A First Course in Linear Algebra

by Jeremy Lane

Contents

About these notes

These lecture notes were created for the course MAT 223 at the University of Toronto, Summer of 2017. This is a two month long accelerated summer course. It was the second time I had taught this course and only the third course I had taught as an instructor. Thus, there are many imperfections and things which could be improved.

These notes feature many in-class exercises and tutorial questions which ask the students to reach conclusions on their own, with the aid of the instructors and teaching assistants. Various constructions, algorithms, proofs, and definitions are not spelled out explicitly in the notes, but rather indicated by examples and exercises. This approach should benefit students who attend and participate. It is recommended that students attempt to write full, detailed answers to all of the in-class exercises and tutorials as part of their study.

These notes haven't been thoroughly edited since they were written (over the course of two months). They are presented as is and may contain errors. I am not maintaining these notes, so please do not contact me with any typographical or factual errors you find.

Some snippets of .tex used in these notes were taken with permission from another set of linear algebra notes written and shared with me by Benjamin Schachter.

<u>On the front page:</u> Ferns have astonishing self-similarities. The Barnsley Fern is a fractal that resembles a fern and can be created using affine transformations. See the webpage `https://en.wikipedia.org/wiki/Barnsley_fern`. This particular photo is my own, taken at Killarney Provincial Park on a camping trip.

Pedagogical remarks

- The only prerequisite for MAT 223 at University of Toronto is high school calculus (and by extension, some high school level geometry and trigonometry). Although some students in the course have much more background, these notes are written with the prerequisite in mind and start from scratch and go slowly. In particular, the notes begin with the assumption that the students' knowledge of mathematical terminology, conventions, etc. is very limited or has been forgotten over time. The first few chapters are designed as if the students are being taught not only math, but a language.

- In Chapter 1, we define the rank of a matrix as the number of leading entries in a reduced echelon form (a proof is given at the end of Chapter 3). This allows us to begin working with rank early in the course.

- The first four chapters are very algebraic by design. This allows students to get familiar working algebraically with definitions. Geometric concepts are introduced and intermingled with algebra in the second half of the course, once students have had time to adopt a more rigorous mindset.

- Being able to manipulate symbols following a set of algebraic rules is a key learning goal of the course. Early on, students are asked to do many of these abstract "calculations" as part of proofs. Later, algebraic manipulations are combined in proofs with more complicated definitions.

- Sets are the source of a lot of confusion. The notion of a set is introduced in Chapter 1 alongside the definition of the set of solutions of a system of equations. We introduce "set-builder notation," as well as enumerative descriptions of finite sets. Both are needed later in the course.

  Starting in Chapter 1, we talk freely about the size of a set, which is intuitive for the students, without giving any overly complicated definition of cardinality. We also introduce the notion of being an element of a set alongside the notion of being a solution to a system of equations.

  Although the proof that the set of solutions of a system of equations is invariant under elementary operations is essentially a proof that two sets are equal, we focus on the concept of showing that two sets are equal or different later, in Chapter 5, where we also describe the concept of set containment.

- In Chapter 6, we give a 'baby' definition of injective and surjective suitable for linear transformations (T is injective if $\ker(T) = \{\mathbf{0}\}$ and T is surjective if $\operatorname{im}(T) = \mathbb{R}^m$). This neatly avoids the complicated "double quantifier" involved in the abstract definition of surjective. We do not define the term "bijection." Invertibility of functions is defined in full generality by the property of having an inverse function. We use the terms 'domain', 'codomain', and 'image'. We avoid use of the word 'range.'

- Throughout the lectures, there are two types of interactive features.

Check your understanding

"Check your understanding" boxes typically follow a new definition or piece of terminology. In class, students are given 1-2 minutes to write down their answers, followed by a short discussion lead by the instructor.

In-class exercise

"In-class exercises" are longer problems that require the students to do a longer computation (they are given 5-10 minutes for these problems). Some of these questions have subtle complications built into them, in order to generate discussion.

Both of these features are designed to prompt interaction from students by giving them opportunities to solve problems and ask questions.

- There is much to say about which topics to include or not in such a course. Although there are many topics I would have liked to have covered, this is what we managed to complete in two months.

Chapter 1

Systems of linear equations

Before beginning this section, it may be helpful to recall the following terminology from high school algebra:

- mathematical expressions
- polynomials
- terms of polynomials
- equations
- coefficient of a polynomial
- variables
- solution of an equation

1.1 Linear equations

You are already familiar with equations involving variables from highschool. For example, part of the Pythagorean Theorem is the equation

$$a^2 + b^2 = c^2.$$

In the Pythagorean Theorem, the symbols a, b, and c are variable quantities (*variables*) that represent the lengths of the sides of a right-triangle (and c is the length of the hypotenuse). The fact that these variables occur in an equation together tells us that they are related by a rule. If we take the right-triangle whose shorter side lengths are $a = 3$ and $b = 4$, then this equation tells us that the length of the hypotenuse must be 5.

Another equation you are familiar with is the equation of a line,

$$y = mx + b$$

where x and y are variables that represent the xy-coordinates of a point in the xy-plane, and m and b are fixed numbers: m is the slope of the line, and b is the y-intercept. For instance, the equation

$$y = 2x + 1$$

describes a line in the xy-plane whose slope is 2 and whose y-intercept is 1. The points on this line are precisely the points whose xy-coordinates (x_0, y_0) solve this equation.

The study of linear equations is concerned with equations similar to the second example (the equation of a line in the xy-plane). However, we do not want to restrict ourselves to only two variables x and y, we want to be able to work with as many variables as we want or need. In general, we define a linear equation in n variables, where n is whatever positive integer we want, the following way.

Definition 1.1. A *linear equation* in n variables is an equation of the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

where a_1, a_2, ..., a_n, b are fixed numbers and x_1, $x_2, \ldots,$ x_n are variables. The numbers a_1, a_2, ..., a_n that multiply the variables are called *coefficients* of the equation and the number b is called the *constant term* of the equation.

Remark 1.2. The word *linear* derives from the latin word for “a line” and means “resembling a line, of or pertaining to a line.” We call these equations linear, since they resemble the equation of a line in the xy-plane.

When we have many variables, we often use letters numbered with subscripts, such as $x_1, \ldots, x_n$ or $y_1, \ldots, y_n$ for our variables. When there are only one, two, or three variables, it is common to use letters such as x, y, z or u, v, w for variables.

Example 1.3. The equation

$$2x + 3y = 6$$

is linear. The variables are x and y. The coefficients are 2 and 3. The constant term is 6.

Example 1.4. The equation

$$x^2 + 2xy + y^2 = 1$$

is not linear in the variables x and y.

Example 1.5. The equation

$$4 + 4z_1 = 6z_2 - \pi z_3 - 10 + 8z_4$$

is linear, even though it is not written precisely the same way as in Definition 1.1. The variables are z_1, z_2, z_3, z_4. We can rearrange this equation so that it looks more like the equation in Definition 1.1:

$$4z_1 - 6z_2 + \pi z_3 - 8z_4 = -6.$$

What are the coefficients of this equation? What is the constant term?

Example 1.6. Most equations that we encounter in math are not linear. For example, the following equations are not linear in the variables x and y:

$$y = x^2,\ y = e^x,\ xy = 1.$$

Check your understanding

Einstein's famous "mass-energy equivalence" equation says that

$$E = mc^2$$

where E is total energy of an object, m is mass of the object, and c is the speed of light.

Is Einstein's equation linear?

Hint: what are the variables and constants in this equation?

$\mathbb{R}^2$ is the set of all pairs (s_1, s_2) of real numbers s_1 and s_2. Geometrically, every pair of numbers (s_1, s_2) in $\mathbb{R}^2$ corresponds to a point in the xy-plane with coordinates $x = s_1$ and $y = s_2$.

Solutions of the equation

$$y = 2x + 1$$

are pairs of numbers $(s_1, s_2) \in \mathbb{R}^2$ such that $x = s_1$ and $y = s_2$ solves the equation, i.e. such that

$$s_2 = 2s_1 + 1.$$

The set of all solutions (s_1, s_2) of this equation corresponds to the set of points on the line in the xy-plane with slope 2 and y-intercept 1. We can also describe the set of solutions using set-builder notation,

$$\left\{(x, y) \in \mathbb{R}^2 \colon y = 2x + 1\right\}.$$

Set-builder notation:

A set is a collection of objects. Examples of sets include the set of all socks in my closet, the set of all apartments in Toronto, and the set of all real numbers.

We describe sets using set-builder notation, which is as follows. The notation

$$\{x : P(x)\}$$

which will be read as "the set of all objects x which satisfy condition P" or "the set of all objects x for which property P is true." If it is important to specify that the objects in a all lie in some particular domain, sometimes the following notation will be used:

$$\{x \in \text{Domain} : P(x)\}.$$

This is read as "the set of all x in Domain for which property P is true." The symbol $\in$ means "is an element of." Some examples include

- $\{x \in \mathbb{R} : x > 0\}$ is the set of all x in the set of real numbers for which x is strictly greater than zero. 1 is an element of this set since $1 > 0$. -1 is not an element of this set. Is 0 and element of this set?

- $\{s \in \text{my shirts} : s \text{ is blue}\}$ is the set of all shirts in my closet that are blue. My white shirts are not elements of this set. Similarly, my socks are not an element of this set. On the other hand, my Blue Jays shirt is an element of this set because it is blue.

Since the set-builder notation describing a set is often lengthy, we often use a letter or symbol to represent a set. For instance, we might call

$$B := \{s \in \text{my shirts} : s \text{ is blue}\}.$$

With this notation, we could describe the fact that my Blue Jays shirt is blue by writing

$$\text{my Blue Jays shirt } \in B.$$

In plain english, this sentence reads, "my Blue Jays shirt is an element of the set of my shirts which are blue."

We will learn more about set-builder notation and sets throughout this course.

More generally, we call a list of real numbers $(s_1, \dots, s_n)$ a *n-tuple of real numbers*[1], and denote by $\mathbb{R}^n$ (pronounced "r-n") the set of all n-tuples of real numbers. For the time being, we will focus on the algebraic aspects of solutions to linear equations.

Definition 1.7. A *solution* of a linear equation

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

is a n-tuple $(s_1, \dots, s_n)$ of real numbers such that

$$a_1s_1 + a_2s_2 + \ldots + a_ns_n = b.$$

In other words, when we let $x_1 = s_1, x_2 = s_2, \dots x_n = s_n$, the equation is true.

Example 1.8. The pair of numbers $(3, 0)$ is a solution of the equation

$$2x + 3y = 6.$$

On the other hand, the pair of numbers $(3, 1)$ is not a solution of the equation. We write

$$(3, 1) \notin \left\{(x, y) \in \mathbb{R}^2 \colon 2x + 3y = 6\right\}$$

which reads "The pair $(3, 1)$ is not an element of the set of pairs (x, y) of real numbers which solve the equation $2x + 3y = 6$." More succinctly, we can simply say "The pair $(3, 1)$ is not a solution of the equation $2x + 3y = 6$."

[1] A 2-tuple is just another name for a pair of numbers, (x, y).

Check your understanding

Which of the following 4-tuples are solutions of the linear equation in Example 1.5?

$$(0, 0, 0, 0),\ (1, -1, 1, -1),\ (2, 0, 0, 1)$$

Try to express your answer using set-builder notation and the symbols $\in$ and $\notin$.

In mathematics, a common goal when given an equation is to describe ALL its solutions. For example, for the equation

$$x^2 = 1,$$

we are not content with the answer $x = 1$ because there is another solution, $x = -1$. The set of solutions to this equation could be described in set-builder notation as

$$\{x \in \mathbb{R} \colon x^2 = 1\}.$$

We know that this set contains exactly two numbers, 1 and -1. In this case, it is more simple to describe this set by listing its elements:

$$\{1, -1\}$$

Enumerative descriptions of sets: There are various ways to describe sets. We have already seen set-builder notation. In the preceding example, we saw that we could express the set $\{x \in \mathbb{R} \colon x^2 = 1\}$ by simply writing $\{1, -1\}$.

What have we done here? When we write $\{1, -1\}$ we simply mean "the set consisting of the elements 1 and -1." We simply list the elements of the set and add curly braces to indicate that we are describing a set.

This is a common way to describe sets that are small such as the one in the example above. In general, a description of a set given in this manner – by listing its elements inside curly braces – is called an *enumerative description.* The name is not particularly important.

As we will see later on, the fact that one set can be described several different ways will lead to some challenging problems. In the example above, we explained why the set-builder notation

$$\{x \in \mathbb{R} \colon x^2 = 1\}$$

and the enumerative description $\{1, -1\}$ are actually describing the same set. We will return to this concept many times throughout the course.

1.2 Systems of linear equations

A system of equations is a list of several equations. For example, one can have a system of equations that includes some non-linear equations, such as

$$x^2 + y^2 = 1,\ x = y$$

or one can have a system of equations that only includes linear equations, such as

$$y = 2x + 1,\ x = y + 2.$$

In this class we are mainly interested in the second kind of system.

Since mathematicians are greedy, we would like to be able to consider systems with as many equations as we want and as many variables as we want.

Definition 1.9. A *system of m linear equations in n variables* is a list of m linear equations of the form

$$m \text{ equations} \left\{ \begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ & & & & \vdots & & & & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array} \right.$$

where $x_1, \ldots, x_n$ are variables, a_{ij} are coefficients, and b_i are constant terms.

An n-tuple $(s_1, \ldots, s_n)$ of numbers is a *solution* of this system of equations if it is a solution of *every* equation in the system.

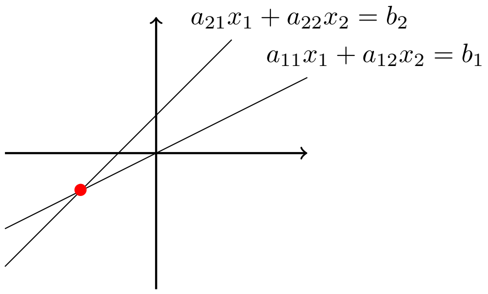
Notation: In the definition above, we write a_{ij} and b_i. What does this mean?

- There are m equations, and we label the equations $1, 2, 3, \ldots$ up to m. In equation 1, we call the constant term b_1. In equation 2 we call the constant term b_2, and so on. If i is an integer between 1 and m, then

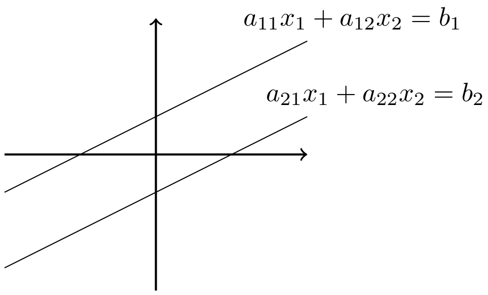
  we write b_i to mean the constant term of equation i.

- There are also n variables, which we have already labelled $x_1, x_2, \ldots$ and so on, up to x_n. If $1 \leq i \leq m$, and $1 \leq j \leq n$, then

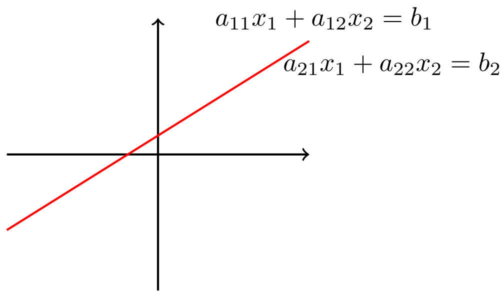
  we write a_{ij} to mean the coefficient of x_j in equation i.



Case a: Different slopes
The intersection is the unique solution (in red).



Case b: Same slopes, non-intersecting
There are no solutions.



Case c: Same slopes, coinciding
There are infintely many solutions (in red).

Example 1.10 (Systems of two linear equations in two variables)*.* Since each linear equation in two variables describes a line in the xy-plane, solutions of a system of linear equations in two variables can be interpreted geometrically as intersections of lines in the xy-plane. For example,

a) The equations in the system

$$\begin{aligned} y &= x \\ y &= 2x + 3 \end{aligned}$$

describe two lines with different slopes which necessarily intersect at one point.

b) The equations in the system

$$\begin{aligned} y &= x + 4 \\ y &= x - 4 \end{aligned}$$

describe two parallel lines with different y-intercepts. Since parallel lines do not intersect, the system has zero solutions.

c) Both equations in the system

$$\begin{aligned} y &= x + 1 \\ 2y &= 2x + 2 \end{aligned}$$

describe the same line. Thus, every point on this line is a solution of the system of linear equations. Since there are infinitely many points on a line, the system has infinitely many solutions.

Example 1.11. Consider the system of two equations in the variables x, y, z,

$$\begin{array}{rcrcrcr} x & + & 2y & - & z & = & 0 \\ 2x & + & y & + & z & = & 2 \end{array}$$

We would like to describe all solutions of this system. In other words, we want to find all 3-tuples (s_1, s_2, s_3) that solve both equations.

If we add the two equations together, the z terms cancel and leave us with

$$3x + 3y = 2$$

which simplifies to

$$x = \frac{2}{3} - y.$$

Substituting this equation back into the first equation in the system, we get

$$\left(\frac{2}{3} - y\right) + 2y - z = 0$$

which simplifies to

$$y = z - \frac{2}{3}.$$

Thus, every solution to this system is of the form

$$\left(\frac{4}{3} - z, z - \frac{2}{3}, z\right).$$

On the other hand, for every real number t, we can check that

$$\left(\frac{4}{3} - t, t - \frac{2}{3}, t\right)$$

is a solution to the system of linear equations. We do this by plugging $x = \frac{4}{3} - t$, $y = t - \frac{2}{3}$ and $z = t$ into the original equations and checking that

$$\left(\frac{4}{3} - t\right) + 2\left(t - \frac{2}{3}\right) - t = 0$$

and

$$2\left(\frac{4}{3} - t\right) + \left(t - \frac{2}{3}\right) + t == 2.$$

Thus we have shown that the set of all solutions of the system of linear equations is the set of all 3-tuples

$$\left(\frac{4}{3} - t, t - \frac{2}{3}, t\right),$$

where t is any real number. We can express the set of all such 3-tuples in set builder notation like this:

$$\left\{\left(\frac{4}{3} - t, t - \frac{2}{3}, t\right) \in \mathbb{R}^3 \colon t \in \mathbb{R}\right\}$$

Another way to describe the set of solutions is by writing

$$\begin{aligned} x &= \frac{4}{3} - t \\ y &= t - \frac{2}{3} \\ z &= t \end{aligned}$$

where t is any real number. We call t a *parameter* of the set of solutions.

Example 1.10.b) also demonstrates the following. It is possible for a system of linear equations to have no solutions. We call a system with no solutions “inconsistent.”

Definition 1.12. A system of linear equations is called *consistent* if it has solutions, otherwise it is called *inconsistent*.

Example 1.10 also demonstrates that, if a system of linear equations is consistent, then it is possible that it can have exactly one solution or infinitely many solutions. In fact, these are the only two possibilities! This is our first theorem of linear algebra.

Theorem 1.13. *If a system of linear equations has more than one solution, then it has infinitely many solutions.*

Proof. Suppose we are given an arbitrary system of m linear equations in n variables, which we can write as

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

In order to prove the theorem, we must show that:

If this system has more than one solution, then it has infinitely many solutions.

This is left as an important exercise. See Part 3 of Tutorial B.1. □

To summarize, for a system of linear equations there are only three possible cases:

(a) The system has no solutions (inconsistent).

(b) The system has solutions (consistent). There are two possibilities:

  (i) The system has exactly one solution.

  (ii) The system has infinitely many solutions.

Check your understanding

The equation

$$x^2 = 1$$

has exactly two solutions. Does this contradict Theorem 1.13?

1.3 The augmented matrix of a system

Suppose we have a system of m linear equations in n variables. We might write the equations out like this:

$$\text{System of equations:} \quad \begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & a_{2n}x_n & = & b_2 \\ & & & & \vdots & & & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & a_{mn}x_n & = & b_m \end{array}$$

This takes up a lot of space, and writing it down can take a lot of time.

In the following section, we will develop a systematic way (called Gaussian elimination) to solve any system of linear equations. When performing Gaussian elimination, one must rewrite the full system of linear equations many times.

To save ourselves time, we can condense all the information of a linear system of equations by writing its *augmented matrix*.

$$\text{Augmented matrix:} \quad \left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ & & \vdots & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right].$$

The augmented matrix records the coefficients and constant terms of a system of equations as a rectangular array of numbers with m rows and $n+1$ columns. The augmented matrix has two pieces, which are separated by a vertical line, called the *coefficient matrix* and the *constant matrix*:

$$\text{Coefficient matrix: } A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & & \vdots & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \text{ constant matrix: } \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

We often denote the coefficient matrix by A, the constant matrix by $\mathbf{b}$ and the augmented matrix by $[A|\mathbf{b}]$.

Example 1.14. Consider the system of equations

$$\begin{array}{rcrcrcr} x & + & 2y & - & z & = & 0, \\ 2x & + & y & + & z & = & 2. \end{array}$$

Its augmented matrix, coefficient matrix, and constant matrix are

$$\left[\begin{array}{ccc|c} 1 & 2 & -1 & 0 \\ 2 & 1 & 1 & 2 \end{array}\right], \left[\begin{array}{ccc} 1 & 2 & -1 \\ 2 & 1 & 1 \end{array}\right], \text{ and } \left[\begin{array}{c} 0 \\ 2 \end{array}\right]$$

respectively.

Check your understanding

Match each system of linear equations from Example 1.10 with its augmented matrix below.

i)

$$\left[\begin{array}{cc|c} -1 & 1 & 4 \\ -1 & 1 & -4 \end{array}\right]$$

ii)

$$\left[\begin{array}{cc|c} -2 & 2 & 2 \\ -1 & 1 & 1 \end{array}\right]$$

iii)

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

iv)

$$\left[\begin{array}{cc|c} -1 & 1 & 1 \\ -2 & 2 & 2 \end{array}\right]$$

As we will see later, the augmented matrix of a system of linear equations is just one instance where rectangular arrays of numbers appear in linear algebra. Let's give these objects a name while we are on the topic.

Definition 1.15. A $m \times n$ *matrix*[2] is a rectangular array of numbers and/or variables that has m rows and n columns.

The rows and columns of a matrix are always numbered the same way: the rows are numbered in increasing order from top to bottom, the columns are numbered in increasing order from left to right.

The i, j entry (or ijth entry) of a matrix is the number that is in the ith row and the jth column.

Example 1.16. (a) A matrix with the same number of rows and columns is called a *square matrix*. For instance, the 3×3 matrix

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

is a square matrix. The $2, 3$ entry of this matrix is 1 whereas the $3, 2$ entry is -1. The 1st row of the matrix is $[1\,0\,0]$ whereas the 3rd row is $[0\ -1\,1]$. What is the 2nd column?

(b) A matrix with only one column is called a *column vector*. For instance, the 4×1 matrix

$$\begin{bmatrix} -2 \\ 11 \\ 6 \\ 7 \end{bmatrix}$$

is a column vector. The $3, 1$ entry is 6.

(c) A matrix with only one row is called a *row vector*. For instance, the 1×4 matrix

$$\begin{bmatrix} -2 & 11 & 6 & 7 \end{bmatrix}$$

is a row vector. The $1, 3$ entry is 6.

[2] The plural of matrix is *matrices*.

In-class exercise

If the augmented matrix of a system of linear equations contains a row of the form

$$[0 \cdots 0\, 1],$$

what can be said about the set of solutions of the system of linear equations?

Warning

The augmented matrix of a system of m linear equations in n variables is **NOT** a $m \times n$ matrix!

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ & & \vdots & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right].$$

It is a $m \times (n+1)$ matrix, since there are n columns of coefficients, and one column of constants. The coefficient matrix is a $m \times n$ matrix.

Unlike augmented matrices of systems of equations, most matrices don't have a line drawn through them. The vertical line in an augmented matrix is just decoration.

1.4 Elementary operations

In order to describe a uniform approach to solving systems of linear equations, we first describe three ways to manipulate a system of equations called *elementary operations*.

The three *elementary operations* on a system of linear equations are:

I) Interchanging two equations,

II) Multiplying an equation by a non-zero number, and

III) Adding a multiple of one equation to another equation.

Example 1.17. Let's perform these operations on the system

$$\begin{array}{lrcrcrcl} \text{Eq. 1} & x & + & 2y & - & z & = & 0 \\ \text{Eq. 2} & 2x & + & y & + & z & = & 2 \end{array}$$

I) **Interchanging two equations.** This just means changing the way we listed the equations. For this example, there is only one possible way to interchange two equations:

$$\begin{array}{lrcrcrcl} \text{Eq. 1} & x & + & 2y & - & z & = & 0 \\ \text{Eq. 2} & 2x & + & y & + & z & = & 2 \end{array} \longrightarrow \begin{array}{lrcrcrcl} \text{Eq. 1} & 2x & + & y & + & z & = & 2 \\ \text{Eq. 2} & x & + & 2y & - & z & = & 0 \end{array}$$

This elementary operation affects the augmented matrix in the following way.

$$\left[\begin{array}{ccc|c} 1 & 2 & -1 & 0 \\ 2 & 1 & 1 & 2 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 2 & 1 & 1 & 2 \\ 1 & 2 & -1 & 0 \end{array}\right]$$

II) **Multiplying an equation by a non-zero number.** For example, we can multiply the second equation by 5.

$$\begin{array}{lrcrcrcl} \text{Eq. 1} & x & + & 2y & - & z & = & 0 \\ \text{Eq. 2} & 2x & + & y & + & z & = & 2 \end{array} \rightarrow \begin{array}{lrcrcrcl} \text{Eq. 1} & x & + & 2y & - & z & = & 0 \\ \text{Eq. 2} & 10x & + & 5y & + & 5z & = & 10 \end{array}$$

This elementary operation affects the augmented matrix in the following way.

$$\left[\begin{array}{ccc|c} 1 & 2 & -1 & 0 \\ 2 & 1 & 1 & 2 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 1 & 2 & -1 & 0 \\ 10 & 5 & 5 & 10 \end{array}\right]$$

III) **Adding a multiple of one equation to another equation.** For example, we can add -2 times equation 1 to equation 2.

$$\begin{array}{lrcrcrcl} \text{Eq. 1} & x & + & 2y & - & z & = & 0 \\ \text{Eq. 2} & 2x & + & y & + & z & = & 2 \end{array} \rightarrow \begin{array}{lrcrcrcl} \text{Eq. 1} & x & + & 2y & - & z & = & 0 \\ \text{Eq. 2} & 0x & + & -3y & + & 3z & = & 2 \end{array}$$

Note that this operation only changes equation 2.

Check your understanding

How does the augmented matrix change when you add -2 times equation 1 to equation 2 in the previous example?

Every elementary operation can be reversed by another elementary operation.

Check your understanding

For each elementary operation in the example above, describe the opposite (or inverse) elementary operation.

As we saw in the previous example, for each elementary operation on a system of linear equations, there is a corresponding operation on the rows of the augmented matrix. We call these *elementary row operations*.

In-class exercise

List the three types of *elementary row operations* on a matrix.

Elementary row operations can be performed on any matrix, it does not have to be the augmented matrix of a system of linear equations!

The most important fact about elementary operations on a system of linear equations is that they do not change the set of solutions.

Theorem 1.18. *The set of solutions of a system of linear equations is not changed by elementary operations.*

Proof. Every solution of a system of linear equations is also a solution of the new system of linear equations obtained by performing an elementary operation. We explain why for each type of operation:

I) Interchanging two equations does not change the set of equations, just the order they are listed. Thus, a solution of the original system is also a solution to the new system if two equations are interchanged.

II) A solution of an equation is also a solution of the equation multiplied by a number. Thus, a solution of the original system is also a solution of the new system obtained by multiplying an equation by a non-zero number.

III) A solution of two equations is also a solution of a multiple of one equation added to the other equation. Thus, a solution of the original system is also a solution of the new system obtained by adding a multiple of one equation to another.

Since all of the elementary operations can be reversed by elementary operations, we have also shown that every solution of the new system is a solution of the original system. Thus, the sets of solutions of the two systems are the same. □

In this proof, we explain that the sets of solutions of two different systems (the original system, and a new system obtained by performing an elementary operation) are the same by explaining two things: every solution of the original system is a solution of the new system, and every solution of the new system is a solution of the original system.

This is a common proof technique for showing to sets are the same. We will return to this concept later in the course.

In-class exercise

Where in the proof above did we use the fact that elementary operation II) is multiplication by a *non-zero* number?

1.5 Gaussian elimination

Gaussian elimination is an algorithm[3] for finding all the solutions of a system of linear equations.

A *leading entry* in a row of a matrix is the first entry from the left that is nonzero.

Definition 1.19 (Lay)**.** A matrix is in *row-echelon form* (REF) if it satisfies three conditions.

a) All nonzero rows are above any rows that contain only zeros.

[3]An algorithm is a set of directions for solving a problem that a computer can follow. To learn more about how algorithms and how they are fundamental in modern life, I recommend the documentary *The secret rules of modern living: Algorithms* hosted by Marcus du Sautoy.

b) Each leading entry of a row is in a column to the right of the leading entry in the row above it.

c) All entries in a column that are below a leading entry are zero.

If a matrix in row-echelon form satisfies the following additional conditions, then it is in *reduced row-echelon form* (RREF):

a) The leading entry in each nonzero row is 1 (we call this a 'leading 1').

b) Each leading 1 is the only nonzero entry in its column.

Example 1.20. Both of these matrices are in row-echelon form:

$$\begin{bmatrix} 3 & 3 & 2 & 0 & 5 \\ 0 & 0 & 4 & 6 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The second of these two matrices is in RREF, but the first is not.

These matrices are not in row-echelon form:

$$\begin{bmatrix} 3 & 3 & 2 & 0 & 5 \\ 0 & 0 & 4 & 6 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

Every column in the augmented matrix of a linear system corresponds to a variable (first column corresponds to first variable, and so on). If the augmented matrix of a system of linear equations is in row echelon form, then certain columns contain leading entries while others do not. A variable is called a *leading variable* if it corresponds to a column in the REF matrix that has contains a leading entry. The other variables are the *non-leading variables.* For instance, if the first matrix in Example 1.20 is the augmented matrix of a system in the variables x_1, x_2, x_3, x_4, then x_1, x_3 are leading variables, and x_2, x_4 are non-leading variables.

We now give the following procedure for solving systems of linear equations.

Solving a system of linear equations

Given a system of linear equations, execute the following steps.

1. Write the augmented matrix of the system of linear equations.

2. Use elementary row operations to put the augmented matrix into row-echelon form (this procedure is called Gaussian elimination).

3. Replace all the non-leading variables with parameters and use the equations from the row-echelon form of the augmented matrix to solve for the leading variables in terms of the parameters.

Example 1.21. Solve the following systems of linear equations.

(a)

$$\begin{aligned} x + 2y + z &= 0 \\ -x + y + z &= 5 \\ 2y + z &= 1 \end{aligned}$$

(b)

$$\begin{aligned} 2x + 2y + z &= 4 \\ -x + z &= 0 \\ x + 2y + 2z &= 0 \end{aligned}$$

(c)

$$\begin{aligned} x + 3y + z &= 4 \\ -x - y &= 2 \\ 2y + z &= 6 \end{aligned}$$

See Appendix A for some worked examples.

In-class exercise

Solve the following system of linear equations.

$$\begin{aligned} 3x + 3y + z &= -1 \\ 7x + 5y + 2z &= -1 \\ 4y + 2y + z &= 0 \end{aligned}$$

1.6 The rank of a matrix

It is a fact that if you start with a matrix A – no matter what order you perform elementary row operations – you will always arrive at the same RREF matrix R. If you are interested, there is a proof of this fact at the end of Chapter 3. One might say,

> "All roads (constructed from row operations) lead to the same reduced row echelon form matrix."

Because of this fact, we can make the following definition.

Definition 1.22. The *rank* of a matrix A is the number of leading entries (or pivots) in any row-echelon form of the matrix. We denote this number by $\operatorname{rank}(A)$.

Example 1.23. In Example 1.17 showed that the matrix

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

can be carried by a row operation to the matrix

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

This new matrix is in REF and has two leading entries. Thus the rank of the original matrix A is 2. We write $\operatorname{rank}(A) = 2$.

Check your understanding

What are the ranks of the matrices in Example 1.20?

Suppose A is a $m \times n$ matrix. Since every leading entry in the REF of A occurs in its own column, and there are n columns, $\operatorname{rank}(A)$ must be less than or equal to n. Since every leading 1 occurs in its own row, and there are m rows, $\operatorname{rank}(A)$ must be less than or equal to m. We can summarize this with two equations (valid for every $m \times n$ matrix A):

$$0 \leq \operatorname{rank}(A) \leq m,$$

$$0 \leq \operatorname{rank}(A) \leq n.$$

Check your understanding

Suppose $[A|\mathbf{b}]$ is the augmented matrix of a system of 3 linear equations in 4 variables.

What are the possible values of $\mathrm{rank}(A)$?

What are the possible values of $\mathrm{rank}([A|\mathbf{b}])$?

In-class exercise

Suppose $[A|\mathbf{b}]$ is the augmented matrix of a system of m linear equations in n variables. If we know that $\mathrm{rank}([A|\mathbf{b}]) = n+1$, what can we say about the solutions of the system of equations?

Recall that every consistent system of linear equations has exactly one solution or infinitely many solutions. Rank is useful because it helps us to distinguish between these two cases.

Theorem 1.24. *Suppose $[A|\mathbf{b}]$ is the augmented matrix of a system of m linear equations in n variables and let r denote the rank of $[A|\mathbf{b}]$. If the system is consistent, then*

a) If $r = n$, the system has exactly one solution.

b) If $r < n$, the system has infinitely many solutions.

Proof. If $r = n$, then every variable is a leading variable, so there is exactly one solution. If $r < n$, then there is at least one non-leading variable, so the set of solutions has at least one parameter, so there are infinitely many solutions. $\square$

Example 1.25. (a) The augmented matrix

$$\left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 0 & 1 \end{array}\right]$$

has rank 2. The corresponding system of linear equations is inconsistent.

(b) The augmented matrix

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

has rank 2. The corresponding system of linear equations is consistent and has exactly one solution.

(c) The augmented matrix

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

has rank 1. The corresponding system of linear equations is consistent and has infinitely many solutions.

(d) The augmented matrix

$$\left[\begin{array}{cc|c} 0 & 0 & 1 \\ 0 & 0 & 0 \end{array}\right]$$

has rank 1. The corresponding system of linear equations is inconsistent.

1.7 Discussion: linear and nonlinear problems in science and mathematics

The most important lesson from this chapter is the following: **every system of linear equations can be solved exactly by Gaussian elimination.** Moreover, Gaussian elimination is a completely algorithmic process: if you give a computer the right general instructions, it will be able to solve any system of linear equations without any thought.

In contrast, equations that are not linear (aka "non-linear" equations) can be much harder, or impossible, to solve precisely. For instance, there is no way to find an exact formula for the simple looking equation

$$1 = \sin(x) + 2x$$

even though one can prove using calculus that a solution does exist.

This leads us to two useful meta-principals for doing math and science:

(a) Problems can often be usefully categorized as **linear** (you can describe the problem using linear equations) or **non-linear** (you cannot describe the problem using linear equations). Roughly, if a problem is linear, then it is very easy from a mathematical perspective. If a problem is non-linear, then it is often a bit trickier or very hard. Examples of non-linear problems include: modelling the shape of a person's face for computer animation or describing the motion of water.

(b) When presented with a non-linear problem, a general strategy is to instead consider a linear problem that is approximately the same, i.e. a **linear**

approximation. For instance, the solution to the equation above can be computed using Newton's method, which relies on repeated linear approximations of the graph of the function $f(x) = \sin(x) + 2x - 1$. A person's face can be approximately modelled by a 3D polygonal mesh, which is just a bunch of solutions of linear equations pieced together.

In order to tackle the difficult non-linear problems, we must first learn to tackle the "easy" linear ones.

Key concepts from Chapter 1

Linear and non-linear equations, systems of linear equations, solutions of systems of equations, the augmented matrix of a system of linear equations, the coefficient matrix of a system of linear equations, the definition of a matrix, elementary row operations do not change the set of solutions of a system of linear equations, Gaussian elimination, row-echelon form, the reduced row-echelon form of a matrix, the rank of a matrix, the possible number of solutions of a system of linear equations.

Chapter 2

Matrix algebra

In highschool algebra, you learned about operations on real numbers (addition, multiplication and division) and the properties of these operations. For example, you may have learned that for any numbers x, y, and z,

$$xy = yx$$

and

$$x(y+z) = xy + xz.$$

These properties are the rules of basic algebra.

These properties are useful in many ways. For instance, one can use them to simplify algebraic expressions using (what some people refer to as) BEDMAS rules for the order of evaluating an expression (brackets, exponents, division, multiplication, addition, subtraction). If a and b are numbers, then we can simplify an expression such as:

$$\begin{aligned} 5(3a + 2b + (6a - a)) &= 5(3a + 2b + 5a) \\ &= 5(8a + 2b) \\ &= 40a + 10b. \end{aligned}$$

Recall from the previous chapter that a $m \times n$ matrix is a rectangular array of numbers and/or variables that has m rows and n columns.

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

We have already seen how matrices are useful in studying solutions to systems of linear equations: the augmented matrix is a nice shorthand way to record the information of a system of linear equations when performing Gaussian elimination.

In fact, the usefulness of matrices in studying systems of linear equations is much greater. One can define various operations for matrices (analogous to operations like multiplication and addition of numbers), and these operations are intimately related to the study of systems of linear equations.

In this chapter, we introduce operations on matrices. As in highschool algebra, operations on matrices have various properties. The rules for these operations are matrix algebra.

2.1 Equality of matrices

Before defining operations on matrices, we should carefully explain what it means for two matrices to be equal.

Two $m \times n$ matrices

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

are *equal* if all of their entries are equal. This means that

$$a_{ij} = b_{ij}$$

for every $1 \leq i \leq n$ and $1 \leq j \leq m$. If they are equal, then we write $A = B$.

Example 2.1. For example, the equality of two matrices

$$\begin{bmatrix} x - 2y \\ 5x^2 + 6yx - z \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

is equivalent to the system of equations

$$\begin{aligned} x - 2y &= 3 \\ 5x^2 + 6yx - z &= 4 \end{aligned}$$

Thus a matrix equation encodes a system of equations!

Check your understanding

Consider the matrices

$$A = \begin{bmatrix} 1 & 0 & x \\ 2 & 6 & x^2 \end{bmatrix}, \; B = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 6 & 1 \end{bmatrix}, \; C = \begin{bmatrix} 1 & 0 \\ 2 & 6 \end{bmatrix}$$

For what values of x is $A = B$? For what values of x is $A = C$?

2.2 Matrix addition and scalar multiplication

The *sum* of two $m \times n$ matrices

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

is the $m \times n$ matrix

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

In other words, the ij entry of $A + B$ is $a_{ij} + b_{ij}$, the sum of the ij entry of A and the ij entry of B. Another name for this operation is *matrix addition*.

Example 2.2. For example,

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

- The formula for the sum of two arbitrary 2×2 matrices is

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{bmatrix}$$

Example 2.3. Let $0_{m\times n}$ denote the $m \times n$ matrix whose entries are all zero. We call this the *zero matrix*. The matrix $0_{m\times n}$ has the special property that for any other $m \times n$ matrix A,

$$A + 0_{m\times n} = 0_{m\times n} + A = A.$$

The *scalar multiplication* of real number (or variable) k and a $m \times n$ matrix

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

is the $m \times n$ matrix

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

In other words, the entries of kA are simply k times the entries of A, ka_{ij}.

Check your understanding

Let $\mathbf{u} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. Compute the vector

$$\mathbf{u} - 3\mathbf{v}.$$

Properties of matrix addition and scalar multiplication

If A and B are $m \times n$ matrices and k, l are real numbers. Then,

(a) (scalar multiplication distributes over matrix addition)

$$k(A + B) = kA + kB$$

(b) (scalar multiplication distributes over scalar addition)

$$(k + l)A = kA + lA$$

(c)
$$(kl)A = k(lA)$$

(d) (matrix addition is commutative)
$$A + B = B + A$$

(e) (matrix addition is associative)
$$A + (B + C) = (A + B) + C$$

Example 2.4. These properties can be used to manipulate expressions involving matrix addition and scalar multiplication. For instance, if A and B are $m \times n$ matrices, then (being careful about order of operations) we can simplify the following expression:

$$\begin{aligned} 5(3A + 2B + (6A - A)) &= 5(3A + 2B + 5A) \\ &= 5(3A + 5A + 2B) \\ &= 5(8A + 2B) \\ &= 40A + 10B. \end{aligned}$$

This is similar to the example from the introduction to the chapter, but it is important to understand that the objects and operations here are different. It is important to check that the properties we are using are valid.

Let's see how we could prove the first property. The others are left as exercises. There are several different ways to describe the proof of the first property.

Proof 1. We want to show two matrices, $k(A + B)$ and $kA + kB$ are equal, so we must explain why each of their entries are equal. Let's explain why their ij entries are equal in a way that applies to any $1 \leq i \leq m$ and $1 \leq j \leq n$.

The matrix $k(A+B)$ is the matrix obtained by first adding $A+B$, then scalar multiplying by k. The ij entry of this matrix is thus $k(a_{ij} + b_{ij})$.

The matrix $kA+kB$ is the matrix obtained by first scalar multiplying kA and kB, then adding the results together. The ij entry of kA is ka_{ij} and the ij entry of kB is kb_{ij}. Thus the ij entry of $kA + kB$ is $ka_{ij} + kb_{ij}$.

Since $k(a_{ij} + b_{ij}) = ka_{ij} + kb_{ij}$ by the distributive property for numbers, the ij entries of the two matrices are equal. Thus

$$k(A + B) = kA + kB.$$

□

Phew. That was pretty wordy! Let's try to make it more succinct. Let's introduce the following shorthand. We write $A = [a_{ij}]_{m\times n}$ to mean the matrix A equals the $m \times n$ matrix with entries a_{ij}. The proof above can then be written more briefly as follows.

Proof 2. Suppose $A = [a_{ij}]_{m\times n}$ and $B = [b_{ij}]_{m\times n}$. Then by definition of matrix addition and scalar multiplication,

$$k(A+B) = k\left([a_{ij}]_{m\times n} + [b_{ij}]_{m\times n}\right) = k[a_{ij}+b_{ij}]_{m\times n} = [k(a_{ij}+b_{ij})]_{m\times n}$$

and

$$kA + kB = k[a_{ij}]_{m\times n} + k[b_{ij}]_{m\times n} = [ka_{ij}]_{m\times n} + [kb_{ij}]_{m\times n} = [ka_{ij}+kb_{ij}]_{m\times n}$$

Since $k(a_{ij}+b_{ij}) = ka_{ij} + kb_{ij}$,

$$[k(a_{ij}+b_{ij})]_{m\times n} = [ka_{ij}+kb_{ij}]_{m\times n}.$$

Thus,

$$k(A+B) = kA + kB.$$

□

We can make it even shorter.

Proof 3. Let $A = [a_{ij}]$ and $B = [b_{ij}]$. Let's drop the $[]_{m\times n}$ from the notation, since we know all the matrices are $m \times n$.

In this proof we start with $k(A+B)$ and manipulate the formulas using definitions and properties of real numbers until we get to $kA + kB$.

$$\begin{aligned}
k(A+B) &= k\left([a_{ij}] + [b_{ij}]\right) \\
&= k[a_{ij}+b_{ij}] && \text{By definition of matrix addition.} \\
&= [k(a_{ij}+b_{ij})] && \text{By definition of scalar multiplication.} \\
&= [ka_{ij}+kb_{ij}] && \text{Since multiplication distributes over addition.} \\
&= [ka_{ij}] + [kb_{ij}] && \text{By definition of matrix addition.} \\
&= k[a_{ij}] + k[b_{ij}] && \text{By definition of scalar multiplication.} \\
&= kA + kB
\end{aligned}$$

□

2.3 Matrix-vector products

Given a $m \times n$ matrix

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

and a $n \times 1$ column vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

the *matrix-vector product* of A and $\mathbf{x}$ is the $m \times 1$ column vector

$$A\mathbf{x} := \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix}$$

Warning

The matrix-vector product $A\mathbf{x}$ is only defined if the number of columns of A equals the number of rows of $\mathbf{x}$. Otherwise writing $A\mathbf{x}$ has no meaning!

Example 2.5. Let's compute a matrix vector product for practice.

$$\begin{bmatrix} 1 & 0 & 4 \\ 2 & 6 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \cdot 1 + 0 \cdot (-1) + 4 \cdot 2 \\ 2 \cdot 1 + 6 \cdot (-1) + (-1) \cdot 2 \end{bmatrix} = \begin{bmatrix} 9 \\ -6 \end{bmatrix}$$

Example 2.6. If $\mathbf{a} = [a_1\, a_2\, \ldots\, a_n]$ is a $1 \times n$ row vector and $\mathbf{x}$ is the $n \times 1$ column vector above, then the matrix-vector product

$$\mathbf{ax} = [a_1\, a_2\, \ldots\, a_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = [a_1x_1 + a_2x_2 + \cdots + a_nx_n]\,.$$

The result on the right hand side is a 1×1 matrix. We often think of a 1×1 matrix as a real number and omit the brackets.

There are two useful ways to think about matrix-vector products. Suppose A is a $m \times n$ matrix as above.

- Suppose that $\mathbf{a}_i = [a_{i1}\, a_{i2}\, \dots\, a_{in}]$ are the rows of A. The formula for matrix vector products and the computation in the previous example tell us that
$$A\mathbf{x} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1\mathbf{x} \\ \mathbf{a}_2\mathbf{x} \\ \vdots \\ \mathbf{a}_m\mathbf{x} \end{bmatrix}$$

- Suppose that
$$\mathbf{c}_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}$$
are the columns of A. Using sums and scalar multiples we can rewrite the matrix-vector product
$$\begin{aligned} A\mathbf{x} &= \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix} \\ &= \begin{bmatrix} a_{11}x_1 \\ a_{21}x_1 \\ \vdots \\ a_{m1}x_1 \end{bmatrix} + \begin{bmatrix} a_{12}x_2 \\ a_{22}x_2 \\ \vdots \\ a_{m2}x_2 \end{bmatrix} + \cdots + \begin{bmatrix} a_{1n}x_n \\ a_{2n}x_n \\ \vdots \\ a_{mn}x_n \end{bmatrix} \\ &= x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} \\ &= x_1\mathbf{c}_1 + x_2\mathbf{c}_2 + \cdots + x_n\mathbf{c}_n. \end{aligned}$$

Properties of matrix-vector multiplication

Suppose that A and B are $m \times n$ matrices, k and l are real numbers, and $\mathbf{u}$ and $\mathbf{v}$ are $n \times 1$ column vectors. Then,

(a)
$$(A+B)\mathbf{u} = A\mathbf{u} + B\mathbf{u}$$

(b)
$$A(\mathbf{u}+\mathbf{v}) = A\mathbf{u} + A\mathbf{v}$$

(c)
$$k(A\mathbf{u}) = A(k\mathbf{u})$$

Proof. We prove the second property, the others are exercises.

Using the second interpretation of the matrix-vector product above, we see that

$$\begin{aligned} A(\mathbf{u}+\mathbf{v}) &= A \begin{bmatrix} u_1 + v_1 \\ \vdots \\ u_n + v_n \end{bmatrix} \\ &= (u_1+v_1)\mathbf{a}_1 + \cdots (u_n+v_n)\mathbf{a}_n \\ &= (u_1\mathbf{a}_1 + \cdots + u_n\mathbf{a}_n) + \cdots + (v_1\mathbf{a}_1 + \cdots v_n\mathbf{a}_n) \\ &= A\mathbf{u} + A\mathbf{v}. \end{aligned}$$

□

In-class exercise

Let A and B be $m \times n$ matrices, and let $\mathbf{u}$ and $\mathbf{v}$ be $n \times 1$ column vectors. Simplify the expression.

$$(A+B)(\mathbf{u}+\mathbf{v}) - B\mathbf{u} + 2A\mathbf{v}$$

2.4 Matrix multiplication

Given a $m \times n$ matrix

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

and a $n \times p$ matrix

$$B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ & & \vdots & \\ b_{n1} & b_{n2} & \cdots & b_{np} \end{bmatrix} = \begin{bmatrix} \mathbf{b}_{11} & \mathbf{b}_{12} & \cdots & \mathbf{b}_{1p} \end{bmatrix}$$

where

$$\mathbf{b}_j = \begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{bmatrix}$$

are the columns of B, the *product*, A *times* B, is the $m \times p$ matrix AB whose jth column is the matrix vector product $A\mathbf{b}_j$. Explicitly,

$$AB := \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & \cdots & A\mathbf{b}_p \end{bmatrix}$$

The ijth entry of this matrix is the ith entry of the matrix-vector product $A\mathbf{b}_j$, which equals

$$\sum_{k=1}^{n} a_{ik}b_{kj} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}.$$

In-class exercise

Suppose

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

Compute the product AB.

Warning

The product AB is only defined if the number of columns of A equals the number of rows of B. Otherwise writing AB has no meaning!

For example, in the preceding exercise, we can compute AB, but not BA, since B has three columns and A has two rows.

Example 2.7. The $n \times n$ *identity matrix* is the matrix whose diagonal entries are 1, and all of whose other entries are zero. We write I_n to denote the $n \times n$ identity matrix. Often the number n is implied by context, so we simply write I. For example, the 2×2 and 3×3 identity matrices are

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The name for this matrix comes from the fact that matrix multiplication by I has no effect. For instance, if B is the matrix from the preceding exercise, then

$$I_2 B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 2 & 1 \\ -1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 2 & 1 \\ -1 & 1 & 0 \end{bmatrix} = B.$$

and

$$BI_3 = \begin{bmatrix} 0 & 2 & 1 \\ -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 2 & 1 \\ -1 & 1 & 0 \end{bmatrix} = B.$$

Check your understanding

What is the product $0_{1\times 2}B$? Be careful, this is not scalar multiplication!

Problem 1. *A square matrix A is* diagonal *if the entries $a_{ij} = 0$ when $i \neq j$. Show that if A and B are $n \times n$ diagonal matrices, then the matrix-product AB is also diagonal.*

Solution. *We need to show that when $i \neq j$, the ij entry of AB is zero. We know that the ij entry of AB is given by the formula*

$$a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}$$

where $[a_{i1}\, a_{i2}\, \cdots\, a_{in}]$ *is the ith row of* A *and*

$$\begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{bmatrix}$$

is the jth column of B.

- *Since* A *is diagonal, all the entries in the ith row of* A *are zero except* a_{ii}.
- *Since* B *is diagonal, all the entries in the jth row of* B *are zero except* b_{jj}.

Thus the formula for the ij *entry of* AB *looks like*

$$0 + \cdots + 0 + a_{ii}b_{ij} + 0 + \cdots + 0 + a_{ij}b_{jj} + 0 + \cdots + 0$$

if $i \neq j$. *But* $b_{ij} = 0$ *and* $a_{ij} = 0$, *so the* ij *entry is*

$$0 + \cdots + 0 + a_{ii}0 + 0 + \cdots + 0 + 0b_{jj} + 0 + \cdots + 0 = 0. \quad \blacklozenge$$

Properties of matrix multiplication

Suppose that A, B and C matrices. Then, whenever the operations below are defined,

(a) (left distributivity)

$$A(B + C) = AB + AC$$

(b) (right distributivity)

$$(A + B)C = AC + BC$$

(c)

$$k(AB) = (kA)B = A(kB)$$

(d) (associativity)

$$A(BC) = (AB)C$$

Proof. (a) Let $\mathbf{b}_j$ denote the columns of B and let $\mathbf{c}_j$ denote the columns of B. Then,

$$\begin{aligned}
A(B+C) &= A[\mathbf{b}_1 + \mathbf{c}_1 \ \cdots \ \mathbf{b}_p + \mathbf{c}_p] \\
&= [A(\mathbf{b}_1 + \mathbf{c}_1) \ \cdots \ A(\mathbf{b}_p + \mathbf{c}_p)] \\
&= [A\mathbf{b}_1 + A\mathbf{c}_1 \ \cdots \ A\mathbf{b}_p + A\mathbf{c}_p] \qquad \text{By prop. of mat-vect mult.} \\
&= [A\mathbf{b}_1 \ \cdots \ A\mathbf{b}_p] + [A\mathbf{c}_1 \ \cdots \ A\mathbf{c}_p] \\
&= A[\mathbf{b}_1 \ \cdots \ \mathbf{b}_p] + A[\mathbf{c}_1 \ \cdots \ \mathbf{c}_p] \\
&= AB + AC.
\end{aligned}$$

(b) Exercise.

(c) Exercise.

(d) Unravelling the definitions, we see that the ij entry of $(AB)D$ is

$$\sum_{l=1}^{p} \left(\sum_{k=1}^{n} a_{ik} b_{kl} \right) d_{lj}$$

whereas the ij entry of $A(BD)$ is

$$\sum_{k=1}^{n} a_{ik} \left(\sum_{l=1}^{p} b_{kl} d_{lj} \right).$$

These two sums are equal, so the ij entries are equal.

□

Because of property (d), we know that for any square matrix A, $A(AA) = (AA)A$. Thus it is ok to simply write AAA because the order in which you multiply the matrices doesn't matter. We write $A^3 = AAA$ for short. In general, we use the notation A^k to mean the product of A by itself k-times.

$$A^k = A \cdot A \cdots A.$$

2.5 Matrix transpose

The *transpose* of a $m \times n$ matrix

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

is the $n \times m$ matrix

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

In plain english, the transpose of A is the matrix A^T whose rows are the columns of A and whose columns are the rows of A.

This is best illustrated with examples.

Example 2.8. Consider the matrices

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 6 & 3 \end{bmatrix}, \; B = \begin{bmatrix} 0 & 2 & 9 \\ 1 & 5 & \pi \\ 6 & 8 & e \end{bmatrix}, \; \mathbf{u} = \begin{bmatrix} 0 \\ -1 \\ 5 \end{bmatrix},$$

The definition of matrix transpose tells us that

$$A^T = \begin{bmatrix} 1 & 1 & 6 \\ 2 & 5 & 3 \end{bmatrix}, \; B^T = \begin{bmatrix} 0 & 1 & 6 \\ 2 & 5 & 8 \\ 9 & \pi & e \end{bmatrix}, \; \mathbf{u}^T = \begin{bmatrix} 0 & -1 & 5 \end{bmatrix},$$

Remark 2.9. Observe that

- The transpose of a square matrix is a square matrix.
- The transpose of a column vector (a $m \times 1$ matrix) is a row vector (a $1 \times m$ matrix).
- The transpose of a row vector (a $1 \times m$ matrix) is a column vector (a $m \times 1$ matrix).

- Since the transpose A^T of a $m \times n$ matrix A is a $n \times m$ matrix, one can always compute the matrix products

$$AA^T \text{ and } A^TA.$$

Check your understanding

If A is a $n \times m$ matrix, how many rows and columns do the matrices AA^T and A^TA have?

In-class exercise

Suppose A is a $n \times m$ matrix and A^T is its transpose. What is the transpose of A^T?

Properties of matrix transpose

Suppose that A and B are $m \times n$ matrices, C is a $n \times p$ matrix, and k is a number. Then,

(a)
$$(A+B)^T = A^T + B^T$$

(b)
$$(AC)^T = C^TA^T$$

(c)
$$kA^T = (kA)^T$$

(d)
$$(A^T)^T = A$$

Proof. (a) We write our explanation using our shorthand for matrices.

$$\begin{aligned}(A+B)^T &= [a_{ij}+b_{ij}]^T_{m\times n} \\ &= [a_{ji}+b_{ji}]_{n\times m} \\ &= [a_{ji}]_{n\times m} + [b_{ji}]_{n\times m} \\ &= A^T + B^T.\end{aligned}$$

(b) Exercise.

(c) Exercise.

(d) Using matrix shorthand, the proof is very brief:

$$(A^T)^T = \left([a_{ij}]^T\right)^T = ([a_{ji}])^T = [a_{ji}]^T = [a_{ij}] = A$$

□

In-class exercise

Explain each step in the proof of property (a).

2.6 Matrix equations

Now that we have defined what it means for two matrices to be equal, and we have introduced various operations on matrices, we can study *matrix equations.*

Example 2.10. Suppose

$$A = \begin{bmatrix} 1 & -1 \\ 0 & 4 \end{bmatrix}, \text{ and } B = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$$

and

$$X = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}$$

where x_{ij} are variables. The matrix equation

$$A + 2X = B$$

can be written explicitly as the equality of matrices

$$\begin{bmatrix} 1 + 2x_{11} & -1 + 2x_{12} \\ 0 + 2x_{21} & 4 + 2x_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$$

Since two matrices are equal when their entries are equal, this equation is true when the variables x_{ij} satisfy the system of equations

$$\begin{aligned} 1 + 2x_{11} &= 1 \\ -1 + 2x_{12} &= 0 \\ 0 + 2x_{21} &= 0 \\ 4 + 2x_{22} &= 3 \end{aligned}$$

Thus, the problem of solving this particular matrix equation reduces to a system of linear equations.

Now suppose that

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

are matrices whose entries are the constants, a_{ij} and b_i, and

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

is a $n \times 1$ column vector whose entries are variables $x_1, \ldots, x_n$.

In-class exercise

Use the matrix vector product to expand the matrix-vector equation

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

What is the resulting system of equations?

Which column vectors $\mathbf{x}$ solve the matrix equation?

Note that matrix equations can also define a system of equations that is not linear. See Tutorial 3.

2.7 Discussion: matrix multiplication is not commutative

As we saw in this chapter, matrix multiplication is not a commutative operation: we don't always have that $AB = BA$. This is actually a big deal.... (add discussion later)

Key concepts from Chapter 2

Equality of matrices, matrix addition and scalar multiplication, matrix-vector multiplication, matrix multiplication, matrix transpose, all the properties of these operations. Matrix equations. Writing a system of linear equations as a matrix equation

Chapter 3

More systems of linear equations

In Chapter 1 we introduced systems of linear equations and in Chapter 2 we introduced operations on matrices: addition, scalar multiplication, and multiplication. In this chapter we show how matrix algebra can be used to study systems of linear equations, and vis versa.

3.1 The matrix equation of a system of linear equations

Recall that solutions of a system of m linear equations in n variables

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ & & & & \vdots & & & & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array} \tag{3.1}$$

are n-tuples of numbers $(s_1, \ldots, s_n)$ that solve every equation in the system. i.e.

$$a_{i1}s_1 + a_{i2}s_2 + \cdots + a_{in}s_n = b_i$$

for every $1 \leq i \leq m$.

Now suppose that

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

are the coefficient matrix and constant matrix of the system, and let

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

be the $n \times 1$ column vector whose entries are the variables $x_1, \ldots, x_n$.

In Section 2.6, we discovered that a vector

$$\mathbf{s} = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix}$$

solves the *matrix equation of the system* (3.1),

$$A\mathbf{x} = \mathbf{b} \tag{3.2}$$

if and only if, the n-tuple $(s_1, \ldots, s_n)$ solves the system of linear equations (3.1).

In other words, the problem of solving the system of linear equations, (3.1) and the problem of solving the matrix equation (3.2) are one and the same!

For this reason, it is common to refer to the equation $A\mathbf{x} = \mathbf{b}$ as a "system of linear equations" even though it is a matrix equation. It <u>encodes</u> a system of linear equations.

As a consequence of this fact, we can use all the nice operations of matrix algebra to study (and prove facts about) systems of linear equations and vis versa! This is one of the main themes of this chapter.

Check your understanding

Write the matrix equation corresponding to the system of linear equations

$$\begin{array}{rcrcrcr} x & + & 2y & - & z & = & 0 \\ 2x & + & y & + & z & = & 2 \end{array}$$

3.2 Span and systems of linear equations

Recall that the matrix vector product $A\mathbf{x}$ can be expanded as

$$A\mathbf{x} = x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}.$$

Thus we see that

The following two statements are equivalent:

- The system $A\mathbf{x} = \mathbf{b}$ is consistent.
- There are numbers $t_1, \ldots, t_n$ such that

$$t_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + t_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + t_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

For the remainder of this section, we denote the column vectors of A by

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

Definition 3.3. Any expression of the form

$$t_1\mathbf{a}_1 + t_2\mathbf{a}_2 + \cdots + t_n\mathbf{a}_n$$

where $t_i \in \mathbb{R}$ and $\mathbf{a}_j$ are column vectors is called a *linear combination* of the column vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$.

Check your understanding

Let $\mathbf{u} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. Can the vector

$$\begin{bmatrix} -2 \\ 6 \end{bmatrix}$$

be written as a linear combination of $\mathbf{u}$ and $\mathbf{v}$?

Definition 3.4. The *span* of a set of vectors, $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$, is the set of all linear combinations of the vectors. We write

$$\operatorname{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} := \{t_1\mathbf{a}_1 + t_2\mathbf{a}_2 + \cdots + t_n\mathbf{a}_n \colon t_1, \ldots, t_n \in \mathbb{R}\}.$$

Example 3.5. Let $\mathbf{a} = \begin{bmatrix} 2 \\ 0 \\ 4 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$. Is it true that

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \in \operatorname{span}\{\mathbf{a}, \mathbf{b}\}?$$

Example 3.6. The column vector $0_{m\times 1}$ is contained in the span of any set of vectors $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$. Indeed, the linear combination

$$0\mathbf{a}_1 + 0\mathbf{a}_2 + \cdots + 0\mathbf{a}_n = 0_{m\times 1}.$$

Example 3.7. Suppose $\mathbf{u}$ and $\mathbf{v}$ are any $m \times 1$ column vectors, then

1. $\mathbf{u} \in \operatorname{span}\{\mathbf{u}, \mathbf{v}\}$ since
$$\mathbf{u} = 1\mathbf{u} + 0\mathbf{v}.$$

2. $\mathbf{u} \in \operatorname{span}\{\mathbf{u} - \mathbf{v}, \mathbf{u} + \mathbf{v}\}$ since
$$\mathbf{u} = \frac{1}{2}(\mathbf{u} - \mathbf{v}) + \frac{1}{2}(\mathbf{u} + \mathbf{v}).$$

With this terminology, we can rephrase our earlier result:

The following two statements are equivalent:

- The system $A\mathbf{x} = \mathbf{b}$ is consistent.
- $\mathbf{b} \in \operatorname{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$, where $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are the columns of A.

In-class exercise

Use what we know about solving systems of linear equations to determine whether

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \in \operatorname{span}\left\{ \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} \right\}.$$

3.3 Homogeneous systems of linear equations

A system of linear equations is called *homogeneous* if all the constant terms are zero:

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & a_{1n}x_n & = & 0 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & a_{2n}x_n & = & 0 \\ & & & & \vdots & & & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & a_{mn}x_n & = & 0 \end{array}$$

The augmented matrix of a homogeneous system has a column of zeros:

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & 0 \\ a_{21} & a_{22} & \cdots & a_{2n} & 0 \\ & & \vdots & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} & 0 \end{array}\right]$$

The matrix equation of a homogeneous system is

$$A\mathbf{x} = \mathbf{0}$$

where $\mathbf{0} = 0_{m\times 1}$ is the $m \times 1$ column vector of zeros.

Homogeneous systems of linear equations are important for several reasons. The first reason is the following fact:

Every homogeneous system of linear equations is consistent.

Just as important as this fact is the reason why it's true: every homogeneous system has the *trivial solution*, $x_1 = 0, x_2 = 0, \ldots, x_n = 0$.

In-class exercise

Consider the following statement:

If a homogeneous system of linear equations has more variables than equations, then it must have infinitely many solutions.

Why is this true? (think back to Chapter 1)

If you remove the word "homogeneous", then the statement is false. Why?

3.4 Linear combinations and basic solutions

If $\mathbf{u}$ and $\mathbf{v}$ are two solutions of a homogeneous matrix equation $A\mathbf{x} = \mathbf{0}$, then we can use matrix algebra to show that $\mathbf{u} + \mathbf{v}$ is also a solution:

$$A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v} = \mathbf{0} + \mathbf{0} = \mathbf{0}.$$

In fact, even more is true. If t and r are any real numbers, then we can show that $t\mathbf{u} + r\mathbf{v}$ is also a solution:

$$A(t\mathbf{u} + r\mathbf{v}) = A(t\mathbf{u}) + A(r\mathbf{v}) = t(A\mathbf{u}) + r(A\mathbf{v}) = t\mathbf{0} + r\mathbf{0} = \mathbf{0}.$$

While we are at it, why stop at only two solutions? If we have k solutions to the equation, $\mathbf{u}_1, \ldots, \mathbf{u}_k$, and k real numbers $t_1, \ldots, t_k$, then the vector

$$t_1\mathbf{u}_1 + t_2\mathbf{u}_2 + \ldots + t_k\mathbf{u}_k$$

is also a solution! We use the same matrix algebra steps as before:

$$\begin{aligned} A\left(t_1\mathbf{u}_1 + t_2\mathbf{u}_2 + \ldots + t_k\mathbf{u}_k\right) &= A\left(t_1\mathbf{u}_1\right) + A\left(t_2\mathbf{u}_2\right) + \ldots + A\left(t_k\mathbf{u}_k\right) \\ &= t_1A\mathbf{u}_1 + t_2A\mathbf{u}_2 + \cdots + t_kA\mathbf{u}_k \\ &= t_1\mathbf{0} + t_2\mathbf{0} + \cdots + t_k\mathbf{0} \\ &= \mathbf{0} \end{aligned}$$

The expression

$$t_1\mathbf{u}_1 + t_2\mathbf{u}_2 + \ldots + t_k\mathbf{u}_k$$

is called a *linear combination* of the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$. In the preceding calculation, we have proven the following fact:

Every linear combination of solutions to a homogeneous system of linear equations is a solution.

Check your understanding

The fact that the matrix equation/system is homogeneous is crucial. Indeed, suppose that we have a non-homogeneous matrix equation

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

where $\mathbf{b}$ is non-zero.

Suppose that $\mathbf{u}$ and $\mathbf{v}$ are solutions to this matrix equation. Use matrix algebra to simplify the expression

$$A(\mathbf{u} + \mathbf{v}).$$

Is $\mathbf{u} + \mathbf{v}$ a solution of the matrix equation?

Example 3.8. Solve the homogeneous system of linear equations

$$\begin{array}{rcrcrcrcl} x_1 & - & 2x_2 & + & 3x_3 & - & 2x_4 & = & 0 \\ -3x_1 & + & 6x_2 & + & x_3 & & & = & 0 \\ -2x_1 & + & 4x_2 & + & 4x_3 & - & 2x_4 & = & 0 \end{array}$$

Follow the steps from Section 1.5

Step 1. We write down the augmented matrix of the system,

$$\left[\begin{array}{cccc|c} 1 & -2 & 3 & -2 & 0 \\ -3 & 6 & 1 & 0 & 0 \\ -2 & 4 & 4 & -2 & 0 \end{array}\right]$$

Step 2. Perform Gaussian elimination and put the augmented matrix into reduced row-echelon form.

$$\left[\begin{array}{cccc|c} 1 & -2 & 3 & -2 & 0 \\ -3 & 6 & 1 & 0 & 0 \\ -2 & 4 & 4 & -2 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{cccc|c} 1 & -2 & 0 & -1/5 & 0 \\ 0 & 0 & 1 & -3/5 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Step 3. Introduce parameters $x_2 = t_1$ and $x_4 = t_2$ for the non-leading variables. From the reduced row-echelon form, we read the equations

$$\begin{aligned} x_1 &= 2t_1 + \frac{1}{5}t_2 \\ x_2 &= t_1 \\ x_3 &= \frac{3}{5}t_2 \\ x_4 &= t_2 \end{aligned}$$

We can write these equations using sums and scalar multiples of column vectors:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 2t_1 + \frac{1}{5}t_2 \\ t_1 \\ \frac{3}{5}t_2 \\ t_2 \end{bmatrix} = t_1 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t_2 \begin{bmatrix} \frac{1}{5} \\ 0 \\ \frac{3}{5} \\ 1 \end{bmatrix}$$

Thus, the set of all solutions to the linear system above can be described as the set of all vectors

$$t_1 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t_2 \begin{bmatrix} \frac{1}{5} \\ 0 \\ \frac{3}{5} \\ 1 \end{bmatrix}$$

where t_1 and t_2 are any real numbers. (you can also describe the solutions as n-tuples). In set-builder notation, we write

$$\left\{ t_1 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t_2 \begin{bmatrix} \frac{1}{5} \\ 0 \\ \frac{3}{5} \\ 1 \end{bmatrix} : t_1, t_2 \in \mathbb{R} \right\}$$

Definition 3.9. Given a system of linear equations whose augmented matrix is in reduced row-echelon form, the *basic solutions* are the vectors obtained by setting all of the free variables to be zero except for one.

In the previous example, the basic solutions to the system of linear equations are the column vectors

$$\begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{1}{5} \\ 0 \\ \frac{3}{5} \\ 1 \end{bmatrix}.$$

The conclusion of the exercise is that the set of solutions of the homogeneous system of equations is precisely the span of the two basic solutions, since

$$\operatorname{span}\left\{\begin{bmatrix}2\\1\\0\\0\end{bmatrix},\begin{bmatrix}\frac{1}{5}\\0\\\frac{3}{5}\\1\end{bmatrix}\right\}=\left\{t_1\begin{bmatrix}2\\1\\0\\0\end{bmatrix}+t_2\begin{bmatrix}\frac{1}{5}\\0\\\frac{3}{5}\\1\end{bmatrix}:t_1,t_2\in\mathbb{R}\right\}.$$

In-class exercise

What are the basic solutions of the following system of linear equations?

$$\begin{array}{rcrcrcl} q_1 & & & + & q_3 & = & 0\\ & + & 2q_2 & + & 3q_3 & = & 0\\ -q_1 & + & 2q_2 & + & 2q_3 & = & 0 \end{array}$$

Let A be a $m\times n$ matrix with rank r and consider the homogeneous matrix equation (system of linear equations) $A\mathbf{x}=\mathbf{0}$. Then,

(a) There are exactly $n-r$ basic solutions.

(b) Every solution of $A\mathbf{x}=\mathbf{0}$ is a linear combination of the basic solutions.

Proof. (a) By definition of basic solutions, there is one basic solution for every non-leading variable.

$$\begin{aligned}\#\text{ of nonleading variables} &= \#\text{ of variables} - \#\text{ of leading variables}\\ &= n-r\end{aligned}$$

since the number of leading variables equals the rank of the coefficient matrix A.

(b) Every solution must satisfy the equations obtained from the RREF of the augmented matrix. Thus it is a linear combination of the basic solutions, just as in Example 3.8.

□

Warning

If the rank of A is n, then the theorem tell us that $A\mathbf{x}=\mathbf{0}$ has $n-n=0$ basic solutions. This does not mean that $A\mathbf{x}=\mathbf{0}$ has zero solutions! Every homogeneous system always has the trivial solution!

Problem 2. *Let A be the matrix*

$$\begin{bmatrix} 1 & -3 & 0 & 2 & 2 \\ -2 & 6 & 1 & 2 & -5 \\ 3 & -9 & -1 & 0 & 7 \\ -3 & 9 & 2 & 6 & -8 \end{bmatrix}$$

Write the set of solutions to the system of equations $A\mathbf{x} = \mathbf{0}$ as the span of a set of vectors.

Solution. <u>*Step 1:*</u> *Row reduce the matrix until it is in reduced row echelon form.*

$$\begin{bmatrix} 1 & -3 & 0 & 2 & 2 \\ -2 & 6 & 1 & 2 & -5 \\ 3 & -9 & -1 & 0 & 7 \\ -3 & 9 & 2 & 6 & -8 \end{bmatrix} \to \cdots \to \begin{bmatrix} 1 & -3 & 0 & 2 & 2 \\ 0 & 0 & 1 & 6 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

<u>*Step 2:*</u> *The free variables are x_2, x_4, x_5. Compute the basic solutions by setting one free variable to be 1 and the others to be zero, then back substituting.*

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

<u>*Step 3:*</u> *Write*

$$\textit{span} \left\{ \begin{bmatrix} 3 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ -6 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

3.5 The associated homogeneous system of linear equations

Suppose we have a system of linear equations whose matrix equation is $A\mathbf{x} = \mathbf{b}$, and $\mathbf{b}$ is non-zero. This is not a homogeneous system, but there is an *associated homogeneous system*, $A\mathbf{x} = \mathbf{0}$, which is obtained simply by replacing $\mathbf{b}$ with $\mathbf{0}$.

The associated homogeneous system is interesting for the following reason.

Suppose $\mathbf{u}_1$ is a solution of $A\mathbf{x} = \mathbf{b}$, we call this a *particular solution*. If $\mathbf{u}_0$ is a solution of the associated homogeneous system, then by properties of matrix algebra,

$$A(\mathbf{u}_0 + \mathbf{u}_1) = A\mathbf{u}_0 + A\mathbf{u}_1 = \mathbf{0} + \mathbf{b} = \mathbf{b}.$$

Thus we have shown that $\mathbf{u}_0 + \mathbf{u}_1$ is also a solution of $A\mathbf{x} = \mathbf{b}$. In fact, all solutions of $A\mathbf{x} = \mathbf{b}$ can be written this way.

Theorem 3.10. *Suppose $\mathbf{u}_1$ is a particular solution of a system $A\mathbf{x} = \mathbf{b}$. Then, every solution of $A\mathbf{x} = \mathbf{b}$ can be written as*

$$\mathbf{u}_1 + \mathbf{u}_0$$

for some solution $\mathbf{u}_0$ of the associated homogeneous system $A\mathbf{x} = \mathbf{0}$.

Proof. Suppose that $\mathbf{v}$ is an arbitrary solution of $A\mathbf{x} = \mathbf{b}$. In other words, $A\mathbf{v} = \mathbf{b}$.

We must show that there is a $\mathbf{u}_0$ solving $A\mathbf{x} = \mathbf{0}$, such that $\mathbf{v} = \mathbf{u}_1 + \mathbf{u}_0$.

We know that $\mathbf{u}_1$ is a solution of a system $A\mathbf{x} = \mathbf{b}$. Thus, by matrix algebra,

$$A(\mathbf{v} - \mathbf{u}_1) = A\mathbf{v} - A\mathbf{u}_1 = \mathbf{b} - \mathbf{b} = \mathbf{0}.$$

This tells us that if we define $\mathbf{u}_0 = \mathbf{v} - \mathbf{u}_1$, then $\mathbf{u}_0$ is a solution of $A\mathbf{x} = \mathbf{0}$.

But now we have found a solution $\mathbf{u}_0$ of $A\mathbf{x} = \mathbf{0}$ such that

$$\mathbf{v} = \mathbf{u}_1 + \mathbf{u}_0.$$

This is what we wanted to do, so we are finished. □

3.6 All roads lead to the same reduced row echelon form

In this section (which is bonus material) we will give a short explanation of the important fact from Chapter 1,

> "All roads (constructed from row operations) lead to the same reduced row echelon form matrix."

In other words, we want to prove the following statement: If C is a matrix that can be carried to reduced row echelon form matrices A and B by row operations, then $A = B$.

We will prove this in a number of steps.

(a) First of all, if $C = 0$ then $C = A = B$. Thus we assume for the rest of the proof that C has some nonzero entries.

(b) Since we can go $C \to A$ by row operations and $C \to B$ by row operations, and we know that all row operations can be reversed, then we can go $A \to C \to B$ using row operations.

(c) The first pivot column of A and B must be the same (this is the first nonzero column from the left, which has a 1 in the top entry and zeros below it). This is true because there is no way to eliminate the first pivot using row operations, and we can get from A to B by row operations.

(d) If $A \neq B$, then they differ in some entries. Let j be the index of the first column from the left where they don't have the same entries. Let's write $\mathbf{a}$ for the jth column of A and $\mathbf{b}$ for the jth column of B.

(e) Delete all the columns of A and B to the right (with index $j + 1$ or larger) and all the non-pivot columns to the left. Now we have two new matrices, lets call them A' and B', with dimensions $n \times k$ ($k \leq j$ is the number of pivot columns we kept). Both A' and B' are in reduced row echelon form, and they differ only in their last column.

(f) Since we deleted all the non-pivot columns to the left of column j, A' and B' are of the form $A' = [e_1|e_2|...|e_{k-1}|\mathbf{a}]$ and $B' = [e_1|e_2|...|e_{k-1}|\mathbf{b}]$.

(g) Since we can go $A \to B$ by row operations, we can go $A' \to B'$ by row operations. Thus, by Theorem 1.18, the systems of equations $A'\mathbf{x} = \mathbf{0}$ and $B'\mathbf{x} = \mathbf{0}$ have the same solutions.

(h) Exercise: Show that if $[e_1|e_2|...|e_{k-1}|\mathbf{a}]\mathbf{x} = \mathbf{0}$ and $[e_1|e_2|...|e_{k-1}|\mathbf{b}]\mathbf{x} = \mathbf{0}$ have the same solutions, then $\mathbf{a} = \mathbf{b}$.

This is awkward. By part (d), we had assumed that $\mathbf{a}$ and $\mathbf{b}$ were the first columns of A and B that were different, but we just used the fact that A and B are row equivalent to explain that $\mathbf{a} = \mathbf{b}$. Where did this contradiction originally come from? It came from the beginning of part (d) where we wrote $A \neq B$! Since this lead to a contradiction, we must conclude that this is wrong. $A = B$ as we wanted to show.

Key concepts from Chapter 3

The matrix equation of a system of linear equations, homogeneous systems, linear combinations, the span of a set of vectors, $A\mathbf{x} = \mathbf{b}$ *is consistent if and only if* $\mathbf{b}$ *is in the span of the columns of* A*, linear combinations of solutions to homogeneous systems are solutions, basic solutions, the associated homogeneous system, particular solutions.*

Chapter 4

Matrix inverses

If a is a nonzero number, then there exists a number, b, such that

$$ab = 1$$

In fact, we know exactly what this number is, so this property is no surprise: $b = \frac{1}{a}$. On the other hand, if $a = 0$, then there is no number b such that $ab = 1$, since

$$ab = 0b = 0.$$

In this chapter we are interested in understanding analogous properties for matrix multiplication. Let's begin with some examples.

In-class exercise

Find u, v, w, z so that

$$\begin{bmatrix} 1 & 1 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} u & v \\ w & z \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

In-class exercise

Suppose a, b, c, d are fixed. When is it possible to find u, v, w, z so that

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} u & v \\ w & z \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}?$$

If you can find solutions for u, v, w, z, how many are there? Exactly one? Infinitely many?

These two exercises demonstrate the following phenomenon: for some matrices A, there exists a matrix B such that

$$AB = I$$

whereas for others there does not. This is similar to the situation in highschool algebra, but slightly more complicated. In this chapter, we study this phenomenon in more detail.

4.1 Invertible matrices

For the rest of this section A and B are $n \times n$ matrices.

Since it is not always true that $AB = BA$, the definition of "invertible" for matrices is a little bit complicated.

Definition 4.1. A $n \times n$ matrix A is *invertible* if there exists a $n \times n$ matrix B such that

$$AB = I \text{ and } BA = I.$$

Example 4.2. We give several examples of invertible and non-invertible matrices.

(a) A 1×1 matrix $[a]$ is just a number. By the discussion at the beginning of the chapter, $[a]$ is invertible if and only if $a \neq 0$.

(b) The identity matrices are invertible since

$$I^2 = I.$$

(c) A diagonal matrix is a matrix of the form

$$D = \begin{bmatrix} a_{11} & & \\ & \ddots & \\ & & a_{nn} \end{bmatrix}$$

where all the entries not along the diagonal are zero. If all of the diagonal entries $a_{11}, \ldots, a_{nn} \neq 0$, then D is invertible. Let B be the diagonal matrix

$$B = \begin{bmatrix} \frac{1}{a_{11}} & & \\ & \ddots & \\ & & \frac{1}{a_{nn}} \end{bmatrix}$$

Then one can compute the matrix products $DB = I$ and $BD = I$, so D is invertible.

(d) As we saw in the introduction, suppose A is the 2×2 matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

where a, b, c, d are arbitrary numbers.

If $ad - bc \neq 0$, then we can define the matrix

$$B = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

(if $ad - bc = 0$, then this formula does not define a matrix B since we have divided by 0). We can check that

$$AB = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \left(\frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \right) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and

$$BA = \left(\frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \right) \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Thus, if $ad - bc \neq 0$, then A is invertible.

On the other hand, as we saw in the introduction to the chapter, if $ad - bc = 0$, then the matrix equation $AB = I$ has no solutions, so A is not invertible.

(e) Suppose A has a row of zeros, say the ith row $= \mathbf{0}_{1\times n}$. Then for any matrix B the ii entry of AB is 0. Thus A is not invertible.

(f) Similarly, if A has a column of zeros, say the jth column, then the jj entry of BA is 0, so A is not invertible.

4.2 The inverse of a matrix

If A is an invertible matrix, then we know that there is a matrix B such that $AB = BA = I$. We saw in the previous section, that if a 2×2 matrix A is invertible, then the matrix

$$B = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

satisfies the matrix equations $AB = I$ and $BA = I$. In fact, the example from the introduction to this chapter tells us that the matrix equation $AB = I$ has exactly one solution.

The same is true in general.

If A is invertible, then there exists a unique matrix B with the property that $AB = BA = I$.

We call this unique matrix the *inverse* of A and denote it by A^{-1}.

An object with a certain property is *unique* if it is the only object with that property. Another way to say the same thing is "there is exactly one object with this property."

For instance, in a large MAT223 lecture,

- There is a unique person in the room that is the instructor.
- There may be several people who share the first name "Jeremy." In this case, there is not a unique person in the class with the first name "Jeremy."

Proof. Suppose that B and C are $n \times n$ matrices such that $CA = I$ and $AB = I$. Then by matrix algebra,

$$B = IB = (CA)B = C(AB) = CI = C.$$

Thus $B = C$. $\square$

If A is invertible, then the system $A\mathbf{x} = \mathbf{b}$ has a unique solution,

$$\mathbf{x} = A^{-1}\mathbf{b}.$$

Proof. Suppose A is invertible. Then $A(A^{-1}\mathbf{b}) = (AA^{-1})\mathbf{b} = \mathbf{b}$ so the equation $A\mathbf{x} = \mathbf{b}$ has a solution, $A^{-1}\mathbf{b}$. In fact, $A^{-1}\mathbf{b}$ is the only solution, since if $\mathbf{u}$ is a solution, then

$$A\mathbf{u} = \mathbf{b} \Rightarrow A^{-1}(A\mathbf{u}) = A^{-1}(\mathbf{b}) \Rightarrow \mathbf{u} = A^{-1}\mathbf{b}$$

Thus, the matrix equation $A\mathbf{x} = \mathbf{b}$ has exactly one solution for all $\mathbf{b}$. $\square$

We record several important properties of matrix inverses, which are a part of matrix algebra.

Properties of matrix inverses

If A and B are invertible $n \times n$ matrices, then

(a) If A is invertible, then A^T is invertible and

$$(A^T)^{-1} = (A^{-1})^T.$$

(b) The product AB is invertible, and

$$(AB)^{-1} = B^{-1}A^{-1}$$

(c) If $k \neq 0$ then kA is invertible and

$$(kA)^{-1} = \frac{1}{k}A^{-1}$$

(d) If A is invertible, then A^{-1} is invertible and

$$(A^{-1})^{-1} = A.$$

Proof. (a) By matrix algebra,

$$A^T(A^{-1})^T = (A^{-1}A)^T = (I)^T = I$$

and

$$(A^{-1})^T A^T = (AA^{-1})^T = I^T = I.$$

(b) Again, using matrix algebra, we see that

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

and

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

(c) Exercise.

(d) Since A is invertible, we know there is an inverse A^{-1} such that

$$AA^{-1} = I \text{ and } A^{-1}A = I.$$

By definition, these two equations tell us that the matrix A^{-1} is invertible and the inverse of A^{-1} is A.

□

Check your understanding

Suppose that A, B and C are invertible $n \times n$ matrices.

Is the product ABC invertible?

In-class exercise

Prove property (c).

Warning

If A and B are invertible $n \times n$ matrices, it is not necessarily true that $A + B$ is invertible. This is true even when A and B are scalars: both 2 and -2 are invertible, but $2 - 2 = 0$ is not invertible.

4.3 Elementary matrices

Example 4.3. Consider the 3×3 matrices

$$E_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & k \end{bmatrix}, \text{ and } E_3 = \begin{bmatrix} 1 & 0 & k \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Let

$$B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}$$

In-class exercise

Compute the products

$$E_1B, \; E_2B, \; \text{and } E_3B.$$

For each matrix product, solve the matrix equation $E_iB = I$ for the variables b_{ij}.

Definition 4.4. An elementary matrix E is a matrix that is obtained by performing a single elementary row operation on the identity matrix I.

Check your understanding

For each of the matrices E_1, E_2, E_3 above, explain how to obtain E_i by performing one elementary row operation on I.

For what values of k is E_2 an elementary matrix?

Observe that the result of multiplying B on the left by an elementary matrix is the same as performing the corresponding elementary row operation on B.

If E_R is an elementary matrix corresponding to an elementary row operation R, then for any $n \times m$ matrix B, the product EB equals the matrix obtained from B by performing the row operation R.

Recall that every elementary row operation R can be reversed by another elementary row operation. We denote the reverse/inverse row operation by R^{-1}.

Every elementary matrix E_R (corresponding to the row operation R) is invertible, and $E_R^{-1} = E_{R^{-1}}$ is the elementary matrix corresponding to the reverse row operation.

Proof. By the preceding fact, we see that the matrix

$$E_{R^{-1}}E_R = E_{R^{-1}}E_RI$$

is the same as the matrix obtained from I by performing the row operation R, then performing the reverse row operation R^{-1}. This sequence of row operations gives us back I, so $E_{R^{-1}}E_R = I$. Similarly, we can check that $E_RE_{R^{-1}} = I$. $\square$

Check your understanding

Write the 3×3 elementary matrices corresponding to the following row operations:

(a) Interchange row 2 and 3.

(b) Interchange row 1 and 3.

(c) Add -2 times row 3 to row 2.

(d) Add -2 times row 2 to row 3.

(e) Multiply row 1 by 6.

What is the inverse of each elementary matrix?

In-class exercise

Suppose that $E_1, E_2, \ldots, E_k$ are elementary matrices.

The product $E_k \cdots E_2 E_1$ is invertible. Why? What is its inverse?

4.4 The matrix inversion algorithm

Given a $n \times n$ matrix A, we can form a $n \times 2n$ matrix $[A|I]$. The matrix inversion algorithm is the following.

The matrix inversion algorithm

The algorithm is simple: perform elementary row operations on the matrix $[A|I]$ until A is in reduced row echelon form.

Once this is finished, there are two possibilities:

1. If the reduced row echelon form of A is I, then the resulting matrix is $[I|A^{-1}]$
2. If the reduced row echelon form of A is not I, then A is not invertible.

Proof. Suppose $R_1, \ldots, R_k$ is a sequence of elementary row operations that carry A to its reduced row-echelon form. Since left multiplication by elementary matrices corresponds to performing elementary row operations, we know that

$$E_k \cdots E_2 E_1 A$$

is in reduced row echelon form.

If the reduced row echelon form is I, then we know that

$$E_k \cdots E_2 E_1 A = I$$

This means that A is invertible, and $A^{-1} = E_k \cdots E_2 E_1$. This is precisely the matrix that will show up on the right hand side of the block matrix $[A|I]$ after performing elementary row operations.

If the reduced row echelon form is not I, then it must have a column of zeros. Thus by the examples in Section 4.1, the matrix

$$E_k \cdots E_2 E_1 A$$

is not invertible. It is possible to conclude from this that A is not invertible. See Tutorial **??**. □

As a consequence of this proof, we also see the following. If A is invertible, then

$$A = (E_k \cdots E_2 E_1)^{-1} = E_1^{-1} \cdots E_k^{-1}.$$

Since inverses of elementary matrices are elementary matrices, we have shown the following useful fact.

Let A be a $n \times n$ matrix. The following statements are equivalent:

1. A is invertible.
2. A is equal to a product of elementary matrices.

Example 4.5. Use the matrix inversion algorithm to determine whether the following matrices are invertible and if so, find their inverses.

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

If the matrix is invertible, write the matrix and its inverse as a product of elementary matrices.

4.5 The big theorem for square matrices (part 1)

The pinnacle of the first four chapters of the course is the following theorem, which demonstrates how several different concepts are related.

Theorem 4.6. *[The big theorem for square matrices, part 1] Suppose A is a $n \times n$ matrix. The following statements are equivalent:*

(a) The matrix A is invertible.

(b) The matrix equation $A\mathbf{x} = \mathbf{b}$ has exactly one solution for all $n \times 1$ column vectors $\mathbf{b}$.

(c) The homogeneous matrix equation $A\mathbf{x} = \mathbf{0}$ has exactly one solution (the trivial solution).

(d) The rank of A is n.

(e) The reduced row-echelon form of A is I.

(f) There exists a $n \times n$ matrix B such that $AB = I$.

(g) There exists a $n \times n$ matrix B such that $BA = I$.

(h) A can be written as a product of elementary matrices.

"the following statements are equivalent" means that if any one of these things are true, then the others are also true. For instance, if you know that (d) is true, then you can conclude that (a) is true.

Proof. Let's prove that (a)-(e) are equivalent. In the next section we will finish the proof by showing that (a), (f), and (g) are equivalent. The fact that (a) and (h) are equivalent was explained in the previous section.

$(a) \Rightarrow (b)$ This was already proven in Section 4.1.

$(b) \Rightarrow (c)$ Suppose the matrix equation $A\mathbf{x} = \mathbf{b}$ has exactly one solution for all $\mathbf{b}$. Then, in particular, it has exactly one solution when $\mathbf{b} = \mathbf{0}$.

$(c) \Rightarrow (d)$ Suppose that $A\mathbf{x} = \mathbf{0}$ has exactly one solution. Then by Theorem 1.24, the rank of A must equal n.

$(d) \Rightarrow (e)$ If the rank of A is n, then the reduced row echelon form of A has n leading 1's. Since there are only n columns, the only $n \times n$ matrix in row echelon form with n leading 1's is the identity matrix, I.

$(e) \Rightarrow (a)$ If the reduced row-echelon form of A is I, then the matrix inversion algorithm of Section 4.4 produces the inverse of A. $\square$

This proof demonstrates one way to prove several different statements are equivalent.

Check your understanding

Suppose A is a 2×2 matrix with rank 2. Which of the following statements are true? Explain your answer using the big theorem for square matrices.

(a) A is invertible.

(b) The reduced row echelon form of A is I.

(c) The matrix equation $A\mathbf{x} = \mathbf{b}$ has exactly one solution when $\mathbf{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

(d) There exists a 2×2 matrix B such that $AB = BA = I$.

(e) There exists a 2×2 matrix B such that $AB = I$.

(f) The rank of A^T is 2.

4.6 $BA = I \implies AB = I$ for square matrices

The definition of an invertible matrix says a square matrix A is invertible matrix if there exists a (square) matrix B such that

$$AB = I \text{ and } BA = I.$$

Until now, this has meant that in order to check a matrix is invertible, we had to compute two matrix products. In this section we show the following.

If A and B are $n \times n$ matrices, then

$$BA = I \implies AB = I.$$

and vis versa,

$$AB = I \implies BA = I$$

Thus it is sufficient to check that only one of the products AB and BA are equal to I, since it immediately follows that the other product equals I as well.

What is surprising is that this fact cannot be proven using only matrix algebra! We need to use something more.

Proof. Assume that $BA = I$. We want to prove that $AB = I$.

- First, the homogeneous system $BA\mathbf{x} = \mathbf{0}$ has exactly one solution, the trivial solution, since

  $$BA\mathbf{x} = I\mathbf{x} = \mathbf{x}$$

  and this equals zero only when $\mathbf{x} = \mathbf{0}$.

- Second, the homogeneous system $A\mathbf{x} = \mathbf{0}$ has exactly one solution, the trivial solution:

  If $\mathbf{u}$ is a solution, then $\mathbf{u}$ satisfies the equation $BA\mathbf{u} = \mathbf{0}$. But,

  $$\mathbf{0} = BA\mathbf{u} = I\mathbf{u} = \mathbf{u}$$

  so $\mathbf{u} = \mathbf{0}$.

- Next, since the homogeneous system $A\mathbf{x} = \mathbf{0}$ has exactly one solution, we can apply the big theorem to conclude that A is invertible. (we are using (c) $\implies$ (a) from the big theorem)

- Since A is invertible, we know that A^{-1} exists. We also know by assumption that

  $$BA = I$$

  Multiplying on the right by A^{-1}, we get

  $$(BA)A^{-1} = IA^{-1}$$

  Thus

  $$B = BI = B(AA^{-1}) = (BA)A^{-1} = IA^{-1} = A^{-1}.$$

- Finally, since $B = A^{-1}$, we know that $AB = AA^{-1} = I$.

The proof of the other direction is similar. Just switch the roles of A and B. $\square$

In general, if $AB = I$, then we say that "B is a right inverse of A," since multiplication of A on the right by B results in I. Similarly, if $BA = I$, then we say that "B is a left inverse of A." The fact above says that for square matrices,

> "If B is a left inverse of A, then B is a right inverse of A".

and

> "If B is a right inverse of A, then B is a left inverse of A".

Warning

If A and B are not square matrices, then this fact is not true. For example, let

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Then $BA = I_2$, but

$$AB = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \neq I_3.$$

Thus B is a left inverse of A, but B is not a right inverse of A.

Thus we have shown that statements (f) and (g) in Theorem 6.25 are equivalent.

Now we can complete the proof of Theorem 6.25 by showing that (a), (f), and (g) are equivalent.

Proof. If (a) is true, then A is invertible. By definition, this means that A has both a left and right inverse. Thus, both (f) and (g) are true.

On the other hand, if one of (f) or (g) are true, then both (f) and (g) are true since (f) and (g) are equivalent. This means that A has both a left inverse and a right inverse. But this implies that A is invertible, so (a) is true. $\square$

Key concepts from Chapter 4

Definition of the inverse of a matrix. The inverse is unique. Formula for the inverse of a 2×2 matrix. Basic properties of matrix inverses. Definition of elementary matrices. Multiplication by elementary matrices corresponds to elementary row operations. Inverses of elementary matrices. If $AB = I$, then $BA = I$ for square matrices. The matrix inversion algorithm. The big theorem for square matrices.

Chapter 5

Geometry of $\mathbb{R}^n$

In Chapter 1 we defined $\mathbb{R}^n$ as the set of all n-tuples of real numbers,

$$(x_1, \ldots, x_n) .$$

When $n = 2$, this set has a familiar geometric interpretation. A 2-tuple, or pair, of real numbers (x, y) can be thought of as a point in the xy-plane. There we have defined many objects: lines, circles, triangles, rectangles etc., and learned about the geometry of these objects (length, angle, area, etc.) in high school.

In this chapter we will define analogous objects in $\mathbb{R}^n$, study their geometry, and begin to explore connections between this geometry and matrix algebra and systems of linear equations.

5.1 Review: geometry of $\mathbb{R}^2$

Let's review some concepts which you may have learned in high school.

First, recall that the *absolute value* $|x|$ of a number $x \in \mathbb{R}$ is the magnitude of the number x, which is defined by the formula

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

A more convenient formula is given by the identity $|x| = \sqrt{x^2}$. The "distance" between two real numbers $x_1, x_2 \in \mathbb{R}$ is the magnitude of their difference,

$$|x_2 - x_1| = \sqrt{(x_2 - x_1)^2}.$$

Calling this quantity distance makes perfect sense. For instance, the distance between a town that is 5km south on the highway, and a village that is 8 km north on the highway is

$$|(-5) - 8| = |-13| = 13 \text{ km}.$$

- Points in the xy-plane are described by their xy coordinates, (x, y). Sometimes we denote a point by a letter, such as P. The point $(0, 0)$ is called the origin and is somtimes denoted by the letter O.
- The *distance between two points* $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ in $\mathbb{R}^2$ is defined by the formula

$$d(P_1, P_2) := \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

  This is a straightforward generalization of the notion of distance between two numbers in $\mathbb{R}$.

  We can use distance to define many new geometric objects. For instance, *the circle centred at P with radius r* is defined as the set

$$\left\{Q \in \mathbb{R}^2 : d(P, Q) = r\right\}.$$

  The circle centred at $(0, 0)$ with radius 1 is often called the *unit circle.*

  The area of a circle of radius r (centred at any point) is given by the formula

$$Area = \pi r^2.$$

- Lines in the xy-plane are hard to define precisely without using equations. There are several ways to define a line using equations:
  - **slope and y-intercept:** The line with slope m and y-intercept b is the set of solutions to the equation

$$y = mx + b.$$

  - **point and slope:** The line with slope m containing the point (x_0, y_0) is the set of solutions to the equation

$$y - y_0 = m(x - x_0).$$

Both of these equations require two pieces of information: the slope of the line, and a point on the line (in the first case, $(0, b)$). Thus we say

> *A point and a slope determines a line.*

There are many lines with different slopes passing through a given point. Given two points on a line, we can calculate the slope of the line as

$$m = \frac{y_1 - y_0}{x_1 - x_0}$$

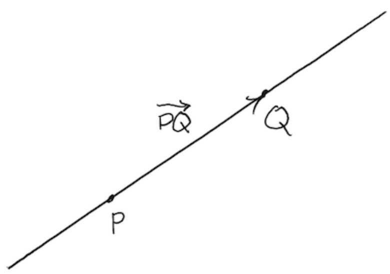
thus we say that

> *Two points determine a line.*

Two lines are called *parallel* if they have the same slope.

- A *line segment* is the portion of a line between two points on the line. We call the two points 'endpoints.' We sometimes write $\overline{PQ}$ to indicate the line segment between the endpoints P and Q. The length of a line segment is the distance between the two endpoints.

- A *directed line segment* is a line segment together with a direction going from one endpoint to the other. The directed line segment from P to Q is denoted by $\overrightarrow{PQ}$. The point P is called the *tail* of the directed line segment and the point Q is called the *tip*.

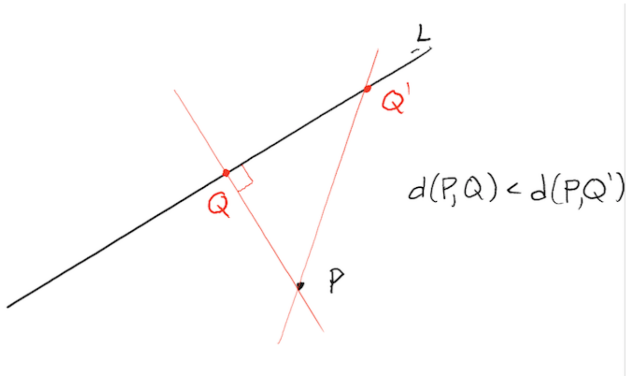


- Two intersecting lines are *perpendicular* if the angles between the lines equal $\pi/2$ (we always use radians).

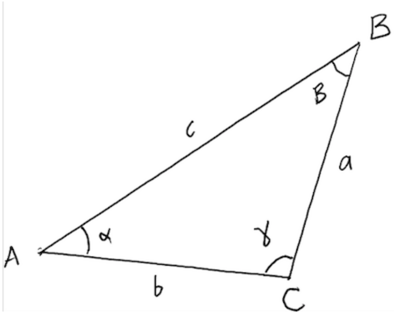
- Given a line L and a point P not on the line, the *closest point to P on L* is the unique point Q on L such that the distance

$$d(P, Q)$$

is as small as possible. It is a geometric fact that this point, Q, can be constructed as the intersection of L and the line through P that is perpendicular to L.



- Three points (that don't line on a line) determine a triangle and these points are the vertices of the triangle. Every triangle has three vertices, three sides, and three interior angles. The sum of the three interior angles is always equal to π.



- The **law of cosines** says that for any triangle as labelled above,

$$c^2 = a^2 + b^2 - 2ab\cos(\gamma)$$

- A triangle is a *right triangle* if one of the interior angles $\gamma = \pi/2$. Right triangles satisfy the Pythagorean theorem:

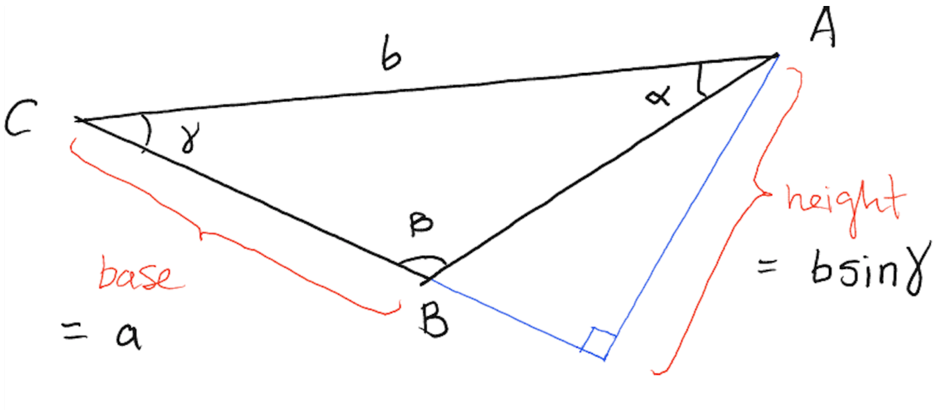
$$a^2 + b^2 = c^2.$$

- The area of a triangle is computed using the formula

$$area = \frac{base \times height}{2}$$

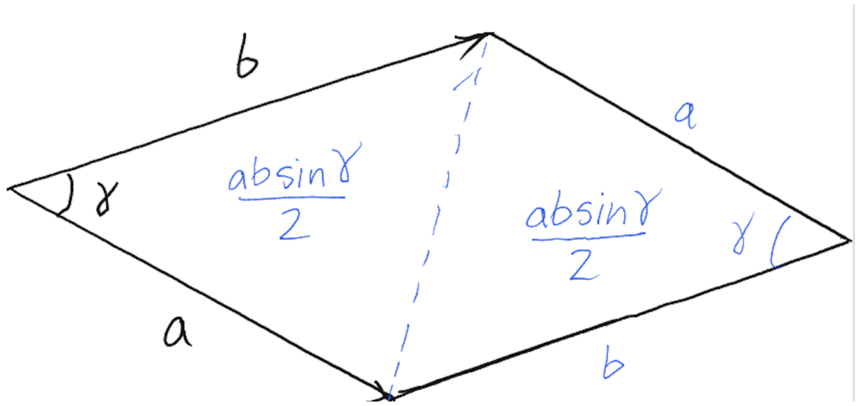
where base and height are the lengths of line segments indicated in the picture. In terms of the angle α, the area of the triangle in the picture is

$$area = \frac{base \times height}{2} = \frac{ab\sin(\gamma)}{2} \tag{5.1}$$



- A parallelogram is a quadrilateral with both pairs of opposite sides parallel. The area of a parallelogram can be computed by dividing it into two triangles as drawn.

$$Area = ab\sin(\gamma) \tag{5.2}$$



5.2 Vectors and points

We have defined $\mathbb{R}^n$ to be the set of all n-tuples of real numbers,

$$\mathbb{R}^n = \{(x_1, \ldots, x_n) : x_1, \ldots, x_n \in \mathbb{R}\}.$$

A *point* in $\mathbb{R}^n$ is just a more picturesque way of saying an element of $\mathbb{R}^n$.

We call the point $(0, \ldots, 0)$ *the origin.*

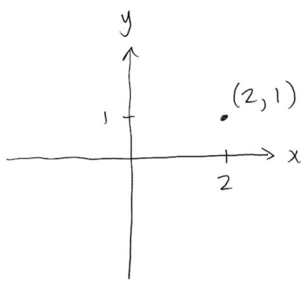
For convenience (and so that we can use matrix algebra to study geometry), we want to use column vectors (and sometimes row vectors) to talk about points and geometry in $\mathbb{R}^n$. This will allow us to connect concepts that we have defined algebraically using matrix algebra to the geometry of $\mathbb{R}^n$.

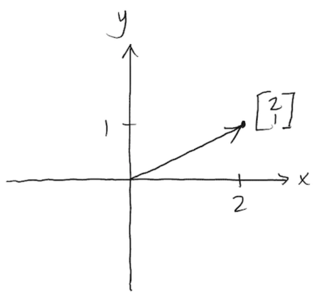
Thus we will often identify a $n \times 1$ column vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

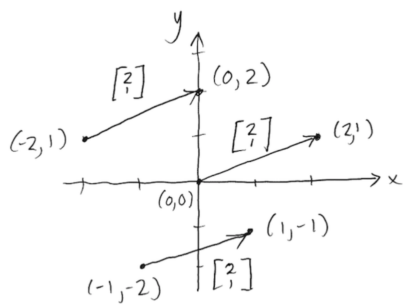
with the point $(x_1, \ldots, x_n)$. With this identification, we can interchangeably think of $\mathbb{R}^n$ as the set of all n-tuples of real numbers, or the set of all $n \times 1$ column vectors.

Another perspective on this identification is that a column vector $\mathbf{x}$ represents the magnitude and direction of a directed line segment in $\mathbb{R}^n$, without specifying the endpoints. If you fix the tail of the directed line segment $\mathbf{x}$ to be the origin, then the tip will be the point $(x_1, \ldots, x_n)$.





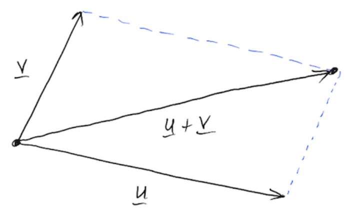
Row vectors can be identified with points and directed line segments in $\mathbb{R}^n$ in the same way.



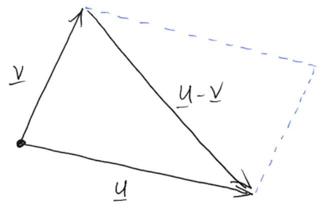
5.3 Addition, scalar multiplication, and lines

We can add and scalar multiply points in $\mathbb{R}^n$ the same way we add and scalar multiply column/row vectors.

The sum of two vectors $\mathbf{u}$ and $\mathbf{v}$ is the vector $\mathbf{u}+\mathbf{v}$ whose tail is the origin and whose tip is the opposite corner of the parallelogram, as drawn.



The difference of two vectors $\mathbf{u}$ and $\mathbf{v}$ is the vector $\mathbf{u}-\mathbf{v}$ that can be represented by the directed line segment whose tail is the tip of $\mathbf{v}$ and whose tip is the tip of $\mathbf{u}$.



If $P_1 = (x_1, \ldots, x_n)$ is represented by $\mathbf{x}$ and $P_2 = (y_1, \ldots, y_n)$ is represented by $\mathbf{y}$, then the directed line segment from P_1 to P_2 can be represented by the vector

$$\mathbf{y} - \mathbf{x} = \begin{bmatrix} y_1 - x_1 \\ \vdots \\ y_n - x_n \end{bmatrix}$$

We can use addition of vectors to define lines in $\mathbb{R}^n$.

- Given a point $\mathbf{x} \in \mathbb{R}^n$ and a vector $\mathbf{v} \in \mathbb{R}^n$, we define *the line through* $\mathbf{x}$ *parallel to* $\mathbf{v}$ to be the set

$$\{\mathbf{x} + t\mathbf{v} \colon t \in \mathbb{R}\} .$$

  This is similar to how a point and a slope determined a line in $\mathbb{R}^2$.

A point and a vector determine a line.

- We can also define a line by specifying two points on the line. If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ are two points, then the line through $\mathbf{x}$ parallel to $\mathbf{y} - \mathbf{x}$ is the set

$$\{\mathbf{x} + t(\mathbf{y} - \mathbf{x}) \colon t \in \mathbb{R}\}.$$

  Note that this line contains $\mathbf{y}$ (set $t = 1$). Thus, as before,

  Two points determine a line.

- A *line through the origin parallel to* $\mathbf{v}$ is the set

$$\{0 + t\mathbf{v} \colon t \in \mathbb{R}\}.$$

  Notice that

$$\{0 + t\mathbf{v} \colon t \in \mathbb{R}\} = \operatorname{span}\{\mathbf{v}\}.$$

  Thus a line through the origin is the span of a set of one vector.

Definition 5.3. Two non-zero vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are *parallel* if there is a scalar k such that $\mathbf{u} = k\mathbf{v}$.

Check your understanding

Which of the following vectors are parallel?

$$\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

If two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are parallel, then there is a scalar $k \neq 0$ such that $\mathbf{u} = k\mathbf{v}$. From this we see that

$$\operatorname{span}\{\mathbf{u}\} = \operatorname{span}\{k\mathbf{v}\} = \operatorname{span}\{\mathbf{v}\}$$

so two parallel vectors span the same line through the origin.

Equality of sets

Two sets are equal if they contain the same objects.

This concept is important because sometimes there are often multiple ways to express a set using set-builder notation (or by other means). Above, we saw that if $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are parallel, then the two sets

$$\operatorname{span}\{\mathbf{u}\} := \{t\mathbf{u} \colon t \in \mathbb{R}\}$$

and

$$\operatorname{span}\{\mathbf{v}\} := \{t\mathbf{v} \colon t \in \mathbb{R}\}$$

are the same, even though the set-builder description of each span looks slightly different.

Here is a similar example. Suppose as before that we are considering sets of shirts that I own. Suppose I own three shirts: my blue Blue Jays shirt (size medium), my blue gym shirt (size medium), and my white Honest Ed's shirt (size large). Consider the following two sets of shirts:

$$\{s \in \text{my shirts} : s \text{ is blue}\}$$

and

$$\{s \in \text{my shirts} : \text{the size of } s \text{ is medium}\}.$$

We see that both of these set-builder descriptions describe the same set of shirts. Thus we write

$$\{s \in \text{my shirts} : s \text{ is blue}\} = \{s \in \text{my shirts} : \text{the size of } s \text{ is medium}\}.$$

There are several ways to show that two sets are equal. We will see more about this later.

Definition 5.4. Two or more points in $\mathbb{R}^n$ are *collinear* if they lie on the same line.

Since any two points $\mathbf{x}$ and $\mathbf{y}$ determine the line

$$L := \{\mathbf{x} + t(\mathbf{y} - \mathbf{x}) \colon t \in \mathbb{R}\},$$

three points $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ are collinear if $\mathbf{z}$ is contained in the line L. In other words, $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ are collinear if there is a number $t \in \mathbb{R}$ such that

$$\mathbf{z} = t\mathbf{y} + (1 - t)\mathbf{x}.$$

In-class exercise

For which values of c are the following three points in $\mathbb{R}^3$ collinear?

$$(1,-1),(-1,1),(0,c)$$

If $P_0 = \mathbf{x}$, $P_1 = \mathbf{y}$ and $P_2 = \mathbf{z}$ are collinear, then the vectors $\overrightarrow{P_0P_1}$ and $\overrightarrow{P_0P_2}$ are parallel.

Proof. Write If $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ are collinear then there is a number $t \in \mathbb{R}$ such that

$$\mathbf{z} = t\mathbf{y} + (1-t)\mathbf{x}.$$

Then

$$\overrightarrow{P_0P_1} = \mathbf{y} - \mathbf{x}$$

and

$$\overrightarrow{P_0P_2} = \mathbf{z} - \mathbf{x} = t\mathbf{y} + (1-t)\mathbf{x} - \mathbf{x} = t(\mathbf{y} - \mathbf{x}).$$

Thus

$$\overrightarrow{P_0P_2} = t\overrightarrow{P_0P_1},$$

so the two vectors are parallel. □

In-class exercise

The vectors, $[1\,2\,0\,5]$ and $[2\,4\,0\,10]$ are parallel. What is the rank of the matrix

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

More generally, suppose that $\mathbf{u} = [u_1 \ \ldots \ u_n]$, $\mathbf{v} = [v_1 \ \ldots \ v_n]$ are two parallel, nonzero vectors. What is the rank of the matrix

$$\begin{bmatrix} u_1 & \cdots & u_n \\ v_1 & \cdots & v_n \end{bmatrix}?$$

Justify your answer.

5.4 The distance between two points

The definition of distance between two points in $\mathbb{R}^2$ generalizes to $\mathbb{R}^n$ in a natural way.

Definition 5.5. If $P_1 = (x_1, \ldots, x_n)$ and $P_2 = (y_1, \ldots, y_n)$ are points in $\mathbb{R}^n$, then the *distance* between the two points is defined as

$$d(P_1, P_2) := \sqrt{(y_1 - x_1)^2 + \cdots + (y_n - x_n)^2}.$$

Definition 5.6. The *length* or *norm* of a vector $\mathbf{v}$ is defined to be

$$||\mathbf{v}|| := \sqrt{v_1^2 + \cdots + v_n^2}.$$

Example 5.7. The length of the zero vector is 0:

$$||\mathbf{0}|| = \sqrt{0^2 + \cdots + 0^2} = 0.$$

Check your understanding

Compute the length of the vector

$$\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} \in \mathbb{R}^3.$$

Compute the distance between the two points $P = (1, 0, 1)$ and $Q = (0, 1, -1)$.

For any vector $\mathbf{x}$,

(a) If $||\mathbf{x}|| = 0$, then $\mathbf{x} = \mathbf{0}$.

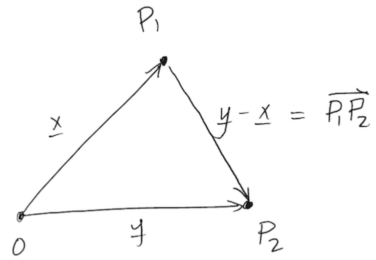
(b) For any scalar k, $||k\mathbf{x}|| = |k| \cdot ||\mathbf{x}||$.

Recall that if $P_1 = (x_1, \ldots, x_n)$ is represented by $\mathbf{x}$ and $P_2 = (y_1, \ldots, y_n)$ is represented by $\mathbf{y}$, then the directed line segment from P_1 to P_2 can be represented by the vector

$$\mathbf{y} - \mathbf{x} = \begin{bmatrix} y_1 - x_1 \\ \vdots \\ y_n - x_n \end{bmatrix}$$

and we see that

$$d(P_1, P_2) = \sqrt{(y_1 - x_1)^2 + \cdots + (y_n - x_n)^2} = ||\mathbf{y} - \mathbf{x}||.$$



In $\mathbb{R}^2$, the unit circle is the set of all points that are distance 1 from the origin. If we view points as vectors, then it is the set of all vectors with norm $= 1$.

Definition 5.8. A vector $\mathbf{x}$ is called a *unit vector* if $||\mathbf{x}|| = 1$.

In-class exercise

Which of the following vectors in $\mathbb{R}^3$ are unit vectors?

$$\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{bmatrix}, \begin{bmatrix} \sin(\alpha) \\ 0 \\ \cos(\alpha) \end{bmatrix}.$$

5.5 The dot product

Definition 5.9. The *dot product* of two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ is the scalar

$$\mathbf{u} \cdot \mathbf{v} := u_1 v_1 + \cdots + u_n v_n.$$

Example 5.10.

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

Matrix multiplication gives us another way to write a dot product:

$$\mathbf{u}^T \mathbf{v} = [u_1 \cdots u_n] \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = u_1 v_1 + \cdots + u_n v_n = \mathbf{u} \cdot \mathbf{v}.$$

There is a simple relationship between the dot product of a vector with itself and its norm,

$$||\mathbf{u}||^2 = \left(\sqrt{u_1^2 + \cdots + u_n^2}\right)^2 = u_1^2 + \cdots + u_n^2 = \mathbf{u} \cdot \mathbf{u}.$$

Properties of the dot-product:

For any vectors $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$.

(a) $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$

(b) $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$

(c) $(k\mathbf{u}) \cdot \mathbf{v} = k(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot (k\mathbf{v})$.

For any two vectors, we see that

$$||\mathbf{u} + \mathbf{v}||^2 = (\mathbf{u} + \mathbf{v}) \cdot (\mathbf{u} + \mathbf{v}) = ||\mathbf{u}||^2 + 2(\mathbf{u} \cdot \mathbf{v}) + ||\mathbf{v}||^2.$$

This is a useful formula for computation.

In-class exercise

Suppose $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ such that $||\mathbf{u}|| = 2$, $||\mathbf{v}|| = 1$, and $||\mathbf{u} + \mathbf{v}|| = 1$. What is $\mathbf{u} \cdot \mathbf{v}$?

Theorem 5.11 (The Cauchy-Schwartz inequality)**.** *For any two vectors* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$*,*

$$|\mathbf{u} \cdot \mathbf{v}| \leq ||\mathbf{u}|| ||\mathbf{v}||.$$

We can use this to prove the triangle inequality:

Theorem 5.12 (The triangle inequality)**.** *For any two vectors* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$*,*

$$||\mathbf{u} + \mathbf{v}|| \leq ||\mathbf{u}|| + ||\mathbf{v}||.$$

Proof. Let's start with the square of the left hand side and show it is less than or equal to the square of the right hand side.

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

Thus $||\mathbf{u}+\mathbf{v}|| \leq ||\mathbf{u}|| + ||\mathbf{v}||$. □

5.6 The angle between two vectors

As in $\mathbb{R}^2$, two non-zero vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ determine two angles: one less than or equal to π and the other larger than or equal to π. The *angle between* $\mathbf{u}$ *and* $\mathbf{v}$ is the angle that is $\leq \pi$.

Applying the law of cosines to the triangle with sides $\mathbf{u}, \mathbf{v}, \mathbf{u}-\mathbf{v}$, we see that

$$||\mathbf{u}-\mathbf{v}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2 - 2||\mathbf{u}||||\mathbf{v}||\cos(\theta)$$

Since

$$||\mathbf{u}-\mathbf{v}||^2 = (\mathbf{u}-\mathbf{v})\cdot(\mathbf{u}-\mathbf{v}) = ||\mathbf{u}||^2 - 2(\mathbf{u}\cdot\mathbf{v}) + ||\mathbf{v}||^2$$

we see that

$$(\mathbf{u}\cdot\mathbf{v}) = ||\mathbf{u}||||\mathbf{v}||\cos(\theta).$$

Since $\mathbf{u}, \mathbf{v}$ are non-zero, we may divide both sides to get the formula

$$\frac{\mathbf{u}\cdot\mathbf{v}}{||\mathbf{u}||||\mathbf{v}||} = \cos(\theta).$$

If the angle $\theta = \pi$ (in other words, the angle is a right angle), then $\cos(\theta) = 0$ so the formula tells us that $\mathbf{u}\cdot\mathbf{v} = 0$.

Definition 5.13. Two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are *orthogonal* if

$$\mathbf{u}\cdot\mathbf{v} = 0.$$

Being "orthogonal" is the $\mathbb{R}^n$ analogue of two vectors being perpendicular in $\mathbb{R}^2$.

Problem 3. *Show that two vectors* $\mathbf{u}$ *and* $\mathbf{v}$ *are orthogonal if and only if*

$$||\mathbf{u}+\mathbf{v}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2.$$

Solution. *From the equation*

$$||\mathbf{u}+\mathbf{v}||^2 = (\mathbf{u}+\mathbf{v})\cdot(\mathbf{u}+\mathbf{v}) = ||\mathbf{u}||^2 + 2(\mathbf{u}\cdot\mathbf{v}) + ||\mathbf{v}||^2$$

we see that the equation

$$||\mathbf{u}+\mathbf{v}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2$$

is equivalent to the equation

$$\mathbf{u}\cdot\mathbf{v} = 0.$$

5.7 Orthogonal projection

The problem of orthogonal projection: Consider the following problem. Given a vector $\mathbf{v}$ and a nonzero vector $\mathbf{u}$, we want to be able to write $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$ where $\mathbf{v}_1$ is parallel to $\mathbf{u}$ and $\mathbf{v}_2$ is orthogonal to $\mathbf{u}$.

Solution: Since $\mathbf{v}_1$ is parallel to $\mathbf{u}$, $\mathbf{v}_1 = k\mathbf{u}$ for some $k \in \mathbb{R}$, so we have

$$\mathbf{v} = k\mathbf{u} + \mathbf{v}_2$$

Taking the dot product of both sides of this equation with $\mathbf{u}$, we get

$$\mathbf{u}\cdot\mathbf{v} = \mathbf{u}\cdot(k\mathbf{u}+\mathbf{v}_2) = k||\mathbf{u}||^2$$

(question: why is the equality on the right hand side true?). This equation can be solved for k, which gives us the solution for $\mathbf{v}_1$.

$$\mathbf{v}_1 = k\mathbf{u} = \frac{\mathbf{u}\cdot\mathbf{v}}{||\mathbf{u}||^2}\mathbf{u}.$$

Knowing $\mathbf{v}_1$, we can then solve for $\mathbf{v}_2$,

$$\mathbf{v}_2 = \mathbf{v} - \mathbf{v}_1 = \mathbf{v} - \frac{\mathbf{u}\cdot\mathbf{v}}{||\mathbf{u}||^2}\mathbf{u}.$$

Thus we have found $\mathbf{v}_1$ and $\mathbf{v}_2$ that solve the problem above.

Check your understanding

At the beginning of the section we said that $\mathbf{u}$ was a nonzero vector. Where did we use this assumption in our solution to the problem of orthogonal projection?

In fact, the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ solving the problem of orthogonal projection are unique (it is worth trying to check this for yourself). These vectors are quite important in linear algebra, so we give them names.

The *orthogonal projection* of $\mathbf{v}$ onto the line through the origin spanned by $\mathbf{u}$ is given by the formula

$$\mathrm{proj}_{\mathbf{u}}(\mathbf{v}) = \frac{\mathbf{u} \cdot \mathbf{v}}{||\mathbf{u}||^2}\mathbf{u}.$$

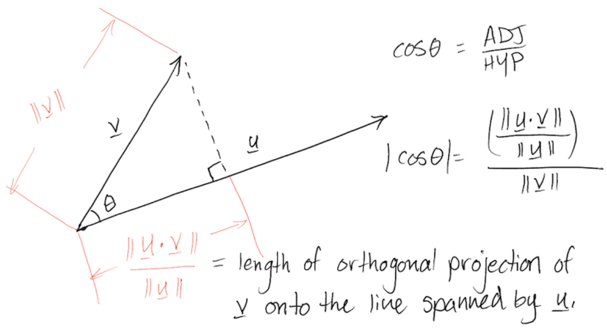
The *projection of* $\mathbf{v}$ *orthogonal to the line spanned by* $\mathbf{u}$ is the vector

$$\mathbf{v} - \mathrm{proj}_{\mathbf{u}}(\mathbf{v}).$$

In-class exercise

In general, what is the angle between $\mathbf{u}$ and $\mathbf{v} - \mathrm{proj}_{\mathbf{u}}(\mathbf{v})$? Can you write a proof?

Hint: use the material from the previous two sections.



Recall that the distance between a point P and a line L is defined as the shortest distance between P and any point on L. The following theorem tells us that given a line and a point, the shortest line segment between the point and the line is given by the projection orthogonal to the line.

Theorem 5.14. *The distance between the line through the origin spanned by* $\mathbf{u}$ *and the point represented by the vector* $\mathbf{v}$ *equals the length of* $\mathbf{v} - \mathrm{proj}_{\mathbf{u}}(\mathbf{v})$ *(the projection of* $\mathbf{v}$ *orthogonal to the line spanned by* $\mathbf{u}$*).*

Proof. We know that $\mathrm{proj}_{\mathbf{u}}(\mathbf{v})$ is a point on the line spanned by $\mathbf{u}$. Using the formula for distance, we compute

$$d\left(\mathbf{v}, \mathrm{proj}_{\mathbf{u}}(\mathbf{v})\right) = ||\mathbf{v} - \mathrm{proj}_{\mathbf{u}}(\mathbf{v})||.$$

In order to show that this is the distance between $\mathbf{v}$ and the line spanned by $\mathbf{u}$, it remains to explain why for any other point $Q' = k\mathbf{u}$ on this line,

$$d\left(\mathbf{v}, \operatorname{proj}_{\mathbf{u}}(\mathbf{v})\right) \leq d(\mathbf{v}, Q').$$

This final step is left an exercise (hint: use pythagoras and the result of the in-class exercise above). $\square$

5.8 Planes in $\mathbb{R}^3$

In $\mathbb{R}^2$, a linear equation $y = mx + b$ determines a line with slope m and y-intercept b. In $\mathbb{R}^3$, a single linear equation determines a plane.

Definition 5.15. A set of points in $\mathbb{R}^3$ defined by a single linear equation,

$$a_1x + a_2y + a_3z = b$$

is called a *plane*.

Example 5.16. (a) The equation $x = 0$ defines the yz-plane

$$\{(x, y, z) \in \mathbb{R}^3 \colon x = 0\} = \{(0, y, z) \in \mathbb{R}^3 \colon y, z \in \mathbb{R}\}.$$

(b) The equation $y = 0$ defines the xz-plane

$$\{(x, y, z) \in \mathbb{R}^3 \colon y = 0\} = \{(x, 0, z) \in \mathbb{R}^3 \colon x, z \in \mathbb{R}\}.$$

(c) The equation $x = 1$ defines the

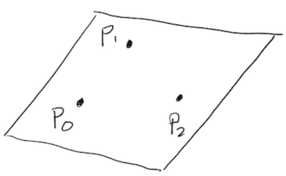
$$\{(x, y, z) \in \mathbb{R}^3 \colon x = 1\} = \{(1, y, z) \in \mathbb{R}^3 \colon y, z \in \mathbb{R}\}.$$

Given three distinct points, P_0, P_1, P_2 not collinear (not lying on a line), there is a unique equation

$$a_1x + a_2y + a_3z = b$$

such that P_0, P_1, P_2 are all solutions to the equation. Thus we say,

> Three non-collinear points determine a plane.



Example 5.17. Find the equation of the plane that contains the three points $P_0 = (1,1,0)$, $P_1 = (1,0,1)$ and $P_2 = (0,0,1)$.

We want to find coefficients a_1, a_2, a_3 and b so that all three points are solutions of the equation

$$a_1x + a_2y + a_3z = b.$$

This problem is a linear system of equations, where the variables are a_1, a_2, a_3 and the coefficients come from the points P_0, P_1, P_2!

$$\begin{aligned} a_1 + a_2 &= b \\ a_1 + a_3 &= b \\ a_3 &= b \end{aligned}$$

The augmented matrix of this system is

$$\left[\begin{array}{ccc|c} 1 & 1 & 0 & b \\ 1 & 0 & 1 & b \\ 0 & 0 & 1 & b \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 1 & 1 & 0 & b \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & b \end{array}\right]$$

Back substituting gives $a_3 = b$, $a_2 = a_3 = b$, and $a_1 = b - a_2 = 0$. Thus (taking $b = 1$), the equation of the plane is

$$y + z = 1.$$

Given a vector $\mathbf{a} = [a_1\, a_2\, a_3]^T$, the plane determined by the equation

$$a_1x + a_2y + a_3z = b$$

is the same as the set

$$\{[x,y,z]^T \in \mathbb{R}^3 \colon \mathbf{a} \cdot [x,y,z]^T = b\}$$

The vector $\mathbf{a}$ is called a normal vector to the plane. Since the plane is determined by $\mathbf{a}$ and b we say

> A normal vector and a number determine a plane.

Definition 5.18. Two planes are called *parallel* if their normal vectors are parallel.

Example 5.19. The two planes

$$2x + 4y + z = 0$$

and

$$-4x - 8y - 2z = 5$$

are parallel.

If a plane is defined by a homogeneous equation,

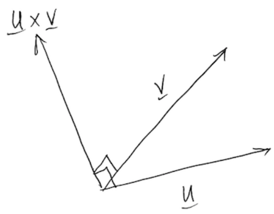
$$a_1x + a_2y + a_3z = 0$$

then the point $(0, 0, 0)$ lies in the plane. We say that the plane contains the origin.

Definition 5.20. The *cross-product* of two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$ is the vector $\mathbf{u} \times \mathbf{v}$ defined by the formula

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \times \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} := \begin{bmatrix} u_2v_3 - u_3v_2 \\ u_3v_1 - u_1v_3 \\ u_1v_2 - u_2v_1 \end{bmatrix}.$$

The cross-product has three beautiful geometric properties.

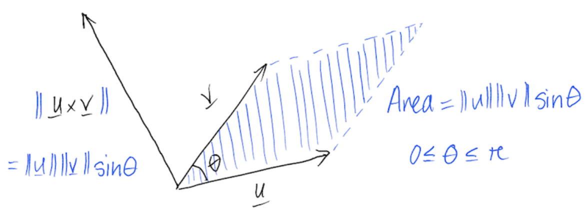


Geometric properties of the cross-product:

(a) $\mathbf{u} \times \mathbf{v}$ is orthogonal to $\mathbf{u}$ and $\mathbf{v}$.

(b) If $\mathbf{u}$ and $\mathbf{v}$ are parallel, then $\mathbf{u} \times \mathbf{v} = 0$.

(c) $||\mathbf{u} \times \mathbf{v}|| = \sqrt{||\mathbf{u}||^2||\mathbf{v}||^2 - (\mathbf{u} \cdot \mathbf{v})^2} = ||\mathbf{u}||||\mathbf{v}|| \sin(\theta)$, where $0 \leq \theta \leq \pi$ is the angle between $\mathbf{u}$ and $\mathbf{v}$.



Note that the right hand side is the formula for the area of the parallelogram with adjacent sides $\mathbf{u}$ and $\mathbf{v}$.

Proof. The proofs of these properties are straightforward applications of the definition. For instance, to prove (a), we simply compute

$$\begin{aligned}(\mathbf{u}\times\mathbf{v})\cdot\mathbf{u} &= \begin{bmatrix} u_2v_3 - u_3v_2 \\ u_3v_1 - u_1v_3 \\ u_1v_2 - u_2v_1 \end{bmatrix} \cdot \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \\ &= u_1(u_2v_3 - u_3v_2) + u_2(u_3v_1 - u_1v_3) + u_3(u_1v_2 - u_2v_1) \\ &= 0.\end{aligned}$$

Since the dot product is 0, the two vectors are orthogonal. The same works to show $(\mathbf{u}\times\mathbf{v})\cdot\mathbf{v} = 0$. $\square$

We can use the cross-product to solve the example above a different way.

Example 5.21. Find the equation of the plane that contains the three points $P_0 = (1, 1, 0)$, $P_1 = (1, 0, 1)$ and $P_2 = (0, 0, 1)$.

First, we compute the vectors

$$\overrightarrow{P_0P_1} = (0, -1, 1)$$
$$\overrightarrow{P_0P_2} = (-1, -1, 1)$$

Then we compute the cross-product,

$$\overrightarrow{P_0P_1} \times \overrightarrow{P_0P_2} = \begin{bmatrix} (-1)(1) - (1)(-1) \\ 0 - 1(-1) \\ 0 - (-1)(-1) \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ -1 \end{bmatrix}.$$

We claim that $\overrightarrow{P_0P_1}\times\overrightarrow{P_0P_2}$ is the normal vector to the plane that contains P_0, P_1, P_2. Indeed, we can check that the three points satisfy the equation

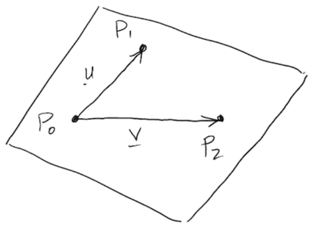
$$0x - y - z = -1.$$

Check your understanding

In examples 5.17 and 5.21 we found the equation of the plane containing the points $P_0 = (1, 1, 0)$, $P_1 = (1, 0, 1)$ and $P_2 = (0, 0, 1)$ two different ways.

Even though the two equations we found look slightly different, they describe the same plane. Why?

As the above example demonstrates, we can record the information of three points P_0, P_1, P_2 as one point P_0 and two vectors, $\overrightarrow{P_0P_1}$, $\overrightarrow{P_0P_2}$. Thus, A point and two non-parallel vectors determine a plane.



5.9 Hyperplanes in $\mathbb{R}^n$

In $\mathbb{R}^2$, a linear equation $y = mx + b$ determines a line with slope m and y-intercept b. In $\mathbb{R}^3$, a single linear equation determines a plane. In $\mathbb{R}^n$, a single linear equation determines a *hyperplane*.

Definition 5.22. The set of points in $\mathbb{R}^n$ defined by a single linear equation,

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

is called a *hyperplane*.

We can view the equation

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

as the matrix equation

$$[a_1\, a_2\, \cdots\, a_n] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = b$$

Thus a hyperplane is the set of solutions of a matrix equation.

We call the vector

$$\mathbf{a} = [a_1\, a_2\, \cdots\, a_n] \text{ or } [a_1\, a_2\, \cdots\, a_n]^T$$

a *normal vector* for the hyperplane.

A hyperplane is determined by the normal vector $\mathbf{a}$ and the number b.

Key concepts from Chapter 5

The parallelogram rule, definition of a line in $\mathbb{R}^n$, the distance between two points, the length of a vector, the dot product, properties of vector length and dot product, the Cauchy Schwartz inequality, the triangle inequality, the definition of orthogonal vectors, the definition of lines and planes in $\mathbb{R}^3$, three non-collinear points determine a plane, a point and two non-parallel vectors determine a plane, the cross-product, properties of the cross-product, definition of a hyperplane in $\mathbb{R}^n$, a point and a normal vector determine a hyperplane.

Chapter 6

Linear transformations

In calculus of several variables, one studies real valued functions of several variables, which are functions $f\colon \mathbb{R}^2 \to \mathbb{R}$, $f\colon \mathbb{R}^3 \to \mathbb{R}$ and so on, and vector fields, which are just functions $F\colon \mathbb{R}^2 \to \mathbb{R}^2$, $F\colon \mathbb{R}^3 \to \mathbb{R}^3$ and so on.

In linear algebra, we are interested in a very special class of functions $T\colon \mathbb{R}^n \to \mathbb{R}^m$ called *linear transformations.* The theory of linear transformations is intimately connected to vector geometry, the study of systems of linear equations and matrix algebra.

6.1 Functions from $\mathbb{R}^n$ to $\mathbb{R}^m$

Definition 6.1. A *function* F from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a rule that assigns a point $F(\mathbf{x}) \in \mathbb{R}^m$ to every point $\mathbf{x} \in \mathbb{R}^n$. The set $\mathbb{R}^n$ is called the *domain* of F and the set $\mathbb{R}^m$ is called the *codomain* of F.

We often use the notation $F\colon \mathbb{R}^n \to \mathbb{R}^m$ to mean "F is a function from $\mathbb{R}^n$ to $\mathbb{R}^m$."

In calculus one often considers functions whose domain is some subset of $\mathbb{R}$. In this class we focus on functions whose domain is all of $\mathbb{R}$ or all of $\mathbb{R}^n$.

Example 6.2. Here are some examples of functions from $\mathbb{R}^n$ to $\mathbb{R}^m$.

(a) A real valued function of one variable, is a function from $\mathbb{R}^1$ to $\mathbb{R}^1$. For instance,

$$f(x) = \sin(x),\ g(x) = x^2,\ h(x) = 2x + 1,\ k(x) = 5x.$$

(b) A real valued function of two variables is a function from $\mathbb{R}^2$ to $\mathbb{R}^1$. For instance,

$$T(x,y) = x^2 + 2xy + 1,\ S(x,y) = 3x - y.$$

If we write elements of $\mathbb{R}^2$ as vectors, then these functions might be written like this:

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = x^2 + 2xy\sin(x) + 1,\ S\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = 3x - y.$$

(c) The general form of a function from $\mathbb{R}^3$ to $\mathbb{R}^3$ is

$$F(x,y,z) = \left(F_1(x,y,z), F_2(x,y,z), F_3(x,y,z)\right).$$

Here $F_1(x,y,z), F_2(x,y,z), F_3(x,y,z)$ are real valued functions of three variables. In other words, each F_I is a map from $\mathbb{R}$ to $\mathbb{R}^3$.

For example, here are some functions from $\mathbb{R}^3$ to $\mathbb{R}^3$.

$$F(x,y,z) = (1,0,0),\ G(x,y,z) = (x+y, z-y, x+2),$$

$$H(x,y,z) = (x-2y, z-x, y+z+x).$$

We can also write these functions in vector form:

$$F\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix},\ G\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x+y \\ z-y \\ x+2 \end{bmatrix},$$

$$H\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x-2y \\ z-x \\ y+z+x \end{bmatrix}.$$

Check your understanding

The formula

$$F(w,x,y,z) = (x+w, y^2)$$

defines a function. What is the domain and codomain of F?

6.2 Linear transformations

In linear algebra, we are primarily interested in a very special kind of function from $\mathbb{R}^n$ to $\mathbb{R}^m$ called a "linear transformation."

Definition 6.3. A function $T : \mathbb{R}^n \to \mathbb{R}^m$ is a *linear transformation* if for all vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and for all scalars $c \in \mathbb{R}$, both of the following conditions are true:

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

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

It is common to say "T is linear" when one means "T is a linear transformation."

If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, then $T(\mathbf{0}) = \mathbf{0}$.

Check your understanding

Let $\mathbf{u} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

Suppose that $T \colon \mathbb{R}^2 \to \mathbb{R}^2$ is a linear transformation such that

$$T(\mathbf{u}) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \; T(\mathbf{v}) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

What is

$$T(3\mathbf{u} + 2\mathbf{v})?$$

If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, then for all $t_1, \cdots t_k \in \mathbb{R}$ and $\mathbf{u}_1, \ldots, \mathbf{u}_k \in \mathbb{R}^n$,

$$T(c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k) = c_1T(\mathbf{u}_1) + \cdots + c_kT(\mathbf{u}_k).$$

In other words, linear transformations send linear combinations to linear combinations.

In-class exercise

For each of the functions below, defined in Example 6.2, determine whether the function is a linear transformation.

- $$f(x) = \sin(x),\ g(x) = x^2,\ h(x) = 2x + 1,\ k(x) = 5x.$$
- $$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = x^2 + 2xy\sin(x) + 1,\ S\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = 3x - y.$$
- $$F(x, y, z) = (1, 0, 0),\ G(x, y, z) = (x + y, z - y, x + 2),$$ $$H(x, y, z) = (x - 2y, z - x, y + z + x).$$

A function $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation if for every $c \in \mathbb{R}$ and $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$,

$$T(\mathbf{u} + c\mathbf{v}) = T(\mathbf{u}) + cT(\mathbf{v}).$$

Proof. Suppose that T satisfies $T(\mathbf{u}+c\mathbf{v}) = T(\mathbf{u})+cT(\mathbf{v})$ for all vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and all scalars $c \in \mathbb{R}$. Then, by setting $c = 1$, since T is linear

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

and by setting $\mathbf{u} = \mathbf{0}$,

$$T(c\mathbf{0}) = cT(\mathbf{v}).$$

Therefore, T is a linear transformation.

□

6.3 Matrix transformations are linear

Given a $m \times n$ matrix A, matrix-vector multiplication defines a map $T_A \colon \mathbb{R}^n \to \mathbb{R}^m$ by the formula

$$T_A(\mathbf{x}) := A\mathbf{x}$$

where the expression on the right hand side is matrix-vector multiplication. Since it is defined using a matrix, we call T_A a *matrix transformation*.

Matrix transformations are linear.

Proof. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^n$ and let c be a scalar. Then,

$$\begin{aligned}
T_A(\mathbf{u}+c\mathbf{v}) &= A(\mathbf{u}+c\mathbf{v}) && \text{(Defintion of } T_A\text{)} \\
&= A\mathbf{u} + A(c\mathbf{v}) && \text{(Property of matrix multiplication)} \\
&= A\mathbf{u} + cA\mathbf{v} && \text{(Property of matrix multiplication)} \\
&= T_A(\mathbf{u}) + cT_A(\mathbf{v}) && \text{(Definition of } T_A\text{).}
\end{aligned}$$

Hence, $T_A(\mathbf{u}+c\mathbf{v}) = T_A(\mathbf{u}) + cT_A(\mathbf{v})$. This means that the function T_A is a linear transformation. □

6.4 Transformations of $\mathbb{R}^2$

Example 6.4. **Rotation around a point by θ degrees is a linear transformation**

$$F(x,y) = (\cos(\theta)x - \sin(\theta)y, \sin(\theta)x + \cos(\theta)y).$$

Example 6.5. **Translation is not a linear transformation.** If $(a,b) \in \mathbb{R}^2$, translation by (a,b) is the function

$$F(x,y) = (x,y) + (a,b) = (x+a, y+b).$$

Example 6.6. **Scaling is a linear transformation.** Given $k \in \mathbb{R}$, we can define the function

$$F(x,y) = (kx, ky).$$

Example 6.7. **Scaling in one coordinate is a linear transformation.**

$$F(x, y) = (kx, y).$$

Example 6.8. **Shear transformation is a linear transformation.**

$$F(x, y) = (x + ay, y).$$

Example 6.9. **Projection to a line through the origin is a linear transformation.**

Example 6.10. **Reflection around a line through the origin is a linear transformation.** Reflection through the y-axis:

$$F(x, y) = (-x, y).$$

Reflection through the x-axis:

$$F(x, y) = (x, -y).$$

Reflection through the line $y = x$:

$$F(x, y) = (y, x).$$

Reflection through the line $y = -x$:

$$F(x, y) = (-y, -x).$$

Example 6.11. **Reflection through the origin is a linear transformation.** Reflection through the origin:

$$F(x, y) = (-x, -y).$$

6.5 The matrix of a linear transformation

In-class exercise

We showed that the function

$$H\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x-2y \\ z-x \\ y+z+x \end{bmatrix}$$

is a linear transformation. Find a 3×3 matrix A so that

$$A\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x-2y \\ z-x \\ y+z+x \end{bmatrix}.$$

In other words, find a matrix A so that $H = T_A$.

In general, if $T\colon \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, we can always find a $m \times n$ matrix A so that $T = T_A$. Since any vector in $\mathbb{R}^n$,

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

we know that since T is a linear transformation,

$$T\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right) = x_1T(\mathbf{e}_1) + x_2T(\mathbf{e}_2) + \cdots + x_nT(\mathbf{e}_n).$$

For any $\mathbf{x} \in \mathbb{R}^n$,

$$
\begin{aligned}
T\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right) &= x_1 T(\mathbf{e}_1) + x_2 T(\mathbf{e}_2) + \cdots + x_n T(\mathbf{e}_n) \\
&= [T(\mathbf{e}_1)|T(\mathbf{e}_2)|\cdots|T(\mathbf{e}_n)] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\
&= A \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}
\end{aligned}
$$

where $A = [T(\mathbf{e}_1)|T(\mathbf{e}_2)|\cdots|T(\mathbf{e}_n)]$ is the $m \times n$ matrix whose columns are $T(\mathbf{e}_1), \ldots, T(\mathbf{e}_n)$.

Every linear transformation is a matrix transformation.

This tells us two things:

(a) The matrix of a linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^m$ is the $m \times n$ matrix whose columns are $T(\mathbf{e}_1), \ldots, T(\mathbf{e}_n)$.

(b) There is a correspondence between linear transformations $T\colon \mathbb{R}^n \to \mathbb{R}^m$ and $m \times n$ matrices.

6.6 Example: orthogonal projection onto a line

Previously, we showed that orthogonal projection onto a line through the origin spanned by a vector $\mathbf{w} \in \mathbb{R}^n$ is given by the formula

$$\operatorname{proj}_{\mathbf{w}}(\mathbf{u}) = \frac{\mathbf{u} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}} \mathbf{w}.$$

We've also showed that orthogonal projection onto a line through the origin in $\mathbb{R}^2$ is a linear transformation. In fact, this is true for orthogonal projections in $\mathbb{R}^n$ as well.

In-class exercise

Prove that the transformation $\mathrm{proj}_{\mathbf{w}} \colon \mathbb{R}^n \to \mathbb{R}^n$ is linear. Check using properties of the dot-product that:

(a) For all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$,

$$\mathrm{proj}_{\mathbf{w}}(\mathbf{u} + \mathbf{v}) = \mathrm{proj}_{\mathbf{w}}(\mathbf{u}) + \mathrm{proj}_{\mathbf{w}}(\mathbf{v})$$

(b) For all $\mathbf{u} \in \mathbb{R}^n$ and $c \in \mathbb{R}$,

$$\mathrm{proj}_{\mathbf{w}}(c\mathbf{u}) = c \cdot \mathrm{proj}_{\mathbf{w}}(\mathbf{u}).$$

As with any linear transformation, we can compute the matrix of an orthogonal projection. We now demonstrate this with an example.

Example 6.12. Let $\mathbf{w} = [1\,3\,2]^T$ Compute the matrix of the linear transformation $\mathrm{proj}_{\mathbf{w}} \colon \mathbb{R}^3 \to \mathbb{R}^3$.

We simply need to compute the image of the three standard basis vectors under this transformation.

$$\mathrm{proj}_{\mathbf{w}}(e_1) = \frac{e_1 \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}} \mathbf{w} = \frac{1}{14} \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$$

$$\mathrm{proj}_{\mathbf{w}}(e_2) = \frac{e_2 \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}} \mathbf{w} = \frac{1}{14} \begin{bmatrix} 3 \\ 9 \\ 6 \end{bmatrix}$$

$$\mathrm{proj}_{\mathbf{w}}(e_3) = \frac{e_3 \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}} \mathbf{w} = \frac{1}{14} \begin{bmatrix} 2 \\ 6 \\ 4 \end{bmatrix}$$

Thus the matrix of this linear transformation is

$$A = \frac{1}{14} \begin{bmatrix} 1 & 3 & 2 \\ 3 & 9 & 6 \\ 2 & 6 & 4 \end{bmatrix}$$

Example 6.13. The orthogonal projection of a vector $\mathbf{x}$ onto the ith coordinate axis is the vector

$$\mathrm{proj}_{e_i}(\mathbf{u}) = \frac{\mathbf{u} \cdot e_i}{e_i \cdot e_i} e_i = x_i e_i.$$

The matrix of this projection is the $n \times n$ matrix, all of whose entries are 0, except $a_{ii} = 1$.

6.7 Composition of linear transformations

If $f\colon \mathbb{R}^n \to \mathbb{R}^m$ and $g\colon \mathbb{R}^m \to \mathbb{R}^p$ are functions, then the composition of g with f is the function $g \circ f\colon \mathbb{R}^n \to \mathbb{R}^p$ defined by the formula

$$g \circ f(\mathbf{x}) := g(f(\mathbf{x})).$$

As always, brackets are important! The expression on the right hand side means "first apply the function f to the vector $\mathbf{x}$ to get $f(\mathbf{x})$, then apply the function g to the vector $f(\mathbf{x})$."

Example 6.14. For example, consider the functions $f, g\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2 + 2$, and $g(x) = x^2$. Then

$$g \circ f(x) = g(f(x)) = g(x^2 + 1) = (x^2 + 1)^2 = x^4 + 2x^2 + 1$$

and

$$f \circ g(x) = f(g(x)) = f(x^2) = (x^2)^2 + 1 = x^4 + 1.$$

Warning

The composition $g \circ f$ is only defined if the codomain of f and the domain of g are the same. For instance, if $f\colon \mathbb{R} \to \mathbb{R}^2$ and $g\colon \mathbb{R} \to \mathbb{R}$, then the composition

$$f \circ g$$

is defined, but the composition

$$g \circ f$$

is not defined.

If A is a $m \times n$ matrix and B is a $p \times m$ matrix, then by properties of matrix algebra, and the definition of composition of functions,

$$T_B \circ T_A(\mathbf{x}) = B(A\mathbf{x}) = (BA)\mathbf{x} = T_{BA}(\mathbf{x})$$

Thus we see that

If A is a $m \times n$ matrix and B is a $p \times m$ matrix, then

$$T_B \circ T_A = T_{BA}$$

This tells us two things:

(a) Composition of linear transformations corresponds to matrix multiplication!

(b) Since T_{BA} is a matrix transformation, and matrix transformations are linear, we know that:

The composition of two linear transformations is a linear transformation.

Warning

If $T\colon \mathbb{R}^n \to \mathbb{R}^n$ and $S\colon \mathbb{R}^n \to \mathbb{R}^n$, then we can define the compositions $T \circ S$ and $S \circ T$, but in general

$$T \circ S \neq S \circ T.$$

Check your understanding

Can you think of two linear transformations T and S where $T \circ S \neq S \circ T$?

6.8 Kernel and image

There are two important sets associated to every linear transformation.

Definition 6.15. The *kernel* of a linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^m$ is the set

$$\ker(T) := \{\mathbf{x} \in \mathbb{R}^n \colon T(\mathbf{x}) = \mathbf{0}\}.$$

The *image* of a linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^m$ is the set

$$\operatorname{im}(T) := \{T(\mathbf{x}) \in \mathbb{R}^m \colon \mathbf{x} \in \mathbb{R}^n\}.$$

Note that the kernel of a linear transformation is a set of vectors in the domain, $\mathbb{R}^n$, whereas the image is a set of vectors in the codomain, $\mathbb{R}^m$.

Since

$$T_A(\mathbf{x}) = A\mathbf{x}$$

we see that the set

$$\ker(T_A) = \{\mathbf{x} \in \mathbb{R}^n \colon A\mathbf{x} = \mathbf{0}\}$$

which is exactly the set of solutions of the system of linear equations $A\mathbf{x} = \mathbf{0}$.

Similarly, saying there exists $\mathbf{x} \in \mathbb{R}^n$ such that $T_A(\mathbf{x}) = \mathbf{b}$ is the same as saying that the system of linear equations

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

is consistent. Equivalently, $\mathbf{b}$ is contained in the span of the columns of A.

Problem 4. *Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be given by*

$$T(x) = \begin{bmatrix} x_1 + x_2 + x_3 \\ x_2 + 2x_3 \\ -x_1 + x_3 \end{bmatrix}.$$

Write the kernel of T as the span of a set of vectors.

Solution. *The matrix of T is*

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

Thus, we want to write the solution set of $Ax = 0$ as the span of a set of vectors. We know that the set of solutions to $Ax = 0$ is the span of the set of basic solutions, so we just need to compute the basic solutions. First, row reducing,

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

Therefore, there is one basic solution, and the solution set is

$$\ker(T) = span \left\{ \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \right\}.$$

Problem 5. *Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be given by*

$$T(x) = \begin{bmatrix} x_1 + x_2 + x_3 \\ x_2 + 2x_3 \\ -x_1 + x_3 \end{bmatrix}.$$

Write the image of T as the span of a set of vectors.

Solution. *We have computed the matrix A of T above. The image of T is the set*

$$\mathrm{im}(T) = \{\mathbf{b} \in \mathbb{R}^3 \colon A\mathbf{x} = \mathbf{b} \textit{ is consistent}\}$$

Since $A\mathbf{x} = \mathbf{b}$ is consistent if and only if $\mathbf{b}$ is contained in the span of the columns of A, we see that

$$\mathrm{im}(T) = \textit{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \right\}.$$

Definition 6.16. A linear transformation $T \colon \mathbb{R}^n \to \mathbb{R}^m$ is called *injective* (or *one-to-one*) if

$$\ker(T) = \{0\}.$$

A linear transformation $T \colon \mathbb{R}^n \to \mathbb{R}^m$ is called *surjective* (or *onto*) if

$$\mathrm{im}(T) = \mathbb{R}^m.$$

The terminology of "injective" and "surjective" for linear transformations comes from a more general definition for maps (linear or not). Since we are only concerned with linear transformations, the definition above is sufficient for the purposes of this course.

- T_A is injective if and only if $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
- T_A is surjective if and only if for every $\mathbf{b}$, the system $A\mathbf{x} = \mathbf{b}$ is consistent.

Example 6.17. The transformation T from the problems above is not injective since $\ker(T) \neq \{0\}$. It is also not surjective: since the rank of A is 2, by Theorem 6.25, the system $A\mathbf{x} = \mathbf{b}$ is not consistent for every $\mathbf{b} \in \mathbb{R}^3$, thus $\mathrm{im}(T) \neq \mathbb{R}^3$.

Example 6.18. Rotation by θ degrees in $\mathbb{R}^2$ is injective and surjective. Indeed, for any value of θ, the matrix of the linear transformation is

$$A = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

The quantity $ad - bc = \cos^2(\theta) + \sin^2(\theta) = 1$, so A is invertible (see Chapter 4). By the big theorem for square matrices, since A is invertible, system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution, so T_A is injective.

Similarly, by the big theorem for square matrices, since A is invertible, for every $\mathbf{b} \in \mathbb{R}^2$, the system $A\mathbf{x} = \mathbf{b}$ is consistent, so T_A is surjective.

Example 6.19. Let

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

The linear transformation $T_A \colon \mathbb{R}^2 \to \mathbb{R}^3$ is injective but not surjective.

The linear transformation $T_B \colon \mathbb{R}^3 \to \mathbb{R}^2$ is surjective but not injective.

Since $BA = I_2$, the composition $T_B \circ T_A = T_{BA} = T_{I_2} \colon \mathbb{R}^2 \to \mathbb{R}^2$ is both injective and surjective.

The composition $T_A \circ T_B = T_{AB} \colon \mathbb{R}^3 \to \mathbb{R}^3$ is neither injective nor surjective, since

$$AB = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

so

$$\ker(T_{AB}) = \left\{ \begin{bmatrix} 0 \\ 0 \\ z \end{bmatrix} : z \in \mathbb{R} \right\} \neq \{0\}$$

and

$$\operatorname{im}(T_{AB}) = \left\{ \begin{bmatrix} x \\ y \\ 0 \end{bmatrix} : x, y \in \mathbb{R} \right\} \neq \mathbb{R}^3.$$

In-class exercise

Let $\mathbf{w} = [1\,3\,2]^T$.
What is the image and kernel of $\operatorname{proj}_{\mathbf{w}}$?

6.9 Invertible linear transformations

Definition 6.20. A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is *invertible* if there is a map $g \colon \mathbb{R}^m \to \mathbb{R}^n$ such that

- For all $x \in \mathbb{R}^n$, $g(f(x)) = x$, and
- For all $x \in \mathbb{R}^m$, $f(g(x)) = x$.

The map g is called the *inverse* of f and sometimes denoted by f^{-1}.

Example 6.21. The function $f \colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = 5x + 2$ is invertible. The inverse $g \colon \mathbb{R} \to \mathbb{R}$ is defined by $g(x) = \frac{1}{5}(x - 2)$, since

$$f(g(x)) = 5\left(\frac{1}{5}(x-2)\right) + 2 = (x-2) + 2 = x$$

$$g(f(x)) = \frac{1}{5}((5x+2) - 2) = \frac{1}{5}(5x) = x.$$

Warning

We have now used the word "invertible" and the notation "$^{-1}$" to mean two different things:

- A square matrix A is invertible if there is a matrix B such that $AB = BA = I$. The matrix B is called the inverse of A and denoted by A^{-1}.
- A function f is invertible if there is a function g such that $f \circ g(y) = y$ and $g \circ f(x) = x$. The function g is called the inverse of f and is denoted by f^{-1}.

These two statements look very similar, and the next example shows that for linear transformations they are intimately related. But the meaning is different!

Theorem 6.22. *The following two statements are equivalent:*

(a) The linear transformation $T \colon \mathbb{R}^n \to \mathbb{R}^n$ is invertible (as a function).

(b) the $n \times n$ matrix A of the linear transformation is invertible (as a matrix).

Warning

If $n \neq m$, then a linear transformation $T \colon \mathbb{R}^n \to \mathbb{R}^m$ cannot ever be invertible. We will see why later in the course.

Proof. "$\Leftarrow$" Suppose A is the matrix of T (i.e. $T = T_A$) and A is invertible. Let $S = T_{A^{-1}}$ be the linear transformation of A^{-1}. Then for all $\mathbf{x} \in \mathbb{R}^n$,

$$T(S(\mathbf{x})) = T_A(T_{A^{-1}}(\mathbf{x})) = A(A^{-1}\mathbf{x}) = (AA^{-1})\mathbf{x} = \mathbf{x}$$

and similarly, we can show that $S(T(\mathbf{x})) = \mathbf{x}$. Thus T is invertible and $S = T^{-1}$.

"$\Rightarrow$" Suppose that T is an invertible linear transformation. By definition, this means that it has an inverse $S\colon \mathbb{R}^n \to \mathbb{R}^n$ such that $T(S(\mathbf{x})) = \mathbf{x}$ and $S(T(\mathbf{x})) = \mathbf{x}$.

- S is linear: for any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and $c \in \mathbb{R}$,

$$\begin{aligned} S(\mathbf{u} + c\mathbf{v}) &= S(T(S(\mathbf{u})) + cT(S(\mathbf{v}))) \\ &= S(T(S(\mathbf{u})) + T(cS(\mathbf{v}))) \\ &= S(T(S(\mathbf{u}) + cS(\mathbf{v}))) \\ &= S(\mathbf{u}) + cS(\mathbf{v}) \end{aligned}$$

- Since S is linear, S is the linear transformation of a $n \times n$ matrix B. Thus, for every $\mathbf{x} \in \mathbb{R}^n$,

$$\mathbf{x} = T(S(\mathbf{x})) = AB\mathbf{x}$$

  and

$$\mathbf{x} = S(T(\mathbf{x})) = BA\mathbf{x}$$

  This can only be true if $AB = AB = I_n$.

  Thus A is invertible, and the inverse of A is B.

$\square$

> If $T\colon \mathbb{R}^n \to \mathbb{R}^n$ is an invertible linear transformation whose matrix is A, then $T^{-1} = T_{A^{-1}}$

Theorem 6.23. *A linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^n$ is invertible if and only if it is injective.*

Proof. Let A be the matrix of the linear transformation T.

By the theorem above, the linear transformation T is invertible if and only if the matrix A is invertible.

By Theorem 6.25, the matrix A is invertible if and only if the system $A\mathbf{x} = 0$ has only the trivial solution.

The system $A\mathbf{x} = 0$ has only the trivial solution if and only if $\ker(T) = \{0\}$. Since T is injective means $\ker(T) = \{0\}$, we are done. $\square$

Theorem 6.24. *A linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^n$ is invertible if and only if it is surjective.*

Proof. Let A be the matrix of the linear transformation T.

By the theorem above, the linear transformation T is invertible if and only if the matrix A is invertible.

By Theorem 6.25, the matrix A is invertible if and only if for every $\mathbf{b} \in \mathbb{R}^n$, the system $A\mathbf{x} = \mathbf{b}$ is consistent.

For every $\mathbf{b} \in \mathbb{R}^n$, the system $A\mathbf{x} = \mathbf{b}$ is consistent if and only if $\mathrm{im}(T) = \mathbb{R}^n$. Since T is surjective means $\mathrm{im}(T) = \mathbb{R}^n$, we are done. □

We can summarize these results by adding three more items to the list of equivalent statements for square matrices:

Theorem 6.25. *[The big theorem for square matrices, part 2] Suppose A is a $n \times n$ matrix. The following statements are equivalent:*

(a) The matrix A is invertible.

(b) The matrix equation $A\mathbf{x} = \mathbf{b}$ has exactly one solution for all $n \times 1$ column vectors $\mathbf{b}$.

(c) The homogeneous matrix equation $A\mathbf{x} = \mathbf{0}$ has exactly one solution (the trivial solution).

(d) The rank of A is n.

(e) The reduced row-echelon form of A is I.

(f) There exists a $n \times n$ matrix B such that $AB = I$.

(g) There exists a $n \times n$ matrix B such that $BA = I$.

(h) A can be written as a product of elementary matrices.

(i) The linear transformation $T_A\colon \mathbb{R}^n \to \mathbb{R}^n$ is invertible.

(j) The linear transformation $T_A\colon \mathbb{R}^n \to \mathbb{R}^n$ is injective ($\ker(T_A) = \{0\}$).

(k) The linear transformation $T_A\colon \mathbb{R}^n \to \mathbb{R}^n$ is surjective ($\mathrm{im}(T_A) = \mathbb{R}^n$).

Key concepts from Chapter 6

Functions $\mathbb{R}^n \to \mathbb{R}^m$, definition of linear transformations, properties of linear transformations, matrix linear transformations, transformations of $\mathbb{R}^2$, computing the matrix of a linear transformation, orthogonal projections onto a line through the origin is a linear transformation, composition of functions, the composition of two linear transformations is a linear transformation, composition of linear transformations corresponds to matrix multiplication, definition of kernel and image of linear transformations, definition of invertible functions, the big theorem for square matrices (part 2).

Chapter 7

Subspaces, span, and linear independence

We have already learned about lines and hyperplanes in $\mathbb{R}^n$. We might think of these as "straight" or "flat" objects in linear algebra. When $n > 3$, there is a natural generalization of lines through the origin and hyperplanes through the origin called "subspaces." Subspaces show up many different ways: in the study of systems of linear equations, vector geometry, matrix algebra, and linear transformations.

Once we have defined what a subspace is, we are immediately interested in describing subspaces in a similar way to how we described lines through the origin (as the span of a vector) and planes in $\mathbb{R}^3$ (as the span of two non-parallel vectors). This leads to a natural question: "what is the fewest vectors we need to span a subspace?" and this question naturally leads to the notions of "linear independence," "basis" and "dimension."

7.1 Subspaces

Definition 7.1. A set of points in $\mathbb{R}^n$, S, is called a *subspace of* $\mathbb{R}^n$ if the following three conditions are satisfied:

I) (*S contains the origin*) $\mathbf{0} \in S$.

II) (*S is closed under vector addition*) If $\mathbf{u}, \mathbf{v} \in S$, then $\mathbf{u} + \mathbf{v} \in S$.

III) (*S is closed under scalar multiplication*) If $\mathbf{u} \in S$ and $c \in \mathbb{R}$, then $c\mathbf{u} \in S$.

Check your understanding

Suppose that S is a subspace of $\mathbb{R}^3$ that contains the vectors

$$\mathbf{u} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.$$

Using the fact that S is a subspace, explain why each of the following vectors is also contained in S.

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

Example 7.2. Let's look at some examples.

1. **Lines through the origin are subspaces:** Given a nonzero vector $\mathbf{w} \in \mathbb{R}^n$, the line through the origin with direction $\mathbf{w}$ is the set

$$L = \{t\mathbf{w} \in \mathbb{R}^n : t \in \mathbb{R}\} = \operatorname{span}\{\mathbf{w}\}.$$

   To show L is a subspace of $\mathbb{R}^n$, we check the three conditions:

   I) Since $0\mathbf{v} = \mathbf{0}$, the zero vector is in L.

   II) If $\mathbf{u}, \mathbf{v} \in L$, then $\mathbf{u} = t\mathbf{w}$ and $\mathbf{v} = s\mathbf{w}$ for some $t, s \in \mathbb{R}$.

   Thus, $\mathbf{u} + \mathbf{v} = t\mathbf{w} + s\mathbf{w} = (t+s)\mathbf{w} \in L$. So L is closed under vector addition.

   III) If $t\mathbf{w} \in L$, then for any $c \in \mathbb{R}$, $ct\mathbf{w} \in L$. Thus L is closed under scalar multiplication.

   Thus we see that the line L is a subspace of $\mathbb{R}^n$.

2. **Lines not through the origin are not subspaces:** A line that does not contain the point $\mathbf{0}$ fails condition I) of the definition.

3. **Planes through the origin in $\mathbb{R}^3$ are subspaces:** This is true because of the next example, which is more general (a hyperplane in $\mathbb{R}^3$ is a plane).

4. **Hyperplanes through the origin are subspaces:** Recall that a hyperplane H in $\mathbb{R}^n$, containing the origin, is defined by a single homogeneous equation:

$$H = \{\mathbf{x} \in \mathbb{R}^n : a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0\}.$$

   To show H is a subspace of $\mathbb{R}^n$, we check the three conditions:

I) Since $a_1 0 + a_2 0 + \cdots + a_n 0 = 0$, we see that the zero vector is in H.

II) If $\mathbf{x}, \mathbf{y} \in H$, then

$$\begin{aligned} a_1(x_1 + y_1) + a_2(x_2 + y_2) + \cdots + a_n(x_n + y_n) &= a_1x_1 + a_2x_2 + \cdots + a_nx_n \\ &\quad + a_1y_1 + a_2y_2 + \cdots + a_ny_n \\ &= 0 + 0 = 0. \end{aligned}$$

Thus, $\mathbf{x} + \mathbf{y} \in H$. So H is closed under vector addition.

III) If $\mathbf{x} \in H$, then for any $c \in \mathbb{R}$,

$$\begin{aligned} a_1(cx_1) + a_2(cx_2) + \cdots + a_n(cx_n) &= c(a_1x_1 + a_2x_2 + \cdots + a_nx_n) \\ &= c(0) = 0 \end{aligned}$$

Thus $c\mathbf{x} \in H$.

Thus we see that the hyperplane H is a subspace of $\mathbb{R}^n$.

5. **The circle in $\mathbb{R}^2$ centred at $(1, 0)$ with radius 1 is not a subspace:** The circle in $\mathbb{R}^2$ centred at $(1, 0)$ with radius 1 is the set of vectors

$$S = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 \colon \sqrt{(x-1)^2 + y^2} = 1 \right\}$$

Let's see if this set satisfies the conditions for being a subspace.

I) S satisfies this condition.

II) S does not satisfy this condition. The points $(1, 1)$ and $(2, 0)$ are contained in S but $(1, 1) + (2, 0) = (3, 1)$ is not contained in S.

III) S does not satisfy this condition. The point $(2, 0)$ is contained in S but the point $2 \cdot (2, 0) = (4, 0)$ is not.

6. **The set $\mathbb{R}^n$ itself a subspace of $\mathbb{R}^n$.** There is almost nothing to check:

I) $\mathbf{0} \in \mathbb{R}^n$.

II) The sum of two vectors in $\mathbb{R}^n$ is a vector in $\mathbb{R}^n$.

III) Scalar multiples of a vector in $\mathbb{R}^n$ is a vector in $\mathbb{R}^n$.

In-class exercise

Is the set $S = \{(0, \ldots, 0)\}$ a subspace of $\mathbb{R}^n$? Check the three conditions.

Problem 6. *Is the set of solutions of the equation*

$$\begin{bmatrix} 1 & 3 & 2 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & -1 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$

a subspace of $\mathbb{R}^4$*?*

Solution. *The set of solutions is not a subspace. For the set of solutions to be a subspace, it must contain the zero vector. But, the zero vector* $(0, 0, 0, 0)$ *does not solve the above equation, so the set of solutions is not a subspace (it fails condition I).* ⧫

To check if a set of vectors in $\mathbb{R}^n$ is a subspace, either the definition can be checked directly or the subspace test can be applied:

The Subspace Test

A set S of points in $\mathbb{R}^n$ is a subspace if it satisfies two conditions:

a) S is non-empty (i.e. there is at least one vector in S), and

b) for all vectors $\mathbf{u}, \mathbf{v} \in S$ and for all scalars $c \in \mathbb{R}$, the vector $\mathbf{u} + c\mathbf{v}$ is in S.

Every line through the origin in $\mathbb{R}^n$ can be expressed as the span of a vector. We saw that lines through the origin are subspaces of $\mathbb{R}^n$. More generally,

Let $\mathbf{u}_1, \dots, \mathbf{u}_k$ be vectors in $\mathbb{R}^n$. Then, the set

$$S = \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$$

is a subspace of $\mathbb{R}^n$.

Proof. Applying the subspace test will verify that S is a subspace. It needs to be checked that for every $\mathbf{w}, \mathbf{v} \in S$ and for every scalar $c \in \mathbb{R}$, the vector $\mathbf{w} + c\mathbf{v}$ is in S.

Because the vectors $\mathbf{w}$ and $\mathbf{v}$ are in $S = \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$, they are both linear combinations of $\mathbf{u}_1, \dots, \mathbf{u}_k$. That is, there exists scalars $s_1, \dots, s_k$ and $t_1, \dots, t_k$ such that

$$\mathbf{w} = s_1\mathbf{u}_1 + s_2\mathbf{u}_2 + \cdots + s_k\mathbf{u}_k \quad \text{and} \quad \mathbf{v} = t_1\mathbf{u}_1 + t_2\mathbf{u}_2 + \cdots + t_k\mathbf{u}_k.$$

Thus,

$$\begin{aligned}\mathbf{w} + c\mathbf{v} &= (s_1\mathbf{u}_1 + s_2\mathbf{u}_2 + \cdots + s_k\mathbf{u}_k) + (t_1\mathbf{u}_1 + t_2\mathbf{u}_2 + \cdots + t_k\mathbf{u}_k) \\ &= (s_1 + ct_1)\mathbf{u}_1 + (s_2 + ct_2)\mathbf{u}_2 + \cdots + (s_k + ct_k)\mathbf{u}_k.\end{aligned}$$

Since $\mathbf{w} + c\mathbf{v}$ can be written as a linear combination of the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$, we have shown that

$$\mathbf{w} + c\mathbf{v} \in \operatorname{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\} = S.$$

By the subspace test, we have shown that S is a subspace. □

Check your understanding

The set of solutions of a system of linear equations $A\mathbf{x} = \mathbf{b}$ is a set of vectors in $\mathbb{R}^n$.

For which values of $\mathbf{b}$ is the set of solutions a subspace of $\mathbb{R}^n$?

7.2 The span of a set of vectors

As we saw in the previous section, the span of a set of vectors in $\mathbb{R}^n$ is a subspace of $\mathbb{R}^n$. There are many different ways to express a subspace as the span of a set of vectors.

Example 7.3. $\mathbb{R}^3$ can be written as a span of a set of vectors many different ways.

1. $\mathbb{R}^3$ can be written as the span of the set of standard basis vectors,

$$\mathbb{R}^3 = \operatorname{span}\left\{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}\right\}.$$

   Since for any vector in $\mathbb{R}^3$,

$$\begin{bmatrix}x\\y\\z\end{bmatrix} = x\begin{bmatrix}1\\0\\0\end{bmatrix} + y\begin{bmatrix}0\\1\\0\end{bmatrix} + z\begin{bmatrix}0\\0\\1\end{bmatrix}.$$

2. $\mathbb{R}^3$ can be written as the span of the set of standard basis vectors along with any other vectors. For instance,

$$\mathbb{R}^3 = \operatorname{span}\left\{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}4\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}, \begin{bmatrix}1\\2\\-4\end{bmatrix}, \begin{bmatrix}3\\1\\1\end{bmatrix}\right\}.$$

Since for any vector in $\mathbb{R}^3$,

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = x \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 0 \begin{bmatrix} 4 \\ 0 \\ 0 \end{bmatrix} + y \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + z \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + 0 \begin{bmatrix} 1 \\ 2 \\ -4 \end{bmatrix} + 0 \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}$$

3. $\mathbb{R}^3$ can be written as the span of three vectors many other ways. For instance,

$$\mathbb{R}^3 = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} \right\}.$$

4. Since two vectors in $\mathbb{R}^3$ either span the set $\{0\}$ (if they are both zero), or a line (if they are parallel), or a plane (if they are not parallel), then for any two vectors in $\mathbb{R}^3$,

$$\mathbb{R}^3 \neq \operatorname{span}\{\mathbf{u}, \mathbf{v}\}.$$

   We need at least 3 vectors to span $\mathbb{R}^3$!

Check your understanding

Consider the two spans,

$$\operatorname{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\},$$

$$\operatorname{span}\left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 5 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \right\}.$$

Are these two subspaces of $\mathbb{R}^3$ the same or different? Can you describe them geometrically?

Example 7.4. Let's look at some more examples.

1. **Two parallel vectors span the same line:** As we saw earlier, if $k \neq 0$, then

$$\operatorname{span}\{k\mathbf{x}\} = \operatorname{span}\{\mathbf{x}\}$$

   On the other hand, if $\mathbf{x}$ is nonzero and $k = 0$, then

$$\operatorname{span}\{k\mathbf{x}\} = \{0\} \neq \operatorname{span}\{\mathbf{x}\}.$$

2. **Adding a linear combination of vectors does not change the span:** Given a set of vectors, $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$, adding any linear combination

$$s_1\mathbf{u}_1 + \ldots + s_k\mathbf{u}_k$$

to the set does not change its span:

$$\text{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\} = \text{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k, s_1\mathbf{u}_1 + \ldots + s_k\mathbf{u}_k\}$$

3. **Listing the vectors in a different order does not change the span:** For instance, given two vectors in $\mathbb{R}^3$,

$$\text{span}\{\mathbf{x}, \mathbf{y}\} = \text{span}\{\mathbf{x}, \mathbf{y}\}$$

4. **Multiplying a vector by a nonzero number does not change the span:** For instance, given two vectors in $\mathbb{R}^3$,

$$\text{span}\{\mathbf{x}, \mathbf{y}\} = \text{span}\{\mathbf{x}, 2\mathbf{y}\}$$

5. **Adding a multiple of one vector to another does not change the span:** For instance, given two vectors in $\mathbb{R}^3$,

$$\text{span}\{\mathbf{x}, \mathbf{y}\} = \text{span}\{\mathbf{x}, 2\mathbf{x} + \mathbf{y}\}$$

Showing two spans are equal:

In general, given two sets of vectors in $\mathbb{R}^n$, $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ and $\{\mathbf{v}_1, \ldots, \mathbf{v}_i\}$,

$$\text{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\} = \text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_i\}$$

if we can show two things:

- All of the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k \in \text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_i\}$.

  If we know this is true, then we can show that every linear combination $t_1\mathbf{u}_1 + \ldots + t_k\mathbf{u}_k \in \text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_i\}$, so every element of $\text{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an element of $\text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_i\}$.

- All of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_i \in \text{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$.

  If we know this is true, then we can show that every linear combination $t_1\mathbf{v}_1 + \ldots + t_i\mathbf{v}_i \in \text{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$, so every element of $\text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_i\}$ is an element of $\text{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$.

Checking all of these inclusions can be quite a chore.

In-class exercise

Consider the two spans,

$$\text{span}\left\{\begin{bmatrix}1\\1\\-1\end{bmatrix}, \begin{bmatrix}0\\1\\1\end{bmatrix}\right\},$$
$$\text{span}\left\{\begin{bmatrix}1\\2\\0\end{bmatrix}, \begin{bmatrix}1\\6\\5\end{bmatrix}\right\}.$$

Are these sets the same, or different?

7.3 Linear independence

Definition 7.5. A set of vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is *linearly independent* if the only solution to the equation

$$x_1\mathbf{u}_1 + \cdots x_k\mathbf{u}_k = \mathbf{0}$$

is the trivial solution, $(0, \ldots, 0)$.

If a set of vectors is not linearly independent, then we say that it is *linearly dependent.*

Check your understanding

Is the set of vectors

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

linearly independent?

Is the set of vectors

$$\left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 5 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \right\}$$

linearly independent?

Example 7.6. Let's look at some examples.

1. **A set of one nonzero vector is linearly independent:** If $\mathbf{u} \neq \mathbf{0}$, then the only solution to the equation
$$x\mathbf{u} = \mathbf{0}$$
is $x = 0$, so by definition, the set $\{\mathbf{u}\}$ is linearly independent.

   On the other hand, the set $\{\mathbf{0}\}$ is linearly dependent, since the equation
$$x\mathbf{0} = \mathbf{0}$$
has many solutions (x can be any number).

2. **Any set containing 0 is linearly dependent:** If a set of vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ contains the vector $\mathbf{0}$, then it is linearly dependent. Indeed, if $\mathbf{u}_j = \mathbf{0}$ for some $1 \leq j \leq k$, then
$$0\mathbf{u}_1 + \cdots + 1\mathbf{u}_j + \cdots + 0\mathbf{u}_k = \mathbf{0} + \cdots + 1\mathbf{0} + \cdots + \mathbf{0} = \mathbf{0},$$
so $(0, \ldots, 1, \ldots, 0)$ is a nontrivial solution to the equation
$$x_1\mathbf{u}_1 + \cdots + x_k\mathbf{u}_k = \mathbf{0},$$
so the set of vectors is not linearly independent. In other words, it is linearly dependent.

3. **A set of two nonzero vectors is linearly independent if and only if they are not parallel:** Suppose we have a set of two nonzero vectors $\{\mathbf{u}, \mathbf{v}\}$ in $\mathbb{R}^n$.

   If $\mathbf{u}$ and $\mathbf{v}$ are parallel, then there is a number k such that $\mathbf{v} = k\mathbf{u}$. If this is true, then
   $$(-k)\mathbf{u} + \mathbf{v} = -\mathbf{v} + \mathbf{v} = \mathbf{0}$$
   so the set $\{\mathbf{u}, \mathbf{v}\}$ is linearly dependent.

   On the other hand, if $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent, then there is a nontrivial solution (x_1, x_2) to the equation
   $$x_1\mathbf{u} + x_2\mathbf{v} = \mathbf{0}$$
   so
   $$\mathbf{u} = -\frac{x_2}{x_1}\mathbf{v}$$
   so $\mathbf{u}$ and $\mathbf{v}$ are parallel.

Warning

In the previous example we explained that for a set of two nonzero vectors $\{\mathbf{u}, \mathbf{v}\}$, the following two statements are equivalent:

(a) $\{\mathbf{u}, \mathbf{v}\}$ is linearly independent.

(b) $\mathbf{u}$ and $\mathbf{v}$ are not parallel.

This doesn't work so simply for three or more vectors, it is best to use the definition of linearly independent when you have a set of three or more vectors.

If we view the equation
$$x_1\mathbf{u}_1 + \cdots x_k\mathbf{u}_k = \mathbf{0}$$
as the matrix-vector equation
$$[\mathbf{u}_1 | \cdots | \mathbf{u}_k] \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{7.7}$$

then we see that

The following statements are equivalent:

- The set $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is linearly independent
- The matrix equation (7.7) has only the trivial solution.
- The matrix $[\mathbf{u}_1|\cdots|\mathbf{u}_k]$ has rank k.

Determining if a set of vectors is linearly independent:

We know that linear independence of a set of vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is equivalent to the matrix $[\mathbf{u}_1|\cdots|\mathbf{u}_k]$ having rank k.

Thus in order to determine if a set of vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is linearly independent, we do the following.

1. Write the vectors as columns of a matrix $A = [\mathbf{u}_1|\cdots|\mathbf{u}_k]$.
2. Use row operations to put the matrix A into REF.
3. From the REF, read off the number of leading entries (rank) of the matrix A.

Once this is done, there are two possibilities:

a. If $\text{rank}(A) = k$ (the number of vectors in the set), then the set $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is linearly independent.

b. If $\text{rank}(A) < k$, then the set $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is linearly dependent.

In-class exercise

The following statement is true:

Every set of more than n vectors in $\mathbb{R}^n$ is linearly dependent.

For instance, every set of 4 vectors in $\mathbb{R}^3$ is linearly dependent.

Can you explain why?

7.4 Basis

Definition 7.8. Let S be a subspace of $\mathbb{R}^n$. A set of vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ in S is a *basis* for S if both of the following conditions are satisfied:

I) $\operatorname{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\} = S$.

II) The set $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is linearly independent.

The vectors in a basis are called *basis vectors*.

Check your understanding

Find a basis for the subspace

$$\operatorname{span}\left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 5 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \right\}$$

From our characterizations of span and linear independence in previous sections, we see that

A set of vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ in S is a basis for S if

I) For every $\mathbf{b} \in S$, the matrix equation

$$[\mathbf{u}_1|\cdots|\mathbf{u}_k]\mathbf{x} = \mathbf{b}$$

is consistent.

II) The matrix equation

$$[\mathbf{u}_1|\cdots|\mathbf{u}_k]\mathbf{x} = \mathbf{0}$$

has exactly one solution.

We can phrase this slightly differently

A set of vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ in S is a basis for S if

I) For every $\mathbf{b} \in S$, the matrix equation

$$[\mathbf{u}_1|\cdots|\mathbf{u}_k]\mathbf{x} = \mathbf{b}$$

is consistent.

II) The matrix $[\mathbf{u}_1|\cdots|\mathbf{u}_k]$ has rank k (the number of vectors).

Example 7.9. Let's look at some examples.

1. **A basis for a line through the origin:** Given a nonzero vector $\mathbf{w} \in \mathbb{R}^n$, the line through the origin with direction $\mathbf{w}$ is the set

$$L = \{t\mathbf{w} \in \mathbb{R}^n : t \in \mathbb{R}\} = \text{span}\{\mathbf{w}\}.$$

   Since $\mathbf{w} \neq \mathbf{0}$, the set $\{\mathbf{w}\}$ is linearly independent.

   Thus, $\{\mathbf{w}\}$ is a basis for L.

2. **A basis for a plane in $\mathbb{R}^3$:** A plane $\mathcal{P}$ in $\mathbb{R}^3$ is spanned by two nonparallel vectors $\mathbf{u}$ and $\mathbf{v}$. Moreover, two nonparallel vectors are linearly independent. Thus, the set $\{\mathbf{u}, \mathbf{v}\}$ is a basis for the plane $\mathcal{P} = \text{span}\{\mathbf{u}, \mathbf{v}\}$.

3. **A basis for the set $\mathbb{R}^n$.** We have shown that the standard basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ span $\mathbb{R}^n$ and are linearly independent, so they are a basis for $\mathbb{R}^n$.

Let B be a basis for a subspace S. Then, each vector in S can be written as a unique linear combination of basis vectors.

Finding a basis for the span of a set of vectors:

Given k vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k \in \mathbb{R}^n$, we know that $S = \text{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is a subspace of $\mathbb{R}^n$.

In order to find a basis for S, we do the following:

1. Write the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ as the columns of a $n \times k$ matrix $A = [\mathbf{u}_1 | \ldots | \mathbf{u}_k]$.
2. Row-reduce the matrix A until it is in row echelon form R.
3. Identify the columns $\mathbf{c}_{j_1}, \ldots, \mathbf{c}_{j_l}$ of the matrix R that contain leading entries.
4. The set of corresponding vectors $\{\mathbf{u}_{j_1}, \ldots, \mathbf{u}_{j_l}\}$ is a basis for S.

Proof. In order to prove that $\{\mathbf{u}_{j_1}, \ldots, \mathbf{u}_{j_l}\}$ is a basis for S, we must prove that

- (span) $\text{span}\{\mathbf{u}_{j_1}, \ldots, \mathbf{u}_{j_l}\} = S$
- (linear independence) $\{\mathbf{u}_{j_1}, \ldots, \mathbf{u}_{j_l}\}$ is linearly independent.

Span: Since R is in row echelon form, the span of the vectors containing leading entries $\mathbf{c}_{j_1}, \ldots, \mathbf{c}_{j_l}$ equals the span of the set of vectors $\mathbf{c}_1, \ldots, \mathbf{c}_k$ (see exercises from earlier in the chapter).

Since $A \to R$ by elementary row operations, there is an invertible product of elementary matrices, call it E, so that $R = EA$. Then $A = E^{-1}R$. Since

$$\text{span}\{\mathbf{c}_{j_1}, \ldots, \mathbf{c}_{j_l}\} = \text{span}\{\mathbf{c}_1, \ldots, \mathbf{c}_k\}$$

it follows that

$$\begin{aligned}\text{span}\{\mathbf{u}_{j_1}, \ldots, \mathbf{u}_{j_l}\} &= \text{span}\{E^{-1}\mathbf{c}_{j_1}, \ldots, E^{-1}\mathbf{c}_{j_l}\} \\ &= \text{span}\{E^{-1}\mathbf{c}_1, \ldots, E^{-1}\mathbf{c}_k\} \\ &= \text{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}\end{aligned}$$

(see the exercise in the tutorial).

Linear independence: Since R is in row echelon form, the set of vectors containing leading entries, $\{\mathbf{c}_{j_1}, \ldots, \mathbf{c}_{j_l}\}$ is linearly independent.

Since E^{-1} is invertible, it follows that the set of vectors

$$\{\mathbf{u}_{j_1}, \ldots, \mathbf{u}_{j_l}\} = \{E^{-1}\mathbf{c}_{j_1}, \ldots, E^{-1}\mathbf{c}_{j_l}\}$$

is linearly independent (see the tutorial). □

In-class exercise

Find a basis for the subspace

$$\operatorname{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 \\ 0 \\ 5 \end{bmatrix} \right\}$$

7.5 Dimension

Subspaces and bases have the following important property:

Suppose S is a subspace of $\mathbb{R}^n$.

Every basis for S has the same number of vectors.

This allows us to make the following definition.

Definition 7.10. The *dimension* of a subspace S is the number of vectors in any basis for S. The dimension of S is denoted $\dim S$.

The dimension of the subspace $\{\mathbf{0}\}$ is 0 by convention.

Example 7.11. Let's look at some examples.

1. **The dimension of a line through the origin is 1:** Given a nonzero vector $\mathbf{u}$, we saw that $\{\mathbf{u}\}$ is a basis for the line $L = \operatorname{span}\{\mathbf{u}\}$. Thus,

$$\dim L = 1.$$

2. **The dimension of $\mathbb{R}^n$ is n:** Since the set of standard basis vectors $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is a basis for $\mathbb{R}^n$,
$$\dim \mathbb{R}^n = n.$$

3. **The dimension of a plane containing the origin in $\mathbb{R}^3$ is 2:** Since two nonparallel vectors are a basis for a plane $\mathcal{P} = \operatorname{span}\{\mathbf{u}, \mathbf{v}\}$ we know that
$$\dim \mathcal{P} = 2.$$

4. **The dimension of a hyperplane containing the origin in $\mathbb{R}^n$ is $n-1$:** We will see why this is true in the next chapter.

Theorem 7.12. *Let S be a subspace with $\dim S = k \neq 0$. Let $\mathcal{U} = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ be a set of m vectors in S. Then,*

(a) If $m < k$, then $\mathcal{U}$ does not span S.

(b) If $m > k$, then $\mathcal{U}$ is not linearly independent.

(c) If $m < k$ and $\mathcal{U}$ is linearly independent, then $\mathcal{U}$ can be extended to a basis for S.

(d) If $m > k$ and $\mathcal{U}$ spans S, then $\mathcal{U}$ contains a basis for S.

(e) If $m = k$ and $\mathcal{U}$ is linearly independent, then $\mathcal{U}$ is a basis for S.

(f) If $m = k$ and $\mathcal{U}$ spans S, then $\mathcal{U}$ is a basis for S.

(g) If S' is a subspace of S and $\dim S' = k$, then $S' = S$.

Key concepts from Chapter 7

Subspace, span of a set of vectors is a subspace, determining whether two spans are the same, definition of linearly independent, determining whether a set of vectors is linearly independent, definition of a basis, finding a basis for the span of a set of vectors, dimension of a subspace, facts about dimension.

Chapter 8

The fundamental subspaces of a matrix

There are three important subspaces associated to a $m\times n$ matrix A: the null space, the column space, and the row space. In this chapter we study these subspaces and how they are related to each other and the other theories studied in this course.

8.1 The null space of a matrix

Definition 8.1. Let A be an $m \times n$ matrix. The *null space of* A is the set

$$\text{null}(A) := \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{0}\}.$$

In other words, the null space of a matrix A is the set of solutions of the homogeneous system

$$Ax = 0.$$

> The null space of $A_{m\times n}$ is a subspace of $\mathbb{R}^n$.

Proof. Apply the subspace test: Suppose that $\mathbf{u}, \mathbf{v} \in S$ and $c \in \mathbb{R}$. $\mathbf{u}, \mathbf{v} \in S$ means that both $\mathbf{u}$ and $\mathbf{v}$ are solutions to the matrix equation $A\mathbf{x} = \mathbf{0}$.

We need to check that $\mathbf{u} + c\mathbf{v} \in S$, i.e. the vector $\mathbf{u} + c\mathbf{v}$ is also a solution to $A\mathbf{x} = \mathbf{0}$. We compute,

$$\begin{aligned} A(\mathbf{u} + c\mathbf{v}) &= A\mathbf{u} + cA\mathbf{v} \\ &= \mathbf{0} + c\mathbf{0} \\ &= \mathbf{0}. \end{aligned}$$

Thus, $\mathbf{u} + c\mathbf{v}$ is also a solution to $A\mathbf{x} = \mathbf{0}$, so $\mathbf{u} + c\mathbf{v} \in S$. By the subspace test, we have shown that S is a subspace of $\mathbb{R}^n$. $\square$

Example 8.2. Lets look at some examples.

1. **The plane through the origin in $\mathbb{R}^3$ normal to a vector.** If $\mathbf{n} \in \mathbb{R}^3$ is the vector normal to a plane through the origin, $\mathcal{P}$, then

$$\mathcal{P} = \left\{\mathbf{x} \in \mathbb{R}^3 \colon \mathbf{n} \cdot \mathbf{x} = 0\right\}$$

If $A = [n_1\, n_2\, n_3]$, then

$$\operatorname{null}(A) = \left\{\mathbf{x} \in \mathbb{R}^3 \colon A\mathbf{x} = 0\right\}.$$

Since $A\mathbf{x} = \mathbf{n} \cdot \mathbf{x}$, these two sets are the same:

$$\operatorname{null}(A) = \mathcal{P}.$$

2. **Hyperplanes through the origin** More generally, if $A = [a_1\, a_2\, \cdots\, a_n]$ is a row vector, then the matrix equation

$$A\mathbf{x} = \mathbf{0}$$

is simply the homogeneous equation

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0.$$

The set of solutions to this equation is precisely the hyperplane through the origin,

$$\operatorname{null}(A) = \{\mathbf{x} \in \mathbb{R}^n \colon a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0\}.$$

Thus hyperplanes through the origin are just an example of the null space of a matrix.

3. **The line through the origin perpendicular to a plane through the origin in $\mathbb{R}^3$.** Recall that a plane through the origin in $\mathbb{R}^3$ can be expressed as the span of two linearly independent vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$:

$$\mathcal{P} = \operatorname{span}\{\mathbf{u}, \mathbf{v}\}.$$

The line through the origin in $\mathbb{R}^3$ perpendicular to $\mathcal{P}$ is the set of all $\mathbf{x} \in \mathbb{R}^3$ that are orthogonal to $\mathbf{u}$ and $\mathbf{v}$:

$$L = \left\{\mathbf{x} \in \mathbb{R}^3 \colon \mathbf{x} \cdot \mathbf{u} = 0 \text{ and } \mathbf{x} \cdot \mathbf{v} = 0\right\}.$$

If

$$A = \begin{bmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{bmatrix}$$

then

$$\operatorname{null}(A) = \left\{\mathbf{x} \in \mathbb{R}^3 \colon A\mathbf{x} = \mathbf{0}\right\} = L.$$

If A is a $m \times n$ matrix, and $T_A\colon \mathbb{R}^n \to \mathbb{R}^m$ is the matrix linear transformation of A, then

$$\text{null}(A) = \ker(T_A).$$

"The null space of A is the kernel of the transformation T_A"

Since every linear transformation is a matrix linear transformation, this also tells us:

The kernel of a linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^m$ is a subspace of $\mathbb{R}^n$.

Problem 7. *Find a basis for* null(A) *where*

$$A = \begin{bmatrix} 1 & 3 & 3 & 2 & -9 \\ -2 & -2 & 2 & -8 & 2 \\ 2 & 3 & 0 & 7 & 1 \\ 3 & 4 & -1 & 11 & -8 \end{bmatrix}.$$

Solution. *We want to find a basis for the set of solutions of the equation $A\mathbf{x} = \mathbf{0}$. Performing elementary row operations on A, we get the reduced row echelon form of A,*

$$\begin{bmatrix} 1 & 3 & 3 & 2 & -9 \\ -2 & -2 & 2 & -8 & 2 \\ 2 & 3 & 0 & 7 & 1 \\ 3 & 4 & -1 & 11 & -8 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -3 & 5 & 0 \\ 0 & 1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

We know that the set of solutions (i.e. the null space of A) to this homogeneous equation is spanned by the set of basic solutions, so

$$\textit{null}(A) = \textit{span}\left\{ \begin{bmatrix} -5 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right\}.$$

Since the two vectors in this set are not parallel, they are linearly independent. Thus, the set

$$\left\{ \begin{bmatrix} -5 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right\}.$$

spans null(A) and is linearly independent, so it is a basis for null(A). Since it has a basis of two vectors, we see that

$$\dim(\textit{null}(A)) = 2.$$

In general, the set of basic solutions is always linearly independent, so

> The set of basic solutions to the equation $A\mathbf{x} = \mathbf{0}$ is a basis for the subspace null(A).

Since the dimension of a subspace equals the number of vectors in a basis, this tells us

> Suppose A is a $m \times n$ matrix with rank r.
>
> The dimension of null(A) is $n - r$ (the number of basic solutions).

Example 8.3. Since a hyperplane in $\mathbb{R}^n$ is the null space of a $1 \times n$ matrix, and for any $1 \times n$, the system $A\mathbf{x} = \mathbf{0}$ has $n - 1$ basic solutions, the dimension of a hyperplane in $\mathbb{R}^n$ is $n - 1$.

> **Finding a basis for null(A):**
>
> In order to find a basis for null(A),
>
> (a) Find the RREF of A.
>
> (b) Using the RREF of A, write down the set of basic solutions.
>
> The set of basic solutions is a basis.

In-class exercise

> Find a basis for the null space of the matrix
>
> $$A = \begin{bmatrix} 2 & 1 & 4 & 2 \\ 2 & 1 & 1 & 1 \end{bmatrix}.$$
>
> What is the dimension of the null space?

8.2 The column space of a matrix

Definition 8.4. Let A be an $m \times n$ matrix with columns $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$. The *column space of* A is the set

$$\operatorname{col}(A) := \operatorname{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}.$$

Since the column space of a matrix is defined as the span of a set of vectors, we automatically know that:

> The column space of $A_{m\times n}$ is a subspace of $\mathbb{R}^m$.

Since

$$\operatorname{col}(A) = \operatorname{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} = \{A\mathbf{x} \in \mathbb{R}^m \colon \mathbf{x} \in \mathbb{R}^n\} = \operatorname{im}(T_A)$$

we see that

> If A is a $m\times n$ matrix, and $T_A\colon \mathbb{R}^n \to \mathbb{R}^m$ is the matrix linear transformation of A, then
>
> $$\operatorname{col}(A) = \operatorname{im}(T_A).$$
>
> *"The column space of A is the image of the transformation T_A"*

Since every linear transformation is a matrix linear transformation, this also tells us that

> The image of a linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^m$ is a subspace of $\mathbb{R}^m$.

Problem 8. *Find a basis for* $\operatorname{col}(A)$ *where*

$$A = \begin{bmatrix} 1 & 3 & 3 & 2 & -9 \\ -2 & -2 & 2 & -8 & 2 \\ 2 & 3 & 0 & 7 & 1 \\ 3 & 4 & -1 & 11 & -8 \end{bmatrix}.$$

Solution. *We want to find a basis for the set*

$$\mathit{col}(A) = \mathit{span}\left\{ \begin{bmatrix} 1 \\ -2 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ -2 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ -8 \\ 7 \\ 11 \end{bmatrix}, \begin{bmatrix} -9 \\ 2 \\ 1 \\ -8 \end{bmatrix} \right\}.$$

This is a set of 5 vectors in $\mathbb{R}^4$, so it is linearly dependent. We need to remove some vectors so that the span is the same, but it is linearly independent.

We know from the previous section that reduced row echelon form of A is

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

The columns of this row echelon form that contain leading entries are columns 1,2,5. By the result of the previous chapter, this tells us that the set

$$\left\{ \begin{bmatrix} 1 \\ -2 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ -2 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} -9 \\ 2 \\ 1 \\ -8 \end{bmatrix} \right\}$$

is a basis for $col(A)$. Since it has a basis of three vectors, we see that

$$\dim(col(A)) = 3.$$

Warning

Elementary row operations may change the span of the columns of a matrix! If $A \to R$ by elementary row operations, then $\mathrm{col}(A) \neq \mathrm{col}(R)$.

Suppose A is a $m \times n$ matrix with rank r.

The dimension of $\mathrm{col}(A)$ is r (the number of linearly independent columns).

Finding a basis for col(A):

Follow the algorithm from Chapter 7 on finding a basis for the span of a set of vectors.

In-class exercise

Find a basis for the column space of the matrix

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

What is the dimension of the column space?

8.3 The row space of a matrix

Definition 8.5. Let A be an $m \times n$ matrix with rows denoted $\mathbf{r}_1, \ldots, \mathbf{r}_m \in \mathbb{R}^n$. The *row space of* A is the set

$$\text{row}(A) := \text{span}\{\mathbf{r}_1, \ldots, \mathbf{r}_m\}.$$

Since the row space of a matrix is defined as the span of a set of vectors, we automatically know that:

The row space of $A_{m\times n}$ is a subspace of $\mathbb{R}^n$.

Since the rows of A are the columns of A^T, we automatically know that

Since the rows of A are the columns of A^T, we see that

$$\text{row}(A) = \text{col}(A^T)$$

"The row space of A is the column space of A^T"

Also, since the columns of A are the rows of A^T,

$$\text{col}(A) = \text{row}(A^T).$$

"The column space of A is the row space of A^T"

Problem 9. *Find a basis for* row(A) *where*

$$A = \begin{bmatrix} 1 & 3 & 3 & 2 & -9 \\ -2 & -2 & 2 & -8 & 2 \\ 2 & 3 & 0 & 7 & 1 \\ 3 & 4 & -1 & 11 & -8 \end{bmatrix}.$$

Solution. *We want to find a basis for the set*

$$row(A) = span\left\{\begin{bmatrix}1\\3\\3\\2\\-9\end{bmatrix}, \begin{bmatrix}-2\\-2\\2\\-8\\2\end{bmatrix}, \begin{bmatrix}2\\3\\0\\7\\1\end{bmatrix}, \begin{bmatrix}3\\4\\-1\\11\\-8\end{bmatrix}\right\}.$$

We know from the previous section that reduced row echelon form of A is

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

Since row operations do not change the span of the rows columns, we know that

$$row(A) = span\left\{\begin{bmatrix}1\\0\\-3\\5\\0\end{bmatrix}, \begin{bmatrix}0\\1\\2\\-1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\0\\1\end{bmatrix}, \begin{bmatrix}0\\0\\0\\0\\0\end{bmatrix}\right\}.$$

Of course, this set contains a zero vector, so it is linearly dependent. If we throw away the zero vector, then the span does not change, so we have

$$row(A) = span\left\{\begin{bmatrix}1\\0\\-3\\5\\0\end{bmatrix}, \begin{bmatrix}0\\1\\2\\-1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\0\\1\end{bmatrix}\right\}.$$

Thus, the set

$$\left\{\begin{bmatrix}1\\0\\-3\\5\\0\end{bmatrix}, \begin{bmatrix}0\\1\\2\\-1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\0\\1\end{bmatrix}\right\}$$

spans $row(A)$ and is linearly independent, so it is a basis for $row(A)$. Since it has a basis of three vectors, we see that

$$\dim(row(A)) = 3.$$

In general,

The set of nonzero row vectors in a row echelon form of the matrix A is a basis for the subspace $\text{row}(A)$.

Suppose A is a $m \times n$ matrix with rank r.

The dimension of $\text{row}(A)$ is r (the number of nonzero rows in a REF of A).

Finding a basis for row(A):

1. Find a REF of A.
2. The non-zero rows of the REF are a basis for $\text{row}(A)$.

In-class exercise

Find a basis for the row space of the matrix

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

What is the dimension of the row space?

8.4 Row and column spaces

Instead of saying "dimension of the null space" some people like to use the word "nullity". I have no idea why.

Definition 8.6. Let A be a matrix. The *nullity of* A is the dimension of the null space of A. That is,

$$\text{nullity}(A) := \dim(\text{null}(A)).$$

Theorem 8.7 (Rank-Nullity Theorem)**.** *Let A be a $m \times n$ matrix. Then,*

$$\text{rank}(A) + \text{nullity}(A) = n = \#\{\text{columns of } A\}.$$

Proof. We saw in the previous sections that if $r = \text{rank}(A)$, then

$$\text{nullity}(A) = \dim(\text{null}(A)) = n - r.$$

Thus

$$\text{rank}(A) + \text{nullity}(A) = r + (n - r) = n.$$

□

Check your understanding

Check that the rank-nullity theorem is true for the exercise from the lecture. Check that

$$\text{rank}\left(\begin{bmatrix} 2 & 1 & 4 & 2 \\ 2 & 1 & 1 & 1 \end{bmatrix}\right) + \text{nullity}\left(\begin{bmatrix} 2 & 1 & 4 & 2 \\ 2 & 1 & 1 & 1 \end{bmatrix}\right) = 4.$$

8.5 Big theorems

First off, we can three more facts to our big list for square matrices.

Theorem 8.8 (The big theorem for $n \times n$ matrices, part 3)**.** *Suppose A is a $n \times n$ matrix. The following statements are equivalent:*

(a) The matrix A is invertible.

(b) The matrix equation $A\mathbf{x} = \mathbf{b}$ has exactly one solution for all $n \times 1$ column vectors $\mathbf{b}$.

(c) The homogeneous matrix equation $A\mathbf{x} = \mathbf{0}$ has exactly one solution (the trivial solution).

(d) The rank of A is n.

(e) The reduced row-echelon form of A is I.

(f) There exists a $n \times n$ matrix B such that $AB = I$.

(g) There exists a $n \times n$ matrix B such that $BA = I$.

(h) A can be written as a product of elementary matrices.

(i) The linear transformation $T_A \colon \mathbb{R}^n \to \mathbb{R}^n$ is invertible.

(j) The linear transformation $T_A \colon \mathbb{R}^n \to \mathbb{R}^n$ is injective ($\ker(T_A) = \{0\}$).

(k) The linear transformation $T_A \colon \mathbb{R}^n \to \mathbb{R}^n$ is surjective ($\operatorname{im}(T_A) = \mathbb{R}^n$).

(l) null$(A) = \{\mathbf{0}\}$.

(m) col$(A) = \mathbb{R}^n$.

(n) row$(A) = \mathbb{R}^n$.

We can also add two other theorems. These theorems apply to matrices that might not be square.

Theorem 8.9 (List of equivalent statements for $m \times n$ matrices)**.** *Suppose A is a $m \times n$ matrix. The following statements are equivalent:*

(a) The rank of A is n.

(b) The homogeneous matrix equation $A\mathbf{x} = \mathbf{0}$ has exactly one solution (the trivial solution).

(c) row$(A) = \mathbb{R}^n$.

(d) The columns of A are linearly independent.

(e) There exists a $n \times m$ matrix B such that $BA = I$.

(f) The linear transformation $T_A \colon \mathbb{R}^n \to \mathbb{R}^m$ is injective ($\ker(T_A) = \{0\}$).

(g) null$(A) = \{\mathbf{0}\}$.

Example 8.10. Let

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

- The rank of A is 2.
- The matrix equation $A\mathbf{x} = \mathbf{0}$ has exactly one solution.
- The rows of A are $[1\,0], [0\,1]$, $[1\ -1]$. The span of these rows is $\mathbb{R}^2$.
- The columns of A are linearly independent.
- The matrix

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

  is a left inverse to A, $BA = I_2$.

- We saw in the previous chapter that T_A is injective.
- The null space of A is $\{0\}$

Theorem 8.11. *[List of equivalent statements for $m \times n$ matrices] Suppose A is a $m \times n$ matrix. The following statements are equivalent:*

(a) The rank of A is m.

(b) The matrix equation $A\mathbf{x} = \mathbf{b}$ is consistent for all $n \times 1$ column vectors $\mathbf{b}$.

(c) col$(A) = \mathbb{R}^m$.

(d) The rows of A are linearly independent.

(e) There exists a $n \times m$ matrix B such that $AB = I$.

(f) The linear transformation $T_A \colon \mathbb{R}^n \to \mathbb{R}^n$ is surjective ($\mathrm{im}(T_A) = \mathbb{R}^m$).

In-class exercise

Give an example of a 2×3 matrix that satisfies the conditions of Theorem 8.11.

Key concepts from Chapter 8

Definition of the null space, finding a basis for the null space, definition of the column space, finding a basis for the column space, definition of the row space, finding a basis for the row space, big theorem for square matrices, two lists of equivalent statements for $m \times n$ *matrices.*

Chapter 9

Determinants

9.1 Definition of the determinant

The determinant is a rule that assigns to each square matrix a number. The most important property of the determinant is that it gives a simple method for determining if a matrix is invertible: If the determinant of a square matrix A is non-zero, the matrix is invertible.

The definition of a the determinant of a $n \times n$ matrix is a bit strange. It is first defined for 1×1 matrices, then for 2×2 matrices, then for 3×3 matrices, and so on.

Definition 9.1. The *determinant* of a 1×1 matrix $[a]$ is a.

Example 9.2. The determinant of $[-3]$ is -3. The determinant of $[0]$ is 0.

Definition 9.3. Let A be an $n \times n$ matrix. The (i, j)*-submatrix of* A is the $(n-1) \times (n-1)$ matrix obtained by removing the i^{th} row and j^{th} column from A. It is denoted by A_{ij}:

$$A_{ij} = \begin{bmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \cdots & a_{nj} & \cdots & a_{nn} \end{bmatrix}$$

Example 9.4. Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$. The A_{12} and A_{33} submatrices of A are

$$A_{12} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 4 & 6 \\ 7 & 9 \end{bmatrix}, \quad A_{33} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix}.$$

Definition 9.5. Let A be a square matrix. The (i, j)-*cofactor of* A is

$$C_{ij}(A) = (-1)^{i+j}\det(A_{ij}).$$

If there is no ambiguity, the (i, j)-cofactor of A will be denoted $C_{ij}(A) = C_{ij}$.

Definition 9.6. Let A be an $n \times n$ matrix. The *determinant of* A is

$$\det(A) = a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_1n,$$

where a_{ij} is the (i, j) entry of A.

The sum in the above definition is called the *cofactor expansion of* A *along the* 1^{st} *row*. The cofactor expansion of A along the i^{th} row is

$$a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}.$$

The cofactor expansion of A along the j^{th} column is

$$a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}.$$

> The cofactor expansion of a square matrix along any row or column is the same.

We demonstrate this with an example.

Problem 10. *Compute the determinant of the matrix*

$$A = \begin{bmatrix} 1 & 3 \\ 4 & 7 \end{bmatrix}$$

by cofactor expansion along the first row, second row, first column, and second column.

Solution. *The submatrices and cofactors of A are computed:*

$$
\begin{aligned}
A_{11} &= \begin{bmatrix} 1 & 3 \\ 4 & 7 \end{bmatrix} = [7] &\implies\quad& c_{11} = (-1)^{1+1}\det(7) = 7 \\
A_{12} &= \begin{bmatrix} 1 & 3 \\ 4 & 7 \end{bmatrix} = [4] &\implies\quad& c_{12} = (-1)^{1+2}\det(4) = -4 \\
A_{21} &= \begin{bmatrix} 1 & 3 \\ 4 & 7 \end{bmatrix} = [3] &\implies\quad& c_{21} = (-1)^{2+1}\det(3) = -3 \\
A_{22} &= \begin{bmatrix} 1 & 3 \\ 4 & 7 \end{bmatrix} = [1] &\implies\quad& c_{22} = (-1)^{2+2}\det(1) = 1
\end{aligned}
$$

Therefore, the determinant of A is

$$
\det(A) = \begin{cases}
a_{11}C_{11} + a_{12}C_{12} = 1 \cdot 7 + 3(-4) = -5 & (1^{st} \textit{ row}) \\
a_{11}C_{11} + a_{21}C_{21} = 1 \cdot 7 + 4(-3) = -5 & (1^{st} \textit{ column}) \\
a_{21}C_{21} + a_{22}C_{22} = 4(-3) + 7 \cdot 1 = -5 & (2^{nd} \textit{ row}) \\
a_{12}C_{12} + a_{22}C_{22} = 3(-4) + 7 \cdot 1 = -5 & (2^{nd} \textit{ column}).
\end{cases} \quad \blacklozenge
$$

The determinant of a 2×2 matrix

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

is

$$
\begin{aligned}
\det(A) &= a_{11}C_{11} + a_{12}C_{12} \\
&= a_{11}a_{22} - a_{12}a_{21}.
\end{aligned}
$$

Check your understanding

Compute $\det \begin{bmatrix} 1 & 3 \\ 6 & -5 \end{bmatrix}$.

Is the matrix invertible?

The following statements are equivalent.

- The determinant of a 2×2 matrix

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

  equals 0.

- The columns of A,

$$\begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix}, \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix}$$

  are parallel (the columns are linearly dependent).

- The rows of A are parallel (the rows are linearly dependent).
- The rank of A is less than 2.
- A is not invertible.

Problem 11. *Compute* det $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$.

Solution. *By definition*

$$\det \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13}.$$

The cofactors are computed:

$$\begin{aligned}
A_{11} &= \begin{bmatrix} 5 & 6 \\ 8 & 9 \end{bmatrix} \quad \Longrightarrow \quad C_{11} = (-1)^{1+1} \det \begin{bmatrix} 5 & 6 \\ 8 & 9 \end{bmatrix} = 5 \cdot 9 - 8 \cdot 6 = -3 \\
A_{12} &= \begin{bmatrix} 4 & 6 \\ 7 & 9 \end{bmatrix} \quad \Longrightarrow \quad C_{12} = (-1)^{1+2} \det \begin{bmatrix} 4 & 6 \\ 7 & 9 \end{bmatrix} = -(4 \cdot 9 - 7 \cdot 6) = 6 \\
A_{13} &= \begin{bmatrix} 4 & 5 \\ 7 & 8 \end{bmatrix} \quad \Longrightarrow \quad C_{13} = (-1)^{1+3} \det \begin{bmatrix} 4 & 5 \\ 7 & 8 \end{bmatrix} = 4 \cdot 8 - 7 \cdot 5 = -3
\end{aligned}$$

Therefore, the determinant of A is

$$\begin{aligned}\det(A) &= a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13} \\ &= 1(-3) + 2(6) + 3(-3) \\ &= 0. \quad \blacklozenge\end{aligned}$$

The determinant of a 3×3 matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

is given by the formula

$$\begin{aligned}\det(A) &= a_{11}\det\begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix} - a_{12}\det\begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix} + a_{13}\det\begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \\ &= a_{11}(a_{22}a_{33} - a_{32}a_{23}) - a_{12}(a_{21}a_{33} - a_{31}a_{23}) + a_{13}(a_{21}a_{32} - a_{31}a_{22}).\end{aligned}$$

Problem 12. *Compute* $\det\begin{bmatrix} -2 & -2 & -2 \\ 4 & 3 & 7 \\ -1 & 0 & 10 \end{bmatrix}$.

Solution. *Using the general formula above,*

$$\begin{aligned}\det\begin{bmatrix} -2 & -2 & -2 \\ 4 & 3 & 7 \\ -1 & 0 & 10 \end{bmatrix} &= -2(3 \cdot 10 - 7 \cdot 0) + 2(4 \cdot 10 - 7(-1)) - 2(4 \cdot 0 - 3(-1)) \\ &= 28. \quad \blacklozenge\end{aligned}$$

In class exercise

Compute the determinant of

$$A = \begin{bmatrix} 1 & 2 & 0 & 4 \\ 5 & 6 & 7 & 8 \\ 0 & 10 & 0 & 0 \\ 13 & 14 & 0 & 16 \end{bmatrix}.$$

Hint: if you choose the right cofactor expansion, the computation will be shorter.

In class exercise

Compute the determinant of

$$A = \begin{bmatrix} 1 & 2 & 0 & 4 \\ 0 & 6 & 7 & 8 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix}.$$

Hint: if you choose the right cofactor expansion, the computation will be shorter.

The determinant of an upper (or lower) trianglular matrix equals the product of the diagonal entries.

9.2 Properties of the determinant

In the previous section, we observed the following fact in some examples.

Theorem 9.7. *Let A be a square matrix. The determinant of A is equal to the cofactor expansion of A along any row or column.*

From this theorem, we can deduce some properties of determinants.

Properties of the determinant:

Let A be a square matrix.

(a) If A has a row or column of zeros, then its determinant equals zero.

(b) $\det(A) = \det(A^T)$.

Proof. (a) Cofactor expanding along the zero row or zero column proves that the determinant is zero.

(b) Since the rows of A are equal to the columns of A^T, cofactor expanding A along rows gives the same result as cofactor expanding A^T along columns. $\square$

Suppose that A is a square matrix with two identical rows. Then the determinant of A is zero.

In class exercise

Compute the determinants of the following matrices.

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \; B = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 5 & 0 \\ 2 & 0 & 0 \end{bmatrix},$$

$$C = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \; D = \begin{bmatrix} 2 & 0 & -1 \\ 0 & 5 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

What elementary row operations were performed to obtain B, C, and D from the matrix A?

The following theorem, in particular, relates determinants to elementary row operations.

Theorem 9.8. *Let A be a square matrix. Elementary row operations performed on the matrix A change the determinant of A in the following ways:*

(a) Interchanging two rows of A changes the sign of the determinant. That is, if $A \xrightarrow{R_i \leftrightarrow R_j} A'$, then

$$\det(A') = -\det(A).$$

(b) Multiplying a row by a scalar c multiplies the determinant by c. That is, if $A \xrightarrow{R_j \to cR_j} A'$, then

$$\det(A') = c\det(A).$$

(c) Adding a multiple of one row to another leaves the determinant unchanged. That is, if $A \xrightarrow{R_i \to R_i + cR_j} A'$, then

$$\det(A') = \det(A).$$

We saw earlier in the course that a 2×2 matrix is invertible if and only if the determinant is not zero. In fact, this is true in general.

Theorem 9.9. *Let A be a $n \times n$ matrix. The following statements are equivalent.*

(a) A is invertible.

(b) $\det(A) \neq 0$.

The next most important property of the determinant is the following multiplicative property.

Theorem 9.10 (Multiplicative property of the determinant)**.** *Let A and B be $n \times n$ matrices. Then,*

$$\det(AB) = \det(A)\det(B).$$

This theorem is actually tricky to prove. The left hand side of the equation in the above theorem involves matrix multiplication while the right hand side involves regular multiplication.

More properties of the determinant:

Let A, $A_1, \ldots, A_k$ be $n \times n$ matrices.

(a) $\det(A_1 A_2 \cdots A_k) = \det(A_1)\det(A_2)\cdots\det(A_k)$.

(b) If k is a positive integer, then $\det(A^k) = \det(A)^k$.

(c) If A is invertible, then $\det(A^{-1}) = \det(A)^{-1}$.

(d) If $c \in \mathbb{R}$, then $\det(cA) = c^n \det(A)$.

(e) $\det(AB) = \det(BA)$, even when $AB \neq BA$.

Warning

In general, there is no nice formula relating $\det(A+B)$ and $\det(A)+\det(B)$.

We finish this section by adding one more item to our long list of equivalent statements for square matrices.

Theorem 9.11 (The big theorem for $n\times n$ matrices, part 3)**.** *Suppose A is a $n\times n$ matrix. The following statements are equivalent:*

(a) The matrix A is invertible.

(b) The matrix equation $A\mathbf{x} = \mathbf{b}$ has exactly one solution for all $n \times 1$ column vectors $\mathbf{b}$.

(c) The homogeneous matrix equation $A\mathbf{x} = \mathbf{0}$ has exactly one solution (the trivial solution).

(d) *The rank of* A *is* n.

(e) *The reduced row-echelon form of* A *is* I.

(f) *There exists a* $n \times n$ *matrix* B *such that* $AB = I$.

(g) *There exists a* $n \times n$ *matrix* B *such that* $BA = I$.

(h) A *can be written as a product of elementary matrices.*

(i) *The linear transformation* $T_A \colon \mathbb{R}^n \to \mathbb{R}^n$ *is invertible.*

(j) *The linear transformation* $T_A \colon \mathbb{R}^n \to \mathbb{R}^n$ *is injective (*$\ker(T_A) = \{0\}$*).*

(k) *The linear transformation* $T_A \colon \mathbb{R}^n \to \mathbb{R}^n$ *is surjective (*$\mathrm{im}(T_A) = \mathbb{R}^n$*).*

(l) *null*$(A) = \{\mathbf{0}\}$.

(m) *col*$(A) = \mathbb{R}^n$.

(n) *row*$(A) = \mathbb{R}^n$.

(o) $\det(A) \neq 0$.

9.3 Determinants and area

There is a simple relationship between determinants and areas.

Theorem 9.12. *Let* D *be a region of finite area in* $\mathbb{R}^2$*. Let* $T_A : \mathbb{R}^2 \to \mathbb{R}^2$ *be the linear transformation* $T_A(\mathbf{x}) = A\mathbf{x}$*. Then,*

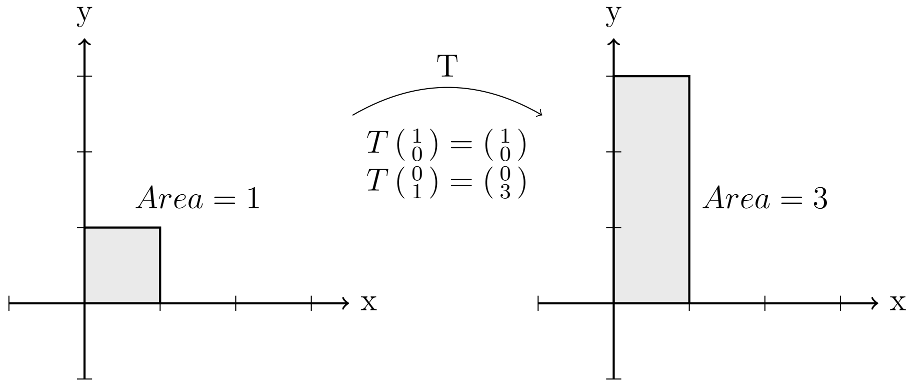
$$\mathrm{area}(T(D)) = |\det(A)|\mathrm{area}(D).$$

This is illustrated in the following examples.

Example 9.13 (Scaling in the y-direction). Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be given by

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x \\ 3y \end{bmatrix}.$$

The linear transformation T maps the unit square as follows:

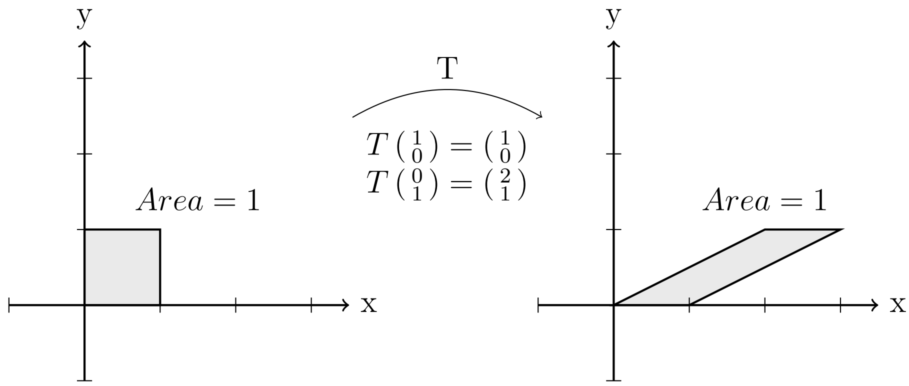


The map T sends the unit square (with area one) to the rectangle (with area three). The map T scales area by three. It is straightforward to check that the determinant of the matrix representing T is three.

Example 9.14 (Shear in the x-direction). Let $T:\mathbb{R}^2\to\mathbb{R}^2$ be given by

$$T\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}1&2\\0&1\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}x+2y\\y\end{bmatrix}.$$

The map T acts on the unit square as in the picture:



The linear transformation T does not change the area, and it is easy to check that the determinant of the matrix representing T is one. (The fact that a shear does not change the areas is an example of "Cavalieri's principle.")

Problem 13. *What is the area of enclosed by the ellipse*

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1?$$

Key concepts from Chapter 9

Definition of the determinant using cofactor expansions, computing determinants using cofactor expansions, properties of determinants, determinants of elementary matrices, determinants of products of matrices, $\det(A) \neq 0$ *iff* A *is invertible, determinants of* 2×2 *matrices and area.*

Chapter 10

Eigenvalues and eigenvectors

10.1 Eigenvalues and Eigenvectors

Definition 10.1. Let A be an $n \times n$ matrix. If there is a scalar $\lambda \in \mathbb{R}$ and a non-zero vector $u \in \mathbb{R}^n$ such that

$$Au = \lambda u,$$

then λ is an *eigenvalue* of A and u is an *eigenvector* of A corresponding to the eigenvalue λ.

Example 10.2. For each of the following 2×2 matrices, find all the eigenvalues and corresponding eigenvectors by using the definition.

(a) $A = \begin{bmatrix} 2 & 0 \\ 0 & -5 \end{bmatrix}$.

(b) $A = \frac{1}{10}\begin{bmatrix} 1 & 3 \\ 3 & 9 \end{bmatrix}$.

(c) $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$.

(d) $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$.

(e) $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

(f) $A = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$.

(g) $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.

(h) $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$.

Give a geometric description of the linear transformation $T_A \colon \mathbb{R}^2 \to \mathbb{R}^2$ corresponding to each matrix A.

10.2 Computing eigenvalues and eigenvectors

To compute the eigenvalues and eigenvectors of a matrix A, it is easiest to first find the eigenvalues of A and then find the corresponding eigenvectors.

> The following statements are equivalent:
>
> (a) $A\mathbf{u} = \lambda\mathbf{u}$ for some non-zero vector $\mathbf{u}$.
>
> (b) $(A - \lambda I)\mathbf{u} = \mathbf{0}$ for some non-zero vector $\mathbf{u}$.
>
> (c) $\text{null}(A - \lambda I) \neq \{\mathbf{0}\}$.
>
> (d) $A - \lambda I$ is not invertible.
>
> (e) $\det(A - \lambda I) = 0$.

The last item on this list of equivalent statments provides a method for computing eigenvalues of a matrix.

Definition 10.3. The polynomial

$$c_A(\lambda) = \det(A - \lambda I)$$

is called the *characteristic polynomial of* A.

> The eigenvalues of A are the roots of the polynomial
>
> $$c_A(\lambda) := \det(A - \lambda I).$$

Computing Eigenvalues of a square matrix

Given a $n \times n$ matrix A, its eigenvalues can be computed in the following way.

1. Compute the characteristic polynomial $c_A(\lambda) = \det(A - \lambda I)$ by computing the determinant of $A - \lambda I$.
2. Compute the roots of $c_A(\lambda)$ (i.e. values of λ that solve the equation $c_A(\lambda) = 0$). The roots are the eigenvalues of A.

Problem 14. *Compute the eigenvalues of the matrix*

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

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

$$\begin{aligned} c_A(\lambda) &= \det\left(\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) \\ &= \det \begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix} \\ &= (2-\lambda)^2 - 1 \\ &= \lambda^2 - 4\lambda + 3 \\ &= (\lambda - 3)(\lambda - 1). \end{aligned}$$

Since the roots of $c_A(\lambda)$ are 3 and 1, the eigenvalues of A are 3 and 1.

In-class exercise

Use the characteristic polynomial to compute the eigenvalues of the following matrices.

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

The last result in this subsection relates the eigenvalues of a square matrix to its invertibility.

Let A be a square matrix. The following statements are equivalent:

(a) The matrix A is invertible.

(b) 0 is not an eigenvalue of A

Check your understanding

Suppose that A is a square matrix and $\det(A) = 3$. Is 0 an eigenvalue of A?

10.3 Eigenspaces

The following statements are equivalent:

(a) $\mathbf{u}$ is an eigenvector of A corresponding to the eigenvalue λ.

(b) $\mathbf{u}$ is a nonzero vector and $\mathbf{u} \in \text{null}(A - \lambda I)$.

Definition 10.4. Let A be a square matrix and let λ be an eigenvalue of A. The *eigenspace of A corresponding to the eigenvalue λ* is the set

$$E_\lambda(A) := \text{null}(A - \lambda I).$$

Since it is defined as the null space of a matrix,

Eigenspaces, $E_\lambda(A)$, of an $n \times n$ matrix A are subspaces of $\mathbb{R}^n$.

Problem 15. *Find a basis for $\mathbb{R}^2$ consisting of eigenvectors of*

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

Solution. *From problem 14, the eigenvalues of A were found to be*

$$\lambda_1 = 3 \quad \textit{and} \quad \lambda_2 = 1.$$

In order to find a basis for $\mathbb{R}^2$ consisting an eigenvectors of A, a basis will be found for each eigenspace of A, and taken together, these will form a basis for A.

The eigenspaces of A are computed as follows:

$$\begin{aligned} E_3(A) &= \textit{null}\left(\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} - 3\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) \\ &= \textit{null}\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \\ &= \textit{span}\left\{\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right\}. \end{aligned}$$

Therefore, a basis for $E_3(A)$ is

$$\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}.$$

Similarly,

$$\begin{aligned} E_1(A) &= null \left(\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \\ &= null \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \\ &= span \left\{ \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}. \end{aligned}$$

Therefore, a basis for $E_1(A)$ is

$$\left\{ \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}.$$

So, a basis for $\mathbb{R}^2$ consisting of eigenvectors of A is

$$\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}.$$

♦

Problem 16. *Find the eigenvalues of the matrix*

$$A = \begin{bmatrix} 1 & 2 & -1 \\ 1 & 0 & 1 \\ 4 & -4 & 5 \end{bmatrix}.$$

Write each eigenspace of A as spanning set.

Solution. *The eigenvalues of A are the roots of its characteristic polynomial:*

$$\begin{aligned} c_A(\lambda) &= \det \left(\begin{bmatrix} 1 & 2 & -1 \\ 1 & 0 & 1 \\ 4 & -4 & 5 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) \\ &= \det \begin{bmatrix} 1-\lambda & 2 & -1 \\ 1 & -\lambda & 1 \\ 4 & -4 & 5-\lambda \end{bmatrix} \\ &= -\lambda^3 + 6\lambda^2 - 11\lambda + 6. \end{aligned}$$

After applying the rational roots theorem and using polynomial long division (or using a calculator), the characteristic polynomial factors as

$$\begin{aligned} c_A(\lambda) &= -\lambda^3 + 6\lambda^2 - 11\lambda + 6 \\ &= -(\lambda - 3)(\lambda - 2)(\lambda - 1). \end{aligned}$$

The eigenvalues of A are the roots of its characteristic polynomial: 3, 2, and 1.

The eigenspaces of A are

$$\begin{aligned} E_3(A) &= \mathit{null}(A - 3I) = \mathit{null} \begin{bmatrix} -2 & 2 & -1 \\ 1 & -3 & 1 \\ 4 & -4 & 2 \end{bmatrix} = \mathit{span} \left\{ \begin{bmatrix} -1 \\ 1 \\ 4 \end{bmatrix} \right\} \\ E_2(A) &= \mathit{null}(A - 2I) = \mathit{null} \begin{bmatrix} -1 & 2 & -1 \\ 1 & -2 & 1 \\ 4 & -4 & 3 \end{bmatrix} = \mathit{span} \left\{ \begin{bmatrix} -2 \\ 1 \\ 4 \end{bmatrix} \right\} \\ E_1(A) &= \mathit{null}(A - I) = \mathit{null} \begin{bmatrix} 0 & 2 & -1 \\ 1 & -1 & 1 \\ 4 & -4 & 4 \end{bmatrix} = \mathit{span} \left\{ \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix} \right\}. \quad \blacklozenge \end{aligned}$$

In-class exercise

Compute the eigenspaces corresponding to each eigenvalue of the following matrices.

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, C = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

10.4 Algebraic multiplicity of eigenvalues

Example 10.5. Let $p(\lambda) = (\lambda - 3)^4(\lambda - 1)(\lambda - 4)^2$. The root 3 has multiplicity 4. The root 1 has multiplicity 1. The root 4 has multiplicity 2.

Definition 10.6. Suppose λ is an eigenvalue of A. The *algebraic multiplicity* of the eigenvalue λ is the multiplicity of λ as a root of the characteristic polynomial, c_A.

Check your understanding

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

What are the eigenvalues of A? What are their algebraic multiplicities?

The next theorem relates the multiplicity of an eigenvalue of a matrix to the dimension of the corresponding eigenspace.

Theorem 10.7. *Let A be a square matrix. Suppose that λ is an eigenvalue of A with algebraic multiplicity m. Then,*

$$1 \leq \dim(E_\lambda(A)) \leq m.$$

In-class exercise

For each matrix below, compare the algebraic multiplicity of the eigenvalue 1, and the dimension of the eigenspace E_1.

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, C = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

10.5 Diagonalization

Diagonal matrices are great for a bunch of reasons:

- They commute. If A and B are both diagonal, then $AB = BA$.
- We can easily see what their eigenvalues and eigenvectors are.
- The linear transformation T_A of a diagonal matrix A acts on $\mathbb{R}^n$ in a very simple way.
- It is easy to compute their determinants and check if they are invertible.
- If a diagonal matrix is invertible, it is easy to compute its inverse.
- Computing products of diagonal matrices takes less time (computationally) than computing a product of two matrices in general.

Definition 10.8. A square matrix A is *diagonalizable* if there exists a diagonal matrix D and an invertible matrix P such that

$$A = PDP^{-1}.$$

A square matrix A is diagonalizable if and only if it has n linearly independent eigenvectors.

Check your understanding

Which of the matrices below are diagonalizable?

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, C = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, D = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

Diagonalizing a matrix:

In order to diagonalize a $n \times n$ matrix A,

1. Compute the eigenvalues of A. (if the characteristic polynomial does not factor completely, then A is not diagonalizable).
2. Compute the eigenspaces of each of the eigenvalues of A.

At this point, there are two possibilities.

(a) If the dimensions of the eigenspaces of A add to n (for each eigenvalue λ, $\dim(E_\lambda) = m_\lambda$), then A is diagonalizable. The matrices P and D can then be computed from a basis for the eigenspaces of A (see examples below).

(b) Otherwise (if the dimensions of the eigenspaces don't add to n), then A is not diagonalizable.

Problem 17. *Let*

$$A = \begin{bmatrix} 1 & 2 & -1 \\ 1 & 0 & 1 \\ 4 & -4 & 5 \end{bmatrix}.$$

Diagonalize A. That is, find a diagonal matrix D and an invertible matrix P such that $A = PDP^{-1}$.

Solution. *It can be verified (as an exercise!) that*

- *The eigenvalues of A are 1, 2, and 3.*
- *The corresponding eigenspaces are*

$$E_1(A) = span\left\{\begin{bmatrix}-1\\1\\2\end{bmatrix}\right\}, \quad E_2(A) = span\left\{\begin{bmatrix}-2\\1\\4\end{bmatrix}\right\}, \quad \textit{and,}$$

$$E_3(A) = span\left\{\begin{bmatrix}-1\\1\\4\end{bmatrix}\right\}.$$

- *A basis for $\mathbb{R}^3$ consisting of eigenvectors of A is*

$$\beta = \left\{\begin{bmatrix}-1\\1\\2\end{bmatrix}, \begin{bmatrix}-2\\1\\4\end{bmatrix}, \begin{bmatrix}-1\\1\\4\end{bmatrix}\right\}.$$

Using the observations from the previous example, choose the change of basis matrix P to be

$$P = \begin{bmatrix}-1 & -2 & -1\\ 1 & 1 & 1\\ 2 & 4 & 4\end{bmatrix} \quad \implies \quad P^{-1} = \begin{bmatrix}0 & 2 & -\frac{1}{2}\\ -1 & -1 & 0\\ 1 & 0 & \frac{1}{2}\end{bmatrix}$$

and choose the diagonal matrix

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

Note that the diagonal entries of D are the eigenvalues of A and the columns of P are eigenvectors of A, which are in a corresponding order to the eigenvalues along the diagonal of D.

It can be checked (as an exercise!) that

$$A = PDP^{-1},$$

which solves the problem. ♦

It is particularly important to note that the diagonalizability of an $n \times n$ matrix depends on whether or not that matrix has n linearly independent eigenvectors. Therefore, it is important to know when a matrix has linearly independent eigenvectors. The following theorems provide helpful information on determining the linear independence of the eigenvectors of a matrix.

Theorem 10.9. *Let A be a square matrix. If $u_1, \ldots, u_k$ are eigenvectors of A corresponding to k distinct eigenvalues, then the set $\{u_1, \ldots, u_k\}$ is linearly independent.*

This theorem will be proven at the end of this subsection. Its proof will be via induction, and can be skipped.

Let A be an $n \times n$ matrix. If A has n distinct eigenvalues, then A is diagonalizable.

Recall that if A is a square matrix and and λ is an eigenvalue of A with multiplicity m, then,

$$1 \leq \dim E_\lambda(A) \leq m.$$

Suppose that A is a square matrix and the characteristic polynomial of A factors fully (that is, it factors entirely into linear terms, with no irreducible quadratic terms). Then, if A has eigenvalues $\lambda_1, \ldots, \lambda_k$ with multiplicities $m_1, \ldots, m_k$, the following statements are equivalent:

(a) the matrix A is diagonalizable

(b) for each $1 \leq i \leq k$,

$$\dim E_{\lambda_i}(A) = m_i \quad \text{for all } i.$$

Problem 18. *Is* $H = \begin{bmatrix} 4 & -1 & -2 \\ -6 & 3 & 4 \\ 8 & -2 & -4 \end{bmatrix}$ *diagonalizable?*

Solution. *The characteristic polynomial of H is*

$$\begin{aligned} c_H(x) &= \det \begin{bmatrix} 4 & -1 & -2 \\ -6 & 3 & 4 \\ 8 & -2 & -4 \end{bmatrix} \\ &= (4-x)[(3-x)(-4-x)+8] + [-6(-4-x)-32] - 2[12-8(3-x)] \\ &= -x(x-1)(x-2). \end{aligned}$$

Therefore, the eigenvalues of H are 2, 1, and 0. Since H is 3×3 and has 3 distinct eigenvalues, by corollary Equation 10.5, it is diagonalizable. ♦

Problem 19. *Is* $A = \begin{bmatrix} 4 & -1 & 3 \\ 0 & 2 & -1 \\ 0 & 0 & 2 \end{bmatrix}$ *diagonalizable?*

Solution. *The matrix A is diagonalizable if it has 3 linearly independent eigenvectors. It can be checked (as an easy exercise!) that the eigenvalues of A are*

$$\begin{aligned} \lambda_1 &= 4 \quad (\textit{multiplicity } 1) \\ \lambda_2 &= 2 \quad (\textit{multiplicity } 2). \end{aligned}$$

Since eigenvectors corresponding to distinct eigenvalues are linearly independent (from theorem Equation 10.9) and $1 \leq \dim E_\lambda(A) \leq [\textit{mult. of } \lambda]$ (from theorem Equation 10.7). Therefore,

(a) $\dim E_4(A) = 1$

(b) $\dim E_2(A) = 1$ *or* 2. *By corollary Equation 10.5, if the dimension is 2, then the matrix A is diagonalizable, and if the dimension is 1, it is not.*

The eigenspace $E_2(A)$ is computed:

$$\begin{aligned} E_2(A) &= \textit{null} \begin{bmatrix} 4-2 & -1 & 3 \\ 0 & 2-2 & -1 \\ 0 & 0 & 2-2 \end{bmatrix} \\ &= \textit{null} \begin{bmatrix} 2 & -1 & 3 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

This corresponds to a system with a one parameter solution set, so $\dim E_2(A) = 1$. *Therefore, the matrix A is not diagonalizable.* ♦

Problem 20. *Is* $B = \begin{bmatrix} 4 & -1 & 3 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ *diagonalizable?*

Solution. *The matrix B is diagonalizable if and only if it has 3 linearly independent eigenvectors. As in the previous example, the eigenvalues of B are*

$$\begin{aligned} \lambda_1 &= 4 \quad (\textit{multiplicity } 1) \\ \lambda_2 &= 2 \quad (\textit{multiplicity } 2). \end{aligned}$$

And, as in the previous problem, the dimension of $E_2(A)$ needs to be determined. If the dimension is 2, then A is diagonalizable. If the dimension is 1, then A is not diagonalizable. The dimension is computed:

$$E_2(A) = null \begin{bmatrix} 2 & -1 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

This is a 2 dimensional subspace. Therefore, A is diagonalizable. ♦

Now, theorem Equation 10.9 is proven:

Proof of theorem Equation 10.9. It needs to be shown that if $u_1, \ldots, u_k$ are eigenvectors of a matrix A corresponding to distinct eigenvalues, then those vectors form a linearly independent set. The proof proceeds by induction.

Base case: Suppose that there is only one eigenvector, u_1. Then, since eigenvectors are non-zero by definition, the set $\{u_1\}$ is linearly independent.

Induction hypothesis: Suppose that the theorem is true for when there are k eigenvectors.

Inductive step: Suppose that $u_1, \ldots, u_k, u_{k+1}$ are $k+1$ eigenvectors for the matrix A corresponding to $k+1$ distinct eigenvalues $\lambda_1, \ldots, \lambda_{k+1}$. Then, these vectors are linearly independent if the only solution to

$$a_1u_1 + \cdots + a_ku_k + a_{k+1}u_{k+1} = 0 \tag{10.10}$$

is $a_1 = \cdots = a_{k+1} = 0$.

The linear map T given by $T(x) = Ax$ is applied to equation (10.10). Since $T(0) = 0$, this equation becomes

$$\begin{aligned} 0 &= A(a_1u_1 + \cdots + a_ku_k + a_{k+1}u_{k+1}) \\ &= \lambda_1a_1u_1 + \cdots + \lambda_ka_ku_k + \lambda_{k+1}a_{k+1}u_{k+1}. \end{aligned} \tag{10.11}$$

Then, subtracting $\lambda_{k+1}(a_1u_1 + \cdots + a_ku_k + a_{k+1}u_{k+1})$ from both sides of equation (10.11), and noting that $a_1u_1 + \cdots + a_ku_k + a_{k+1}u_{k+1} = 0$ by assumption, yields

$$\begin{aligned} 0 &= a_1(\lambda_1 - \lambda_{k+1})u_1 + \cdots + a_k(\lambda_k - \lambda_{k+1})u_k + a_{k+1}(\lambda_{k+1} - \lambda_{k+1})u_{k+1} \\ &= a_1(\lambda_1 - \lambda_{k+1})u_1 + \cdots + a_k(\lambda_k - \lambda_{k+1})u_k. \end{aligned} \tag{10.12}$$

Then, by the induction hypothesis, all of the coefficients in equation (10.12) must be zero. Since all of the eigenvalues are distinct, this implies that $a_1 = \cdots = a_k = 0$. Hence, equation (10.10) reduces to

$$0 = a_{k+1}u_{k+1},$$

which is only possible if $a_{k+1} = 0$. Therefore, all of the coefficients $a_1, \ldots, a_{k+1}$ are zero, and the set of eigenvectors is linearly independent, which concludes the proof. □

10.6 Application: google page rank

10.7 Application: waves and schrodinger's equation

Key concepts from Chapter 10

Definition of eigenvalues and eigenvectors, the characteristic polynomial, computing eigenvalues, computing eigenspaces, properties of eigenvalues and eigenspaces, diagonalizing a matrix.

Chapter 11

Orthogonality

Orthogonal projection in $\mathbb{R}^2$. Pictures.

11.1 Orthogonality

Recall that two nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ are called "orthogonal" if $\mathbf{u} \cdot \mathbf{v} = 0$. If the two vectors are non-zero, then this means that the angle formed by two orthogonal vectors is $\pi/2$.

Warning

The vector $\mathbf{0}$ is orthogonal to every vector, since $\mathbf{0} \cdot \mathbf{u} = 0$.

Definition 11.1. A set $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ of k vectors in $\mathbb{R}^n$ is called an *orthogonal set* if whenever $i \neq j$, the vectors $\mathbf{u}_i$ and $\mathbf{u}_j$ are orthogonal.

Definition 11.2. A set $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ of k non-zero vectors in $\mathbb{R}^n$ is an *orthonormal set* if it is an orthogonal set with the additional property that every vector in the set is a unit vector.

In-class exercise

Is $\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -3 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 5 \\ -4 \end{bmatrix} \right\}$ an orthogonal set?

Is $\left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \right\}$ an orthogonal set?

Theorem 11.3. *An orthogonal set of nonzero vectors is linearly independent.*

Proof. Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ be an orthogonal set of nonzero vectors. It needs to be shown that the only solution to the equation

$$\mathbf{0} = x_1\mathbf{u}_1 + \cdots + x_k\mathbf{u}_k$$

is the trivial solution $x_1 = \cdots = x_k = 0$.

For each i, we can take the dot product of this equation with $\mathbf{u}_i$,

$$\begin{aligned}
\mathbf{0} &= x_1\mathbf{u}_1 + \cdots + x_k\mathbf{u}_k \\
\Rightarrow \mathbf{0} \cdot \mathbf{u}_i &= (x_1\mathbf{u}_1 + \cdots + x_k\mathbf{u}_k) \cdot \mathbf{u}_i \\
\Rightarrow 0 &= x_1(\mathbf{u}_1 \cdot \mathbf{u}_i) + \cdots + x_i(\mathbf{u}_i \cdot \mathbf{u}_i) + \cdots + x_k(\mathbf{u}_k \cdot \mathbf{u}_i) \\
\Rightarrow 0 &= x_1(0) + \cdots + x_i(\mathbf{u}_i \cdot \mathbf{u}_i) + \cdots + x_k(0) \text{ since } \mathbf{u}_i \cdot \mathbf{u}_j = 0 \\
\Rightarrow 0 &= x_i(\mathbf{u}_i \cdot \mathbf{u}_i) \\
\Rightarrow 0 &= x_i \text{ since } \mathbf{u}_i \neq \mathbf{0}
\end{aligned}$$

Thus, for every i, $x_i = 0$, so we have shown that the only solution to the equation

$$\mathbf{0} = x_1\mathbf{u}_1 + \cdots + x_k\mathbf{u}_k$$

is the trivial solution; in other words, the set $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is linearly independent.
□

Recall that a set is a basis for a subspace if it spans the subspace and is linearly independent. Thus, the theorem above tells us that any orthogonal set which spans a subspace is automatically a basis for that subspace.

Definition 11.4. An orthogonal set that is also a basis for a subspace S is called an *orthogonal basis of* S. An orthonormal set that is also a basis for a subspace S is called an *orthonormal basis of* S.

Example 11.5. The set of standard unit vectors $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is an orthonormal basis for $\mathbb{R}^n$ since

- The set $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is orthogonal, since $\mathbf{e}_i \cdot \mathbf{e}_j = 0$ whenever $i \neq j$.
- The set $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is a basis for $\mathbb{R}^n$ (we checked this in Chapter 7).
- The vectors are unit vectors: $||\mathbf{e}_i|| = 1$ for every $1 \leq i \leq n$.

Example 11.6. As we saw in the previous example, the standard basis for $\mathbb{R}^n$ is an orthonormal basis. Of course, there are many other orthogonal and orthonormal bases. In this example we describe different orthogonal bases of $\mathbb{R}^2$ and $\mathbb{R}^3$.

(a) As we saw above, the standard basis $\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$ is an orthonormal basis for $\mathbb{R}^2$.

(b) If we rotate the two vectors in the standard basis for $\mathbb{R}^2$ by some angle θ, then the result is a different orthonormal basis. Using the rotation matrix,

$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \end{bmatrix}$$

$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -\sin(\theta) \\ \cos(\theta) \end{bmatrix}$$

The new set of vectors is thus

$$\left\{ \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \end{bmatrix}, \begin{bmatrix} -\sin(\theta) \\ \cos(\theta) \end{bmatrix} \right\}.$$

This set is orthogonal since

$$\begin{bmatrix} \cos(\theta) \\ \sin(\theta) \end{bmatrix} \cdot \begin{bmatrix} -\sin(\theta) \\ \cos(\theta) \end{bmatrix} = -\sin(\theta)\cos(\theta) + \sin(\theta)\cos(\theta) = 0.$$

Thus this new set gives an orthogonal basis for $\mathbb{R}^2$. Furthermore, for each of the two vectors in this set, we see that

$$\left\| \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \end{bmatrix} \right\| = \sqrt{\cos^2(\theta) + \sin^2(\theta)} = 1$$

$$\left\| \begin{bmatrix} -\sin(\theta) \\ \cos(\theta) \end{bmatrix} \right\| = \sqrt{(-\sin(\theta))^2 + \cos^2(\theta)} = 1$$

(c) If we scale the two vectors in the standard basis for $\mathbb{R}^2$ by non-zero numbers then we still have an orthogonal basis. For instance $\left\{ \begin{bmatrix} 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \end{bmatrix} \right\}$ is an orthogonal basis for $\mathbb{R}^2$, but it is not an orthonormal basis.

(d) The basis $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$ is an orthonormal basis for the subspace $S \subseteq \mathbb{R}^2$ defined as

$$S = \left\{ \begin{bmatrix} x \\ y \\ 0 \end{bmatrix} \in \mathbb{R}^3 \colon x, y \in \mathbb{R} \right\}$$

which is a plane in $\mathbb{R}^3$ (sometimes called the xy-plane in $\mathbb{R}^3$).

The linear transformation of $\mathbb{R}^3$ "rotation by an angle of θ in the xy-plane" is described by the matrix

$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

If we apply this transformation to the vectors in the basis for S above, we get a different orthonormal basis for S,

$$\left\{ \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \\ 0 \end{bmatrix}, \begin{bmatrix} -\sin(\theta) \\ \cos(\theta) \\ 0 \end{bmatrix} \right\}$$

There is a straightforward formula for writing a vector $\mathbf{u}$ as a linear combination of the vectors $\mathbf{s}_1, \ldots, \mathbf{s}_k$ in an orthogonal basis.

> Let $\beta = \{\mathbf{s}_1, \ldots, \mathbf{s}_k\}$ be an orthogonal basis for a subspace S. If $\mathbf{u} \in S$, then
>
> $$\begin{aligned} \mathbf{u} &= \left(\frac{\mathbf{u} \cdot \mathbf{s}_1}{\|\mathbf{s}_1\|^2} \right) \mathbf{s}_1 + \left(\frac{\mathbf{u} \cdot \mathbf{s}_2}{\|\mathbf{s}_2\|^2} \right) \mathbf{s}_2 + \cdots + \left(\frac{\mathbf{u} \cdot \mathbf{s}_n}{\|\mathbf{s}_n\|^2} \right) \mathbf{s}_n \\ &= \mathrm{proj}_{\mathbf{s_1}} \mathbf{u} + \mathrm{proj}_{\mathbf{s_2}} \mathbf{u} + \cdots + \mathrm{proj}_{\mathbf{s}_n} \mathbf{u}. \end{aligned}$$

Proof. Since β is a basis, we know the vector x can be written uniquely as

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

for some coefficients $c_1, \ldots, c_n \in \mathbb{R}$. By taking the dot product of $\mathbf{u}$ with the basis vector $\mathbf{s}_i$, the coefficient c_i can be computed, as follows.

$$\begin{aligned} \mathbf{u} \cdot \mathbf{s}_i &= (c_1 \mathbf{s}_1 + c_2 \mathbf{s}_2 + \cdots + c_n \mathbf{s}_n) \cdot \mathbf{s}_i \\ &= c_1(\mathbf{s}_1 \cdot \mathbf{s}_i) + \cdots + c_i(\mathbf{s}_i \cdot \mathbf{s}_i) + \cdots + c_n(\mathbf{s}_n \cdot \mathbf{s}_i) \\ &= c_i(\mathbf{s}_i \cdot \mathbf{s}_i) && \text{(since } \mathbf{s}_j \cdot \mathbf{s}_i = 0 \text{ for all } i \neq j) \\ &= c_i \|\mathbf{s}_i\|^2. \end{aligned}$$

Rearranging yields

$$c_i = \frac{\mathbf{u} \cdot \mathbf{s}_i}{\|\mathbf{s}_i\|^2}.$$

□

In-class exercise

Show that the set

$$\left\{ \mathbf{s}_1 = \begin{bmatrix} 1 \\ -1 \\ 3 \\ 0 \end{bmatrix}, \mathbf{s}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{s}_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \mathbf{s}_4 = \begin{bmatrix} 1 \\ 10 \\ 3 \\ -11 \end{bmatrix} \right\}.$$

is orthogonal by computing dot-products of every pair of vectors in the set.

Normalize the vectors in this basis to turn it into an orthonormal basis for $\mathbb{R}^4$.

Since the set β in the previous exercise is orthogonal, we know that

- Since β is orthogonal, it is linearly independent.
- Since β is a set of 4 linearly independent vectors in $\mathbb{R}^4$, it is a basis for $\mathbb{R}^4$.
- Thus we can conclude that β is an orthogonal basis for $\mathbb{R}^4$.

Problem 21. *Write the vector* $(1, 3, 2, 4)$ *as a linear combination of the vectors in the orthogonal basis* β *from the previous exercise.*

Solution. *The formula above is applied:*

$$\begin{bmatrix} 1 \\ 3 \\ 2 \\ 4 \end{bmatrix} = \frac{4}{11} \begin{bmatrix} 1 \\ -1 \\ 3 \\ 0 \end{bmatrix} + \frac{7}{7} \begin{bmatrix} -2 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \frac{8}{3} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} + \frac{-7}{231} \begin{bmatrix} 1 \\ 10 \\ 3 \\ -11 \end{bmatrix}. \quad \blacklozenge$$

11.2 Orthogonal complements

Definition 11.7. Let S be a subspace of $\mathbb{R}^n$. The *orthogonal complement of* S is the set

$$S^{\perp} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \cdot \mathbf{s} = 0 \text{ for all } \mathbf{s} \in S\}.$$

Example 11.8. The orthogonal complement of a subspace of $\mathbb{R}^2$ or $\mathbb{R}^3$ is easy to visualize and understand geometrically.

(a) **In $\mathbb{R}^2$:**

- The orthogonal complement of the trivial subspace $\{\mathbf{0}\}$ is the set

$$\{\mathbf{0}\}^{\perp} = \{\mathbf{x} \in \mathbb{R}^2 \colon \mathbf{x} \cdot \mathbf{0} = 0\} = \mathbb{R}^2.$$

- The orthogonal complement of a line through the origin spanned by a non-zero vector $\mathbf{u}$, $L = \operatorname{span}\{\mathbf{u}\} = \{t\mathbf{u} \colon t \in \mathbb{R}\}$, is the set

$$L^{\perp} = \{\mathbf{x} \in \mathbb{R}^2 \colon \mathbf{x} \cdot (t\mathbf{u}) = 0 \text{ for all } t \in \mathbb{R}\} = \{\mathbf{x} \in \mathbb{R}^2 \colon \mathbf{x} \cdot \mathbf{u} = 0\}.$$

  This set is simply the line through the origin orthogonal to $\mathbf{u}$.

- The orthogonal complement of the subspace $\mathbb{R}^2$ is the set

$$(\mathbb{R}^2)^{\perp} = \{\mathbf{x} \in \mathbb{R}^2 \colon \mathbf{x} \cdot \mathbf{y} = 0 \text{ for all } \mathbf{y} \in \mathbb{R}^2\} = \{\mathbf{0}\}.$$

(b) **In $\mathbb{R}^3$:**

- Similar to the case in $\mathbb{R}^2$, the orthogonal complement of the trivial subspace $\{\mathbf{0}\}$ is the set

$$\{\mathbf{0}\}^{\perp} = \{\mathbf{x} \in \mathbb{R}^3 \colon \mathbf{x} \cdot \mathbf{0} = 0\} = \mathbb{R}^3.$$

- The orthogonal complement of a line through the origin in $\mathbb{R}^3$ spanned by a non-zero vector $\mathbf{u}$, $L = \operatorname{span}\{\mathbf{u}\} = \{t\mathbf{u} \colon t \in \mathbb{R}\}$, is the set

$$L^{\perp} = \{\mathbf{x} \in \mathbb{R}^3 \colon \mathbf{x} \cdot (t\mathbf{u}) = 0 \text{ for all } t \in \mathbb{R}\} = \{\mathbf{x} \in \mathbb{R}^3 \colon \mathbf{x} \cdot \mathbf{u} = 0\}.$$

  This set is simply the plane through the origin orthogonal to $\mathbf{u}$. This is the plane through the origin in $\mathbb{R}^3$ determined by the normal vector $\mathbf{u}$.

- Suppose that $\mathcal{P} = \operatorname{span}\{\mathbf{u}, \mathbf{v}\}$ is a plane through the origin in $\mathbb{R}^3$ (in particular the two vectors are linearly independent). The orthogonal complement of $\mathcal{P}$ is

$$\begin{aligned}\mathcal{P}^{\perp} &= \big\{\mathbf{x} \in \mathbb{R}^3 \colon \mathbf{x} \cdot (s\mathbf{u} + t\mathbf{v}) = 0 \text{ for all } s, t \in \mathbb{R}\big\} \\ &= \big\{\mathbf{x} \in \mathbb{R}^3 \colon \mathbf{x} \cdot \mathbf{u} = 0 \text{ and } \mathbf{x} \cdot \mathbf{v} = 0\big\}\end{aligned}$$

  There are several ways to express this set. One particularly instructive thing is to observe that

$$\big\{\mathbf{x} \in \mathbb{R}^3 \colon \mathbf{x} \cdot \mathbf{u} = 0 \text{ and } \mathbf{x} \cdot \mathbf{v} = 0\big\} = \operatorname{null}\left(\begin{bmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{bmatrix}\right).$$

  We also know from the tutorials that when $\mathbf{u}$ and $\mathbf{v}$ are not parallel

$$\operatorname{null}\left(\begin{bmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{bmatrix}\right) = \operatorname{span}\{\mathbf{u} \times \mathbf{v}\}.$$

  In other words, the subspace $\mathcal{P}^{\perp}$ equals the line through the origin orthogonal to $\mathcal{P}$.

- Finally, $(\mathbb{R}^3)^\perp = \{\mathbf{0}\}$.

The following fact is useful for computing the orthogonal complement of a subspace.

Theorem 11.9. *Let A be a $m \times n$ matrix. Then*

$$\mathit{row}(A)^\perp = \mathit{null}(A).$$

Proof. To show these two subspaces are equal, we show both inclusions: $\mathrm{row}(A)^\perp \subseteq \mathrm{null}(A)$ and $\mathrm{null}(A) \subseteq \mathrm{row}(A)^\perp$.

The key step in this proof is to remember that, by definition of matrix multiplication, if A is a matrix with rows $\mathbf{u}_1, \ldots, \mathbf{u}_m$, then

$$A\mathbf{x} = \begin{bmatrix} \mathbf{u}_1 \cdot \mathbf{x} \\ \vdots \\ \mathbf{u}_m \cdot \mathbf{x} \end{bmatrix}.$$

- First, we show that $\mathrm{row}(A)^\perp \subseteq \mathrm{null}(A)$.

  If $\mathbf{x} \in \mathrm{row}(A)^\perp$, then $\mathbf{u}_1 \cdot \mathbf{x} = 0$ for all $1 \leq i \leq m$, so

  $$A\mathbf{x} = \begin{bmatrix} \mathbf{u}_1 \cdot \mathbf{x} \\ \vdots \\ \mathbf{u}_m \cdot \mathbf{x} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix},$$

  so $\mathbf{x} \in \mathrm{null}(A)$.

- Second, we show that $\mathrm{null}(A) \subseteq \mathrm{row}(A)^\perp$.

  If $\mathbf{x} \in \mathrm{null}(A)$, then we know that $A\mathbf{x} = \mathbf{0}$. Thus we know that

  $$\begin{bmatrix} \mathbf{u}_1 \cdot \mathbf{x} \\ \vdots \\ \mathbf{u}_m \cdot \mathbf{x} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}.$$

  This tells us that $\mathbf{x}$ is orthogonal to the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_m$.

Since $\text{row}(A) = \text{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$, we know that for any $\mathbf{y} = t_1\mathbf{u}_1 + \cdots + t_m\mathbf{u}_m \in \text{row}(A)$,

$$\begin{aligned}\mathbf{x} \cdot \mathbf{y} &= \mathbf{x} \cdot (t_1\mathbf{u}_1 + \cdots + t_m\mathbf{u}_m)\\ &= t_1\mathbf{x} \cdot \mathbf{u}_1 + \cdots + t_m\mathbf{x} \cdot \mathbf{u}_m\\ &= 0\end{aligned}$$

Thus $\mathbf{x} \in \text{row}(A)^\perp$.

□

Any subspace $S \subseteq \mathbb{R}^n$ can be expressed as the row space of a matrix (just take the vectors in a spanning set and write them as the rows of a matrix). Since the null space of a matrix is a subspace of $\mathbb{R}^n$, Theorem 11.9 immediately tells us that

> If S is a subspace of $\mathbb{R}^n$, then the orthogonal complement, $S^\perp$, is also a subspace of $\mathbb{R}^n$.[a]
>
> [a]In fact, one can similarly define the orthogonal complement of a set and show that the orthogonal complement of any set is a subspace.

By the rank-nullity theorem, we know that for a $m \times n$ matrix A,

$$\dim(\text{row}(A)) + \dim(\text{null}(A)) = n.$$

Combining this with Theorem 11.9, we see that

$$\dim(\text{row}(A)) + \dim(\text{row}(A)^\perp) = \dim(\text{row}(A)) + \dim(\text{null}(A)) = n.$$

This gives us the following result.

> If S is a subspace of $\mathbb{R}^n$, then
>
> $$\dim(S) + \dim(S^\perp) = n.$$

Theorem 11.9 also gives us an algorithm for computing a basis for the orthogonal complement of a subspace.

Algorithm for computing the orthogonal complement of a subspace:

If we are given a subspace $S \subseteq \mathbb{R}^n$ as the span of a set of vectors $S = \text{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$, then we can compute a basis for $S^\perp$ as follows:

1. Make the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_m$ the rows of a $m \times n$ matrix A.
2. Compute a basis for $\text{null}(A) = S^\perp$ by computing the basic solutions.

Problem 22. *Let*

$$S = \text{span}\left\{ b_1 = \begin{bmatrix} 1 \\ -2 \\ -2 \\ 1 \\ 0 \end{bmatrix}, b_2 = \begin{bmatrix} -2 \\ 4 \\ 3 \\ -1 \\ 2 \end{bmatrix} \right\}.$$

Find a basis for $S^\perp$.

Solution. *By Theorem 11.9,*

$$S^\perp = null \begin{bmatrix} 1 & -2 & -2 & 1 & 0 \\ -2 & 4 & 3 & -1 & 2 \end{bmatrix}.$$

Computing basic solutions of this matrix, we see that

$$\left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix} \right\}$$

is a basis for this null space, which makes it a basis for $S^\perp$. ♦

We end this section by observing that since for any matrix A,

$$\text{row}(A^T) = \text{col}(A),$$

Theorem 11.9 immediately implies the following fact.

If A is a $m \times n$ matrix, then

$$\mathrm{col}(A)^{\perp} = \mathrm{null}(A^T).$$

(note that these are both subspaces of $\mathbb{R}^m$)

Proof. Applying Theorem 11.9 to the matrix A^T we have that

$$\mathrm{col}(A)^{\perp} = \mathrm{row}(A^T)^{\perp} = \mathrm{null}(A^T).$$

□

11.3 Orthogonal projection onto a subspace

Recall that a nonzero vector $\mathbf{u} \in \mathbb{R}^n$ spans a line $L = \mathrm{span}\{\mathbf{u}\}$. In Chapter 5 we studied the following problem:

The problem of orthogonal projection onto a line:

Given a vector $\mathbf{v} \in \mathbb{R}^n$, find vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ so that

- $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$
- $\mathbf{v}_1$ is parallel to $\mathbf{u}$ (contained in the line L) and $\mathbf{v}_2$ is orthogonal to $\mathbf{u}$ (contained in the hyperplane normal to $\mathbf{u}$).

In Chapter 5, we discovered that we can always find such vectors, and they are given by formulas:

$$\begin{aligned} \mathbf{v}_1 &= \mathrm{proj}_{\mathbf{u}}\mathbf{v} = \frac{\mathbf{v} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u} \\ \mathbf{v}_2 &= \mathbf{v} - \mathrm{proj}_{\mathbf{u}}\mathbf{v} = \mathbf{v} - \frac{\mathbf{v} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u}. \end{aligned}$$

Given a subspace S of $\mathbb{R}^n$, we would like to solve a similar problem.

The problem of orthogonal projection onto a subspace:

Given a vector $\mathbf{v} \in \mathbb{R}^n$, find vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ so that

- $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$
- $\mathbf{v}_1$ is contained in S and $\mathbf{v}_2$ is orthogonal to S (contained in $S^\perp$).

The solution to this problem is similar to the solution to the problem of orthogonal projection onto a line.

- First, suppose that $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is an orthogonal basis for S. Since it is a basis, we know that $\mathbf{v}_1 \in S$ can be written as

$$\mathbf{v}_1 = x_1\mathbf{u}_1 + \cdots + x_k\mathbf{u}_k$$

  for some coefficients $x_1, \dots, x_k$.

- We can solve for the coefficients $x_1, \dots, x_k$ one by one. For each $\mathbf{u}_i$, we see that

$$\begin{aligned}
\mathbf{u}_i \cdot \mathbf{v} &= \mathbf{u}_i \cdot (\mathbf{v}_1 + \mathbf{v}_2) \\
&= \mathbf{u}_i \cdot \mathbf{v}_1 + \mathbf{u}_i \cdot \mathbf{v}_2 \\
&= \mathbf{u}_i \cdot (x_1\mathbf{u}_1 + \cdots + x_k\mathbf{u}_k) \\
&= x_1(\mathbf{u}_i \cdot \mathbf{u}_1) + \cdots + x_k(\mathbf{u}_i \cdot \mathbf{u}_k) \\
&= x_i(\mathbf{u}_i \cdot \mathbf{u}_i)
\end{aligned}$$

  Thus we see that

$$x_i = \frac{\mathbf{u}_i \cdot \mathbf{v}}{\mathbf{u}_i \cdot \mathbf{u}_i}$$

  so

$$\mathbf{v}_1 = \frac{\mathbf{u}_1 \cdot \mathbf{v}}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \cdots + \frac{\mathbf{u}_k \cdot \mathbf{v}}{\mathbf{u}_k \cdot \mathbf{u}_k}\mathbf{u}_k = \mathrm{proj}_{\mathbf{u}_1}\mathbf{v} + \cdots + \mathrm{proj}_{\mathbf{u}_k}\mathbf{v}$$

Definition 11.10. Let S be a non-zero subspace. Let $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ be an orthogonal basis for S. The *orthogonal projection of* $\mathbf{v}$ *onto* S is denoted $\mathrm{proj}_S\mathbf{v}$ and given by the formula

$$\mathrm{proj}_S\mathbf{v} = \frac{\mathbf{u}_1 \cdot \mathbf{v}}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \cdots + \frac{\mathbf{u}_k \cdot \mathbf{v}}{\mathbf{u}_k \cdot \mathbf{u}_k}\mathbf{u}_k = \mathrm{proj}_{\mathbf{u}_1}\mathbf{v} + \cdots + \mathrm{proj}_{\mathbf{u}_k}\mathbf{v}.$$

Example 11.11. Consider the subspace $S = \{(x, y, 0) \in \mathbb{R}^3 \colon x, y \in \mathbb{R}\}$. The orthogonal projection of $\mathbf{v} = [8\,6\,-9]$ onto S can be computed using the basis $\{\mathbf{e}_1, \mathbf{e}_2\}$ for S:

$$\text{proj}_S\mathbf{v} = \text{proj}_{\mathbf{e}_1}\mathbf{v} + \text{proj}_{\mathbf{e}_2}\mathbf{v} = \begin{bmatrix} 8 \\ 6 \\ 0 \end{bmatrix} \in S$$

- The vector $\mathbf{v}_2$ is then found by rearranging the equation

$$\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$$

  to solve for $\mathbf{v}_2$,

$$\mathbf{v}_2 = \mathbf{v} - \mathbf{v}_1 = \mathbf{v} - \text{proj}_{\mathbf{u}_1}\mathbf{v} - \cdots - \text{proj}_{\mathbf{u}_k}\mathbf{v}$$

- The vector $\mathbf{v}_2$ is orthogonal to S, since for every $\mathbf{u}_i$ in the orthogonal basis for S above,

$$\begin{aligned} \mathbf{u}_i \cdot \mathbf{v}_2 &= \mathbf{u}_i \cdot (\mathbf{v} - \text{proj}_{\mathbf{u}_1}\mathbf{v} - \cdots - \text{proj}_{\mathbf{u}_k}\mathbf{v}) \\ &= \mathbf{u}_i \cdot \mathbf{v} - \mathbf{u}_i \cdot \text{proj}_{\mathbf{u}_1}\mathbf{v} - \cdots - \mathbf{u}_i \cdot \text{proj}_{\mathbf{u}_k}\mathbf{v} \\ &= \mathbf{u}_i \cdot \mathbf{v} - \mathbf{u}_i \cdot \text{proj}_{\mathbf{u}_i}\mathbf{v} \\ &= \mathbf{u}_i \cdot \mathbf{v} - \mathbf{u}_i \cdot \left(\frac{\mathbf{u}_i \cdot \mathbf{v}_i}{\mathbf{u}_i \cdot \mathbf{u}_i}\mathbf{u}_i \right) = 0 \end{aligned}$$

  Thus $\mathbf{v}_2 \in S^\perp$.

We call the vector $\mathbf{v}_2$ the *projection of* $\mathbf{v}$ *orthogonal to* S. It is sometimes denoted $\text{proj}_{S^\perp}\mathbf{v}$.

Example 11.12. Continuing from the previous example, the subspace $S^\perp = \{(0, 0, z) \in \mathbb{R}^3 \colon z \in \mathbb{R}\}$. The projection of $\mathbf{v} = [8\,6\,-9]$ orthogonal to S is

$$\text{proj}_{S^\perp}\mathbf{v} = \mathbf{v} - \text{proj}_S\mathbf{v} = \begin{bmatrix} 0 \\ 0 \\ -9 \end{bmatrix} \in S$$

Similar to orthogonal projection onto lines, orthogonal projection onto subspaces has several properties:

Properties of orthogonal projection:

Suppose S is a subspace of $\mathbb{R}^n$. Then

- $T(\mathbf{x}) = \text{proj}_S(\mathbf{x})$ is a linear transformation with
  - $\ker(T) = S^{\perp}$
  - $\text{im}(T) = S$.
- For any vector $\mathbf{v} \in \mathbb{R}^n$, $\text{proj}_S(\mathbf{v}) \in S$.
- If $\mathbf{v} \in S$, then $\text{proj}_S(\mathbf{v}) = \mathbf{v}$.
- If $\mathbf{v} \in S^{\perp}$, then $\text{proj}_S(\mathbf{v}) = \mathbf{0}$.

Problem 23. *The set*

$$\left\{ \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix} \right\}$$

is orthogonal. Compute the projection of $\begin{bmatrix} 7 \\ 7 \\ 4 \end{bmatrix}$ *onto* $S = span \left\{ \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix} \right\}$.

Solution. *Apply the definition of projection onto a subspace (Equation 11.10):*

$$proj_s \begin{bmatrix} 7 \\ 7 \\ 4 \end{bmatrix} = \frac{4}{9} \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix} + \frac{13}{9} \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}. \quad \blacklozenge$$

11.4 The Gram-Schmidt algorithm

In this section, we describe an algorithm for solving the following problem.

> Given a basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ for a subspace $S \subseteq \mathbb{R}^n$, find an orthogonal basis for S.

Let's try to do this for a simple example.

Problem 24. *Find an orthogonal basis for the subspace* $S = span \left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$.

The Gram-schmidt algorithm

Suppose that $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is a basis for a subspace $S \subseteq \mathbb{R}^n$. In order to compute an orthogonal basis for S, compute

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{u}_1\\
\mathbf{v}_2 &= \mathbf{u}_2 - \mathrm{proj}_{\mathbf{v}_1}\mathbf{u}_2\\
\mathbf{v}_3 &= \mathbf{u}_3 - \mathrm{proj}_{\mathbf{v}_1}\mathbf{u}_3 - \mathrm{proj}_{\mathbf{v}_2}\mathbf{u}_3\\
&\vdots\\
\mathbf{v}_k &= \mathbf{u}_k - \mathrm{proj}_{\mathbf{v}_1}\mathbf{u}_k - \mathrm{proj}_{\mathbf{v}_2}\mathbf{u}_k - \cdots - \mathrm{proj}_{\mathbf{v}_{k-1}}\mathbf{u}_k.
\end{aligned}$$

By construction, the set $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$, is an orthogonal basis for S.

Further, the set

$$\left\{\frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}, \dots, \frac{\mathbf{v}_k}{\|\mathbf{v}_k\|}\right\}.$$

is an orthonormal basis for S.

Problem 25. *The set*

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

is a basis for its span, which is a subspace of $\mathbb{R}^4$. *Transform this basis into an orthogonal basis.*

Solution. *Apply the Gram-Schmidt algoritheorem: The first orthogonal basis vector is*

$$v_1 = \begin{bmatrix}1\\1\\1\\0\end{bmatrix}.$$

The second is

$$
\begin{aligned}
v_2 &= \begin{bmatrix} -1 \\ 0 \\ -1 \\ 1 \end{bmatrix} - proj_{\left(\begin{smallmatrix} 1 \\ 1 \\ 1 \\ 0 \end{smallmatrix}\right)} \begin{bmatrix} -1 \\ 0 \\ -1 \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} -1 \\ 0 \\ -1 \\ 1 \end{bmatrix} - \frac{\left(\begin{smallmatrix} -1 \\ 0 \\ -1 \\ 1 \end{smallmatrix}\right) \cdot \left(\begin{smallmatrix} 1 \\ 1 \\ 1 \\ 0 \end{smallmatrix}\right)}{\left\| \left(\begin{smallmatrix} 1 \\ 1 \\ 1 \\ 0 \end{smallmatrix}\right) \right\|^2} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} \\
&= \frac{1}{3} \begin{bmatrix} -1 \\ 2 \\ -1 \\ 3 \end{bmatrix}.
\end{aligned}
$$

And the third is

$$
\begin{aligned}
v_3 &= \begin{bmatrix} -1 \\ 0 \\ 0 \\ -1 \end{bmatrix} - proj_{\left(\begin{smallmatrix} 1 \\ 1 \\ 1 \\ 0 \end{smallmatrix}\right)} \begin{bmatrix} -1 \\ 0 \\ 0 \\ -1 \end{bmatrix} - proj_{\left(\begin{smallmatrix} -1 \\ 2 \\ -1 \\ 3 \end{smallmatrix}\right)} \begin{bmatrix} -1 \\ 0 \\ 0 \\ -1 \end{bmatrix} \\
&= \begin{bmatrix} -1 \\ 0 \\ 0 \\ -1 \end{bmatrix} - \frac{\left(\begin{smallmatrix} -1 \\ 0 \\ 0 \\ -1 \end{smallmatrix}\right) \cdot \left(\begin{smallmatrix} 1 \\ 1 \\ 1 \\ 0 \end{smallmatrix}\right)}{\left\| \left(\begin{smallmatrix} 1 \\ 1 \\ 1 \\ 0 \end{smallmatrix}\right) \right\|^2} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} - \frac{\left(\begin{smallmatrix} -1 \\ 0 \\ 0 \\ -1 \end{smallmatrix}\right) \cdot \left(\begin{smallmatrix} -1 \\ 2 \\ -1 \\ 3 \end{smallmatrix}\right)}{\left\| \left(\begin{smallmatrix} -1 \\ 2 \\ -1 \\ 3 \end{smallmatrix}\right) \right\|^2} \begin{bmatrix} -1 \\ 2 \\ -1 \\ 3 \end{bmatrix} \\
&= \frac{1}{5} \begin{bmatrix} -4 \\ 3 \\ 1 \\ -3 \end{bmatrix}.
\end{aligned}
$$

Therefore, the set

$$
\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ -1 \\ 3 \end{bmatrix}, \begin{bmatrix} -4 \\ 3 \\ 1 \\ -3 \end{bmatrix} \right\}
$$

is an orthogonal basis for the subspace given in the problem. ♦

11.5 Application: QR factorization and linear regression

Key concepts from Chapter 11

Definition of orthogonal basis, definition of orthonormal basis, examples of orthogonal and orthonormal bases, orthogonal complements, Theorem 11.9, properties of orthogonal complements, orthogonal projection onto a subspace, properties of orthogonal projection onto a subspace, the Gram-Schmidt algorithm.

Appendix A

Extra examples for Chapter 1

In this appendix we provide several additional worked computational examples.

Problem 26 (system with a unique solution). *Solve the system*

$$\begin{aligned} x + y &= 7 \\ x - y &= 1. \end{aligned}$$

Solution. *Adding the two equations yields*

$$2x = 8 \qquad \Longrightarrow \qquad x = 4.$$

Substituting this into either original equation yields $y = 3$. Therefore, the unique solution is $(x, y) = (4, 3)$. ♦

Problem 27 (System with infinitely many solutions). *Solve the system*

$$\begin{aligned} &x + y = 7 && (1) \\ &-3x - 3y = -21 && (2). \end{aligned}$$

Solution. *Since equation (2) is -3 times equation (1), if (x, y) solves (1), then it also solves (2), and vice-versa. So, the solutions are*

$$x = t,\ y = 7 - t \qquad \Longrightarrow \qquad (x, y) = (t, 7 - t),\ t \in \mathbb{R}.$$

The set of solutions can be written using set-builder notation as follows.

$$\left\{(t, 7 - t) \in \mathbb{R}^2 \colon t \in \mathbb{R}\right\}. \quad ♦$$

Problem 28 (System with no solutions). *Solve the system*

$$x + y = 4 \qquad (1)$$
$$2x + 2y = -4 \qquad (2)$$

Solution. *Subtracting* 2 *times equation (1) from equation (2) yields*

$$0x + 0y = -12,$$

which has no solutions. Hence, this system is inconsistent. ♦

Problem 29. *Solve the system*

$$\begin{aligned} 4x_1 \qquad + x_3 \qquad\qquad &= 4 \\ x_2 + 3x_3 \qquad - x_5 &= 3 \\ x_4 + 2x_5 &= 5. \end{aligned}$$

Solution. *The augmented matrix of this system is already in row-echelon form, so we don't need to perform Gaussian elimination.*

The leading variables are x_1, x_2, and x_4. The free variables are x_3 and x_5. Thus we,

1. *Set free variables equal to parameters: $x_3 = s$ and $x_5 = t$.*

2. *Back substitute to solve for leading variables:*

$$\begin{aligned} x_4 + 2t = 5 \quad &\Longrightarrow \quad x_4 = 5 - 2t \\ x_2 + 3s - t = 3 \quad &\Longrightarrow \quad x_2 = 3 + t - 3s \\ 4x_1 + s = 4 \quad &\Longrightarrow \quad x_1 = 1 - \frac{s}{4}. \end{aligned}$$

Therefore, the general form of a solution is

$$(1 - \tfrac{s}{4}, 3 + t - 3s, s, 5 - 2t, t)$$

where t and s are parameters. The solution set is

$$\left\{ \left(1 - \frac{s}{4}, 3 + t - 3s, s, 5 - 2t, t\right) \in \mathbb{R}^5 : t, s \in \mathbb{R} \right\}. \quad ♦$$

Problem 30. *Solve the system of linear equations*

$$\begin{aligned} 2x_1 + x_2 - x_3 &= 1 \\ x_1 + x_2 + x_3 &= 6 \\ -3x_1 + 2x_2 + 4x_3 &= 13. \end{aligned}$$

Solution. *The corresponding augmented matrix is simplified with elementary row operations:*

$$\left[\begin{array}{ccc|c} 2 & 1 & -1 & 1 \\ 1 & 1 & 1 & 6 \\ -3 & 2 & 4 & 13 \end{array}\right] \xrightarrow{R_1 \leftrightarrow R_2} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 6 \\ 2 & 1 & -1 & 1 \\ -3 & 2 & 4 & 13 \end{array}\right] \xrightarrow[R_3+3R_1]{R_2-2R_1} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 6 \\ 0 & -1 & -3 & -11 \\ 0 & 5 & 7 & 31 \end{array}\right]$$

$$\xrightarrow{R_3+5R_2} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 6 \\ 0 & -1 & -3 & -11 \\ 0 & 0 & -8 & -24 \end{array}\right].$$

This matrix is in row-echelon form. From this augmented matrix we obtain the equations:

$$\begin{aligned} x_1 + x_2 + x_3 &= 6 \\ -x_2 - 3x_3 &= -11 \\ -8x_3 &= -24. \end{aligned}$$

Every variable is a leading variable, so there are no parameters. We solve these equations by back-substitution. From the third equation, we obtain:

$$x_3 = 3$$

then we substitute this into the second equation to obtain:

$$-x_2 = -11 + 3(3) = -2 \implies x_2 = 2.$$

Finally, substituting these equations into the first equation we obtain:

$$x_1 = 6 - (2) - (1) = 1.$$

Thus there is exactly one solution,

$$x_1 = 1, \quad x_2 = 2, \quad x_3 = 3.$$

The set of solutions is $\{(1, 2, 3)\}$. ♦

Problem 31. *Solve the system*

$$\begin{aligned} x + y + z &= 6 \\ 2x + y - z &= 1 \\ x + 2y + 4z &= 15. \end{aligned}$$

Solution. *This system has corresponding augmented matrix*

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 6 \\ 2 & 1 & -1 & 1 \\ 1 & 2 & 4 & 15 \end{array}\right].$$

The $(1,1)$ *entry will be used to eliminate the remaining terms in the first column:*

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 6 \\ 2 & 1 & -1 & 1 \\ 1 & 2 & 4 & 15 \end{array}\right] \underset{R_3-R_1}{\overset{R_2-2R_1}{\longrightarrow}} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 6 \\ 0 & -1 & -3 & -11 \\ 0 & 1 & 3 & 9 \end{array}\right].$$

Then, using the $(2,2)$ *entry to reduce the matrix to echelon form yields*

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 6 \\ 0 & -1 & -3 & -11 \\ 0 & 1 & 3 & 9 \end{array}\right] \overset{R_3+R_2}{\longrightarrow} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 6 \\ 0 & -1 & -3 & -11 \\ 0 & 0 & 0 & -2 \end{array}\right]$$

The last row corresponds to the equation $0x+0y+0z=-2$*, which has no solutions! Thus the system is inconsistent and has no solutions.* ♦

Problem 32. *Find the reduced row-echelon form of the matrix*

$$\begin{bmatrix} 1 & 1 & 1 & 6 \\ 0 & -1 & -3 & -11 \\ 0 & 0 & 0 & -2 \end{bmatrix}.$$

Solution. *The solution is given by row reducing. Since the matrix is already in echelon form, all that is left is to use row operations to change the pivots to both be 1 and to clear out the remaining terms in the pivot columns. The* $(2,1)$ *entry is used to clear out its column and then is changed to be 1:*

$$\begin{bmatrix} 1 & 1 & 1 & 6 \\ 0 & -1 & -3 & -11 \\ 0 & 0 & 0 & -2 \end{bmatrix} \overset{R_1+R_2}{\longrightarrow} \begin{bmatrix} 1 & 0 & -2 & -5 \\ 0 & -1 & -3 & -11 \\ 0 & 0 & 0 & -2 \end{bmatrix} \overset{-R_2}{\longrightarrow} \begin{bmatrix} 1 & 0 & -2 & -5 \\ 0 & 1 & 3 & 11 \\ 0 & 0 & 0 & -2 \end{bmatrix}.$$

The $(3,4)$ *entry is used to clear its column:*

$$\begin{bmatrix} 1 & 0 & -2 & -5 \\ 0 & 1 & 3 & 11 \\ 0 & 0 & 0 & -2 \end{bmatrix} \overset{-\frac{1}{2}R_3}{\longrightarrow} \begin{bmatrix} 1 & 0 & -2 & -5 \\ 0 & 1 & 3 & 11 \\ 0 & 0 & 0 & 1 \end{bmatrix} \underset{R_2-11R_3}{\overset{R_1+5R_3}{\longrightarrow}} \begin{bmatrix} 1 & 0 & -2 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Note that the pivot columns are all zero except for the pivots. ♦

Problem 33. *For what values of (a,b,c,d) is the following system consistent?*

$$\begin{aligned} x_1 \qquad\quad - 2x_3 \qquad\quad &= a \\ x_1 + 4x_2 - 4x_3 - x_4 &= b \\ x_1 + 2x_2 \qquad\quad - 2x_4 &= c \\ x_2 - x_3 \qquad\quad &= d. \end{aligned}$$

Solution. *This system has a solution if and only if its echelon form is consistent. Row reducing the corresponding augmented matrix yields the following. The pivots being used to clear a column are boxed.*

$$\left[\begin{array}{cccc|c} \boxed{1} & 0 & -2 & 0 & a \\ 1 & 4 & -4 & -1 & b \\ 1 & 2 & 0 & -2 & c \\ 0 & 1 & -1 & 0 & d \end{array}\right] \xrightarrow[R_3-R_1]{R_2-R_1} \left[\begin{array}{cccc|c} 1 & 0 & -2 & 0 & a \\ 0 & 4 & -2 & -1 & b-a \\ 0 & 2 & 2 & -2 & c-a \\ 0 & 1 & -1 & 0 & d \end{array}\right]$$

$$\xrightarrow[R_2\leftrightarrow R_4]{\frac{1}{2}R_3} \left[\begin{array}{cccc|c} 1 & 0 & -2 & 0 & a \\ 0 & \boxed{1} & -1 & 0 & d \\ 0 & 1 & 1 & -1 & \frac{1}{2}(c-a) \\ 0 & 4 & -2 & -1 & b-a \end{array}\right] \xrightarrow[R_4-4R_2]{R_3-R_2} \left[\begin{array}{cccc|c} 1 & 0 & -2 & 0 & a \\ 0 & 1 & -1 & 0 & d \\ 0 & 0 & 2 & -1 & \frac{1}{2}(c-a)-d \\ 0 & 0 & 2 & -1 & b-a-4d \end{array}\right]$$

Observing that rows 3 and 4 are identical to the left of the augmentation, subtracting row 3 from 4 gives

$$\left[\begin{array}{cccc|c} 1 & 0 & -2 & 0 & a \\ 0 & 1 & -1 & 0 & d \\ 0 & 0 & 2 & -1 & \frac{1}{2}(c-a)-d \\ 0 & 0 & 2 & -1 & b-a-4d \end{array}\right] \xrightarrow{R_4-R_3} \left[\begin{array}{cccc|c} 1 & 0 & -2 & 0 & a \\ 0 & 1 & -1 & 0 & d \\ 0 & 0 & 2 & -1 & \frac{1}{2}(c-a)-d \\ 0 & 0 & 0 & 0 & b-a-3d-\frac{1}{2}(c-a) \end{array}\right]$$

Therefore, the system is consistent if and only if $b-a-3d-\frac{1}{2}(c-a)=0$.

The equation $b-a-3d-\frac{1}{2}(c-a)=0$ *is itself a system of* one *linear equation with 4 variables,* a,b,c,d. *The solution set can be described with three parameters* $b=q, c=r, d=s$ *and can be written with set-builder notation as*

$$\left\{(2q-r-6s, q, r, s) \in \mathbb{R}^4 \colon q,r,s \in \mathbb{R}\right\}. \quad \blacklozenge$$

Appendix B

Tutorials

B.1 Tutorial 1: solving systems of linear equations

In this tutorial we will practice using Gaussian elimination to solve a system of linear equations and describing the set of solutions. We will also check our understanding of some terminology and finish the proof of Theorem 1.13 in the lecture notes.

Background: Chapter 1. If you have forgotten the definition of a word in this worksheet, try looking it up in the index.

1. Consider the following system of linear equations in the variables x_1, x_2, x_3, x_4, x_5.

$$\begin{aligned} 3x_1 + 2x_2 + x_5 &= 4 + x_3 \\ x_2 &= 4 + x_4 \\ +x_3 &= 5 + x_1 + x_2 + 3x_4 \end{aligned}$$

(a) Write the augmented matrix for this system and use Gaussian elimination to put the matrix in row-echelon form.

(b) What is the reduced row-echelon form of the coefficient matrix?

(c) What is the rank of the coefficient matrix for this system?

(d) Use the result from part (a) to describe the set of solutions to this system of linear equations.

(e) Is the system consistent or inconsistent?

(f) How many solutions does this system of linear equations have? Relate your answer to part (c).

2. Give examples of systems of linear equations in three variables (call them x, y, and z) with each of the following properties, or explain why no example exists. Justify your answers.

(a) A system with exactly one solution, $(1, 1, -1)$.

(b) A system that is inconsistent.

(c) A system with infinitely many solutions, such that the solutions satisfy the equation $x + y = z$.

(d) A system with exactly four solutions.

(e) A system whose augmented matrix has rank 2.

(f) A system whose augmented matrix has rank 0.

(g) A system whose augmented matrix has rank 4.

(h) A system whose augmented matrix has rank 5.

3. Suppose that a, b, c, d, e, and f are numbers and

$$\begin{aligned} ax + by &= e \\ cx + dy &= f \end{aligned}$$

is a system of linear equations in the variables x and y.

(a) Prove that if (x_0, y_0) and (x_1, y_1) are solutions to this system, then for every real number $t \in \mathbb{R}$, the pair

$$(tx_1 + (1-t)x_0,\ ty_1 + (1-t)y_0)$$

is a solution.

(b) Use part (a) to explain why a 2×2 system of linear equations can only have 0 solutions, 1 solution or infinitely many solutions.

(c) Your answer to part (b) is a proof of Theorem 1.13 for the special case of two equations in two variables. Can you adapt this proof to the general case (m equations in n variables)?

4. The second elementary operation on systems of linear equations is "multiply an equation by a non-zero number." Why is it not "multiply an equation by a number"?

5. Suppose that the rank of the augmented matrix $[A|\mathbf{b}]$ of an inconsistent system of linear equations is r. What is the rank of the coefficient matrix A?

   What if the system of linear equations is consistent?

6. (food for thought) In class we learned that the important feature of elementary operations on a system of linear equations is Theorem 1.18: they do not change the set of solutions of the system.

   Of course, we can define many other 'operations' on systems of linear equations that are not elementary operations. For each of the following operations on systems of equations, discuss whether it will change the set of solutions. You may want to try the operations on some simple examples.

   For simplicity, suppose we are only discussing systems of linear equations in three variables, x_1, x_2, x_3.

   (a) Add a new equation to the system of linear equations.

   (b) Delete an equation from the system of linear equations.

   (c) Multiply the coefficient of x_1 in each equation by 5.

   (d) In each equation, switch the coefficients of x_1 and x_3. For example,

   $$x_1 - 2x_2 + 5x_3 = 0 \to 5x_1 - 2x_2 + x_3 = 0.$$

   (e) Replace the constant terms in each equation with 0.

   (f) In each equation, add the coefficient of x_1 to the coefficient of x_2. For example,

   $$x_1 - 2x_2 + x_3 = 0 \to x_1 - x_2 + x_3 = 0.$$

   (g) In each equation, set $x_3 = 1$.

B.2 Tutorial 2: matrix algebra

In this tutorial we will explore matrix algebra, with a focus on the comparison with highschool algebra.

Background: Highschool algebra and material from Chapter 2.

1. Suppose that A is a 2×3 matrix, B is a 3×2 matrix, u is a 3×1 column vector, and w is a 1×3 row vector. Which the following expressions are equal? Justify your answers using properties of matrix algebra.

(i) $uw(B - A^T)$

(ii) $w^T u^T \left(B - A^T\right)$

(iii) $-\left(Aw^T u^T - B^T w^T u^T\right)^T$

(iv) $(B^T uw)^T - (uw)^T A^T$

(v) BB^T

(vi) $B^T B$

2. Suppose A is the 2×2 matrix

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

Find all 2×2 matrices B that solve the matrix equation

$$AB = BA.$$

Use set-builder notation to describe your final answer.

Hint: Convert the matrix equation into a system of linear equations. Solve the system of equations. Convert your answer back into a matrix.

3. Consider the following list of algebraic identities and facts that are true when x, y, and z are real numbers.

(a) $x + y = y + x$

(b) $xy = yx$ (hint: compare this equation with question 2)

(c) If $xy = xz$ and $x \neq 0$, then $y = z$.

(d) $x^2 + 2xy + y^2 = (x + y)^2$

(e) $x^2y^2 = xyxy$

(f) $x^2 \geq 0$

(g) If $x^2 = 1$, then $x = \pm 1$. (compare with your answer to question 4).

(h) If $xy = 0$, then $x = 0$ or $y = 0$.

Which of these algebraic identities/facts are true if the numbers x, y, and z are replaced with matrices A, B, and C? Be careful to specify what the number of rows and columns of the matrices must be in order for the equations to make any sense.

For example: for part (a), the corresponding equation with matrices is $A + B = B + A$. For this equation to make sense, A and B must both have the same number of rows and columns. Is this equation always true for any $m \times n$ matrices? Can you cite a fact from class?

If the identity for matrices is true, explain why. If it is not true, give an example that demonstrates it is false.

4. We know from highschool algebra that the roots of a quadratic polynomial $p(x) = ax^2 + bx + c$ are given by the formula

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Using matrix multiplication, addition, and scalar multiplication, we can write an analogous matrix equation for $n \times n$ matrices X, and scalars a, b, and c,

$$aX^2 + bX + cI_n = 0.$$

Here I_n is the $n \times n$ identity matrix and 0 is the $n \times n$ zero matrix. Although these matrix equations look like quadratic polynomials, the set of solutions is very different.

(a) Can you find some or all solutions to the matrix equation $X^2 - I = 0$ when X is a 2×2 matrix? How does your answer compare to the solutions to the equation $x^2 - 1 = 0$? Describe your answer with set-builder notation.

(b) Is it possible to use Gaussian elimination to find the set of all solutions to the matrix equation in part (a)?

We know from highschool that the quadratic polynomial $p(x) = ax^2 + bx + c$ only has real roots if the discriminant $\Delta := b^2 - 4ac$ is a non-negative number ($\Delta \geq 0$). For example, the descriminant of the polynomial

$$q(x) = x^2 + 1$$

is $\Delta = -4$ and $q(x)$ has no real roots.

Suppose X is a 2×2 matrix. Unlike $q(x) = 0$, the matrix equation

$$X^2 + I = 0$$

does have solutions!

(c) Find all solutions of the matrix equation $X^2 + I = 0$. How does your answer compare to the set of solutions of the equation $x^2 + 1 = 0$? Use set-builder notation to describe your answer.

5. A square matrix $A = [a_{ij}]_{n \times n}$ is called *upper triangular* if

$$a_{ij} = 0 \text{ when } i > j.$$

(i) Write down the general form of a 2×2 and 3×3 upper triangular matrix.

(ii) Show that the product of two upper-triangular matrices is upper-triangular. (Hint: try the cases $n = 2, 3$ first, then see if you can make your argument general).

6. A matrix A is called *purple*[1] if $A = A^T$. For each of the following statements, say whether it is true or false. Justify your answers with an explanation.

(i) The matrix

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

is purple.

[1]The real word for this property is *symmetric*.

(ii) The matrix

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

is purple.

(iii) The matrix

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

is purple.

(iv) If A is a purple matrix, then A is a square matrix.

(v) If A and B are purple $n \times n$ matrices, then $A + B$ is purple.

(vi) If A and B are purple $n \times n$ matrices, then $3A + 8B^T$ is purple.

(vii) If B is a $m \times n$ matrix, then BB^T is a purple $m \times m$ matrix and $B^T B$ is a purple $n \times n$ matrix.

(viii) If A is a purple matrix, then A^2 is a purple matrix.

(ix) If A is a purple matrix, then A^3 is a purple matrix.

(x) If A and B are purple $n \times n$ matrices, then AB is purple.

(xi) If A and B are $n \times n$ matrices and B is purple, then ABA is purple.

(xii) If A and B are $n \times n$ matrices and B is purple, then ABA^T is purple.

7. (food for thought) Operations on matrices effect their rank in interesting ways. Here are some questions to think about (we will return to them later in the course, after test 1).

For each item below, suppose A and B are matrices so that the operations make sense.

(i) What is the relation between $\text{rank}(A)$ and $\text{rank}(A^T)$?

(ii) What is the relation between $\text{rank}(A)$, $\text{rank}(B)$ and $\text{rank}(A + B)$?

(iii) What is the relation between $\text{rank}(A)$, $\text{rank}(B)$ and $\text{rank}(AB)$?

For each of these questions, you can explore what is true by trying out various examples with different A and B.

In the following two questions, we explore some matrix operations other than the ones introduced in chapter 2. These operations have their own interesting properties and are useful in many applications. Let's explore these operations.

8. If A and B are matrices with the same dimension $n \times m$, then we can define the element-wise multiplication $A \vee B$ (this notation is not standard) to be the $n \times m$ matrix whose i, jth entry is simply $a_{i,j} b_{i,j}$.

(i) Compute the element-wise multiplication of

$$A = \begin{bmatrix} 0 & 1 \\ 3 & 4 \\ 0 & 9 \end{bmatrix}$$

and

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

(ii) What properties does element-wise multiplication have? For instance, is it always true that $A \vee B = B \vee A$? Is it always true that $A \vee (B + C) = A \vee B + A \vee C$ (provided the operations are defined)? Can you think of other properties to check?

(iii) Describe the differences between matrix multiplication AB and element-wise matrix multiplication $A \vee B$.

(iv) If A is a square matrix, then you can compute $A \vee A$. What can you say about the matrix $A \vee A$?

B.3 Tutorial 3: More systems of linear equations

In this tutorial we will practice: using matrix algebra to study systems of linear equations, and determining whether a vector is contained in a span.

Background: Chapter 3 of the lecture notes.

1. (warm-up) Compute the basic solutions of the following homogeneous system of equations.

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

Compare your answer with Problem 1 from Tutorial 1. Write the solutions to Problem 1 from Tutorial 1 as the sum of a particular solution plus a homogeneous solution.

2. Suppose A is a 2×4 matrix and $\mathbf{u} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}$ are basic solutions of the homogeneous system $A\mathbf{x} = \mathbf{0}$.

Are the follow statements true or false? Justify your answers.

(i) $\begin{bmatrix} 0 \\ -1 \\ 1 \\ 1 \end{bmatrix}$ is a solution.

(ii) $\begin{bmatrix} 6 \\ 3 \\ 3 \\ 3 \end{bmatrix}$ is a solution.

(iii) $\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$ is a solution.

(iv) There are 3 basic solutions.

(v) There are 2 basic solutions.

(vi) The rank of A is 2.

(vii) Every solution is a linear combination of $\mathbf{u}$ and $\mathbf{v}$.

3. Use matrix algebra to prove the following theorem.

> **Theorem:** Suppose that $\mathbf{u}_1, \ldots \mathbf{u}_k$ are solutions of the matrix equation $A\mathbf{x} = \mathbf{b}$. If $t_1, \ldots, t_k$ are real numbers such that
>
> $$t_1 + t_2 + \cdots + t_k = 1,$$
>
> then the linear combination
>
> $$t_1\mathbf{u}_1 + t_2\mathbf{u}_2 + \cdots + t_k\mathbf{u}_k$$
>
> is a solution of the matrix equation $A\mathbf{x} = \mathbf{b}$.

Once you are done proving this theorem, go back and use this result to answer question 3 in tutorial 1 again.

5. Suppose $\mathbf{u}$ and $\mathbf{v}$ are $m\times 1$ column vectors. Which of the following statements are always true? Justify your answers.

(i) $\mathbf{0}_{m\times 1} \in \operatorname{span}\{\mathbf{u}, \mathbf{v}\}$

(ii) $\mathbf{u} \in \operatorname{span}\{\mathbf{u}, \mathbf{v}\}$

(iii) $2\mathbf{u} - \mathbf{v} \in \operatorname{span}\{\mathbf{u}, \mathbf{v}\}$

(iv) $2\mathbf{u} - \mathbf{v} \in \operatorname{span}\{2\mathbf{u} - \mathbf{v}, \mathbf{u} + \mathbf{v}\}$

(v) $\mathbf{u} \in \operatorname{span}\{\mathbf{u} + \mathbf{v}, 2\mathbf{u} + 2\mathbf{v}\}$

(vi) $\mathbf{b} \in \operatorname{span}\{\mathbf{u}, \mathbf{v}\}$ for every $m \times 1$ column vector $\mathbf{b}$.

Suppose A is a $m \times 2$ matrix with columns $\mathbf{u}$ and $\mathbf{v}$ as above. Which of the following statements are always true? Justify your answers.

(i) The homogeneous system $A\mathbf{x} = \mathbf{0}$ is consistent.

(ii) The system $A\mathbf{x} = \mathbf{u}$ is consistent.

(iii) The system $A\mathbf{x} = 2\mathbf{u} - \mathbf{v}$ is consistent.

(iv) The system $A\mathbf{x} = \mathbf{u} + \mathbf{v}$ is consistent.

(v) The system $A\mathbf{x} = \mathbf{b}$ is consistent for every $m \times 1$ column vector $\mathbf{b}$.

(vi) The system $A\mathbf{x} = \mathbf{u} + \mathbf{b}$ is consistent for every $m \times 1$ column vector $\mathbf{b}$.

B.4 Tutorial 4: Matrix inverses

In this tutorial we will practice finding matrix inverses, using the big theorem for square matrices, and using properties of matrix inverses.

Background: Chapter 4 of the lecture notes.

1. (warm-up) For each of the following matrices, determine whether the matrix is invertible and, if it is invertible, find its inverse.

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

If the matrix is invertible, write the matrix and its inverse as a product of elementary matrices.

2. Consider a homogeneous system of two equations in the variables x_1, x_2 with arbitrary coefficients a, b, c, d,

$$\begin{aligned} ax_1 + bx_2 &= 0 \\ cx_1 + dx_2 &= 0 \end{aligned}$$

For what values of a, b, c, d does this system have exactly one solution? Give a complete explanation of your answer.

3. Definition: A square matrix A is called *orange*[2] if

$$A^T A = I \text{ and } AA^T = I.$$

(i) Which elementary matrices are orange?

(ii) Find two examples of orange matrices that aren't elementary matrices (hint: try products of elementary matrices).

(iii) Is the identity matrix orange? Why?

(iv) A orange matrix is always invertible. What is the inverse?

(v) Show that if A is orange, then A^T is orange.

(vi) Show that the product of two $n \times n$ orange matrices is orange.

4. Suppose that A is a $n \times n$ matrix such that $A\mathbf{x} = \mathbf{b}$ has a solution for every $\mathbf{b} \in \mathbb{R}^n$. Which of the following statements are true? Explain your answer.

(i) A is invertible.

(ii) $A\mathbf{x} = \mathbf{0}$ is consistent.

(iii) The span of the columns of A is $\mathbb{R}^n$.

(iv) There are elementary matrices $E_1, \ldots, E_k$ such that $E_k \cdots E_1 A = I$.

(v) A^T is invertible.

(vi) The span of the columns of A^T is $\mathbb{R}^n$.

(vii) The equation $A^{-1}\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b} \in \mathbb{R}^n$.

(viii) The rank of A is n.

[2]The real word for this property is *orthogonal*.

(ix) A^3 is invertible.

5. Use the Big Theorem for Square Matrices to prove the following statement.

If A is invertible, then A does not have a row of zeros.

On the other hand, consider the following statement:

If A does not have a row of zeros, then A is invertible.

Is this statement true?

Hint: Look at your results from the warm-up.

6. In this question A and B are $n \times n$ matrices.

(i) Show that if AB is invertible, then B is invertible.

Warning: Be careful that your logic is not circular. If you write B^{-1} in your proof, then you have implicitly assumed B is invertible while trying to explain why B is invertible.

(ii) Suppose that A is invertible and AB is not invertible. Explain why B cannot be invertible.

(iii) Use part (ii) to complete the proof from the section on the matrix inversion algorithm. Show that if $E_1, \ldots, E_k$ are elementary matrices, and $E_k \cdots E_2 E_1 A$ is not invertible, then A is not invertible.

7. (bonus) Suppose that A is an invertible $n \times n$ matrix and $\mathbf{u}, \mathbf{v}$ are nonzero $n \times 1$ column vectors.

(i) The Sherman-Morrison formula says that if $1 + \mathbf{v}^T A^{-1} \mathbf{u} \neq 0$, then $A + \mathbf{u}\mathbf{v}^T$ is invertible and the inverse is given by the formula

$$(A + \mathbf{u}\mathbf{v}^T)^{-1} = A^{-1} - \frac{1}{1 + \mathbf{v}^T A^{-1} \mathbf{u}} A^{-1} \mathbf{u}\mathbf{v}^T A^{-1}.$$

Verify the Sherman-Morrison formula.

This formula is used in many applications for quickly computing inverses of large matrices: If you have a matrix $C = A + B$ where A is a matrix whose inverse you can easily compute (maybe A is a diagonal matrix, for example) and B is a matrix of rank 1 (i.e. $B = \mathbf{u}\mathbf{v}^T$), then the Sherman-Morrison formula gives you a fast way to compute the inverse of C.

(ii) Suppose that $1 + \mathbf{v}^T A^{-1}\mathbf{u} = 0$. Is $A + \mathbf{u}\mathbf{v}^T$ invertible? Justify your answer.

B.5 Tutorial 5: Vector geometry

Background: Chapter 5 of the lecture notes.

1. (i) Find the equation of the plane that contains the points

$$(2, 4, 4), (2, -2, 2), (4, 2, 2) \in \mathbb{R}^3.$$

(ii) Find the equation of the plane that contains the points

$$(2, 1, 3), (3, 0, 2), (3, 3, 3) \in \mathbb{R}^3.$$

(iii) Show that the two planes from part (a) and part (b) are the same.

(iv) (food for thought) What trick did I use come up with the three points for part (b) using the three points from part (a) so that I knew the two planes would be the same without having to check?

2. (i) Show that the three points

$$P_0 = (5, 6, 1),\ P_1 = (-1, -2, 3),\ P_2 = (11, 14, -1)$$

are collinear.

(ii) Describe all the planes that contain all three points, P_0, P_1, P_2 (there are infinitely many).
Hint: use normal vectors.

3. (i) Show that if two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are parallel, then

$$\text{proj}_{\mathbf{u}}\mathbf{v} = \mathbf{v}.$$

(ii) Show that if $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are orthogonal, then

$$\text{proj}_{\mathbf{u}}\mathbf{v} = \mathbf{0}.$$

(iii) Show that for any vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$,

$$||\text{proj}_{\mathbf{u}}\mathbf{v}|| \leq ||\mathbf{v}||.$$

Hint: Use the definition of proj$_{\mathbf{u}}\mathbf{v}$ *and properties of the dot-product.*

4. Show that for any two vectors, $\mathbf{u} = [u_1\, u_2\, u_3]$ and $\mathbf{v} = [v_1\, v_2\, v_3]$, the vector

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

is a solution of the matrix equation

$$\begin{bmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{bmatrix} \mathbf{x} = \mathbf{0}.$$

5. Three points, $P_0, P_1, P_2 \in \mathbb{R}^3$ are called *collinear* if they all lie on the same line.

(i) Are the points $(0, 0, 0)$, $(1, 0, 0)$, $(-10, 0, 0)$ collinear?

(ii) Are the points $(0, -1, 4)$, $(0, -1, 7)$, $(0, -1, 10)$ collinear?

(iii) Are the points $(2, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$ collinear?

(iv) Show that the following two statements are equivalent:

- $P_0 = (x_0, y_0, z_0), P_1 = (x_1, y_1, z_1), P_2 = (x_2, y_2, z_2)$ are collinear.

- There is a number $t \in \mathbb{R}$ such that

$$\begin{aligned} x_2 &= tx_1 + (1-t)x_0 \\ y_2 &= ty_1 + (1-t)y_0 \\ z_2 &= tz_1 + (1-t)z_0 \end{aligned}$$

Hint: use the fact that a point and a vector determine a line.

(v) Use part (d) to show that if P_0, P_1, P_2 are collinear, then the matrix

$$\begin{bmatrix} x_0 & y_0 & z_0 \\ x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{bmatrix}$$

has rank ≤ 2.

(vi) Compute the rank of the matrix from part (e) for the three points from parts (a)-(c).

6. Consider the two vectors

$$\mathbf{u} = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

(i) Show that $\mathbf{u}$ and $\mathbf{v}$ are not parallel.

(ii) Write the equation of the plane through the origin determined by the vectors $\mathbf{u}$ and $\mathbf{v}$ (use the cross-product to find a normal vector).

(iii) Show that the plane through the origin determined by the vectors $\mathbf{u}$ and $\mathbf{v}$ equals the set

$$\text{span}\{\mathbf{u}, \mathbf{v}\}.$$

This has two steps:

- Show that any linear combination $t\mathbf{u}+s\mathbf{v}$ is a solution to the equation from part (b).
- Show that any point that solves the equation from part (b) can be written as a linear combination $t\mathbf{u}+s\mathbf{v}$.

7. Let $Q_1 = (1, -1)$ and $Q_2 = (-1, 1)$. Describe the set of points $P = (x, y)$ such that

$$d(P, Q_1) = d(P, Q_2).$$

8. (bonus) The area of a parallelogram in $\mathbb{R}^2$ with adjacent sides $\mathbf{u}$ and $\mathbf{v}$ can be computed with the formula

$$Area = ||\mathbf{u}||||\mathbf{v}|| \sin(\theta)$$

where θ is the angle between $\mathbf{u}$ and $\mathbf{v}$.

Use this formula and other properties of angles and dot products to show that if $\mathbf{u} = \begin{bmatrix} a \\ c \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} b \\ d \end{bmatrix}$, then

$$Area = ad - bc.$$

One approach: we know $Area^2 = ||\mathbf{u}||^2||\mathbf{v}||^2 \sin^2(\theta) = ||\mathbf{u}||^2||\mathbf{v}||^2(1-\cos^2(\theta))$. *Plug in formulas and simplify.*

B.6 Tutorial 6: Linear transformations

Background: Chapter 6 of the lecture notes.

1. Suppose that

$$\mathbf{u} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \mathbf{w} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

Let T be the linear transformation such that

$$T(\mathbf{u}) = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, T(\mathbf{v}) = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, T(\mathbf{w}) = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

(i) Use properties of linear transformations to find the matrix of T.

Hint: first, write the standard basis vectors e_1, e_2, e_3 *as linear combinations of* $\mathbf{u}, \mathbf{v}, \mathbf{w}$. *Second, use properties of linear transformations to compute the vectors* $T(e_1), T(e_2), T(e_3)$.

(ii) This linear transformation is invertible. Use the matrix from part (a) to compute the matrix of T^{-1}.

(iii) There is another (quicker) way to find the matrix of T^{-1} without using the matrix from part (a). What is it?

2. Consider two linear transformations of $\mathbb{R}^2$,

$$T_\theta\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$$
$$F\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$$

The transformation T_θ rotates $\mathbb{R}^2$ by an angle of θ in a counterclockwise direction. The transformation F reflects points in $\mathbb{R}^2$ about the y-axis.

(i) Consider the three vertices of a triangle, $P_0 = (1, 0)$, $P_1 = (-1, 1)$ and $P_2 = (-1, -1)$.

When $\theta = \pi/2$, $T_{\pi/2}$ describes a counter-clockwise rotation of $\mathbb{R}^2$ by an angle of $\pi/2$.

For all three points, P_0, P_1, P_2, compute

$$T_{\pi/2}(P_0),\ T_{\pi/2}(P_1),\ T_{\pi/2}(P_2).$$

Illustrate your computation by drawing the three points, and their image under $T_{\pi/2}$ in $\mathbb{R}^2$.

Draw a picture of the image of the triangle under $T_{\pi/2}$.

(ii) For all three points, P_0, P_1, P_2, compute

$$F(P_0),\ F(P_1),\ F(P_2).$$

Illustrate your computation by drawing the three points, and their image under F in $\mathbb{R}^2$.

Draw a picture of the image of the triangle under F

(iii) We can compose F and $T_{\pi/2}$ two ways.

For all three points, P_0, P_1, P_2, compute their image under the transformation $F \circ T_{\pi/2}$. Draw a picture of the result.

For all three points, P_0, P_1, P_2, compute their image under the transformation $T_{\pi/2} \circ F$. Draw a picture of the result.

(iv) Compare the two pictures from part (c). Does $F \circ T_{\pi/2} = T_{\pi/2} \circ F$?

3. For each item below, give an example of a linear transformation with that property or explain why no linear transformation with that property exists.

(i) T is injective and surjective.

(ii) T is injective but not surjective.

(iii) T is surjective but not injective.

(iv) $T\colon \mathbb{R}^3 \to \mathbb{R}^3$ is invertible.

(v) $T\colon \mathbb{R}^3 \to \mathbb{R}^3$ is invertible but not injective.

(vi) $T\colon \mathbb{R}^3 \to \mathbb{R}^3$ is injective but not surjective

(vii) For every nonzero vector $\mathbf{x}$, $T(\mathbf{x}) \neq \mathbf{x}$ and $T \circ T(\mathbf{x}) = \mathbf{x}$.

Hint: can you think of a matrix $A \neq I$ so that $A^2 = I$?

(viii) $T\colon \mathbb{R}^3 \to \mathbb{R}^3$ and the image of T is $\text{span}\{[1\,5\,2]^T\}$.

(ix) $T\colon \mathbb{R} \to \mathbb{R}^3$ and the image of T is $\text{span}\{[1\,5\,2]^T\}$.

(x) $T\colon \mathbb{R}^2 \to \mathbb{R}^3$ and the image of T is $\text{span}\{[1\,5\,2]^T\}$.

4. Given a column vector $\mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \in \mathbb{R}^n$.

We know that $\mathbf{u}^T\mathbf{u}$ is a number, and $\mathbf{u}\mathbf{u}^T$ is a $n \times n$ matrix.

(i) Use the definition of the orthogonal projection $\text{proj}_{\mathbf{u}}\mathbf{x}$ to show that the matrix of the linear transformation $T(\mathbf{x}) = \text{proj}_{\mathbf{u}}\mathbf{x}$ is given by the formula

$$A = \frac{1}{\mathbf{u}^T\mathbf{u}}\mathbf{u}\mathbf{u}^T.$$

(ii) Use the result from part (a) to compute the matrix of the orthogonal projection $\text{proj}_{\mathbf{u}}\colon \mathbb{R}^3 \to \mathbb{R}^3$ for the vector $\mathbf{u} = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$.

(iii) For the orthogonal projection in part (b), write the sets $\ker(\text{proj}_{\mathbf{u}})$ and $\text{im}(\text{proj}_{\mathbf{u}})$ as the span of a set of vectors.

(iv) Use the result from part (a), and properties of matrix algebra, to show that the matrix A of orthogonal projection onto a line is always purple (Remember, A is purple if $A = A^T$).

5. A linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^n$ is called *covfefe* if for every vector $\mathbf{x} \in \mathbb{R}^n$,

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

(i) Which of the following linear transformations $\mathbb{R}^2 \to \mathbb{R}^2$ are covfefe?

i. Rotation by θ degrees.

ii. Reflection through a coordinate axis.

iii. Scaling (dilation) by a factor of k.

iv. The shear transformation with matrix

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

v. Orthogonal projection onto a line through the origin.

(ii) Describe in words what a covfefe transformation does geometrically.

(iii) Recall that a square matrix A is called orange if $AA^T = A^TA = I$. Show that if A is orange, then the linear transformation $T_A(\mathbf{x}) = A\mathbf{x}$ is covfefe.

(iv) Show that if a linear transformation T is covfefe, then $\ker(T) = \{0\}$.

(v) Use the result of part (d) and the Theorems from section 6.9 of the lecture notes to explain why every covfefe linear transformation is invertible.

(vi) Use the result of part (e) and the Theorems from section 6.9 of the lecture notes to explain why every covfefe linear transformation is surjective.

6. (bonus) It's really quite remarkable that composition of liner transformations corresponds to matrix multiplication. Let's explore this for an example.

Recall that the matrix transformation

$$T_\theta\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$$

describes rotation of $\mathbb{R}^2$ around the origin by an angle of θ.

(i) Convince yourself geometrically (perhaps by drawing a picture) that if θ and α are two angles, then the compositions

$$T_\theta \circ T_\alpha = T_\alpha \circ T_\theta = T_{\theta+\alpha}.$$

(ii) We can also check that this is true algebraically by doing matrix multiplication. Using matrix multiplication, show that

$$T_\theta \circ T_\alpha = T_\alpha \circ T_\theta = T_{\theta+\alpha}.$$

Hint: you will need to use some trig identities (angle sum formulas). You don't need to memorize these specific trig formulas for this class. Just look them up for this problem.

B.7 Tutorial 7: Subspaces, spanning, and linear independence

Background: Chapter 7 of the lecture notes.

1. Let

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

and

$$\mathbf{u} = \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \mathbf{w} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

(i) Compare the subspaces $S = \text{span}\{\mathbf{u}\}$ and $S' = \text{span}\{E\mathbf{u}\}$.

(ii) Compare the subspaces $S = \text{span}\{\mathbf{v}\}$ and $S' = \text{span}\{E\mathbf{v}\}$.

(iii) Compare the subspaces $S = \text{span}\{\mathbf{u}, \mathbf{v}\}$ and $S' = \text{span}\{E\mathbf{u}, E\mathbf{v}\}$.

(iv) Compare the subspaces $S = \text{span}\{\mathbf{u}, \mathbf{w}\}$ and $S' = \text{span}\{E\mathbf{u}, E\mathbf{w}\}$.

(v) Compare the subspaces $S = \text{span}\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ and $S' = \text{span}\{E\mathbf{u}, E\mathbf{v}, E\mathbf{w}\}$.

2. Which of the following sets are subspaces of $\mathbb{R}^2$? Justify your answer. Draw a picture of what the set looks like.

(i) A line.

(ii) A point.

(iii) A circle of radius 2 centred at $(0, 2)$.

(iv) $\mathbb{R}^2$.

(v) The set of points equidistant from $(1, 1)$ and $(-1, -1)$.

(vi) The set of points $\{(x, y) \in \mathbb{R}^2 \colon y = x^2\}$.

(vii) The set of points $\{(x, y) \in \mathbb{R}^2 \colon x \geq 0\}$.

(viii) The set of points $\{(x, y) \in \mathbb{R}^2 \colon x^2 \geq 0\}$.

3. True or false.

(i) If S is a subspace of $\mathbb{R}^n$, then $\dim(S) < n$.

(ii) If S is a subspace of $\mathbb{R}^n$, then $\dim(S) > 0$.

(iii) If $\{\mathbf{u}, \mathbf{v}\}$ is a set of two vectors in $\mathbb{R}^3$, then there is a vector $\mathbf{w}$ so that $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is a basis for $\mathbb{R}^3$.

(iv) If $\{\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{x}\}$ is a set of vectors in $\mathbb{R}^3$, then it is linearly dependent.

(v) If $\mathbf{0}$ is not contained in a set of vectors $\mathcal{U}$, then $\mathcal{U}$ is linearly independent.

(vi) If S and S' are subspaces of $\mathbb{R}^n$ with $\dim(S) = \dim(S')$, then $S = S'$.

4. Suppose S is a set of points in $\mathbb{R}^n$ and $T \colon \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation.

Define the set

$$S' = \{T(\mathbf{x}) \colon \mathbf{x} \in S\} \subseteq \mathbb{R}^m.$$

We call S' the *image of S under T*.

(i) Let S be the line through the origin in $\mathbb{R}^2$ spanned by $\mathbf{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

i. If T is the linear transformation "rotate by an angle of $\pi/2$ radians clockwise" then what is the image of S under T?

ii. If T_A is the linear transformation corresponding to the matrix

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

then what is the image of S under T?

(ii) Show using the definition of subspace that if S is a subspace of $\mathbb{R}^n$, then the image S' of S under a linear transformation T is a subspace of $\mathbb{R}^m$.

(iii) Show that if $S = \text{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$, then $S' = \text{span}\{T(\mathbf{u}_1), \ldots, T(\mathbf{u}_k)\}$.

(iv) Show that if S is a subspace of $\mathbb{R}^n$, then $\dim(S') \leq \dim(S)$.

5. In this question we will finish a detail from the proof of the algorithm for finding a basis of the span of a set of vectors.

Show that if

- $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is a set of linearly independent vectors in $\mathbb{R}^n$, and
- A is an invertible $n \times n$ matrix,

then the set

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

is linearly independent.

B.8 Tutorial 8: The fundamental subspaces of a matrix

Background: Chapter 8 of the lecture notes.

1. Let

$$A = \begin{bmatrix} 1 & -2 & 0 & 0 & 3 \\ 2 & -5 & -3 & -2 & 6 \\ 0 & 5 & 15 & 10 & 0 \\ 2 & 6 & 18 & 8 & 6 \end{bmatrix}$$

The reduced row echelon form of A is

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

(i) What is the rank of A?

(ii) Find a basis for null(A). What is the nullity of A?

(iii) Find a basis for col(A). What is the dimension of the column space of A?

(iv) Find a basis for row(A). What is the dimension of the row space of A?

(v) Check the rank-nullity theorem for A.

(vi) Write a basis for the column and row space of A^T.

(hint: the rows of A are the columns of A^T and vis versa)

(vii) What is the rank of A^T? Why?

(It's important to note that you know the rank of A^T without having to row-reduce A^T)

(viii) What is the dimension of the null space of A^T?
(It's important to note that you can answer this question without finding a basis for the null space of A^T)

2. In this question we will explore examples and theory related to Theorem 8.9 in the lecture notes.

(i) Are the statements in Theorem 8.9 true for the matrix

$$A = \begin{bmatrix} 1 & -2 \\ 2 & -5 \\ 0 & 5 \\ 2 & 6 \end{bmatrix}?$$

(ii) Are the statements in Theorem 8.9 true for the matrix

$$A = \begin{bmatrix} 1 & -2 & 1 \\ 2 & -5 & 0 \end{bmatrix}?$$

(iii) Are the statements in Theorem 8.9 true for the matrix

$$A = \begin{bmatrix} 1 & 7 \\ 2 & -5 \end{bmatrix}?$$

(iv) Is it possible for any of the statements in Theorem 8.9 to be true for a matrix with fewer rows than columns ($m < n$)? Why?

(v) Use the result from part (d) to explain the following fact.

Fact: If A is a $m \times n$ matrix and $m < n$, then the matrix equation

$$XA = I_n$$

has no solutions (i.e. there is no $n \times m$ matrix X that solves this equation). In other words, A cannot have a left inverse.

(vi) Give an example of a 4×3 matrix A such that every statement in Theorem 8.9 is false for A.

(vii) Give an example of a linear transformation $T\colon \mathbb{R}^3 \to \mathbb{R}^4$ that is not injective.

3. In this question we will explore examples and theory related to Theorem 8.11 in the lecture notes.

(i) Are the statements in Theorem 8.11 true for the matrix

$$A = \begin{bmatrix} 1 & -2 \\ 2 & -5 \\ 0 & 5 \\ 2 & 6 \end{bmatrix}?$$

(ii) Are the statements in Theorem 8.11 true for the matrix

$$A = \begin{bmatrix} 1 & -2 & 1 \\ 2 & -5 & 0 \end{bmatrix}?$$

(iii) Are the statements in Theorem 8.11 true for the matrix

$$A = \begin{bmatrix} 1 & 7 \\ 2 & -5 \end{bmatrix}?$$

(iv) Is it possible for any of the statements in Theorem 8.9 to be true for a matrix with fewer columns than rows ($n < m$)? Why?

(v) Use the result from part (d) to explain the following fact.

Fact: If A is a $m \times n$ matrix and $n < m$, then the matrix equation

$$AX = I_m$$

has no solutions (i.e. there is no $n \times m$ matrix X that solves this equation). In other words, A cannot have a right inverse.

(vi) Give an example of a 3×4 matrix A such that every statement in Theorem 8.11 is false for A.

(vii) Give an example of a linear transformation $T\colon \mathbb{R}^4 \to \mathbb{R}^3$ that is not surjective.

4. In this question we will explore how matrix multiplication effects the null space and rank of a matrix.

The proofs of the theorems at the end of this question are bonus exercises, but the statements of the theorems are something that everyone should know.

(i) Consider the matrices

$$A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, \; B = \begin{bmatrix} 1 & -2 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

Compute a basis for null(B) and null(AB). Use the result of your computation to explain that

$$\text{null}(B) = \text{null}(AB).$$

(ii) Consider the matrices

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \ B = \begin{bmatrix} 1 & -2 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

Compute a basis for null(B) and null(AB). Use the result of your computation to explain that

$$\text{null}(B) \subseteq \text{null}(AB).$$

but the two subspaces are not equal.

(iii) Show that, in general, if A is a $m \times n$ matrix and B is a $n \times p$ matrix, then

$$\text{null}(B) \subseteq \text{null}(AB).$$

Hint: You simply need to show that every element of null(B) is an element of null(AB). In order to do this, you need to use the definition of "null space" and a simple matrix algebra argument.

(iv) (bonus) Use part (c) and the rank-nullity theorem to conclude the following theorem:

Theorem: If A is a $m \times n$ matrix and B is a $n \times p$ matrix, then

$$\text{rank}(B) \geq \text{rank}(AB).$$

(v) There is a similar but slightly different fact:

Theorem: If C is a $n \times m$ matrix and D is a $p \times n$ matrix, then

$$\text{rank}(D) \geq \text{rank}(DC).$$

(bonus) Prove this theorem.

Hint: Although this theorem is different from the previous one, you can use the previous one to prove it. Look at the theorem from the previous question. Use the fact that taking transposes does not change rank, but it does change the order of multiplication.

(vi) In part (a), we saw an example where A was invertible and we observed that null(B) = null(AB).

In fact, this is true in general:

Theorem: If A is an invertible $n \times n$ matrix and B is a $n \times p$ matrix, then

$$\text{null}(B) = \text{null}(AB).$$

(bonus) Prove this theorem.

Hint: By the result from part (c), we know that null(B) $\subseteq$ null(AB). To show that null(AB) $\subseteq$ null(B), use the fact that the homogeneous system $A\mathbf{x} = \mathbf{0}$ has a unique solution when when A is invertible.

Here is a hint for a different, slightly more complicated proof.

Hint for proof 2: First, use the fact that A can be written as a product of elementary matrices to explain why the rank of A and AB must be the same. Second, combine the rank-nullity theorem, the result from part (c), and Theorem 7.12(g) from the lecture notes.

B.9 Tutorial 9: Determinants

Background: Chapter 9 of the lecture notes.

1. (warm-up) Use cofactor expansion to compute the determinant of the following matrices.

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

(ii) $B = \begin{bmatrix} 1 & 4 & 2 \\ 6 & 1 & -1 \\ 3 & 1 & 1 \end{bmatrix}$.

(iii) $C = \begin{bmatrix} 1 & 2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 2 & 1 & 1 & 4 & 2 \\ 0 & 0 & 6 & 1 & -1 \\ 1 & 0 & 3 & 1 & 1 \end{bmatrix}$.

(iv) The matrix C in part (c) is an example of a matrix that is "block lower triangular." This means that C looks like

$$C = \begin{bmatrix} A & 0_{m\times n} \\ D & B \end{bmatrix}.$$

where A is an $m \times m$ submatrix, B is a $n \times n$ submatrix, and D is a $n \times m$ submatrix.

Based on the result of part (c), can you guess a general formula for $\det(C)$ in terms of $\det(A)$ and $\det(B)$ when C is block lower triangular? (you don't need to prove it)

2. Find the determinant of the following matrices by inspection (don't use co-factor expansion, instead use properties of determinants and to quickly "see" the determinant without computing).

(i) $\begin{bmatrix} 4 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 8 \end{bmatrix}.$

(ii) $\begin{bmatrix} 4 & 3 & 43243 & 5 & 121321 \\ 0 & 2 & 2 & 1000 & 10 \\ 0 & 0 & 2 & 82 & 9392234 \\ 0 & 0 & 0 & 3 & 65 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix}.$

(iii) $\begin{bmatrix} 1 & 0 & 4 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & -2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 6 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$

(iv) $\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$

(v) $\begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 100 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$.

(vi) $\begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 100 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$.

3. Suppose that A and B are 3×3 matrices with $\det(A) = -1$ and $\det(B) = 2$. Compute each of the following determinants using properties of the determinant.

(i) $\det(2A^2)$
(ii) $\det(A^3B^5)$
(iii) $\det(BABABAB^2)$
(iv) $\det(3A^2B^{-1})$
(v) $\det(3(A^2B^{-1})^T)$

4. Consider the matrix

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

(i) Compute $\det(AA^T)$
(ii) Compute $\det(A^TA)$
(iii) Here is a fake proof that $\det(AA^T) = \det(A^TA)$ (which we know is false by our computations in part (a) and (b)). Try to explain the mistake in this fake proof.

Fake proof: We show that $\det(AA^T) = \det(A^TA)$ using properties of determinants.

$$\begin{aligned} \det(AA^T) &= \det(A)\det(A^T) \text{ since } \det(AB) = \det(A)\det(B) \\ &= \det(A^T)\det(A) \text{ since the order of multiplying two numbers doesn't matter} \\ &= \det(A^TA) \text{ since } \det(AB) = \det(A)\det(B) \end{aligned}$$

$\square$

(iv) (food for thought) The computations from part (a) and (b) tell us that one of the matrices AA^T and A^TA is invertible, but the other is not. Can you relate this to Theorems 8.9 and 8.11 in the lecture notes?

B.10 Tutorial 10: Eigenvalues, eigenevectors, and diagonalization

Background: Chapter 10 of the lecture notes.

1. (warm-up) Compute the eigenvalues and corresponding eigenspaces for the matrices below.

$$\begin{bmatrix} 7 & 4 \\ -3 & -1 \end{bmatrix}, \begin{bmatrix} 4 & -2 \\ -3 & 9 \end{bmatrix}, \begin{bmatrix} 4 & 2 & 3 \\ -1 & 1 & -3 \\ 2 & 4 & 9 \end{bmatrix}, \begin{bmatrix} 4 & 0 & 0 \\ -1 & 4 & -3 \\ 0 & 0 & 1 \end{bmatrix}.$$

Which of the matrices are diagonalizable? For each matrix that is diagonalizable, compute its diagonalization.

2. In this question we explore an application of matrix diagonalization: computing large powers of a matrix[3]

(i) If $A = PDP^{-1}$, then there is a simple formula for A^k (where k is a positive integer). What is it?

(ii) For each of the matrices from question 1 that are diagonalizable, use formula from part (a) to compute A^{2017} (you don't need to simplify powers of numbers that appear in your final answer).

3. Suppose that A and B are 3×3 matrices. Which of the following statements are true or false, and why?

[3] Computing the matrix product AB (where A and B are $n \times n$ matrices) using the definition of matrix multiplication is an algorithm that takes an order of n^3 operations (one is counted every time you add, multiply two numbers together). For instance, if you try to compute A^{1000}, and A is a 10×10 matrix, you will use roughly $1,000,000$ operations. However, if A is diagonalized, you only need to compute powers of diagonal entries, which is much faster.

It is interesting to note that there are clever algorithms for matrix multiplication that are more efficient than the definition when taking products of large matrices (even if the matrices aren't diagonalizable). See the Strassen algorithm on wikipedia.

(i) If -2 is an eigenvalue of A, then -8 is an eigenvalue of A^3.

(ii) If -2 is an eigenvalue of A, then -8 is an eigenvalue of $2A^3$.

(iii) If -2 is an eigenvalue of A and -2 is an eigenvalue of B, then 4 is an eigenvalue of AB.

(iv) If -2 is an eigenvalue of A and -2 is an eigenvalue of B, then -2 is an eigenvalue of $A + B$.

(v) If $A^2 = I$, then the only eigenvalues of A are 1 or -1.

(vi) If $A^2 = A$, then the only eigenvalues of A are 1.

(vii) If A has two distinct eigenvalues, then A is diagonalizable.

(viii) If A is invertible, then A is diagonalizable.

4. In this question we will add another fact to our list of equivalent statements for $n \times n$ matrices.

Theorem Suppose that A is a $n \times n$ matrix. The following two statements are equivalent.

(i) A is invertible.

(ii) 0 is not an eigenvalue of A.

Hint: use the characteristic polynomial, and the list of statements equivalent to (a).

5. Suppose $\mathbf{u} \in \mathbb{R}^n$ is nonzero. Orthogonal projection onto the line spanned by $\mathbf{u}$ is the linear transformation $T(\mathbf{x}) = \text{proj}_{\mathbf{u}}(\mathbf{x})$.

Recall that we showed in Tutorial 6, question 4 that the matrix of this linear transformation is given by the formula

$$A = \frac{1}{\mathbf{u}^T\mathbf{u}}\mathbf{u}\mathbf{u}^T.$$

(i) Show that 1 and 0 are eigenvalues of A by using the definition of eigenvalues (find some eigenvectors for these eigenvalues). *Hint: this question may be easier if you think about the geometric properties of* $\text{proj}_{\mathbf{u}}$.

(ii) What are the eigenspaces $E_0(A)$ and $E_1(A)$? Can you describe them geometrically? What are their dimensions?

(iii) Use the result of parts (a) and (b), together with what we know about eigenvalues and eigenvectors, to conclude that 0 and 1 are the only eigenvalues of A. *Hint: the facts in section 10.4 of the lecture notes might be useful.*

(iv) Use the result of part (c) to explain why A is diagonalizable. What is the matrix D of the diagonalization? (you don't need to find a formula for P)

B.11 Tutorial 11: Orthogonality

Background: Chapter 11 of the lecture notes.

1. Let

$$S = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ -2 \\ 0 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 4 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 4 \end{bmatrix} \right\} \subseteq \mathbb{R}^5$$

(i) Use Gram-Schmidt to compute an orthogonal basis for the subspace S.

(ii) Compute a basis for $S^\perp$ using the fact that $\text{row}(A)^\perp = \text{null}(A)$. Then use Gram-Schmidt to compute an orthogonal basis for $S^\perp$.

(iii) Use the computations from parts (a) and (b) to confirm the formula

$$\dim(S) + \dim(S^\perp) = n$$

for this example.

(iv) Check that if the vectors from the orthogonal basis obtained in part (a) for S and the vectors from the orthogonal basis obtained in part (b) for $S^\perp$ are combined into one set, then the resulting set of 5 vectors is an orthogonal basis for $\mathbb{R}^5$.

2. Let

$$\mathbf{v} = \begin{bmatrix} 2 \\ 3 \\ -1 \\ 1 \\ 4 \end{bmatrix}.$$

(i) Use the result of Question 1.(a) to compute the orthogonal projection $\text{proj}_S(\mathbf{v})$.

(ii) Use the formula

$$\text{proj}_{S^\perp}(\mathbf{v}) = \mathbf{v} - \text{proj}_S(\mathbf{v})$$

and the result from 2.(a) to compute $\text{proj}_{S^\perp}(\mathbf{v})$.

(iii) Check that the vectors computed in 2.(a) and 2.(b) are orthogonal.

(iv) Use the orthonormal basis for $S^\perp$ from question 1.(b) to compute $\text{proj}_{S^\perp}(\mathbf{v})$ by using the definition of orthogonal projection onto the subspace $S^\perp$ (you should get the same answer as in 2.(b)).

Would it have been possible to compute $\text{proj}_S(\mathbf{v})$ and $\text{proj}_{S^\perp}(\mathbf{v})$ without having first computed an orthogonal basis for S in question 1?

3. Suppose that S is a subspace of $\mathbb{R}^n$ with dimension k. For each statement below, say whether it is true or false. Can you explain why?

(i) $\dim(S^\perp) = n - k$.

(ii) If $\mathbf{x}$ is contained in both S and $S^\perp$, then $\mathbf{x} = \mathbf{0}$.

(iii) The orthogonal complement of $S^\perp$ is S. In other words, $(S^\perp)^\perp = S$.

(iv) If $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthogonal set, then it is linearly independent.

(v) For any vector $\mathbf{x} \in \mathbb{R}^n$,

$$||\mathbf{x}||^2 = ||\text{proj}_S\mathbf{x}||^2 + ||\text{proj}_{S^\perp}\mathbf{x}||^2.$$

4. Recall from earlier tutorials that we called a $n \times n$ matrix A is orange[4] if $AA^T = A^TA = I$.

(i) Show that A is orange if and only if the columns of A are an orthonormal basis for $\mathbb{R}^n$. *Hint: Recall that the ij-entry of AB is the dot-product of the ith row of A and the jth column of B.*

(ii) Show that if $\mathbf{u}$ and $\mathbf{v}$ are orthogonal, and A is orange, then $A\mathbf{u}$ and $A\mathbf{v}$ are orthogonal.

(iii) Show that if $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthogonal set, then $\{A\mathbf{u}_1, \ldots, A\mathbf{u}_k\}$ is an orthogonal set.

[4] In the lecture notes, we called this orthogonal.

(iv) Recall from an earlier tutorial that we showed the following: If A is orthogonal, then for any vector $\mathbf{u}$, $||A\mathbf{u}|| = ||\mathbf{u}||$.

(v) Show that if $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthonormal set, then $\{A\mathbf{u}_1, \ldots, A\mathbf{u}_k\}$ is an orthonormal set.

Appendix C

Tutorial sample solutions

C.1 Tutorial 1-4

Tutorial 1

1. **(a)** The augmented matrix of this system is:

$$\left[\begin{array}{ccccc|c} 3 & 2 & -1 & 0 & 1 & 4 \\ 0 & 1 & 0 & -1 & 0 & 4 \\ -1 & -1 & 1 & -3 & 0 & 5 \end{array}\right]$$

(b) It has RREF:

$$\left[\begin{array}{ccccc|c} 1 & 0 & 0 & -1 & \frac{1}{2} & \frac{5}{2} \\ 0 & 1 & 0 & -1 & 0 & 4 \\ 0 & 0 & 1 & -5 & \frac{1}{2} & \frac{23}{2} \end{array}\right]$$

- The rank of the coefficient matrix is 3.
- The set of solutions to this SLE is:

$$\left\{ \left(\frac{5}{2} + s - \frac{1}{2}t, 4 + s, \frac{23}{2} + 5s - \frac{1}{2}t, s, t \right) \in \mathbb{R}^5 \mid s, t \in \mathbb{R} \right\}.$$

- The system is consistent.
- There are infinitely many solutions to this system. This is because the rank (3) of the coefficient matrix is less than the number of columns in the coefficient matrix (5).

2. Give examples of systems of linear equations in three variables (call them x, y, and z) with each of the following properties, or explain why no example exists. Justify your answers.

There are many valid answers to these questions, here are the most simple answers:

(a) A system with exactly one solution, $(1, 1, -1)$. $x = 1, y = 1, z = -1$

(b) A system that is inconsistent. $0x + 0y + 0z = 1$

(c) A system with infinitely many solutions, such that the solutions satisfy the equation $x + y = z$. $x + y = z$

(d) A system with exactly four solutions. Impossible. If a system of linear equations has more than one solution, then it has infinitely many solutions.

(e) A system whose augmented matrix has rank 2. $x = 1$, $y = 1$.

(f) A system whose augmented matrix has rank 0. $0x + 0y + 0z = 0$ (this is not a very interesting system of linear equations, but it is a system)

(g) A system whose augmented matrix has rank 4. $x = 1$, $y = 1$, $z = 1$, $0x + 0y + 0z = 1$.

(h) A system whose augmented matrix has rank 5. Impossible, rank must be less than 5 (.

3. Suppose that a, b, c, d, e, and f are numbers and

$$\begin{aligned} ax + by &= e \\ cx + dy &= f \end{aligned}$$

is a system of linear equations in the variables x and y.

(a) Prove that if (x_0, y_0) and (x_1, y_1) are solutions to this system, then for every real number $t \in \mathbb{R}$, the pair

$$(tx_1 + (1-t)x_0,\ ty_1 + (1-t)y_0)$$

is a solution.

Proof. To show that for any $t \in \mathbb{R}$,

$$(tx_1 + (1-t)x_0,\ ty_1 + (1-t)y_0)$$

is a solution to the system of equations, we must plug these pairs of numbers into both equations and show that the equation is true.

For the first equation,

$$\begin{aligned} a(tx_1 + (1-t)x_0) + b(ty_1 + (1-t)y_0) &= t(ax_1 + by_1) + (1-t)(ax_0 + by_0) \\ &= t(e) + (1-t)(e) \\ &= e \end{aligned}$$

so the equation is true.

For the second equation,

$$\begin{aligned} c(tx_1 + (1-t)x_0) + d(ty_1 + (1-t)y_0) &= t(cx_1 + dy_1) + (1-t)(cx_0 + dy_0) \\ &= t(f) + (1-t)(f) \\ &= f \end{aligned}$$

so the equation is true.

Since the pair satisfies both equations in the system, it is a solution of the system. □

(b) Use part (a) to explain why a 2×2 system of linear equations can only have 0 solutions, 1 solution or infinitely many solutions.

Proof. First, without explaining anything, we know that there are three possibilities: the system has 0 solutions, the system has 1 solution, or the system has more than one solution.

We just need to explain why in the last case, if a system has more than one solution, then it has infinitely many solutions.

If the system has more than one solution, then it has at least two solutions. Let's call those solutions (x_0, y_0), (x_1, y_1). Part (a) tells us that if these two pairs are solutions, then every pair

$$(tx_1 + (1-t)x_0,\ ty_1 + (1-t)y_0)$$

is a solution as well. But since there are infinitely many numbers $t \in \mathbb{R}$, this means there are infinitely many pairs

$$(tx_1 + (1-t)x_0,\ ty_1 + (1-t)y_0)\,.$$

So the system has infinitely many solutions. $\square$

4. If we multiply an equation by 0, then we can potentially introduce new solutions to the system. For example, consider the system

$$\begin{aligned} x + y &= 2 \\ x - y &= 0. \end{aligned}$$

It has a unique solution, $(1, 1)$.

If we multiply the second equation in this system by 0, the system becomes just $x + y = 2$, whose solutions are of the form $(1 + t, 1 - t)$.

Thus, multiplying an equation by 0 can change the set of solutions to the system.

The point of elementary operations is that *they don't change the set of solutions.* This is why the second elementary operation is multiplying an equation by a *nonzero number.*

5. The following two statements are equivalent:

- The augmented matrix $[A|b]$, corresponds to an inconsistent system of linear equations.
- A row echelon form of $[A|b]$ contains the row $[0 \cdots 0\,1]$.

If a A row echelon form of $[A|b]$ contains the row $[0 \cdots 0\,1]$, then there is a leading entry in the last column of the row echelon form.

Since the row echelon form of $[A|b]$ has r leading entries (rank = r), and one of those leading entries is in the last column, A has $r - 1$ leading entries.

If the system is consistent, then all the leading 1's in a row echelon form of $[A|b]$ are in the coefficient part of the augmented matrix (not the last column), so $\text{rank}(A) = r = \text{rank}([A|b])$.

For example, the augmented matrix in REF

$$[A|b] = \left[\begin{array}{ccccc|c} 1 & 0 & 0 & -1 & \frac{1}{2} & \frac{5}{2} \\ 0 & 1 & 0 & -1 & 0 & 4 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

corresponds to an inconsistent system of equations because the third row is $[0\,0\,0\,0\,0\,1]$.

The rank of $[A|b]$ is 3. The coefficient matrix

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

has rank $3 - 1 = 2$.

Tutorial 2

1. (a) and (c) are equal. (b) and (d) are equal. In general, (e) and (f) are not equal (they don't even have the same number of rows and columns! One matrix is $m \times m$, the other is $n \times n$).

2. First, let

$$B = \begin{bmatrix} w & x \\ y & z \end{bmatrix}$$

where w, x, y, z are variables we want to solve for.

Next, evaluate both sides of the matrix equation $AB = BA$,

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} w & x \\ y & z \end{bmatrix} = \begin{bmatrix} w & x \\ y & z \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

$$\begin{bmatrix} w+y & x+z \\ w & x \end{bmatrix} = \begin{bmatrix} w+x & w \\ y+z & y \end{bmatrix}.$$

This matrix equation is equivalent to the system of equations

$$\begin{aligned} w + y &= w + x \\ x + z &= w \\ w &= y + z \\ x &= y \end{aligned}$$

After a little work, we can solve this system and find that the solutions are

$$\begin{aligned} w &= s + t \\ x &= s \\ y &= s \\ z &= t \end{aligned}$$

for any $s, t \in \mathbb{R}$. Thus the set of matrices B that solve the matrix equation is

$$\left\{ \begin{bmatrix} s+t & s \\ s & t \end{bmatrix} : s, t \in \mathbb{R} \right\}$$

3. Consider the following list of algebraic identities and facts that are true when x, y, and z are real numbers.

(a) $A + B = B + A$. By properties of matrix addition, this is always true when A, B are $m \times n$ matrices.

(b) $AB = BA$. This is not always true! For instance, take

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \; B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

(c) If $AB = AC$ and $A \neq 0$, then $B = C$. This is not always true! For instance, take

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \; B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \; C = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

We see that $A \neq 0$ and $B \neq C$, but $AB = AC$.

4. (a) Can you find some or all solutions to the matrix equation $X^2 - I = 0$? There are a couple quick solutions, namely $X = I$ and $X = -I$. For the others, Write X as

$$X = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Then we have that

$$X^2 = \begin{bmatrix} a^2 + bc & b(a+d) \\ c(a+d) & d^2 + bc \end{bmatrix}.$$

For this to be equal to I, we need the following equations to all be satisfied:

$$a^2 + bc = 1 \tag{C.1}$$
$$b(a+d) = 0 \tag{C.2}$$
$$c(a+d) = 0 \tag{C.3}$$
$$d^2 + bc = 1 \tag{C.4}$$
$$\tag{C.5}$$

We will solve this case by case. Since both $b(a+d) = 0$ and $c(a+d) = 0$, either both of b and c must be 0, or $a+d = 0$. If b and c are 0, then we get that $a^2 = d^2 = 1$, which give the solutions:

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

If $a + d = 0$, then we have that $a = -d$. Then for any choice of a, b, c that satisfy $bc = 1 - a^2$, the matrix

$$X = \begin{bmatrix} a & b \\ c & -a \end{bmatrix}$$

will square to I.

(b) Is it possible to use Guassian elimination to solve this system of equations? No. The system of equations (C.1) above is not a system of linear equations in the variables (a, b, c, d).

6. A matrix A is called *purple*[1] if $A = A^T$. For each of the following statements, say whether it is true or false. Justify your answers with an explanation.

[1] The real word for this property is *symmetric*.

(i) False.

(ii) True.

(iii) False.

(iv) True.

(v) True.
Proof: If A and B are purple, then

$$\begin{aligned} (A+B)^T &= A^T + B^T && \text{By properties of transpose.} \\ &= A + B && \text{Since } A,\ B \text{ are purple.} \end{aligned}$$

so $A+B$ is purple.

(vi) True.

(vii) True.
Proof: We show BB^T is purple. For any matrix B,

$$\begin{aligned} (BB^T)^T &= (B^T)^T B^T && \text{By properties of transpose.} \\ &= BB^T && \text{By properties of transpose.} \end{aligned}$$

so BB^T is purple.

(viii) True.

(ix) True.

(x) False.

Explanation: In general, properties of matrix algebra tell us that

$$\begin{aligned} (AB)^T &= B^T A^T && \text{By properties of transpose.} \\ &= BA && \text{Since } A \text{ and } B \text{ are purple.} \end{aligned}$$

but it is not always true that $AB = BA$! So it is not always true that AB is purple.

(xi) True.

Proof: If A and B are purple, then

$$\begin{aligned} (ABA)^T &= A^T B^T A^T && \text{By properties of transpose.} \\ &= ABA && \text{Since } A,\ B \text{ are purple.} \end{aligned}$$

so ABA is purple.

(xii) True.

Tutorial 3

1. This system has augmented matrix:

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

It has RREF:

$$\left[\begin{array}{ccccc|c} 1 & 0 & 0 & -1 & \frac{1}{2} & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & -5 & \frac{1}{2} & 0 \end{array}\right]$$

Therefore, the set of solutions is

$$\left\{ \left(s - \frac{1}{2}t, s, 5s - \frac{1}{2}t, s, t \right) |s, t \in \mathbb{R} \right\}.$$

The solutions to the system in question 1, tutorial 1 will be given by any particular solution to that system plus a solution to this system.

An example of a particular solution to the non-homogeneous system of equations from question 1, tutorial 1 is

$$\left(\frac{5}{2}, 4, \frac{23}{2}, 0, 0 \right).$$

2. (i) True.

Explanation:

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

so it is a linear combination of basic solutions, so it is a solution.

(ii) True.

Explanation:

$$\begin{bmatrix} 6 \\ 3 \\ 3 \\ 3 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + 3 \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

so it is a linear combination of basic solutions, so it is a solution.

(iii) False.

Explanation: The matrix equation

$$\begin{bmatrix} 0 \\ -1 \\ 1 \\ 1 \end{bmatrix} = s \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

has no solutions.

(iv) False.

Explanation: Looking at the form of the two basic solutions we are given, we can deduce that x_1, x_3 are leading variables and x_2, x_4 are free variables.

Thus there are exactly two basic solutions (one for each free variable).

(v) True.

(vi) True.

Explanation: The rank of A equals the number of leading variables, which is 2.

(vii) True.

Explanation: We know from our explanation in part (d) that there are only two basic solutions: $\mathbf{u}, \mathbf{v}$.

We know that the set of solutions to a homogeneous system equals the span of the basic solutions.

3. *Proof.* Suppose that $\mathbf{u}_1, \ldots \mathbf{u}_k$ are solutions of the matrix equation $A\mathbf{x} = \mathbf{b}$. If $t_1, \ldots, t_k$ are real numbers such that

$$t_1 + t_2 + \cdots + t_k = 1.$$

To show that

$$t_1\mathbf{u}_1 + t_2\mathbf{u}_2 + \cdots + t_k\mathbf{u}_k$$

is a solution of the matrix equation $A\mathbf{x} = \mathbf{b}$ we just need to plug it into the left hand side of the equation and show that the equation is true.

$$\begin{aligned}
A(t_1\mathbf{u}_1 + t_2\mathbf{u}_2 + \cdots + t_k\mathbf{u}_k) &= A(t_1\mathbf{u}_1) + A(t_2\mathbf{u}_2) + \cdots + A(t_k\mathbf{u}_k) && \text{matrix algebra} \\
&= t_1(A\mathbf{u}_1) + t_2(A\mathbf{u}_2) + \cdots + t_k(A\mathbf{u}_k) && \text{matrix algebra} \\
&= t_1(\mathbf{b}) + t_2(\mathbf{b}) + \cdots + t_k(\mathbf{b}) && \text{since } A\mathbf{u}_i = \mathbf{b} \\
&= (t_1 + t_2 + \cdots t_k)\mathbf{b} && \text{matrix algebra} \\
&= 1\mathbf{b} && \text{since } t_1 + t_2 + \cdots + t_k = 1 \\
&= \mathbf{b}.
\end{aligned}$$

Thus, we have shown that

$$t_1\mathbf{u}_1 + t_2\mathbf{u}_2 + \cdots + t_k\mathbf{u}_k$$

is a solution of the matrix equation $A\mathbf{x} = \mathbf{b}$. □

4. If a system $A\mathbf{x} = \mathbf{b}$ has the trivial solution, then it is homogeneous. Explain why this is true.

The system has the trivial solution means that $\mathbf{x} = \mathbf{0}$ is a solution.

This means that the equation $A\mathbf{0} = \mathbf{b}$ is true.

But $\mathbf{0} = A\mathbf{0} = \mathbf{b}$, so it must be the case that $\mathbf{b} = \mathbf{0}$.

Since $\mathbf{b} = \mathbf{0}$, the original equation $A\mathbf{x} = \mathbf{b}$ is homogeneous.

5. Suppose $\mathbf{u}$ and $\mathbf{v}$ are $m \times 1$ column vectors. Which of the following statements are always true? Justify your answers.

(i) True. $\mathbf{0} = 0\mathbf{u} + 0\mathbf{v}$, so $\mathbf{u}$ is a linear combination of $\mathbf{u}$ and $\mathbf{v}$.

(ii) True. $\mathbf{u} = 1\mathbf{u} + 0\mathbf{v}$.

(iii) True. $2\mathbf{u} - \mathbf{v}$ is a linear combination of $\mathbf{u}$ and $\mathbf{v}$.

(iv) True. $2\mathbf{u} - \mathbf{v} = 1(2\mathbf{u} - \mathbf{v}) + 0(\mathbf{u} + \mathbf{v})$. So $2\mathbf{u} - \mathbf{v}$ is a linear combination of $2\mathbf{u} - \mathbf{v}$ and $\mathbf{u} + \mathbf{v}$.

(v) False. This is not always true. For example, $\mathbf{u} = [1\,0]^T$, $\mathbf{v} = [1\,1]^T$.

(vi) False. This is not always true. For example, $\mathbf{u} = [1\,0]^T$, $\mathbf{v} = [2\,0]^T$. Then $\mathbf{b} = [0\,1]^T$ is not in the span.

Suppose A is a $m \times 2$ matrix with columns $\mathbf{u}$ and $\mathbf{v}$ as above. Which of the following statements are always true? Justify your answers.

(i) True.
(ii) True.
(iii) True.
(iv) True.
(v) False. Compare with (f) above.
(vi) False.

Tutorial 4

1. A is invertible, B is not invertible.

2. By the big theorem for square matrices, the system has exactly one solution if and only if A is invertible.

We know that the matrix
$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$
is invertible if and only if $ad - bc \neq 0$.

Thus, the system has exactly one solution if and only if $ad - bc \neq 0$.

3. Definition: A square matrix A is called *orange*[2] if

$$A^T A = I \text{ and } AA^T = I.$$

(i) The only elementary matrices that are orange are: the identity matrix, the elementary matrix corresponding to multiplying a row by -1, and the elementary matrices corresponding to interchanging two rows.

(ii) Find two examples of orange matrices that aren't elementary matrices (hint: try products of elementary matrices).

(iii) First, $I^T = I$. Second, $I^T I = II = I$ and $II^T = II = I$.

(iv) A orange matrix is always invertible. What is the inverse? A^T.

(v) *Explanation:* By definition, A^T is invertible if there is a matrix B such that $A^T B = I$ and $BA^T = I$.

We already know that $A^T A = I$ and $AA^T = I$, so we know that A^T is invertible.

(vi) *Proof:* Suppose that A and B are orange matrices. We need to show that $(AB)(AB)^T = I$ and $(AB)^T(AB) = I$.

First,

$$\begin{aligned}
(AB)(AB)^T &= (AB)(B^T A^T) && \text{By properties of transpose.}\\
&= A(BB^T)A^T && \text{Matrix algebra.}\\
&= A(I)A^T && B \text{ is orange.}\\
&= (AI)A^T && \text{Matrix algebra.}\\
&= AA^T && \text{Matrix algebra.}\\
&= I && A \text{ is orange.}
\end{aligned}$$

[2] The real word for this property is *orthogonal*.

Second,

$$\begin{aligned}
(AB)^T(AB) &= (B^TA^T)(AB) && \text{By properties of transpose.}\\
&= B^T(A^TA)B && \text{Matrix algebra.}\\
&= B^T(I)B && A \text{ is orange.}\\
&= B^T(IB) && \text{Matrix algebra.}\\
&= B^TB && \text{Matrix algebra.}\\
&= I && B \text{ is orange.}
\end{aligned}$$

4. Note first that the criteria that $A\mathbf{x} = \mathbf{b}$ has a solution for all $\mathbf{b} \in \mathbb{R}^n$ is equivalent to A being invertible by the big theorem on square matrices (part 1).

(i) True. (by big theorem)

(ii) True. (always true)

(iii) True. (as the $\mathbf{b}$ that make the system consistent are precisely the elements of the span of the columns).

(iv) True. (by big theorem and in particular the matrix inversion algorithm)

(v) True. (the transpose of an invertible matrix is invertible)

(vi) True. (similar to (c))

(vii) True. (by big theorem and the fact that A^{-1} is invertible).

(viii) True. (by big theorem)

(ix) True. (since A is invertible, A^3 is invertible).

5. Use the Big Theorem for Square Matrices to prove the following statement.

If A is invertible, then A does not have a row of zeros.

Proof. By the big theorem, since A is invertible, we know that the rank of A is n.

Since the rank of A is n, any REF of A has n leading entries.

Since A is $n \times n$, this means that every row in the REF of A contains a leading entry.

Thus, A cannot have a row of zeros (since there would be a row of zeros in the REF). $\square$

> If A does not have a row of zeros, then A is invertible.

This statement is false. For example, the matrix

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

does not have a row of zeros, and it is not invertible.

C.2 Tutorial 5-8

Tutorial 5: Vector geometry

Background: Chapter 5 of the lecture notes.

1. (i) Find the equation of the plane that contains the points

$$(2, 4, 4), (2, -2, 2), (4, 2, 2) \in \mathbb{R}^3.$$

Solution: We will give these points names: $P_0 = (2, 4, 4)$, $P_1 = (2, -2, 2)$, and $P_2 = (4, 2, 2)$.

We calculate that $\overrightarrow{P_0P_1} = \begin{bmatrix} 0 \\ -6 \\ -2 \end{bmatrix}$, and $\overrightarrow{P_1P_2} = \begin{bmatrix} 2 \\ 4 \\ 0 \end{bmatrix}$.

Let $\mathbf{a} = \overrightarrow{P_0P_1} \times \overrightarrow{P_1P_2} = \begin{bmatrix} 8 \\ -4 \\ 12 \end{bmatrix}$.

Therefore, the equation of the plane containing P_0, P_1, and P_2 is $\mathbf{a} \cdot \mathbf{x} = b$, and we just need to calculate b. To do this we plug in one point. P_1 will give us: $b = (8)(2) + (-4)(-2) + (12)(2) = 48$. Therefore the equation of the plane is

$$8x - 4y + 12z = 48.$$

(ii) Find the equation of the plane that contains the points

$$Q_0 = (2, 1, 3), Q_1 = (3, 0, 2), Q_2 = (3, 3, 3) \in \mathbb{R}^3.$$

Solution: First,

$$\overrightarrow{Q_0Q_1} = \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}, \overrightarrow{Q_0Q_2} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}.$$

Second,

$$\mathbf{a} = \overrightarrow{Q_0Q_1} \times \overrightarrow{Q_0Q_2} = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}.$$

The equation of the plane is then

$$\mathbf{a} \cdot \mathbf{x} = \mathbf{a} \cdot Q_0$$

which gives

$$2x_1 - x_2 + 3x_3 = 12.$$

(iii) Show that the two planes from part (a) and part (b) are the same.

Solution: The equation we obtained in part (a) was

$$8x - 4y + 12z = 48.$$

The equation we obtained in part (b) was

$$2x_1 - x_2 + 3x_3 = 12.$$

The first equation equals 4 times the second equation, so the set of solutions to these two equations is the same. (in other words, these two systems of one equation differ by a single elementary operation).

2. (i) Show that the three points

$$P_0 = (5, 6, 1),\ P_1 = (-1, -2, 3),\ P_2 = (11, 14, -1)$$

are collinear.

Solution: We can see by inspection that

$$P_2 = 2P_0 - P_1 = (1 - (-1))P_0 + (-1)P_1,$$

so the three points are collinear.

(ii) Describe all the planes that contain all three points, P_0, P_1, P_2 (there are infinitely many).

Solution: The line containing the three points is parallel to the vector $\mathbf{d} = \overrightarrow{P_0P_1} = (-6, -8, 2)$.

The set of vectors that are orthogonal to $\mathbf{d}$ (normal vectors to planes that contain the line) is the set

$$\{\mathbf{n} = (n_1, n_2, n_3) \in \mathbb{R}^3 \colon \mathbf{n} \cdot \mathbf{d} = 0\}$$

Which is just the set of solutions of the equation

$$-6n_1 - 8n_2 + 2n_3 = 0.$$

After solving this system we find

$$\{\mathbf{n} = (n_1, n_2, n_3) \in \mathbb{R}^3 \colon \mathbf{n}{\cdot}\mathbf{d} = 0\} = \{\mathbf{n} = ((t{-}4s)/3, s, t) \in \mathbb{R}^3 \colon s, t \in \mathbb{R}\}.$$

All the planes containing the line are of the form

$$\mathbf{n} \cdot \mathbf{x} = \mathbf{n} \cdot P_0$$
$$\frac{t-4s}{3}x_1 + sx_2 + tx_3 = \frac{5(t-4s)}{3} + 6s + t.$$

where for each plane the numbers $s, t \in \mathbb{R}$ have been fixed, and x_1, x_2, x_3 are the variables.

Notice that some of these equations will determine the same plane, but there is still infinitely many planes that contain this line.

3. (i) Show that if two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are parallel, then

$$\mathrm{proj}_{\mathbf{u}}\mathbf{v} = \mathbf{v}.$$

Solution: If $\mathbf{u}$ and $\mathbf{v}$ are parallel, that means that $k\mathbf{u} = \mathbf{v}$ for some scalar k. Therefore

$$\begin{aligned}
\mathrm{proj}_{\mathbf{u}}(\mathbf{v}) &= \frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u} \text{ by definition of proj} \\
&= k\frac{\mathbf{u} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u} \text{ since } k\mathbf{u} = \mathbf{v} \\
&= k\mathbf{u} \text{ cancelling} \\
&= \mathbf{v} \text{ since } k\mathbf{u} = \mathbf{v}
\end{aligned}$$

(ii) Show that if $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are orthogonal, then

$$\text{proj}_{\mathbf{u}}\mathbf{v} = \mathbf{0}.$$

Solution: If $\mathbf{u}$ and $\mathbf{v}$ are orthogonal then $\mathbf{u} \cdot \mathbf{v} = 0$, so

$$\text{proj}_{\mathbf{u}}(\mathbf{v}) = \frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u} = \frac{0}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u} = 0\mathbf{u} = \mathbf{0}.$$

(iii) Show that for any vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$,

$$||\text{proj}_{\mathbf{u}}\mathbf{v}|| \leq ||\mathbf{v}||.$$

Solution 1: We use the Cauchy-Schwartz inequality in the key step.

$$\begin{aligned}
||\text{proj}_{\mathbf{u}}(v)|| &= ||\frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u}|| \text{ definition of proj} \\
&= |\frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{u} \cdot \mathbf{u}}|||\mathbf{u}|| \text{ since } \frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{u} \cdot \mathbf{u}} \text{ is a scalar.} \\
&= \frac{|\mathbf{u} \cdot \mathbf{v}|}{|\mathbf{u} \cdot \mathbf{u}|}||\mathbf{u}|| \\
&= \frac{|\mathbf{u} \cdot \mathbf{v}|}{||\mathbf{u}||^2}||\mathbf{u}|| \\
&= \frac{|\mathbf{u} \cdot \mathbf{v}|}{||\mathbf{u}||} \\
&\leq \frac{||\mathbf{u}||||\mathbf{v}||}{||\mathbf{u}||} \text{ Cauchy-schwartz} \\
&= ||\mathbf{v}||
\end{aligned}$$

Solution 2: $||\text{proj}_{\mathbf{u}}(v)||^2 = (\frac{\mathbf{u \cdot v}}{\mathbf{u \cdot u}}\mathbf{u}) \cdot (\frac{\mathbf{u \cdot v}}{\mathbf{u \cdot u}}\mathbf{u}) = \frac{(\mathbf{u \cdot v})^2}{\mathbf{u \cdot u}} = \frac{||\mathbf{u}||^2||\mathbf{v}||^2 \cos^2\theta}{||\mathbf{u}||^2} = ||\mathbf{v}||^2 \cos^2\theta \leq ||\mathbf{v}||^2$

4. Show that for any two vectors, $\mathbf{u} = [u_1\, u_2\, u_3]$ and $\mathbf{v} = [v_1\, v_2\, v_3]$, the vector

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

is a solution of the matrix equation

$$\begin{bmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{bmatrix} \mathbf{x} = \mathbf{0}.$$

Solution: Note that by definition of matrix-vector multiplication,

$$\begin{bmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{u} \cdot \mathbf{x} \\ \mathbf{v} \cdot \mathbf{x} \end{bmatrix}.$$

By properties of the cross-product, we know that $(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{u} = 0$ and $(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{v} = 0$. Thus

$$\begin{bmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{bmatrix} (\mathbf{u} \times \mathbf{v}) = \begin{bmatrix} \mathbf{u} \cdot (\mathbf{u} \times \mathbf{v}) \\ \mathbf{v} \cdot (\mathbf{u} \times \mathbf{v}) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

so $\mathbf{u} \times \mathbf{v}$ is a solution.

5. Three points, $P_0, P_1, P_2 \in \mathbb{R}^3$ are called *collinear* if they all lie on the same line.

(i) Are the points $(0, 0, 0)$, $(1, 0, 0)$, $(-10, 0, 0)$ collinear? *Solution:* Yes

(ii) Are the points $(0, -1, 4)$, $(0, -1, 7)$, $(0, -1, 10)$ collinear? *Solution:* Yes

(iii) Are the points $(2, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$ collinear? *Solution:* No

(iv) Show that the following two statements are equivalent:

- $P_0 = (x_0, y_0, z_0), P_1 = (x_1, y_1, z_1), P_2 = (x_2, y_2, z_2)$ are collinear.

- There is a number $t \in \mathbb{R}$ such that

$$\begin{aligned} x_2 &= tx_1 + (1 - t)x_0 \\ y_2 &= ty_1 + (1 - t)y_0 \\ z_2 &= tz_1 + (1 - t)z_0 \end{aligned}$$

Solution: P_0, P_1, P_2 collinear means that P_2 lies on the line determined by P_0 and P_1. This means that P_2 is contained in the line

$$L = \left\{ P_0 + t\overrightarrow{P_0P_1} : t \in \mathbb{R} \right\}.$$

$P_2 \in L$ means there is a number $t \in \mathbb{R}$ so that (using vector notation)

$$P_2 = P_0 + t\overrightarrow{P_0P_1} = \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} + t \begin{bmatrix} x_1 - x_0 \\ y_1 - y_0 \\ z_1 - z_0 \end{bmatrix} = \begin{bmatrix} tx_1 + (1 - t)x_0 \\ ty_1 + (1 - t)y_0 \\ tz_1 + (1 - t)z_0 \end{bmatrix}.$$

(v) Use part (d) to show that if P_0, P_1, P_2 are collinear, then the matrix

$$\begin{bmatrix} x_0 & y_0 & z_0 \\ x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{bmatrix}$$

has rank ≤ 2.

Solution: By part (d), we know that if P_0, P_1, P_2 are collinear, then there is a number $t \in \mathbb{R}$ so that

$$\begin{aligned} x_2 &= tx_1 + (1-t)x_0 \\ y_2 &= ty_1 + (1-t)y_0 \\ z_2 &= tz_1 + (1-t)z_0 \end{aligned}$$

Thus

$$\begin{bmatrix} x_0 & y_0 & z_0 \\ x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{bmatrix} = \begin{bmatrix} x_0 & y_0 & z_0 \\ x_1 & y_1 & z_1 \\ tx_1 + (1-t)x_0 & ty_1 + (1-t)y_0 & tz_1 + (1-t)z_0 \end{bmatrix}$$

Thus if we perform the row operations "subtract t times R_2 to R_3" and "subtract $1-t$ times R_1 to R_3", then the result will be

$$\begin{bmatrix} x_0 & y_0 & z_0 \\ x_1 & y_1 & z_1 \\ 0 & 0 & 0 \end{bmatrix}$$

so the matrix must have rank ≤ 2.

This fact can be useful for checking whether three points are collinear, since it is often easy to compute the rank of a matrix.

6. Consider the two vectors

$$\mathbf{u} = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

(i) Show that $\mathbf{u}$ and $\mathbf{v}$ are not parallel.

Solution: If $k \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$, we have that simultaneously k must be 0 and $1/2$, so there can be no such k. Therefore these vectors are not parallel.

(ii) Write the equation of the plane through the origin determined by the vectors $\mathbf{u}$ and $\mathbf{v}$ (use the cross-product to find a normal vector).

Solution: $\mathbf{u} \times \mathbf{v} = \begin{bmatrix} 2 \cdot 1 - (-1) \cdot 1 \\ (-1) \cdot 0 - 1 \cdot 1 \\ 1 \cdot 1 - 2 \cdot 0 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix}$. Therefore the plane going through the origin determined by these two vectors is

$$(\mathbf{u} \times \mathbf{v}) \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 3x - y + z = 0.$$

(iii) Show that the plane through the origin determined by the vectors $\mathbf{u}$ and $\mathbf{v}$ equals the set

$$\operatorname{span}\{\mathbf{u}, \mathbf{v}\}.$$

This has two steps:

- Show that any linear combination $t\mathbf{u} + s\mathbf{v}$ is a solution to the equation from part (b).
- Show that any point that solves the equation from part (b) can be written as a linear combination $t\mathbf{u} + s\mathbf{v}$.

Solution: Let's use the notation

$$\mathcal{P} = \left\{(x, y, z) \in \mathbb{R}^3 \colon 3x - y + z = 0\right\}.$$

- Any linear combination $t\mathbf{u} + s\mathbf{v}$ is a solution to the equation from part (b), since

$$(\mathbf{u} \times \mathbf{v}) \cdot (t\mathbf{u} + s\mathbf{v}) = s(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{u} + t(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{v} = s0 + t0 = 0.$$

  Thus we know that $\operatorname{span}\{\mathbf{u}, \mathbf{v}\} \subseteq \mathcal{P}$. (note that this argument works for any plane through the origin since we just used properties of dot and cross products).

7. Let $Q_1 = (1, -1)$ and $Q_2 = (-1, 1)$. Describe the set of points $P = (x, y)$ such that

$$d(P, Q_1) = d(P, Q_2).$$

Solution: For $P = (x, y)$, we have

$$d(P, Q_1) = \sqrt{(x-1)^2 + (y+1)^2}$$

$$d(P, Q_2) = \sqrt{(x+1)^2 + (y-1)^2}$$

Thus the condition

$$d(P, Q_1) = d(P, Q_2)$$

simplifies to

$$(x-1)^2 + (y+1)^2 = (x+1)^2 + (y-1)^2$$

Expanding both sides gives

$$x^2 - 2x + 1 + y^2 + 2y + 1 = x^2 + 2x + 1 + y^2 - 2y + 1$$

and this simplifies to

$$x = y.$$

Thus the set of all points is the line

$$L = \{(x, y) \in \mathbb{R}^2 \colon x = y\}.$$

8. (bonus) The area of a parallelogram in $\mathbb{R}^2$ with adjacent sides $\mathbf{u}$ and $\mathbf{v}$ can be computed with the formula

$$Area = ||\mathbf{u}||||\mathbf{v}|| \sin(\theta)$$

where θ is the angle between $\mathbf{u}$ and $\mathbf{v}$.

Use this formula and other properties of angles and dot products to show that if $\mathbf{u} = \begin{bmatrix} a \\ c \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} b \\ d \end{bmatrix}$, then

$$Area = ad - bc.$$

Solution: We know that the area of the triangle satisfies

$$Area^2 = ||\mathbf{u}||^2||\mathbf{v}||^2 \sin^2(\theta) = ||\mathbf{u}||^2||\mathbf{v}||^2(1 - \cos^2(\theta)) = ||\mathbf{u}||^2||\mathbf{v}||^2 - (\mathbf{u} \cdot \mathbf{v})^2.$$

Plugging in $\mathbf{u} = \begin{bmatrix} a \\ c \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} b \\ d \end{bmatrix}$, we get

$$\begin{aligned} Area^2 &= ||\mathbf{u}||^2||\mathbf{v}||^2 - (\mathbf{u} \cdot \mathbf{v})^2 \\ &= (a^2 + c^2)(b^2 + d^2) - (ab + cd)^2 \\ &= a^2b^2 + a^2d^2 + c^2b^2 + c^2d^2 - a^2b^2 - 2abcd - c^2d^2 \\ &= a^2d^2 - 2abcd + b^2c^2 \\ &= (ad - bc)^2 \end{aligned}$$

Thus

$$Area = ad - bc.$$

Tutorial 6: Linear transformations

Background: Chapter 6 of the lecture notes.

1. Suppose that

$$\mathbf{u} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \mathbf{w} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

Let T be the linear transformation such that

$$T(\mathbf{u}) = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, T(\mathbf{v}) = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, T(\mathbf{w}) = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

(i) Use properties of linear transformations to find the matrix of T.

Solution: First, after some thought, we see that

$$\begin{aligned}
\mathbf{e}_1 &= \mathbf{w} - \frac{1}{2}\mathbf{u} + \frac{1}{2}\mathbf{v} \\
\mathbf{e}_2 &= \frac{1}{2}\mathbf{u} - \frac{1}{2}\mathbf{v} \\
\mathbf{e}_3 &= \frac{1}{2}\mathbf{v} - \mathbf{w} + \frac{1}{2}\mathbf{u}
\end{aligned}$$

By properties of linear transformations, we know that

$$\begin{aligned}
T(\mathbf{e}_1) &= T(\mathbf{w} - \frac{1}{2}\mathbf{u} + \frac{1}{2}\mathbf{v}) \\
&= T(\mathbf{w}) - \frac{1}{2}T(\mathbf{u}) + \frac{1}{2}T(\mathbf{v}) \\
&= \begin{bmatrix} -1/2 \\ 1/2 \\ 1 \end{bmatrix} \\
T(\mathbf{e}_2) &= T(\frac{1}{2}\mathbf{u} - \frac{1}{2}\mathbf{v}) \\
&= \frac{1}{2}T(\mathbf{u}) - \frac{1}{2}T(\mathbf{v}) \\
&= \begin{bmatrix} 1/2 \\ -1/2 \\ 0 \end{bmatrix} \\
T(\mathbf{e}_3) &= T(\frac{1}{2}\mathbf{v} - \mathbf{w} + \frac{1}{2}\mathbf{u}) \\
&= \frac{1}{2}T(\mathbf{v}) - T(\mathbf{w}) + \frac{1}{2}T(\mathbf{u}) \\
&= \begin{bmatrix} 1/2 \\ 1/2 \\ -1 \end{bmatrix}
\end{aligned}$$

Thus, the matrix of T is

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

(ii) This linear transformation is invertible. Use the matrix from part (a) to compute the matrix of T^{-1}.

Solution: Using matrix inversion algorithm to compute the inverse of the matrix from part (a).

(iii) There is another (quicker) way to find the matrix of T^{-1} without using the matrix from part (a). What is it?

Solution: The inverse linear transformation T^{-1} satisfies $T^{-1}(T(\mathbf{x})) = \mathbf{x}$. Thus from what we are given, we know that

$$T^{-1}(e_1) = \mathbf{u},\ T^{-1}(e_2) = \mathbf{v},\ T^{-1}(e_3) = \mathbf{w}.$$

Thus the matrix of T^{-1} is

$$[\mathbf{u}|\mathbf{v}|\mathbf{w}] = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

3. For each item below, give an example of a linear transformation with that property or explain why no linear transformation with that property exists.

(i) T is injective and surjective.

Solution: $T(x, y) = (x, y)$ is both injective and surjective.

(ii) T is injective but not surjective.

Solution: $T\colon \mathbb{R}^2 \to \mathbb{R}^3$ given by $T(x, y) = (x, y, 0)$ is injective but not surjective.

(iii) T is surjective but not injective.

Solution: $T\colon \mathbb{R}^3 \to \mathbb{R}^2$ given by $T(x, y, z) = (x, y)$ is surjective but not injective.

(iv) $T\colon \mathbb{R}^3 \to \mathbb{R}^3$ is invertible.

Solution: $T\colon \mathbb{R}^3 \to \mathbb{R}^3$ given by $T(x, y, z) = (x, y, z)$ is invertible.

(v) $T\colon \mathbb{R}^3 \to \mathbb{R}^3$ is invertible but not injective.

Solution: This is impossible by Theorem in the lecture notes.

(vi) $T\colon \mathbb{R}^3 \to \mathbb{R}^3$ is injective but not surjective.

Solution: This is impossible. By Theorem , since T is injective, it is invertible. By Theorem , since T is invertible, it is surjective.

(vii) For every nonzero vector $\mathbf{x}$, $T(\mathbf{x}) \neq \mathbf{x}$ and $T \circ T(\mathbf{x}) = \mathbf{x}$.

Solution: $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by $T(x, y) = (-x, -y)$.

(viii) $T\colon \mathbb{R}^3 \to \mathbb{R}^3$ and the image of T is $\operatorname{span}\{[1\,5\,2]^T\}$.

Solution: $T\colon \mathbb{R}^3 \to \mathbb{R}^3$ given by orthogonal projection onto the line spanned by $[1\,5\,2]^T$.

(ix) $T\colon \mathbb{R} \to \mathbb{R}^3$ and the image of T is $\operatorname{span}\{[1\,5\,2]^T\}$.

Solution: Take T_A where

$$A = \begin{bmatrix} 1 \\ 5 \\ 2 \end{bmatrix}.$$

Solution:

(x) $T\colon \mathbb{R}^2 \to \mathbb{R}^3$ and the image of T is $\operatorname{span}\{[1\,5\,2]^T\}$.

Solution: Take T_A where

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

4. Given a column vector $\mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \in \mathbb{R}^n$.

We know that $\mathbf{u}^T\mathbf{u}$ is a number, and $\mathbf{u}\mathbf{u}^T$ is a $n \times n$ matrix.

(i) Use the definition of the orthogonal projection $\mathrm{proj}_{\mathbf{u}}\mathbf{x}$ to show that the matrix of the linear transformation $T(\mathbf{x}) = \mathrm{proj}_{\mathbf{u}}\mathbf{x}$ is given by the formula

$$A = \frac{1}{\mathbf{u}^T\mathbf{u}}\mathbf{u}\mathbf{u}^T.$$

Solution: We simply need to check that for all $\mathbf{x}$,

$$\mathrm{proj}_{\mathbf{u}}\mathbf{x} = \frac{1}{\mathbf{u}^T\mathbf{u}}\mathbf{u}\mathbf{u}^T\mathbf{x}.$$

This is true since

$$\begin{aligned}
\mathrm{proj}_{\mathbf{u}}\mathbf{x} &= \frac{\mathbf{u}\cdot\mathbf{x}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u} \text{ by definition of orthogonal projection} \\
&= \mathbf{u}\left(\frac{\mathbf{u}^T\mathbf{x}}{\mathbf{u}^T\mathbf{u}}\right) \text{ properties of dot-product} \\
&= \frac{\mathbf{u}(\mathbf{u}^T\mathbf{x})}{\mathbf{u}^T\mathbf{u}} \text{ rearranging} \\
&= \frac{(\mathbf{u}\mathbf{u}^T)\mathbf{x}}{\mathbf{u}^T\mathbf{u}} \text{ associativity of matrix mult.} \\
&= \frac{1}{\mathbf{u}^T\mathbf{u}}\mathbf{u}\mathbf{u}^T\mathbf{x} \text{ rearranging again.}
\end{aligned}$$

(ii) Use the result from part (a) to compute the matrix of the orthogonal projection $\mathrm{proj}_{\mathbf{u}}\colon \mathbb{R}^3 \to \mathbb{R}^3$ for the vector $\mathbf{u} = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$.

Solution: The matrix

$$A = \frac{1}{\mathbf{u}^T\mathbf{u}}\mathbf{u}\mathbf{u}^T = \frac{1}{6}\begin{bmatrix} 1 & -2 & -1 \\ -2 & 4 & 2 \\ -1 & 2 & 1 \end{bmatrix}.$$

(iii) For the orthogonal projection in part (b), write the sets $\ker(\mathrm{proj}_{\mathbf{u}})$ and $\mathrm{im}(\mathrm{proj}_{\mathbf{u}})$ as the span of a set of vectors.

Solution:

- The image of orthogonal projection onto a line is the line itself, so

$$\mathrm{image}(T) = \mathrm{span}\left\{\begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}\right\}.$$

- The kernel of orthogonal projection on to the line equals the set of solutions to the equation $A\mathbf{x} = \mathbf{0}$. We can find a spanning set for this set by computing the basic solutions of this homogeneous system, which gives

$$\ker(T) = \operatorname{span}\left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

  Another way to understand this solution is that we have found a spanning set for the plane through O with normal vector $\mathbf{u}$.

(iv) Use the result from part (a), and properties of matrix algebra, to show that the matrix A of orthogonal projection onto a line is always purple (Remember, A is purple if $A = A^T$).

Solution: Using properties of matrix algebra, we check that

$$\begin{aligned} A^T &= \left(\frac{1}{\mathbf{u}^T\mathbf{u}} \mathbf{u}\mathbf{u}^T \right)^T \\ &= \frac{1}{\mathbf{u}^T\mathbf{u}} \left(\mathbf{u}\mathbf{u}^T \right)^T \text{ since } \frac{1}{\mathbf{u}^T\mathbf{u}} \text{ is a scalar} \\ &= \frac{1}{\mathbf{u}^T\mathbf{u}} (\mathbf{u}^T)^T \mathbf{u}^T \text{ properties of matrix transpose} \\ &= \frac{1}{\mathbf{u}^T\mathbf{u}} \mathbf{u}\mathbf{u}^T \text{ properties of matrix transpose} \\ &= A. \end{aligned}$$

5. A linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^n$ is called *covfefe* if for every vector $\mathbf{x} \in \mathbb{R}^n$,

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

(i) Which of the following linear transformations $\mathbb{R}^2 \to \mathbb{R}^2$ are covfefe?

i. Rotation by θ degrees is covfefe.

ii. Reflection through a coordinate axis is covfefe.

iii. Scaling (dilation) by a factor of k is only covfefe if $k = \pm 1$.

iv. The shear transformation with matrix

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

is only covfefe if $a = 0$.

v. Orthogonal projection onto a line through the origin is never covfefe.

(ii) Describe in words what a covfefe transformation does geometrically.

Solution: A covfefe transformation preserves the lengths of vectors (alterantively, one could say that a covfefe transformation preserves the distance of points from the origin).

(iii) Recall that a square matrix A is called orange if $AA^T = A^T A = I$. Show that if A is orange, then the linear transformation $T_A(\mathbf{x}) = A\mathbf{x}$ is covfefe.

Solution: If A is orange, then we check

$$\begin{aligned}
||T_A(\mathbf{x})||^2 &= ||A\mathbf{x}||^2 \text{ by definition of } T_A \\
&= (A\mathbf{x}) \cdot (A\mathbf{x}) \text{ properties of vector length} \\
&= (A\mathbf{x})^T (A\mathbf{x}) \text{ properties of dot product} \\
&= \mathbf{x}^T A^T A\mathbf{x} \text{ properties of matrix transpose} \\
&= \mathbf{x}^T \mathbf{x} \text{ since } A \text{ is orange} \\
&= ||\mathbf{x}||^2
\end{aligned}$$

Thus $||T_A(\mathbf{x})|| = ||\mathbf{x}||$.

(iv) Show that if a linear transformation T is covfefe, then $\ker(T) = \{0\}$.

Solution: In order to show equality of two sets, we must show that $\ker(T) \subseteq \{0\}$ and $\{0\} \subseteq \ker(T)$. The second inclusion is true for any linear transformation, so we only need to check the first one.

To see that $\ker(T) \subseteq \{0\}$, we use the definition of $\ker(T)$.

If $\mathbf{x} \in \ker(T)$, then $T(\mathbf{x}) = \mathbf{0}$.

Since T is covfefe, this implies that

$$||\mathbf{x}|| = ||T(\mathbf{x})|| = ||\mathbf{0}|| = 0.$$

But if $||\mathbf{x}|| = 0$, then $\mathbf{x} = \mathbf{0}$.

(v) Use the result of part (d) and the Theorems from section 6.9 of the lecture notes to explain why every covfefe linear transformation is invertible.

Solution: By part (d), we know that T is injective. By Theorem 6.23, this tells us that T is invertible.

(vi) Use the result of part (e) and the Theorems from section 6.9 of the lecture notes to explain why every covfefe linear transformation is surjective.

Solution: By part (e) we know that T is invertible. By Theorem 6.24 in the lecture notes, this tells us that T is surjective.

Tutorial 7:

Background: Chapter 7 of the lecture notes.

1. Let

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

and

$$\mathbf{u} = \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \mathbf{w} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

(i) Compare the subspaces $S = \text{span}\{\mathbf{u}\}$ and $S' = \text{span}\{E\mathbf{u}\}$. *Solution:* They are different.

(ii) Compare the subspaces $S = \text{span}\{\mathbf{v}\}$ and $S' = \text{span}\{E\mathbf{v}\}$. *Solution:* They are the same.

(iii) Compare the subspaces $S = \text{span}\{\mathbf{u}, \mathbf{v}\}$ and $S' = \text{span}\{E\mathbf{u}, E\mathbf{v}\}$.

Solution: They are different. Indeed,

$$S = \text{span}\{\mathbf{u}, \mathbf{v}\} = \text{span}\left\{ \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \right\}$$

and

$$S' = \operatorname{span}\{E\mathbf{u}, E\mathbf{v}\} = \operatorname{span}\left\{ \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \right\}.$$

The vector $\begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} \in S'$, but it is not contained in S since

$$\begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} = s \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix} + t \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 2s+t \\ 2t \\ t-s \end{bmatrix}$$

has no solutions. (first, t must be 0, but then the third and first entries are the equations $-1 = 2s$ and $2 = -s$, which has no solutions).

Since S' contains a vector that is not in S, the two sets are different.

(iv) Compare the subspaces $S = \operatorname{span}\{\mathbf{u}, \mathbf{w}\}$ and $S' = \operatorname{span}\{E\mathbf{u}, E\mathbf{w}\}$.

Solution: They are the same. Indeed,

$$S = \operatorname{span}\{\mathbf{u}, \mathbf{w}\} = \operatorname{span}\left\{ \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\}$$

$$S' = \operatorname{span}\{E\mathbf{u}, E\mathbf{w}\} = \operatorname{span}\left\{ \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

If you look carefully at both of these spans, you will see that both of them equal the subspace

$$\left\{ \begin{bmatrix} x \\ 0 \\ z \end{bmatrix} \in \mathbb{R}^3 : x, z \in \mathbb{R} \right\}$$

In other words, the sets $\{\mathbf{u}, \mathbf{w}\}$ and $\{E\mathbf{u}, E\mathbf{w}\}$ are two different bases for the same subspace.

One way to describe what is happening here is that the transformation given by E, $T_E : \mathbb{R}^3 \to \mathbb{R}^3$ preserves the subspace above, since for any vector in this subspace,

$$T_E\left(\begin{bmatrix} x \\ 0 \\ z \end{bmatrix}\right) = E\begin{bmatrix} x \\ 0 \\ z \end{bmatrix} = \begin{bmatrix} z \\ 0 \\ x \end{bmatrix}$$

which is still in the subspace.

(v) Compare the subspaces $S = \mathrm{span}\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ and $S' = \mathrm{span}\{E\mathbf{u}, E\mathbf{v}, E\mathbf{w}\}$. *Solution:* They are the same. They are both 3 dimensional spaces in $\mathbb{R}^3$, so they both must be $\mathbb{R}^3$.

2. Which of the following sets are subspaces of $\mathbb{R}^2$? Justify your answer. Draw a picture of what the set looks like.

(i) A line is not a subspace unless it contains the origin.

(ii) A point is not a subspace unless it is the origin.

(iii) A circle of radius 2 centred at $(0, 2)$ is not a subspace.

(iv) $\mathbb{R}^2$ is a subspace.

(v) The set of points equidistant from $(1, 1)$ and $(-1, -1)$ is a subspace (it is the line $y = -x$).

(vi) The set of points $\{(x, y) \in \mathbb{R}^2 : y = x^2\}$ is not a subspace (it fails conditions II) and III).

(vii) The set of points $\{(x, y) \in \mathbb{R}^2 : x \geq 0\}$ is not a subspace. It fails condition III).

(viii) The set of points $\{(x, y) \in \mathbb{R}^2 : x^2 \geq 0\}$ equals $\mathbb{R}^2$, which is a subspace.

3. True or false.

(i) If S is a subspace of $\mathbb{R}^n$, then $\dim(S) < n$. *Solution:* False. S could be $\mathbb{R}^n$ in which case $\dim(S) = n$. Otherwise $\dim(S) < n$.

(ii) If S is a subspace of $\mathbb{R}^n$, then $\dim(S) > 0$. *Solution:* False. S could be the zero subspace, $\{\mathbf{0}\}$, in which case $\dim(S) = 0$. Otherwise $\dim(S) > 0$.

(iii) If $\{\mathbf{u}, \mathbf{v}\}$ is a set of two vectors in $\mathbb{R}^3$, then there is a vector $\mathbf{w}$ so that $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is a basis for $\mathbb{R}^3$. *Solution:* False. It is only true if $\{\mathbf{u}, \mathbf{v}\}$ is linearly independent.

(iv) If $\{\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{x}\}$ is a set of vectors in $\mathbb{R}^3$, then it is linearly dependent. *Solution:* True. The maximum size of a linearly independent subset of $\mathbb{R}^3$ is 3.

(v) If $\mathbf{0}$ is not contained in a set of vectors $\mathcal{U}$, then $\mathcal{U}$ is linearly independent. *Solution:* False. For example $\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}-1\\0\end{bmatrix}\}$ is linearly dependent.

(vi) If S and S' are subspaces of $\mathbb{R}^n$ with $\dim(S) = \dim(S')$, then $S = S'$. *Solution:* . False. For example, any two lines through the origin in $\mathbb{R}^2$ have dimension 1.

4. Suppose S is a set of points in $\mathbb{R}^n$ and $T\colon \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation.

Define the set

$$S' = \{T(\mathbf{x})\colon\ \mathbf{x} \in S\} \subseteq \mathbb{R}^m.$$

We call S' the *image of S under T*.

(i) Let S be the line through the origin in $\mathbb{R}^2$ spanned by $\mathbf{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

i. If T is the linear transformation "rotate by an angle of $\pi/2$ radians clockwise" then what is the image of S under T?

Solution: The image of S under T is the line span$\{\begin{bmatrix}1\\-2\end{bmatrix}\}$.

ii. If T_A is the linear transformation corresponding to the matrix

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

then what is the image of S under T?

Solution: Since $A\begin{bmatrix}2\\1\end{bmatrix} = 0$, so the image of S under T_A is span$\{A\begin{bmatrix}1\\2\end{bmatrix}\} = \{\mathbf{0}\}$.

(ii) Show using the definition of subspace that if S is a subspace of $\mathbb{R}^n$, then the image S' of S under a linear transformation T is a subspace of $\mathbb{R}^m$. *Solution:* We check each condition of being a subspace.

I) Since S is a subspace, $\mathbf{0} \in S$. Since T is a linear transformation, $T(\mathbf{0}) = \mathbf{0}$. Thus $\mathbf{0} \in S'$.

II) If $\mathbf{u}, \mathbf{v} \in S'$, then $\mathbf{u} = T(\mathbf{x})$ and $\mathbf{v} = T(\mathbf{y})$ for $\mathbf{x}, \mathbf{y} \in S$.

Since S is a subspace, $\mathbf{x} + \mathbf{y} \in S$, so $T(\mathbf{x} + \mathbf{y}) \in S'$. Finally, since T is a linear transformation,

$$T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y}) = \mathbf{u} + \mathbf{v}.$$

Thus $\mathbf{u} + \mathbf{v} \in S'$.

III) Suppose $\mathbf{u} = T(\mathbf{x}) \in S'$ and $k \in \mathbb{R}$.

Since S is a subspace, $k\mathbf{x} \in S$. Thus $T(k\mathbf{x}) \in S'$. Finally, since T is a linear transformation,

$$T(k\mathbf{x}) = kT(\mathbf{x}).$$

Thus $kT(\mathbf{x}) \in S'$.

Solution 2: We can use the subspace test.

If $\mathbf{y}, \mathbf{y}'$ are vectors in S', and $k \in \mathbb{R}$, then there are vectors $\mathbf{x}, \mathbf{x}'$ in S so that $T(\mathbf{x}) = \mathbf{y}$, and $T(\mathbf{x}') = \mathbf{y}'$. Since S is a subspace, $\mathbf{x} + k\mathbf{x}' \in S$, so therefore $\mathbf{y} + k\mathbf{y}' = T(\mathbf{x} + k\mathbf{x}')$, so $\mathbf{y} + k\mathbf{y}' \in S'$, showing that S' is a subspace.

(iii) Show that if $S = \text{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$, then $S' = \text{span}\{T(\mathbf{u}_1), \ldots, T(\mathbf{u}_k)\}$.

Solution: Suppose that $\mathbf{y} \in S'$. Then $\mathbf{y} = T(\mathbf{x})$ for some $\mathbf{x} \in S$. Since $S = \text{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\ \}$, there are scalars $t_1, \ldots, t_k$, so that $\mathbf{x} = t_1\mathbf{u}_1 + \ldots + t_k\mathbf{u}_k$. Therefore $\mathbf{y} = T(\mathbf{x}) = T(t_1\mathbf{u}_1 + \ldots + t_k\mathbf{u}_k) = t_1T(\mathbf{u}_1) + \ldots + t_kT(\mathbf{u}_k)$.

(iv) Show that if S is a subspace of $\mathbb{R}^n$, then $\dim(S') \leq \dim(S)$.

Solution: If $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\ \}$ is a basis for S, then by the result of the previous item, $\{T(\mathbf{u})_1, \ldots, T(\mathbf{u}_k)\ \}$ contains a basis for S'. Therefore $\dim(S) = k \geq \dim(S')$.

5. In this question we will finish a detail from the proof of the algorithm for finding a basis of the span of a set of vectors.

Show that if

- $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is a set of linearly independent vectors in $\mathbb{R}^n$, and
- A is an invertible $n \times n$ matrix,

then the set

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

is linearly independent.

Solution: To show that $\{A\mathbf{u}_1, \ldots, A\mathbf{u}_k\}$ is linearly independent, we show that the only solution to the equation

$$x_1(A\mathbf{u}_1) + \cdots + x_k(A\mathbf{u}_k) = 0$$

is $x_1, \ldots, x_k = 0$.

By matrix algebra,

$$x_1(A\mathbf{u}_1) + \cdots + x_k(A\mathbf{u}_k) = A(x_1\mathbf{u}_1 + \cdots + x_k\mathbf{u}_k).$$

Since A is invertible, if $A(x_1\mathbf{u}_1 + \cdots + x_k\mathbf{u}_k) = 0$, then $x_1\mathbf{u}_1 + \cdots + x_k\mathbf{u}_k = \mathbf{0}$.

Finally, since $\{A\mathbf{u}_1, \ldots, A\mathbf{u}_k\}$ is linearly independent, if $x_1\mathbf{u}_1 + \cdots + x_k\mathbf{u}_k = \mathbf{0}$, then $x_1, \ldots, x_k = 0$.

Tutorial 8

Background: Chapter 8 of the lecture notes.

1. Let

$$A = \begin{bmatrix} 1 & -2 & 0 & 0 & 3 \\ 2 & -5 & -3 & -2 & 6 \\ 0 & 5 & 15 & 10 & 0 \\ 2 & 6 & 18 & 8 & 6 \end{bmatrix}$$

The reduced row echelon form of A is

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

(i) What is the rank of A? 3

(ii) Find a basis for null(A). What is the nullity of A?

null(A) $= 2$

(iii) Find a basis for col(A). What is the dimension of the column space of A?

$\dim(\text{col}(A)) = 3$

(iv) Find a basis for row(A). What is the dimension of the row space of A?

$\dim(\text{row}(A)) = 3$

(v) Check the rank-nullity theorem for A.

$2 + 3 = 5$

(vi) Write a basis for the column and row space of A^T.

(hint: the rows of A are the columns of A^T and vis versa)

(vii) What is the rank of A^T? Why?

$\text{rank}(A^T) = \text{rank}(A) = 3$

(viii) What is the dimension of the null space of A^T?

$\dim(\text{null}(A^T) = 4 - \text{rank}(A^T) = 4 - 3 = 1.$

2. In this question we will explore examples and theory related to Theorem 8.9 in the lecture notes.

(i) Are the statements in Theorem 8.9 true for the matrix

$$A = \begin{bmatrix} 1 & -2 \\ 2 & -5 \\ 0 & 5 \\ 2 & 6 \end{bmatrix}?$$

Solution: Yes

(ii) Are the statements in Theorem 8.9 true for the matrix

$$A = \begin{bmatrix} 1 & -2 & 1 \\ 2 & -5 & 0 \end{bmatrix}?$$

Solution: No

(iii) Are the statements in Theorem 8.9 true for the matrix

$$A = \begin{bmatrix} 1 & 7 \\ 2 & -5 \end{bmatrix}?$$

Solution: Yes. But this matrix is also invertible so the statements in Theorem 8.8 are also true.

(iv) Is it possible for any of the statements in Theorem 8.9 to be true for a matrix with fewer columns than rows ($m < n$)? Why?

Solution: No. If $m < n$, then $\operatorname{rank}(A) \leq m < n$. Thus $\operatorname{rank}(A) = n$ is impossible.

(v) Use the result from part (d) to explain the following fact.

Fact: If A is a $m \times n$ matrix and $m < n$, then the matrix equation

$$XA = I_n$$

has no solutions (i.e. there is no $n \times m$ matrix X that solves this equation). In other words, A cannot have a left inverse.

Solution: By part (d), all of the equivalent statements in Theorem 8.9 are false for A.

(vi) Give an example of a 4×3 matrix A such that every statement in Theorem 8.9 is false for A.

Solution:

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

(vii) Give an example of a linear transformation $T\colon \mathbb{R}^3 \to \mathbb{R}^4$ that is not injective.

Solution: Just take T_A, where A is the matrix from the solution to part (f).

4. In this question we will explore how matrix multiplication effects the null space and rank of a matrix.

(i) Consider the matrices

$$A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, \; B = \begin{bmatrix} 1 & -2 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

Compute a basis for $\mathrm{null}(B)$ and $\mathrm{null}(AB)$. Use the result of your computation to explain that

$$\mathrm{null}(B) = \mathrm{null}(AB).$$

(ii) Consider the matrices

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \; B = \begin{bmatrix} 1 & -2 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

Compute a basis for $\mathrm{null}(B)$ and $\mathrm{null}(AB)$. Use the result of your computation to explain that

$$\mathrm{null}(B) \subseteq \mathrm{null}(AB).$$

but the two subspaces are not equal.

(iii) Show that, in general, if A is a $m \times n$ matrix and B is a $n \times p$ matrix, then

$$\text{null}(B) \subseteq \text{null}(AB).$$

Proof. Suppose that $\mathbf{v} \in \text{null}(B)$ (this means that $B\mathbf{v} = \mathbf{0}$).

Then $AB\mathbf{v} = A(B\mathbf{v}) = A\mathbf{0} = \mathbf{0}$.

Therefore $\mathbf{v} \in \text{null}(AB)$.

Since we have shown that every element of $\text{null}(B)$ is an element of $\text{null}(AB)$, we have shown that

$$\text{null}(B) \subseteq \text{null}(AB).$$

□

(iv) (bonus) Use part (c) and the rank-nullity theorem to conclude the following theorem:

Theorem: If A is a $m \times n$ matrix and B is a $n \times p$ matrix, then

$$\text{rank}(B) \geq \text{rank}(AB).$$

Proof. The rank-nullity theorem for the $m \times p$ matrix AB tells us that

$$p = \text{rank}(AB) + \text{nullity}(AB).$$

The rank-nullity theorem for the $n \times p$ matrix B tells us that

$$p = \text{rank}(B) + \text{nullity}(B).$$

Combining these two equations gives us the equation

$$\text{rank}(AB) + \text{nullity}(AB) = \text{rank}(B) + \text{nullity}(B).$$

By part (c), we know that $\text{null}(B) \subseteq \text{null}(AB)$, so $\text{nullity}(B) \leq \text{nullity}(AB)$.

Therefore since

$$\text{rank}(AB) + \text{nullity}(AB) = \text{rank}(B) + \text{nullity}(B)$$

and $\text{nullity}(B) \leq \text{nullity}(AB)$, we know that

$$\text{rank}(B) \geq \text{rank}(AB).$$

□

(v) There is a similar but slightly different fact:

Theorem: If C is a $n \times m$ matrix and D is a $p \times n$ matrix, then

$$\text{rank}(D) \geq \text{rank}(DC).$$

Solution:

Proof. We can use part (d) and the magic of matrix transposes. We see that

$$\begin{aligned}
\text{rank}(DC) &= \text{rank}(C^T D^T) \text{ since transpose does not change rank} \\
&\leq \text{rank}(D^T) \text{ by part (d)} \\
&= \text{rank}(D) \text{ since transpose does not change rank}
\end{aligned}$$

□

(vi) In part (a), we saw an example where A was invertible and we observed that $\text{null}(B) = \text{null}(AB)$.

In fact, this is true in general:

Theorem: If A is an invertible $n \times n$ matrix and B is a $n \times p$ matrix, then

$$\text{null}(B) = \text{null}(AB).$$

Prove this theorem.

Proof 1: To show that the two sets are equal, we show two things. First, by part (c), we already know that $\text{null}(B) \subseteq \text{null}(AB)$.

Second, we can show that if A is invertible, then $\text{null}(AB) \subseteq \text{null}(B)$.

Indeed, if A is invertible, and $\mathbf{x} \in \text{null}(AB)$, then

$$B\mathbf{x} = (A^{-1}A)B\mathbf{x} = A^{-1}(AB\mathbf{x}) = A^{-1}(\mathbf{0}) = \mathbf{0}.$$

Thus, $\mathbf{x} \in \text{null}(B)$.

finally, since we have shown that $\text{null}(B) \subseteq \text{null}(AB)$ and $\text{null}(AB) \subseteq \text{null}(B)$, we know that

$$\text{null}(B) = \text{null}(AB).$$

Proof 2: By part (c), we know that $\text{null}(B) \subseteq \text{null}(AB)$.

Thus, in order to show that $\text{null}(B) = \text{null}(AB)$, it is sufficient to explain why they have the same dimension (since by Theorem 7.12(g), if $S \subseteq S'$ and $\dim(S) = \dim(S')$, $S = S'$).

Since A is invertible, we know that $\text{rank}(AB) = \text{rank}(B)$ (this is true since A is a product of elementary matrices). By the rank-nullity theorem, this tells us that

$$\dim(\text{null}(AB)) = n - \text{rank}(AB) = n - \text{rank}(B) = \dim(\text{null}(B)).$$

This completes the proof.

C.3 Tutorial 9-11

Tutorial 9: Determinants

Background: Chapter 9 of the lecture notes.

1. (warm-up) Use cofactor expansion to compute the determinant of the following matrices.

(i) $\det(A) = 3$

(ii) $\det(B) = -28$

(iii) $\det(C) = -84$

(iv) In general, $\det\left(\begin{bmatrix} A & 0_{m\times n} \\ D & B \end{bmatrix}\right) = \det(A)\det(B)$.

2. Find the determinant of the following matrices by inspection (don't use cofactor expansion, instead use properties of determinants and to quickly "see" the determinant without computing).

$$\text{(i)}\ \det \begin{bmatrix} 4 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 8 \end{bmatrix} = 4(-1)(2)(2)(8).$$

$$\text{(ii)}\ \det \begin{bmatrix} 4 & 3 & 43243 & 5 & 121321 \\ 0 & 2 & 2 & 1000 & 10 \\ 0 & 0 & 2 & 82 & 9392234 \\ 0 & 0 & 0 & 3 & 65 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix} = (4)(2)(2)(3)(-1)$$

$$\text{(iii)}\ \det \begin{bmatrix} 1 & 0 & 4 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & -2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 6 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} = 6$$

$$\text{(iv)}\ \det \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} = (-1)(-1)(1) = 1$$

$$\text{(v)}\ \det \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 100 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} = (-1)(100) = -100$$

$$\text{(vi)}\ \det \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 100 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} = (-1)(0)(100) = 0$$

3. Suppose that A and B are 3×3 matrices with $\det(A) = -1$ and $\det(B) = 2$. Compute each of the following determinants using properties of the determinant.

(i) $\det(2A^2) = 2^3(-1)(-1)$

(ii) $\det(A^3B^5) = (-1)^3(2)^5$

(iii) $\det(BABABAB^2) = (-1)^3(2)^5$

(iv) $\det(3A^2B^{-1}) = 3^3(-1)^2(1/2)$

(v) $\det(3(A^2B^{-1})^T) = \det(3A^2B^{-1}) = 3^3(-1)^2(1/2)$

4. Consider the matrix
$$A = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 0 & 3 \end{bmatrix}.$$

(i)
$$\det(AA^T) = 49$$

(ii)
$$\det(A^TA) = 0$$

(iii) The problem with the fake proof is the step $\det(AA^T) = \det(A)\det(A^T)$. Since A is not a square matrix, the expressions $\det(A)$ and $\det(A^T)$ aren't defined, so this step is nonsense.

Tutorial 10: Eigenvalues, eignevectors, and diagonalization

Background: Chapter 10 of the lecture notes.

1. Compute the eigenvalues and corresponding eigenspaces for the matrices below. Which of the matrices are diagonalizable? For each matrix that is diagonalizable, compute its diagonalization.

(i)
$$\begin{bmatrix} 7 & 4 \\ -3 & -1 \end{bmatrix}$$

Eigenvalues: 1,5.

$$E_1 = \operatorname{span}\left\{ \begin{bmatrix} -2 \\ 3 \end{bmatrix} \right\}$$
$$E_5 = \operatorname{span}\left\{ \begin{bmatrix} -2 \\ 1 \end{bmatrix} \right\}$$

There are two distinct eigenvalues, so this matrix is diagonalizable. One possible diagonalization is

$$\begin{bmatrix} 7 & 4 \\ -3 & -1 \end{bmatrix} = \begin{bmatrix} -2 & -2 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} 1/4 & 1/2 \\ -3/4 & -1/2 \end{bmatrix}$$

(ii)

$$\begin{bmatrix} 4 & -2 \\ -3 & 9 \end{bmatrix}$$

Eigenvalues: $3, 10$.

$$E_3 = \operatorname{span}\left\{ \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right\}$$

$$E_{10} = \operatorname{span}\left\{ \begin{bmatrix} -1 \\ 3 \end{bmatrix} \right\}$$

There are two distinct eigenvalues, so this matrix is diagonalizable.

$$\begin{bmatrix} 7 & 4 \\ -3 & -1 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 10 \end{bmatrix} \begin{bmatrix} \frac{3}{7} & \frac{1}{7} \\ \frac{-1}{7} & \frac{2}{7} \end{bmatrix}$$

(iii)

$$\begin{bmatrix} 4 & 2 & 3 \\ -1 & 1 & -3 \\ 2 & 4 & 9 \end{bmatrix}$$

Eigenvalues: 3 (algebraic multiplicity 2), 8 (algebraic multiplicity 1).

$$E_3 = \operatorname{span}\left\{ \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \right\}$$

$$E_8 = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} \right\}$$

$\dim(E_3) + \dim(E_8) = 2 + 1 = 3$, so this matrix is diagonalizable. One possible diagonalization is

$$\begin{bmatrix} 4 & 2 & 3 \\ -1 & 1 & -3 \\ 2 & 4 & 9 \end{bmatrix} = \begin{bmatrix} -3 & -2 & 1 \\ 0 & 1 & -1 \\ 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 8 \end{bmatrix} \begin{bmatrix} \frac{-2}{5} & \frac{-4}{5} & \frac{-5}{5} \\ \frac{1}{5} & \frac{7}{5} & \frac{3}{5} \\ \frac{1}{5} & \frac{2}{5} & \frac{3}{5} \end{bmatrix}$$

(iv)

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

Eigenvalues: 4 (algebraic multiplicity 2), 1 (algebraic multiplicity 1).

$$E_4 = \operatorname{span}\left\{\begin{bmatrix}0\\1\\0\end{bmatrix}\right\}$$

$$E_1 = \operatorname{span}\left\{\begin{bmatrix}0\\1\\1\end{bmatrix}\right\}$$

$\dim(E_4)+\dim(E_1) = 1+1 = 2 < 3$, so this matrix is not diagonalizable.

2. In this question we explore an application of matrix diagonalization: computing large powers of a matrix.

(i) If $A = PDP^{-1}$ then using associativity of matrix multiplication, we see that

$$\begin{aligned}A^k &= (PDP^{-1})^k = (PDP^{-1})(PDP^{-1})\cdots(PDP^{-1})(PDP^{-1})\\ &= PD(P^{-1}P)D(P^{-1}P)\cdots(P^{-1}P)D(P^{-1}P)DP^{-1} = PD^kP^{-1}\end{aligned}$$

(ii) For each of the matrices from question 1 that are diagonalizable, use formula from part (a) to compute A^{2017} (you don't need to simplify powers of numbers that appear in your final answer).

i.

$$\begin{aligned}\begin{bmatrix}7 & 4\\-3 & -1\end{bmatrix}^{2017} &= \begin{bmatrix}-2 & -2\\3 & 1\end{bmatrix}\begin{bmatrix}1 & 0\\0 & 5\end{bmatrix}^{2017}\begin{bmatrix}1/4 & 1/2\\-3/4 & -1/2\end{bmatrix}\\ &= \begin{bmatrix}-2 & -2\\3 & 1\end{bmatrix}\begin{bmatrix}1 & 0\\0 & 5^{2017}\end{bmatrix}\begin{bmatrix}1/4 & 1/2\\-3/4 & -1/2\end{bmatrix}\\ &= \frac{1}{4}\begin{bmatrix}-2 & -2\cdot 5^{2017}\\-6 & 5^{2017}\end{bmatrix}\begin{bmatrix}1 & 2\\-3 & -1\end{bmatrix}\\ &= \frac{1}{4}\begin{bmatrix}-2+6\cdot 5^{2017} & -4+2\cdot 5^{2017}\\-6-3\cdot 5^{2017} & -12-5^{2017}\end{bmatrix}\end{aligned}$$

ii.

$$\begin{aligned}\begin{bmatrix}7 & 4\\-3 & -1\end{bmatrix}^{2017} &= \begin{bmatrix}2 & -1\\1 & 3\end{bmatrix}\begin{bmatrix}3^{2017} & 0\\0 & 10^{2017}\end{bmatrix}\begin{bmatrix}\frac{3}{7} & \frac{1}{7}\\\frac{-1}{7} & \frac{2}{7}\end{bmatrix}\\ &= \frac{1}{7}\begin{bmatrix}6\cdot 3^{2017}+10^{2017} & 2\cdot 3^{2017}-2\cdot 10^{2017}\\3^{2018}-3\cdot 10^{2017} & 3^{2017}+6\cdot 10^{2017}\end{bmatrix}\end{aligned}$$

3. Suppose that A and B are 3×3 matrices. Which of the following statements are true or false, and why?

(i) If -2 is an eigenvalue of A, then -8 is an eigenvalue of A^3. True.

(ii) If -2 is an eigenvalue of A, then -8 is an eigenvalue of $2A^3$. False.

(iii) If -2 is an eigenvalue of A and -2 is an eigenvalue of B, then 4 is an eigenvalue of AB. False.

(iv) If -2 is an eigenvalue of A and -2 is an eigenvalue of B, then -2 is an eigenvalue of $A + B$. False.

(v) If $A^2 = I$, then the only eigenvalues of A are 1 or -1. True.

(vi) If $A^2 = A$, then the only eigenvalues of A are 1. False.

(vii) If A has two distinct eigenvalues, then A is diagonalizable. False.

(viii) If A is invertible, then A is diagonalizable. False.

4. In this question we will add another fact to our list of equivalent statements for $n \times n$ matrices.

Theorem Suppose that A is a $n \times n$ matrix. The following two statements are equivalent.

(i) A is invertible.

(ii) 0 is not an eigenvalue of A.

Hint: use the characteristic polynomial, and the list of statements equivalent to (a).

Proof. First, by the big theorem for square matrices, we know that "A is invertible" is equivalent to the statement "$\det(A) \neq 0$".

Thus, A is invertible if and only if

$$\det(A - 0I) = \det(A) \neq 0.$$

This equation means that 0 is not a root of the characteristic polynomial of A, so A is invertible if and only if 0 is not an eigenvalue of A. □

5. Suppose $\mathbf{u} \in \mathbb{R}^n$ is nonzero. Orthogonal projection onto the line spanned by $\mathbf{u}$ is the linear transformation $T(\mathbf{x}) = \text{proj}_{\mathbf{u}}(\mathbf{x})$.

Recall that we showed in Tutorial 6, question 4 that the matrix of this linear transformation is given by the formula

$$A = \frac{1}{\mathbf{u}^T\mathbf{u}}\mathbf{u}\mathbf{u}^T.$$

(i) Show that 1 and 0 are eigenvalues of A by using the definition of eigenvalues (find some eigenvectors for these eigenvalues). *Hint: this question may be easier if you think about the geometric properties of* $\mathit{proj}_{\mathbf{u}}$.

Solution: The way to show that 0 and 1 are eigenvalues for A is to use the definition and find eigenvectors.

First, observe that

$$A\mathbf{u} = \left(\frac{1}{\mathbf{u}^T\mathbf{u}}\mathbf{u}\mathbf{u}^T\right)\mathbf{u} = \frac{1}{\mathbf{u}^T\mathbf{u}}\mathbf{u}(\mathbf{u}^T\mathbf{u}) = \mathbf{u} = 1\mathbf{u}$$

so $\lambda = 1$ is an eigenvalue of A, and $\mathbf{u}$ is an eigenvector for this eigenvalue.

Second, if $\mathbf{v}$ is nonzero and orthogonal to $\mathbf{u}$, then

$$A\mathbf{v} = \left(\frac{1}{\mathbf{u}^T\mathbf{u}}\mathbf{u}\mathbf{u}^T\right)\mathbf{v} = \frac{1}{\mathbf{u}^T\mathbf{u}}\mathbf{u}(\mathbf{u}^T\mathbf{v}) = \mathbf{0} = 0\mathbf{v}$$

so $\lambda = 0$ is an eigenvalue of A, and $\mathbf{v}$ is an eigenvector for this eigenvalue.

Tutorial 11: Orthogonality

Background: Chapter 11 of the lecture notes.

1. Let

$$S = \operatorname{span}\left\{\begin{bmatrix}1\\-2\\0\\0\\3\end{bmatrix}, \begin{bmatrix}0\\2\\1\\0\\1\end{bmatrix}, \begin{bmatrix}1\\0\\4\\0\\0\end{bmatrix}, \begin{bmatrix}1\\0\\1\\0\\4\end{bmatrix}\right\} \subseteq \mathbb{R}^5$$

(i) Use Gram-Schmidt to compute an orthogonal basis for the subspace S. *Final answer:* Note that the fourth vector in the set above is a linear combination of the previous vectors in the set, so when performing

Gram-Schmidt, the vector $\mathbf{s}_4$ will compute to be $\mathbf{0}$ and we throw it out.

One way to write the final answer is:

$$(1,-2,0,0,3),(1,26,14,0,17),(73,-94,275,0,-87).$$

(ii) Compute a basis for $S^\perp$ using the fact that $\mathrm{row}(A)^\perp = \mathrm{null}(A)$. Then use Gram-Schmidt to compute an orthogonal basis for $S^\perp$.
After row-reducing, we see that the basic solutions are

$$(-32,-7,8,0,6),(0,0,0,1,0).$$

This is already an orthogonal basis.

(iii) Use the computations from parts (a) and (b) to confirm the formula

$$\dim(S)+\dim(S^\perp)=n$$

for this example.

We check,

$$\dim(S)+\dim(S^\perp)=3+2=5.$$

2. Let

$$\mathbf{v}=\begin{bmatrix}2\\3\\-1\\1\\4\end{bmatrix}.$$

(i) Use the result of Question 1.(a) to compute the orthogonal projection $\mathrm{proj}_S(\mathbf{v})$.

$$\begin{aligned}\mathrm{proj}_S(\mathbf{v}) &= \frac{8}{14}\begin{bmatrix}1\\-2\\0\\0\\3\end{bmatrix}+\frac{2+78-14+4\cdot 17}{1162}\begin{bmatrix}1\\26\\14\\0\\17\end{bmatrix}+\frac{2\cdot 73-3\cdot 94-275-4\cdot 87}{97359}\begin{bmatrix}73\\-94\\275\\0\\-87\end{bmatrix}\\ &= \frac{1}{17}\begin{bmatrix}2\\44\\-9\\0\\74\end{bmatrix}\end{aligned}$$

(I realize the numbers in this computation get a little out of hand. We promise not to put a problem where you have to multiply numbers this large on the final)

(ii) Use the formula

$$\text{proj}_{S^\perp}(\mathbf{v}) = \mathbf{v} - \text{proj}_S(\mathbf{v})$$

and the result from 2.(a) to compute $\text{proj}_{S^\perp}(\mathbf{v})$.

$$\text{proj}_{S^\perp}(\mathbf{v}) = \begin{bmatrix} 2 \\ 3 \\ -1 \\ 1 \\ 4 \end{bmatrix} - \frac{1}{17}\begin{bmatrix} 2 \\ 44 \\ -9 \\ 0 \\ 74 \end{bmatrix} = \frac{1}{17}\begin{bmatrix} 32 \\ 7 \\ -8 \\ 1 \\ -6 \end{bmatrix}$$

(iii) Check that the vectors computed in 2.(a) and 2.(b) are orthogonal.

$$\frac{1}{17}\begin{bmatrix} 2 \\ 44 \\ -9 \\ 0 \\ 74 \end{bmatrix} \cdot \frac{1}{17}\begin{bmatrix} 32 \\ 7 \\ -8 \\ 1 \\ -6 \end{bmatrix} = \frac{1}{17^2}(2\cdot 32+44\cdot 7+(-9)\cdot(-8)+0+74\cdot(-6)) = 0$$

(iv) Use the orthonormal basis for $S^\perp$ from question 1.(b) to compute $\text{proj}_{S^\perp}(\mathbf{v})$ by using the definition of orthogonal projection onto the subspace $S^\perp$ (you should get the same answer as in 2.(b)).

$$\text{proj}_{S^\perp}(\mathbf{v}) = \frac{2(-32)+3(-7)+(-1)8+4(6)}{1173}\begin{bmatrix} -32 \\ -7 \\ 8 \\ 0 \\ 6 \end{bmatrix} + \frac{1}{1}\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \frac{1}{17}\begin{bmatrix} 32 \\ 7 \\ -8 \\ 1 \\ -6 \end{bmatrix}$$

3. Suppose that S is a subspace of $\mathbb{R}^n$ with dimension k. For each statement below, say whether it is true or false. Can you explain why?

(i) $\dim(S^\perp) = n - k$.

True. This follows from the rank-nullity theorem (see the lecture notes).

(ii) If $\mathbf{x}$ is contained in both S and $S^{\perp}$, then $\mathbf{x} = \mathbf{0}$.

True. Since vectors in $S^{\perp}$ are orthogonal to vectors in S, if $\mathbf{x}$ is in both S and $S^{\perp}$, then it must be true that

$$\mathbf{x} \cdot \mathbf{x} = 0.$$

But the only vector with this property is $\mathbf{0}$, so $\mathbf{x} = \mathbf{0}$.

(iii) The orthogonal complement of $S^{\perp}$ is S. In other words, $(S^{\perp})^{\perp} = S$.

True.

(iv) If $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is an orthogonal set, then it is linearly independent.

False, it is possible that one of the vectors is zero.

(v) For any vector $\mathbf{x} \in \mathbb{R}^n$,

$$||\mathbf{x}||^2 = ||\text{proj}_S \mathbf{x}||^2 + ||\text{proj}_{S^{\perp}} \mathbf{x}||^2.$$

True by the Pythagorean Theorem, since $\text{proj}_S \mathbf{x}$ and $\text{proj}_{S^{\perp}} \mathbf{x}$ are orthogonal.

4. Recall from earlier tutorials that we called a $n \times n$ matrix A is orange[3] if $AA^T = A^T A = I$.

(i) Show that A is orange if and only if the columns of A are an orthonormal basis for $\mathbb{R}^n$. *Hint: Recall that the ij-entry of AB is the dot-product of the ith row of A and the jth column of B.*

Proof. Let $\mathbf{a}_1, \cdots, \mathbf{a}_n$ denote the columns of A.

Since the ith row of A^T equals $\mathbf{a}_i^T$, we see from the definition of matrix multiplication that the (i, j)-entry of $A^T A$ is $\mathbf{a}_i^T \mathbf{a}_j = \mathbf{a}_i \cdot \mathbf{a}_j$.

Thus the following statements are equivalent:

[3] In the lecture notes, we called this orthogonal.

- $A^T A = I$.
- When $i \neq j$, $\mathbf{a}_i \cdot \mathbf{a}_j = 0$, and for $1 \leq i \leq n$, $\mathbf{a}_i \cdot \mathbf{a}_j = 1$.

The second statement is precisely the statement that the set $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ is an orthonormal set. Since it is a set of n vectors in $\mathbb{R}^n$, this is equivalent to the statement that the set $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ is an orthonormal basis for $\mathbb{R}^n$. $\square$

(ii) Show that if $\mathbf{u}$ and $\mathbf{v}$ are orthogonal, and A is orange, then $A\mathbf{u}$ and $A\mathbf{v}$ are orthogonal.

Proof. If $\mathbf{u}$ and $\mathbf{v}$ are orthogonal, and A is orange, then

$$\begin{aligned} A\mathbf{u} \cdot A\mathbf{v} &= (A\mathbf{u})^T(A\mathbf{v}) \text{ by properties of dot-product} \\ &= \mathbf{u}^T A^T A\mathbf{v} \text{ by properties of matrix transpose} \\ &= \mathbf{u}^T I\mathbf{v} \text{ since } A \text{ is orange} \\ &= \mathbf{u}^T\mathbf{v} \\ &= 0 \text{ since } \mathbf{u} \text{ and } \mathbf{v} \text{ are orthogonal.} \end{aligned}$$

$\square$

(iii) Show that if $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthogonal set, then $\{A\mathbf{u}_1, \ldots, A\mathbf{u}_k\}$ is an orthogonal set.

Proof. $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthogonal set means that $\mathbf{u}_i$ and $\mathbf{u}_j$ are orthogonal for each $i \neq j$.

By part (b), we know that since A is orange, and $\mathbf{u}_i$ and $\mathbf{u}_j$ are orthogonal, then $A\mathbf{u}_i$ and $A\mathbf{u}_j$ are orthogonal for each $i \neq j$. This means (by definition), that $\{A\mathbf{u}_1, \ldots, A\mathbf{u}_k\}$ is an orthogonal set. $\square$

(iv) Recall from an earlier tutorial that we showed the following: If A is orthogonal, then for any vector $\mathbf{u}$, $||A\mathbf{u}|| = ||\mathbf{u}||$.

In case you forgot the proof, here it is:

Proof. If A is orthogonal and $\mathbf{u} \in \mathbb{R}^n$, then

$$\begin{aligned}
||A\mathbf{u}|| &= \sqrt{(A\mathbf{u}) \cdot (A\mathbf{u})} \text{ by properties of dot-product, length} \\
&= \sqrt{(A\mathbf{u})^T (A\mathbf{u})} \text{ by properties of dot-product} \\
&= \sqrt{\mathbf{u}^T A^T A\mathbf{u}} \text{ by properties of matrix transpose} \\
&= \sqrt{\mathbf{u}^T \mathbf{u}} \text{ since } A \text{ is orange} \\
&= ||\mathbf{u}|| \text{ by properties of dot-product, length.}
\end{aligned}$$

□

(v) Show that if $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthonormal set, then $\{A\mathbf{u}_1, \ldots, A\mathbf{u}_k\}$ is an orthonormal set.

Proof. We already showed in part (c) that if $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthogonal set and A is orange, then $\{A\mathbf{u}_1, \ldots, A\mathbf{u}_k\}$ is an orthogonal set.

An orthonormal set is an orthogonal set with every vector a unit vector.

It remains to show that if $||\mathbf{u}_i|| = 1$ for $1 \leq i \leq n$, then $||A\mathbf{u}_i|| = 1$ for $1 \leq i \leq n$.

But we know from part (d) that if A is orange, then $||A\mathbf{u}_i|| = ||\mathbf{u}_i|| = 1$ for $1 \leq i \leq n$. Thus $\{A\mathbf{u}_1, \ldots, A\mathbf{u}_k\}$ is an orthonormal set.

□

Index