits the data?
- i) Use the model of part (h) to predict the value of  $z$  when  $x = 4$  and  $y = 5$ .
24. Suppose you are looking at a curve on an oscilloscope which for theoretical reasons should be the graph of some function of the form

$$y = a \cos x + b \sin x + c \cos 2x + d \sin 2x$$

where  $a, b, c$  and  $d$  are constants. If you see the following  $(x, y)$  points on your oscilloscope:

$$(0, 0) \quad (.2, -1.6) \quad (.7, 5.1) \quad (1, 8) \quad (1.3, 8.7) \quad (1.8, 4.4) \quad (2.2, -1.5)$$

use least-squares to compute (decimal approximations to) the values of  $a, b, c$  and  $d$  which best fit this data.

25. Suppose you are looking at data which is supposed to fit an exponential equation, i.e. a model of the form

$$y = Ce^{kx}$$

where  $C$  and  $k$  are constants. Suppose your data points are  $(2, 5)$ ,  $(3, 8)$  and  $(4, 17)$ . Use least-squares to find (decimal approximations to) the values of  $C$  and  $k$  which best fit this model.

*How to proceed:* First, take the natural log of both sides of the model to obtain what is called a “log-log” equation. Then, use log rules to rewrite the log-log equation as a linear equation in terms of the variables  $\ln C$  and  $k$ . Then use least-squares to compute  $\ln C$  and  $k$ ; last, find  $C$ .

26. Find the projection of  $(2, 1, -2, 4, -1)$  onto the subspace spanned by the three vectors  $(1, 2, -1, 0, 4)$ ,  $(1, -5, -1, 8, 2)$  and  $(0, 0, 2, 3, -2)$ .

## Chapter 7

---

# Determinants

---

### 7.1 Definition and properties

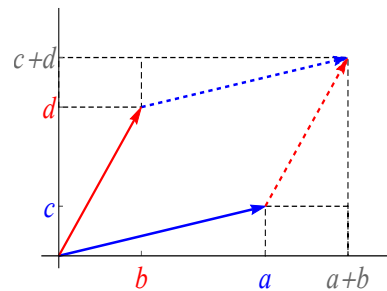
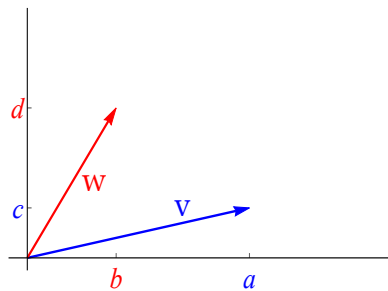
#### GOAL

Assign, to every square matrix, a real number which tells you something about the matrix.

As motivation, consider a  $2 \times 2$  square matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Treat the columns of  $A$  as vectors: let  $\mathbf{v} = (a, c)$  and  $\mathbf{w} = (b, d)$ . Use those vectors to make a parallelogram:



#### QUESTION?

What is the area of this parallelogram?

**Definition 7.1** Let  $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(\mathbb{R})$ . The **determinant** of  $A$ , denoted  $\det A$ , is the number  $\det A = ad - bc$ .

The determinant of  $A$  is also denoted  $|A|$ , but this a poor choice of notation that should be avoided.

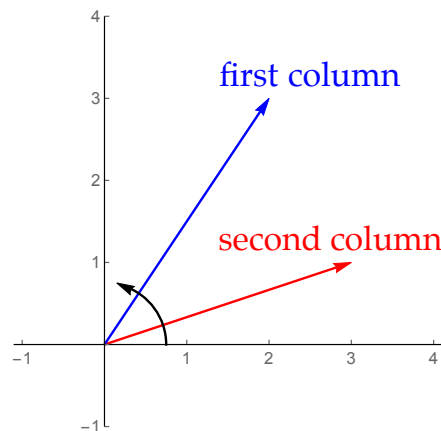
What we have shown is that if you have a parallelogram in  $\mathbb{R}^2$ , and you treat the sides of the parallelogram as vectors, and put those vectors in as columns of a matrix  $A$ , then

$$\text{area of the parallelogram} = |\det A|.$$

EXAMPLE 1

Compute the determinant of  $\begin{pmatrix} 2 & 3 \\ 3 & 1 \end{pmatrix}$ .

What is the significance of the fact that this determinant is negative?

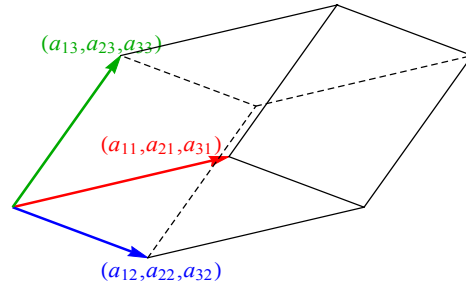


In general, the columns of an  $n \times n$  matrix can be thought of as  $n$  vectors in  $\mathbb{R}^n$ . Those  $n$  vectors can be used to make a  $n$ -dimensional box called a **parallelepiped**.

If you take the same vectors and put them in as columns of a matrix, this makes an  $n \times n$  matrix  $A$ . The **determinant** of this matrix is either  $+1$  or  $-1$  times the volume of this parallelepiped. Whether or not it is positive or negative depends on the “orientation” of the column vectors, i.e. which order they are written in. In

$\mathbb{R}^3$ , this goes like this:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \Rightarrow \det A = \pm \text{volume of}$$



**Remark:** There is a more rigorous definition of the determinant of a matrix, but that requires mathematics beyond the scope of this course.

**Note:** If  $A$  is not square, then  $\det A$  is undefined.

**Theorem 7.2 (Properties of determinants)** Let  $A, B \in M_n(\mathbb{R})$ . Then:

1.  $\det A = \det A^T$ .
2. If  $B$  is obtained from  $A$  by swapping two columns of  $A$ , then  $\det B = -\det A$ .
3. If  $B$  is obtained from  $A$  by swapping two rows of  $A$ , then  $\det B = -\det A$ .
4. If two rows or columns of  $A$  coincide, then  $\det A = 0$ .
5.  $\det I = 1$ , where  $I$  is the identity matrix.
6. If  $B$  is obtained from  $A$  by multiplying one row (or column) of  $A$  by a constant  $r$ , then  $\det B = r \det A$ .
7.  $\det(rA) = r^n \det A$ .
8.  $\det(AB) = \det A \cdot \det B$ .
9.  $A$  is invertible if and only if  $\det A \neq 0$ , in which case  $\det(A^{-1}) = \frac{1}{\det A}$ .

**Fact:** The only function  $M_n(\mathbb{R}) \rightarrow \mathbb{R}$  which satisfies (1), (2), (5) and (6) in the above theorem is the determinant.

## JUSTIFICATION OF SOME OF THESE

- (2): Swapping two columns of  $A$  doesn't change the parallelepiped whose volume is being computed, but does change the order in which the columns are written, causing the determinant to be multiplied by  $-1$ .
- (4): If two columns of  $A$  coincide, then the parallelepiped is "degenerate", i.e. has no  $n$ -dimensional volume. Since the absolute value of  $\det A$  is this volume, we have  $|\det A| = 0$ , i.e.  $\det A = 0$ .  
If two rows of  $A$  coincide, then two columns of  $A^T$  coincide so by the above,  $\det A^T = 0$  so by (1)  $\det A = 0$ .
- (5): Treating the columns of  $I$  as the sides of a parallelepiped, we get a  $1 \times 1 \times 1 \cdots 1$  box with the sides orthogonal. Such a box clearly has volume 1, so  $|\det I| = 1$ .  $\det I = 1$  (as opposed to  $-1$ ) because the columns of  $I$  are written in "the right order" (this is one part of the more rigorous definition of determinant).
- (6): Multiplying one column of a matrix by  $r$  makes the parallelepiped " $r$  times as big" since one side of the parallelepiped is multiplied by  $r$ . Thus the determinant is also multiplied by  $r$ , since the determinant of a matrix measures (up to the sign) this volume.
- (7): This follows from (6).
- (9): ( $\Rightarrow$ ) If  $A$  is invertible, then  $\exists A^{-1}$  such that  $AA^{-1} = I$ . Then by taking determinants of both sides and applying (5) and (8), we see

$$\det A \cdot (\det A^{-1}) = 1,$$

so  $\det A \neq 0$  (and  $\det A^{-1} = \frac{1}{\det A}$ ).  $\square$

**Warning:** Many of these arguments are only sketches of ideas; to formally justify many of them one needs to define determinant more rigorously.

## 7.2 Computing determinants

### $1 \times 1$ matrices

The determinant of a  $1 \times 1$  matrix is just the number in the matrix:

$$\text{If } A = (a), \text{ then } \det A = a.$$

### $2 \times 2$ matrices

As we have seen, the determinant of a  $2 \times 2$  matrix is given by the following formula:

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$$

#### EXAMPLE 2

$$\det \begin{pmatrix} 7 & 12 \\ 3 & 5 \end{pmatrix} = 7(5) - 12(3) = \boxed{-1}.$$

### $3 \times 3$ matrices

**Theorem 7.3** Let  $A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \in M_3(\mathbb{R})$ . Then

$$\det A = aei + cdh + bfg - bdi - ceg - afh.$$

The proof of this theorem is beyond the scope of this course.

The formula in Theorem 7.3 is difficult to remember, but there is a trick called the **Rule of Sarrus**. To implement this rule, given a  $3 \times 3$  matrix, first copy the first two columns to the right of the matrix.

Then multiply along the diagonals, and add the “upper” and “lower” products. The determinant is the bottom sum minus the top sum.

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \begin{matrix} a & b \\ d & e \\ g & h \end{matrix}$$



**WARNING:** The Rule of Sarrus does not work for  $4 \times 4$  and larger matrices.

EXAMPLE 3

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

### $4 \times 4$ and larger matrices

The easiest and most efficient way to compute a determinant of a large matrix is to use technology (see the appendix of this packet). However, there is a way to compute determinants by hand called **evaluation by minors**.

**Definition 7.4** Let  $A \in M_n(\mathbb{R})$  where  $n \geq 2$  and let  $i, j \in \{1, \dots, n\}$ .

The  $(i, j)$ -**minor** of  $A$ , denoted  $A_{i,j}$ , is the  $(n - 1) \times (n - 1)$  matrix obtained by deleting the  $i^{\text{th}}$  row and  $j^{\text{th}}$  column from  $A$ .

(Recall that  $a_{i,j}$  is the  $(i, j)$ -entry of  $A$ .)

EXAMPLE 3

Let  $A = \begin{pmatrix} 1 & 3 & 5 & 7 \\ 2 & -1 & 1 & 0 \\ 0 & 3 & 1 & 2 \\ 2 & 4 & -2 & -4 \end{pmatrix}$ . Compute  $A_{1,1}$ ,  $a_{1,1}$  and  $A_{3,2}$ .

Here are formulas that show you how to compute determinants of large matrices, in terms of determinants of smaller matrices (the proofs of these theorems are beyond the scope of this class):

**Theorem 7.5 (Evaluation of determinants via minors along a row)** *Let  $A$  be an  $n \times n$  matrix. Then for any  $i \in \{1, 2, \dots, n\}$ ,*

$$\det A = \sum_{j=1}^n (-1)^{i+j} a_{i,j} \det(A_{i,j}).$$

**Theorem 7.6 (Evaluation of determinants via minors along a column)** *Let  $A \in M_n(\mathbb{R})$ . Then for any  $j \in \{1, 2, \dots, n\}$ ,*

$$\det A = \sum_{i=1}^n (-1)^{i+j} a_{i,j} \det(A_{i,j}).$$

**Note:** The values of  $(-1)^{i+j}$  should be thought of as the following:

$$\begin{pmatrix} + & - & + & \cdots \\ - & + & - & \ddots \\ + & - & \ddots & \\ - & \ddots & & \\ \vdots & & & \end{pmatrix}$$

#### EXAMPLE 4

Compute the determinant of each matrix:

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

$$\text{b) } A = \begin{pmatrix} 1 & 1 & -2 & 3 \\ 2 & 0 & 4 & -1 \\ 3 & 0 & -1 & 1 \\ 3 & 2 & 0 & 1 \end{pmatrix}$$

$$\text{c) } A = \begin{pmatrix} 0 & 0 & -1 & 0 & 3 \\ 1 & 0 & 2 & 4 & 1 \\ 0 & 1 & -3 & 0 & 2 \\ 1 & 1 & 2 & 0 & -2 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

$$\text{d) } A = \begin{pmatrix} 3 & 1 & 4 & -2 \\ 0 & 2 & 1 & 5 \\ 0 & 0 & -2 & 5 \\ 0 & 0 & 0 & 10 \end{pmatrix}$$

This example generalizes:

**Theorem 7.7** *If  $A$  is a triangular matrix, then  $\det A$  is the product of its diagonal entries.*

## Using technology to compute determinants

### Determinants on *Mathematica*:

Type in the matrix (saving it as, say  $A$ ) and then execute `Det[A]`.

### Determinants on TI-83/84 type calculators:

First type in the matrix and save it, following the directions of Chapter 2.

Then hit `[MATRX]`, go to the right to find `MATH`, then go down to `det(`. Hit `[ENTER]`, then pull up the matrix you have saved by hitting `MATRX]` and scrolling down to the matrix you want. Then execute this command; your calculator will spit out the determinant of  $A$ .

## 7.3 Determinants and cross products

RECALL FROM CHAPTER 4

If  $\mathbf{v}$  and  $\mathbf{w}$  are two vectors in  $\mathbb{R}^3$ , a vector  $\mathbf{n}$  orthogonal to both  $\mathbf{v}$  and  $\mathbf{w}$  is given by the cross product

$$\mathbf{n} = \mathbf{v} \times \mathbf{w}.$$

We learned a formula for cross products that is hard to memorize. Here is an easier method to remember:

**Theorem 7.8** Let  $\mathbf{v} = (v_1, v_2, v_3)$  and  $\mathbf{w} = (w_1, w_2, w_3)$  be two vectors in  $\mathbb{R}^3$ . Then

$$\mathbf{v} \times \mathbf{w} = \det \begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{pmatrix}.$$

PROOF Evaluating the determinant by minors across the top row, we see

$$\begin{aligned} \det \begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{pmatrix} &= \mathbf{e}_1(v_2w_3 - w_2v_3) - \mathbf{e}_2(v_1w_3 - w_1v_3) + \mathbf{e}_3(v_1w_2 - w_1v_2) \\ &= (v_2w_3 - w_2v_3, 0, 0) - (0, v_1w_3 - w_1v_3, 0) + (0, 0, v_1w_2 - w_1v_2) \\ &= (v_2w_3 - w_2v_3, w_1v_3 - v_1w_3, v_1w_2 - w_1v_2) \end{aligned}$$

which is the same formula we already defined as  $\mathbf{v} \times \mathbf{w}$ .  $\square$

**Warning:** Strictly speaking, the expression  $\begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{pmatrix}$  is gibberish, be-

cause it is a matrix with some entries that are vectors and some entries that are numbers.

But if you use the Rule of Sarrus or some other determinant formula on this gibberish, you get the correct cross product, so it's "okay" to use.

EXAMPLE 5

Compute  $(2, 3, -1) \times (1, 4, -3)$ .

EXAMPLE 6

Write a normal equation of the plane containing the points  $(1, 3, -2)$ ,  $(0, 2, 1)$  and  $(5, -1, 3)$ .

## 7.4 Homework exercises for Chapter 7

1. Find the determinant of each given matrix, if it exists (you can use a computer or calculator, but you should be able to compute this by hand):

a)

$$(-3)$$

b)

$$\begin{pmatrix} 1 & 2 \end{pmatrix}$$

c)

$$\begin{pmatrix} -5 & 2 \\ 8 & -7 \end{pmatrix}$$

d)

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

e)

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

2. Find the determinant of each given matrix, if it exists (you can use a computer or calculator, but you should be able to compute this by hand):

a)

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

c)

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

b)

$$\begin{pmatrix} -3 & 1 & 1 \\ -1 & 2 & -1 \\ 4 & 0 & -2 \\ 1 & 7 & -3 \end{pmatrix}$$

d)

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

3. Suppose that  $A$ ,  $B$  and  $C$  are  $4 \times 4$  matrices such that  $\det A = 2$ ,  $\det B = \frac{-2}{3}$  and  $\det C = 0$ .

a) Find  $\det(4A)$ .b) Find  $\det(B^{-1})$ .c) Find  $2 \det(3A)$ .d) Find  $\det(BAB)$ .e) Which one or ones of the matrices  $A$ ,  $B$  and  $C$  are invertible?f) Is the matrix  $AC$  invertible? Is the matrix  $CA$  invertible?

4. Let  $A \in M_n(\mathbb{R})$  be an orthogonal matrix (see Problem 9 of Chapter 5). What are the possible values of  $\det A$ ? Explain.
5. Suppose that  $A$  and  $B$  are  $n \times n$  square matrices.
- Prove that if  $A$  and  $B$  have rank  $n$ , then  $AB$  also has rank  $n$ .  
*Hint:* use determinants.
  - Prove that if  $AB$  has rank  $n$ , then both  $A$  and  $B$  must both have rank  $n$ .  
*Hint:* use determinants.
6. Use determinants to compute each cross product:
- $(4, 0, -3) \times (2, 1, 7)$
  - $(-1, -5, 2) \times (11, -3, 4)$



## Chapter 8

---

# Eigentheory

---

### 8.1 Motivation: a differential equation

#### RECALL

One of our motivating problems from Chapter 1 was the following differential equation, modeling the charge  $q$  in an RLC electrical circuit at time  $t$ :

$$q''(t) + 6q'(t) + 5q(t) = 2 \sin t. \quad (8.1)$$

Let's relate this to linear algebra. First, define  $C^\infty(\mathbb{R}, \mathbb{R})$  to be the set of functions from  $\mathbb{R}$  to  $\mathbb{R}$  that are **infinitely differentiable**. This set  $C^\infty(\mathbb{R}, \mathbb{R})$  forms a vector space (it is closed under  $+$  and scalar  $\cdot$ ).

Next, define  $T : C^\infty(\mathbb{R}, \mathbb{R}) \rightarrow C^\infty(\mathbb{R}, \mathbb{R})$  by

$$T(f) = f'' + 6f' + 5f, \text{ e.g.}$$

This  $T$  is a linear transformation (proof - HW).

So if we let  $b(t) = 2 \sin t$ , then equation (8.1) above is of the form

$$T(q) = b,$$

which is exactly the kind of equation we have been studying this semester.

#### QUESTION

---

What do we know about the solution of such an equation?

---

 FROM THE PREVIOUS PAGE
 

---

$$q''(t) + 6q'(t) + 5q(t) = 2 \sin t \quad (8.1)$$

$$T(q) = b$$

This equation has as its solution  $q = \ker(T) + q_p$ , where  $q_p$  is any one particular solution of the equation.

 QUESTION A
 

---

How do you find a particular solution  $q_p$ ?

Since  $T(q)$  is supposed to be a sinusoidal function, it stands to reason  $q_p$  might itself be a trig function. Let's try

$$q_p(t) = A \sin t$$

where  $A$  is a constant which I don't yet know. Then, by plugging  $A \sin t$  into the formula for  $T$ , we get

$$\begin{aligned} T(q_p) &= (A \sin t)'' + 6(A \sin t)' + 5(A \sin t) \\ &= -A \sin t + 6A \cos t + 5A \sin t \\ &= 4A \sin t + 6A \cos t \quad \leftarrow \text{this is supposed to be } b = 2 \sin t \end{aligned}$$

There is no choice of  $A$  which makes this expression equal  $b = 2 \sin t$  (since it has a cosine term in it), so this  $q_p$  can't be correct. But, if we try

$$q_p(t) = A \sin t + B \cos t,$$

then

$$\begin{aligned} T(q_p) &= (A \sin t + B \cos t)'' + 6(A \sin t + B \cos t)' + 5(A \sin t + B \cos t) \\ &= -A \sin t - B \cos t + 6A \cos t - 6B \sin t + 5A \sin t + 4B \cos t \\ &= (4A - 6B) \sin t + (6A + 3B) \cos t \quad \leftarrow \text{supposed to be } b = 2 \sin t \end{aligned}$$

This will work out to be  $b = 2 \sin t = 2 \sin t + 0 \cos t$  exactly if

$$\begin{cases} 4A - 6B = 2 \\ 6A + 3B = 0 \end{cases} \Rightarrow (A, B) = \left(\frac{1}{8}, \frac{-1}{4}\right).$$

Thus a particular solution  $q_p$  to our original equation  $T(q) = b$  is

$$q_p(t) = \frac{1}{8} \sin t - \frac{1}{4} \cos t.$$

This finishes Question A, and tells us that the solution set of the differential equation

$$q''(t) + 6q'(t) + 5q(t) = 2 \sin t \tag{8.1}$$

has the form

$$\frac{1}{8} \sin t - \frac{1}{4} \cos t + \ker(T)$$

where  $T(q) = q'' + 6q' + 5q$ .

#### QUESTION B

How do you find the kernel of  $T$ , where  $T(q) = q'' + 6q' + 5q$ ?

**Simpler motivating example:** Let  $U : C^\infty(\mathbb{R}, \mathbb{R}) \rightarrow C^\infty(\mathbb{R}, \mathbb{R})$  be defined by  $U(f) = f' - \lambda f$ , where  $\lambda \in \mathbb{R}$  is a constant. What is  $\ker(U)$ ?

$$\begin{aligned} y = f(x) \in \ker(U) &\iff U(f) = 0 \\ &\iff \frac{dy}{dx} - \lambda y = 0 \end{aligned}$$

**Theorem 8.1** Let  $\lambda \in \mathbb{R}$  be a constant, and define  $U : C^\infty(\mathbb{R}, \mathbb{R}) \rightarrow C^\infty(\mathbb{R}, \mathbb{R})$  by  $U(f) = f' - \lambda f$ . Then

$$\ker(U) = \text{Span}(e^{\lambda t}) = \{C e^{\lambda t} : C \in \mathbb{R}\}.$$

There is a way to rephrase this simpler example that is **very** useful.

Define  $D : C^\infty(\mathbb{R}, \mathbb{R}) \rightarrow C^\infty(\mathbb{R}, \mathbb{R})$  by  $D(f) = f'$  (“ $D$ ” stands for derivative). This  $D$  is a linear transformation. Any multiple of the function  $f(t) = e^{\lambda t}$  has a special property:

In other words, **applying the linear transformation  $D$  to  $f$  is the same as multiplying  $f$  by a constant.**

Given any linear transformation (not just differentiation), identifying the vectors for which the linear transformation is the same as multiplying those vectors by a constant is key to understanding important behavior of that linear transformation.

## 8.2 Eigenvalues and eigenvectors

**Definition 8.2** Let  $V$  be a vector space, and let  $T : V \rightarrow V$  be a linear transformation. If  $\mathbf{v} \in V$  is a **nonzero vector** and  $\lambda \in \mathbb{R}$  is a **constant** such that

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

then we say  $\lambda$  is an **eigenvalue** of  $T$  and that  $\mathbf{v}$  is an **eigenvector** of  $T$  corresponding to  $\lambda$ .

In the situation where  $T$  is given by matrix multiplication by  $A$  (i.e.  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is given by  $T(\mathbf{x}) = A\mathbf{x}$  for an  $n \times n$  matrix  $A$ ), we say  $\lambda$  is an **eigenvalue** of  $A$  and  $\mathbf{v}$  is an **eigenvector** of  $A$ .

If the elements of  $V$  are functions, the word **eigenfunction** can be used interchangeably with **eigenvector**.

The **spectrum** of a linear transformation is the set of its eigenvalues.

**Theorem 8.3 (Basic properties of eigenvalues and eigenvectors)** *Let  $V$  be a vector space, and let  $T : V \rightarrow V$  be a linear transformation.*

1. *Let  $\lambda$  be any eigenvalue of  $T$ . The set of eigenvectors corresponding to  $\lambda$ , together with the zero vector, forms a subspace of  $V$ . In particular:*
  - a) *If  $\mathbf{v}, \mathbf{w} \in V$  are both eigenvectors corresponding to  $\lambda$ , so is  $\mathbf{v} + \mathbf{w}$ ;*
  - b) *If  $\mathbf{v} \in V$  is an eigenvector corresponding to  $\lambda$ , so is  $c\mathbf{v}$  for any  $c \in \mathbb{R}$ .*
2. *If  $\lambda_1, \lambda_2, \dots, \lambda_n$  are all **different** eigenvalues of  $T$  with respective eigenvectors  $\mathbf{v}_1, \dots, \mathbf{v}_n$ , then  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  are linearly independent.*
3. *The maximum number of eigenvalues of  $T$  is  $\dim V$ .*

PROOF The proof of (1) is a homework exercise.

For (2), notice that for all  $j \in \{1, \dots, n\}$  we have

$$T(\mathbf{v}_j) = \lambda_j \mathbf{v}_j.$$

To show  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  are lin. ind., suppose not, i.e. that they are lin. dep.

Then let  $\mathbf{v}_k$  be the **first** dependent vector.

That means  $\{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}\}$  are lin. ind. and  $\exists c_1, \dots, c_{k-1}$  such that

$$\mathbf{v}_k = c_1 \mathbf{v}_1 + \dots + c_{k-1} \mathbf{v}_{k-1}. \quad (8.2)$$

Apply  $T$  to both sides of equation (8.2) to get

$$T(\mathbf{v}_k) = T(c_1 \mathbf{v}_1 + \dots + c_{k-1} \mathbf{v}_{k-1})$$

$$T(\mathbf{v}_k) = c_1 T(\mathbf{v}_1) + \dots + c_{k-1} T(\mathbf{v}_{k-1})$$

(since  $T$  is linear)

$$\lambda_k \mathbf{v}_k = c_1 \lambda_1 \mathbf{v}_1 + \dots + c_{k-1} \lambda_{k-1} \mathbf{v}_{k-1}$$

(since  $\mathbf{v}_j$ s are eigenvectors)

$$\lambda_k (c_1 \mathbf{v}_1 + \dots + c_{k-1} \mathbf{v}_{k-1}) = c_1 \lambda_1 \mathbf{v}_1 + \dots + c_{k-1} \lambda_{k-1} \mathbf{v}_{k-1}$$

(by the dependence relation (8.2))

$$c_1 \lambda_k \mathbf{v}_1 + \dots + c_{k-1} \lambda_k \mathbf{v}_{k-1} = c_1 \lambda_1 \mathbf{v}_1 + \dots + c_{k-1} \lambda_{k-1} \mathbf{v}_{k-1}$$

$$\mathbf{0} = c_1 (\lambda_1 - \lambda_k) \mathbf{v}_1 + \dots + c_{k-1} (\lambda_{k-1} - \lambda_k) \mathbf{v}_{k-1}.$$

Thus  $\mathbf{v}_1, \dots, \mathbf{v}_{k-1}$  are lin. dep., contradicting the hypothesis.

Therefore  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  are lin. ind.

Finally, statement (3) follows from (2) by applying the Exchange Lemma.  $\square$

## EXAMPLE 1

Let  $D : C^\infty(\mathbb{R}, \mathbb{R}) \rightarrow C^\infty(\mathbb{R}, \mathbb{R})$  be  $D(f) = f'$ . Then, every  $\lambda \in \mathbb{R}$  is an eigenvalue, with corresponding eigenvector

EXAMPLE 2

---

Let  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be  $T(\mathbf{x}) = A\mathbf{x}$ , where  $A = \begin{pmatrix} 4 & -7 \\ -2 & -9 \end{pmatrix}$ . Then:

$$T(1, 2) = \begin{pmatrix} 4 & -7 \\ -2 & -9 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} =$$

$$T(7, -1) = \begin{pmatrix} 4 & -7 \\ -2 & -9 \end{pmatrix} \begin{pmatrix} 7 \\ -1 \end{pmatrix} =$$

RETURN TO QUESTION B FROM EARLIER

---

$T(q) = q'' + 6q' + 5q$ ;  $\ker(T) = ?$

As above, let  $D : C^\infty(\mathbb{R}, \mathbb{R}) \rightarrow C^\infty(\mathbb{R}, \mathbb{R})$  be  $D(q) = q'$ . Then notice

So if  $q$  is an eigenvector of  $D$  with eigenvalue  $\lambda$  (i.e.  $q(t) = ce^{\lambda t}$ ), we have

$$T(q) =$$

## CONCLUSION

if  $\lambda^2 + 6\lambda + 5 = 0$ , then any eigenvector of  $D$  corresponding to  $\lambda$  (those eigenvectors are functions of the form  $ce^{\lambda t}$ ) is in the kernel of  $T$ .

Notice  $\{e^{-5t}, e^{-t}\}$  are linearly independent, which means

$$\text{Span}(e^{-t}, e^{-5t}) \subseteq \ker(T), \text{ implying } \dim(\ker(T)) \geq 2.$$

## LAST QUESTION

Is there anything else in  $\ker(T)$ , other than  $\text{Span}(e^{-5t}, e^{-t})$ ?

To determine this, let  $U_1, U_5 : C^\infty(\mathbb{R}, \mathbb{R}) \rightarrow C^\infty(\mathbb{R}, \mathbb{R})$  be  $U_1(q) = q' + q$  and  $U_5(q) = q' + 5q$ . We know (from the “simpler motivating example” in Section 8.1) that

- $\ker(U_1) =$
- $\ker(U_5) =$

Now let  $f(t) = \frac{-1}{4}e^{-5t}$ . Observe

$$U_1(f) = f' + f = \frac{5}{4}e^{-5t} - \frac{1}{4}e^{-5t} = e^{-5t},$$

so  $e^{-5t} \in \text{im}(U_1)$ , meaning  $\text{Span}\left(\frac{-1}{4}e^{-5t}\right) = \text{Span}(e^{-5t}) = \ker(U_5) \subseteq \text{im}(U_1)$ .

Therefore, we can apply the result of Homework Exercise 34 of Chapter 6 to conclude

$$\dim(\ker(U_5 \circ U_1)) = \dim(\ker(U_5)) + \dim(\ker(U_1)) = 1 + 1 = 2.$$

**But**, notice that

$$(U_5 \circ U_1)(q) = U_5(U_1(q)) =$$

From our previous work, we have the following facts (for  $T(q) = q'' + 6q' + 5q$ ):

- $\{e^{-5t}, e^{-t}\} \subseteq \ker T$ ;
- $T = U_5 \circ U_1$  (where  $U_\lambda(f) = f' + \lambda f$ );
- $\ker(T) = \ker(U_5 \circ U_1)$ ;
- $\dim \ker(U_5 \circ U_1) = 2$ .

We can conclude:

and so we have finally answered Question B: if  $T(q) = q'' + 6q' + 5q$ , then

$$\ker(T) = \text{Span}(e^{-5t}, e^{-t}) = \{C_1 e^{-5t} + C_2 e^{-t} : C_1, C_2 \in \mathbb{R}\}.$$

FINALLY!

---

After several pages of work, we have determined that the solution of

$$q'' + 6q' + 5q = 2 \sin t \tag{8.1}$$

is

$$\begin{aligned} q(t) &= \ker(T) + q_p \\ &= \text{Span}(e^{-5t}, e^{-t}) + \frac{1}{8} \sin t - \frac{1}{4} \cos t \\ &= \left\{ C_1 e^{-5t} + C_2 e^{-t} + \frac{1}{8} \sin t - \frac{1}{4} \cos t : C_1, C_2 \in \mathbb{R} \right\}. \end{aligned}$$

All this machinery generalizes as indicated on the next page (although to prove this rigorously, you need complex numbers... take MATH 330 to see proofs):



**Theorem 8.4** Let  $T : C^\infty(\mathbb{R}, \mathbb{R}) \rightarrow C^\infty(\mathbb{R}, \mathbb{R})$  be

$$T(f) = c_n f^{(n)} + c_{n-1} f^{(n-1)} + \dots + c_1 f' + c_0 f$$

where  $c_0, c_1, \dots, c_n$  are all constants. (Such a  $T$  is called an  $n^{\text{th}}$ -order constant-coefficient, linear differential operator.) Then:

1.  $\dim(\ker(T)) = n$ .
2. For any  $\lambda \in \mathbb{R}$  which is a root of the **characteristic equation** of  $T$ , which is

$$c_n \lambda^n + c_{n-1} \lambda^{n-1} + \dots + c_1 \lambda + c_0 = 0,$$

we have  $e^{\lambda t} \in \ker(T)$ .

**EXAMPLE 1**

Find all functions  $y(t)$  which satisfy the differential equation

$$y''(t) - 4y'(t) - 21y(t) = -9e^t.$$

## 8.3 Eigentheory of matrices

## RECALL

We have seen that understanding the eigenvalues and eigenvectors of  $D(f) = f'$  helped us understand solutions to  $T(f) = b$  where  $T(f) = f'' + 6f' + 5f = D \circ D(f) + 6D(f) + 5f$ .

More generally, understanding the eigenvalues and eigenvectors of  $D(f) = f'$  gives us information about any transformation  $T : C^\infty(\mathbb{R}, \mathbb{R}) \rightarrow C^\infty(\mathbb{R}, \mathbb{R})$  of the form

$$T(f) = c_n D^n(f) + c_{n-1} D^{n-1}(f) + \dots + c_1 D(f) + c_0 f$$

where  $c_0, \dots, c_n$  are constants and by  $D^r$ , we mean  $D \circ D \circ D \circ \dots \circ D$  ( $r$  times).

## NEW SETTING

Consider the linear transformation  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  given by  $T(\mathbf{x}) = A\mathbf{x}$  for some square matrix  $A \in M_n(\mathbb{R})$ . To study the transformation  $T^r = T \circ T \circ T \circ \dots \circ T$ , which is given by matrix  $A^r$ , it therefore behooves us to study the eigenvalues and eigenvectors of  $A$ .

**Who cares?** Suppose that at each time  $n \in \{0, 1, 2, \dots\}$ , you are measuring regularly two quantities  $x_n$  and  $y_n$  coming from some physical or biological process, or from economics or actuarial science. For example, let  $x_n$  and  $y_n$  be the number of male and female chipmunks in a certain region at time  $n$ . Initially (when  $n = 0$ ), there are  $x_0 = 4$  and  $y_0 = 5$  chipmunks of each sex. Suppose that

$$\begin{cases} x_{n+1} = 2x_n + 3y_n \\ y_{n+1} = x_n + y_n \end{cases}$$

This problem can be phrased in linear algebra language. For each  $n$ , let  $\mathbf{x}_n = \begin{pmatrix} x_n \\ y_n \end{pmatrix}$ . Our equations above can then be rewritten as

That means that, say at time 73, the number of chipmunks of each sex satisfies

Now for  $A = \begin{pmatrix} 2 & 3 \\ 1 & 1 \end{pmatrix}$ , we see (by direct calculation) that

$$A^2 = AA = \begin{pmatrix} 7 & 9 \\ 3 & 4 \end{pmatrix} \quad A^3 = A^2A = \begin{pmatrix} 23 & 30 \\ 10 & 13 \end{pmatrix} \quad A^4 = \begin{pmatrix} 76 & 99 \\ 33 & 43 \end{pmatrix} \quad \text{etc.}$$

The entries of these powers of  $A$  don't seem to have anything to do with  $A$ . And if you were asked to compute  $A^{73}$ , for now, you'd be screwed (other than that the numbers would be really large).

## Diagonalization

### EXAMPLE 2

Let  $\Lambda = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 3 \end{pmatrix}$ . ( $\Lambda$  is the Greek capital letter lambda.) Then

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

$$\Lambda^3 = \Lambda^2\Lambda = \begin{pmatrix} 4 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 9 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

$$\Lambda^r =$$

Example 2 shows that it is easy to compute the powers of a diagonal matrix:

**Theorem 8.5** Let  $\Lambda = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$  be a diagonal  $n \times n$  matrix.

Then for any  $r \in \{0, 1, 2, \dots\}$ ,  $\Lambda^r = \begin{pmatrix} \lambda_1^r & & & \\ & \lambda_2^r & & \\ & & \ddots & \\ & & & \lambda_n^r \end{pmatrix}$ .

Now, suppose that given a square matrix  $A \in M_n(\mathbb{R})$ , you can (somehow) find a diagonal matrix  $\Lambda \in M_n(\mathbb{R})$  and an invertible matrix  $S \in M_n(\mathbb{R})$  such that

$$A = S\Lambda S^{-1}.$$

Then, we can compute powers of  $A$  as follows:

$$A^r = (S\Lambda S^{-1})^r =$$

This idea motivates the following definition and theorem:

**Definition 8.6** A square  $n \times n$  matrix  $A$  is called **diagonalizable** (a.k.a. **similar to a diagonal matrix**) if there is an **invertible** matrix  $S \in M_n(\mathbb{R})$  and a **diagonal** matrix  $\Lambda \in M_n(\mathbb{R})$  such that

$$A = S\Lambda S^{-1}.$$

Writing  $A$  as  $S\Lambda S^{-1}$  is called **diagonalizing**  $A$ .

**Theorem 8.7** If  $A \in M_n(\mathbb{R})$  is diagonalizable, then for any  $r \in \{0, 1, 2, \dots\}$ ,

$$A^r = S\Lambda^r S^{-1}.$$

Therefore, to compute the powers of a diagonalizable matrix  $A$ , we could to find the  $S$  and the  $\Lambda$  and apply the formula in the preceding theorem.

QUESTION

Given  $A \in M_n(\mathbb{R})$ , is  $A$  diagonalizable? If so, what are  $S$  and  $\Lambda$ ?

To answer this question theoretically, we first assume we can diagonalize  $A$ :

$$S\Lambda S^{-1} = A$$

Multiply both sides of this equation by  $S$  on the right to obtain

$$S\Lambda = AS. \tag{8.3}$$

Write the columns of  $S$  as  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  so that

$$S = \begin{pmatrix} \uparrow & \uparrow & \cdots & \uparrow \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \\ \downarrow & \downarrow & \cdots & \downarrow \end{pmatrix}.$$

Since  $S$  is invertible,  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  must be lin. ind.

On the previous page we found  $S\Lambda = AS$ . The left-hand side of this is

$$\begin{aligned} S\Lambda &= \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \cdots & s_{nn} \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1 s_{11} & \lambda_2 s_{12} & \cdots & \lambda_n s_{1n} \\ \lambda_1 s_{21} & \lambda_2 s_{22} & \cdots & \lambda_n s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1 s_{n1} & \lambda_2 s_{n2} & \cdots & \lambda_n s_{nn} \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1 \mathbf{x}_1 & \lambda_2 \mathbf{x}_2 & \cdots & \lambda_n \mathbf{x}_n \end{pmatrix}. \end{aligned}$$

The right-hand side of  $S\Lambda = AS$  is

$$\begin{aligned} AS &= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \cdots & s_{nn} \end{pmatrix} \\ &= \begin{pmatrix} a_{11}s_{11} + a_{12}s_{21} + \cdots + a_{1n}s_{n1} & \cdots & \cdots & a_{11}s_{1n} + a_{12}s_{2n} + \cdots + a_{1n}s_{nn} \\ a_{21}s_{11} + a_{22}s_{21} + \cdots + a_{2n}s_{n1} & & & a_{21}s_{1n} + a_{22}s_{2n} + \cdots + a_{2n}s_{nn} \\ \vdots & & & \vdots \\ a_{n1}s_{11} + a_{n2}s_{21} + \cdots + a_{nn}s_{n1} & \cdots & \cdots & a_{n1}s_{1n} + a_{n2}s_{2n} + \cdots + a_{nn}s_{nn} \end{pmatrix} \\ &= \begin{pmatrix} A\mathbf{x}_1 & A\mathbf{x}_2 & \cdots & A\mathbf{x}_n \end{pmatrix}. \end{aligned}$$

Equating the columns of the left- and right-hand sides of  $S\Lambda = AS$ , we see that

$$\lambda_j \mathbf{x}_j = A\mathbf{x}_j \tag{8.4}$$

for all  $j$ . In other words:

We have proven:

**Theorem 8.8 (Diagonalization Theorem)** Let  $A \in M_n(\mathbb{R})$  be a square matrix.

1.  $A$  is diagonalizable  $\iff A$  has  $n$  linearly independent eigenvectors.
2. If  $A$  is diagonalizable, and  $A = SAS^{-1}$  where  $\Lambda$  is diagonal, then

- the diagonal entries of  $\Lambda$  are the eigenvalues of  $A$ , and
- the columns of  $S$  are the corresponding eigenvectors (written in the same order as the eigenvalues are written in  $\Lambda$ ).

**Theorem 8.9** Let  $A \in M_n(\mathbb{R})$  be a square matrix. If  $A$  has  $n$  different eigenvalues, then  $A$  is diagonalizable.

PROOF If  $A$  has  $n$  different eigenvalues, then  $A$  has  $n$  linearly independent eigenvectors by statement (2) of Theorem 8.3.  $A$  is therefore diagonalizable by the preceding theorem.  $\square$

## 8.4 Computing eigenvalues and eigenvectors

### QUESTION

Given square matrix  $A$ , how do you compute eigenvalues and eigenvectors of  $A$ ?

### Answer # 1: use technology

If the size of a matrix is large (larger than  $3 \times 3$ , usually use a computer. In *Mathematica*, you can find the eigenvalues and eigenvectors of a matrix  $A$  by typing in the matrix and saving it as  $A$ , then executing

`Eigensystem[A]`

For example, suppose  $A = \begin{pmatrix} 2 & -5 \\ -1 & 2 \end{pmatrix}$ . After typing in this matrix, the command `Eigensystem[A]` produces the output

`{{2 +  $\sqrt{5}$ , 2 -  $\sqrt{5}$ }, {{- $\sqrt{5}$ , 1}, { $\sqrt{5}$ , 1}}}`

This means the eigenvalues of  $A$  are  $\lambda_1 = 2 + \sqrt{5}$  and  $\lambda_2 = 2 - \sqrt{5}$ , and the corresponding eigenvectors are  $\mathbf{v}_1 = (-\sqrt{5}, 1)$  and  $\mathbf{v}_2 = (\sqrt{5}, 1)$ , i.e.  $A = S\Lambda S^{-1}$  where

$$S = \begin{pmatrix} -\sqrt{5} & \sqrt{5} \\ 1 & 1 \end{pmatrix} \quad \Lambda = \begin{pmatrix} 2 + \sqrt{5} & 0 \\ 0 & 2 - \sqrt{5} \end{pmatrix}$$

**Answer # 2: factor the characteristic polynomial**

Suppose  $\lambda_j$  is an eigenvalue of  $A$ . Then  $A\mathbf{v}_j = \lambda_j\mathbf{v}_j$  for some nonzero vector  $\mathbf{v}_j$ . Rewriting this, we get

$$\mathbf{0} = A\mathbf{v}_j - \lambda_j\mathbf{v}_j =$$

Thus  $\mathbf{v}_j \in N(A - \lambda_j I)$ . Since  $\mathbf{v}_j \neq \mathbf{0}$ , that means  $N(A - \lambda_j I) \neq \{\mathbf{0}\}$  so  $A - \lambda_j I$  is not invertible. That means  $\det(A - \lambda_j I) = 0$ . We have shown:

**Definition 8.10** Let  $A \in M_n(\mathbb{R})$ . The **characteristic polynomial** of  $A$  is the expression

$$p_A(x) = \det(A - xI).$$

(This  $p_A$  is a polynomial of degree  $n$ .)

**Theorem 8.11** Let  $A \in M_n(\mathbb{R})$ .  $\lambda$  is an eigenvalue of  $A$  if and only if it is a root of the characteristic polynomial of  $A$  (i.e.  $\det(A - \lambda I) = 0$ ).

**Note:** The *characteristic equation* of a differential equation was defined earlier. That's actually the same polynomial as this, but you first have to know how to turn a differential equation into a matrix (which I teach you how to do in MATH 330).

## The multiplicity of an eigenvalue

**Remark:** Suppose  $A$  is some matrix such that  $p_A(x)$  factors as

$$p_A(x) = (x + 1)(x - 2)(x - 3)^2(x - 6)^4.$$

Then (for example),  $\lambda = 3$  is an eigenvalue of  $A$ . Since 3 is a “double root” of  $p_A(x)$ , we say  $\lambda = 3$  is an **eigenvalue of multiplicity 2**. Similarly, for this matrix  $A$ , its eigenvalues are

The right way to list the eigenvalues of a matrix is to repeat them according to their multiplicities, i.e. the eigenvalues of the  $A$  with the characteristic polynomial as above are

**Theorem 8.12** *Assume the eigenvalues of a matrix  $A$  are listed according to their multiplicities. Then:*

1. *The sum of the eigenvalues is  $\text{tr}(A)$ .*
2. *The product of the eigenvalues is  $\det(A)$ .*

The proof of this theorem is beyond the scope of this course.

### EXAMPLE 3

For the  $A$  whose characteristic polynomial is

$$p_A(x) = (x + 1)(x - 2)(x - 3)^2(x - 6)^4,$$

we have

$$\text{tr}(A) =$$

and

$$\det(A) = \boxed{-23328}.$$



## 8.4. Computing eigenvalues and eigenvectors

---

### EXAMPLE 4

---

Compute eigenvalues and eigenvectors for the matrix  $A = \begin{pmatrix} 13 & -4 \\ -4 & 7 \end{pmatrix}$ .

**Putting all this together****EXAMPLE 5**

Compute  $A^{45}$ , if  $A = \begin{pmatrix} 5 & 8 & 16 \\ 4 & 1 & 8 \\ -4 & -4 & -11 \end{pmatrix}$ .

**EIGENVALUES:**

$$\begin{aligned}
 p_A(\lambda) = \det(A - \lambda I) &= \det \begin{pmatrix} 5 - \lambda & 8 & 16 \\ 4 & 1 - \lambda & 8 \\ -4 & -4 & -11 - \lambda \end{pmatrix} = \begin{vmatrix} 5 - \lambda & 8 \\ 4 & 1 - \lambda \\ -4 & -4 \end{vmatrix} \\
 &= (5 - \lambda)(1 - \lambda)(-11 - \lambda) - 512 + 64(1 - \lambda) + 32(5 - \lambda) - 32(-11 - \lambda) \\
 &= (5 - \lambda)(1 - \lambda)(-11 - \lambda) - 512 + 64(1 - \lambda) + 512 \\
 &= (1 - \lambda)[(5 - \lambda)(-11 - \lambda) + 64] \\
 &= (1 - \lambda)(-55 + 6\lambda + \lambda^2 + 64) \\
 &= (1 - \lambda)(\lambda^2 + 6\lambda + 9) \\
 &= (1 - \lambda)(\lambda + 3)^2
 \end{aligned}$$

**EIGENVECTORS:** First, an eigenvector to go with  $\lambda = 1$ . Write the eigenvector  $\mathbf{v}$  as  $(x, y, z)$ . Then  $A\mathbf{v} = 1\mathbf{v}$  gives

$$\begin{cases} 5x + 8y + 16z = 1x \\ 4x + y + 8z = 1y \\ -4x - 4y - 11z = 1z \end{cases} \Rightarrow \begin{cases} x + 2y + 4z = 0 \\ x + 2z = 0 \\ x + y + 3z = 0 \end{cases} \Rightarrow \begin{cases} x = -2z \\ 2y + 2z = 0 \\ y = -z \\ x = -2z \end{cases}$$

Now for  $\lambda = -3$ :  $A\mathbf{v} = -3\mathbf{v}$  gives

$$\begin{cases} 5x + 8y + 16z = -3x \\ 4x + y + 8z = -3y \\ -4x - 4y - 11z = -3z \end{cases} \Rightarrow \begin{cases} x + y + 2z = 0 \\ x + y + 2z = 0 \\ x + y + 2z = 0 \end{cases}$$

So eigenvalues and eigenvectors of  $A$  are

$$\lambda = 1 \leftrightarrow (-2, -1, 1) \quad \lambda = -3 \leftrightarrow (2, 0, -1) \quad \lambda = -3 \leftrightarrow (0, 2, -1).$$

**DIAGONALIZATION:**  $A = SAS^{-1}$  where

$$S = \begin{pmatrix} -2 & 2 & 0 \\ -1 & 0 & 2 \\ 1 & -1 & -1 \end{pmatrix} \quad \Lambda = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -3 \end{pmatrix} \quad S^{-1} = \frac{-1}{2} \begin{pmatrix} 2 & 2 & 4 \\ 1 & 2 & 4 \\ 1 & 0 & 2 \end{pmatrix}$$

**MATRIX POWER:**

$$\begin{aligned}
 A^{45} &= SA^{45}S^{-1} = \begin{pmatrix} -2 & 2 & 0 \\ -1 & 0 & 2 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & (-3)^{45} & 0 \\ 0 & 0 & (-3)^{45} \end{pmatrix} \frac{-1}{2} \begin{pmatrix} 2 & 2 & 4 \\ 1 & 2 & 4 \\ 1 & 0 & 2 \end{pmatrix} \\
 &= \begin{pmatrix} 2 + 3^{45} & 2 + 2 \cdot 3^{45} & 4 + 4 \cdot 3^{45} \\ 1 + 3^{45} & 1 & 2 + 2 \cdot 3^{45} \\ -1 - 3^{45} & -1 - 3^{45} & -2 - 3 \cdot 3^{45} \end{pmatrix}.
 \end{aligned}$$

**EXAMPLE 6**

Compute the eigenvalues of the matrix

$$A = \begin{pmatrix} 3 & 0 & 1 & 4 \\ 0 & -5 & 1 & 3 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 7 \end{pmatrix}.$$

*Solution:*

$$\begin{aligned}
 p_A(\lambda) &= \det(A - \lambda I) = \det \begin{pmatrix} 3 - \lambda & 0 & 1 & 4 \\ 0 & -5 - \lambda & 1 & 3 \\ 0 & 0 & 2 - \lambda & 2 \\ 0 & 0 & 0 & 7 - \lambda \end{pmatrix} \\
 &= (3 - \lambda)(-5 - \lambda)(2 - \lambda)(7 - \lambda)
 \end{aligned}$$

This example generalizes:

**Theorem 8.13** *The eigenvalues of a triangular matrix are its diagonal entries.*

EXAMPLE 7

Find eigenvalues and eigenvectors for the matrix

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

*EIGENVALUES:*

*EIGENVECTORS:*

## 8.4. Computing eigenvalues and eigenvectors

---

### EXAMPLE 8

The Fibonacci numbers  $\{f_n\}$  are defined by setting  $f_0 = 0$ ,  $f_1 = 1$  and setting  $f_n = f_{n-2} + f_{n-1}$  for  $n \geq 2$ . Find a formula for the 200<sup>th</sup> Fibonacci number.

All together,

$$f_{200} = -\frac{1}{\sqrt{5}} (-\sigma^{200} + \tau^{200}) = \frac{\sigma^{200} - \tau^{200}}{\sqrt{5}} = \boxed{\frac{1}{\sqrt{5}} \left[ \left( \frac{1 + \sqrt{5}}{2} \right)^{200} - \left( \frac{1 - \sqrt{5}}{2} \right)^{200} \right]}.$$

## 8.5 Matrix exponentials

## RECALL

In Calculus 2, you learn the following formula for the exponential function  $e^x$ :

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{1}{2}x^2 + \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + \dots$$

This formula can be used as the definition of the exponential of a square matrix  $A$ :

**Definition 8.14** Let  $A \in M_n(\mathbb{R})$  be a square matrix. The **exponential** of  $A$ , denoted  $\exp(A)$  or  $e^A$ , is the  $n \times n$  matrix

$$e^A = \exp(A) = \sum_{n=0}^{\infty} \frac{1}{n!} A^n = I + A + \frac{1}{2}A^2 + \frac{1}{3!}A^3 + \frac{1}{4!}A^4 + \dots$$

One can show that this series “converges” for every matrix  $A$  (but that is beyond the scope of MATH 322).

**Theorem 8.15 (Properties of matrix exponentials)** 1. If  $\Lambda$  is diagonal with diagonal entries  $\lambda_1, \lambda_2, \dots, \lambda_n$ , then

$$\exp(\Lambda) = \begin{pmatrix} e^{\lambda_1} & 0 & \dots & 0 \\ 0 & e^{\lambda_2} & 0 & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & \dots & 0 & e^{\lambda_n} \end{pmatrix}.$$

2. If  $A = S\Lambda S^{-1}$ , then  $\exp(A) = S \exp(\Lambda) S^{-1}$ .

3. If  $t$  is a scalar, then  $\exp(tA) = \exp(At)$ .

PROOF (1) If  $\Lambda = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$ , then  $\Lambda^r = \begin{pmatrix} \lambda_1^r & & & \\ & \lambda_2^r & & \\ & & \ddots & \\ & & & \lambda_n^r \end{pmatrix}$ .

Therefore

(2) Recall that if  $A = S\Lambda S^{-1}$ , then  $A^n = S\Lambda^n S^{-1}$  for all  $n$ . Therefore

$$\begin{aligned} e^A &= e^{S\Lambda S^{-1}} \\ &= \sum_{n=0}^{\infty} \frac{(S\Lambda S^{-1})^n}{n!} \end{aligned}$$

(3) Follows from the fact that  $tA = At$ .  $\square$

**WARNING:**  $e^{A+B} \neq e^A e^B$  in general (it is true  $e^{A+B} = e^A e^B$  if  $AB = BA$ ).

Theorem 8.15 suggests that to compute the exponential of a matrix, you should diagonalize it (similar to how you would compute a power of a matrix).

#### EXAMPLE 9

Find  $\exp(A)$  if

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

*EIGENVALUES AND EIGENVECTORS: (work omitted)*

$$\det(A - xI) = (5 - x)(3 - x)(-1 - x) \Rightarrow \lambda = 5, \lambda = 3, \lambda = -1$$

$$\lambda = 5 \leftrightarrow (1, -1, 1)$$

$$\lambda = 3 \leftrightarrow (1, 1, 0)$$

$$\lambda = -1 \leftrightarrow (1, -1, -2)$$

*DIAGONALIZATION:*

$$A = S\Lambda S^{-1} = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & -1 \\ 1 & 0 & -2 \end{pmatrix} \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & -1 \\ 1 & 0 & -2 \end{pmatrix}^{-1}$$

**MATRIX EXPONENTIAL:**

$$\begin{aligned} \exp(A) &= S \exp(\Lambda) S^{-1} \\ &= \begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & -1 \\ 1 & 0 & -2 \end{pmatrix} \begin{pmatrix} e^5 & 0 & 0 \\ 0 & e^3 & 0 \\ 0 & 0 & e^{-1} \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & -1 \\ 1 & 0 & -2 \end{pmatrix}^{-1} \\ &= \begin{pmatrix} \frac{-1}{2}e^{-1} + \frac{1}{2}e^3 + e^5 & \frac{1}{2}e^{-1} + \frac{1}{2}e^3 - e^5 & e^{-1} - e^5 \\ \frac{1}{2}e^{-1} + \frac{1}{2}e^3 - e^5 & \frac{-1}{2}e^{-1} + \frac{1}{2}e^3 + e^5 & -e^{-1} + e^5 \\ -e^{-1} + e^5 & e^{-1} - e^5 & 2e^{-1} - e^5 \end{pmatrix}. \end{aligned}$$

**SAME EXAMPLE, SLIGHTLY MODIFIED**

If you were asked to compute  $\exp(At)$ , you would start with the same eigenvalues and eigenvectors as above. Then the diagonalization would be

$$At = S(\Lambda t)S^{-1} = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & -1 \\ 1 & 0 & -2 \end{pmatrix} \begin{pmatrix} 5t & 0 & 0 \\ 0 & 3t & 0 \\ 0 & 0 & -t \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & -1 \\ 1 & 0 & -2 \end{pmatrix}^{-1}$$

and the matrix exponential would be

$$\begin{aligned} \exp(At) &= S \exp(\Lambda t) S^{-1} \\ &= \begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & -1 \\ 1 & 0 & -2 \end{pmatrix} \begin{pmatrix} e^{5t} & 0 & 0 \\ 0 & e^{3t} & 0 \\ 0 & 0 & e^{-t} \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & -1 \\ 1 & 0 & -2 \end{pmatrix}^{-1} \\ &= \begin{pmatrix} \frac{-1}{2}e^{-t} + \frac{1}{2}e^{3t} + e^{5t} & \frac{1}{2}e^{-t} + \frac{1}{2}e^{3t} - e^{5t} & e^{-t} - e^{5t} \\ \frac{1}{2}e^{-t} + \frac{1}{2}e^{3t} - e^{5t} & \frac{-1}{2}e^{-t} + \frac{1}{2}e^{3t} + e^{5t} & -e^{-t} + e^{5t} \\ -e^{-t} + e^{5t} & e^{-t} - e^{5t} & 2e^{-1} - e^{5t} \end{pmatrix}. \end{aligned}$$

**QUESTION**

Why are we interested in computing the exponential of a matrix?

*Answer:*

**EXAMPLE (FROM EARLIER IN THIS CHAPTER)**

Suppose  $y$  is a function of  $t$  which satisfies  $y'(t) = \lambda y(t)$  where  $\lambda$  is a constant. Then  $y(t)$  is an eigenvalue of the transformation  $D(y) = y'$  corresponding to  $\lambda$ , so by repeating the argument we did in the first section of this chapter, we know



## EXAMPLE 10

---

Suppose  $y_1$  and  $y_2$  are functions of  $t$  such that

$$(*) \begin{cases} y_1'(t) = a_{11}y_1(t) + a_{12}y_2(t) \\ y_2'(t) = a_{21}y_1(t) + a_{22}y_2(t) \end{cases} .$$

for constants  $a_{11}$ ,  $a_{12}$ ,  $a_{21}$  and  $a_{22}$ . Describe a method for finding  $y_1(t)$  and  $y_2(t)$ .

## EXAMPLE 11

---

Suppose  $y_1$  and  $y_2$  are functions of  $t$  such that

$$\begin{cases} y_1'(t) = 2y_1(t) - 4y_2(t) \\ y_2'(t) = -y_1(t) - y_2(t) \end{cases} .$$

Assuming that  $y_1(0) = 2$  and  $y_2(0) = 1$ , find  $y_1(t)$  and  $y_2(t)$ .

## 8.6 Homework exercises for Chapter 8

1. Let  $T : C^\infty(\mathbb{R}, \mathbb{R}) \rightarrow C^\infty(\mathbb{R}, \mathbb{R})$  be defined by  $T(f) = f'' + 6f' + 5f$ .
  - a) Compute and simplify  $T(e^{2x})$ .
  - b) Compute and simplify  $T(5e^{-3x})$ .
  - c) Prove that  $T$  is a linear transformation.
2. Prove that for any scalars  $c_0, \dots, c_n \in \mathbb{R}$ , the function  $T : C^\infty(\mathbb{R}, \mathbb{R}) \rightarrow C^\infty(\mathbb{R}, \mathbb{R})$  defined by

$$T(f) = c_n f^{(n)} + c_{n-1} f^{(n-1)} + \dots + c_2 f'' + c_1 f' + c_0 f$$

is a linear transformation.

3. Prove statement (1) of Theorem 8.3.
4. Show that  $g(x) = x^{4/3}e^{-x^2/6}$  is an eigenfunction for the linear transformation  $T : C^\infty(\mathbb{R}, \mathbb{R}) \rightarrow C^\infty(\mathbb{R}, \mathbb{R})$  defined by  $T(f) = 3xf'(x) + x^2f(x)$ . Find the eigenvalue to which  $g$  corresponds.
5. Find two linearly independent eigenfunctions corresponding to eigenvalue  $\lambda = 1$  for the linear transformation  $T : C^\infty(\mathbb{R}, \mathbb{R}) \rightarrow C^\infty(\mathbb{R}, \mathbb{R})$  defined by  $T(f) = f''$ . Explain why your two functions are indeed linearly independent.
6. Find two linearly independent eigenfunctions corresponding to eigenvalue  $\lambda = -1$  for the linear transformation  $T : C^\infty(\mathbb{R}, \mathbb{R}) \rightarrow C^\infty(\mathbb{R}, \mathbb{R})$  defined by  $T(f) = f''$ . Explain why your two functions are indeed linearly independent.
7. In a certain type of electrical circuit (containing a battery, resistor, inductor and capacitor in series), Kirchoff's Law says that the charge on the capacitor at time  $t$  is a function  $Q(t)$  which satisfies

$$LQ''(t) + RQ'(t) + \frac{1}{C}Q(t) = E(t)$$

where  $L, R$  and  $C$  are constants and  $E(t)$  is a function. Suppose that  $L = 2$ ,  $R = 10$ ,  $C = \frac{1}{12}$  and  $E(t) = 0$ . Find the general solution  $Q(t)$  of this equation.

8. Find all functions  $y = y(t)$  such that  $y'' + 4y' - 60y = 215 \cos t - 476 \sin t$ .
9. Find a function  $y = y(x)$  which satisfies  $y(0) = 2$ ,  $y'(0) = -3$ ,  $y''(0) = 1$  and

$$y''' - y'' - 4y' + 4y = 0.$$

10. (This problem must be done by hand.) Let  $A = \begin{pmatrix} 2 & 6 \\ 7 & 1 \end{pmatrix}$ .

- a) Find the characteristic polynomial of this matrix.  
b) Find the eigenvalues and eigenvectors of this matrix.
11. Find the eigenvalues and eigenvectors of the matrix  $\begin{pmatrix} 2 & -1 & 1 \\ 2 & -1 & -8 \\ 4 & -2 & -7 \end{pmatrix}$ .
12. a) Suppose a  $2 \times 2$  matrix has trace  $-2$  and determinant  $-15$ .  
i. Find the eigenvalues of the matrix.  
ii. Is the matrix diagonalizable? Why or why not?  
b) Suppose a  $3 \times 3$  matrix has determinant  $-32$  and trace  $6$ . If the matrix has two eigenvalues, one of multiplicity one and one of multiplicity two, find the eigenvalues of the matrix (stating which one is of which multiplicity).
13. Suppose a  $3 \times 3$  matrix  $A$  has the following eigenvalues and eigenvectors:
- $$\lambda = 2 \leftrightarrow \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \quad \lambda = -1 \leftrightarrow \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix} \quad \lambda = 4 \leftrightarrow \begin{pmatrix} 2 \\ -3 \\ 2 \end{pmatrix}$$
- Find  $A$ .
14. Compute the exact value of  $A^{25}$  if  $A = \begin{pmatrix} -4 & 6 \\ -3 & 5 \end{pmatrix}$ .
15. Compute the exact value of  $e^A$  if  $A = \begin{pmatrix} 0 & 1 \\ 2 & 1 \end{pmatrix}$ .
16. Let  $B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ .
- a) Show that  $0$  is the only eigenvalue of  $B$ .  
b) Prove that  $B$  is not diagonalizable by showing that it does not have two linearly independent eigenvectors.  
c) Compute  $B^2$ .  
d) Compute  $B^n$  for all  $n \geq 2$ .  
e) Compute  $\exp(B)$ .
17. Suppose  $\{x_n\}$  is a sequence defined by setting  $x_0 = 2$ ;  $x_1 = -3$  and for  $n \geq 2$ , setting  $x_n = -10x_{n-2} + 7x_{n-1}$ .
- a) Find  $x_2$ ,  $x_3$  and  $x_4$ .

- b) Find the exact value of  $x_{2000}$ .
18. Denote the owl population and wood rat population in a certain ecosystem at time  $k$  by  $\mathbf{x}_k = (O_k, R_k)$ . Suppose biologists determine that these populations evolve by the equations

$$\begin{cases} O_{k+1} = .5 O_k + .4 R_k \\ R_{k+1} = -p O_k + 1.1 R_k \end{cases}$$

where  $p > 0$  is some unknown parameter representing the rate of deaths of rats due to predation by owls. Suppose initially that there are 250 owls and 2000 rats in the ecosystem.

- a) Show that this set of recursive relations is equivalent to the matrix equation  $\mathbf{x}_{k+1} = A\mathbf{x}_k$  for some matrix  $A$ . Find  $A$ .
- b) Explain how you would compute the number of rats in the ecosystem when  $k = 100$ . (You don't actually have to do the problem, but your explanation should be thorough.)
19. Suppose that the number of foxes  $f(t)$  and number of rabbits  $r(t)$  in a forest at time  $t$  is modeled by the following system of differential equations:

$$\begin{cases} f'(t) = 4f(t) + r(t) \\ r'(t) = -2f(t) + r(t) \end{cases}$$

If there are initially 13 foxes and 7 rabbits in the forest, find formulas for  $f(t)$  and  $r(t)$ .

## Appendix A

---

# All about sets

---

### A.1 Sets

The fundamental objects of mathematics are called **sets**. A set is really just a list of objects (in math, the objects are usually numbers, or vectors, or functions). Here are some examples of sets (observe that sets are usually denoted by capital letters):

$$A = \{3, 5, 7, 9, 11\}$$

$$B = \{1, 2, 3, 4, 5, 6\}$$

$$C = \{3, 5, 7\}$$

The members of a set are called the **elements** of that set. For example, the elements of set  $C$  described above are 3, 5 and 7. If  $x$  is an element of set  $S$ , we write " $x \in S$ ". If  $x$  is not an element of set  $S$ , we write " $x \notin S$ ". For example, for the set  $A$  above,  $3 \in A$  and  $5 \in A$  but  $8 \notin A$ .

We often define a set without listing the elements (using English language). For example, the sets  $A$ ,  $B$  and  $C$  given above could be described, respectively, by saying

"let  $A$  be the set of odd numbers from 3 to 11";

"let  $B$  be the set of integers from 1 to 6";

"let  $C$  be the set of odd numbers from 3 to 7".

We also describe sets by using what is called **set-builder notation**: to describe the same sets  $A$ ,  $B$ ,  $C$  as above using set-builder notation, we would write (or say)

$$A = \{x : 3 \leq x \leq 11 \text{ and } x \text{ is odd}\}$$

$$B = \{x : 1 \leq x \leq 6 \text{ and } x \text{ is an integer}\}$$

$$C = \{x : 3 \leq x \leq 7 \text{ and } x \text{ is odd}\}.$$

The first statement above is interpreted as follows: it says that set  $A$  is equal to the set of numbers  $x$  such that (the colon means “such that” in mathematics)  $3 \leq x \leq 11$  and  $x$  is odd. Notice that this is exactly the set  $\{3, 5, 7, 9, 11\}$ .

Each of the sets  $A$ ,  $B$  and  $C$  above are **finite** sets, because they have finitely many members. Sets can have infinitely many members, however. For example, consider the interval

$$\begin{aligned} D &= [3, 5] \\ &= \text{the set of real numbers that are at least 3 and at most 5} \\ &= \{x : 3 \leq x \leq 5\}. \end{aligned}$$

This set contains not just 3, 4 and 5, but things like 3.2,  $\pi$ ,  $\sqrt{21}$ , since there is no requirement in the definition of  $D$  that  $x$  is an integer. Using the proper notation, we’d write things like  $\pi \in D$ ,  $3.75 \in D$ ,  $8 \notin D$ , etc.

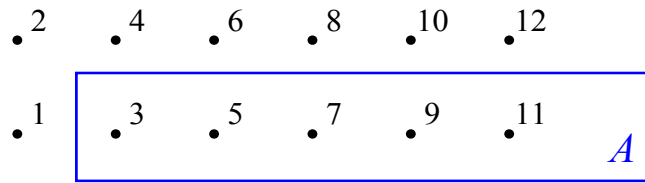
To show you a different kind of example: if you were defining some set of functions (instead of a set of numbers), then instead of  $x$  you’d write  $f$ , and then after the colon you’d describe what has to be true about  $f$  for the function  $f$  to be in the set. For example, the set  $E$  of functions whose derivative at  $x = 2$  is positive could be described by writing

$$E = \{f : f'(2) > 0\}.$$

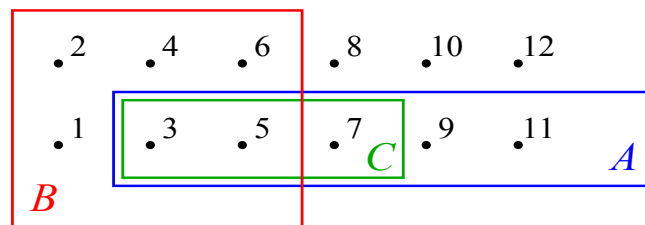
For this set  $E$ , it would be valid to say that if  $g(x) = x^3$ , then  $g \in E$  (because  $g'(2) = 3(2^2) = 12 > 0$ ) but if  $h(x) = 3 - 4x$ , then  $h \notin E$  (because  $h'(2) = -4 \leq 0$ ).

## A.2 Venn diagrams

A useful way to think about sets is to draw pictures called **Venn diagrams**. To draw a Venn diagram, represent each set you’re thinking about by a circle (or an oval, or a square, or a rectangle, or some other shape); think of an object as being an element of the set if and only if it is inside the shape corresponding to the set. For example, a Venn diagram for the set  $A$  described above (recall that  $A = \{3, 5, 7, 9, 11\}$ ) would be given by something like



because the box describing  $A$  contains exactly the elements of  $A$  (nothing more and nothing less). Similarly, a Venn diagram representing the  $A$ ,  $B$  and  $C$  we defined earlier, all at the same time, would be something like



### A.3 Relationships between, and operations on, sets

#### Subset

Let  $S$  and  $T$  be two sets. If all the elements of set  $S$  are also elements of set  $T$ , then we say “ $S$  is a **subset** of  $T$ ” and write  $S \subseteq T$ .

(Note the difference between the symbols  $\in$  and  $\subseteq$ : the first symbol should be preceded by an element, but the second symbol should be preceded by a set.)

If you draw a Venn diagram, to say  $S \subseteq T$  means that the shape corresponding to set  $S$  is completely inside the shape corresponding to set  $T$ . For example, for the sets  $A$  and  $C$  given in the preceding section,  $C \subseteq A$  since every element of  $C$  is also in  $A$ .

#### Equality of sets

To say two sets are **equal** means they have exactly the same elements; this is the same thing as requiring that each of the two sets is a subset of the other set. If two sets  $S$  and  $T$  are equal, we write “ $S = T$ ”.



## Union

The **union** of two (or more) sets is the set consisting of elements which belong to at least one of the given sets. This union is denoted  $S \cup T$  (or  $\cup_j S_j$  if you have a bunch of sets which are subscripted).

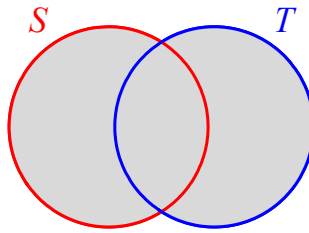
Another way to think of the union of two sets is that  $S \cup T$  is the set of things in  $S$  **or**  $T$  or both. So  $\cup$  is **mathematical code for the word or**.

For example, using the sets described above,

$$A \cup B = \{1, 2, 3, 4, 5, 6, 7, 9, 11\}$$

because all of the numbers listed there either are in  $A$ , or in  $B$ , or both.

In terms of Venn diagrams, the union is usually thought of as a “MasterCard-symbol” shaped region that encompasses the sets whose union you are taking. For example, the union of  $A$  and  $B$  in the figure below is exactly the shaded region (the set  $A$  is just the circle on the left;  $B$  is the circle on the right):



## Intersection

The **intersection** of two (or more) sets is the set consisting of elements which belong to all of the given sets. This intersection is denoted  $S \cap T$  (or  $\cap_j S_j$  if the sets are subscripted).

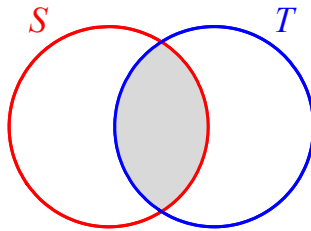
Another way to think of the intersection of two sets is that  $S \cap T$  is the set of things in  $S$  **and**  $T$  or both. So  $\cap$  is **mathematical code for the word and**.

For example, using the sets described  $A$  and  $B$  earlier,

$$A \cap B = \{3, 5\}$$

because the only numbers lying in both  $A$  and  $B$  are 3 and 5.

In terms of Venn diagrams, the intersection of two sets is the overlap of the shapes representing the sets (below, the shaded region is  $A \cap B$ ):



### Frequently used special sets with specific names

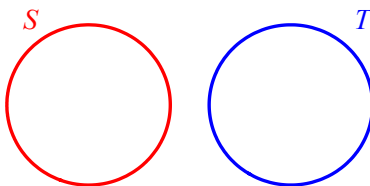
- The set which has no elements is called the **empty set** and is denoted  $\emptyset$ .
- The set of real numbers is denoted  $\mathbb{R}$ .
- The set of integers is denoted  $\mathbb{Z}$  (i.e.  $\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$ ).
- The set of natural numbers is denoted  $\mathbb{N}$  (i.e.  $\mathbb{N} = \{1, 2, 3, 4, \dots\}$ ).
- The set of ordered pairs  $(x, y)$  of real numbers is denoted  $\mathbb{R}^2$ , and the set of  $n$ -tuples of real numbers is called  $\mathbb{R}^n$ .

### Disjointness

Two sets  $S$  and  $T$  are called **disjoint** (or **mutually exclusive**) if their intersection is the empty set.

Equivalently, this means there are no objects which are both elements of  $S$  and elements of  $T$ .

If you drew a Venn diagram with sets  $S$  and  $T$  where  $S$  and  $T$  are disjoint, then the shapes corresponding to  $S$  and  $T$  should not overlap: the Venn diagram would look like this:



## A.4 Sums and differences of sets

Suppose you have two sets (say  $S$  and  $T$ ) where the elements of these objects are vectors (i.e. objects that can be added to one another and multiplied by constants).

Define a new set, called the **sum** of  $S$  and  $T$  and denoted  $S + T$ , to be the set of all objects which are the sum of some element of  $S$  and some element of  $T$ . In set-builder notation, this means

$$S + T = \{s + t : s \in S, t \in T\}.$$

Also, define the set  $S - T$  to be the set of all objects which can be written as some element of  $S$  minus some element of  $T$ . In set-builder notation, this means

$$S - T = \{s - t : s \in S, t \in T\}.$$

If  $x$  is an element, we write  $S + x$  and  $S - x$  when we mean  $S + \{x\}$  and  $S - \{x\}$ , respectively.

Last, given a set  $S$  and a real number  $c$ , define the set  $cS$  to be the set of all objects which are  $c$  times something in  $S$ .

Here are some examples, where  $F$  is given as  $\{0, 2, 4\}$  and  $G$  is given as  $\{0, 15\}$ :

$5F$	$= \{0, 10, 20\}$	(obtained by multiplying each element in $F$ by 5)
$F + G$	$= \{0, 2, 4, 15, 17, 19\}$	(obtained by adding each element in $F$ to each element of $G$ )
$-G$	$= \{0, -15\}$	(obtained by multiplying each element in $G$ by $-1$ )
$F + 3$	$= \{3, 5, 7\}$	(obtained by adding 3 to each element in $F$ ).

## Appendix B

---

# *Mathematica* information

---

### B.1 What is *Mathematica*?

*Mathematica* is an extremely useful and powerful software package / programming language *Mathematica* is an extremely useful and powerful software package / programming language invented by a mathematician named Stephen Wolfram. Early versions of *Mathematica* came out in the late 1980s and early 1990s; as of 2023, the most recent version available to you is *Mathematica* 13.

*Mathematica* does symbolic manipulation of mathematical expressions; it solves all kinds of equations; it has a library of important functions from mathematics which it recognizes while doing computations; it does 2– and 3–dimensional graphics; it has a built-in word processor tool; it works well with Java and C++; etc. One thing it doesn't do is prove theorems, so it is less useful for a theoretical mathematician than it is for an engineer or college student.

#### **A bit about how *Mathematica* works**

When you use the *Mathematica* program, you are actually running two programs. The “front end” of *Mathematica* is the part that you type on and the part you see. The “kernel” is the part of *Mathematica* that actually does the calculations. If you type in  $2 + 2$  and hit [ENTER] (actually [SHIFT]+[ENTER]; see below), the front end “sends” that information to the kernel which actually does the computation. The kernel then “sends” the result back to the front end, which displays 4 on the screen.

#### **About *Mathematica* notebooks and cells**

The actual files that *Mathematica* produces that you can edit and save are called *notebooks* and carry the file designation \*.nb; they take up little space and can easily

be saved to Google docs or on a flash drive, or emailed to yourself if you want them somewhere you can retrieve them.

**Suggestion:** when saving any file, include the date in the file name (so it is easier to remember which file you are supposed to be open).

A *Mathematica* notebook is broken into *cells*. A cell can contain text, input, or output. A cell is indicated by a dark blue, right bracket (a “]”) on the right-hand side of the notebook. To select a cell, click that bracket. This highlights the “]” in blue. Once selected, you can cut/copy/paste/delete cells as you would highlighted blocks of text in a Word document.

To change the formatting of a cell, select the cell, then click “Format / Style” and select the style you want. You may want to play around with this to see what the various styles look like. There are three particularly important styles:

- **input:** this is the default style for new cells you type
- **output:** this is the default style for cells the kernel produces from your commands
- **text:** changing a cell to text style allows you to make comments in between the calculations

To execute an input cell, put the cursor anywhere in the cell and hit [SHIFT]+[ENTER] (or the [ENTER] on the numeric keypad at the far-right edge of the keyboard). The [ENTER] next to the apostrophe key (a.k.a. [RETURN]) gives you only a carriage return.

## B.2 Important general concepts re: *Mathematica* syntax

**Executing mathematical commands:** To execute an input cell, put the cursor anywhere in the cell and hit [SHIFT]+[ENTER] (or the [ENTER] on the numeric keypad at the far-right edge of the keyboard). The [ENTER] next to the apostrophe key (a.k.a. [RETURN]) gives you only a carriage return.

**Multiplication:** use a star or a space:  $2 * 3$  or  $2\ 3$  will multiply numbers;  $a \times$  means  $a$  times  $x$ ;  $ax$  means the variable  $ax$  (in *Mathematica*, variables do not have to be named after one letter; they can be named by words or other strings of characters as well).

**Parentheses:** used for grouping only. Parentheses mean “times” in *Mathematica*.

**Brackets:** used to enclose all functions and *Mathematica* commands. For example, to evaluate a function  $f(x)$ , you would type  $f[x]$ ; for  $\sin x$  you type  $\text{Sin}[x]$ ; etc.. **Brackets** mean “of” in *Mathematica* and cannot be used for multiplication.

**Capitalization:** All *Mathematica* commands and built-in functions begin with **capital letters**. For example, to find the sine of  $\pi$ , typing  $\text{sin}(\text{pi})$  does you no good (this would be the variable “sin” times the variable “pi”). The correct syntax is  $\text{Sin}[\text{Pi}]$ .

**Spaces:** *Mathematica* commands do not have spaces in them; for example, the inverse function of sine is  $\text{ArcSin}$ , not  $\text{Arc Sin}$  or  $\text{Arcsin}$ .

**Pallettes:** Lots of useful commands are available on the Basic Math Assistant Palette, which can be brought up by clicking “Pallettes / Basic Math Assistant” on the toolbar. If you click on a button in the palette, what you see appears in the cell.

**Commands *Mathematica* knows:**  $\text{Sqrt}$ ,  $\text{Sin}$ ,  $\text{Cos}$ ,  $\text{Tan}$ ,  $\text{Csc}$ ,  $\text{Cot}$ ,  $\text{Sec}$ ,  $\text{ArcSin}$ ,  $\text{ArcCos}$ ,  $\text{ArcTan}$ ,  $\text{ArcCsc}$ ,  $\text{ArcSec}$ ,  $\text{ArcCot}$ , ! (for factorial). It knows what  $\text{Pi}$  and  $\text{E}$  are (but not  $\text{pi}$  or  $\text{e}$ ).

**Logarithms:**  $\text{Log}[ ]$  means natural logarithm (base  $e$ );  $\text{Log10}[ ]$  means common logarithm (base 10).

$\%$  refers to the last output (like ANS on a TI-calculator).

**Exact answers versus decimal approximations:** *Mathematica* gives exact answers for everything if possible. If you need a decimal approximation, click “numerical value” or use the command  $\text{N}[ ]$ . For example,  $\text{N}[\text{Pi}]$  spits out 3.14159...

**To solve an equation:** make sure there are two equals signs (“ $==$ ”) in your equation.

**Getting help from the program:** To get help on a command, type `?` followed by the command you don't understand (with no space between the `?` and the command).

**To export graphics:** Once *Mathematica* produces a graphic, you can right-click the graphic, and select "Copy Graphic". Then you can go in a Word document or a PowerPoint, and paste the graphic. You can subsequently resize it and/or move it around as you see fit.

**Troubleshooting:** For a command to run correctly, you usually want everything in your command to be black. **If anything is purple or red, that suggests where the problem is.** Variables that don't have values should be blue. Next, check that everything is capitalized appropriately. Next, check that you aren't missing a space if you are trying to multiply two variables. Next, if you are using variables in your code, try clearing the variables by executing something like `Clear[x]` (if your variable is  $x$ ). Then re-run the command that is giving you trouble.

If *Mathematica* freezes up in the middle of a calculation and you see "Running..." at the top of your screen, click "Evaluation / Abort Evaluation" on the toolbar. If this doesn't help, kill the program and restart it.

**To get help:** Email me, and **attach your *Mathematica* file (or a screenshot) to your email.** I can troubleshoot things pretty quickly if the file is attached. If the file isn't attached, it is hard for me to figure out what you are doing wrong. Alternatively, seek assistance from another math major who has experience with *Mathematica*.

## B.3 *Mathematica* quick reference guides

### General tasks

TASK	MATHEMATICA SYNTAX
To call the preceding output	%
To get a decimal approximation to the preceding output	N[%] (or click numerical value)

### Algebraic manipulations

TASK	MATHEMATICA SYNTAX
To factor an expression	Factor[ ]
To multiply out an expression (i.e. FOIL an expression)	Expand[ ]
Partial fraction decomposition	Apart[ ]
To combine rational terms (i.e. “undo” a partial fraction decomp)	Together[ ]
To simplify an answer	Simplify[ ] (or FullSimplify[ ])

### Solving equations

GOAL	MATHEMATICA SYNTAX
Find exact solution(s) to equation of form $lhs = rhs$ (assuming the variable is $x$ )	Solve[ $lhs == rhs, x$ ] (two equals signs) (works only with polynomials or other relatively “easy” equations)
Find decimal approx. to solutions of equation $lhs = rhs$	NSolve[ $lhs == rhs, x$ ] (two equals signs) (works only with “easy” equations)
Find decimal approx. to solutions of equation $lhs = rhs$	FindRoot[ $lhs == rhs, \{x, guess\}$ ] (two equals signs)
Solve two (or more) equations together, like $\begin{cases} lhs_1 = rhs_1 \\ lhs_2 = rhs_2 \end{cases}$ (assuming variables are $x$ and $y$ )	Solve[ $\{lhs_1 == rhs_1, lhs_2 == rhs_2\}, \{x, y\}$ ]



**Precalculus operations**

	EXPRESSION	MATHEMATICA SYNTAX
SPECIAL SYMBOLS	$e$	<b>E</b> (not e) (or use Basic Math Assistant palette)
	$\pi$	<b>Pi</b> (or use Basic Math Assistant)
	$\infty$	<b>Infinity</b> (or use Basic Math Assistant) (or type [Esc] inf [Esc])
	$i = \sqrt{-1}$	<b>I</b> (not i) (or use Basic Math Assistant)
ARITHMETIC	$3 + 4x$	<b>3 + 4x</b>
	$5 - 27$	<b>5 - 27</b>
	$12x$	<b>12x</b> or <b>12 x</b> or <b>12 * x</b>
	$xy$	<b>x y</b> (don't forget the space)
	$\frac{x}{y}$	<b>x/y</b> (or use Basic Math Assistant palette) (or type [CTRL]+/ to get $\frac{\square}{\square}$ )
	$\sqrt{32}$	<b>Sqrt[32]</b> (or use Basic Math Assistant) (or type [CTRL]+2 for the $\sqrt{\quad}$ sign)
	$\sqrt[4]{40}$	<b>40^(1/4)</b> (or use Basic Math Assistant)
	$ x - 3 $	<b>Abs[x-3]</b>
	<b>30! (factorial)</b>	<b>30!</b>
TRIG	$\sin \pi$	<b>Sin[Pi]</b>
	$\cos(x(y + 1))$	<b>Cos[x(y+1)]</b>
	$\cos 60^\circ$	<b>Cos[60 Degree]</b> (or use Basic Math Assistant)
	$\cot\left(\frac{2\pi}{3} + \frac{3\pi}{4}\right)$	<b>Cot[2 Pi/3 + 3 Pi/4]</b>
	$\sin^2 x$	<b>Sin[x]^2</b> (not <b>Sin^2[x]</b> )
	$\arctan 1$	<b>ArcTan[1]</b>
EXPS / LOGS	$\ln 3$	<b>Log[3]</b>
	$\log_6 63$	<b>Log[6,63]</b>
	$\log 18$	<b>Log10[18]</b> or <b>Log[10, 18]</b>
	$2^{7y}$	<b>2^(7y)</b> (or use Basic Math Assistant) (or type [CTRL]+6 to get $\square^\square$ )
	$e^{x-5+x^2}$	<b>E^(x-5+x^2)</b> or <b>Exp[x-5+x^2]</b> (or use Basic Math Assistant)

## Defining functions

CLASS OF FUNCTION	SYNTAX TO DEFINE FUNCTION
Calculus 1 function $f : \mathbb{R} \rightarrow \mathbb{R}$ $x \xrightarrow{f} y$  Ex: $f(x) = 3 \cos(x^{2-x})$	$f[x\_ ] = formula$ (one equals sign, underscore after the x) $f[x\_ ] = 3 \text{Cos}[x^{(2-x)}]$

## Algebraic operations on functions

All these commands assume you have previously defined the function(s) as outlined above.

EXPRESSION	MATHEMATICA SYNTAX
Generate table of values for $f$	<code>Table[{x, f[x]}, {x, xmin, xmax, step}]</code> (put <code>//TableForm</code> after this command to format the output in a table)
$f(x + 3)$	<code>f[x+3]</code>
$xf(2x) - x^2f(x)$	<code>x f[2x] - x^2 f[x]</code> (spaces important)
Composition $(f \circ g)(x)$	<code>f[g[x]]</code>
Addition $(f + g)(x)$	<code>f[x] + g[x]</code>
Multiplication $(fg)(x)$	<code>f[x] g[x]</code>
Powers $f^n(x)$	<code>(f[x])^n</code> (or just <code>f[x]^n</code> )
Ex: $\sin^2 x$	<code>Sin[x]^2</code>

## Graphs

The basic command to graph a function is `Plot[f(x), {x, xmin, xmax}]`; the examples below describe how to adapt the `Plot[ ]` command:

GOAL	HOW TO ADAPT THE <code>Plot[ ]</code> COMMAND
Plot multiple graphs at once	<code>Plot[{formula, formula, ..., formula}, {x, xmin, xmax}]</code>
Plot the graph of $f(x) = formula$ with range of $y$ -values specified	<code>Plot[formula, {x, xmin, xmax}, PlotRange -&gt; {ymin, ymax}]</code>
Plot the graph of $f(x) = formula$ with $x$ - and $y$ -axes on same scale	<code>Plot[formula, {x,xmin,xmax}, PlotRange -&gt; ymin,ymax, AspectRatio -&gt; Automatic]</code>
Plot the graph of $f(x) = formula$ with a red, dashed curve	<code>Plot[formula, {x,xmin,xmax}, PlotStyle -&gt; {Red, Dashed}]</code>

## Single-variable calculus

EXPRESSION	MATHEMATICA SYNTAX
$\lim_{x \rightarrow 4} f(x)$	<code>Limit[f[x], x -&gt; 4]</code>
$f'(3)$	<code>f'[3]</code>
$h'(x)$	<code>D[h[x], x]</code>
$\frac{d}{dx}(\cos x)$	<code>D[Cos[x], x]</code>
$g'''(x)$	<code>g' '[x]</code> or <code>D[g[x], {x,3}]</code>
$\int x^2 dx$	<code>Integrate[x^2, x]</code> (or use Basic Math Assistant palette) <b>Note: answer will be missing the "+C"</b>
$\int_2^5 \cos x dx$	For an exact answer: <code>Integrate[Cos[x], {x, 2, 5}]</code> (or use Basic Math Assistant) For a decimal approximation: <code>NIntegrate[Cos[x], {x, 2, 5}]</code>
$\sum_{k=1}^{12} f(k)$	<code>Sum[f[k], {k, 1, 12}]</code> (or use Basic Math Assistant)
$\sum_{n=3}^{\infty} blah$	<code>Sum[blah, {n, 3, Infinity}]</code> (or use Basic Math Assistant)

## B.4 Linear algebra in *Mathematica*

### Typing in vectors

To type in a vector, use a set of squiggly braces to surround the components, and separate the components with commas: for example, to save the vector  $v = (2, -3, 5, 0, 1)$ , execute

$$v = \{2, -3, 5, 0, 1\}$$

### Typing in matrices

To type in a matrix, use one of two methods:

1. Use squiggly braces and commas to separate the entries. Each row should be surrounded by a squiggly brace, and the entire matrix should be surrounded by a set of squiggly braces, and everything should be separated by commas. For example, to define  $A$  as

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

one could execute

$$A = \{\{1, 2\}, \{3, 4\}\}$$

Note that if you have a column matrix like  $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ , this matrix can be defined

by thinking of  $B$  as a vector and typing something like  $B = \{1, 2, 3\}$  (instead of having to type  $B = \{\{1\}, \{2\}, \{3\}\}$ ).

2. On the Basic Math Assistant palette, under Basic Commands, click the matrix. Then type  $A =$ , then click the matrix in the palette. To add rows and columns, click AddRow or AddColumn until the matrix is the appropriate size. Then go into the matrix and type in each entry, moving between the locations using the [TAB] key or clicking on the location you want.

If the entries of the matrix are functions, then define the matrix as a function by executing  $A[t\_ ] = \dots$  instead of  $A = \dots$

Once you have defined all necessary matrices and vectors, *Mathematica* commands for operations on those matrices are given in the quick reference guides for vectors and matrices on the next two pages.

**Vector operations**

EXPRESSION	MATHEMATICA SYNTAX
Vector addition / subtraction: $(2, 5, -1) + (5, 0, 2)$ $\mathbf{v} - \mathbf{w}$	$\{2,5,-1\} + \{5,0,2\}$ $\mathbf{v} - \mathbf{w}$
Scalar multiplication: $3(1, 2, -3, -4)$ $5\mathbf{v} - 3\mathbf{w}$	$3\{1,2,-3,-4\}$ $5\mathbf{v} - 3\mathbf{w}$
Dot product $\mathbf{v} \cdot \mathbf{w}$ $(3, -4) \cdot (2, 10)$	$\mathbf{v}.\mathbf{w}$ $\{3,-4\}.\{2,10\}$
Norm $\ \mathbf{v}\ $	$\text{Norm}[\mathbf{v}]$
Unit vector in same direction as $\mathbf{v}$ (i.e. $\frac{\mathbf{v}}{\ \mathbf{v}\ }$ )	$\text{Normalize}[\mathbf{v}]$
Projection of $\mathbf{v}$ onto $\mathbf{w}$ (i.e. $\pi_{\mathbf{w}}\mathbf{v}$ )	$\text{Projection}[\mathbf{v},\mathbf{w}]$
Angle between two vectors	$\text{VectorAngle}[\mathbf{v},\mathbf{w}]$ (answer is in radians) (to get degrees, click degree measure in the suggestions bar)
Cross product $\mathbf{v} \times \mathbf{w}$	$\text{Cross}[\mathbf{v},\mathbf{w}]$
To get the number of components of $\mathbf{v}$	$\text{Length}[\mathbf{v}]$
To get the $i^{\text{th}}$ component of $\mathbf{v}$	$\mathbf{v}[[i]]$

## Matrix operations

To make *Mathematica* display an answer as a matrix:

1. follow your command with `// MatrixForm`, or
2. once you've executed the command, choose Display as... matrix from the suggestions bar.

EXPRESSION	MATHEMATICA SYNTAX
Matrix addition / subtraction $A + B$ $A - B$	$A + B$ $A - B$
Scalar multiplication $3A$ $nA$ $-5A + \frac{1}{2}B$	$3A$ $n A$ (space important) $-5A + (1/2)B$
Matrix product $AB$ $A^2$ $A^7$	$A.B$ (the period is important) $A.A$ or <code>MatrixPower[A,2]</code> ( <b>not</b> $A^2$ ) <code>MatrixPower[A,7]</code>
Trace $tr(A)$	<code>Tr[A]</code>
Determinant $\det A$	<code>Det[A]</code>
Transpose $A^T$	<code>Transpose[A]</code>
To get the entry of matrix $A$ in the $i^{th}$ row and $j^{th}$ column	<code>A[[i,j]]</code>
To call the $n \times n$ identity matrix $I$	<code>IdentityMatrix[n]</code>
Find derivative of a matrix of functions term-by-term	<code>A'[t]</code>
Matrix exponential $e^A = \exp(A)$	<code>MatrixExp[A]</code>
Eigenvalues and eigenvectors of $A$ Just the eigenvalues of $A$ Just the eigenvectors of $A$ Find $\det(A - xI)$	<code>Eigensystem[A]</code> <code>Eigenvalues[A]</code> <code>Eigenvectors[A]</code> <code>CharacteristicPolynomial[A,x]</code>
Determine if $A$ is diagonalizable	<code>DiagonalizableMatrixQ[A]</code>
Determine if $A$ is positive definite	<code>PositiveDefiniteMatrixQ[A]</code>
Determine if $A$ is negative definite	<code>NegativeDefiniteMatrixQ[A]</code>

---

# Index

---

- $C^\infty(\mathbb{R}, \mathbb{R})$ , 241
- $T(\mathbf{x}) = \mathbf{b}$ , solutions of, 154
- $\mathbb{R}^n$ , definition of, 22
- $\mathbb{R}^n$ , pictorial representation, 23
- $\exists$ , 25
- $\forall$ , 25
- $\pi_W(\mathbf{v})$ , 107
- $\pi_{\mathbf{w}}\mathbf{v}$ , 104
- $proj_{\mathbf{w}}\mathbf{v}$ , 105
  
- absolute value (of a vector), 95
- addition (in a vector space), 19
- addition (of matrices), 37
- affine subspace, 62
- affine subspaces of  $\mathbb{R}$ ,  $\mathbb{R}^2$  and  $\mathbb{R}^3$ , 83
- angle (between vectors), 117
- associated subspace (of an affine subspace), 62
- augmented matrix, 196
  
- basis, 74
- basis (of column space), 208
- basis (of null space), 208
- basis (of row space), 208
- Basis Extension Theorem, 81
- Basis Theorem, 82
- bijjective, 157
- bijjectivity, equivalent properties, 159, 194
- bilinearity (dot product), 93
  
- brute-force method (for determining whether a subset is a subspace), 50
- brute-force method (for proving that a transformation is linear), 145
  
- Cauchy-Schwarz Inequality, 116
- characteristic equation (of a differential equation), 249
- characteristic polynomial (of a matrix), 255
- column space, 163
- column space, basis of, 208
- column space, dimension of, 165
- column vector, 38
- columns (of a matrix), 35
- component of a vector orthogonal to a subspace, 107
- composition of linear transformations, 143
- computing determinants, 232
- computing eigenvalues, 255
- consistent, 185
- counterexample, 55
- cross product, 122
- cross product, properties, 122
  
- definiteness (distance), 97
- definiteness (dot product), 93
- definiteness (norm), 96

- description of solutions to  $T(\mathbf{x}) = \mathbf{b}$ , 154  
 description of the solution set of a linear system, 188  
 determinant ( $2 \times 2$  matrix), 229  
 determinant, evaluation via minors, 234  
 determinant, properties of, 230  
 determinants on *Mathematica*, 236  
 determinants on calculators, 236  
 determinants, computing, 232  
 diagonal (matrix), 34  
 diagonal entries (of a matrix), 34  
 diagonalizable (matrix), 252  
 diagonalizing a matrix, 252  
 dimension (of a vector space), 78  
 dimension (of an affine subspace), 79  
 Dimension Formula, 109  
 Dimension Theorem, 78  
 dimensions of fundamental subspaces, 165  
 direction (of a vector), 60  
 distance (between vectors), 97  
 distance from vector to subspace, 110  
 distance, properties of, 97  
 dot product ( $\mathbb{R}^n$ ), 91  
 dot product (of functions), 92  
 dot product (of matrices), 92  
 dot product, properties of, 93  
 downward reduction, 201  
 Dual Relations, 95  
  
 echelon form, 198  
 eigenfunction, 244  
 eigenvalue, 244  
 eigenvalues, basic properties, 245  
 eigenvalues, computing, 255  
 eigenvector, 244  
 eigenvectors, basic properties, 245  
 equality (of matrices), 33  
 equivalent (linear systems), 185  
 equivalent properties to bijectivity, 159, 194  
  
 equivalent properties to injectivity, 158, 193  
 equivalent properties to surjectivity, 158, 192  
 evaluation of determinants via minors, 234  
 Exchange Lemma, 77  
 existence/uniqueness of solutions to  $T(\mathbf{x}) = \mathbf{b}$ , 154  
 exists, 25  
 exponential (of a matrix), 262  
  
 finite dimensional (vector space), 78  
 for all, 25  
 free column, 199  
 FTLA, 164  
 fundamental subspaces (of a matrix), 163  
 fundamental subspaces, dimension of, 165  
 Fundamental Theorem of Linear Algebra, 164  
  
 Gaussian elimination, 195, 201  
 Generalized Triangle Inequality, 117  
 Gram-Schmidt Theorem, 114  
  
 head-to-tail addition (of vectors), 23  
 hyperplane, 79, 118  
 hyperplane, normal equation, 118  
  
 identity matrix, 35  
 image, 152  
 Image-Kernel Theorem, 155  
 infinite dimensional (vector space), 78  
 infinitely differentiable, 241  
 injective, 157  
 injectivity, equivalent properties, 158, 193  
 inner product space, 93  
 inverse (of a function), 157  
 inverse (of a matrix), 167  
 inverse (of a matrix), computing, 211  
 inverse of  $1 \times 1$  matrix, 168, 169



- inverse of  $2 \times 2$  matrix, 169  
 inverses, properties of, 168  
 invertible, 157  
 invertible (matrix), 167  
  
 kernel, 152  
 key picture associated to linear transformations, 153  
  
 leading entry, 198  
 least-squares approximation, 215  
 least-squares solution, 217  
 left null space, 163  
 left null space, dimension of, 165  
 length (of a vector), 95  
 line, 62, 79  
 linear combination, 69  
 linear equation, 185  
 linear system, 185  
 linear transformation, 132  
 linear transformations, compositions of, 143  
 linear transformations, examples of, 149  
 linear transformations, key picture, 153  
 linearly dependent, 69  
 linearly independent, 71  
 Linearly Independent Set Theorem, 81  
 lower triangular, 34  
  
 magnitude, 95  
 matrix, 33  
 matrix exponential, 262  
 matrix exponential, properties of, 262  
 matrix inverse, computing, 211  
 matrix multiplication, 39  
 matrix operations on *Mathematica*, 44  
 matrix operations on calculators, 45  
 matrix operations, properties of, 38, 42  
 minor, 233  
 multiplication of matrices, 39  
 multiplicativity (distance), 97  
 multiplicativity (norm), 96  
 multiplicity (eigenvalue), 256  
  
 nontrivial combination, 69  
 norm, 95  
 normal equation (of hyperplane), 118  
 normalized version (of a vector), 99  
 norms, properties of, 96  
 null space, 163  
 null space, basis of, 208  
 null space, dimension of, 165  
  
 one-to-one, 157  
 onto, 157  
 orthogonal, 100  
 orthogonal (set of vectors), 112  
 orthogonal complement, 102  
 orthogonal complements are subspaces, 104  
 Orthogonal Decomposition Theorem (dimension 1), 104  
 Orthogonal Decomposition Theorem (general case), 107  
 orthogonal decomposition, uniqueness of, 108  
 orthonormal (basis), 112  
 orthonormal (set of vectors), 112  
  
 pairwise orthogonal, 112  
 parallel (vectors), 60  
 parallelepiped, 229  
 parallelogram addition (of vectors), 23  
 Parallelogram Law, 98  
 parametric equations (of a line), 63  
 parametric equations (of a plane), 64  
 particular solution, 154  
 perp (of a subspace), 102  
 perpendicular, 100  
 pivot, 198  
 pivot column, 199  
 plane, 64, 79  
 point, 79  
 Polarization Identity, 98  
 positivity (distance), 97  
 positivity (dot product), 93  
 positivity (norm), 96

- product (of matrices), 39  
 projection (onto a vector), 105  
 projection formula, 113  
 projection matrix, 217  
 projection of a vector onto a subspace, 107  
 projection onto a column space, 217  
 projections, properties of, 107  
 properties of cross product, 122  
 properties of determinants, 230  
 properties of distance, 97  
 properties of dot product, 93  
 properties of matrix exponentials, 262  
 properties of matrix inverses, 168  
 properties of matrix operations, 38, 42  
 properties of norms, 96  
 properties of projections, 107  
 Pythagorean Theorem, 101  
  
 rank (of a linear transformation), 153  
 rank (of a matrix), 165  
 Rank Theorem, 165, 199  
 Rank-Nullity Theorem, 156  
 real vector space, 19  
 reduced row-echelon form, 198  
 ref, 198  
 row equivalent (matrices), 198  
 row operations, 197  
 row reduction, 201  
 row space, 163  
 row space, basis of, 208  
 row space, dimension of, 165  
 row-echelon form, 198  
 rows (of a matrix), 35  
 rref, 198  
 Rule of Sarrus, 232  
  
 Sarrus, 232  
 scalar, 19  
 scalar multiplication, 19  
 scalar multiplication (of matrices), 37  
 solution (of a system), 185  
 solution set (of a system), 185  
  
 solution set of linear system, description of, 188  
 solving systems of equations on *Mathematica*, 219  
 solving systems of equations on calculators, 220  
 span (as a verb), 57  
 span (of a set of vectors), 57  
 span (of a single vector), 56  
 Spanning Set Theorem, 81  
 spectrum, 244  
 square (matrix), 34  
 standard basis of  $\mathbb{R}^n$ , 135  
 standard equation (of hyperplane), 118  
 standard matrix (of a linear transformation), 135  
 subspace, 49, 50  
 subspace, generic picture of, 67  
 subspaces of  $\mathbb{R}$ ,  $\mathbb{R}^2$  and  $\mathbb{R}^3$ , 83  
 surjective, 157  
 surjectivity, equivalent properties, 158, 192  
 symmetry (distance), 97  
 symmetry (dot product), 93  
 system of linear equations, 185  
  
 there exists, 25  
 trace, 34  
 translation vector, 62  
 transpose, 34  
 triangle inequality, 96, 97, 117  
 Triangle Inequality, Generalized, 117  
 triangle inequality, proof, 117  
 triangular, 34  
 trivial combination, 69  
  
 Unique Representation Theorem, 75  
 uniqueness of orthogonal decompositions, 108  
 unit vector, 99  
 upper triangular, 34  
 upward reduction, 201

vector, 19  
vector space, 19  
vector spaces, examples of, 25  
zero matrix, 38