Introduction to Linear Algebra

Adam Clay

Contents

List of Figures

Preface

I believe that university textbooks are too expensive.

Please report any errors to adamjosephclay@gmail.com. This text is designed around the MATH 133 curriculum at McGill University, though it is taught at a more introductory level, and I have attempted to use simple English throughout.

Chapter 1

Vectors and Geometry

1.1 Basic properties of vectors

In this section we will cover all the basic ideas needed to work with vectors. By the end of the section we'll have the necessary tools to tackle some more interesting geometric problems, so look to the next section for applications of the ideas learned here.

1.1.1 Adding and subtracting vectors

A vector $\mathbf{v}$ in two dimensions is an arrow in the plane that records a direction and a length. If $\mathbf{v}$ starts at the point A and ends at the point B, it's written as $\mathbf{v} = \overrightarrow{AB}$. The point A is called the *tail*, and B the *tip*. By convention, if the point A is the origin then we don't have to bother with writing 'A', and instead we just write the coordinates of the tip B inside of square brackets, e.g. $\mathbf{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$. The numbers in the vector are called *entries*, and we count them from top to bottom. So in the case of the vector $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$, 2 is the first entry and 3 is the second entry.

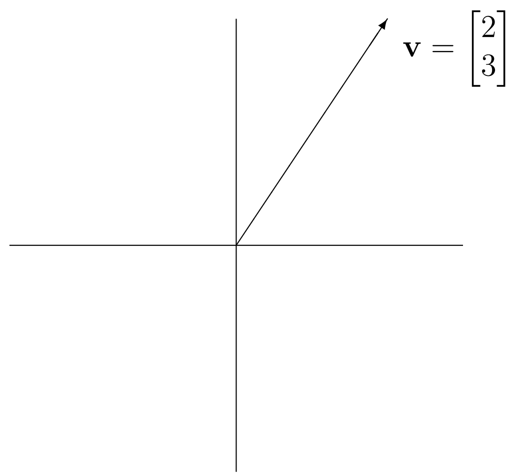

Figure 1.1: An example of a vector in two dimensions, here the tail is $(0, 0)$ and the tip is $(2, 3)$.

When vectors are written this way, they can be added to one another and subtracted from one another. The rules are:

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

and

$$\begin{bmatrix} a \\ b \end{bmatrix} - \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} a - c \\ b - d \end{bmatrix}.$$

In the language we've just introduced, these rules are best explained as: In order to add vectors you add their corresponding entries, in order to subtract vectors you subtract their corresponding entries.

The addition and subtraction of vectors both have good geometric interpretations. See the pictures below for an explanation of what the vectors $\mathbf{v} + \mathbf{w}$ and $\mathbf{v} - \mathbf{w}$ represent.

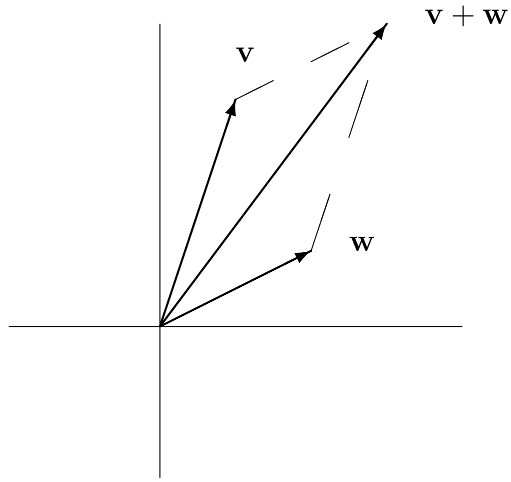

Figure 1.2: The sum of two vectors gives the direction of the diagonal of a parallelogram.

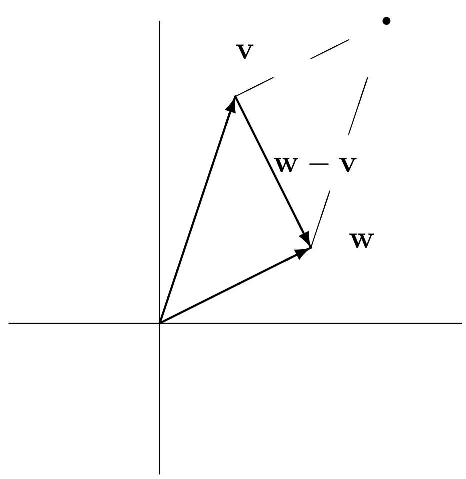

Figure 1.3: The difference of two vectors gives the direction of the other diagonal.

From the subtraction example, we can figure out more. If we have two points A and B in the plane, and we want to know the vector $\overrightarrow{AB}$ which gives the direction from point A to point B, then we can use vector subtraction to find $\overrightarrow{AB}$.

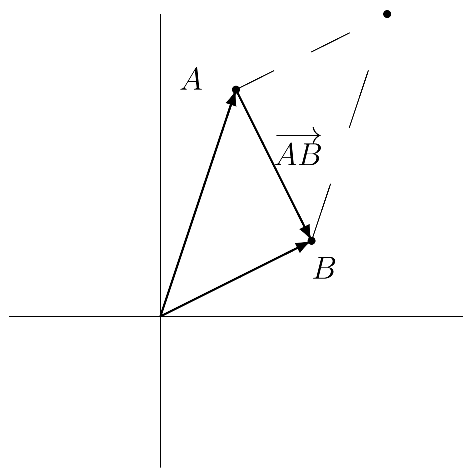

Figure 1.4: The vector from A to B gives the direction you must travel to get from point A to point B.

In the picture above, each of the points $A = (1, 3)$ and $B = (2, 1)$ has a corresponding vector $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ respectively, whose tail is at the origin and whose tip is A or B. Then you subtract to find that $\overrightarrow{AB} = \begin{bmatrix} 2-1 \\ 1-3 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$.

Example 1. Find the midpoint of the line segment between the points $A = (4, 5)$ and $B = (-2, 1)$.

Solution. Start every vector problem by drawing a picture, if you can. In this case we get:

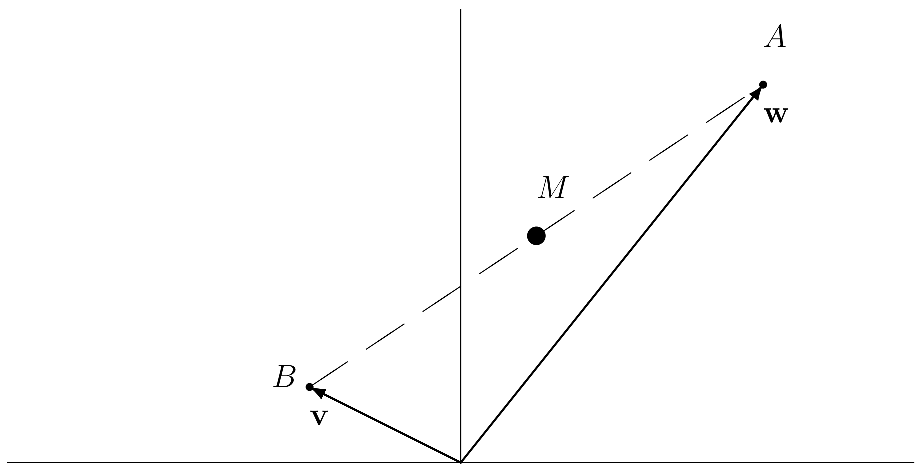

Figure 1.5: The points $A = (4, 5)$ and $B = (-2, 1)$ with corresponding vectors $\mathbf{v}$ and $\mathbf{w}$ respectively, and the line segement between them with midpoint M.

To get to the point M from the origin, we can travel first in the direction of $\mathbf{v}$ towards the point $A = (4, 5)$, then turn and travel in the direction of the vector $\overrightarrow{AB}$ for half of the distance to the point $B = (-2, 1)$. This instruction can be coded in equations as:

$$\begin{bmatrix}4\\5\end{bmatrix} + \frac{1}{2}\overrightarrow{AB} = \begin{bmatrix}4\\5\end{bmatrix} + \frac{1}{2}\begin{bmatrix}-2-4\\1-5\end{bmatrix} = \begin{bmatrix}4\\5\end{bmatrix} + \begin{bmatrix}-3\\-2\end{bmatrix} = \begin{bmatrix}1\\3\end{bmatrix}.$$

So, the vector $\begin{bmatrix}1\\3\end{bmatrix}$ gives us the coordinates of the point $M = (1, 3)$. ■

This distinction between points and vectors (with their tip at the given point) is subtle, because every point has a corresponding vector and every vector ends at a corresponding point. The importance of this distinction is that vectors can be added and subtracted, eventually we will see that they can be multiplied by matrices, etc. On the other hand points cannot have these operations done to them. In this way the distinction between points and vectors is similar to something you may see in computer programming, where you can have different data structures storing the same information. For example, the string `dog:cat:mouse` is very different than the list `(dog, cat, mouse)`, and to change the string into a list we would have to split the string at the colons. Yet there are many functions that we can perform on strings that we can't perform on lists (and vice versa), even though both contain the same information of dog, cat, mouse.

1.1.2 The length of a vector

Every vector also has a length. The length of a vector is written as $||\mathbf{v}||$, or in the case that we're given numbers, $\left\|\begin{matrix}2\\3\end{matrix}\right\|$. For vectors in two dimensions, we can calculate their length using the Pythagorean Theorem. The length of the vector $\begin{bmatrix}a\\b\end{bmatrix}$ is

$$\left\|\begin{matrix}a\\b\end{matrix}\right\| = \sqrt{a^2+b^2} = \sqrt{c^2} = c,$$

as the picture indicates.

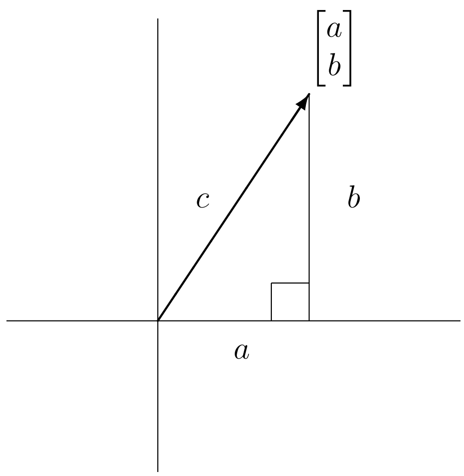

Figure 1.6: The length of a vector in two dimensions.

Besides adding vectors to one another and subtracting vectors from one another, we can also multiply a vector by a number. The formula is:

$$d\begin{bmatrix}a\\b\end{bmatrix} = \begin{bmatrix}da\\db\end{bmatrix},$$

so multiplying a vector by a number d is the same as multiplying all its entries by d. Look at how multiplying by d changes the length of a vector. The length before is

$$\left\|\begin{matrix}a\\b\end{matrix}\right\| = \sqrt{a^2+b^2},$$

the length after is

$$\left\|\begin{matrix}da\\db\end{matrix}\right\| = \sqrt{(da)^2+(db)^2} = \sqrt{d^2(a^2+b^2)} = \sqrt{d^2}\sqrt{a^2+b^2} = |d|\sqrt{a^2+b^2}.$$

We can see from our calculation that the length of a vector is changed by $|d|$ whenever you multiply the vector by d. This scaling works just like the figure below.

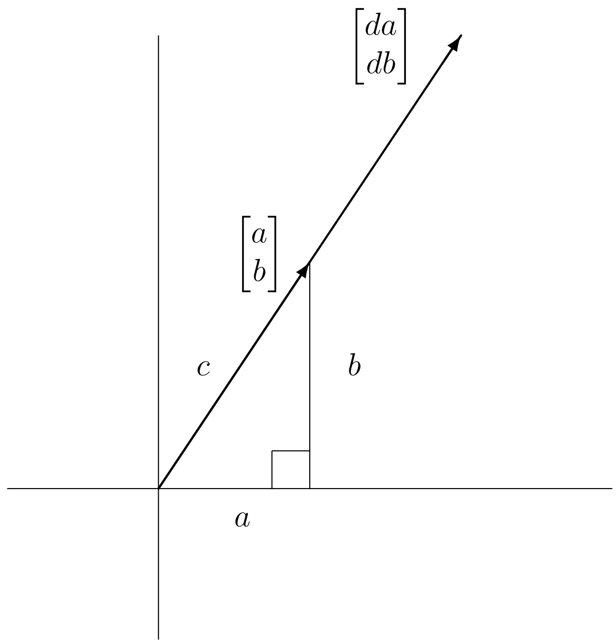

Figure 1.7: The length of a vector changing after scalar multiplication. Here, d looks to be about two, since the vector after scaling is about twice as long.

Multiplying a vector by a number is called *scalar multiplication*, because the number d is a *scalar*.

Example 2. Find a vector of length 1 in the same direction as $\mathbf{v} = \begin{bmatrix} 4 \\ 7 \end{bmatrix}$.

Solution. The length of the vector $\mathbf{v}$ is

$$||\mathbf{v}|| = \sqrt{4^2 + 7^2} = \sqrt{65},$$

so $\mathbf{v}$ is $\sqrt{65} \cong 8.062...$ times too long. To correct this, we multiply $\mathbf{v}$ by $\frac{1}{\sqrt{65}}$. This gives a new vector $\mathbf{w} = \frac{1}{\sqrt{65}}\mathbf{v}$, and the length of $\mathbf{w}$ is

$$||\mathbf{w}|| = ||\frac{1}{\sqrt{65}}\mathbf{v}|| = \frac{1}{\sqrt{65}}||\mathbf{v}|| = \frac{1}{\sqrt{65}}(\sqrt{65}) = 1.$$

Note that in our calculation, the step $||\frac{1}{\sqrt{65}}\mathbf{v}|| = \frac{1}{\sqrt{65}}||\mathbf{v}||$ uses what we've already learned: when you multiply a vector by a number d, the length of the vector is multiplied by $|d|$.

$\blacksquare$

1.1.3 The dot product

There is an operation that can be done on two vectors in order to figure out the angle between them, called the *dot product*. The dot product of two vectors $\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} c \\ d \end{bmatrix}$ is the number

$$\mathbf{v} \cdot \mathbf{w} = \begin{bmatrix} a \\ b \end{bmatrix} \cdot \begin{bmatrix} c \\ d \end{bmatrix} = ac + bd.$$

This number is related to the angle θ between the vectors $\mathbf{v}$ and $\mathbf{w}$ by the formula

$$\mathbf{v} \cdot \mathbf{w} = ||\mathbf{v}||||\mathbf{w}|| \cos(\theta).$$

The number θ is always between 0 and π, or possibly equal to 0 or to π.

You can also check from this formula that $\mathbf{v} \cdot \mathbf{v} = ||\mathbf{v}||^2$ for every vector $\mathbf{v}$, or you can check directly that $\mathbf{v} \cdot \mathbf{v} = ||\mathbf{v}||^2$ from the definition of the dot product.

Example 3. Show that the angle between the vectors $\begin{bmatrix} 2 \\ -5 \end{bmatrix}$ and $\begin{bmatrix} 5 \\ 2 \end{bmatrix}$ is $\pi/2$.

Solution. We can use the dot product to figure out the angle. We calculate that

$$\begin{bmatrix} 2 \\ -5 \end{bmatrix} \cdot \begin{bmatrix} 5 \\ 2 \end{bmatrix} = 2(5) + (-5)(2) = 0,$$

so that $0 = ||\mathbf{v}||||\mathbf{w}|| \cos(\theta)$. Since we have three numbers $||\mathbf{v}||$, $||\mathbf{w}||$, and $\cos(\theta)$ multiplying together to give zero, one of them must be equal to zero. Both $||\mathbf{v}||$ and $||\mathbf{w}||$ can't be zero since they're equal to the lengths of our vectors, so we must have $\cos(\theta) = 0$. This means that $\theta = \pi/2$. $\blacksquare$

This is actually a special case of a general fact: the angle between two vectors $||\mathbf{v}||$ and $||\mathbf{w}||$ is $\pi/2$ exactly when $\mathbf{v} \cdot \mathbf{w} = 0$. Vectors with an angle of $\pi/2$ between them are called *orthogonal*. So, from this calculation above we would conclude that the vectors $\begin{bmatrix} 2 \\ -5 \end{bmatrix}$ and $\begin{bmatrix} 5 \\ 2 \end{bmatrix}$ are orthogonal.

The dot product behaves a lot like multiplication of numbers, because we have the following formulas:

1. $\mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v}$, so the order of vectors in a dot product doesn't matter.
2. $\mathbf{v} \cdot \mathbf{0} = 0$, here $\mathbf{0}$ is the zero vector. So it's similar to when you multiply any number by zero and you get zero.
3. $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$, so it behaves like multiplying numbers since you can distribute $\mathbf{u}$ over the vectors inside the brackets.

1.1.4 Projection of one vector onto another

When you have two vectors $\mathbf{v}$ and $\mathbf{w}$, you can 'project one vector onto the other'. The result of projecting the vector $\mathbf{v}$ onto the vector $\mathbf{w}$ is a new vector that points in the same direction as $\mathbf{w}$, but it has a different length than $\mathbf{w}$. This new vector is denoted $\text{proj}_{\mathbf{w}}(\mathbf{v})$ and is nonzero as long as $\mathbf{w}$ and $\mathbf{v}$ are not orthogonal.

The best way of describing the relationship between $\mathbf{v}$, $\mathbf{w}$ and $\text{proj}_{\mathbf{w}}(\mathbf{v})$ is with the picture below.

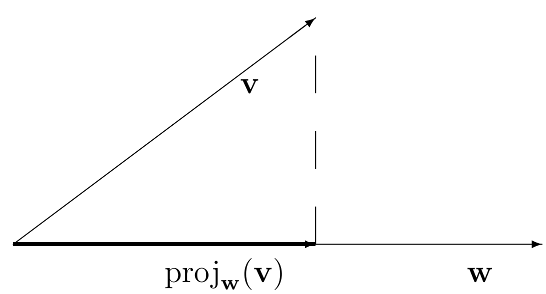

Figure 1.8: The projection of $\mathbf{v}$ onto $\mathbf{w}$ is often referred to as the 'shadow' of $\mathbf{v}$ on $\mathbf{w}$, as though there were a light shining from directly above $\mathbf{w}$.

The formula for projection is

$$\text{proj}_{\mathbf{w}}(\mathbf{v}) = \left(\frac{\mathbf{v} \cdot \mathbf{w}}{||\mathbf{w}||^2} \right) \mathbf{w}.$$

Observe that this formula has two parts, the number $\dfrac{\mathbf{v} \cdot \mathbf{w}}{||\mathbf{w}||^2}$ and the vector $\mathbf{w}$. This makes sense: from Figure 1.8 you would probably guess that $\text{proj}_{\mathbf{w}}(\mathbf{v})$ is going to be equal to $a\mathbf{w}$ for some appropriate scalar a, since both vectors are pointing in the same direction. It turns out that this is exactly the case, and the amount that you have to scale $\mathbf{w}$ in order to make it the right length is $a = \dfrac{\mathbf{v}\cdot\mathbf{w}}{||\mathbf{w}||^2}$. You can arrive at this formula for the number a by using the formula $\mathbf{v}\cdot\mathbf{w} = ||\mathbf{v}||||\mathbf{w}||\cos(\theta)$ above and doing a bit of trigonometry, if you are so inclined.

Example 4. Let $\mathbf{w} = \begin{bmatrix}-3\\1\end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix}3\\4\end{bmatrix}$. Calculate $\text{proj}_{\mathbf{w}}(\mathbf{v})$.

Solution. The solution to this problem is to simply apply the formula, but let's begin by drawing a picture anyway.

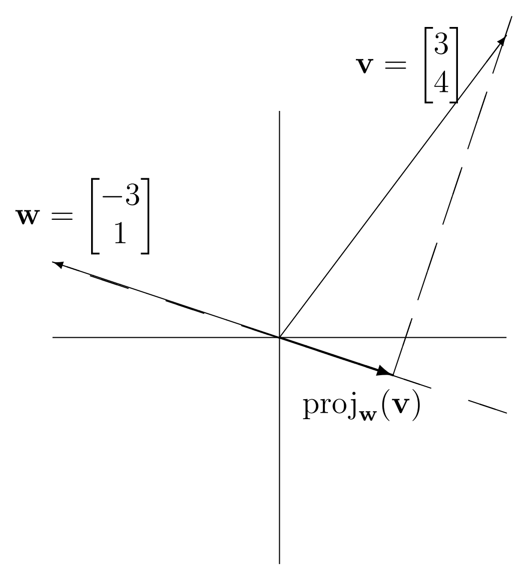

Figure 1.9: Projecting the vector $\mathbf{v}$ onto $\mathbf{w}$.

From our picture it looks like something different is happening than in the standard picture (by 'standard picture' I mean Figure 1.8). What's happening is that our scaling factor $\dfrac{\mathbf{v}\cdot\mathbf{w}}{||\mathbf{w}||^2}$ is going to be negative, so in this case the projection will actually point in the opposite direction of $\mathbf{w}$. Let us calculate now to check this claim:

$$\text{proj}_{\mathbf{w}}(\mathbf{v}) = \left(\frac{\begin{bmatrix}3\\4\end{bmatrix}\cdot\begin{bmatrix}-3\\1\end{bmatrix}}{\left\|\begin{matrix}-3\\1\end{matrix}\right\|^2}\right)\mathbf{w} = \frac{(3)(-3)+(1)(4)}{(-3)^2+1^2}\mathbf{w} = -\frac{1}{2}\mathbf{w}.$$

$\blacksquare$

1.1.5 Basics of vectors in three dimensions

All of the properties we have discussed so far extend naturally to three dimensional vectors. A vector in three dimensions is an arrow in space that indicates a direction. We add three dimensional vectors by adding the entries, and subtract by subtracting the entries. The addition and subtraction of 3-d vectors have the same geometric interpretation as the pictures before, but we would have a much harder time drawing the pictures now–since the pictures would have to be 3-d, but these pages are 2-d. The length of three dimensional vectors is calculated more or less the same way as two dimensional vectors:

$$\left|\left|\begin{bmatrix} a \\ b \\ c \end{bmatrix}\right|\right| = \sqrt{a^2 + b^2 + c^2}.$$

The reason this formula works again comes from the Pythagorean Theorem, but it's not as direct as before. The dot product also works the same way, as we'll see in this short example.

Example 5. Calculate the angle between the vectors $\begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$.

Solution. The dot product of these two vectors is

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

From the formula $\mathbf{v} \cdot \mathbf{w} = ||\mathbf{v}||||\mathbf{w}|| \cos(\theta)$, we solve for $\cos(\theta)$ and substitute:

$$\cos(\theta) = \frac{\mathbf{v} \cdot \mathbf{w}}{||\mathbf{v}||||\mathbf{w}||} = \frac{3}{\sqrt{6}\sqrt{6}} = \frac{1}{2}.$$

Therefore $\cos(\theta)$ comes from the triangle

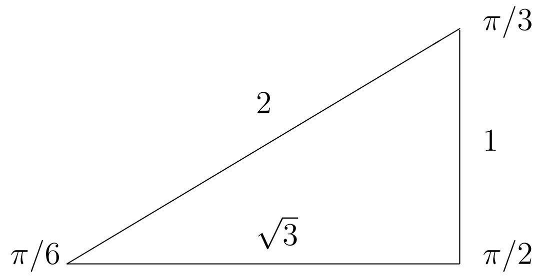

Figure 1.10: The triangle with angles $\pi/2$, $\pi/3$ and $\pi/6$.

So $\theta = \pi/3$. ■

Since the formula

$$\mathbf{v} \cdot \mathbf{w} = ||\mathbf{v}||||\mathbf{w}|| \cos(\theta)$$

works for vectors in three dimensions as well, we know that two 3-d vectors $\mathbf{v}$ and $\mathbf{w}$ have an angle of $\pi/2$ between them precisely when their dot product is zero. In this case $\mathbf{v}$ and $\mathbf{w}$ are said to be *orthogonal*.

Example 6. Calculate the projection of $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$ onto $\mathbf{w} = \begin{bmatrix} -1 \\ -1 \\ 3 \end{bmatrix}$.

Solution. The dot product of $\mathbf{v}$ and $\mathbf{w}$ is

$$\mathbf{v} \cdot \mathbf{w} = (1)(-1) + (1)(-1) + (2)(3) = 4.$$

The length of $\mathbf{w}$ squared is

$$||\mathbf{w}||^2 = (-1)^2 + (-1)^2 + 3^2 = 11.$$

So, the projection of $\mathbf{v}$ onto $\mathbf{w}$ is

$$\text{proj}_{\mathbf{w}}(\mathbf{v}) = \left(\frac{\mathbf{v} \cdot \mathbf{w}}{||\mathbf{w}||^2} \right) \mathbf{w} = \frac{4}{11} \begin{bmatrix} -1 \\ -1 \\ 3 \end{bmatrix}.$$

■

1.1.6 The cross product

Everything we have discussed so far–projection, dot product, addition, subtraction and scalar multiplication–works for both two and three dimensional vectors. The cross product is the first thing we'll discuss that only works for three dimensional vectors, and not two dimensional ones.

The purpose of the cross product of two 3-dimensional vectors $\mathbf{v}$ and $\mathbf{w}$ is to provide a new vector $\mathbf{v} \times \mathbf{w}$ that is orthogonal to both $\mathbf{v}$ and $\mathbf{w}$ and whose length is a special quantity. If $\mathbf{v} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} d \\ e \\ f \end{bmatrix}$, then the cross product formula is

$$\mathbf{v} \times \mathbf{w} = \begin{bmatrix} bf - ce \\ cd - fa \\ ae - bd \end{bmatrix}.$$

The way it is written here, this formula is hard to remember. Before the end of the book we will see two more formulas for $\mathbf{v} \times \mathbf{w}$ which are much easier to remember, but the other formulas require a knowledge of matrices and determinants. For that reason, we'll work with this formula for the time being.

A picture of what $\mathbf{v} \times \mathbf{w}$ represents is:

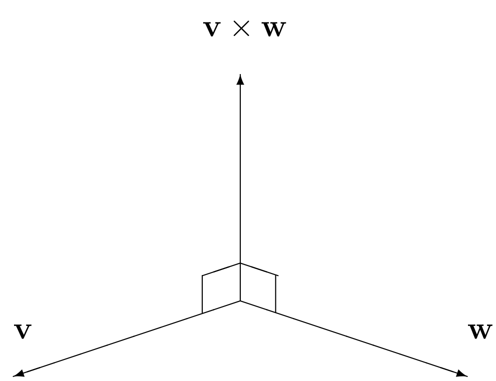

Figure 1.11: The cross product of two vectors.

Now observe that in the sentence before the formula for $\mathbf{v} \times \mathbf{w}$, and in the picture above, we're claiming that $\mathbf{v} \times \mathbf{w}$ is orthogonal to $\mathbf{v}$ and $\mathbf{w}$. This is certainly not obvious from the formula, but there is a way we can check to make sure that this is true.

Remember that two vectors are orthogonal exactly when their dot product is zero. So to check that the formula for $\mathbf{v} \times \mathbf{w}$ actually gives a vector that is orthogonal to $\mathbf{v}$ and $\mathbf{w}$, we can calculate the dot products:

$$\mathbf{v} \cdot (\mathbf{v} \times \mathbf{w}) = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \cdot \begin{bmatrix} bf - ce \\ cd - fa \\ ae - bd \end{bmatrix} = abf - ace + bcd - bfa + cae - cbd = 0,$$

$$\mathbf{w} \cdot (\mathbf{v} \times \mathbf{w}) = \begin{bmatrix} d \\ e \\ f \end{bmatrix} \cdot \begin{bmatrix} bf - ce \\ cd - fa \\ ae - bd \end{bmatrix} = dbf - dce + ecd - efa + fae - fbd = 0.$$

Miraculously, everything cancels just as we'd hoped, so the vectors are orthogonal.

The cross product has two important formulas that come with it. It is very reasonable to ask if there is any relationship between the cross product and the dot product. There is a relationship, and it comes in the form of the *Lagrange identity*, which is this famous formula:

$$||\mathbf{v} \times \mathbf{w}||^2 = ||\mathbf{v}||^2||\mathbf{w}||^2 - (\mathbf{v} \cdot \mathbf{w})^2.$$

Example 7. Calculate the cross product of two parallel vectors.

Solution. This example sounds impossible at first, until you call on the Lagrange identity. Saying that two vectors $\mathbf{v}$ and $\mathbf{w}$ are parallel means that the angle between them, θ, is zero. So in the Lagrange identity above, you can substitute

$$\mathbf{v} \cdot \mathbf{w} = ||\mathbf{v}||||\mathbf{w}|| \cos(\theta) = ||\mathbf{v}||||\mathbf{w}|| \cos(0) = ||\mathbf{v}||||\mathbf{w}||(1) = ||\mathbf{v}||||\mathbf{w}||.$$

With this substitution, the Lagrange identity changes into

$$||\mathbf{v} \times \mathbf{w}||^2 = ||\mathbf{v}||^2||\mathbf{w}||^2 - (||\mathbf{v}||||\mathbf{w}||)^2 = 0.$$

So, if $\mathbf{v}$ and $\mathbf{w}$ are parallel then $||\mathbf{v} \times \mathbf{w}||^2 = 0$, in other words the length of the vector $\mathbf{v} \times \mathbf{w}$ is zero. The only vector with length zero is $\mathbf{0}$, so $\mathbf{v} \times \mathbf{w} = \mathbf{0}$. In fact, vectors are parallel exactly when their cross product is zero. ■

We can also use the Lagrange identity to relate the length of the cross product to the angle θ between the two vectors. The relationship is

$$||\mathbf{v} \times \mathbf{w}|| = ||\mathbf{v}||||\mathbf{w}|| \sin(\theta).$$

You can get this formula from the Lagrange identity by replacing the dot product with $||\mathbf{v}||||\mathbf{w}|| \cos(\theta)$ and applying a trig identity (try it!).

1.1.7 The right-hand rule

Let us return to the picture of the cross product that was used in the last section, and point out a curious fact. First let's describe the picture carefully to ensure that we're all imagining it in 3-d in the same way. The vector **v** should be imagined as coming out of the page and pointing at your left shoulder, and the vector **w** should be imagined to be coming out of the page and pointing at your right shoulder. The vertical vectors are lying on the page.

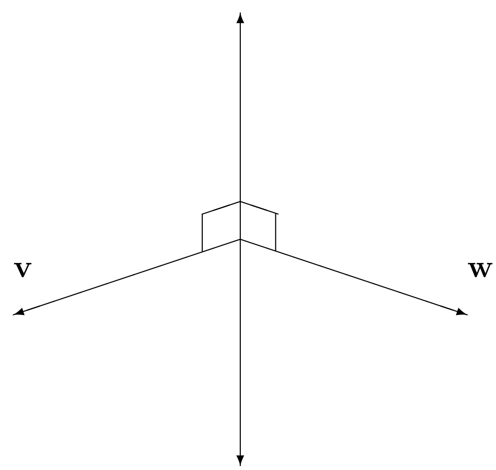

Figure 1.12: There are two choices for a vector orthogonal to **v** and **w**.

Now, imagining **v** and **w** as just described, if I asked you to give me a vector which is orthogonal to both **v** and **w**, which one would you choose? Should you choose the vector which points up the page, or the one that points down the page?

One way of choosing a vector that is orthogonal to both **v** and **w** is to use the cross product formula

$$\mathbf{v} \times \mathbf{w} = \begin{bmatrix} bf - ce \\ cd - fa \\ ae - bd \end{bmatrix}$$

to calculate a vector which is orthogonal to both **v** and **w**. But which vector will this formula give us, the one which points up the page, or the one which points down the page?

The answer is: The formula will give us the vector which points up along the page, not the one that points down. This fact is called 'the right-hand rule,' we say that the cross product 'obeys the right-hand rule.'

The reason it is called the right-hand rule is because an alternative way of picturing the cross product is as follows: Using your right hand, with the vector **v** pointing along your index finger, and **w** pointing along your middle finger, the vector $\mathbf{v} \times \mathbf{w}$ points along your thumb.

Example 8. Holding your arms straight out in front of you, suppose that your left arm is **v** and your right arm is **w**. Does the vector $\mathbf{v} \times \mathbf{w}$ point at the ceiling or the floor?

Solution. On the left is a picture of the cross product, which obeys the right hand rule. On the right is a picture of a person with their left arm labeled as **v** and their right arm labeled as **w**.

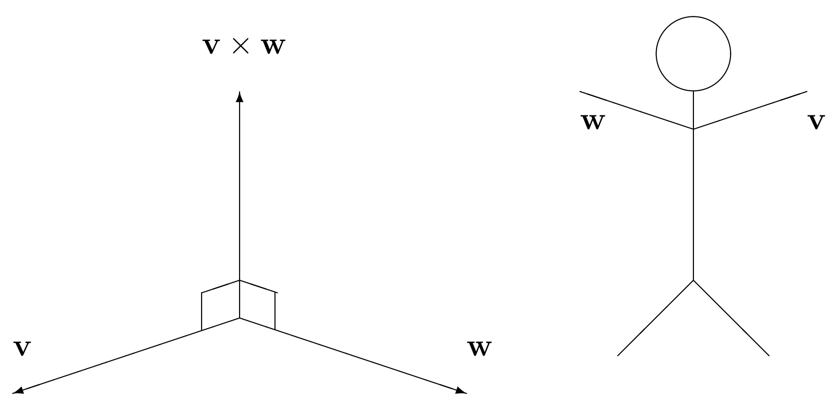

Figure 1.13: A person facing you with their arms labeled.

In order to match the person's arms with the vectors in the cross product picture, we have to turn the cross product diagram upside down:

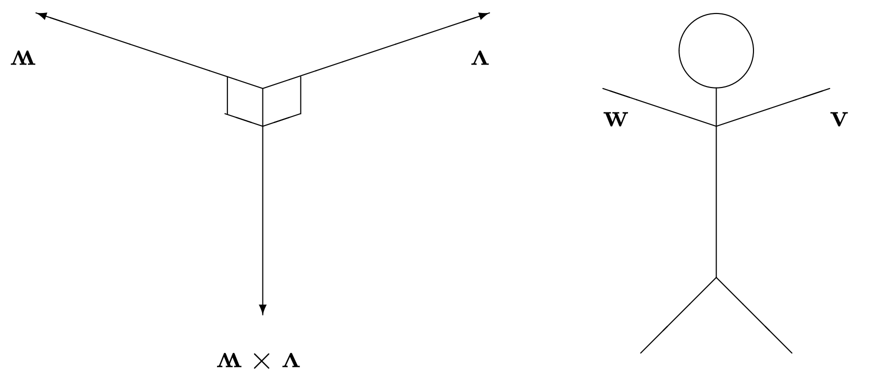

Figure 1.14: A person with arms labeled in a way that matches our cross product picture.

With this new perspective we can see that the cross product must point down towards the floor. ■

1.2 Lines, Planes, and Geometry

In this section we'll learn the equations of lines and planes, and how to apply our knowledge of vectors to these objects. In particular we will learn how to calculate the distances between two points, two lines, two planes, a point and a line, a line and a line, etc.

1.2.1 The equations of a line

We'll work in 3 dimensions, but everything we learn here applies in two dimensions as well. A line L that passes through the origin can be described by a vector equation

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = t \begin{bmatrix} a \\ b \\ c \end{bmatrix},$$

or if $\mathbf{p} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ and $\mathbf{d} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$, we can write $\mathbf{p} = t\mathbf{d}$ for short. We use the letter '$\mathbf{d}$' because the vector $\mathbf{d}$ is called the *direction vector* of the line L. Each value of t that we plug in gives us a vector that corresponds to a point on the line L, as illustrated in the picture below.

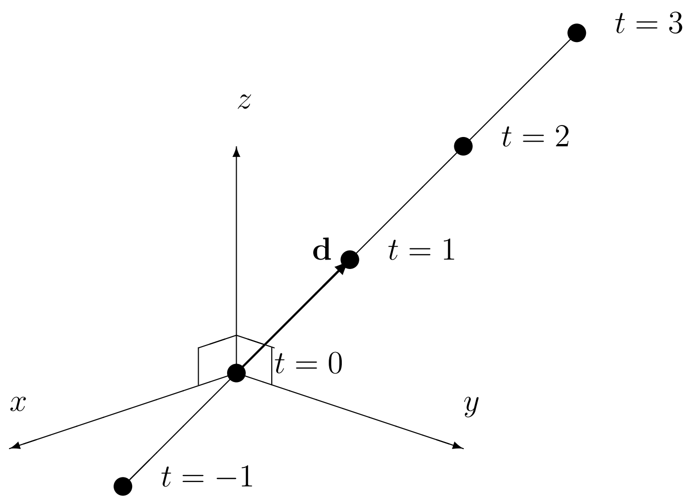

Figure 1.15: Different values of t give different points on a line through the origin.

If we want to describe a line that doesn't pass through the origin, then we have to add a nonzero *position vector* to the equation. The vector equation of a line that passes through the point (x_0, y_0, z_0) is

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} + t \begin{bmatrix} a \\ b \\ c \end{bmatrix},$$

or if we set $\mathbf{p}_0 = \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix}$, then we write $\mathbf{p} = \mathbf{p}_0 + t\mathbf{d}$ for short. In pictures, adding the vector $\mathbf{p}_0$ corresponds to shifting the picture of our line away from the origin:

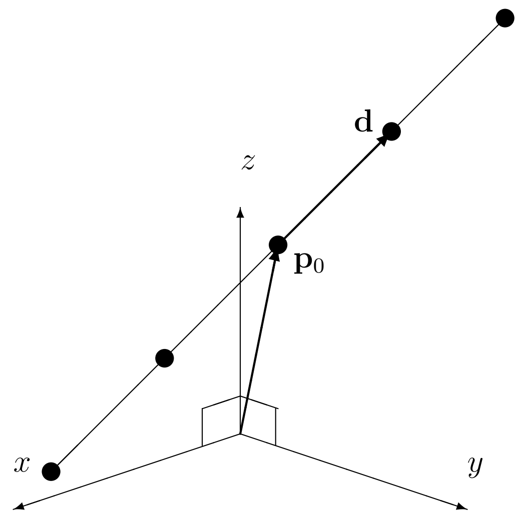

Figure 1.16: By adding the vector $\mathbf{p_0}$, we shift the line away from the origin.

So, to completely determine the equation of a line L, we need the direction vector $\mathbf{d}$ of L and we need a point on the line L. From the point we make our position vector $\mathbf{p}_0$.

Every line also has a corresponding set of *scalar equations*. The scalar equations are another way of presenting the same information we've already covered. If a line L has vector equation

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} + t \begin{bmatrix} a \\ b \\ c \end{bmatrix},$$

then its corresponding scalar equations are

$$\begin{aligned} x &= x_0 + at \\ y &= y_0 + bt \\ z &= z_0 + ct. \end{aligned}$$

Scalar equations are sometimes called *parametric equations*, and the variable t is the *parameter*.

Basically, the scalar equations are what you get by reading across the first, second and third entries of the vector equation. Again, this is something that's like the distinction between vectors and points, where they both hold the same information but are different things. The reason we sometimes use scalar equations instead of vector equations is that sometimes we'll want to refer specifically to the equation for the x entry, or the equation for the y entry, etc. With scalar equations it's easier to do that, whereas with a vector equation you would have to say 'take the equation you get from reading across the first entries of the vector equation' every time you want to talk about the x coordinate alone.

Example 9. Find the scalar and vector equations for the line passing through the points $A = (-1, 1, 5)$ and $B = (2, -1, 2)$.

Solution. If a line passes through the points A and B, then it's parallel to the vector

$$\mathbf{d} = \overrightarrow{AB} = \begin{bmatrix} 2-(-1) \\ -1-1 \\ 2-5 \end{bmatrix} = \begin{bmatrix} 3 \\ -2 \\ -3 \end{bmatrix}.$$

We can use $A = (-1, 1, 5)$ as a point on the line which tells us to shift by the vector $\mathbf{p}_0 = \begin{bmatrix} -1 \\ 1 \\ 5 \end{bmatrix}$. Then the vector equation for the line is

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 5 \end{bmatrix} + t \begin{bmatrix} 3 \\ -2 \\ -3 \end{bmatrix}.$$

Note that we can actually use any point on the line to choose our position vector $\mathbf{p}_0$. So usually we just choose the easiest or most obvious point. From here writing down the scalar equations is easy, they are:

$$\begin{aligned} x &= -1 + 3t \\ y &= 1 - 2t \\ z &= 5 - 3t. \end{aligned}$$

■

1.2.2 The equations of a plane

The equation of a plane can be written in several ways. We'll see how to go back and forth between the two most common equations, so that you can use whichever equation is most convenient.

A plane P passing through the origin can be described by one vector $\mathbf{n}$, which is called the *normal vector* of the plane. The points on P are all those points whose corresponding vector is orthogonal to $\mathbf{n} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$, so the equation of the plane is

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} = 0.$$

We write $\mathbf{p} \cdot \mathbf{n} = 0$ for short. Remember, the dot product between these two vectors being equal to zero means they're orthogonal.

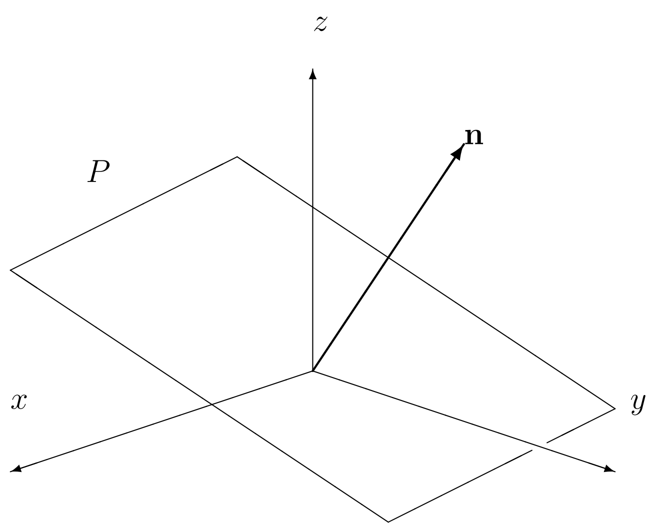

Figure 1.17: A piece of the plane orthogonal to $\mathbf{n}$.

Now if we want to move the plane away from $(0, 0, 0)$, it's more complicated than adding a position vector as in the case of a line. Suppose we want the plane to pass through the point (x_0, y_0, z_0). Then the equation for the plane is

$$\begin{bmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} = 0,$$

or in vectors we can write this as $(\mathbf{p} - \mathbf{p}_0) \cdot \mathbf{n} = 0$. Thinking about the points (x, y, z) and (x_0, y_0, z_0) corresponding to the vectors $\mathbf{p}$ and $\mathbf{p}_0$, this equation says that the vector $\mathbf{p} - \mathbf{p}_0$ which points from (x_0, y_0, z_0) to (x, y, z) is orthogonal to the normal vector $\mathbf{n}$. Putting this into a picture, we get:

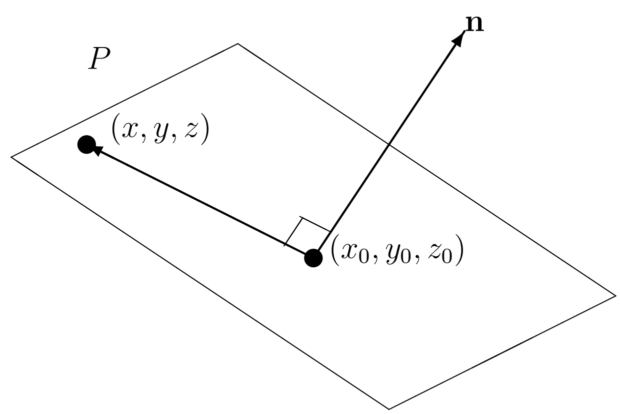

Figure 1.18: A piece of the plane orthogonal to $\mathbf{n}$, passing through (x_0, y_0, z_0). The point (x, y, z) is in the plane because the vector that points from (x_0, y_0, z_0) to (x, y, z) is orthogonal to $\mathbf{n}$.

So we see that a plane is described by two pieces of information: the normal vector $\mathbf{n}$, and a point (x_0, y_0, z_0).

We get the second form of the equation of a plane by multiplying out the dot product. The equation

$$\begin{bmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} = 0$$

becomes

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0,$$

which is often rearranged to look like

$$ax + by + cz = k.$$

In this last equation, k is a constant that is equal to $ax_0 + by_0 + cz_0$. This equation is called *the scalar equation* of the plane.

On the other hand, when you are given an equation that looks like

$$ax + by + cz = k$$

you may want to rewrite this equation in vector form. The normal vector to the plane is $\mathbf{n} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$. The constant k determines whether or not the plane passes through the origin. If $k = 0$ then the plane passes through the origin, since we can plug in $x = 0, y = 0, z = 0$ and get $0 = 0$. Otherwise, if k is not zero it means the plane doesn't pass through the origin and we have to find a point (x_0, y_0, z_0) to use in our vector equation:

$$\begin{bmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} = 0.$$

As in the case of determining the equation of a line, it doesn't matter what point on the plane you use to write your equation. So we try to find just one point by picking numbers and plugging them into the equation. Try plugging in zeroes or ones for two of the variables, and then solve for the third. For example, if c is not zero we can plug $x = 0, y = 0$ into $ax + by + cz = k$ and get $0 + 0 + cz = k$, or $z = \frac{k}{c}$. Then as our point on the plane we can use $(0, 0, \frac{k}{c})$ and write our equation as

$$\begin{bmatrix} x \\ y \\ z - \frac{k}{c} \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} = 0.$$

Example 10. Find the scalar equation of the plane P passing through the points $A = (1, 1, 2), B = (-1, 0, 5)$ and $C = (-2, 3, 0)$.

Solution. In order to write down the equation for P, we need to find a point on P, and P's normal vector $\mathbf{n}$. Obviously any one of the points $(1, 1, 2), (-1, 0, 5)$ or $(-2, 3, 0)$ can serve as our point on P.

For the normal vector we proceed as follows. Each of the vectors

$$\overrightarrow{AB} = \begin{bmatrix} (-1) - 1 \\ 0 - 1 \\ 5 - 2 \end{bmatrix} = \begin{bmatrix} -2 \\ -1 \\ 3 \end{bmatrix},$$

and

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

are parallel to P, since the points A, B and C lie in P. From the picture below, we can see that the normal vector we want is orthogonal to both $\overrightarrow{AB}$ and $\overrightarrow{AC}$.

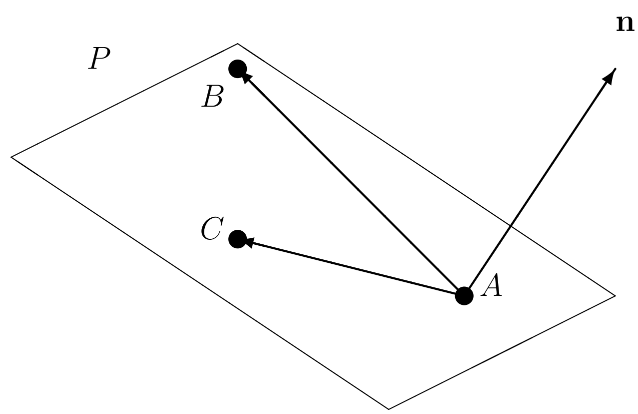

Figure 1.19: The normal vector $\mathbf{n}$ that we want is orthogonal to both vectors.

This situation is exactly the reason we learned the cross product, which will find a normal vector for us. We use as our normal vector:

$$\mathbf{n} = \overrightarrow{AB} \times \overrightarrow{AC} = \begin{bmatrix} -2 \\ -1 \\ 3 \end{bmatrix} \times \begin{bmatrix} -3 \\ 2 \\ -2 \end{bmatrix} = \begin{bmatrix} -4 \\ -13 \\ -7 \end{bmatrix}.$$

So using $A = (1, 1, 2)$ as our point on the plane, by plugging into the vector equation we get

$$\begin{bmatrix} x-1 \\ y-1 \\ z-2 \end{bmatrix} \cdot \begin{bmatrix} -4 \\ -13 \\ -7 \end{bmatrix} = 0.$$

This multiplies out to give

$$-4(x-1) - 13(y-1) - 7(z-2) = 0,$$

or

$$-4x - 13y - 7z = -4 - 13 - 14 = -31.$$

■

1.2.3 Distances between points, lines and planes

One of the standard problems that arises in geometry is to find the distance between objects. In this subsection, we'll find the distances between points, lines and planes by several different methods. In the six examples that follow, each case will be covered according to the entries in the table below.

	Point	Line	Plane
Point	Example 11	Example 12	Example 13
Line		Example 14	Example 15
Plane			Example 16

Example 11. Find the distance between the two points $A = (-1, 2, 3)$ and $B = (4, 4, 3)$.

Solution. The distance between the points A and B is the length of the vector $\overrightarrow{AB}$. We calculate:

$$||\overrightarrow{AB}|| = \left\| \begin{matrix} 4-(-1) \\ 4-2 \\ 3-3 \end{matrix} \right\| = \left\| \begin{matrix} 5 \\ 2 \\ 0 \end{matrix} \right\| = \sqrt{5^2+2^2} = \sqrt{29}.$$

■

Example 12. Find the distance between the point $A = (1, 1, 1)$ and the line L with equations

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

Also find the point C on L that is closest to the point A.

Solution. From the scalar equations of the line, we can see that the line passes through the point $B = (1, 2, -1)$, with direction vector $\mathbf{d} = \begin{bmatrix} 2 \\ -1 \\ -1 \end{bmatrix}$. The picture you should have in mind is something like this:

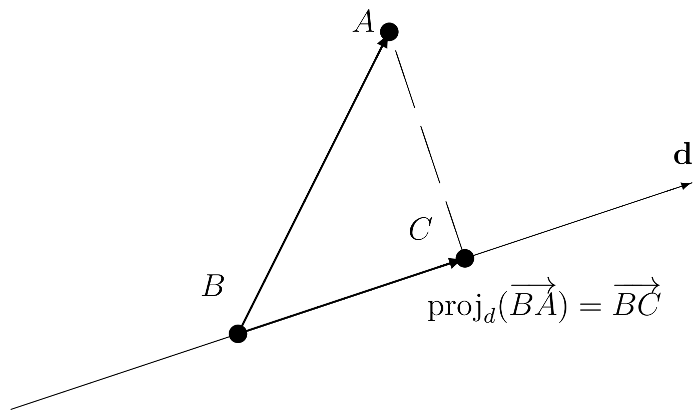

Figure 1.20: The distance we want is the length of the dotted line. The point we want to find is the point C.

From the picture we can see that the distance we want is the length of the dotted line, which is the length of the vector $\overrightarrow{CA}$. To find this vector, we'll use projection as indicated the picture. We calculate $\overrightarrow{BA} = \begin{bmatrix} 1-1 \\ 2-1 \\ -1-1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix}$, and so

$$\overrightarrow{BC} = \text{proj}_d(\overrightarrow{BA}) = \left(\frac{\overrightarrow{BA} \cdot \mathbf{d}}{||\mathbf{d}||^2}\right)\mathbf{d} = \frac{0-1+2}{4+1+1}\mathbf{d} = \frac{1}{6}\mathbf{d} = \begin{bmatrix} 1/3 \\ -1/6 \\ -1/6 \end{bmatrix}.$$

Now we can calculate the point C by adding the vector $\overrightarrow{BC}$ to the vector corresponding to the point $B = (1, 2, -1)$:

$$\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} + \begin{bmatrix} 1/3 \\ -1/6 \\ -1/6 \end{bmatrix} = \begin{bmatrix} 4/3 \\ 11/6 \\ -7/6 \end{bmatrix}.$$

So we have $C = (4/3, 11/6, -7/6)$. Now, the problem of finding the distance between the point A and the line L has been simplified to the problem of finding the distance between the point C and the point $A = (1, 1, 1)$. As in Example 11, we calculate

$$||\overrightarrow{AC}|| = \left\|\begin{matrix} 4/3-1 \\ 11/6-1 \\ -7/6-1 \end{matrix}\right\| = \left\|\begin{matrix} 1/3 \\ 7/6 \\ -11/6 \end{matrix}\right\| = \sqrt{(1/3)^2 + (7/6)^2 + (-11/6)^2} = \sqrt{29/6}.$$

In the last step, we simplified

$$\sqrt{(1/3)^2 + (7/6)^2 + (-11/6)^2} = \sqrt{4/36 + 49/36 + 121/36} = \sqrt{174/36} = \sqrt{29/6}.$$

$\blacksquare$

Example 13. Find the distance between the point $A = (1, 1, 2)$ and the plane P with equation $2x - y + 3z = 0$.

Solution. **(First method)** Choose a point on the plane, say $B = (2, 1, -1)$. Draw the normal vector to the plane, $\mathbf{n} = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}$, with its tail at the point B. The point on the plane that's closest to A will be called C.

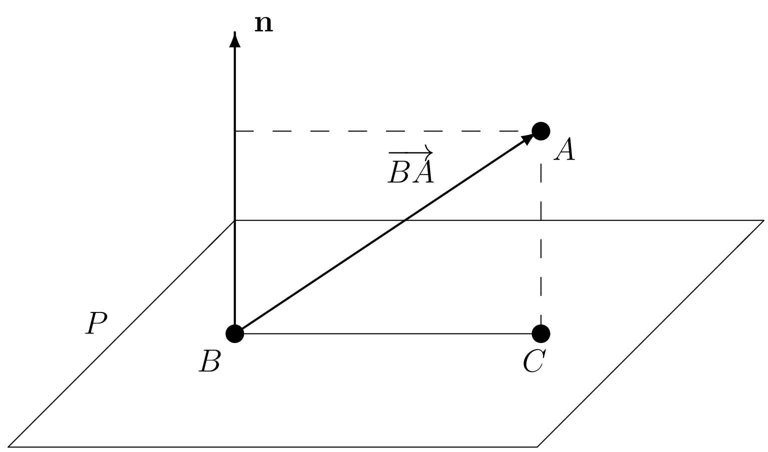

Figure 1.21: The distance we want to find is the distance from A to C.

From Figure 1.25, we can see that the distance from A to C equal to the length of $\text{proj}_{\mathbf{n}}(\overrightarrow{BA})$. So we calculate

$$\overrightarrow{BA} = \begin{bmatrix} 1-2 \\ 1-(1) \\ 2-(-1) \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 3 \end{bmatrix},$$

and then compute

$$\text{proj}_{\mathbf{n}}(\overrightarrow{BA}) = \left(\frac{\overrightarrow{BA} \cdot \mathbf{n}}{||\mathbf{n}||^2} \right) \mathbf{n} = \frac{-2+0+9}{4+1+9}\mathbf{n} = \frac{1}{2}\mathbf{n}.$$

Therefore, the distance is

$$||\text{proj}_{\mathbf{n}}(\overrightarrow{BA})|| = ||\frac{1}{2}\mathbf{n}|| = \frac{1}{2}||\mathbf{n}|| = \frac{1}{2}\sqrt{14}.$$

If we're also asked to find the coordinates of the point C, we proceed as follows. Observe that we can get to the point $A = (1, 1, 2)$ by first traveling along the vector $\begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$, then we can get from A to C by traveling in the direction of the vector $-\text{proj}_{\mathbf{n}}(\overrightarrow{BA}) = -\frac{1}{2}\mathbf{n}$. Note that the minus sign on $-\frac{1}{2}\mathbf{n}$ means we're traveling backwards along the vector (see the picture to help you visualize this). Therefore, we get to the point C by traveling along the vector

$$\begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} - \frac{1}{2}\mathbf{n} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 3/2 \\ 1/2 \end{bmatrix}.$$

So C is the point $(0, 3/2, 1/2)$.

(Second method) For the second method, refer again to Figure 1.25. There's a line that passes through A and C, and we'll call it L. The line L has direction vector $\mathbf{n}$, and we can use the point A on L to write down the scalar equations of the line, they are

$$\begin{aligned} x &= 1 + 2t \\ y &= 1 - t \\ z &= 2 + 3t. \end{aligned}$$

Now, we can think of the point C as the point of intersection of the line L and the plane P. To find the point of intersection of L and P, we plug the scalar equations of L into the equation of P. In other words, in the equation $2x - y + 3z = 0$ we replace x with $1 + 2t$, y with $1 - t$, and z with $2 + 3t$. This gives us an equation in only one variable (namely t), so we can solve for t. We get

$$\begin{aligned} 2(1 + 2t) - (1 - t) + 3(2 + 3t) &= 0 \\ 14t + 7 &= 0 \\ t &= -\frac{1}{2}. \end{aligned}$$

Now, we plug this value of t back into the equation for L, to find out what point this gives on the line. We get

$$x = 1 + 2\left(-\frac{1}{2}\right) = 0$$
$$y = 1 - \left(-\frac{1}{2}\right) = \frac{3}{2}$$
$$z = 2 + 3\left(-\frac{1}{2}\right) = \frac{1}{2}.$$

So the point on the plane that is closest to A is the point $C = (0, 3/2, 1/2)$. The distance from A to this point is

$$\overrightarrow{AC} = \left\|\begin{matrix} 0-1 \\ 3/2-1 \\ 1/2-2 \end{matrix}\right\| = \left\|\begin{matrix} 0-1 \\ 3/2-1 \\ 1/2-2 \end{matrix}\right\| = \sqrt{1 + (1/2)^2 + (-3/2)^2} = \frac{\sqrt{14}}{2}.$$

■

Example 14. Consider two lines in three dimensional space, L_1 with scalar equations

$$x = -1 + 2t$$
$$y = 1 - t$$
$$z = -2 - t$$

and L_2 with scalar equations

$$x = 3 + 2t$$
$$y = 2 + t$$
$$z = -1 - 3t.$$

Find the distance between L_1 and L_2, and for the second method, also find the points B on L_1 and C on L_2 that are closest to one another.

Solution. Before we begin, it should be pointed out that there's always the possiblity that the two lines intersect. If this is the case, then either of the methods below will give a distance of zero. The second method below will also find the intersection point of the lines.

(First method) The first method is simply to apply the formula

$$||\mathrm{proj}_{\mathbf{d}_1 \times \mathbf{d}_2}(\overrightarrow{A_1A_2})||,$$

but we first have to understand it before we use it. Imagine the two lines L_1 and L_2 in three dimensional space, with a the vector $\overrightarrow{CB}$ connecting their closest points B and C. The vector $\overrightarrow{CB}$ forms a right angle with each line, here's why:

Suppose that it *didn't* make a right angle with each line, say the angle with L_2 was less than $\pi/2$. That means close to the line L_2, we'd have something that looks like this:

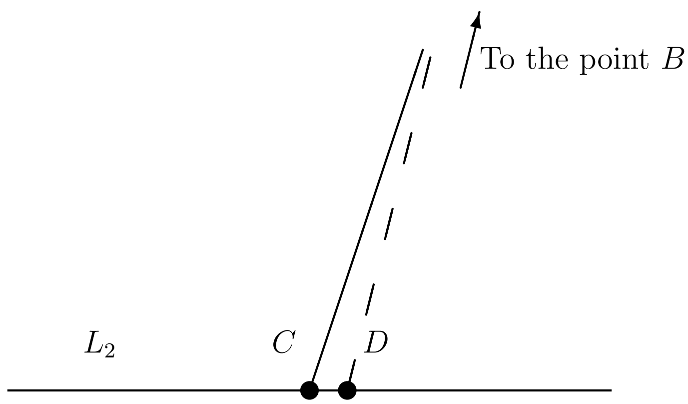

Figure 1.22: The vector $\overrightarrow{CB}$ must form right angles with the lines L_1 and L_2.

From Figure 1.22, observe that a new point D slightly to the right of C (in the direction of the acute angle between $\overrightarrow{CB}$ and L_2) will be closer to B, as indicated by the dotted line. This is not allowed, since B and C are supposed to be points on L_1 and L_2 that are as close as possible.

We conclude that $\overrightarrow{CB}$ must be orthogonal to both the direction vector $\mathbf{d}_1$ of L_1 and the direction vector $\mathbf{d}_2$ of L_2. So it's parallel to

$$\mathbf{d}_1 \times \mathbf{d}_2 = \begin{bmatrix} 2 \\ -1 \\ -1 \end{bmatrix} \times \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix} = \begin{bmatrix} (-1)(-3) - (-1)(1) \\ (-1)(2) - (-3)(2) \\ (2)(1) - (-1)(2) \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 4 \end{bmatrix}$$

Finally, we take two arbitrary points A_1 and A_2 on L_1 and L_2 respectively, project $\overrightarrow{A_1A_2}$ onto this orthogonal vector, and take the length. We choose $A_1 = (-1, 1, -2)$ and $A_2 = (3, 2, -1)$ so that

$$\overrightarrow{A_1A_2} = \begin{bmatrix} 3 - (-1) \\ 2 - 1 \\ (-1) - (-2) \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \\ 1 \end{bmatrix}$$

and then calculate

$$||\text{proj}_{\mathbf{d}_1 \times \mathbf{d}_2}(\overrightarrow{A_1 A_2})|| = ||\frac{16+4+4}{16+16+16}(\mathbf{d}_1 \times \mathbf{d}_2)|| = ||\frac{1}{2}(\mathbf{d}_1 \times \mathbf{d}_2)|| = \frac{1}{2}\sqrt{48} = 2\sqrt{3}.$$

So the distance from L_1 to L_2 is $2\sqrt{3}$.

(Second method) This method will find the coordinates of the points B and C that are closest to one another. First, we change the scalar equations of L_2 from having a parameter t to having a parameter s. We do this because we're about to do a calculation where the parameters of L_1 and L_2 will both appear together in the same equation, and if both of the parameters are t then we won't be able to tell them apart. So L_1 has equations

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

and L_2's equations are changed to

$$\begin{aligned} x &= 3 + 2s \\ y &= 2 + s \\ z &= -1 - 3s. \end{aligned}$$

Now consider two points: a point $B(t)$ on L_1

$$B(t) = (-1 + 2t, 1 - t, -2 - t)$$

and $C(s)$ on L_2

$$C(s) = (3 + 2s, 2 + s, -1 - 3s).$$

The vector $\overrightarrow{B(t)C(s)}$ points from $B(t)$ to $C(s)$, and when this vector is orthogonal to the direction vectors of both lines it will point along the shortest path between L_1 and L_2. So we solve for values of t and s that make $\overrightarrow{B(t)C(s)}$ orthogonal to both lines. First,

$$\overrightarrow{B(t)C(s)} = \begin{bmatrix} 3 + 2s - (-1 + 2t) \\ 2 + s - (1 - t) \\ -1 - 3s - (-2 - t) \end{bmatrix} = \begin{bmatrix} 4 + 2s - 2t \\ 1 + s + t \\ 1 - 3s + t \end{bmatrix}$$

This vector is orthogonal to $\mathbf{d}_1$ and $\mathbf{d}_2$ if $\overrightarrow{B(t)C(s)} \cdot \mathbf{d}_1 = 0$, i.e.

$$\begin{aligned}\overrightarrow{B(t)C(s)} \cdot \mathbf{d}_1 &= \begin{bmatrix} 4+2s-2t \\ 1+s+t \\ 1-3s+t \end{bmatrix} \cdot \begin{bmatrix} 2 \\ -1 \\ -1 \end{bmatrix} \\ &= 2(4+2s-2t) + (-1)(1+s+t) + (-1)(1-3s+t) \\ &= -6t+6s+6 \\ &= 0\end{aligned}$$

and $\overrightarrow{B(t)C(s)} \cdot \mathbf{d}_2 = 0$,

$$\begin{aligned}\overrightarrow{B(t)C(s)} \cdot \mathbf{d}_2 &= \begin{bmatrix} 4+2s-2t \\ 1+s+t \\ 1-3s+t \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix} \\ &= 2(4+2s-2t) + (1)(1+s+t) + (-3)(1-3s+t) \\ &= -6t+14s+6 \\ &= 0.\end{aligned}$$

Now we solve the equations

$$\begin{aligned}-6t + 14s &= -6 \\ -6t + 6s &= -6\end{aligned}$$

for t and s. The second equation rearranges to give $s = -1 + t$, which we plug into the first equation to get $-6t + 14(-1 + t) = -6$. We find $t = 1$, and so $s = 0$. So the points B and C on L_1 and L_2 that are closest are

$$B = B(1) = (-1 + 2(1), 1 - 1, -2 - 1) = (1, 0, -3)$$

and

$$C = C(0) = (3 + 2(0), 2 + 0, -1 - 3(0)) = (3, 2, -1),$$

because these values of t and s make the vector $\overrightarrow{B(t)C(s)}$ orthogonal to $\mathbf{d}_1$ and $\mathbf{d}_2$. So, the distance between the two lines is $||\overrightarrow{B(1)C(0)}|| = \sqrt{2^2 + 2^2 + 2^2} = \sqrt{12} = 2\sqrt{3}$. ∎

Example 15. Find the distance between the line L

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

and the plane P with equation $2x + 4y - z = 3$.

Solution. There are three possibilities. Either line L is inside the plane P, or it intersects the plane P exactly once, or it's parallel to P. These three possibilities are illustrated in the figures below:

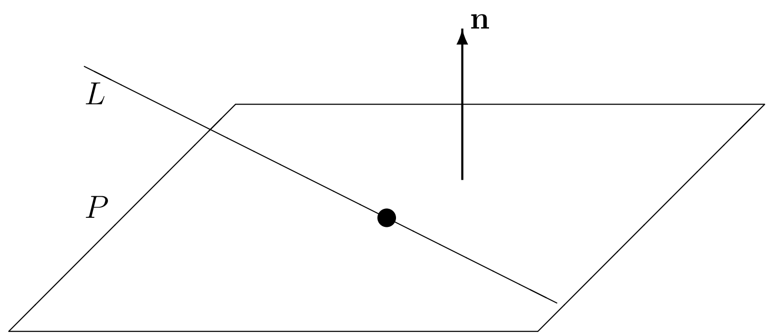

Figure 1.23: A line L intersecting the plane P

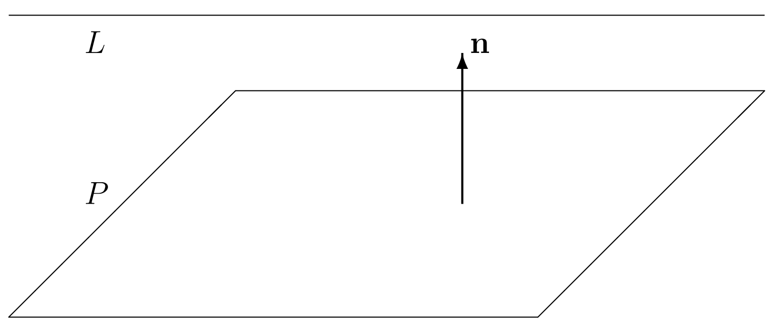

Figure 1.24: A line L parallel to the plane P

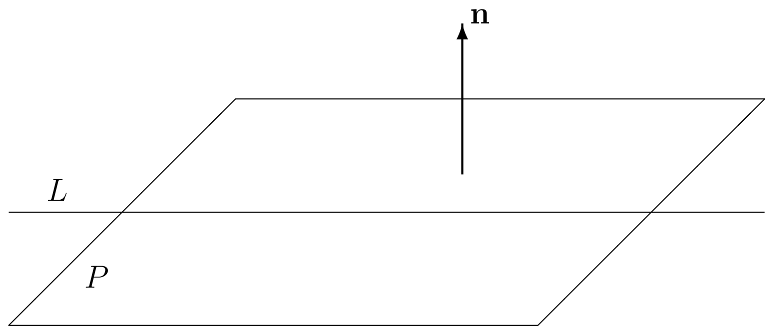

Figure 1.25: A line L inside the plane P

We can see that in the cases where L is parallel to P or inside P, the direction vector of L must be orthogonal to the normal vector of P. If the direction vector is not orthogonal to the normal vector, then they must intersect. So, to check which case we're in we have to see if the direction vector of L and the normal of P are orthogonal or not. The direction vector of L is $\mathbf{d} = \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix}$, and the normal vector of P is $\mathbf{n} = \begin{bmatrix} 2 \\ 4 \\ -1 \end{bmatrix}$. Since

$$\mathbf{n} \cdot \mathbf{d} = \begin{bmatrix} 2 \\ 4 \\ -1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix} = (2)(1) + (4)(-1) + (-1)(-2) = 0,$$

we know the vectors $\mathbf{d}$ and $\mathbf{n}$ are orthogonal, so that L is either parallel to P or inside P. We check the point $(-1, 1, -2)$ on L and find that when we plug it into the equation for P, we get

$$2(-1) + 4(1) - (-2) = 4 \neq 3.$$

So, the point $(-1, 1, -2)$ on L isn't on the plane P, so L isn't inside the plane P. They must be parallel. Now we are ready to find the distance.

Because L is parallel to P, every point on L is the same distance from P. Therefore we can pick an arbitrary point on L and calculate its distance from P as in Example 13. For our point on L we will use $A = (-1, 1, 2)$.

As in Example 13, we pick an arbitrary point on P, say $B = (0, 0, -1/3)$. Then

$$\overrightarrow{BA} = \begin{bmatrix} -1 - 0 \\ 1 - 0 \\ 2 - (-1/3) \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 7/3 \end{bmatrix},$$

and

$$\text{proj}_{\mathbf{n}}(\overrightarrow{BA}) = \left(\frac{\overrightarrow{BA} \cdot \mathbf{n}}{||\mathbf{n}||^2} \right) \mathbf{n} = \frac{-2 + 4 + (-7/3)}{4 + 16 + 1} \mathbf{n} = \frac{1}{63} \mathbf{n}.$$

Therefore, the distance is

$$||\text{proj}_{\mathbf{n}}(\overrightarrow{BA})|| = ||-\frac{1}{63}\mathbf{n}|| = \frac{1}{63}||\mathbf{n}|| = \frac{1}{63}\sqrt{21}.$$

■

Example 16. Calculate the line of intersection of the plane $2x - 4y + z = 1$ and $x - y - z = 5$.

Solution. First, a remark. When you have two planes, it is possible that they don't intersect at all. This is the case when their normal vectors are parallel. In order to find the distance between them, you can simply choose a point on one plane and then find the point-plane distance as in Example 13.

The remaining case is when the two planes intersect in a line, as with the two planes given above. First we need to find a point that lies in both planes. So, solve for x in the second equation

$$x = 5 + y + z$$

and use this to eliminate x from the other:

$$2(5 + y + z) - 4y + z = 1.$$

The equation $2(5 + y + z) - 4y + z = 1$ simplifies to $-2y + 3z = -9$. Of course, there is not a unique solution for y and z in this case. But all we need is one point on the line of intersection of the two planes, so we choose a solution of this equation $y = 0$ and $z = -3$. Now plug these values of y and z back into the original plane equations to find x:

$$x - y - z = 5 \text{ becomes } x - 0 - (-3) = 5,$$

so $x = 2$. Therefore a point which lies on both planes is $(2, 0, -3)$ (check this!), so this point is on their line of intersection.

Now we need the direction vector for the line of intersection. It's orthogonal to the normal vectors of both planes, so the direction vector is

$$\mathbf{d} = \mathbf{n}_1 \times \mathbf{n}_2 = \begin{bmatrix} 2 \\ -4 \\ 1 \end{bmatrix} \times \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \\ 2 \end{bmatrix}.$$

So the equation of the line of intersection is

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = t \begin{bmatrix} 5 \\ 3 \\ 2 \end{bmatrix} + \begin{bmatrix} 2 \\ 0 \\ -3 \end{bmatrix}.$$

■

Chapter 2

Calculating with Matrices

The purpose of this chapter is, simply put, to show you how to do all the calculations one needs to know at our level. All of these calculations have theoretical meanings that will be explained in Chapter 3. If you master these calculations before moving on to Chapter 3, then the theoretical discussion will be much easier to handle than if we had tried to do the theory and the calculations at the same time.

2.1 Solving equations with matrices

2.1.1 What is a matrix?

A matrix is a rectangular array of numbers. Matrices are named using capitals letters, so an example of a matrix would be

$$A = \begin{bmatrix} 2 & 7 & 3 \\ 4 & 0 & 1 \end{bmatrix}.$$

A vector is a special case of a matrix which has only one column. Vectors come naturally from considering directions and lengths of arrows in space, but matrices come naturally from systems of equations. If we have the system of equations

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

then by simply forgetting the variables and the other mathematical symbols and recording only the numbers, we get the matrix

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

Note that the 0 in the leftmost column indicates that there are no x's in the second equation. Usually, when a matrix comes from a system of equations like this we add a vertical line to indicate where the equals sign was:

$$\left[\begin{array}{ccc|c} 5 & 3 & -1 & 1 \\ 0 & 1 & 4 & -2 \end{array}\right].$$

The numbers in a matrix are called entries, and the rows of a matrix are numbered from top to bottom, the columns numbered from left to right. So in the matrix A below

$$A = \left[\begin{array}{ccc|c} 5 & 3 & -1 & 1 \\ 0 & 1 & 4 & -2 \end{array}\right]$$

the third column is $\begin{bmatrix} -1 \\ 4 \end{bmatrix}$, and the first row is $\begin{bmatrix} 5 & 3 & -1 & 1 \end{bmatrix}$.

The *size* of a matrix means the number of rows and the number of columns in a matrix, listed in that order. So the matrix above is a 2×4 matrix, because it has two rows and four columns (the symbols '2×4' should be read as 'two by four') . If you want to specify a single entry in a matrix, you give its row and column. For example, the $(2, 3)$-entry in the matrix above is 4.

Because matrices are named with capital letters, the entries in the matrix are named with lowercase letters. For example in the matrix A if the $(2, 1)$-entry is unknown, we would denote it by the variable $a_{2,1}$. This means as a whole, the matrix A would look like

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots \\ a_{2,1} & a_{2,2} & \cdots \\ \vdots & \vdots & \end{bmatrix}.$$

The dots are a common way of indicating that the pattern continues on for some number of entries. Sometimes instead of writing

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots \\ a_{2,1} & a_{2,2} & \cdots \\ \vdots & \vdots & \end{bmatrix}$$

we simply write $A = [a_{i,j}]$ to indicate that A is a matrix, and its entries are named $a_{1,1}, a_{1,2}, \cdots$, etc.

2.1.2 Row reduction

Row reduction, also called *Gaussian elimination*, is a way of solving a system of linear equations by using matrices. Roughly, the steps in this process are:

1. Replacing the system of linear equations with a matrix A.

2. Changing the matrix A into a new matrix according to a recipe.

3. Transforming the new matrix back into a set of solutions for the system of linear equations.

Obviously each of steps (2) and (3) need more explaining. In order to explain step (2), we will work through an example below. The important thing to learn from this example is how each of the operations that we do on a system of equations corresponds to a certain way of changing a matrix. After we finish the example, then we can go on to explain the general procedure.

Suppose we are going to solve

$$\begin{aligned} -2x + 3y &= 8 \\ 3x - y &= -5 \end{aligned}$$

This system corresponds to the matrix

$$\left[\begin{array}{cc|c} -2 & 3 & 8 \\ 3 & -1 & -5 \end{array}\right],$$

which is called the *augmented matrix* of the system. The matrix $\begin{bmatrix} -2 & 3 \\ 3 & -1 \end{bmatrix}$ is called the *coefficient matrix* of the system and the matrix $\begin{bmatrix} 8 \\ -5 \end{bmatrix}$ is called the *constant matrix* (or sometimes the *constant vector*).

In the table below, we solve the system in the column on the left. In the column on the right we translate each step into a way of changing an augmented matrix. The steps in the table below are not the fastest or easiest steps one could choose in order to solve for x and y, but they illustrate a particular example of the algorithm we'll develop a few pages from now. This algorithm will work on any system of linear equations, including very large systems where ad-hoc steps could lead to confusion.

System of equations	Augmented matrix
$$\begin{aligned} -2x + 3y &= 8 \\ 3x - y &= -5 \end{aligned}$$	$$\left[\begin{array}{cc\|c} -2 & 3 & 8 \\ 3 & -1 & -5 \end{array}\right]$$
Multiply the first equation by $-1/2$ $$\begin{aligned} x + (-3/2)y &= -4 \\ 3x - y &= -5 \end{aligned}$$	Multiply the first row by $-1/2$ $$\left[\begin{array}{cc\|c} 1 & -3/2 & -4 \\ 3 & -1 & -5 \end{array}\right]$$
Subtract 3 times the first equation from the second $$\begin{aligned} x + (-3/2)y &= -4 \\ (7/2)y &= 7 \end{aligned}$$	Substract 3 times the first row from the second row $$\left[\begin{array}{cc\|c} 1 & -3/2 & -4 \\ 0 & 7/2 & 7 \end{array}\right]$$
Multiply the second equation by $2/7$ $$\begin{aligned} x + (-3/2)y &= -4 \\ y &= 2 \end{aligned}$$	Multiply the second row by $2/7$ $$\left[\begin{array}{cc\|c} 1 & -3/2 & -4 \\ 0 & 1 & 2 \end{array}\right]$$
Add $3/2$ times the second equation to the first $$\begin{aligned} x &= -1 \\ y &= 2 \end{aligned}$$	Add $3/2$ times the second row to the first row $$\left[\begin{array}{cc\|c} 1 & 0 & -1 \\ 0 & 1 & 2 \end{array}\right]$$

So you see from this example, instead of writing the system of equations at each step we could write only the matrix instead. Instead of writing the way in which we changed the system of equations we can write the way in which we changed the rows of the augmented matrix.

2.1.3 Row operations and the row reduction algorithm

In order to describe the row reduction algorithm we need to introduce two pieces of terminology.

An *elementary row operation* is a way of changing the rows of a matrix. There are three types of elementary row operations, and we have a shorthand way of writing each one. They are:

I. Swapping two rows. If we swap row i and row j in a matrix, we'll write $R_i \Leftrightarrow R_j$.

II. Multiplying a row by a nonzero number. If we multuply row i by a number c, we'll write $R_i \Rightarrow cR_i$.

III. Adding a multiple of one row to another. If we add c times row i to row j, we'll write $R_i \Rightarrow R_i + cR_j$.

In each of the row operations above, the arrows '$\Rightarrow$' can be read aloud as the word 'becomes.' This will help the shorthand notation make sense. For example, the elementary row operation '$R_i \Rightarrow R_i + cR_j$' should be read aloud as 'row i becomes row i plus c times row j.'

Now we introduce the row reduction algorithm. In order to make its description simpler, we will call the first nonzero entry in a row the *leading entry* in that row. If the first nonzero entry is a 1, we'll call it *a leading* 1.

Given a matrix A, here is how you perform row reduction:

The Algorithm.

1. Find the leftmost column in A which has a nonzero entry. Pick one nonzero entry in that column. By swapping rows, move the row containing that entry to the top. (Use row operations of type I)

2. If the top row now has leading entry $k \neq 0$, multiply the whole top row by $1/k$ to make the leading entry into a 1. (Use row operations of type II)

3. Make all the entries below the leading 1 from step two into zeroes. This is done by adding appropriate multiples of the top row to those rows below it. (Use row operations of type III)

4. Ignore the top row, which now has a leading 1. Repeat steps (1)-(4) on the rows of A which haven't been changed by steps 1-3 in order to have leading ones. Proceed to step (5) once every row has a leading one with zeroes below it.

5. Make all the entries above every leading 1 into zeroes. This is done by adding appropriate multiples of each row containing a leading one to those rows above it. (Use row operations of type III)

I cannot stress enough how important this algorithm is. Anyone studying linear algebra must completely master these steps in order to proceed with any of the material that comes later in this book. We will see shortly how this algorithm can be used to solve a system of equations, but first we will practice once on a matrix that does not come from a system of equations.

Example 17. Row reduce the matrix

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

Solution. We will follow the steps outlined above exactly.

Step 1. The first nonzero column from the left is column 2. It has a nonzero entry in the second row, we move it to the top and write the step like this:

$$\begin{bmatrix} 0 & 0 & -1 & 3 \\ 0 & 2 & 4 & -3 \\ 0 & 1 & 3 & 6 \end{bmatrix} \xrightarrow{R_1 \Leftrightarrow R_2} \begin{bmatrix} 0 & 2 & 4 & -3 \\ 0 & 0 & -1 & 3 \\ 0 & 1 & 3 & 6 \end{bmatrix}$$

Step 2. The top row now has leading entry 2, so we scale the top row by 1/2.

$$\begin{bmatrix} 0 & 2 & 4 & -3 \\ 0 & 0 & -1 & 3 \\ 0 & 1 & 3 & 6 \end{bmatrix} \xrightarrow{R_2 \Rightarrow (1/2)R_2} \begin{bmatrix} 0 & 1 & 2 & -3/2 \\ 0 & 0 & -1 & 3 \\ 0 & 1 & 3 & 6 \end{bmatrix}$$

Step 3. Make zeroes below the leading one we just created. So we have to make the $(3,2)$-entry into a zero, we can do this by subtracting R_1 from R_3 so that the two leading ones will cancel.

$$\begin{bmatrix} 0 & 1 & 2 & -3/2 \\ 0 & 0 & -1 & 3 \\ 0 & 1 & 3 & 6 \end{bmatrix} \xrightarrow{R_3 \Rightarrow R_3 - R_1} \begin{bmatrix} 0 & 1 & 2 & -3/2 \\ 0 & 0 & -1 & 3 \\ 0 & 0 & 1 & 15/2 \end{bmatrix}$$

Step 4. Now we focus on the last two rows of our matrix, because we haven't engineered them to have leading ones by using steps 1-3 yet. We write the entire matrix at each step, but just focus on the last two rows and repeat steps 1-4.

Substep 4.1. The leftmost column that has a nonzero entry in the last two rows is column 3. The $(2,3)$-entry is -1, which is not zero. So substep 4.1, which says "move the row containing that nonzero entry to the top" does not require us to swap any rows, because the nonzero entry is already at the top. (Remember since we are only focusing on the last two rows, 'top' here means row 2!)

Substep 4.2. Make the nonzero leading entry from the last step into a 1.

$$\begin{bmatrix} 0 & 1 & 2 & -3/2 \\ 0 & 0 & -1 & 3 \\ 0 & 0 & 1 & 15/2 \end{bmatrix} \xrightarrow{R_2 \Rightarrow (-1)R_2} \begin{bmatrix} 0 & 1 & 2 & -3/2 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 1 & 15/2 \end{bmatrix}$$

Substep 4.3. Below the leading one we created in the last step, make all the entries zero.

$$\begin{bmatrix} 0 & 1 & 2 & -3/2 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 1 & 15/2 \end{bmatrix} \xrightarrow{R_3 \Rightarrow R_3 - R_2} \begin{bmatrix} 0 & 1 & 2 & -3/2 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 21/2 \end{bmatrix}$$

Substep 4.4. Last, we repeat steps 1-4 on the remaining row that does not have a leading one with zeroes below it (row 3). If we do steps 1-4 on row 3, the only step which makes any changes is step 2, where we scale the row to have a leading 1:

$$\begin{bmatrix} 0 & 1 & 2 & -3/2 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 21/2 \end{bmatrix} \xrightarrow{R_3 \Rightarrow (2/21)R_3} \begin{bmatrix} 0 & 1 & 2 & -3/2 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Step 5. Make all the entries above the leading ones into zeroes.

$$\begin{bmatrix} 0 & 1 & 2 & -3/2 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{R_2 \Rightarrow R_2 + 3R_3} \begin{bmatrix} 0 & 1 & 2 & -3/2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{R_1 \Rightarrow R_1 + (3/2)R_3} \begin{bmatrix} 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{R_1 \Rightarrow R_1 - 2R_2} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Here, the algorithm stops. Congratulations, you have row-reduced your first matrix. ■

2.1.4 Solving systems of equations with row reduction

At the beginning of the section on row reduction we saw that row reduction would help us solve systems in 3 steps:

1. Replacing the system of linear equations with a matrix A.
2. Changing the matrix A according to a recipe.
3. Transforming the matrix back into a set of solutions for the system of linear equations.

We already know how to do step (1), and step (2) was just covered in the last section. Now we learn how to do step (3). Again, we need some important terminology before we can proceed.

A matrix A is said to be in *echelon form* if:

1. If there are any rows of zeroes in A, they are at the bottom.
2. Every nonzero row has a leading 1.
3. Every leading 1 is to the right of all the leading 1's above it.

If a matrix has had steps (1)-(4) of the row reduction algorithm done to it, then it will be in echelon form. If you do step (5) of the row reduction algorithm to a matrix in echelon form, then it will also have the property:

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

An echelon form matrix which has this additional property is said to be in *reduced echelon form*.

Now we can apply this terminology to describe the last step in solving systems of equations . Suppose you have a system of equations, which you translate into a matrix. Then you use row-reduction to bring the matrix to reduced echelon form. Once you have a matrix that is in reduced echelon form, you can do the steps which follow in order to write the answer to your system of equations. It is extremely important that you only do these steps to reduced echelon matrices! That's the whole reason for naming reduced echeclon matrices before introducing these steps.

Writing your answer.

1. Translate the reduced echelon matrix back into a system of equations.

2. Write the variables in a vector.

3. For each variable in the vector that corresponds to a leading 1 in the reduced matrix, use one of the equations to substitute other variables in its place. If a variable corresponds to a column with no leading 1 then it should not be touched!

4. Factor the vector of equations as a sum of vectors each multiplied by a single variable.

Of course in order for this to make sense we need to see it in action.

Example 18. Solve the system of equations

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

Solution. First we translate the system into an augmented matrix

$$\left[\begin{array}{ccc|c} -1 & 1 & -2 & 3 \\ 3 & -1 & 1 & -5 \end{array}\right]$$

Now we row reduce the matrix to bring it to reduced echelon form.

$$\left[\begin{array}{ccc|c} -1 & 1 & -2 & 3 \\ 3 & -1 & 1 & -5 \end{array}\right] \xrightarrow{R_1 \Rightarrow (-1)R_1} \left[\begin{array}{ccc|c} 1 & -1 & 2 & -3 \\ 3 & -1 & 1 & -5 \end{array}\right]$$

$$\left[\begin{array}{ccc|c} -1 & 1 & -2 & 3 \\ 3 & -1 & 1 & -5 \end{array}\right] \xrightarrow{R_2 \Rightarrow R_2 - 3R_1} \left[\begin{array}{ccc|c} 1 & -1 & 2 & -3 \\ 0 & 2 & -5 & 1 \end{array}\right]$$

$$\left[\begin{array}{ccc|c} 1 & -1 & 2 & -3 \\ 0 & 2 & -5 & 1 \end{array}\right] \xrightarrow{R_2 \Rightarrow (1/2)R_2} \left[\begin{array}{ccc|c} 1 & -1 & 2 & -3 \\ 0 & 1 & -5/2 & 1/2 \end{array}\right]$$

Now the matrix is in echelon form. We bring it to reduced echelon form:

$$\left[\begin{array}{ccc|c} 1 & -1 & 2 & -3 \\ 0 & 1 & -5/2 & 1/2 \end{array}\right] \xrightarrow{R_1 \Rightarrow R_1 + R_2} \left[\begin{array}{ccc|c} 1 & 0 & -1/2 & -5/2 \\ 0 & 1 & -5/2 & 1/2 \end{array}\right]$$

Finally, here are the steps to write our answer. First step (1) says to translate our matrix back into equations.

$$\begin{aligned} x - (1/2)z &= -5/2 \\ y - (5/2)z &= 1/2 \end{aligned}$$

Next, step (2) says we write our variables in a vector:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Step (3) says to substitute away the variables which correspond to columns containing leading 1's. There are leading 1's in columns 1 and 2 of the reduced echelon matrix, which correspond to the variables x and y. So we use $x - (1/2)z = -5/2$ to substitute $-5/2+(1/2)z$ for x, and we use $y-(5/2)z = 1/2$ to substitute $1/2+(5/2)z$ for y. We leave the variable z in the third entry untouched.

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -5/2 + (1/2)z \\ 1/2 + (5/2)z \\ z \end{bmatrix}$$

Then according to step (4) we factor our answer:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -5/2 + (1/2)z \\ 1/2 + (5/2)z \\ z \end{bmatrix} = \begin{bmatrix} -5/2 \\ 1/2 \\ 0 \end{bmatrix} + z \begin{bmatrix} 1/2 \\ 5/2 \\ 1 \end{bmatrix}$$

This form of writing the solution is called the *general solution* to the system of equations. ■

2.1.5 Important examples, concepts and terminology

Here is a list of important terms and concepts related to solving systems of linear equations. Each new word or idea links to an example that shows the relevant concept. You can also go over these examples if you want to practice row reduction.

Inconsistent systems. When a system of equations has a solution, it is called *consistent*. All the systems we have seen so far have been consistent. When a system of equations has no solution, it is called *inconsistent*. See Example 19 for an example of an inconsistent system.

Constant vectors. When you find the general solution of a system it will be of the form

$$X = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n + \mathbf{c}$$

where the x_i's are variables and the $\mathbf{v}_j$'s are vectors. There is another vector, $\mathbf{c}$, which is not multiplied by a variable x_i and is called a *constant vector*. See Example 20 for an example of solving a system with nonzero constant vector.

Basic solutions. When you solve a homogeneous system you get an answer of the form

$$X = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n + \mathbf{c}$$

where the x_i's are variables and the $\mathbf{v}_j$'s are vectors the vectors $\mathbf{v}_j$ are called *basic solutions*. See Example 22 for an example which highlights the notion of basic solutions.

Rank. The *rank* of a matrix is the number of leading 1's in the matrix when it's in row-reduced echelon form. When the rank of a matrix A is 3, for short one writes $\operatorname{rank}(A) = 3$. See Example 20 for an example of solving a system and calculating the rank of the associated matrix.

Free and non-free variables. Once you row reduce the matrix corresponding to a system of equations to reduced echelon form, every column of the matrix either has a leading 1, or it doesn't have a leading 1. Those columns that have no leading 1 correspond to variables that are called *free variables*, or sometimes *parameters.* A variable whose corresponding column contains a leading 1 is called a *non-free variable.* See Example 21 for an example of this.

Homogeneous systems. A system of linear equations is called *homogeneous* if the numbers on the right hand side of the equals sign are all zero. A system with non-zero numbers on the right hand side of the equals sign is called non-homogeneous. When you solve a homogeneous system you get an answer of the form

$$X = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n$$

What is special about this solution is that there is no constant vector, the general solution contains only vectors $\mathbf{v}_j$ which are the basic solutions. See Example 22 for an example of a homogeneous system.

Trivial and nontrivial solutions. If

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

is a solution to some system of equations, then it is called the *trivial solution.* Any solution that is not the trivial solution is called a *nontrivial solution.*

Example 19. Solve the system whose augmented matrix is

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

Solution. Row reduce the augmented matrix to find the general solution.

$$\left[\begin{array}{ccc|c} 1 & 1 & 3 & 1 \\ 1 & -1 & 4 & 2 \\ 0 & -2 & 1 & 0 \end{array}\right] \xrightarrow{R_2 \Rightarrow R_2 - R_1} \left[\begin{array}{ccc|c} 1 & 1 & 3 & 1 \\ 0 & -2 & 1 & 1 \\ 0 & -2 & 1 & 0 \end{array}\right]$$

$$\left[\begin{array}{ccc|c} 1 & 1 & 3 & 1 \\ 0 & -2 & 1 & 1 \\ 0 & -2 & 1 & 0 \end{array}\right] \xrightarrow{R_2 \Rightarrow (-1/2) R_2} \left[\begin{array}{ccc|c} 1 & 1 & 3 & 1 \\ 0 & 1 & -1/2 & -1/2 \\ 0 & -2 & 1 & 0 \end{array}\right]$$

$$\left[\begin{array}{ccc|c} 1 & 1 & 3 & 1 \\ 0 & -2 & 1 & 1 \\ 0 & -2 & 1 & 0 \end{array}\right] \xrightarrow{R_2 \Rightarrow (-1/2) R_2} \left[\begin{array}{ccc|c} 1 & 1 & 3 & 1 \\ 0 & 1 & -1/2 & -1/2 \\ 0 & -2 & 1 & 0 \end{array}\right]$$

$$\left[\begin{array}{ccc|c} 1 & 1 & 3 & 1 \\ 0 & 1 & -1/2 & -1/2 \\ 0 & -2 & 1 & 0 \end{array}\right] \xrightarrow{R_3 \Rightarrow R_3 + 2R_2} \left[\begin{array}{ccc|c} 1 & 1 & 3 & 1 \\ 0 & 1 & -1/2 & -1/2 \\ 0 & 0 & 0 & -1 \end{array}\right]$$

$$\left[\begin{array}{ccc|c} 1 & 1 & 3 & 1 \\ 0 & 1 & -1/2 & -1/2 \\ 0 & 0 & 0 & -1 \end{array}\right] \xrightarrow{R_3 \Rightarrow -R_3} \left[\begin{array}{ccc|c} 1 & 1 & 3 & 1 \\ 0 & 1 & -1/2 & -1/2 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

$$\left[\begin{array}{ccc|c} 1 & 1 & 3 & 1 \\ 0 & 1 & -1/2 & -1/2 \\ 0 & 0 & 0 & 1 \end{array}\right] \xrightarrow{R_1 \Rightarrow R_1 - R_2} \left[\begin{array}{ccc|c} 1 & 0 & 7/2 & 3/2 \\ 0 & 1 & -1/2 & -1/2 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

Now the matrix is in reduced echelon form, and we translate it back into equations. We get

$$\begin{aligned} x_1 + (7/2)x_3 &= 3/2 \\ x_2 - (1/2)x_3 &= -1/2 \\ 0 &= 1 \end{aligned}$$

It is obvious that the equation $0 = 1$ is not possible, meaning there is no solution and the system is called *inconsistent*. It is important to note that this example is not a special case: whenever a system is inconsistent, row reduction will always result in the last equation being $0 = 1$. So this is how every example of an inconsistent system will end. ■

Example 20. Solve the system of equations

$$\begin{aligned} x_1 - 2x_2 - x_3 \qquad &= 1 \\ 2x_2 - x_3 - 6x_4 &= 0 \\ -x_1 \qquad + 2x_3 + 6x_4 &= -1 \end{aligned}$$

What is the rank of the associated coefficient matrix?

Solution. We row reduce the augmented matrix to find the general solution.

$$\left[\begin{array}{cccc|c} 1 & -2 & -1 & 0 & 1 \\ 0 & 2 & -1 & -6 & 0 \\ -1 & 0 & 2 & 6 & -1 \end{array}\right] \xrightarrow{R_3 \Rightarrow R_3 + R_1} \left[\begin{array}{cccc|c} 1 & -2 & -1 & 0 & 1 \\ 0 & 2 & -1 & -6 & 0 \\ 0 & -2 & 1 & 6 & 0 \end{array}\right]$$

$$\left[\begin{array}{cccc|c} 1 & -2 & -1 & 0 & 1 \\ 0 & 2 & -1 & -6 & 0 \\ 0 & -2 & 1 & 6 & 0 \end{array}\right] \xrightarrow{R_2 \Rightarrow (1/2) R_2} \left[\begin{array}{cccc|c} 1 & -2 & -1 & 0 & 1 \\ 0 & 1 & -1/2 & -3 & 0 \\ 0 & -2 & 1 & 6 & 0 \end{array}\right]$$

$$\left[\begin{array}{cccc|c} 1 & -2 & -1 & 0 & 1 \\ 0 & 1 & -1/2 & -3 & 0 \\ 0 & -2 & 1 & 6 & 0 \end{array}\right] \xrightarrow{R_3 \Rightarrow R_3 + 2R_2} \left[\begin{array}{cccc|c} 1 & -2 & -1 & 0 & 1 \\ 0 & 1 & -1/2 & -3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

$$\left[\begin{array}{cccc|c} 1 & -2 & -1 & 0 & 1 \\ 0 & 1 & -1/2 & -3 & 0 \\ 0 & -2 & 1 & 6 & 0 \end{array}\right] \xrightarrow{R_1 \Rightarrow R_1 + 2R_2} \left[\begin{array}{cccc|c} 1 & 0 & -2 & -6 & 1 \\ 0 & 1 & -1/2 & -3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Now the matrix is in reduced echelon form. To write the soluton, we translate back into equations:

$$\begin{aligned} x_1 - 2x_3 - 6x_4 &= 1 \\ x_2 - (1/2)x_3 - 3x_4 &= 0 \end{aligned}$$

Write a vector of variables and substitute away those variables corresponding to leading 1's. So in this case we substitute away x_1 and x_2.

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 + 2x_3 + 6x_4 \\ (1/2)x_3 + 3x_4 \\ x_3 \\ x_4 \end{bmatrix}$$

Now to write our general solution in its final form, we factor it into vectors. Note that in this example we have a *constant vector*, that is, a vector that isn't multiplied by a variable.

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 2 \\ 1/2 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 6 \\ 3 \\ 0 \\ 1 \end{bmatrix}$$

To find the rank of the coefficient matrix, we only need to count the number of leading ones in the reduced echelon form of the coefficient matrix. After our row reduction, the coefficient matrix became

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

Since there are two leading ones in this matrix, its rank is 2.

■

Example 21. Solve the system whose augmented matrix is

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

Solution. We row reduce the augmented matrix according to the algorithm.

$$\left[\begin{array}{cccc|c} 0 & -1 & -1 & 0 & 1 \\ 0 & 0 & 0 & -6 & 2 \\ 0 & -2 & 1 & 2 & 0 \end{array}\right] \xrightarrow{R_1 \Rightarrow -R_1} \left[\begin{array}{cccc|c} 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & -6 & 2 \\ 0 & -2 & 1 & 2 & 0 \end{array}\right]$$

$$\left[\begin{array}{cccc|c} 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & -6 & 2 \\ 0 & -2 & 1 & 2 & 0 \end{array}\right] \xrightarrow{R_3 \Rightarrow R_3 + 2R_1} \left[\begin{array}{cccc|c} 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & -6 & 2 \\ 0 & 0 & 3 & 2 & -2 \end{array}\right]$$

$$\left[\begin{array}{cccc|c} 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & -6 & 2 \\ 0 & -2 & 1 & 2 & 0 \end{array}\right] \xrightarrow{R_3 \Leftrightarrow R_2} \left[\begin{array}{cccc|c} 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 3 & 2 & -2 \\ 0 & 0 & 0 & -6 & 2 \end{array}\right]$$

$$\left[\begin{array}{cccc|c} 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 3 & 2 & -2 \\ 0 & 0 & 0 & -6 & 2 \end{array}\right] \xrightarrow{R_2 \Rightarrow (1/3)R_2} \left[\begin{array}{cccc|c} 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 1 & 2/3 & -2/3 \\ 0 & 0 & 0 & -6 & 2 \end{array}\right]$$

$$\left[\begin{array}{cccc|c} 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 1 & 2/3 & -2/3 \\ 0 & 0 & 0 & -6 & 2 \end{array}\right] \xrightarrow{R_3 \Rightarrow (-1/6)R_3} \left[\begin{array}{cccc|c} 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 1 & 2/3 & -2/3 \\ 0 & 0 & 0 & 1 & -1/3 \end{array}\right]$$

Now the matrix is in echelon form, but we take it one step further and put it in reduced echelon form:

$$\left[\begin{array}{cccc|c} 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 1 & 2/3 & -2/3 \\ 0 & 0 & 0 & 1 & -1/3 \end{array}\right] \xrightarrow{R_1 \Rightarrow R_1 - R_2} \left[\begin{array}{cccc|c} 0 & 1 & 0 & -2/3 & -1/3 \\ 0 & 0 & 1 & 2/3 & -2/3 \\ 0 & 0 & 0 & 1 & -1/3 \end{array}\right]$$

$$\left[\begin{array}{cccc|c} 0 & 1 & 0 & -2/3 & -1/3 \\ 0 & 0 & 1 & 2/3 & -2/3 \\ 0 & 0 & 0 & 1 & -1/3 \end{array}\right] \xrightarrow{R_1 \Rightarrow R_1 + (2/3)R_3} \left[\begin{array}{cccc|c} 0 & 1 & 0 & 0 & -5/9 \\ 0 & 0 & 1 & 2/3 & -2/3 \\ 0 & 0 & 0 & 1 & -1/3 \end{array}\right]$$

$$\left[\begin{array}{cccc|c} 0 & 1 & 0 & 0 & -5/9 \\ 0 & 0 & 1 & 2/3 & -2/3 \\ 0 & 0 & 0 & 1 & -1/3 \end{array}\right] \xrightarrow{R_2 \Rightarrow R_2 - (2/3)R_3} \left[\begin{array}{cccc|c} 0 & 1 & 0 & 0 & -5/9 \\ 0 & 0 & 1 & 0 & -4/9 \\ 0 & 0 & 0 & 1 & -1/3 \end{array}\right]$$

Now the matrix is in reduced echelon form. We can now write the solution, first we translate this matrix back into equations:

$$x_2 = -5/9, \quad x_3 = -4/9, \quad x_4 = -1/3.$$

Now we write the variables in a vector and use these equations to substitute away x_2, x_3 and x_4, leaving the free variables alone; and we factor the result as a sum of vectors:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_1 \\ -5/9 \\ -4/9 \\ -1/3 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ -5/9 \\ -4/9 \\ -1/3 \end{bmatrix}$$

An important part of this problem is that the first column of the augmented matrix is a column of zeroes, so x_1 is a free variable. Many students make mistakes in their calculations when working with a matrix whose first column is all zeroes, because it is hard to interpret when you think of the corresponding system of equations. Be careful of this common mistake and remember that columns of zeroes give free variables. ■

Example 22. Solve the homogeneous system of equations

$$\begin{aligned} 2x_1 + 3x_2 - x_3 + x_4 + 20x_5 &= 0 \\ -x_1 - x_2 + x_3 + 10x_4 - x_5 &= 0 \end{aligned}$$

Solution. We row reduce the associated augmented matrix.

$$\left[\begin{array}{ccccc|c} 2 & 3 & -1 & 1 & 20 & 0 \\ -1 & -1 & 1 & 10 & -1 & 0 \end{array}\right] \xrightarrow{R_1 \Rightarrow (1/2)R_1} \left[\begin{array}{ccccc|c} 1 & 3/2 & -1/2 & 1/2 & 10 & 0 \\ -1 & -1 & 1 & 10 & -1 & 0 \end{array}\right]$$

$$\left[\begin{array}{ccccc|c} 2 & 3 & -1 & 1 & 20 & 0 \\ -1 & -1 & 1 & 10 & -1 & 0 \end{array}\right] \xrightarrow{R_2 \Rightarrow R_2 - R_1} \left[\begin{array}{ccccc|c} 1 & 3/2 & -1/2 & 1/2 & 10 & 0 \\ 0 & -5/2 & 3/2 & 19/2 & -11 & 0 \end{array}\right]$$

$$\left[\begin{array}{ccccc|c} 1 & 3/2 & -1/2 & 1/2 & 10 & 0 \\ 0 & -5/2 & 3/2 & 19/2 & -11 & 0 \end{array}\right] \xrightarrow{R_2 \Rightarrow -(2/5)R_2} \left[\begin{array}{ccccc|c} 1 & 3/2 & -1/2 & 1/2 & 10 & 0 \\ 0 & 1 & -3/5 & -19/5 & 22/5 & 0 \end{array}\right]$$

$$\left[\begin{array}{ccccc|c} 1 & 3/2 & -1/2 & 1/2 & 10 & 0 \\ 0 & 1 & -3/5 & -19/5 & 22/5 & 0 \end{array}\right] \xrightarrow{R_1 \Rightarrow R_1 - (3/2)R_2} \left[\begin{array}{ccccc|c} 1 & 0 & 2/5 & 31/5 & 17/5 & 0 \\ 0 & 1 & -3/5 & -19/5 & 22/5 & 0 \end{array}\right]$$

In order to get the solution we write out the corresponding equations.

$$\begin{aligned} x_1 &= -(2/5)x_3 - (31/5)x_4 + (17/5)x_5 \\ x_2 &= (3/5)x_3 + (19/5)x_4 + (22/5)x_5 \end{aligned}$$

Finally, we write a vector of variables and substitute away those variables whose columns have leading ones (the non-free variables). In this case, we substitute away the variables x_1 and x_2 using the equations above.

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} -(2/5)x_3 - (31/5)x_4 + (17/5)x_5 \\ (3/5)x_3 + (19/5)x_4 + (22/5)x_5 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}$$

Next we factor it into basic solutions.

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = x_3 \begin{bmatrix} -(2/5) \\ 3/5 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -(31/5) \\ 19/5 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 17/5 \\ 22/5 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

■

2.2 Basic matrix algebra

In this section we introduce matrix algebra. The rules of matrix algebra are similar to the rules of algebra that you may already know, with a few notable differences. Recall that the entries of a matrix A are labeled as $a_{i,j}$.

2.2.1 Adding and subtracting matrices, scalar multiplication

Matrices can be added or subtracted from one another, as long as they have the same size. You add or subtract matrices by adding or subtracting their corresponding entries. The general formula for adding matrices $A = [a_{i,j}]$ and $B = [b_{i,j}]$ which have the same number of rows and columns is

$$A + B = [a_{i,j} + b_{i,j}],$$

or in order to subtract we use the formula

$$A + B = [a_{i,j} - b_{i,j}].$$

Let us show what this means in an example.

Example 23. Suppose that A, B and C are the matrices

$$A = \begin{bmatrix} 1 & -1 \\ 0 & 5 \end{bmatrix}, B = \begin{bmatrix} 4 & -2 & 3 \\ 4 & 5 & 7 \end{bmatrix}, C = \begin{bmatrix} -1 & 6 \\ 4 & 1 \end{bmatrix}.$$

Calculate each of the matrices below, if it is possible.

1. $A + C$
2. $A + B$
3. $C - A$

Solution. We proceed in each case by adding or subtracting corresponding entries:

1. $$A + C = \begin{bmatrix} 1 & -1 \\ 0 & 5 \end{bmatrix} + \begin{bmatrix} -1 & 6 \\ 4 & 1 \end{bmatrix} = \begin{bmatrix} 1 + (-1) & (-1) + 6 \\ 0 + 4 & 5 + 1 \end{bmatrix} = \begin{bmatrix} 0 & 5 \\ 4 & 6 \end{bmatrix}.$$
2. Adding the matrices A and B is not possible, becuase they do not have the same size. The matrix A is 2×2, but the matrix B is 2×3.
3. $$C - A = \begin{bmatrix} -1 & 6 \\ 4 & 1 \end{bmatrix} - \begin{bmatrix} 1 & -1 \\ 0 & 5 \end{bmatrix} = \begin{bmatrix} (-1) - 1 & 6 - (-1) \\ 4 - 0 & 1 - 5 \end{bmatrix} = \begin{bmatrix} -2 & 7 \\ 4 & -4 \end{bmatrix}.$$

■

Matrices can also be multiplied by scalars. If c is any real number and $A = [a_{i,j}]$ is a matrix, then $cA = c[a_{i,j}] = [ca_{i,j}]$. This formula means that in order to multiply a matrix by a scalar c, we multiply each entry by c. Here's an example.

Example 24. If

$$A = \begin{bmatrix} 1 & -1 \\ 0 & 5 \end{bmatrix},$$

what is $5A$?

Solution. We multiply each entry by 5. This gives

$$5A = \begin{bmatrix} 5 & -5 \\ 0 & 25 \end{bmatrix}.$$

■

As a shorthand, we'll write $-A$ in place of the matrix $(-1)A$, so $-A = [-a_{i,j}]$. This way when we add together A and $-A$, we get a matrix of zeroes.

So, with these rules we can treat matrices much like we would treat numbers, as the following example shows. The only exception is that we **cannot** yet divide by a matrix, or multiply by a matrix. These are topics that will be covered in the later sections.

Example 25. Solve for the matrix X if

$$\frac{3}{5}\left(X - 25\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}\right) = 2X - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Solution. The fraction 3/5 multiplies through the brackets to give

$$\frac{3}{5}X - 15\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} = 2X - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and so we rearrange to get

$$\frac{3}{5}X - 2X = -\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + 15\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}.$$

Now

$$-\frac{5}{7}A = \begin{bmatrix} 14 & 15 \\ 0 & -1 \end{bmatrix}$$

so

$$A = -\frac{7}{5}\begin{bmatrix} 14 & 15 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} -10 & -75/7 \\ 0 & 5/7 \end{bmatrix}$$

■

2.2.2 The transpose of a matrix

There is one final basic operation, called transposition. The *transpose* of a matrix A is a new matrix A^T that you create from the matrix A. The entries in the first row of A^T are the same as the entries in the first column of A, the entries of the second row of A^T are the entries of the second column of A, etc. In our shorthand notation, if $A = [a_{i,j}]$ then the transpose is given by the formula $A^T = [a_{j,i}]$. Note that the j and i are switched in the second equation, this indicates that columns change to rows as described in the last two sentences.

Example 26. If

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

what is A^T?

Solution. The first row of A^T has the same entries as the first column of A. So the first row of A^T is $\begin{bmatrix} 2 & 3 \end{bmatrix}$. The second row of A^T is the second column of A, so it's $\begin{bmatrix} -1 & 5 \end{bmatrix}$. The third row of A^T is $\begin{bmatrix} 4 & 8 \end{bmatrix}$. Putting these together

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

■

An important property of the transpose is that if A is an $m \times n$ matrix, then A^T is an $n \times m$ matrix. So, transposition changes the size of a matrix (but in a very predictable way).

Sometimes transposition of a matrix A is though of as 'flipping A along the main diagonal.' To explain this, first we need to know that the *main diagonal* of a matrix A means all the entries $a_{1,1}, a_{2,2}, a_{3,3}, \cdots$ etc. The entries of the main diagonal are bold in the matrix below.

$$A = \begin{bmatrix} \mathbf{a_{1,1}} & a_{1,2} & a_{1,3} & \dots \\ a_{2,1} & \mathbf{a_{2,2}} & a_{2,3} & \dots \\ a_{3,1} & a_{3,2} & \mathbf{a_{3,3}} & \dots \\ \vdots & \vdots & \vdots & \end{bmatrix}$$

Then 'flipping along the main diagonal' is supposed to mean that we switch the entries $a_{1,2}$ and $a_{2,1}$, the entries $a_{1,3}$ and $a_{3,1}$, and so on. Switching all these entries looks a lot like reflecting the matrix across the diagonal:

$$A^T = \begin{bmatrix} \mathbf{a_{1,1}} & a_{2,1} & a_{3,1} & \cdots \\ a_{1,2} & \mathbf{a_{2,2}} & a_{3,2} & \cdots \\ a_{1,3} & a_{2,3} & \mathbf{a_{3,3}} & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix}$$

Because of this idea of 'flipping' or 'reflecting,' a matrix that doesn't change when you take its transpose is called *symmetric*. In equations, A is called a symmetric matrix if $A = A^T$.

2.2.3 Matrix multiplication

Matrix multiplication is a way of taking two matrices A and B, and making a new matrix AB. The formula for multiplying matrices depends on the formula for the dot product of two vectors, so we recall that formula here. Suppose that we have a vector A with n entries

$$A = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

and a vector B with n entries

$$B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

then the dot product $A \cdot B$ is given by the formula

$$A \cdot B = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n.$$

We build on this formula in order to give a rule for multiplying two matrices. Suppose that A is an $m \times n$ matrix

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & a_{2,3} & \cdots & a_{2,n} \\ a_{3,1} & a_{3,2} & a_{3,3} & \cdots & a_{3,n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m,1} & a_{m,2} & a_{m,3} & \cdots & a_{m,n} \end{bmatrix}$$

and write R_i for the i-th row of A. What this means is

$$R_i = \begin{bmatrix} a_{i,1} & a_{i,2} & a_{i,3} & \dots & a_{i,n} \end{bmatrix}$$

and you can think of the matrix A as being built out of the rows $R_1, R_2, \dots, R_m$:

$$A = \begin{bmatrix} R_1 \\ R_2 \\ R_3 \\ \vdots \\ R_m \end{bmatrix}$$

Now take a second matrix B that is size $p \times q$

$$B = \begin{bmatrix} b_{1,1} & b_{1,2} & b_{1,3} & \dots & b_{1,q} \\ b_{2,1} & b_{2,2} & b_{2,3} & \dots & b_{2,q} \\ b_{3,1} & b_{3,2} & b_{3,3} & \dots & b_{3,q} \\ \vdots & \vdots & \vdots & & \vdots \\ b_{p,1} & b_{p,2} & b_{p,3} & \dots & b_{p,q} \end{bmatrix}$$

and write C_j for the j-th column of B. What this means is

$$C_j = \begin{bmatrix} b_{1,j} \\ b_{2,j} \\ b_{3,j} \\ \vdots \\ b_{p,j} \end{bmatrix}$$

and you can think of the matrix B as being built out of the columns $C_1, C_2, \dots, C_q$:

$$B = \begin{bmatrix} C_1 & C_2 & C_3 & \cdots & C_q \end{bmatrix}$$

Now we are ready to describe the matrix AB. The (i, j) entry of the matrix AB is the dot product of row i from matrix A with column j of matrix B, in other words:

$$AB = \begin{bmatrix} R_1 \cdot C_1 & R_1 \cdot C_2 & \dots & R_1 \cdot C_q \\ R_2 \cdot C_1 & R_2 \cdot C_3 & \dots & R_2 \cdot C_q \\ \vdots & \vdots & & \vdots \\ R_m \cdot C_1 & R_m \cdot C_2 & \dots & R_m \cdot C_q \end{bmatrix}$$

Because of this formula, you cannot multiply matrices of certain sizes. This is because the formula is based on the dot product of vectors, and you cannot take the dot product of two vectors that have a different number of entries. So, in order for the matrix multiplication formula to work, every row of A must have the same number of entries as every column of B (since we are dotting rows of A with columns of B). If A is $m \times n$ and B is $p \times q$, this means $m = p$ is required in order for the product AB to be defined. We can also see from the formula that the matrix AB has size $m \times q$.

Example 27. If

$$A = \begin{bmatrix} 2 & 4 \\ 0 & -1 \end{bmatrix} \text{ and } B = \begin{bmatrix} 6 & 7 & 8 \\ 1 & 2 & 3 \end{bmatrix}$$

calculate AB.

Solution. According to the formula,

$$AB = \begin{bmatrix} (2)(6) + (4)(1) & (2)(7) + (4)(2) & (2)(8) + (4)(3) \\ (0)(6) + (-1)(1) & (0)(7) + (-1)(2) & (0)(8) + (-1)(3) \end{bmatrix} = \begin{bmatrix} 16 & 22 & 28 \\ -1 & -2 & -3 \end{bmatrix}$$

■

Example 28. If

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

calculate AB and BA.

Solution. According to the formula,

$$AB = \begin{bmatrix} 1 & -1 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 4 & -2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 4+0 & -2-1 \\ 8+0 & -4+3 \end{bmatrix} = \begin{bmatrix} 4 & -3 \\ 8 & -1 \end{bmatrix}$$

and

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

Observe that AB and BA are different, so in general BA and AB are not the same matrix. Sometimes it is possible to have $AB = BA$ for different matrices A and B, but this is a special occurence. ■

There is also a special matrix I, called the identity matrix

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

which has ones on the diagonal and zeroes everywhere else. If A is any other matrix, then the identity matrix obeys the law

$$AI = IA = A$$

In other words, multiplying A on the right or on the left of I does not change the matrix A. You should check that you believe this claim, by writing out a matrix A and the matrix I and performing the multiplication AI and IA. You will get back the matrix A each time.

2.2.4 Systems of equations and matrix multiplication

Using matrix multiplication, we can rewrite systems of equations in a compact form. Consider the system of equations from Example 20

$$\begin{aligned} x_1 - 2x_2 - x_3 &= 1 \\ 2x_2 - x_3 - 6x_4 &= 0 \\ -x_1 + 2x_3 + 6x_4 &= -1 \end{aligned}$$

and do the following easy trick. Set

$$A = \begin{bmatrix} 1 & -2 & -1 & 0 \\ 0 & 2 & -1 & -6 \\ -1 & 0 & 2 & 6 \end{bmatrix}, X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}.$$

Then the system of equations can be written compactly as $AX = B$. This is because when we do the matrix multiplication for AX, we get

$$AX = \begin{bmatrix} 1 & -2 & -1 & 0 \\ 0 & 2 & -1 & -6 \\ -1 & 0 & 2 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} (1)x_1 + (-2)x_2 + (-1)x_3 + (0)x_4 \\ (0)x_1 + (2)x_2 + (-1)x_3 + (-6)x_4 \\ (-1)x_1 + (0)x_2 + (2)x_3 + (6)x_4 \end{bmatrix}$$

Upon setting this equal to the vector B, we get back our original system of equations.

Because of this compact way of rewriting, instead of being asked to solve a system of equations, you will often be asked to "solve the system $AX = B$ for X." To do this you row reduce the augmented matrix $[A|B]$ and proceed exactly as before.

2.2.5 The inverse of a matrix

Suppose we start with a matrix A. If there is a matrix C so that

$$AC = CA = I$$

then C is called the *inverse of* A, and A is called *invertible*. We write $C = A^{-1}$.

There are restrictions on the size of A if it is invertible. Suppose that A is $m \times n$ and C is $p \times q$. According to the definition, we have to have $AC = CA = I$, and I is a square matrix (has the same number of columns as rows). From $AC = CA$, the sizes of AC and CA must be the same. So $m \times q$ must be the same as $p \times n$, and we get $m = p$ and $q = n$. From $AC = I$ and $CA = I$ we know that both AC and CA are square, so $m = q$ and $p = n$. Now $m = n = p = q$. In other words, an invertible matrix A is square, and A^{-1} is also square.

From this we know that some matrices cannot have inverses since only square matrices can have inverses. However being square is not enough to guarantee that a matrix has an inverse, for example a square matrix with all its entries equal to 0 cannot have an inverse. We will see more complex examples soon, once we learn how to calculate inverses.

Next we explain how to calculate A^{-1}. After explaining the idea, we will give a faster method that uses the ideas we develop here. First, think of calculating A^{-1} as solving the equation $AC = I$ for the matrix C. Write $C_1, C_2, \dots, C_n$ for the columns of the matrix C, and write E_i for the i-th column of I, so

$$E_i = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$

where the 1 is in position i.

Then the formula for matrix multiplication allows us to change $AC = I$ into

$$AC = A\begin{bmatrix} C_1 & C_2 & \cdots & C_n \end{bmatrix} = \begin{bmatrix} AC_1 & AC_2 & \cdots & AC_n \end{bmatrix} = \begin{bmatrix} E_1 & E_2 & \cdots & E_n \end{bmatrix} = I$$

So, in order to solve for the matrix C we need to find each of its columns by solving $AC_i = E_i$ for $i = 1, 2, \dots, n$.

Example 29. Calculate the inverse of

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

Solution. Denote the columns of A^{-1} by C_1 and C_2. To find the first column of A^{-1}, we solve $AC_1 = E_1$, or with $C_1 = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$ we get

$$\begin{bmatrix} 1 & 2 \\ -1 & 4 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

So we row reduce the augmented matrix

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

and find

$$\left[\begin{array}{cc|c} 1 & 0 & 2/3 \\ 0 & 1 & 1/6 \end{array}\right]$$

so that $C_1 = \begin{bmatrix} 2/3 \\ 1/6 \end{bmatrix}$.

Now we solve $AC_2 = E_2$ by row reducing the augmented matrix

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

and get

$$\left[\begin{array}{cc|c} 1 & 0 & -1/3 \\ 0 & 1 & 1/6 \end{array}\right]$$

so $C_2 = \begin{bmatrix} -1/3 \\ 1/6 \end{bmatrix}$. Therefore

$$A^{-1} = \begin{bmatrix} C_1 & C_2 \end{bmatrix} = \begin{bmatrix} 2/3 & -1/3 \\ 1/6 & 1/6 \end{bmatrix}$$

We can check that this matrix is the correct answer, by multiplying A and A^{-1} and checking that we get I. We check and find:

$$AA^{-1} = \begin{bmatrix} 1 & 2 \\ -1 & 4 \end{bmatrix} \begin{bmatrix} 2/3 & -1/3 \\ 1/6 & 1/6 \end{bmatrix} = \begin{bmatrix} 1(2/3) + 2(1/6) & 1(-1/3) + 2(1/6) \\ -1(2/3) + 4(1/6) & -1(-1/3) + 4(1/6) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Similarly we can check that $A^{-1}A = I$. ■

Now we present an algorithm which is more efficient than solving the equation $AC_i = E_i$ many times.

Given a $n \times n$ matrix A, in order to calculate A^{-1} you make an augmented matrix

$$\left[A \mid I\right]$$

That is, you make a matrix whose first n columns are the matrix A, and whose last n columns are the identity matrix I. Now bring the matrix $\left[A \mid I\right]$ to row-reduced echelon form by doing the row reduction algorithm. This is like solving all of the equations $AC_i = E_i$ at the same time. Once $\left[A \mid I\right]$ is row-reduced, there are two possibilities:

1. The row-reduced form of the matrix $\left[A \mid I\right]$ has an identity matrix where the matrix A used to be. In other words, after row reduction $\left[A \mid I\right]$ became a matrix $\left[I \mid C\right]$. In this case, C is A^{-1}.

2. The row-reduced form of the matrix $\left[A \mid I\right]$ does not have an identity matrix where the matrix A used to be. In this case, the matrix A does not have an inverse. We say "A^{-1} does not exist" or "A is not invertible" or "A is singular", which all mean the same thing.

Example 30. Find the inverse of the matrix

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

Solution. Form the augmented matrix

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

and row-reduce this matrix to reduced echelon form. We may assume that $a \neq 0$ in the calculations below, because it were 0, then we would do $R_1 \Leftrightarrow R_2$ to bring a non-zero entry to the top left. Since we assume $a \neq 0$ we can begin by diving the top row by a.

$$\left[\begin{array}{cc|cc} a & b & 1 & 0 \\ c & d & 0 & 1 \end{array}\right] \xrightarrow{R_1 \Rightarrow \frac{1}{a} R_1} \left[\begin{array}{cc|cc} 1 & \frac{b}{a} & \frac{1}{a} & 0 \\ c & d & 0 & 1 \end{array}\right]$$

$$\left[\begin{array}{cc|cc} 1 & \frac{b}{a} & \frac{1}{a} & 0 \\ c & d & 0 & 1 \end{array}\right] \xrightarrow{R_2 \Rightarrow R_2 - cR_1} \left[\begin{array}{cc|cc} 1 & \frac{b}{a} & \frac{1}{a} & 0 \\ 0 & \frac{ad-bc}{a} & \frac{-c}{a} & 1 \end{array}\right]$$

$$\left[\begin{array}{cc|cc} 1 & \frac{b}{a} & \frac{1}{a} & 0 \\ 0 & \frac{ad-bc}{a} & \frac{-c}{a} & 1 \end{array}\right] \xrightarrow{R_2 \Rightarrow \frac{a}{ad-bc} R_2} \left[\begin{array}{cc|cc} 1 & \frac{b}{a} & \frac{1}{a} & 0 \\ 0 & 1 & \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{array}\right]$$

$$\left[\begin{array}{cc|cc} 1 & \frac{b}{a} & \frac{1}{a} & 0 \\ 0 & 1 & \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{array}\right] \xrightarrow{R_2 \Rightarrow R_1 - \frac{b}{a} R_2} \left[\begin{array}{cc|cc} 1 & 0 & \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ 0 & 1 & \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{array}\right]$$

So, according to our algorithm the inverse of

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

is

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

This formula, and the row operations above, do not work if $ad - bc = 0$. If $ad - bc = 0$ then our third row operation is division by zero, which is not allowed. On the other hand, if $ad - bc \neq 0$ then this calculation gives us a formula for the inverse of a 2×2 matrix. ■

Example 31. Find the inverse of the matrix

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

Solution. Create the matrix

$$\left[\begin{array}{ccc|ccc} 1 & 3 & 1 & 1 & 0 & 0 \\ 1 & 1 & 2 & 0 & 1 & 0 \\ 2 & 3 & 4 & 0 & 0 & 1 \end{array}\right]$$

and row reduce it. Its reduced echelon form is

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 2 & 9 & -5 \\ 0 & 1 & 0 & 0 & -2 & 1 \\ 0 & 0 & 1 & -1 & -3 & 2 \end{array}\right]$$

so the inverse of the matrix A is

$$A^{-1} = \begin{bmatrix} 2 & 9 & -5 \\ 0 & -2 & 1 \\ -1 & -3 & 2 \end{bmatrix}$$

■

Example 32. Find the inverse of the matrix

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

Solution. We find the inverse by row reducing the matrix

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

After row reduction, we arrive at

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

Because the first three columns of this matrix did not row reduce to give the identity matrix, the matrix A is not invertible. ■

2.2.6 Solving systems using inverses

Suppose that you want to solve the system $AX = B$. If A is an invertible matrix, there is a very fast way of understanding the solution. By multiplying both sides of the equation $AX = B$ by A^{-1}, you get

$$A^{-1}AX = A^{-1}B$$

and since $A^{-1}A = I$, this gives $IX = A^{-1}B$. Since multiplication by the identity matrix does not change X, this means $X = A^{-1}B$. So when A is invertible, the unique solution to $AX = B$ is $X = A^{-1}B$. A special case that is often highlighted is that when A is invertible, the unique solution to $AX = 0$ is $X = A^{-1}0 = 0$.

Example 33. Solve the system

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

Solution. We rewrite this system as $AX = B$, and we get

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

In the last section we calculated that

$$A^{-1} = \begin{bmatrix} 2 & 9 & -5 \\ 0 & -2 & 1 \\ -1 & -3 & 2 \end{bmatrix}$$

and so

$$X = A^{-1}B = \begin{bmatrix} 2 & 9 & -5 \\ 0 & -2 & 1 \\ -1 & -3 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 5 \end{bmatrix} = \begin{bmatrix} -23 \\ 5 \\ 9 \end{bmatrix}$$

■

2.2.7 Determinants

We saw in the last section that sometimes matrices are invertible, sometimes they are not. The *determinant* of a matrix A is a number that one calculates in order to figure out whether or not A has an inverse, this number is written $\det(A)$. The rule is that a matrix is invertible if and only if its determinant is non-zero.

We already saw in the last section that the inverse of

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

is

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

but this formula only works if $ad - bc \neq 0$. Therefore if A is the 2×2 matrix

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

then $\det(A) = ad - bc$, since A has an inverse exactly when this quantity is nonzero.

Unfortunately for $n \times n$ matrices there is no easy formula when $n > 2$, so aside from the 2×2 case every determinant calculation will be a fair bit of work. The strategy we will use in this case is to break down the determinant calculation for a very big matrix into many calculations done with smaller matrices. Here is how we will get smaller matrices from larger ones:

Given an $n \times n$ matrix A, write $A_{i,j}$ for the $(n-1) \times (n-1)$ matrix one gets from A by deleting row i and column j. So for example, if

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

then

$$A_{1,1} = \begin{bmatrix} 5 & 6 \\ 8 & 9 \end{bmatrix}, A_{1,2} = \begin{bmatrix} 4 & 6 \\ 7 & 9 \end{bmatrix}, A_{1,3} = \begin{bmatrix} 4 & 5 \\ 7 & 8 \end{bmatrix}$$

These are the smaller matrices that will appear in the determinant formula. They appear as part of what is called a *cofactor*. The formula for the (i,j)-cofactor of a matrix A is

$$C_{i,j}(A) = (-1)^{i+j}\det(A_{i,j})$$

Note that we can calculate the cofactors of a 3×3 matrix because the cofactors only contain 2×2 determinants, and we already have a formula for those.

Example 34. Calculate $C_{1,1}(A), C_{1,2}(A)$, and $C_{1,3}(A)$ if

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

Solution. We already know the matrices $A_{1,1}, A_{1,2}, A_{1,3}$ from above. So, we need only apply the formula for the cofactors. We find

$$C_{1,1}(A) = (-1)^{1+1}\det A_{1,1} = (-1)^2 \det \begin{bmatrix} 5 & 6 \\ 8 & 9 \end{bmatrix} = (1)(5\cdot 9 - 8\cdot 6) = -3$$

$$C_{1,2}(A) = (-1)^{1+2}\det A_{1,2} = (-1)^3 \det \begin{bmatrix} 4 & 6 \\ 7 & 9 \end{bmatrix} = (-1)(4\cdot 9 - 7\cdot 6) = 6$$

$$C_{1,3}(A) = (-1)^{1+3}\det A_{1,3} = (-1)^4 \det \begin{bmatrix} 4 & 5 \\ 7 & 8 \end{bmatrix} = (1)(4\cdot 8 - 7\cdot 5) = -3$$

■

The determinant of a matrix A is then calculated from the cofactors of the matrix. The formula is

$$\det(A) = a_{1,1}C_{1,1}(A) + a_{1,2}C_{1,2}(A) + \ldots + a_{1,n}C_{1,n}(A).$$

Example 35. Calculate the determinant of

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

Solution. According to the formula

$$\det(A) = a_{1,1}C_{1,1}(A) + a_{1,2}C_{1,2}(A) + a_{1,3}C_{1,3}(A)$$

For this matrix, we see that $a_{1,1} = 1, a_{1,2} = 2, a_{1,3} = 3$, while the numbers $C_{1,1}(A), C_{1,2}(A)$ and $C_{1,3}(A)$ were all calculated above. So we get

$$\det(A) = (1)(-3) + 2(6) + 3(-3) = 0,$$

so in fact the matrix A is not invertible. ■

We can describe the formula for the determinant as follows: The determinant is the number you get upon multiplying each number in the first row of the matrix by its corresponding cofactor, and then summing the results.

In fact in this description there is nothing special about the first row. You can use any row or column of a matrix to calculate the determinant, so the rule becomes: The determinant of a matrix is the number you get by choosing any row (or column), then multiplying each entry in that row (or column) by its corresponding cofactor and summing the results. This method of determinant calculation is called *cofactor expansion.*

Example 36. Calculate the determinant of

$$A = \begin{bmatrix} 3 & -1 & 5 & -2 \\ 0 & 1 & 2 & 3 \\ 0 & 4 & 5 & 6 \\ 0 & 7 & 8 & 9 \end{bmatrix}$$

Solution. In this case we choose to do cofactor expansion down the first column in order to simplify calculations. That way, the cofactor expansion formula becomes

$$\begin{aligned} \det(A) &= a_{1,1}C_{1,1}(A) + a_{2,1}C_{2,1}(A) + a_{3,1}C_{3,1}(A) + a_{4,1}C_{4,1}(A) \\ &= a_{1,1}C_{1,1}(A) + 0 \cdot C_{2,1}(A) + 0 \cdot C_{3,1}(A) + 0 \cdot C_{4,1}(A) \\ &= a_{1,1}C_{1,1}(A) \\ &= a_{1,1}(-1)^{1+1} \det(A_{1,1}) \end{aligned}$$

However $A_{1,1}$ is the matrix

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

whose determinant we calculated in the previous example, we found that it is 0. So here we have

$$\det(A) = a_{1,1}(-1)^{1+1} \det(A_{1,1}) = 3 \cdot (-1)^2 \cdot 0 = 0$$

■

The lesson one should take away from the last example is that it greatly simplifies matters if you 'aim for zeroes" when doing cofactor expansion. Sometimes this is not possible, and you just have to do the (long) calculation.

Example 37. For what values of the variable h is the matrix

$$A = \begin{bmatrix} 1 & h & 3 \\ 4 & 5 & 6 \\ 0 & h & h \end{bmatrix}$$

invertible?

Solution. The determinant of this matrix is zero if and only if A is not invertible. So, we solve for the values of h which make the determinant equal to zero, those values are the ones which are not allowed. We use cofactor expansion down the first column to take advantage of the zero appearing there. According to the derminant formula

$$\begin{aligned}
\det(A) &= a_{1,1}C_{1,1}(A) + a_{2,1}C_{2,1}(A) + a_{3,1}C_{3,1}(A) \\
&= 1 \cdot (-1)^{1+1} \cdot \det \begin{bmatrix} 5 & 6 \\ h & h \end{bmatrix} + 4 \cdot (-1)^{2+1} \cdot \det \begin{bmatrix} h & 3 \\ h & h \end{bmatrix} + 0 \\
&= 1 \cdot (-1)^2 \cdot (5h - 6h) + 4 \cdot (-1)^3 (h^2 - 3h) \\
&= -4h^2 + 4h
\end{aligned}$$

The determinant is zero when $-4h^2 + 4h = 4h(-h+1) = 0$, which happens when either $h = 0$ or $h = 1$. So, the matrix A is invertible for all values of h except $h = 0$ and $h = 1$. ■

Row operations and determinants: faster calculations

Determinant calculations can evidently take a very long time if the matrices do not have zeroes in them to make it easier. A trick that one can do in order to make things go faster is to take the matrix A, do some row operations to it in order to create a new matrix that has some zeroes, then calculate the determinant. If one chooses clever row operations, the determinant calculation can go much faster after having created some zeroes. The only problem is that by doing row operations to A, you change it into a new matrix whose determinant might be different than the original matrix. Thankfully, every row operation changes the determinant in a predictable way, so as long as you keep track of the changes you can work out the determinant of the original matrix A.

The way that each kind of row operation changes the determinant is (the numbering of the operations and notation are the same as Section 2.1.3):

I. If a matrix A is changed into a matrix B by swapping row i and row j ($R_i \Leftrightarrow R_j$), then $\det(B) = -\det(A)$.

II. If a matrix A is changed into a matrix B by multiplying a row by a nonzero number c ($R_i \Rightarrow cR_i$), then $\det(B) = c\det(A)$.

III. If a matrix A is changed into a matrix B by adding a multiple of one row to another ($R_i \Rightarrow R_i + cR_j$), then $\det(B) = \det(A)$.

Here is an example of how one can use these rules to speed up determinant calculations.

Example 38. Calculate the determinant of

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

Solution. We can do a few steps in the row reduction algorithm to simplify the matrix.

$$\begin{bmatrix} 1 & 1 & -1 & 3 \\ 4 & 4 & 6 & 3 \\ -2 & -2 & 3 & 1 \\ 3 & 1 & -4 & 1 \end{bmatrix} \xrightarrow{R_2 \Rightarrow R_2 - 4R_1} \begin{bmatrix} 1 & 1 & -1 & 3 \\ 0 & 0 & 10 & -9 \\ -2 & -2 & 3 & 1 \\ 3 & 1 & -4 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & -1 & 3 \\ 0 & 0 & 10 & -9 \\ -2 & -2 & 3 & 1 \\ 3 & 1 & -4 & 1 \end{bmatrix} \xrightarrow{R_3 \Rightarrow R_3 + 2R_1} \begin{bmatrix} 1 & 1 & -1 & 3 \\ 0 & 0 & 10 & -9 \\ 0 & 0 & 1 & 7 \\ 3 & 1 & -4 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & -1 & 3 \\ 0 & 0 & 10 & -9 \\ 0 & 0 & 1 & 7 \\ 3 & 1 & -4 & 1 \end{bmatrix} \xrightarrow{R_2 \Rightarrow R_2 - 10R_3} \begin{bmatrix} 1 & 1 & -1 & 3 \\ 0 & 0 & 0 & -79 \\ 0 & 0 & 1 & 7 \\ 3 & 1 & -4 & 1 \end{bmatrix}$$

Note that every row operation we did is of type III, which according to the rules above does not change the determinant. Therefore

$$\det(A) = \det \begin{bmatrix} 1 & 1 & -1 & 3 \\ 4 & 4 & 6 & 3 \\ -2 & -2 & 3 & 1 \\ 3 & 1 & -4 & 1 \end{bmatrix} = \det \begin{bmatrix} 1 & 1 & -1 & 3 \\ 0 & 0 & 0 & -79 \\ 0 & 0 & 1 & 7 \\ 3 & 1 & -4 & 1 \end{bmatrix}$$

And the determinant of the latter matrix is easier to calculate. By doing cofactor expansion across the second row, we get

$$\det \begin{bmatrix} 1 & 1 & -1 & 3 \\ 0 & 0 & 0 & -79 \\ 0 & 0 & 1 & 7 \\ 3 & 1 & -4 & 1 \end{bmatrix} = -79 \cdot (-1)^{2+4} \cdot \det \begin{bmatrix} 1 & 1 & -1 \\ 0 & 0 & 1 \\ 3 & 1 & -4 \end{bmatrix}$$

Again, we can do cofact expansion along the second row of the 3×3 matrix in the equation above, so we get

$$\begin{aligned}
\det \begin{bmatrix} 1 & 1 & -1 & 3 \\ 0 & 0 & 0 & -79 \\ 0 & 0 & 1 & 7 \\ 3 & 1 & -4 & 1 \end{bmatrix} &= -79 \cdot (-1)^{2+4} \cdot \det \begin{bmatrix} 1 & 1 & -1 \\ 0 & 0 & 1 \\ 3 & 1 & -4 \end{bmatrix} \\
&= -79 \cdot (-1)^{2+4} \left(1 \cdot (-1)^{2+3} \cdot \det \begin{bmatrix} 1 & 1 \\ 3 & 1 \end{bmatrix} \right) \\
&= -79 \cdot (1)(1 \cdot (-1) \cdot (1-3)) \\
&= -158
\end{aligned}$$

■

Algebraic properties of the determinant

There are two essential properties of determinants that one uses often. They are

I. If A is a square matrix, then $\det(A) = \det(A^T)$.

II. If A and B are square matrices, then $\det(AB) = \det(A)\det(B)$.

Property I is useful because it allows us to use 'column operations' in order to simplify a matrix before attempting a determinant calculation. This is because all of the row operations listed in the last section become column operations upon taking the transpose, and taking the transpose does not change the determinant. Here is an example. We use the obvious notation '$C_i \Rightarrow C_i + cC_j$' to denote column operations.

Example 39. Calculate the determinant of

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

Solution. We can do a column operation in order to simplify A

$$\begin{bmatrix} 1 & -2 & 2 \\ -2 & 4 & -2 \\ 3 & 7 & 1 \end{bmatrix} \xrightarrow{C_2 \Rightarrow C_2 + 2C_3} \begin{bmatrix} 1 & 0 & 2 \\ -2 & 0 & -2 \\ 3 & 13 & 1 \end{bmatrix}$$

In analogy with rule III for row operations, this column operation does not change the determinant of the matrix. Therefore

$$
\begin{aligned}
\det \begin{bmatrix} 1 & -2 & 2 \\ -2 & 4 & -2 \\ 3 & 7 & 1 \end{bmatrix} &= \begin{bmatrix} 1 & 0 & 2 \\ -2 & 0 & -2 \\ 3 & 13 & 1 \end{bmatrix} \\
&= 13 \cdot (-1)^{3+2} \cdot \det \begin{bmatrix} 1 & 2 \\ -2 & -2 \end{bmatrix} \\
&= 13 \cdot (-1)^{3+2} \cdot (-2 - (-4)) \\
&= -26
\end{aligned}
$$

■

Here is why property II is useful. Often in practical applications, you will find yourself calculating the determinant of a product of many matrices. By multiplying them together first, the product results in a complicated matrix for which cofactor expansion will take a long time. However, by computing determinants of each matrix in the product *before* multiplying, you are often in a position to use zeroes and row/column operation tricks to make the determinant calculations easier.

Example 40. Find the determinant of the product:

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

Solution. Each matrix in the product has a very simple determinant, because each matrix in the product comes from the identity matrix by doing a single row operation. So, we observe that $\det(I) = 1$ (check this!), and then use the row operation rules to find:

I. The first matrix in the product comes from I by doing the row operation $R_3 \Rightarrow R_3 + 3R_1$, so according to rule III above, its determinant is the same as the identity matrix and so is 1.

II. The second matrix in the product above comes from I by doing the row operation $R_2 \Leftrightarrow R_3$, and so according to rule I above its determinant is negative the determinant of I, so it is -1.

III. The third matrix in the product above comes from I by doing the row operation $R_2 \Rightarrow 3R_2$, and so according to rule II above its determinant is 3 times the determinant of I, so it is 3.

All together we have

$$\det(A) = \det\begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\3 & 0 & 1\end{bmatrix}\det\begin{bmatrix}1 & 0 & 0\\0 & 0 & 1\\0 & 1 & 0\end{bmatrix}\det\begin{bmatrix}1 & 0 & 0\\0 & -3 & 0\\0 & 0 & 1\end{bmatrix} = (1)\cdot(-1)\cdot(3) = -3.$$

■

We can also use the relationship between determinants and products to derive new formulas.

Example 41. The determinant of an invertible matrix A is 5. What is the determinant of A^{-1}?

Solution. We can apply the property $\det(AB) = \det(A)\cdot\det(B)$ to the equation $AA^{-1} = I$. Taking determinants of both sides, we get $\det(AA^{-1}) = \det(I) = 1$. Here, we use the fact that the determinant of the identity matrix is 1. As a result, $\det(A)\cdot\det(A^{-1}) = 1$, and so since $\det(A) = 5$, we must have $\det(A^{-1}) = 1/5$. ■

In general, we learn from the last example that $\det(A^{-1}) = \dfrac{1}{\det(A)}$.

Determinants and matrix inverses: the adjoint formula

Since a matrix is invertible if and only if its determinant is not zero, you might expect there to be a formula relating determinants to matrix inverses. Indeed there is such a formula, but we need to introduce a new idea first.

The new idea is the *adjoint* of a matrix A, which is a new matrix that we will call $\mathrm{adj}(A)$. Each entry in the adjoint matrix is a cofactor $C_{i,j}(A)$ of the matrix A:

$$A = \begin{bmatrix}C_{1,1}(A) & C_{2,1}(A) & C_{3,1}(A) & \dots\\C_{1,2}(A) & C_{2,2}(A) & C_{3,2}(A) & \dots\\C_{1,3}(A) & C_{2,3}(A) & C_{3,3}(A) & \dots\\\vdots & \vdots & \vdots & \end{bmatrix}$$

However, note that the (i,j)-cofactor is not the (i,j)-entry, it is the (j,i)-entry. In short, the adjoint is

$$\mathrm{adj}(A) = [C_{i,j}(A)]^T$$

Now the formula that relates the determinant to the inverse of a matrix is:

$$A^{-1} = \frac{1}{\det(A)}\mathrm{adj}(A)$$

Example 42. Using the adjoint formula, calculate the inverse of

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

Solution. We must calculate every one of the cofactors. The calculations are as follows:

$$C_{1,1}(A) = (-1)^{1+1} \det \begin{bmatrix} 0 & 2 \\ 4 & 1 \end{bmatrix} = 1 \cdot (0 - 8) = -8$$

$$C_{1,2}(A) = (-1)^{1+2} \det \begin{bmatrix} 2 & 2 \\ 3 & 1 \end{bmatrix} = (-1) \cdot (2 - 6) = 4$$

$$C_{1,3}(A) = (-1)^{1+3} \det \begin{bmatrix} 2 & 0 \\ 3 & 4 \end{bmatrix} = 1 \cdot (8 - 0) = 8$$

$$C_{2,1}(A) = (-1)^{2+1} \det \begin{bmatrix} 4 & 3 \\ 4 & 1 \end{bmatrix} = (-1) \cdot (4 - 12) = 8$$

$$C_{2,2}(A) = (-1)^{2+2} \det \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} = 1 \cdot (1 - 9) = -8$$

$$C_{2,3}(A) = (-1)^{2+3} \det \begin{bmatrix} 1 & 4 \\ 3 & 4 \end{bmatrix} = (-1) \cdot (4 - 12) = 8$$

$$C_{3,1}(A) = (-1)^{3+1} \det \begin{bmatrix} 4 & 3 \\ 0 & 2 \end{bmatrix} = 1 \cdot (8 - 0) = 8$$

$$C_{3,2}(A) = (-1)^{3+2} \det \begin{bmatrix} 1 & 3 \\ 2 & 2 \end{bmatrix} = (-1) \cdot (2 - 6) = 4$$

$$C_{3,3}(A) = (-1)^{3+3} \det \begin{bmatrix} 1 & 4 \\ 2 & 0 \end{bmatrix} = 1 \cdot (0 - 8) = -8$$

We can use these calculations to find the determinant by cofactor expansion down the middle column:

$$\det(A) = a_{1,2}C_{1,2}(A) + a_{2,2}C_{2,2}(A) + a_{3,2}C_{3,2}(A) = 4 \cdot (4) + 0 + 4 \cdot (4) = 32.$$

Plugging all of these numbers into the adjoint formula gives

$$A^{-1} = \frac{1}{32} \begin{bmatrix} -8 & 8 & 8 \\ 4 & -8 & 4 \\ 8 & 8 & -8 \end{bmatrix}$$

■

Cramer's rule

There is also a method of solving a matrix equation $AX = B$ using determinants, as long as A is invertible. The new method lets you solve for one variable at a time, the method is called *Cramer's rule*.

If we are working with an equation $AX = B$, write $A_i(B)$ for the new matrix you create upon replacing column i of A with B. The vector X is

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

Then Cramer's rule says that as long as A is invertible, the formula for x_i is

$$x_i = \frac{\det(A_i(B))}{\det(A)}$$

Example 43. If

$$A = \begin{bmatrix} 1 & 4 & 5 \\ 2 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

solve the equation $AX = B$ for x_3.

Solution. To use Cramer's rule we need to calculate $\det(A)$ and $\det(A_3(B))$. First we cofactor expand along column 3 to find $\det(A_3(B))$

$$\det(A_3(B)) = \det \begin{bmatrix} 1 & 4 & 1 \\ 2 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix} = 1 \cdot (-1)^{1+3} \det \begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix} 0 + 0 = 2$$

and then cofactor expand down the middle column to calculate $\det(A)$.

$$\begin{aligned} \det(A) = \det \begin{bmatrix} 1 & 4 & 5 \\ 2 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} &= 4 \cdot (-1)^{1+2} \cdot \det \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} + 0 + 1 \cdot (-1)^{3+2} \cdot \det \begin{bmatrix} 1 & 5 \\ 2 & 1 \end{bmatrix} \\ &= -4(2-1) - (1-10) \\ &= 5 \end{aligned}$$

Therefore according to Cramer's rule, $x_3 = 2/5$. ■

2.3 Eigenvalues and diagonalizing

When the number of columns of a matrix A is equal to the number of entries in a vector $\mathbf{v}$, they can be multiplied in order to get a new vector $A\mathbf{v}$. Sometimes the new vector $A\mathbf{v}$ is not new at all, but instead it is just another copy of $\mathbf{v}$ that has been stretched by a certain amount. In equations, we would write

$$A\mathbf{v} = \lambda\mathbf{v}$$

where λ indicates the "stretch factor". This is exactly what it means to be an eigenvector: A nonzero vector v is called an *eigenvector* of the matrix A if there is a number λ such that $A\mathbf{v} = \lambda\mathbf{v}$. The number λ is called the associated *eigenvalue*.

Some matrices have eigenvectors, and others do not. For example, only square matrices can have eigenvectors, because $A\mathbf{v} = \lambda\mathbf{v}$ forces the number of columns and the number of rows of A to both equal the number of entries in the vector $\mathbf{v}$. Even if A is a square matrix it might not have any eigenvalues.[1] In this section we'll learn how to find the eigenvalues and eigenvectors associated to a matrix, when they exist.

In general, the way you calculate the eigenvalues and eigenvectors of a square matrix is to find the eigenvalues first. So we'll start with that.

2.3.1 Calculating eigenvalues

In order to find an eigenvalue and eigenvector, we must find a nonzero solution to the equation

$$A\mathbf{v} = \lambda\mathbf{v}$$

which we can rewrite as $A\mathbf{v} - \lambda\mathbf{v} = (A - \lambda I)\mathbf{v} = 0$, where I is the identity matrix. So, we will only be able to find an eigenvector and an eigenvalue when there's a nonzero solution to the equation $(A - \lambda I)\mathbf{v} = 0$.

If the matrix $A - \lambda I$ is invertible, then the only solution to $(A - \lambda I)\mathbf{v} = 0$ is the zero vector, as explained in Section 2.2.6. So for $(A - \lambda I)\mathbf{v} = 0$ to have a nonzero solution, we need $A - \lambda I$ to be noninvertible—this happens exactly when $\det(A - \lambda I) = 0$. So we solve this equation to find the eigenvalues of a matrix A.

Example 44. If

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

find the eigenvalues of A.

[1]If we allow ourselves to use complex numbers, then every square matrix has at least one eigenvector which possibly contains complex numbers. However, we are only considering real numbers.

Solution. The matrix $A - \lambda I$ is

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

We calculate the determinant of this matrix using the formula for 2×2 matrices:

$$\det(A - \lambda I) = \det \begin{bmatrix} 2-\lambda & 4 \\ 5 & 1-\lambda \end{bmatrix} = (2-\lambda)(1-\lambda) - 20 = \lambda^2 - 3\lambda - 18$$

Setting this expression equal to zero and factoring, we get $(\lambda - 6)(\lambda + 3) = 0$, so the eigenvalues of A are -3 and 6.

■

Example 45. If

$$A = \begin{bmatrix} 0 & 1 & 5 \\ 1 & 3 & 1 \\ 5 & 1 & 0 \end{bmatrix}$$

find the eigenvalues of A.

Solution. The matrix $A - \lambda I$ is

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

We calculate the determinant of this matrix by doing cofactor expansion down the first column:

$$\begin{aligned} \det(A - \lambda I) &= \det \begin{bmatrix} -\lambda & 1 & 5 \\ 1 & 3-\lambda & 1 \\ 5 & 1 & -\lambda \end{bmatrix} \\ &= (-\lambda) \cdot (-1)^{1+1} \cdot \det \begin{bmatrix} 3-\lambda & 1 \\ 1 & -\lambda \end{bmatrix} + 1 \cdot (-1)^{2+1} \cdot \det \begin{bmatrix} 1 & 5 \\ 1 & -\lambda \end{bmatrix} \\ &\qquad + 5 \cdot (-1)^{3+1} \cdot \det \begin{bmatrix} 1 & 5 \\ 3-\lambda & 1 \end{bmatrix} \\ &= (-\lambda)((3-\lambda)(-\lambda) - 1) - ((-\lambda) - 5) + 5(1 - 5(3-\lambda))) \\ &= (-\lambda^3 + 3\lambda^2 + \lambda) + (\lambda + 5) + (-70 + 25\lambda) \\ &= -\lambda^3 + 3\lambda^2 + 27\lambda - 65 \end{aligned}$$

Now we solve $-\lambda^3 + 3\lambda^2 + 27\lambda - 65 = 0$. If the roots of this polynomial are integers, then they will divide the number 65. [2] So we plug the numbers which divide 65 into our polynomial: $\pm 1, \pm 5, \pm 13$ and ± 65. When we plug in $\lambda = -5$, we get zero, so it is a root. Then we can factor, using polynomial long division:

$$-\lambda^3 + 3\lambda^2 + 27\lambda - 65 = -(\lambda + 5)(\lambda^2 - 8\lambda + 13) = 0$$

and use the quadratic equation on $\lambda^2 - 8\lambda + 13 = 0$ to find its roots: $4 \pm \sqrt{3}$. So, the eigenvalues of this matrix are $5, 4 + \sqrt{3}$ and $4 - \sqrt{3}$. ■

There are two important facts to remember when solving for eigenvalues: First, remember that taking a determinant can sometimes be simpler if one does a few cleverly chosen row or column operations first. Second, once you have a polynomial in λ and you set it equal to zero, remember that you may not find as many real solutions as you expect—sometimes none. These two cases are illustrated in the examples below.

Example 46. If

$$A = \begin{bmatrix} 0 & 1 & 5 \\ 1 & 3 & 1 \\ 5 & 1 & 0 \end{bmatrix}$$

find the eigenvalues of A, using column operations to simply the process. (This is the same matrix as the previous example, but we will use a different method for the sake of comparison).

Solution. Instead of cofactor expansion of $A - \lambda I$, we first do the column operation $C_1 \Rightarrow C_1 - C_3$ on $A - \lambda I$. Then the matrix $A - \lambda I$ then becomes

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

Now if we do cofactor expansion down the first column, we get

$$\begin{aligned} \det \begin{bmatrix} -\lambda - 5 & 1 & 5 \\ 0 & 3 - \lambda & 1 \\ 5 + \lambda & 1 & -\lambda \end{bmatrix} &= (-\lambda - 5) \cdot (-1)^{1+1} \cdot \det \begin{bmatrix} 3 - \lambda & 1 \\ 1 & -\lambda \end{bmatrix} \\ &\quad + 0 \cdot (-1)^{2+1} \cdot \det \begin{bmatrix} 1 & 5 \\ 1 & -\lambda \end{bmatrix} + (5 + \lambda) \cdot (-1)^{3+1} \cdot \det \begin{bmatrix} 1 & 5 \\ 3 - \lambda & 1 \end{bmatrix} \\ &= (-\lambda - 5)((3 - \lambda)(-\lambda) - 1) + (\lambda + 5)(1 - 5(3 - \lambda))) \\ &= (\lambda + 5)(-\lambda^2 + 3\lambda + 1) + (\lambda + 5)(-14 + 5\lambda) \\ &= (\lambda + 5)(-\lambda^2 + 8\lambda - 13) \end{aligned}$$

[2] This is true in general if you have integer coefficients and integer roots: the roots of a polynomial with leading coefficient ± 1 will divide the constant term, so test its divisors!

From here, we use the quadratic equation as in the previous example. However, note that our clever choice of column operation saved us from having to factor a cubic equation.

■

Example 47. Here is an example of a matrix which has no eigenvalues. If

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

find the eigenvalues of A.

Solution. The matrix $A - \lambda I$ is

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

We calculate the determinant of this matrix using the formula for 2×2 matrices:

$$\det(A - \lambda I) = (1-\lambda)(-1-\lambda) - (-6) = \lambda^2 + 5$$

Setting this expression equal to zero gives us $\lambda = \sqrt{-5}$, and so we cannot solve for *any* eigenvalues of A in this case (unless we allow complex numbers, a topic not covered in these notes).

■

2.3.2 Solving for eigenvectors

We started Section 2.3 by introducing the equation $A\mathbf{v} = \lambda\mathbf{v}$. Then we studied how to find the possible values of λ in this equation. Now that we know how to find the possible values of λ, called eigenvalues, we will find the possible values of $\mathbf{v}$, the eigenvectors.

In order to do this, we start with a list of all eigenvalues. Say you have a matrix A and you found eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ by solving $\det(A - \lambda A)$. Then for each eigenvalue λ_i, we need to solve

$$A\mathbf{v} = \lambda_i\mathbf{v}$$

to find nonzero solutions for $\mathbf{v}$ (remember, $\mathbf{v}$ is not an eigenvector if it is zero). However, since $\lambda_i\mathbf{v} = \lambda_i I\mathbf{v}$, this is the same as solving $(A - \lambda_i I)\mathbf{v} = 0$ for $\mathbf{v}$. So we need only solve $(A - \lambda_i I)\mathbf{v} = 0$ for $\mathbf{v}$ in order to find the eigenvectors that go with λ_i. Solving equations like this is something we already covered, and it can be done with row reduction.

Example 48. If

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

find the eigenvalues and corresponding eigenvectors of A.

Solution. In the last section we already saw that this matrix has two eigenvalues, -3 and 6. Name the eigenvalues, so that $\lambda_1 = -3$ and $\lambda_2 = 6$. Now to find the eigenvector that goes with $\lambda_1 = -3$, we solve

$$(A - (-3)I)\mathbf{v} = 0.$$

Plugging in, we have

$$\left(\begin{bmatrix} 2 & 4 \\ 5 & 1 \end{bmatrix} + 3 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \mathbf{v} = \begin{bmatrix} 5 & 4 \\ 5 & 4 \end{bmatrix} \mathbf{v} = 0$$

This corresponds to the system whose augmented matrix is

$$\left[\begin{array}{cc|c} 5 & 4 & 0 \\ 5 & 4 & 0 \end{array}\right].$$

We solve this system by following the usual steps, and find $\mathbf{v} = x_2 \begin{bmatrix} -4/5 \\ 1 \end{bmatrix}$. This means that any multiple of the vector $\begin{bmatrix} -4/5 \\ 1 \end{bmatrix}$ will do as an eigenvector corresponding to $\lambda_1 = -3$. For example, if we want to avoid messy fractions we can scale by 5 and take $\begin{bmatrix} -4 \\ 5 \end{bmatrix}$ to be our eigenvector. Then we say that $\begin{bmatrix} -4 \\ 5 \end{bmatrix}$ is the eigenvector associated to $\lambda_1 = -3$ (note we do not keep the parameter x_2).

We do the same for $\lambda_2 = 6$. We solve $(A - 6I)\mathbf{v} = 0$, and plugging in we find

$$\left(\begin{bmatrix} 2 & 4 \\ 5 & 1 \end{bmatrix} - 6 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \mathbf{v} = \begin{bmatrix} -4 & 4 \\ 5 & -5 \end{bmatrix} \mathbf{v} = 0$$

This corresponds to the system whose augmented matrix is

$$\left[\begin{array}{cc|c} -4 & 4 & 0 \\ 5 & -5 & 0 \end{array}\right].$$

Solving by the usual steps, $\mathbf{v} = x_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. So any multiple of the vector $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ will do as an eigenvector corresponding to $\lambda_2 = 6$, and we say that $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is the eigenvector associated to $\lambda_2 = 6$. ■

Notice that in the last example, our first eigenvector initially had a fraction that we were able to eliminate by scaling. In many eigenvector/eigenvalue problems, you should always try taking multiples of your eigenvectors if it will help make your calculations easier. In particular, if you are trying to follow worked examples online or from textbooks, your eigenvectors may differ from the given eigenvectors by a scalar, and this is completely normal.

Example 49. If

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

find the eigenvalues and eigenvectors of A.

Solution. First, we need to find the eigenvalues of A in order to move on to the eigenvectors. So we solve

$$\begin{aligned}
\det \begin{bmatrix} -1-\lambda & 1 & -1 \\ 2 & 1-\lambda & 2 \\ 2 & 1 & 2-\lambda \end{bmatrix} &= (-\lambda - 1) \cdot \det \begin{bmatrix} 1-\lambda & 2 \\ 1 & 2-\lambda \end{bmatrix} \\
&\quad - 2 \cdot \det \begin{bmatrix} 1 & -1 \\ 1 & 2-\lambda \end{bmatrix} + 2 \cdot \det \begin{bmatrix} 1 & -1 \\ 1-\lambda & 2 \end{bmatrix} \\
&= (-\lambda - 1)((1-\lambda)(2-\lambda) - 2) - 2((2-\lambda)+1) \\
&\qquad + 2(2 + (1-\lambda)) \\
&= (-\lambda - 1)(\lambda^2 - 3\lambda) \\
&= -\lambda(1+\lambda)(\lambda - 3).
\end{aligned}$$

So the eigenvalues are $\lambda_1 = 0$, $\lambda_2 = -1$ and $\lambda_3 = 3$. Now, to find the corresponding eigenvectors we need to solve $(A - \lambda_i I)\mathbf{v} = 0$ for each λ_i.

For $\lambda_1 = 0$, we find:

$$(A - \lambda_1 I)\mathbf{v} = (A - 0I)\mathbf{v} = A\mathbf{v} = 0.$$

Then $A\mathbf{v} = 0$ corresponds to the augmented matrix

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

and we row reduce to find

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

Solving in the traditional way yields an eigenvector $\mathbf{v} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$.

For $\lambda_2 = -1$, we find:

$$(A - \lambda_2 I)\mathbf{v} = (A - (-1)I)\mathbf{v} = (A + I)\mathbf{v} = 0.$$

Then $(A + I)\mathbf{v} = 0$ corresponds to the augmented matrix

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

and we row reduce to find

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

Solving in the traditional way yields an eigenvector $\mathbf{v} = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$.

For $\lambda_3 = 3$, we find:

$$(A - \lambda_3 I)\mathbf{v} = (A - (3)I)\mathbf{v} = 0.$$

Then $(A - 3I)\mathbf{v} = 0$ corresponds to the augmented matrix

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

and we row reduce to find

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

Solving in the traditional way yields an eigenvector $\mathbf{v} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$. ■

Last, we need to point out an exception that can happen. In the last two examples, we had an $n \times n$ matrix and when solving for eigenvalues, we found n distinct eigenvalues. Sometimes you find fewer eigenvalues than the size of the matrix, like in the example below.

Example 50. If

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

find the eigenvalues and corresponding eigenvectors of A.

Solution. First, we need to find the eigenvalues of A in order to move on to the eigenvectors. So we solve

$$\begin{aligned} \det \begin{bmatrix} 1-\lambda & -1 \\ 1 & 3-\lambda \end{bmatrix} &= (1-\lambda)(3-\lambda)+1 \\ &= \lambda^2 - 4\lambda + 4 \\ &= (\lambda-2)^2 \end{aligned}$$

So there is only one eigevalue $\lambda = 2$, which is repeated twice in the sense that $(\lambda - 2)$ is raised to the power 2.

To find the eigenvectors that go with $\lambda = 2$, we solve $(A - 2I)\mathbf{v} = 0$, which gives

$$\begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix} \mathbf{v} = 0$$

This corresponds to the system whose augmented matrix is

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

We solve this system by following the usual steps, and find an eigenvector of $\mathbf{v} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$. ■

The important difference to notice between the previous example and the ones before it is that we only found one eigenvector, even though the matrix A is 2×2. In the two earlier examples, we found as many eigenvalues and eigenvectors as the size of the matrix. This difference will be very important when we discuss diagonalizing matrices.