Supplementary Notes for Math133

Pete L. Clark

November 23, 2005
Chapter 1

Vector Geometry

1.1 First notions of vector geometry

Standard position: Our geometric notion of a vector in the plane is that of a directed line segment, with an initial point \( A \) and a terminal point \( B \). We view \( v = \overrightarrow{AB} \) as representing the displacement from \( A \) to \( B \). A vector is said to be in standard position if \( A = 0 \); then the directed line segment \( \overrightarrow{OB} \) can be identified with the point in the plane \( B \).

Any vector \( \overrightarrow{AB} \) can be “carried back” to standard position to give the vector \( \overrightarrow{OB} = \overrightarrow{OB} - \overrightarrow{OA} \). We (or at least, I) picture the vector as being an actual arrow made out of wood, so that it has a well-defined length. When we carry its tail back to the origin, we are careful not to turn the arrow, but instead keep it pointing in the same direction.

It is a useful geometric notion to entertain vectors which are not in standard position. E.g., to add the two standard position vectors \( a = \overrightarrow{OA} \) and \( b = \overrightarrow{OB} \), we carry \( b \) without changing its direction so that the tail of \( b \) coincides with the head of \( a \); then the head of the standard position vector \( a + b \) is where the head of the carried vector \( b \) is.

Scaling: if \( v \) is a vector in the plane or in space, and \( c \) is a real number, \( cv \) scales \( v \) — if \( c > 0 \), \( cv \) has the same direction as \( v \) but is \( c \) times as long, while if \( c < 0 \), \( cv \) points in the opposite direction as \( v \) and is \( |c| = -c \) times as long. Because of this, the act of multiplying a vector \( v \) by a number \( c \) is referred to as scalar multiplication.

\( v - w \): Since \(-w\) is obtained by scaling \( w \) by \(-1\), it has the same magnitude as \( w \) but the opposite direction. Thus to get \( v - w \), we flip \( w \) then carry its tail to the head of \( v \).
Another way of saying this is that the vector $v - w$ is the vector starting at the head of $w$ and ending at the head of $v$. A good way of checking that we got the order right is to make sure that $v - w$, when added to $w$, yields $v$.

### 1.2 The dot product

For $n$-dimensional vectors $v = (v_1, \ldots, v_n)$, $w = (w_1, \ldots, w_n)$, the dot product $v \cdot w$ is a real number: its algebraic definition involves multiplying each coordinate of $v$ by the corresponding coordinate of $w$, and adding together all these products:

$$v \cdot w := v_1 w_1 + \ldots + v_n w_n.$$

We should ask: why in the world have we defined an operation? The fact that we have “multiplied” two vectors and gotten a scalar seems weird at first; but we will see that there is great geometric significance to the dot product, whereas nothing useful would come from forming the vector $(v_1 w_1, \ldots, v_n w_n)$.

The answer is the following beautiful geometric formula:

$$v \cdot w = ||v|| ||w|| \cos \theta,$$

where $\theta$ is the angle between $v$ and $w$. In the text, this formula is derived using the law of cosines, but for a moment forget that you ever heard of the law of cosines\(^1\); let’s see how we we would work up to this formula.

Scaling properties of the dot product: We have $(cv) \cdot w = c(v \cdot w)$. Indeed, $(cv) \cdot w = (cv_1, \ldots, cv_n) \cdot (w_1, \ldots, w_n) = cv_1 w_1 + \ldots + cv_n w_n = c(v_1 w_1 + \ldots + v_n w_n) = c(v \cdot w)$. Similarly $v \cdot (cw) = c(v \cdot w)$. So the dot product of two vectors keeps track of their length in the following way: if $v$ is any (nonzero) vector, write $v_u$ for the unit vector in the same direction as $v$: $v_u = \frac{1}{||v||} v$, so $v = ||v|| v_u$. So

$$v \cdot w = (||v|| v_u) \cdot (||w|| w_u) = ||v|| ||w|| (v_u \cdot w_u).$$

Now we just want to see that if $v$ and $w$ are unit vectors – i.e., lying on the unit circle\(^2\) – so suppose that $v = (\cos \alpha, \sin \alpha)$ and $w = (\cos \beta, \sin \beta)$, for some angles $0 \leq \alpha, \beta \leq 2\pi$. Then

$$v \cdot w = \cos \alpha \cos \beta + \sin \alpha \sin \beta = \cos(\alpha - \beta) = \cos(\beta - \alpha),$$

so indeed the dot product of unit vectors is the cosine of the angle between them. Altogether this proves the dot product formula.

---

\(^1\)Not very difficult, was it?

\(^2\)We implicitly assume that $v$ and $w$ are vectors in the plane; if they are in three space, there is still a plane containing both of them, and it is in this plane that the angle between them is measured.
Press release for the dot product: The dot product formula shows that the dot product of two vectors \( v \) and \( w \) takes into account all of the “intrinsic information” between \( v \) and \( w \) – namely, the length of \( v \), the length of \( w \), and their relative position, as measured by the angle in between them. In particular, the dot product of two vectors would not change if we translated, rotated or reflected the coordinate system. Here is an example to drive this point home: 

\[ v = (1, 0), \quad w = (0, 1). \]

Then \( v \cdot w = 0 \), so that \( v \) and \( w \) are perpendicular. Now rotate everything in sight by 45 degrees (or \( \pi/4 \) radians): the vector \( v \) becomes the unit vector in the direction of \((1, 1)\), which we computed in class to be \((\sqrt{2}/2, \sqrt{2}/2)\), and the vector \( w \) becomes the unit vector in the direction of \((-1, 1)\), or \((-\sqrt{2}/2, \sqrt{2}/2)\). The dot product of these new vectors is

\[ (\sqrt{2}/2, \sqrt{2}/2) \cdot (-\sqrt{2}/2, \sqrt{2}/2) = -(\sqrt{2}/2)^2 + (\sqrt{2}/2)^2 = 0, \]

still.

This example can show us why the operation \( v \ast w = (v_1w_1, \ldots, v_nw_n) \) isn’t worth anything. Indeed, \((1, 0) \ast (0, 1) = (0, 0)\), but rotating 45 degrees, we get \((\sqrt{2}/2, \sqrt{2}/2) \ast (-\sqrt{2}/2, \sqrt{2}/2) = -(\sqrt{2}/2)^2 + (\sqrt{2}/2)^2 \) is certainly not zero. Indeed this “weird product” of two vectors will only be zero if at least one of the two vectors has zero \( x \)-coordinate, at least one has zero \( y \)-coordinate and at least one has zero \( z \)-coordinate. This is not capturing anything interesting about \( v \) and \( w \).

(View of things to come: in about a week we will learn about the cross product of two vectors, which is only defined for vectors in \( \mathbb{R}^3 \) – given vectors \( v \) and \( w \), \( v \times w \) is another vector which is perpendicular to both \( v \) and \( w \).)

Projection of one vector onto another: many of the geometric constructions in this course will come down to the concept of projecting one vector \( v \) onto another vector \( u \). What this means: imagine tilting our heads so that the vector \( u \) is horizontal. Then the vector \( v \) has two coordinates, its \( x \)-coordinate – the piece of \( v \) which goes in the direction of \( u \) – and its coordinate in the “\( y \)”-direction, i.e. the direction perpendicular to \( u \). By definition, the projection of \( v \) onto \( u \) is the \( u \)-piece of the vector \( v \) – in particular, it is a vector parallel to \( u \) whose length is somewhere between zero – if \( v \) is perpendicular to \( u \) – and \( v \) itself – if \( v \) is parallel to \( u \). Denoting this projection by \( p \), we draw a triangle and see that the length of \( p \) is \( ||v|| \cos \theta \) – notice that this formula is predicted in words by the previous sentence. So \( p \) is obtained by starting with a unit vector in the direction of \( u \), namely \( u/||u|| \) and multiplying by \( ||v|| \cos \theta \), to get

\[ p = \frac{||v|| \cos \theta}{||u||} u. \]

Recall that \( \cos \theta = \frac{u \cdot v}{||u|| ||v||} \); substituting this in, we get

\[ p = \frac{u \cdot v}{||u||^2} u = \left( \frac{u \cdot v}{u \cdot u} \right) u. \]

I know math teachers are always saying, “Don’t memorize this formula; just know how to derive it when you need to.” This seems patently ridiculous: the
point of a “formula” is that it is epigrammatic enough so that learning it saves time. But this formula you do not want to blindly memorize, since there are two many instances of $u$ and $v$ – you’re liable to screw it up.

I can remember this formula in steps as follows: first, we are getting a vector in the direction of $u$, so we need $c(u, v)u$, where $c$ is some scalar expressed in terms of $u$ and $v$; to get a scalar out of the vectors $u$ and $v$ there must be some dot products involved; one of them had better be $u \cdot v$. Could $p = (u \cdot v)u$? No, because the formula should scale in $v$ but not in $u$ – that is, if I replace $v$ by $cv$, then the length of the projection should also scale by $c$ – this is what happens in the current formula. But the length of the vector $u$ onto which we’re projecting does not matter – we’re just projecting onto the line that it generates – so final formula should be independent of the length of $u$, whereas in the formula $(u \cdot v)u$, if we replace $u \mapsto cu$, we get $(cu \cdot v)(cu) = c^2(u \cdot v)u$.

Exercise 29, p.27: Show that no two of the four diagonals of a cube are perpendicular.

Solution: We may as well take the cube to have sides of unit length, so that the eight vertices are $(0,0,0)$, $(1,0,0)$, $(0,1,0)$, $(1,1,0)$, $(0,0,1)$, $(1,0,1)$, $(0,1,1)$, $(1,1,1)$. The diagonals of the cube are determined by the pairs of vertices all three of whose coordinates are different, so they are:

I. $D_1 = (1,1,1) - (0,0,0) = (1,1,1)$.
II. $D_2 = (0,1,1) - (1,0,0) = (-1,1,1)$.
III. $D_3 = (1,0,1) - (0,1,0) = (1,-1,1)$.
IV. $D_4 = (0,0,1) - (1,1,0) = (-1,-1,1)$.

We are then asked to show that the dot product of any two of these four vectors equals zero. This is certainly no problem: $D_1 \cdot D_2 = (1,1,1) \cdot (-1,1,1) = (1)(-1) + (1)(1) + (1)(1) = -1 + 1 + 1 \neq 0$. At this point, we could compute the other 5 dot products, but it’s more interesting to observe that in all of the dot products we are going to get a sum of three numbers each of which is either $+1$ or $-1$, so indeed we can be sure that the dot product will be an odd integer – so it certainly is not zero!

Extra credit: Are any two of the (8) diagonals of a 4-dimensional cube perpendicular? Are any two of the $(2^{10} = 1024)$ diagonals of an 11-dimensional cube perpendicular? (Hint: 11, like 3, is odd.)

1.3 Worked problems in vector geometry

Remark on these problems: there is a distinction between working out these problems using vector geometry and working out these problems using Cartesian geometry – i.e., by laying down $x$ and $y$ axes and converting to an algebra problem. Instead we want to work “coordinate-free,” which gives prettier formulas.

---

3Where by “you,” I mean “me,” of course.
1) Give a vector description of the point $P$ which is one-third of the way from $A$ to $B$ on $AB$. Generalize.

Solution: The point in question is \( \frac{2}{3}A + \frac{1}{3}B \). In general, if \( 0 \leq c \leq 1 \), then the point which is the 100\(c\) percent of the way from $A$ to $B$ is \( (1-c)A + cB \).

2) Prove that the line segment joining the midpoints of two sides of a triangle is parallel to the third side and half as long. (Refer to Figure 2 on p. 30).

Solution: \( PQ = Q - P = \frac{1}{2}C + \frac{1}{2}B - \frac{1}{2}C - \frac{1}{2}A = \frac{1}{2}(B - A) \), whereas the third side in question is $AB = B - A$. (We see the elegance of the vector method – the algebra is very clean.)

3) Prove that the quadrilateral $PQRS$, whose sides are the midpoints of an arbitrary quadrilateral $ABCD$, is a parallelogram.

Solution: That is, we are supposed to prove that $PS$ and $QR$ are parallel and that $PQ$ and $SR$ are parallel. We’ll do the first one; the second is (exactly) the same only with different letters. So:

\[
PS = S - P = \frac{1}{2}(A + D) - \frac{1}{2}(A + B) = \frac{1}{2}(D - B).
\]

\[
QR = R - Q = \frac{1}{2}(C + D) - \frac{1}{2}(B + C) = \frac{1}{2}(D - B).
\]

(Not only are the opposite sides parallel, but they have the same length.)

4) A median of a triangle is a line segment from a vertex to the midpoint of the opposite side. Prove, using Figure 5, that the medians of any triangle are concurrent.

Solution: As the hint suggests, we will show that the common intersection point of the three medians is the point which is $\frac{2}{3}$ of the way along any one of them. Now the point which is $\frac{2}{3}$ of the way from $A$ to $P = \frac{1}{2}(B + C)$ is by 1 equal to $\frac{1}{3}A + \frac{2}{3}(\frac{1}{2}(B + C)) = \frac{1}{3}(A + B + C)$. Already from the symmetric form of the answer, it’s clear that we’re going to get the same thing with the other two medians. We’ll do one more, leaving the last to you: the point which is $\frac{2}{3}$ of the way from $C$ to $R = \frac{1}{2}(A + B)$ is $\frac{1}{3}C + \frac{2}{3}(\frac{1}{2}(A + B)) = \frac{1}{3}C + \frac{1}{3}(A + B) = \frac{1}{3}(A + B + C)$.

5) An altitude of a triangle is the line segment from a vertex to the opposite side which is perpendicular. Prove that the three altitudes are concurrent, in a point called the orthocenter.

Solution: We know that $\overrightarrow{AH} \cdot \overrightarrow{BC} = \overrightarrow{BH} \cdot \overrightarrow{AC} = 0$. We will also use that, \[
\overrightarrow{BC} = \overrightarrow{BH} + \overrightarrow{HC} = \overrightarrow{AH} + \overrightarrow{HC}, \overrightarrow{KP} \cdot \overrightarrow{CB} = \overrightarrow{HC},
\]
so that \[
\overrightarrow{HC} = \overrightarrow{BC} - \overrightarrow{BH} = \overrightarrow{AC} - \overrightarrow{AH}.
\]
Following the hint, it is enough to show that $\overrightarrow{HC} \cdot \overrightarrow{AB} = 0$. We compute:

$$
\overrightarrow{HC} \cdot \overrightarrow{AB} = \overrightarrow{HC} \cdot (\overrightarrow{AH} + \overrightarrow{HB}) = 
\overrightarrow{AH} \cdot (\overrightarrow{BH} - \overrightarrow{BC}) + \overrightarrow{HB} \cdot (\overrightarrow{AC} - \overrightarrow{AH}) = 
-(\overrightarrow{AH} \cdot \overrightarrow{BH}) - (\overrightarrow{HB} \cdot \overrightarrow{AH}) = -(\overrightarrow{AH} \cdot \overrightarrow{BH}) + (\overrightarrow{AH} \cdot \overrightarrow{BH}) = 0.
$$

6) Prove that the perpendicular bisectors of the three sides of a triangle are concurrent, in a point called the **circumcenter**.

Solution: (Don’t worry, this is more involved than the problems you will be asked to do!) We will show, following the hint, that $\overrightarrow{KR} \cdot \overrightarrow{AB} = 0$. The two pieces of information we know are that (I) $\overrightarrow{KP} \cdot \overrightarrow{CB} = 0$ and that (II) $\overrightarrow{KQ} \cdot \overrightarrow{CA} = 0$. We also have $P = \frac{1}{2}(B + C)$, $Q = \frac{1}{2}(A + C)$, $R = \frac{1}{2}(A + B)$. Using (I), we compute

$$
0 = \overrightarrow{KP} \cdot \overrightarrow{CB} = (P - K) \cdot (B - C) = P \cdot B - P \cdot C - B \cdot K + K \cdot C = 
1/2(B \cdot B) - 1/2(C \cdot C) - K \cdot B + K \cdot C = 0.
$$

Using (II) we compute

$$
0 = \overrightarrow{KQ} \cdot \overrightarrow{CA} = (Q - K) \cdot (A - C) = Q \cdot A - Q \cdot C - K \cdot A + K \cdot C = 
1/2(A \cdot A) - 1/2(C \cdot C) - K \cdot A + K \cdot C = 0.
$$

Subtracting the second equation from the first we get that

$$
1/2(B \cdot B) - 1/2(A \cdot A) - K \cdot B + K \cdot A = 0.
$$

But in fact

$$
\overrightarrow{KR} \cdot \overrightarrow{AB} = (R - K) \cdot (B - A) = R \cdot B - R \cdot A - K \cdot B + K \cdot A = 
1/2(B \cdot B) - 1/2(A \cdot A) - K \cdot B + K \cdot A,
$$

a quantity that we saw above was zero.

7) If $A$ and $B$ are the endpoints of a diameter of a circle, and $C$ is any point on the circle, prove that $\angle ACB$ is a right angle.

Solution: We have $\overrightarrow{AC} \cdot \overrightarrow{BC} = (c - a) \cdot (c - b) = ||c||^2 - (c \cdot b) - (a \cdot c) + (a \cdot b)$. But since $b = -a$, this simplifies to $||c||^2 - ||a||^2$, which is zero since both $||c||$ and $||a||$ are the radius of the circle.

8) Prove that the line segments joining the midpoints of opposite sides of a quadrilateral bisect each other.

Solution: It will be enough to show that the midpoints of $\overrightarrow{PR}$ and of $\overrightarrow{SQ}$ coincide. The midpoint of $\overrightarrow{PR}$ is $1/2(P + R) = 1/2(1/2(A + B) + 1/2(C + D)) = 1/4(A + B + C + D)$. By the symmetrical form of the answer, we are undoubtedly going to get the same answer for the midpoint of $\overrightarrow{SQ}$; I leave it to you to confirm this.

6
1.4 Equations of lines and planes

Our first order of business is to find equations to for lines in $\mathbb{R}^2$, for planes in $\mathbb{R}^3$, and for lines in $\mathbb{R}^3$.

Let $v = (a, b)$ be a (nonzero) vector in $\mathbb{R}^2$. Consider the locus of all points $P = (x, y)$ in the plane such that $v \cdot P = 0$. What is this? It is the set of all points $P$ such that the vector $\overrightarrow{OP}$ is perpendicular to $v$, and this forms a line through the origin: indeed if we take any nonzero vector $P_0 = (x_0, y_0)$ such that $P_0 \cdot v = 0$, then all the points $P$ are obtained by taking (positive, negative and zero) scalar multiples of $P_0$. Algebraically, we are just looking for points $(x, y)$ such that

$$(a, b) \cdot (x, y) = ax + by = 0.$$

This is the familiar equation of a line through the origin.

What about the equation of a line in the plane that does not pass through the origin? Suppose we wanted the locus of all points passing through, say, $(1, 3)$, and perpendicular to $v = (a, b)$. This locus is a line passing through $(1, 3)$. Indeed, if $P = (x, y)$ is any point on the line, we want the vector from $(1, 3)$ to $(x, y)$ to be perpendicular to $v$; this vector is $(x - 1, y - 3)$, so the equation we get is

$$(a, b) \cdot (x - 1, y - 3) = a(x - 1) + b(y - 3) = 0.$$

Another way to say what has been done is that we have just translated the coordinate axes: moving the $x$-axis to the right by 1 and the $y$-axis to the right by 3 so that the point which was called $(1, 3)$ is now the origin. Calling these new coordinates $x'$ and $y'$, we have $x' = x - 1$ and $y' = y - 3$, and the equation of the line in the new coordinates is the same as above: $ax' + by' = 0$.

Notice also that the equation $a(x - 1) + b(y - 3) = 0$ can be simplified to $ax + by = a + 3b = C$, so the equation of a line in the plane perpendicular to $v = (a, b)$ is of the form

$$ax + by = C$$

for some $C$, which will be zero exactly when the line passes through the origin. We can determine the value of $C$ by plugging in any one point that the line passes through.

Suppose we go now to $\mathbb{R}^3$ and ask the same question: let $v = (a, b, c)$ be a vector in $\mathbb{R}^3$: what is the locus of all points $P$ such that $v \cdot P = 0$? It is now a plane in $\mathbb{R}^3$: for instance, if $v = (0, 0, 1)$ the set of points perpendicular to $v$ is the $xy$-plane. But the algebra is exactly the same: if $P = (x, y, z)$, then we get the equation of the plane

$$v \cdot P = (a, b, c) \cdot (x, y, z) = ax + by + cz = 0.$$
We call $v$ the **normal vector** to the plane, and we see that a plane through the origin is determined by its normal vector. We get the equation of a plane passing through some other point $(x_0, y_0, z_0)$ with normal vector $v$ via the same discussion as above: we want the set of all points $(x, y, z)$ such that $(x_0, y_0, z_0)(x, y, z) = (x - x_0, y - y_0, z - z_0) \cdot v = 0$ and we get

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

and in general, this equation looks like

$$ax + by + cz = ax_0 + by_0 + cz_0 = C.$$

We remark that the planes $ax + by + cz = C_1$, $ax + by + cz = C_2$ are parallel – certainly if $C_1 \neq C_2$ there are no common solutions to both equations! – and the equations of any two parallel planes can be put in this form.

Suppose on the other hand that we want the equation of a line in $\mathbb{R}^3$. If we wanted to get the equation using dot products, we would have to use two equations: e.g., consider the $z$-axis, or the set of all scalar multiples of $(0, 0, 1)$. We can specify this as the set of all $(x, y, z)$ which are perpendicular both to the $x$-axis and the $y$-axis, i.e. $(x, y, z) \cdot (1, 0, 0) = 0$ and $(x, y, z) \cdot (0, 1, 0) = 0$: we just get the two equations $x = 0$, $y = 0$, whose geometric meaning is that the $z$-axis is obtained as the intersection of the $yz$-plane ($x = 0$) with the $xz$-plane ($y = 0$). This is not a very efficient way to proceed: indeed, we have already described this line: it is the set of all $t(0, 0, 1) = (0, 0, t)$, i.e. the set of all scalar multiples of a given vector $(0, 0, 1)$.

This immediately generalizes to show that a line in $\mathbb{R}^3$ through the origin is given by starting with a single vector $v$ – the direction – and taking all scalar multiples $\ell = \{tv \mid t \in \mathbb{R}\}$. If we want a line passing instead through the point $(x_0, y_0, z_0)$ with direction $v$, then we just take

$$\ell = (x_0, y_0, z_0) + tv.$$

Let’s check that every vector whose head and tail lie on this line has direction $v$: if the tail is, say, $(x_0, y_0, z_0) + t_1 v$ and the head is, say, $(x_0, y_0, z_0) + t_2 v$, then the vector is

$$(x_0, y_0, z_0) + t_2 v - ((x_0, y_0, z_0) + t_1 v) = (t_2 - t_1)v,$$

which is indeed in the direction of $v$.

Example (two points determine a line): Find the equation of the line connecting the two points $(1, 2, 3)$ and $(-4, 6, 0)$ in $\mathbb{R}^3$. 

8
Solution: The direction of a line is given by taking any vector whose tail and head are both points on the line, so the direction is given by \((-4, 6, 0) - (1, 2, 3) = (-5, 4, -3) = v\). So the line consists of all points of the form

\[(1, 2, 3) + t(-5, 4, -3) = (1 - 5t, 2 + 4t, 3 - 3t).
\]

As a matter of terminology (and little else), we call the left hand side the **vector equation** of the line and the right hand side the **parametric equations** of the line: it is really three equations

\[x = 1 - 5t, \quad y = 2 + 4t, \quad z = 3 - 3t,
\]

and we can think of \(t\), the parameter, as a “time coordinate”: in general an expression of the form \((f(t), g(t), h(t))\) traces out some curve in three-dimensional space whose \(x\), \(y\) and \(z\) coordinates are being given independently (imagine a three-dimensional Etch-A-Sketch wielded with consummate skill); in this case, since all three functions are linear functions of \(t\), we get a straight line.

Three points determine a plane: How can we find the equation of a plane passing through three points in \(\mathbb{R}^3\), say \(P, Q, R\)? Our formula for the equation of a plane requires knowing a normal vector – how do we get this? Inside the plane lie the vectors \(\overrightarrow{PQ} = Q - P\), \(\overrightarrow{PR} = R - P\). Assuming these two vectors are not scalar multiples one another – i.e., assuming \(P, Q\) and \(R\) do not all lie along the same line in \(\mathbb{R}^3\), which we must certainly do – then a little thought shows that the locus of all points perpendicular to both \(\overrightarrow{PQ}\) and \(\overrightarrow{PR}\) is a line. (Indeed, the equations \(\overrightarrow{PQ} \cdot v = 0\) and \(\overrightarrow{PR} \cdot v = 0\) are nonparallel planes, so their intersection is a line.) It would certainly be nice if there were some formula which, given two vectors \(w_1 = (a, b, c),\ w_2 = (x, y, z)\ in \(\mathbb{R}^3\), would magically spit out a vector perpendicular to both \(w_1\) and \(w_2\).

There is! It is called the cross product \(w_1 \times w_2\). Before we define it, let us understand that it can only exist in \(\mathbb{R}^3\), because it is only in \(\mathbb{R}^3\) that the locus of points perpendicular to two vectors is one-dimensional (i.e., \(2 + 1 = 3\)). We just give the formula for the cross-product:

\[(a, b, c) \times (x, y, z) = (bz - cy, -az - cx, ay - bx).
\]

But we should check that it has the advertised property:

\[
((a, b, c) \times (x, y, z)) \cdot (a, b, c) = abz - acy - abz + bcx + acy - bcx = 0,
\]

\[
(a, b, c) \times (x, y, z) = bxz - cxy - axz + cxy + ayz - bxz = 0
\]

okay. We record without proof a formula for the magnitude of the cross product,

\[||w_1 \times w_2|| = ||w_1|| ||w_2|| \sin \theta,
\]

where \(\theta\) is the angle between \(w_1\) and \(w_2\).
Example: Find the equation of the plane passing through the three points
\( P = (1, 0, -1), \ Q = (2, 3, 0), \ R = (-1, 3, 7) \).

Solution: We compute the crossproduct of \( Q - P \) and \( R - P \) to find a normal vector for the plane. So:
\[
Q - P = (1, 3, 1), \quad R - P = (-2, 3, 8) \quad \text{and} \quad (Q - P) \times (R - P) = (3\times 8 - 1\times 3, -(1\times 8 - 1\times (-2), 1\times 3 - 3\times (-2)) = (21, -10, 9) = v.
\]
So the equation of the plane is
\[
v \cdot ((x, y, z) - P) = 0,
\]
or
\[
(21, -10, 9) \cdot (x, y, z) = v \cdot P = (21, -10, 9) \cdot (1, 0, -1),
\]
\[
21x - 10y + 9z = 12.
\]

1.5 Distances

You will be asked to find distances from points to lines, from points to planes, between lines and between (parallel) planes. The one piece of advice which goes for all of these problems is that: the distance will always be minimized by enforcing one or more perpendicularity conditions. That is, you should figure out what dot products to set to zero.

To give the game away, the correct solution will proceed in one of two ways (sometimes both will work, but sometimes not): when you want to define the distance from something to a line (in \( \mathbb{R}^2 \) or \( \mathbb{R}^3 \)), projection onto the line will work. When you want to find the distance from something to a plane, you will almost certainly use the fact that a plane in \( \mathbb{R}^3 \) has a unique normal direction, and that – better still – from the equation of a plane \( ax + by + cz = C \), we can pick off the normal vector immediately as \( n = (a, b, c) \). Finally, remember that a line in \( \mathbb{R}^2 \) and a plane in \( \mathbb{R}^3 \) are cognate entities: they are both given by setting a single dot product equal to zero. But a line in \( \mathbb{R}^3 \) is different, because it is given not by a single equation but by three (parametric) equations.

Example: Find the distance from the point \((x_0, y_0, z_0) = P\) to the plane \( ax + by + cz = C \).

Solution: There will be a unique point \( Q \) in the plane such that \( |PQ| \) is minimal among all points \( Q \) in the plane, and that point will be such that \( PQ \) is perpendicular to the plane. In other words, we want to take the line through \( P \) whose direction is the normal direction of the plane and intersect that line with the plane. But we already know the normal direction: it is \( v = (a, b, c) \). So the line we want is
\[
\ell = (x_0, y_0, z_0) + t(a, b, c) = (x_0 + ta, y_0 + tb, z_0 + tc),
\]
and we plug it into the equation of the plane to find the value of $t$:

$$a(x_0 + ta) + b(y_0 + tb) + c(z_0 + tc) = C,$$

so $t(a^2 + b^2 + c^2) = C - (ax_0 + by_0 + cz_0)$ and

$$t = \frac{C - (ax_0 + by_0 + cz_0)}{a^2 + b^2 + c^2}.$$

So the line segment whose length we want starts at $P$ and ends at $P + t(a, b, c)$, so its length is the length of $t(a, b, c)$, so

$$|t| \sqrt{a^2 + b^2 + c^2} = \frac{|C - (ax_0 + by_0 + cz_0)|}{\sqrt{a^2 + b^2 + c^2}}.$$

Example: Find the distance from the point $P = (x_0, y_0)$ to the line $ax + by = C$.

Solution: The method of the previous example will also work here: namely, we know that there is a unique point $Q$ on the line which minimizes the distance $||PQ||$ and this point will occur when the direction of $PQ$ is the normal direction to the line, namely $n = (a, b)$. So we know that the line joining $P$ and $Q$ is the line $(x_0, y_0) + t(a, b) = (x_0 + at, y_0 + bt)$, and as above we plug the $x$ and $y$-coordinates into the equation $ax + by = C$. The calculation is exactly the same as above, but without a $z$-coordinate, and the final answer is that the minimal distance is

$$\frac{|C - (ax_0 + by_0)|}{\sqrt{a^2 + b^2}}.$$

Example: Find the distance between a point $P$ and the line in $\mathbb{R}^3$ given by $Q_1 + sV$. 

In this case, however, there is another solution: let $Q_1$ be any point on the line, and consider the line segment $PQ_1$. If we project $PQ_1$ onto the line $\ell$, then we will have $PQ_1 = proj_{\ell}PQ_1 + w$, where by construction $w$ will be a line segment joining $P$ to $\ell$ and perpendicular to $\ell$ – its length is what we want. To compute this, we need to pick two points on the line – set $y = 0$, say, to get $Q_1 = (C/a, 0)$ and $x = 0$ to get $Q_2 = (0, C/b)$. Therefore $u = Q_2 - Q_1 = (C/b, -C/a)$ is a point on the line. The vector we are projecting is $v = Q_1 - P = (C/a - x_0, -y_0)$, so theoretically we can use the projection formula. Unfortunately the algebra becomes rather unpleasant, so we will exhibit the method with particular values: say the line is $x + y = 1$ and the point is $P = (-3, 0)$. Then we find that $proj_{\ell}v = (-2, 2)$, so that $v - proj_{\ell}v = (-2, -2)$, whose length is $\sqrt{8}$. Compare with the formula we derived in the last paragraph: $\sqrt{\frac{(1-x)^2}{2}} = 2/\sqrt{2} = 2^{1/2} = \sqrt{2} = \sqrt{8}$, the same answer.

Example: Find the distance between a point $P$ and the line in $\mathbb{R}^3$ given by $Q_1 + sV$. 

11
Solution: Now it is the method of “plugging in the perpendicular” which will not apply, since a line in $\mathbb{R}^3$ does not have a unique normal direction. We must therefore use the method of projection discussed in the previous example.

Example: Find the distance between the parallel planes $ax + by + cz = C_1$ and $ax + by + cz = C_2$.

Solution: Because the planes are parallel, given any point $P_1$ lying in the first plane, the distance from $P_1$ to the second plane gives us what we want. Thus picking any point in the first plane, say $P_1 = (0, 0, C_1/c)$, we are reduced to a previous example. (There is a nice explicit formula, but we leave its derivation to the interested reader.)

Example: Find the distance between the two parallel lines given by parametric equations $P_1 + sv, P_2 + sv$.

Solution: Similarly, we can choose any point on the first line – like $P_1$ – and reduce to finding the distance between $P_1$ and the line $P_2 + sv$, which can be done by projection.

Example: Find the distance between the skew lines $P_1 + sv, P_2 + tw$.

Solution: This is the hardest problem of this type, which does not reduce to the previous ones – we have two non-parallel, non-intersecting lines in $\mathbb{R}^3$: you should take a moment to picture this situation and see that there will be unique points $Q_1 = P_1 + sv$ on the first line and $Q_2 = P_2 + tw$ on the second line such that $Q_1Q_2$ has the shortest length. \(^4\)

To get the solution we need to enforce two perpendicularity conditions: the line segment $Q_1Q_2$ we are looking for is the one which is perpendicular to the first line and to the second line. That is, we need to solve the equations

\[(P_2 + tw - (P_1 + sv)) \cdot v = 0, \quad (P_2 + tw - (P_1 + sv)) \cdot w = 0.\]

In fact, once we give ourselves numerical values for $P_1$, $P_2$, $v$, and $w$, these two equations will reduce to a system of two linear equations in two unknowns. E.g., say $P_1 = (1, 2, 3)$, $P_2 = (2, 0, 1)$, $v = (-1, 1, 1)$, $w = (-2, 3, 7)$. Then the first equation is $(1 + s - 2t, -2 - s + 3t, -2 - s + 7t) \cdot (-1, 1, 1) = 0$, or

\[-3s + 12t = 5.\]

The second equation is

\[(1 - 2t + s, -2 + 3t - s, -2 + 7t - s) \cdot (-2, 3, 7) = 0,\]

\(^4\)Here’s an easy to picture example, an “overpass”: $\ell_1 = O + t(0, 1, 0) = (0, t, 0)$ is the $y$-axis in the $xy$-plane, $\ell_2 = (0, 0, 1) + t(1, 0, 0) = (t, 0, 1)$ is the line which lies one unit above the $x$-axis in the $xy$-plane. Then $(0, 0, 0)$ on $\ell_1$ and $(0, 0, 1)$ on $\ell_2$ are one unit away from each other, and any other pair of points will be farther away.
or

\[-12s + 62t = 22.\]

Thus we have reduced to what we will be studying for most of the rest of the course: solving systems of linear equations. In geometric language, we want the unique intersection point \((s, t)\) of the two nonparallel lines

\[-3s + 12t = 5\]
\[-12s + 62t = 22.\]

Multiplying the first equation by \(-4\) and adding it to the second equation to cancel the \(s\)'s, we get \(14t = 2\), or \(t = 1/7\). Plugging this into the first equation, we get \(-3s + 12/7 = 5\), or \(s = -23/21\). Knowing these values of \(s\) and \(t\), we know that \(\overrightarrow{Q_1Q_2} = (P_2 + tw) - (P_1 + sv) = [-8/21, -10/21, 2/21]\), and its length is \(\sqrt{(-8/21)^2 + (-10/21)^2 + (2/21)^2} = \sqrt{8/21}\).

Remark: The calculations are rather tedious, and you should feel free to use a calculator (as e.g. on the webwork package). When you get to the end, you should check that indeed \(\overrightarrow{Q_1Q_2} \cdot v = \overrightarrow{Q_1Q_2} \cdot w = 0\), so that you know your answer is correct. Good luck!

1.6 Final comments and warnings

Since I’ve never taught linear algebra before, here are a couple of mistakes that I didn’t think to warn against until now:

1.6.1 Direction and sense

When a student brought up the notion of “sense” of a vector in class last week, I had never heard of it, but now I think it’s a good idea to distinguish between direction and sense as follows:

We say that two vectors \(v\) and \(w\) have the same direction if there is a scalar \(t\) such that \(v = tw\). More geometrically, two vectors have the same direction if their scalings generate the same line. On the other hand, if \(v = tw\) with \(t > 0\) we say they have the same sense, whereas if \(v = tw\) with \(t < 0\) we say they have the same direction but opposite sense. This terminology really is a little better: e.g., a plane in \(R^3\) has a unique normal direction, but two different normal senses.

More importantly, we need to be able to test whether two vectors \(v\) and \(w\) point in the same direction. E.g., consider the planes given by equations

\[x - 2y + 5z = 3,\]
\[3x + 10z = 0.\]
Are they parallel? They will be if and only if their normal vectors $\mathbf{n}_1 = [1, -2, 5]$ and $\mathbf{n}_2 = [3, 0, 10]$ have the same direction. What that means is there is some real number $t$ such that

$$[3, 0, 10] = t[1, -2, 5] = [t, -2t, 5t].$$

This is really three equations: $3 = t$, $0 = -2t$, $10 = 5t$. In other words, we need $t$ to be, simultaneously, $3$, $0$ and $2$. This is clearly not possible, so the planes are not parallel.

1.6.2 Dot products

First of all, in the business we say dot product, scalar product and inner product to mean exactly the same thing. Secondly, I cannot stress strongly enough that the dot product $\mathbf{v} \cdot \mathbf{w}$ of two vectors is a scalar. Many of you have asked me what it represents geometrically. The best answer I can give you is the dot product formula, that

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

You should interpret this formula as follows: if you know the lengths of $\mathbf{v}$ and $\mathbf{w}$, then knowing $\mathbf{v} \cdot \mathbf{w}$ is equivalent to knowing the angle $\theta$ in between them. In particular, we can interpret geometrically the statement that $\mathbf{v} \cdot \mathbf{w} > 0$: namely, that the angle between $\mathbf{v}$ and $\mathbf{w}$ is acute (less than $90^\circ$); that $\mathbf{v} \cdot \mathbf{w} = 0$: $\mathbf{v}$ and $\mathbf{w}$ are perpendicular (or one of them is the zero vector); and $\mathbf{v} \cdot \mathbf{w} < 0$: the angle between $\mathbf{v}$ and $\mathbf{w}$ is obtuse (greater than $90^\circ$). But there really is no vector out there somewhere which is the dot product of $\mathbf{v}$ and $\mathbf{w}$—sorry!

In the previous paragraph we lamented the fact that the dot product $\mathbf{v} \cdot \mathbf{w}$ doesn’t satisfy even the most basic property that we would require of a “product operation” on the set of vectors in $\mathbb{R}^n$: namely, that it doesn’t return another vector in $\mathbb{R}^n$. This should make us skeptical that the dot product may not satisfy other “identities” that a product of numbers would. It turns out that the dot product does have some nice properties, namely:

Commutativity: all $\mathbf{v}$ and $\mathbf{w}$, $\mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v}$.

(Indeed, if $\mathbf{v} = (v_1, \ldots, v_n)$, $\mathbf{w} = (w_1, \ldots, w_n)$, then

$v \cdot w = v_1w_1 + \cdots + v_n w_n$)

For all $\mathbf{v}$, $\mathbf{w}$ and any scalar $\mathbf{w}$, $(a\mathbf{v}) \cdot \mathbf{w} = a(\mathbf{v} \cdot \mathbf{w})$:

$a(v_1, \ldots, v_n) \cdot (w_1, \ldots, w_n) = av_1w_1 + \cdots + av_n w_n = a(v_1w_1 + \cdots + v_n w_n) = a(v \cdot w).$

Distributivity: $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = (\mathbf{u} \cdot \mathbf{w}) + (\mathbf{v} \cdot \mathbf{w})$.

$(u_1+ v_1, \ldots, u_n+ v_n) \cdot (w_1, \ldots, w_n) = (u_1+ v_1)w_1 + \cdots + (u_n+ v_n)w_n =$
\[(u_1w_1 + \ldots + u_nw_n) + (v_1w_1 + \ldots + v_nw_n) = (u \cdot w) + (v \cdot w).\]

But there is no cancellation law for the dot product: if \(u \cdot v = u \cdot w\), then – even when \(u \neq 0\) – we certainly do not need to have \(v = w\). Indeed,

\[u \cdot v = u \cdot w(u \cdot v) - (u \cdot w) = 0 \rightarrow u \cdot (v - w) = 0,\]

i.e., what we get is that \(u\) is perpendicular to \(v - w\). For example, if \(u = (0, 1)\), \(v = (1, 1)\) and \(w = (100, 1)\), then \(u \cdot v = 1 = u \cdot w\), but \(v\) and \(w\) are not the same vector – both their magnitudes and directions are different.
Chapter 2

Solving systems of linear equations

2.1 From geometry to algebra

In order to find the distance between two skew lines, we had to solve a system of equations of the form

\[ As + Bt = E \]
\[ Cs + Dt = F. \]

In this chapter, we will study systems of linear equations in any number of variables and develop algebraic techniques for finding (and parameterizing) their solutions.

First we must understand what we mean by a linear equation. A linear equation in the variables \( x_1, x_2, \ldots, x_n \) over the real numbers is precisely an equation of the form

\[ a_1x_1 + a_2x_2 + \ldots + a_nx_n = C, \]

where \( a_1, a_2, a_3, \ldots, C \) are all real numbers. For instance

\[ \frac{3}{4}x + \pi y + \left( \sum_{n=1}^{\infty} \frac{1}{n^3} \right)z = 1.1 \]

is a linear equation in \( x, y \) and \( z \), because all the coefficients are real numbers. The simpler-looking equation

\[ x^2 + yz = 2 \]

is not linear because the variables are multiplied together.
A **linear system** is just several (finitely many!) linear equations in a common set of variables $x_1, \ldots, x_n$. We speak of a system of $m$ equations in $n$ unknowns. For instance, here is a system of 3 equations in 3 unknowns:

\[
\begin{align*}
    x - y - z &= 2 \\
    3x - 3y + 2z &= 16 \\
    2x - y + z &= 9.
\end{align*}
\]

Finally, a **solution** to a linear system in the variables $x_1, \ldots, x_n$ is just an $n$-tuple $(x_1, \ldots, x_n)$ which satisfies each equation in the system.

The most important single question to ask about a linear system is whether or not it has any solutions. A linear system is **consistent** if it has at least one solution; otherwise it is said to be **inconsistent**. This terminology may seem harsh, but consider: we would have no qualms labeling the equation

\[0 = 1\]

as inconsistent. We will see however that every linear system without a solution can be “put into a form” in which one of the equations reads $0 = 1$.

We also speak of the **solution set** of a linear system, which is just the set of all solutions $(x_1, \ldots, x_n)$, viewed as a subset of $\mathbb{R}^n$. The big question is: what can the solution set of a linear system look like?

Let’s start by thinking about linear systems in just two variables $x$ and $y$: any linear equation $ax + by = C$ is just the equation of a line in $\mathbb{R}^2$, so a solution to the linear system

\[
\begin{align*}
    a_1x + b_1y &= C_1, \\
    a_2x + b_2y &= C_2, \\
    \ldots \\
    a_mx + b_my &= C_m
\end{align*}
\]

corresponds to what we get by **intersecting** finitely many lines in the plane. We understand this perfectly well, and we know that the nature of the solution set **usually** depends on how many lines we are intersecting.

**Case 1** ($m < n$): If we have more variables than equations, then here since we have two variables we must only have one equation. In other words, the locus is just some line $ax + by = C$ – in particular, there are infinitely many solutions. But, technically, there is an exception to this: if $a = b = 0$, when our equation is just $0 = C$. There are no variables here: if $C = 0$ the equation is always true and we are forced to admit – rather legalistically – that the solution
locus is the entire plane: it is indeed true that for every point \((x, y)\) in \(R^2\) we have \(0x + 0Y = 0\). If \(C \neq 0\) it is just as clear that there is no solution.

Case 2 \((m = n)\): So we are intersecting two lines in the plane:

\[
\begin{align*}
  a_1x + b_1y &= C_1 \\
  a_2x + b_2y &= C_2.
\end{align*}
\]

Usually two lines in the plane intersect in a unique solution – indeed this happens precisely when the lines are not parallel. In the language of vectors, it does not occur if and only if the (normal) vectors \((a_1, b_1)\) and \((a_2, b_2)\) are scalar multiples of one another. Say that \((a_2, b_2) = t(a_1, b_1)\) for some \(t\). Then we could multiply the first equation through by \(t\) and get

\[
\begin{align*}
  (ta_1 = a_2)x + (tb_1 = b_2)y &= tC_1 \\
  a_2x + b_2y &= C_2.
\end{align*}
\]

and we have two equations whose left hand sides are equal. Then, if \(tC_1 = C_2\), we really have the same equation – i.e., the solution set is a line (unless the line is degenerate!). If \(tC_1 \neq C_2\) we have distinct parallel lines, and there is no solution at all.

To summarize the solution locus to a linear system in two variables and two unknowns: most of the time there is a unique solution, unless the two lines are parallel, in which case there will either be infinitely many solutions – either a line, or in the ridiculous case where \(a_1 = b_1 = C_1 = a_2 = b_2 = C_2 = 0\), the whole plane – or there will be no solutions at all.

Case 3 \((m > n)\): Suppose we have three or more lines in the plane. We now expect that “most of the time” there is no solution: indeed, two of the three lines should intersect in a point, and it is “lucky” if the third line passes through that intersection point. Of course, all the lines could pass through the same point – for instance if \(C_1 = C_2 = C_3 = 0\) all three lines pass through the origin – or, what would be even “luckier” – all three lines could be the same line, in which case a line would be the solution locus, or “luckiest of all” we could have every coefficient in sight being equal to zero, in which case the solution set is the entire plane.

To summarize: In every case, the solution set of a linear system in the plane is empty, a point, a line, or the entire plane – in particular, we never have exactly 2, or 3, or any finite number of solutions greater than 1. When the number of equations is less than the number of variables, it is likely that we will have infinitely many solutions – indeed one linear equation should give a line! When we have two equations in two unknowns, we expect a unique solution, but it is possible that there could be more or less. When we have more than two equations we expect that there will be no solutions, but again if we are “lucky”
there could be a point, a line, or the entire plane as the solution locus.

Let us now – somewhat more briefly – carry out the same discussion with linear systems in three variables, i.e. systems of equations of intersecting planes in $R^3$.

We see:

When we intersect one plane in $R^3$, we expect the solution will be a plane (and it will be unless in the equation $ax + by + cz = C$, $a = b = c = 0$). When we intersect two planes in $R^3$, we expect the solution will be a line, but it could be empty (if the planes are parallel or the equations are degenerate) or a plane (if the planes are the same) or all of $R^3$. An interesting observation: however lucky we get, we cannot make two planes in $R^3$ intersect in a point.

When we intersect three planes in $R^3$ we expect the solution to be a point, because we expect two planes to intersect in a line and we expect a line and a plane to intersect in a point: to see this, consider the condition that a line $P_0 + t(a, b, c) = (x_0 + ta, y_0 + tb, z_0 + tc)$ intersect the $yz$-plane: we need $z_0 + tc = 0$ to have a solution. This will have a solution unless $c = 0$ (and $z_0 \neq 0$), i.e., unless the $z$-component of the direction of the line is zero – this is rare. Again three planes could fail to intersect at all, or they could intersect in a line, or they could all be the same plane, or everything could vanish and we’d get all of $R^3$.

When we intersect more than three planes in $R^3$ we expect to have no solution, but again we could get lucky – e.g., take the $xy$-plane and “spin it around” on the $x$-axis to get a whole circle’s worth of planes all of which intersect in the $x$-axis.

Summary/prediction: If we have a linear system of $m$ equations in $n$ unknowns, then it is natural to divide into three cases: If the number of equations is less than the number of unknowns ($m < n$), then we call the system underdetermined and we expect that we will have infinitely many solutions. Indeed, looking at the linear system

$$x_1 = 0, x_2 = 0, \ldots, x_m = 0$$

in $R^n$, the solution set is all vectors whose first $m$ coordinates are all zero, and whose remaining $n - m$ coordinates are arbitrary, so the solution space can itself be identified with $R^{n-m}$. In general, we expect the solution space to be “$n - m$-dimensional,” although we have not yet attached a precise meaning to this term. Indeed, the idea that the solution space should be $n - m$-dimensional serves us well in the other cases as well: when $n = m$ this means we expect it to be a single point, and when $m > n$, we take the “negative-dimensionality” of the solution space to mean that we expect it to be empty. We call such a system overdetermined.

---

1But we will!
The goals for the next piece of the course are to flesh out the considerations of the last paragraph. Specifically, we will see the following:

- How to determine whether our linear system has any solutions (is consistent) or no solution at all (is inconsistent).
- That the solution set of any consistent linear system is either a single point, or a “linear space” (a line, a plane, ...) with a well-defined dimension \( d \). (In particular, if there is more than one solution there are always infinitely many.)
- We will also learn how to find and parameterize this solution space explicitly.
- In so doing, we will understand what is going on when our expectations are violated: i.e., we will understand what allows a linear system to have a larger or smaller solution set than we expect.

The first step is to let go of the geometry and enter the realm of pure algebra: we write our linear system

\[
\begin{align*}
  x - y - z &= 2 \\
  3x - 3y + 2z &= 16 \\
  2x - y + z &= 9
\end{align*}
\]

as a matrix:

\[
\begin{bmatrix}
  1 & -1 & 1 & \mid & 2 \\
  3 & -3 & 2 & \mid & 16 \\
  2 & -1 & 1 & \mid & 9
\end{bmatrix}
\]

### 2.2 Row reduction of matrices

In this section we will see how the usual operations that we use to simplify a system of linear equations carry over to row operations on the corresponding matrix. For now a matrix is just a rectangular array of numbers, and just by copying down the coefficients we represent our linear system of \( m \) equations in \( n \) unknowns

\[
\begin{align*}
  a_{11}x_1 + \ldots + a_{1n}x_n &= C_1 \\
  a_{21}x_1 + \ldots + a_{2n}x_n &= C_2 \\
  \vdots \\
  a_{m1}x_1 + \ldots + a_{mn}x_n &= C_m
\end{align*}
\]

as a matrix
\[
\begin{bmatrix}
  a_{11} & \ldots & a_{1n} \\
  a_{21} & \ldots & a_{2n} \\
  \vdots \\
  a_{m1} & a_{mn} & | C_m
\end{bmatrix}
\]

with \( m \) rows and \( n + 1 \) columns.

There are three \textbf{elementary row operations} that we can perform on our matrix:

\((O1)\) Interchange any two of the rows.
\((O2)\) Multiply every entry in a row by any nonzero scalar.
\((O3)\) Take any given row, multiply it by a nonzero scalar, and add it to any other row.

Notice that all three of these operations take our linear system to an \textbf{equivalent} one: i.e., they are acceptable manipulations of the equations in the sense that they do not change the solution space. Indeed, \((O2)\) just corresponds to multiplying both sides of one of the equations through by a nonzero scalar, which is certainly permissible, and \((O3)\) corresponds to doing this and then adding two of the equations, which is again permissible, and indeed exactly the sort of thing one does to solve linear systems. The third operation just corresponds to swapping the order of two of our equations, which even more obviously does not change the solution set (it may seem strange that this could be a useful operation, but we will see that it is).

It is very important that each of our elementary row operations is \textbf{invertible}, which means that there is another elementary row operation that exactly undoes it. In fact \((O1)\) undoes itself: if we swap two rows and then swap them again, we get back to where we started. To undo multiplying a row by a nonzero \( c \), we just multiply by \( c^{-1} \).  

\[ R_j \mapsto cR_i + R_j \]

meaning that we replace the \( j \)th row with \( c \) times the \( i \)th row plus the \( j \)th row. The inverse of this is to multiply the \( i \)th row by \( -c \) and add it to the \( j \)th row, indeed:

\[ R_j \overset{O3}{\mapsto} cR_i + R_j \overset{O3}{\mapsto} -cR_i + (cR_i + R_j) = R_j. \]

We say that two matrices (with the same number of rows and columns) are \textbf{row equivalent} if there is a sequence of elementary row operations we can perform on the first matrix to arrive at the second matrix.

To recast the entire business of solving linear systems in terms of matrices, then, we take our linear system, right down the associated matrix (which, again, is

\[ \text{This makes clear why the scalar we multiply by needs to be nonzero!} \]
just obtained by writing the coefficients of the linear system in a rectangular array of numbers), and perform our so-called elementary row operations on our matrix until we get a matrix whose form is simple enough to enable us to write down the solution space of the linear system.

This last point needs elaboration: what kind of form of a matrix are we going for? Roughly, we want a matrix whose entries towards the bottom left are zero. For example, consider the matrix

\[
\begin{bmatrix}
1 & -1 & -1 & 2 \\
0 & 1 & 3 & 5 \\
0 & 0 & 5 & 10 \\
\end{bmatrix}
\]

This corresponds to the linear system

\[
\begin{align*}
x - y - z &= 2 \\
y + 3z &= 5 \\
5z &= 10.
\end{align*}
\]

But this is easy to solve! The value for \( z \) is staring us in the face, and solving for it, we will then get the value for \( y \), and then for \( x \). Indeed, \( z = 2 \), so \( y + 3 \times 2 = 5 \), so \( y = -1 \), so \( x - (-1) - (2) = 2 \), so \( x = 3 \). That is, the unique solution is the point \((3, -1, 2)\).

Here’s another example when back substitution works well:

\[
\begin{bmatrix}
1 & 2 & 3 & 0 \\
0 & 1 & 2 & 1 \\
\end{bmatrix}
\]

Notice that this is a system corresponds to two intersecting planes in \( \mathbb{R}^3 \), so we should expect a one-parameter family of solutions, i.e., a line. The second equation reads

\[ y + 2z = 1. \]

The trick is to let \( z \) be our parameter: i.e., to solve for \( x \) and \( y \) in terms of \( z \). So \( y = 1 - 2z \), and then we plug this into the first equation,

\[ x + 2y + 3z = 0 \]

to get

\[ x + 2(1 - 2z) + 3z = 0, \]

or

\[ x = z - 2. \]

That is, the general solution is all points of the form \((-2 + z, 1 - 2z, z) = (-2, 1, 0) + z(1, -2, 1)\). This is visibly the vector equation of a line in \( \mathbb{R}^3 \).

Definition: The **leading entry** of a row of a matrix is just the first entry of the row which is nonzero. (A row consisting entirely of zeroes has no leading entry.)
Definition (row echelon form): A matrix is said to be in **row echelon form** if it satisfies two conditions: first, any rows which consist entirely of zeros – if any – lie below all other rows of the matrix. Secondly, the leading entry in any row must lie to the left of the leading entries of all rows beneath it. That is, the leading entries must go down and to the right in staircase – or “echelon” formation.

Notice that our two examples above are in row echelon form, and it is for matrices in row echelon form that back substitution leads to parameterizations of the solution set. Luckily for us,

**Theorem 1** Every matrix is row-equivalent to a matrix in row echelon form.

Rather than just knowing this as a true fact, we actually want a procedure for putting a matrix in row echelon form. Here’s one way to do it (the procedure is not unique, because the row echelon form is not unique: e.g. all matrices of the form \[
\begin{pmatrix}
1 & a \\
0 & 1
\end{pmatrix}
\] are row-equivalent to each other and in row echelon form):

We work column by column, starting from the left. First, look at the first column of the matrix: if all entries are zero, then the same will be true for any row-equivalent matrix, and we must move on to the second column. Eventually we will find a column whose entries are not all zero (or our matrix is the zero matrix, which is in row echelon form!): we can mentally push the all-zero columns off to the side and thus assume that the first column has a nonzero entry in some row. Swap that row up to the first row, so that now the upper left hand corner of our matrix (we are ignoring the zero columns on the left of everything) has a nonzero entry in some row. Swap that row up to the first row, so that now the upper left hand corner of our matrix (we are ignoring the zero columns on the left of everything) has a nonzero entry. It is not necessary but convenient to multiply through so that this leading entry is a 1 – so let’s do this. Then we can use this entry to make zero all the entries in the same column below it. Let’s look at an example:

\[
\begin{bmatrix}
0 & 1 & 2 & 3 & | & 0 \\
0 & -3 & 1 & 1 & | & 2 \\
0 & 2 & -1 & 1 & | & 1
\end{bmatrix}
\]

So if we multiply the first row by 3 and add it to the second row, we’ll get a zero entry just below the 1 and similarly we multiply the first row by $-2$ and add it to the third row to kill the 2 two entries below our leading entry. We’ll get:

\[
\begin{bmatrix}
0 & 1 & 2 & 3 & | & 0 \\
0 & 0 & 7 & 10 & | & 2 \\
0 & 0 & -5 & -5 & | & 1
\end{bmatrix}
\]

Now – in general – we mentally cross out the column and row containing the pivot and start again with the smaller matrix that remains. This completes the description of the algorithm, but let’s carry it through in this example: first we make our leading entry 7 a 1 by multiplying the second row through by $1/7$, to get:
If we now just multiply the second row by 5 and add it to the third row, we’ll get a matrix in row echelon form; while we’re at it, we’ll make the final leading entry 1, getting

\[
\begin{bmatrix}
0 & 1 & 2 & 3 & 0 \\
0 & 0 & 1 & 10/7 & 2/7 \\
0 & 0 & -5 & -5 & 1 \\
\end{bmatrix}
\]

Finally, we need to discuss how, in general, to write down the solution space from a matrix in row echelon form. I recommend the following procedure: write above each column (but the last) the name of the corresponding variable, so in our example \(x_1, x_2, x_3, x_4\). Also circle the leading entries in the columns and identify which variables correspond to columns with a leading entry, the \textbf{leading variables} The remaining variables are the \textbf{free variables}, and however many there are will be the number of parameters of the solution space. We solve for the leading variables in terms of the free variables.

Unfortunately our particular ref matrix is not such a great example, because in this case the \(x_1\) variable simply does not appear in any of the equations, so of course it can be arbitrary in the solution space. The remaining three variables can be solved for uniquely: \(z = 17/15, y + (10/7) \cdot (17/15) = 2/7, \) so \(y = -4/3, x + 2 \cdot (-4/3) + 3 \cdot (17/15) = 0, \) so \(x = -11/15. \) So the general solution is \((0, -11/15, -4/3, 17/15) + x_1(1, 0, 0, 0).\)

For a better example, the matrix

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

It can be put in row echelon form as

\[
\begin{bmatrix}
1 & -1 & -1 & 2 & 1 \\
0 & 0 & 1 & -1 & 1 \\
0 & 0 & 0 & 0 & 0 \\
\end{bmatrix}
\]

In this case, we circle the variables \(x_1\) and \(x_3\), the leading variables, and we get the equations

\[
x_3 - x_4 = 1, \quad x_1 - x_2 - x_3 + 2x_4 = 1,
\]

We write \(x_3 = 1 + x_4, \) and \(x_1 = 1 + x_2 + x_3 - 2x_4 = 1 + x_2 + (1 + x_4) - 2x_4 = 2 + x_2 - x_4. \) It follows that the solution set is parameterized as

\[
(2 + x_2 - x_4, x_2, 1 + x_4, x_4) = (2, 0, 1, 0) + x_2(1, 1, 0, 0) + x_4(-1, 0, 1, 1),
\]

so is a plane in \(R^4. \) (Notice that it is surprising that the solution set is two-dimensional, since \(n - m = 4 - 3 = 1. \) The row reduction process makes clear
that the three equations we started with are “equivalent” to just two equations.)

Inconsistent systems: But our system might not have a solution. How can we tell this from the row reduction process? Consider an example of an obviously inconsistent system, e.g.
\begin{align*}
    x + 2y &= 3 \\
    x + 2y &= 4
\end{align*}

The corresponding matrix is
\[
\begin{bmatrix}
    1 & 2 & 3 \\
    1 & 2 & 4
\end{bmatrix}
\]
and by performing $R_2 \mapsto -R_1 + R_2$ we get it in row echelon form, as
\[
\begin{bmatrix}
    1 & 2 & 3 \\
    0 & 0 & 1
\end{bmatrix}
\]
The second row reads $0 = 0x + 0y = 1$, the *ne plus ultra* of inconsistent equations.

In fact this example is typical: a linear system is inconsistent if and only if when put in row echelon form it has a row consisting entirely of zeros except for the last entry, which is nonzero – for any other matrix in row echelon form, back substitution can be used to find at least one solution.

By the way, the method of solving a linear system (or discovering that it has no solution) by putting the matrix in row echelon form and then back substituting to solve for the leading variables in terms of the free variables is called *Gaussian elimination*, after the 19th century German mathematician Karl Friedrich Gauss\(^3\).

Homogeneous systems: Consider the linear system
\begin{align*}
    x - 2y + 3z &= 0 \\
    2x + y + 17z &= 0 \\
    13x - 3/2y + 12z &= 0 \\
    2x - 8y - 101z &= 0 \\
    x - 0y + z &= 0
\end{align*}
There are five equations in three unknowns, so according to our philosophy the system is “unlikely to have a solution.” But there is an obvious solution to the system: $x = y = z = 0$. A system of linear equations in which the constants on

\(^3\)Gauss is considered by many to be the greatest mathematician of all time, his traditional rivals being Newton and Archimedes. But the next time this argument arises at your local bar, try naming Gerd Faltings or Jean-Pierre Serre instead. Among numerous other achievements, these last two have the virtue of still being alive.
the right hand side are all zero is called homogeneous. Homogeneous systems
are always consistent, because setting all the variables equal to zero will give a
solution. The question becomes whether there are any more solutions – in this
case there are not, as you can check by putting the matrix in row echelon form.

2.3 Reduced row echelon form

As we noted above, the row echelon form of a matrix is not uniquely determined
– we gave the example \[
\begin{bmatrix}
1 & a \\
0 & 1
\end{bmatrix}
\]. The way to see that, no matter what \(a\) is,
all these matrices are row-equivalent is just to use the leading entry 1 at the
bottom right to clear out the \(a\) – by multiplying the second row by \(-a\) and
adding it to the first, we get \[
\begin{bmatrix}
1 & 0 \\
0 & 1
\end{bmatrix}
\].

In point of fact, we could have cleared entries above leading entries as well
as below leading entries all along. Together with making sure that the leading
entries are actually 1 (rather than just nonzero), this gives us the notion of
reduced row echelon form. Reduced row echelon form is also nice: the more
zero entries we have, the easier it is to back-substitute. Moreover, we do not
have to worry about the ambiguity of different reduced row echelon forms:

**Theorem 2** *The reduced row echelon form of a matrix is unique.*

In fact we “expect” the reduced row echelon form of a square matrix to be the
matrix in which there is a leading entry, 1, in each row and column (as in our
above example for \(n = 2\)): this is why we expect a linear system with the same
number of variables as equations to have a unique solution.

When we discuss determinants we will be able to understand what is going
on when this expectation is not fulfilled.

The method of going all the way to reduced row echelon form before backsub-
stituting is called **Gauss-Jordan elimination**, after the 19th century “geode-
siast” Wilhelm Jordan⁴.

2.4 An application: (re)discovering the cross prod-
uct

In this section we will give an application of our newfound ability to solve lin-
ear systems in any number of unknowns: namely, we will (re)discover the cross
product.

⁴Although Gauss was certainly capable of putting a matrix in reduced row echelon form
unaided!
Let \( v = (a, b, c) \) and \( w = (d, e, f) \) be vectors in \( \mathbb{R}^3 \). Consider the locus of all vectors \( (x, y, z) \) which are simultaneously perpendicular to both \( v \) and \( w \): that is \[ v \cdot (x, y, z) = w \cdot (x, y, z) = 0. \]

If we write this out, we get a homogeneous system of two equations in three unknowns:

\[
ax + by + cz = 0 \\
dx + ey + fz = 0.
\]

Consider the associated matrix

\[
\begin{bmatrix}
a & b & c & 0. \\ 0 & e - \frac{bd}{a} & f - \frac{cd}{a} & 0.
\end{bmatrix}
\]

We row-reduce, not to reduced row echelon form, but to a row echelon form that avoids denominators. (Actually, we will assume that \( a \) is not zero. Certainly we want to assume that \( v \) is not zero, otherwise there will be at least two parameters in the solution set. So one of \( a, b, c \) is not zero; if it were \( b \) or \( c \), we could perform an entirely similar computation.) Anyway, we leave the first row alone

\[
\begin{bmatrix}
a & b & c & 0. \\ 0 & e - \frac{bd}{a} & f - \frac{cd}{a} & 0.
\end{bmatrix}
\]

To get rid of the denominators, multiply the second row by \( a \):

\[
\begin{bmatrix}
a & b & c & 0. \\ 0 & (ae - bd) & (af - cd) & 0.
\end{bmatrix}
\]

Notice that we are already in row echelon form, and \( z \) is our free variable. we have \((ae - bd)y + (af - cd)z = 0\), so \( y = \frac{-(af - cd)}{ae - bd}z \). Motivated by a solution without denominators, why not take \( z = (ae - bd) \)? Then \( y = -(af - cd) \).

Substituting this back into the first equation, we get

\[
ax + b(-(af - cd)) + c( ae - bd) = 0,
\]

which simplifies to

\[
ax = abf - bcd + bcd - ace = abf - ace, \quad x = bf - ce,
\]

so we have found a solution to the system, namely the vector \((bf - ce, -(af - cd), ae - bd)\). This is precisely our formula for the cross product of \( v \) and \( w \).

### 2.5 Spans, Linear Independence

We now have an algorithm, Gauss(-Jordan) row reduction, that enables us to determine whether a given system of linear equations has a solution, and if it does to write down all solutions in terms of finitely many parameters (namely, one for each nonleading variable). Still, questions remain: we expect a linear system of \( m \) equations in \( n \) unknowns to have \( n - m \) solution parameters (recall that when \( n - m < 0 \), this means we expect there is no solution). But we
have seen that this expectation is not always fulfilled, and we have a right to
understand why!

Let us again look for intuition from the geometry of lines in the plane and
planes in three-dimensional space. We do understand why a linear system
\[ \begin{align*}
ax + by &= C_1 \\
 cx + dy &= C_2
\end{align*} \]
would fail to have a unique solution: this happens precisely when the lines are
parallel (which includes the coincident case, i.e., when we have the same line
twice). In terms of the normal vectors \(n_1 = (a, b),\ n_2 = (c, d)\), this means that
\(n_2 = \alpha n_1\) for some scalar \(\alpha\) – the normal vectors are scalar multiples of
each other. So this is the problem: usually, when we pick two vectors in the plane,
they do not point in the same (or opposite!) direction. If two vectors \(v_1\) and \(v_2\)
in the plane are not scalar multiples of each other, then we can use them
to parameterize the plane, in the sense that every vector \(w\) in the plane can be
written in terms of \(v_1\) and \(v_2\): there exist scalars \(s\) and \(t\) such that
\[ w = sv_1 + tv_2. \]
Indeed, if \(w = (w_1, w_2)\), we are trying to solve
\[ (w_1, w_2) = s(a, b) + t(c, d) = (as + ct, bs + dt) \]
We get one equation by setting the \(x\)-coordinates equal and another by setting
the \(y\)-coordinates equal:
\[ \begin{align*}
as + bt &= w_1 \\
cs + dt &= w_2
\end{align*} \]
and we are right back where we started! Since these lines are not parallel, no
matter what the right-hand side of the equation is, we can solve for unique \(s\)
and \(t\) – in other words, we have shown that any two vectors in the plane which
are not scalar multiples of each other are such that every vector in the plane
can uniquely be written in terms of them. On the other hand, if \(v_2 = \alpha v_1\), then
\[ sv_1 + tv_2 = sv_1 + t\alpha v_1 = (s + t\alpha)v_1, \]
and we only get vectors on the line with direction given by \(v_1\) or by \(v_2\). The problem is that the second vector \(v_2\) can already be written in terms of a scaling of \(v_1\), so it is redundant to give \(v_2\)
when we already have \(v_1\). Looking back in terms of the linear system, this also
explains why \(v_2 = \alpha v_1\) can be the cause of either too many or too few solutions: since (after multiplying through the second equation by \(\alpha\)) the left hand
side of the second equation is the same as the left hand side of the first equation,
we look to the right hand sides: if \(\alpha C_2 = C_1\) then we have the same equation
twice, which does not impose any new restrictions; whereas if \(\alpha C_2 \neq C_1\) then
the second equation imposes a restriction which is inconsistent with the second
equation.
Consider now the case of three planes intersecting in \( \mathbb{R}^3 \), i.e., a system of three equations in \( x, y \) and \( z \). It is equally true that two planes are parallel exactly when their normal vectors \( n_1 \) and \( n_2 \) are scalar multiples of each other. So if indeed we found ourselves with a linear system like

\[
ax + by + cz = C_1 \\
dx + ey + fz = C_2 \\
\alpha dx + \alpha ey + \alpha fz = C_3
\]

we would recognize a “violation of expectations” since the second and third planes are parallel. But starting in three dimensions (and more so in four, five, . . .) things are more subtle: we could fail to have a unique solution even though no two of the planes are parallel, e.g.

\[
x + y + 0z = 1 \\
0x + y + z = 0 \\
x + 2y + z = 2
\]

whose reduced row echelon form matrix is

\[
\begin{bmatrix}
1 & 0 & 0 & 1 \\
0 & 1 & 1 & 0 \\
0 & 0 & 0 & 2
\end{bmatrix}, \text{ an inconsistent system.}
\]

If we look again at the system, we can check that no two of the left hand sides are scalar multiples of each other, but in fact the third left hand side is precisely the sum of the first two, whereas the third right hand side is \textbf{not} just the sum of the first two. Think about it in terms of the normal vectors \( n_1 = (1, 1, 0) \), \( n_2 = (0, 1, 1) \), \( n_3 = (1, 2, 1) \). We have \( n_3 = n_1 + n_2 \): the last normal vector can be written in terms of the other two, which is why the third equation must either be redundant (as would be \( x + 2y + z = 1 \)) or inconsistent (as it is). It would be equally so if the third equation read \( n_3 = sn_1 + tn_2 \) for any scalars \( s \) and \( t \), because we would be imposing a condition that is \textbf{dependent} on the first two conditions; according to the right hand side, it will either be redundant or inconsistent.\(^5\)

---

\(^5\)Imagine a customs agent is inquiring into the contents of your suitcase. “How many shirts do you have in there?” You say 3. “How many books do you have?” You say 2. If he then asks, “What is the total number of shirts and books in your suitcase?” then he is without a doubt wasting your time, since the answer to this question is clearly determined by the previous two questions. If in a fit of pique you reply, “Seventeen,” then – if he is smarter than the average customs agent – he shouldn’t have to look in your bag to realize you are lying. On the other hand, a third question of, “What is twice the number of shirts minus the number of loaded firearms in your bag?” is (bizarre but) as reasonable a method of determining exactly how many of each of these three quantities you possess as any (maybe the extra layer of mathematics will trick you into saying 5, and he will have outsmarted you after all).
If the third normal vector $n_3$ can be written as $sn_1 + tn_2$, it means all three normal vectors lie in a single plane generated by just $n_1$ and $n_2$. If one is adept at picturing three-dimensional geometry\(^6\), you might try to “see” that three planes in $\mathbb{R}^3$ with no common intersection point necessarily have all three normal vectors lying in a single plane.

These considerations have led us inexorably to the notions of linear independence and spanning. The key notion linking both of them is that of linear combination, which we have been using implicitly but should spell out:

A linear combination of the vectors $v_1, \ldots, v_n$ is any vector of the form $a_1v_1 + \ldots + a_nv_n$, where the $a_1, \ldots, a_n$ can be any scalars. Thus the set of all possible linear combinations of a nonzero vector $v$ is the line generated by $v$. The set of linear combinations of two vectors $v_1$ and $v_2$ generates a plane unless $v_2$ is itself a linear combination of $v_1$ – i.e., a scalar multiple.

Definition: A (finite) collection of vectors $v_1, \ldots, v_m$ is linearly dependent if some vector in the set can be written as a linear combination of the other vectors. A collection is said to be linearly independent if it is not linearly dependent, so that no vector in the set can be written as a linear combination of the other vectors.

The notion of linearly dependent vectors is exactly what we need to make our idea of the left hand sides of a linear system being redundant precise. Indeed:

Two vectors $v_1, v_2$ are linearly dependent precisely when each is a scalar multiple of the other.

Three vectors $v_1, v_2, v_3$ are linearly dependent precisely when there is a single plane that contains all of them.

A characterization of linear independence: Suppose we have vectors $v_1, \ldots, v_m \in \mathbb{R}^n$ such that the zero vector $0$ is a linear combination of them. Well, big deal; we can always write

$$0 = 0v_1 + 0v_2 + \ldots + 0v_m.$$  

But now suppose that $0$ is a linear combination of the $v_1, \ldots, v_m$ and that the coefficients are not all zero, e.g.

$$0 = 0v_1 + 2v_2 - 3/2v_3 + v_4.$$  

(Some of the coefficients can be zero, just not all of them.) But this equation is equivalent to

$$-2v_2 = -3/2v_3 + v_4.$$  

\(^6\)Actually one of the principal merits of the algebraic viewpoint is that as the number of spatial dimensions increases, it becomes successively harder to visualize what is going on, but the algebraic methods work regardless of the number of dimensions.
or

\[ v_2 = \frac{3}{4}v_3 + v_4, \]

and one of the vectors has been written as a linear combination of the rest – the set is linearly dependent. The converse is equally valid: if the vectors \( v_1, \ldots, v_m \) are linearly dependent, then one of them – say \( v_m \) – can be written as a linear combination of the rest,

\[ v_m = a_1 v_1 + a_2 v_2 + \ldots + a_{m-1} v_{m-1}. \]

Bringing the \( v_m \) to the other side we get

\[ a_1 v_1 + a_2 v_2 + \ldots + a_{m-1} v_{m-1} - v_m = 0, \]

and at least one of the coefficients – namely the coefficient of \( v_m \) is not zero. In summary, a set of vectors is linearly dependent if and only if the only way of expressing the zero vector as a linear combination of them is by taking all the coefficients to be zero (the “trivial solution”).

Example: Are the vectors \( v_1 = [1, 2, 3], \ v_2 = [0, 3, 7], \ v_3 = [2, 1, -1] \) in \( \mathbb{R}^3 \) linearly independent?

Solution: To find out consider the equation

\[ a_1 v_1 + a_2 v_2 + a_3 v_3 = 0. \]

This is

\[ (a_1 + 2a_3, 2a_1 + 3a_2 + a_3, 3a_1 + 7a_2 - a_3) = (0, 0, 0). \]

Thus, it is equivalent to the homogeneous linear system

\[ a_1 + 0a_2 + 2a_3 = 0. \]
\[ 2a_1 + 3a_2 + a_3 = 0. \]
\[ 3a_1 + 7a_2 - a_3 = 0. \]

The reduced row echelon form is

\[
\begin{bmatrix}
1 & 0 & 2 & 0 \\
0 & 1 & -1 & 0 \\
0 & 0 & 0 & 0 \\
\end{bmatrix}
\]

Since the \( a_3 \)-column does not contain a leading entry, there will be infinitely many solutions, parameterized as \((-2a_3, a_3, a_3)\). So e.g. we find that \(-2v_1 + v_2 + v_3 = 0\) and the vectors are linearly dependent.

Note well what happened: to test whether vectors \( v_1, \ldots, v_m \) are linearly dependent, we are led to consider a homogeneous linear system which, in matrix form, is obtained by taking as the columns of the matrix the vectors \( v_1, v_2, \ldots, v_m \) and the final column 0. Then the vectors will be linearly independent if and
only if this homogeneous system has a nontrivial (not all zero) solution.

Example: Show that four vectors in $\mathbb{R}^3$ are always linearly dependent.

Solution: When we write down the linear system and put it in matrix form, the matrix (not including the final column of zeros) will have three rows and four columns. Since in the best of all possible worlds we could only have one leading entry in each row, whenever there are more columns than rows we must therefore have at least one column without a leading entry. Since homogeneous systems are always consistent, this means we will always have at least a one parameter family of solutions, hence a nonzero one.

Notice that we did not mention 4 or 3 once in the solution. What we actually showed was:

**Proposition 3** *More than n vectors in $\mathbb{R}^n$ are always linearly dependent.*

Remark: A set of vectors containing the zero vector is always linearly dependent: since 

$$1(\overline{0}) + 0v_2 + \ldots + 0v_m = 0.$$  

Surprisingly, this will come in handy later.

### 2.5.1 Spanning sets

Closely related to the notion of linear independence is that of the span of a set of vectors.

Definition: The **span** of a set of vectors $v_1, \ldots, v_m$ in $\mathbb{R}^n$ is the set of all linear combinations of them, i.e., the set of all $a_1v_1 + \ldots + a_m v_m$ as $a_1, \ldots, a_m$ range through all possible values.

So the span of any nonzero vector is a line, and the span of two vectors is usually a plane, unless the two vectors are linearly independent, in which case their span is a line.

Example: Is the vector $w = (3, -4)$ in the span of $v_1 = (1, 2)$ and $v_2 = (0, 5)$?

Solution: All roads lead to row reduction! Indeed, the question is whether we can find scalars $a$ and $b$ such that 

$$av_1 + bv_2 = w,$$

or 

$$(a + 0b, 2a + 5b) = (3, -4),$$

which is equivalent to the linear system 

$$a + 0b = 3$$
\[2a + 5b = -4\]

Note well that the matrix comes out as \([v_1 \ v_2 \ | \ w]\), so \(w\) being in the span is equivalent to the consistency of the linear system, i.e., row reduction will give us the answer. Indeed in this case we know in advance that the answer is yes, since the system corresponds to equations of two non-parallel lines, which always intersect in a unique solution.

Example: Find a vector \(w\) in \(\mathbb{R}^3\) which is not in the span of \(v_1 = (1, 2, 3)\) and \(v_2 = (-1, 1, 1)\).

Solution: Since the linearly independent vectors \(v_1\) and \(v_2\) span a plane in \(\mathbb{R}^3\), most any vector that we pick at random will not lie in the plane, and hence will not be in the span of \(v_1\) and \(v_2\). Let’s proceed algebraically instead: consider an arbitrary vector \(w = (a, b, c) = sv_1 + tv_2\) in the span. What can be said about \(a\), \(b\), \(c\)? The equations we get are
\[\begin{align*}
    s - t &= a \\
    2s + t &= b \\
    3s + t &= c
\end{align*}\]

The corresponding matrix in row echelon form is
\[
\begin{bmatrix}
1 & 0 & a - (2a + b)/3 \\
0 & 1 & (-2a + b)/3 \\
0 & 0 & -3a + c - 4/3(-2a + b)
\end{bmatrix}
\]

We find, therefore, that the system is consistent if and only if
\[-1/3a - 4/3b + c = 0.\]

2.6 A closer look at linear independence, spanning and dimension

Although our motivation for introducing linear independence was to figure out why we have more or fewer solutions than the “expected” \(n - m\), we have not yet seen the full connection between linear independence and the number of parameters of the solution space. What is in fact true is the following result.

**Theorem 4** Consider a linear system
\[\begin{align*}
    a_{11}x_1 + \ldots + a_{1n}x_n &= C_1 \\
    \vdots \\
    a_{m1}x_1 + \ldots + a_{mn}x_n &= C_n.
\end{align*}\]
Suppose that the row vectors \( v_1 = (a_{11}, \ldots, a_{1n}), \ldots, v_n = (a_{m1}, \ldots, a_{mn}) \) are linearly independent. Then the system is consistent, and the solution space has exactly \( n - m \) parameters.

We want to prove this theorem by contemplating a matrix in row echelon form. In order to do this however, we need to be sure that if we start with a linearly independent set of row vectors and perform elementary row operations, then the row vectors remain linearly independent. In fact a similar statement is true about the span: the set of all vectors in \( \mathbb{R}^n \) spanned by the row vectors \( v_1, \ldots, v_m \) of a matrix is (quite reasonably) called the row space of the matrix.

We now have the following result:

**Proposition 5** Let \( M_1 \) and \( M_2 \) be two \( m \times n \) matrices which are row-equivalent (i.e., we can get from one to the other by a sequence of elementary row operations). Then the row spaces of \( M_1 \) and \( M_2 \) are equal, and the row vectors of \( M_1 \) form a linearly independent set if and only if the row vectors of \( M_2 \) form a linearly independent set.

Proof: Since by definition one goes between row-equivalent matrices by a sequence of elementary row operations, it will be enough to show that if we perform any of the three elementary row operations (O1), (O2) or (O3), then we do not change the row space, nor the linear in/dependence of the row vectors. Let’s show both of these properties for each of the three row operations in turn.

Recall that (O1) just swaps two rows of our matrix. But since a linear combination of vectors \( a_1v_1 + \ldots + a_mv_m \) does not depend on the order in which the vectors are written, it is quite clear that switching the order of the two vectors will not change which vectors we can write as linear combinations of our vectors or whether we can write the zero vector as a nontrivial linear combination of our vectors.

The second operation (O2) is just scaling one of the row vectors \( v_i \) by a nonzero constant \( c \), and this is also easily seen to be harmless: if \( w = a_1v_1 + \ldots + a_iv_i + \ldots + a_mv_m \) is in the span of \( v_1, \ldots, v_m \) then it is certainly also in the span of \( cv_1, \ldots, cv_m \), namely \( w = a_1v_1 + \ldots + a_i/c_i(v_i) + \ldots a_mv_m \). Similarly, we can just multiply and divide by \( c \) to see that \( v_1, \ldots, v_m \) is linearly dependent if and only if \( cv_1, \ldots, cv_m, v_m \) is linearly dependent.

The only one which is a little tricky is the third operation, which replaces some vector \( v_j \) by \( cv_i + v_j \) where \( v_i \) is another (row) vector and \( c \) is any nonzero scalar. We want to show first that any vector \( w \) which is in the span of \( v_1, \ldots, v_m \) is also in the span of \( v_1, \ldots, cv_i + v_j, \ldots, v_m \) and conversely. By suitably labeling things (or, if you like, by applying (O1) to swap the rows into the way we want them to be), we can arrange that the row operation replaces the second row \( v_2 \) with \( cv_1 + v_2 \). If

\[
  w = a_1v_1 + a_2v_2 + a_3v_3 + \ldots + a_mv_m
\]

then

\[
  cv_1 + v_2 = a_1v_1 + a_2(v_2 + cv_1) + a_3v_3 + \ldots + a_mv_m
\]

Notice that

\[
  cv_1 + v_2 = a_1v_1 + a_2v_2 + a_3v_3 + \ldots + a_mv_m
\]

which is in the span of \( v_1, \ldots, v_m \).
then \( w - (a_3v_3 + \ldots + a_mv_m) = a_1v_1 + a_2v_2 = (a_1 - ca_2)v_1 + a_2(cv_1 + v_2) \), so
\[
 w = (a_1 - ca_2)v_1 + a_2(cv_1 + v_2) + a_3v_3 + \ldots + a_mv_m \quad (2.2)
\]
is now seen to be in the span of the vectors \( v_1, (cv_1 + v_2), v_3, \ldots, v_m \). Similarly, if
\[
 w = a_1v_1 + a_2(cv_1 + v_2) + a_3v_3 + \ldots + a_mv_m \quad (2.3)
\]
is a linear combination of the second set of vectors, it is (more easily) seen to be a linear combination of the first, namely
\[
 w = (a_1 + a_2c)v_1 + a_2v_2 + a_3v_3 \ldots + a_mv_m. \quad (2.4)
\]
Thus we have shown that the third row operation does not affect the row space. The proof that the third row operation doesn’t affect linear independence is similar: we assume that the second set of vectors is linearly dependent, and – taking \( w = 0 \), use the same formulas to express 0 as a linear combination of the first set of vectors, and conversely. We just have to make sure that the condition of at least one of the coefficients being nonzero is preserved: i.e., if as long as one of the coefficients in (2.1) is nonzero, (we must check that) so is one of the coefficients in (2.2), and as long as one of the coefficients in (2.3) is nonzero, so is one of the coefficients in (2.4). If any \( a_3, \ldots, a_m \) is nonzero, great, because we’re not changing these coefficients. If in the expression
\[
 0 = a_1v_1 + a_2v_2 + \ldots + a_mv_m
\]
\( a_2 \) is not zero, then since \( a_2 \) is the coefficient of \( cv_1 + v_2 \) in the second set of vectors, we’re also okay. The last possibility to worry about is if \( a_2 = a_3 = \ldots = a_m = 0 \) and it is only \( a_1 \) that is nonzero. But then \( a_2 = 0 \) and the coefficient of \( v_1 \) in (2.2) is \( a_1 + ca_2 = a_1 \neq 0 \). Similarly, we can see that as long as one of the coefficients of (2.2) is nonzero, one of the coefficients of (2.1) is nonzero. (Similarly with (2.3) and (2.4)) This completes the proof of the proposition!

After that rather technical work, the proof of Theorem 4 becomes surprisingly easy. Suppose we have a linear system of equations with linearly independent row vectors. Put that system in row echelon form. By what we just showed, the row vectors will still be linearly independent. I claim that each row has a leading entry – indeed in any matrix a row can fail to have a leading entry only if it consists entirely of zeros. But any set of vectors containing the zero vector is linearly dependent! This shows immediately that the system is consistent, because the only way for a row echelon form system to be inconsistent is to have a zero row and then, to the right of the vertical line, a nonzero entry. Moreover, the number of parameters is equal to the number of free variables, which is the total number of variables, \( n \), minus the number of leading variables, which we have just argued is in our case equal to the number of rows \( m \). So the number of solution parameters is \( n - m \), as we expected.