Chapter 6. Determinants

This material is in Chapter 2 of Anton & Rorres (or most of it is there).

6.1 Introductory remarks

The determinant of a square matrix A is a number $\det(A)$ associated with the matrix A, and one of its main propertiues is that A^{-1} exists exactly when $\det(A) \neq 0$.

Unfortunatley the calculation of $\det(A)$, and the explanation of what it is, turns out to be tricky. Certainly it is harder than the trace of A. Very vaguley $\det(A)$ is the number you end up dividing by when you compute A^{-1} (and that ties in with the fact that you can't divide by it if it is zero, so that the inverse matrix of A won't make sense if $\det(A) = 0$).

We can make that rather less vague for 2×2 matrices

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

In this case you can calculate A^{-1} as a formula. You can do it either by row-reducing

$$[A \,|\, I_2] = \left[\begin{array}{cccc} a_{11} & a_{12} & : & 1 \quad 0 \\ a_{21} & a_{22} & : & 0 \quad 1 \end{array} \right]$$

and you should end up with

$$A^{-1} = \frac{1}{a_{11}a_{22} - a_{21}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

Having got this formula somehow, you could also check that it works out. (To do that, multiply the supposed A^{-1} by A to see you do indeed get I_2.)

6.1.1 Definition. For a 2×2 matrix $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, the *determinant* of A is defined to be the number

$$\det(A) = a_{11}a_{22} - a_{21}a_{21}$$

In words, this is the product of the two diagonal entries minus the product of the two off-diagonal entries.

It is possible to work out a formula for the inverse of a 3×3 matrix, though it would is quite a bit more messy. There are a number of ways to say what $\det(A)$ is for matrices that are larger than 2×2. I think there is no one way that is really neat. All the approaches either use a lot of ancillary theory, or else have some significant drawback. The way we will choose now is easy enough to explain, but tricky enough to use as a way of showing that determinants do what they are meant to do. In other words proofs are not so nice when we start the way we are going to do, but we won't really notice that problem because we will skip the proofs!

6.2 Cofactor expansion approach to determinants

A quick way to define a determinant is via what is called cofactor expansion along the first row. For 3×3 matrices this means

$$\det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = a_{11} \det \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix} - a_{12} \det \begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix} + a_{13} \det \begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$$

In words the idea is to multiply each entry of the first row times the determinant of the matrix you get by convering over the first row and the column of the entry. Then add these up with alternating signs $+, -, + \ldots$.

When we start with a 3×3 matrix A, we end up with $\det(A)$ in terms of 2×2 determinants. And we already know how to evaluate them.

For the 4×4 case, this idea works out like this

$$\begin{aligned} \det \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} &= a_{11} \det \begin{bmatrix} a_{22} & a_{23} & a_{24} \\ a_{32} & a_{33} & a_{34} \\ a_{42} & a_{43} & a_{44} \end{bmatrix} - a_{12} \det \begin{bmatrix} a_{21} & a_{23} & a_{24} \\ a_{31} & a_{33} & a_{34} \\ a_{41} & a_{43} & a_{44} \end{bmatrix} \\ &\quad + a_{13} \det \begin{bmatrix} a_{21} & a_{22} & a_{24} \\ a_{31} & a_{32} & a_{34} \\ a_{41} & a_{42} & a_{44} \end{bmatrix} - a_{14} \det \begin{bmatrix} a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{bmatrix} \end{aligned}$$

You should note the alternating signs and that what we now end up with is 4 determinants of size 3×3 to calculate. If we expand each of these via cofactors along the first row, we end up with $12 = 4 \times 3$ determinants of size 2×2 to calculate.

If we use the same approach for 5×5 determinants, we end up with even more work to do. So this method may be simple enough in principle but it is laborious. We will soon explain a more efficient approach for large matrices.

6.3 A formula for determinants

While the above explanation is fine, it is what is called a reduction formula for a determinant. It says how to work out a determinant (of an $n \times n$ matrix A with $n \geq 3$) in terms of smaller determinants. When you keep using the reduction formula enough you get down to 2×2 determinants and we have a nice tidy formula for them. You might like to have a formula for bigger determinants, not a reduction formula.

Such a thing is available and is described in the book by Anton & Rorres in §2.4. One snag with it is that it requires a bit of theory to explain how it works. I'll outline it below.

In fact the cofactor expansion idea (the reduction formula) works out for 2×2 determinants as well. If you wanted to use it on a 2×2 determinant, it would tell you the answer in terms of 1×1 determinants! It would say

$$\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11} \det[a_{22}] - a_{12} \det[a_{21}]$$

and we might have to worry a bit about what the 1×1 determinants $\det[a_{22}]$ and $\det[a_{21}]$ might mean. Well, a 1×1 matrix has just one entry, one number in it. If you want to be fussy you can insist that a 1×1 matrix is a matrix and an number is a scalar, not the same as a matrix. But actually there is rarely a reason to be so fussy and we can usually safely ignore the difference between a 1×1 matrix and a scalar. The determinant of a 1×1 matrix is just that scalar. Then the reduction formula works out the right determinant for 2×2 matrices!

Technically, we should have said what determinants of 1×1 matrices are and starting with 2×2 meant that we were not being totally complete. So we've filled in that small detail now, though it is not real important.

Back to a formula for general determinants. If you think about it for a while, it is not hard to see that what you get when you expand out $\det(A)$ completely is a sum of products of entries of A times ± 1. In fact what happens is that, if A is an $n \times n$ matrix, then all products of n entries of A show up which satisfy the restriction that the product contains just one etry from each row of A and one from each column. This is kind of apparent from the cofactor expansion approach. At the beginning we get an entry from the first row times a determinant of a matric where the first row is no longer there (and the column of the entry you have is also no longer present in the smaller determinant).

By arguing in this way, you can establish that what you would get if you multiplied out all the reduction formulae for

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

would be a big sum of terms of the form

$$\pm a_{1j_1} a_{2j_2} \cdots a_{nj_n}$$

where $j_1, j_2, \ldots, j_n$ are all of the n column numbers in some order.

So $j_1, j_2, \ldots, j_n$ must be all n of the column numbers $1, 2, \ldots, n$, but not necessarily in that order. In fact all possible orders appear. The possible way to reorder $1, 2, \ldots, n$ are called the permuations of these n numbers. It is possible to see fairly easily that the total number of these permuations is a number called 'n factorial'. We write in as $n!$ and it is the product of the numbers $1, 2, \ldots, n$.

$$n! = n(n-1)(n-2)\cdots(3)(2)(1)$$

So this approach gives a formula for the determiant, a formula with $n!$ terms. One snag is that $n!$ gets big pretty fast.

$$3! = 6, \quad 5! = 120, \quad 10! = 3628800$$

So for a 10×10 matrix this formula would have more than 3.6 million terms, a lot. Even for 5×5, you'd have more than 100 terms, each involving a product of 5 terms.

Then there is the problem of which terms get a plus sign and which get a minus. There is a theory about this, and it comes down to something called the 'sign' of a permutation. It would be a digression for us to try and explain what that is in a satisfactory way. So here is a quick explanation. Srtaing with a permutation

$$j_1, j_2, \dots, j_n$$

of $1, 2, \dots, n$, (so that $j_1, j_2, \dots, j_n$ are all the first n whole numbers written in some order), we are going to write down a matrix called the matrix for this permuation. In each row (and column) the permutation matrix has just one single entry equal to 1, all the others are 0. To be specific, in row number i, there is a 1 in column j_i, and zeros elsewhere. (Another way to say it is that there are entries $= 1$ at the positions (i, j_i) for $i = 1, 2, \dots, n$, but every other entry is 0.) The sign of the permutation is the same as the determinant of its permuatation matrix.

Well, that is a true statement, but it is a bit unsatisfactory. Our long formula for a determinant still has some determinants in it, the ones that give the ± 1 signs.

There is a way to say how the whole formula works out for 3×3 matrices, and it is a fairly satisfactory way of working out 3×3 determinants. The drawback is that it does not extend to bigger determinants in any very similar way.

Starting with a 3×3 matrix

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

write it down with the first two columns repeated

$$\begin{tikzcd}
a_{11} \arrow[dr] & a_{12} \arrow[dr] & a_{13} \arrow[dr] \arrow[dl, dashed] & a_{11} \arrow[dl, dashed] & a_{12} \arrow[dl, dashed] \\
a_{21} \arrow[dr] & a_{22} \arrow[dr] \arrow[dl, dashed] & a_{23} \arrow[dr] \arrow[dl, dashed] & a_{21} \arrow[dl, dashed] & a_{22} \\
a_{31} & a_{32} & a_{33} & a_{31} & a_{32}
\end{tikzcd}$$

Add the products diagonally to the right and subtract those diagonally to the left as indicated by the arrows

$$\begin{aligned} \det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} &= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ &\quad - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} \end{aligned}$$

Notice that there are the correct number of terms here ($3! = 6$). And each product of 3 has one entry from each row, one entry from each column.

As mentioned above, this method of repeating the first and second columns does **not** work for sizes apart from 3×3, and there is nothing really like this for 4×4 or bigger matrices. The cofactor expansion method does work for any size of (square) matrix.

6.4 Properties of determinants

Here are the key properties of dertminants. We'll explain why they are true in the case of 2×2 determinants, and give short shift to the explanations of why these properties still work for $n \times n$ determinants.

(i) $\det(I_2) = 1$

This is pretty easy to see.

(ii) $\det(AB) = \det(A)\det(B)$

Proof. We should show that this is true for any 2×2 matrices A and B, without knowing what the entries are in the matrices. What we do is write out the matrices

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}, \quad AB = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix}$$

Then multiply out

$$\det(AB) = (a_{11}b_{11} + a_{12}b_{21})(a_{21}b_{12} + a_{22}b_{22}) - (a_{11}b_{12} + a_{12}b_{22})(a_{21}b_{11} + a_{22}b_{21})$$

and

$$\det(A)\det(B) = (a_{11}a_{22} - a_{12}a_{21})(b_{11}b_{22} - b_{12}b_{21})$$

to show that the answers are the same.

It is not really hard to do, though maybe not worth writing the remaining steps out. You might like to convince yourself that it does work out as claimed. $\square$

(iii) $\det(A^{-1}) = \dfrac{1}{\det(A)}$

Proof. Using the previous result

$$\det(A^{-1})\det(A) = \det(A^{-1}A) = \det(I_2) = 1$$

and so $\det(A^{-1}) = 1/\det(A)$. $\square$

(iv) $\det(A^t) = \det(A)$

This is not at all hard (for the 2×2 case).

(v) The determinants of elementary matrices are as follows

$1'$. E the elementary matrix for the row operation "multiply row 1 by $k \neq 0$"

$E = \begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}$, $\det(E) = k$.

1″. E the elementary matrix for the row operation "multiply row 1 by $k \neq 0$"
$E = \begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix}$, $\det(E) = k$.

2′. E the elementary matrix for the row operation "add k times row 2 to row 1"
$E = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$, $\det(E) = 1$.

2″. E the elementary matrix for the row operation "add k times row 1 to row 2"
$E = \begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}$, $\det(E) = 1$.

3. E the elementary matrix for the row operation "swop rows 1 and 2"
$E = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $\det(E) = -1$.

For general $n \times n$ determinants, all these statements remain true, though the last one needs to be restated:

6.4.1 Lemma. *The determinants of elementary matrices are as follows*

1. *E the elementary matrix for the row operation "multiply a row by $k \neq 0$" has* $\det(E) = k$.

2. *E the elementary matrix for the row operation "add k times one row to a different row" has* $\det(E) = 1$.

3. *E the elementary matrix for the row operation "swop two specified rows" has* $\det(E) = -1$.

This leads us to a way to calculate $n \times n$ determinants. Starting with the matrix A, do row operations on A to row reduce A. At the first row operation we are replacing the matrix A by EA for some elementary matrix A. So the determinant of the matrix we have changes to

$$\det(EA) = \det(E)\det(A)$$

We've seen that $\det(E)$ is easy to figure out. So it is quite easy to keep track of the changes in the determinant as we do each row operation. (We'll organise this better soon.)

If we keep doing row operations, keeping track of how the determinant changes as we go along, we will know that if A is invertible row operations will lead to the identity matrix. And that has determinant 1.

On the other hand if the reduced row echelon form of the $n \times n$ matrix A is not the $n \times n$ identity matrix I_n, then it is easy to see it is a matrix with a row of zeros in the last row (and maybe rows of zeros above that too). Now it is not all that hard to see that a matrix with a row of zeros has zero determinant. So it follows that if A is not invertible then $\det(A) = 0$. We can state the combination of that fact with the fact that $\det(A^{-1}) = 1/\det(A)$ (so that $\det(A)$ can't be zero if A is invertible) as a theorem.

6.4.2 Theorem. *If A is an $n \times n$ matrix, then the following are two equivalent statements about A:*

(a) *A is invertible*

(b) $\det(A) \neq 0$.

This can then be added as an extra part to Theorem 5.13.2.

Here now comes a theorem listing the properties of determinants, including tidier statements of what was mentioned above about using row operations to evaluate determinants. The statements are listed in an order where a proof is possible, even though we will not really prove them all. The list of properties is also extended beyond what we have said above.

6.4.3 Theorem. *(Simplification rules and properties of determinants) Let A be an $n \times n$ matrix throughout.*

(i) *Adding a multiple of one row of A to another row results in a matrix with unchanged determinant.*

(ii) *Factoring a scalar $k \neq 0$ out from a single row of A divides the determinant by k. That is*

$$\det \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & & a_{2n} \\ \vdots & & & \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = k \det \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & & a_{2n} \\ \vdots & & & \\ a_{i1}/k & a_{i2}/k & \cdots & a_{in}/k \\ \vdots & & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

(iii) *Swopping two rows of A changes the determinant by a factor -1.*

(iv) *If A has a row of all zeros, then* $\det(A) = 0$.

(v) *If A has two rows where one row is a multiple of the other, then* $\det(A) = 0$.

(vi) *If A is upper triangular or if A is lower triangular then* $\det(A)$ *is the product of the diagonal entries.*

(vii) *An efficient way to work out* $\det(A)$ *is to use Gauss-Jordan elimination to row reduce A to row-echelon form, keeping track of how the determinat changes after each row operation (see properties (i) – (iii)). Notice that the row-echelon form will be upper triangular with either all 1's on the diagonal or some diagonal entries are 0 (so that the determinant of the row-echelon form is 1 in the case where A is invertible and 0 if A is not invertible).*

(viii) *A is invertible exactly when* $\det(A) \neq 0$. *(That's Theorem 6.4.2.)*

(ix) $\det(AB) = \det(A)\det(B)$ *if B is $n \times n$.*

(x) $\det(A^{-1}) = 1/\det(A)$ *if A is invertible.*

(xi) $\det(A^t) = \det(A)$.

(xii) Properties (i) – (v) remain true if the word "row" is replaced by column everywhere. Thus one can also simlify determinants by using colimns operations, as well as or instead of row operations.

Proof. This is not really a proof, just some remarks about how a proof could go. Some of the easier steps are done more or less in full, but other parts if the argument are missing quite a few details.

The messiest part of the proof is to show properties (i) – (iii). It is possible to show them starting from the definition we gave, defining $\det(A)$ by cofactor expansion along the first row, but the proof is a bit messy. For that reason they are often proved via the machinery with permutations that we mention above. However, there is quite a bit of work needed to straighten out the theory needed to understand the sign of a permutation. So that method is not so simple either.

What (i) – (iii) tell you is that $\det(EA) = \det(E)\det(A)$ for an elementary matrix E.

Property (iv) is easy enough using (ii) — just factor 2 out of the row of zeros. On the one hand it has no effect on the matrix and so leaves the determinant unchanged. But by (ii) it has to change the determinant by a factor of 2. The only way both can be true is if $\det(A) = 0$.

Property (v) is quite easy from (i) and (iv).

For (vi), the lower triangular case is fairly obvious from the cofactor expansion definition. For the upper triangular case, it is a bit less obvious but still not real hard.

The fact the (vii) works as a method follows from the previous results. The fact that it is efficient requires an interpretation. What it means is that the number of aritmetic operations innvolved in carrying out the method is much smaller than would be involved in expanding the determinant by cofactors — if n is anyway big.

(viii) follows from an analysis of the method (vii) together with the fact we know already that invertible matrices row-reduce to the identity.

For (ix) we can use this again. If A is invertible, then there are elementary row operations to row-deduce it to the identity. That means there is a sequence of row operatiosn, a corresponding sequence of elementary matrices $E_1, E_2, \ldots, E_r$ so that

$$E_r E_{r-1} \ldots E_2 E_1 A = I_n$$

That tells us

$$A = E_1^{-1} E_2^{-1} \ldots E_r^{-1}$$

and from (i) – (iii), we also have

$$\det(E_r)\det(E_{r-1}) \cdots \det(E_2)\det(E_1)\det(A) = \det(I_n) = 1$$

So

$$\det(A) = \frac{1}{\det(E_1)\det(E_2)\cdots\det(E_r)}$$

If B is also invertible then we can make a similar statement about $\det(B)$ and then show (ix). If one of A or B is not invertible, we need a slightly different argument to show AB can't be invertible either and so $\det(AB) = 0$.

A proof for (x) can be just the same as the proof we gave in the 2×2 case.

For (xi) we can use the idea about products of elementary matrices again, and first check that $\det(E^t) = \det(E)$ when E is elementary. In the case of invertible A,

$$A = E_1^{-1} E_2^{-1} \dots E_r^{-1} \Rightarrow A^t = (E_r^{-1})^t (E_{r-1}^{-1})^t \dots (E_1^{-1})^t$$

and so

$$\det(A^t) = \det(E_r^{-1}) \det(E_{r-1}^{-1}) \dots \det(E_1^{-1}) = \frac{1}{\det(E_1) \det(E_2) \cdots \det(E_r)} = \det(A)$$

The last property (xii) is a consequence of (xi), since column operations on A correspond to row operations on A^t. $\square$

6.4.4 Examples. 1. Find $\det \begin{bmatrix} 0 & 1 & 5 \\ 1 & 2 & 4 \\ 4 & 5 & 7 \end{bmatrix}$ via row reduction.

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

2. Show $\det \begin{bmatrix} 1 & 1 & 1 \\ x & y & z \\ x^2 & y^2 & z^2 \end{bmatrix} \neq 0$ if x, y and z are all different.

   (Notice that the determinant would certainly be zero if any two of x, y and z were equal. In this case the matrix would have tow identitcal columns, and so determinant zero.)

   To solve this we will first transpose the matrix and then use row operations on the transpose. We need not do this as we are allowed use column operations on the original matrix, but we are more used to row operations.

$$
\begin{aligned}
\det\begin{bmatrix} 1 & 1 & 1 \\ x & y & z \\ x^2 & y^2 & z^2 \end{bmatrix} &= \det\begin{bmatrix} 1 & x & x^2 \\ 1 & y & y^2 \\ 1 & z & z^2 \end{bmatrix} \\
&= \det\begin{bmatrix} 1 & x & x^2 \\ 0 & y-x & y^2-x^2 \\ 1 & z & z^2 \end{bmatrix} \\
&= \det\begin{bmatrix} 1 & x & x^2 \\ 0 & y-x & y^2-x^2 \\ 0 & z-x & z^2-x^2 \end{bmatrix} \\
&= \det\begin{bmatrix} 1 & x & x^2 \\ 0 & y-x & (y-x)(y+x) \\ 0 & z-x & (z-x)(z+x) \end{bmatrix} \\
&= (y-x)\det\begin{bmatrix} 1 & x & x^2 \\ 0 & 1 & y+x \\ 0 & z-x & (z-x)(z+x) \end{bmatrix} \\
&= (y-x)(z-x)\det\begin{bmatrix} 1 & x & x^2 \\ 0 & 1 & y+x \\ 0 & 1 & z+x \end{bmatrix} \\
&= (y-x)(z-x)\det\begin{bmatrix} 1 & x & x^2 \\ 0 & 1 & y+x \\ 0 & 0 & z+x-(y+x) \end{bmatrix} \\
&= (y-x)(z-x)\det\begin{bmatrix} 1 & x & x^2 \\ 0 & 1 & y+x \\ 0 & 0 & z-y \end{bmatrix} \\
&= (y-x)(z-x)(z-y)
\end{aligned}
$$

If x, y and z are all different, then all 3 factors in the determinant are different from 0. So their product is not zero.

By the way, this determinant we just worked out has a special name. It is called a *Vandermonde determinant*.

6.5 Geometrical view of determinants

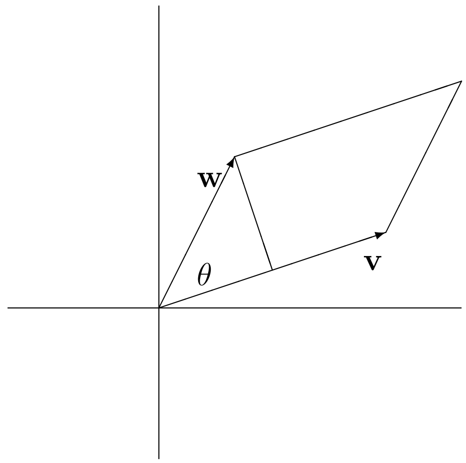
For 2×2 matrices and 3×3 matrices, there is a graphical interpretation of the determinant.

Consider the case of 2×2 matrices first, and think of the rows of the matrix as components of vectors in the plane $\mathbb{R}^2$.

$$
\det\begin{bmatrix} v_1 & v_2 \\ w_1 & w_2 \end{bmatrix}
$$

So we are thinking of the two vectors $\mathbf{v} = (v_1, v_2) = v_1\mathbf{i} + v_2\mathbf{j}$ and $\mathbf{w} = (w_1, w_2) = w_1\mathbf{i} + w_2\mathbf{j}$. If we draw the two vectors from the origin (the position vectors of the points (v_1, v_2) and (w_1, w_2)) and fill in the parallelogram that has $\mathbf{v}$ and $\mathbf{w}$ as sides, then

$$\det \begin{bmatrix} v_1 & v_2 \\ w_1 & w_2 \end{bmatrix} = \pm(\text{the area of the parallelogram})$$



The area of that parallelogram is

$$\text{base } \times \text{ perpendicular height} = \|\mathbf{v}\|(\|\mathbf{w}\| \sin\theta)$$

We could work out $\sin\theta$ by first working from $\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\|\|\mathbf{w}\| \cos\theta$, and then using $\sin^2\theta + \cos^2\theta = 1$. Here is an alternative way that needs less algebra.

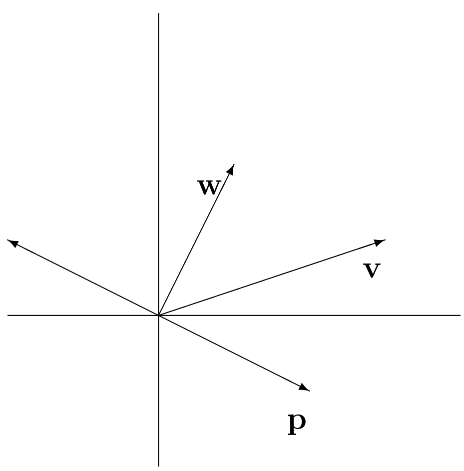
We know $\sin\theta = \cos\left(\frac{\pi}{2} - \theta\right)$. The vector

$$\mathbf{p} = w_2\mathbf{i} - w_1\mathbf{j}$$

has the same length $\|\mathbf{p}\| = \sqrt{w_2^2 + w_1^2} = \|\mathbf{w}\|$ as $\mathbf{w}$ and it is perpendicular to $\mathbf{w}$ because

$$\mathbf{p} \cdot \mathbf{w} = w_2 w_1 - w_1 w_2 = 0.$$

There are two vectors in the plane with that length and direction perpendicular to $\mathbf{w}$.



In our picture, the one that makes an angle $(\pi/2) - \theta$ with $\mathbf{v}$ is $\mathbf{p}$ (but that depends on the fact that $\mathbf{w}$ is θ radians anticlockwise from $\mathbf{v}$ — if it was clockwise then the right one would be $-\mathbf{p}$ instead).

So we have

$$\mathbf{v} \cdot \mathbf{p} = \|\mathbf{v}\| \|\mathbf{p}\| \cos\left(\frac{\pi}{2} - \theta\right) = \|\mathbf{v}\| \|\mathbf{w}\| \sin\theta$$

and so

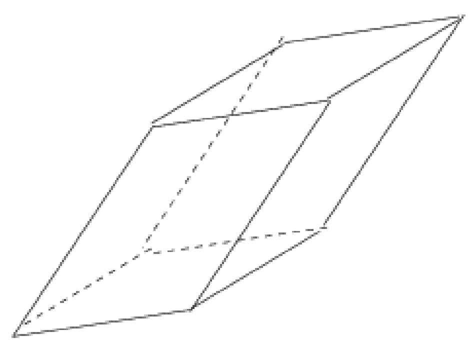
$$v_1 w_2 - v_2 w_1 = \|\mathbf{v}\| \|\mathbf{w}\| \sin\theta = \text{ area of parallelogram}$$

$$\det\begin{bmatrix} v_1 & v_2 \\ w_1 & w_2 \end{bmatrix} = \text{ the area of the parallelogram}$$

As we mentioned, we don't always get this with a plus sign. Sometimes we get a minus (when $\mathbf{w}$ is clockwise from $\mathbf{v}$).

Notice that if $\mathbf{v}$ is parallel to $\mathbf{w}$, or if they are in eactly opposite directions, the parallelogram collapses to a line, and so we get zero area. In this case one row of the matrix is a multiple of the other.

When we move to 3 dimensions, we can get a similar result, but we need a three dimensional version of a parallelogram. Three vectors in space will span a *parallelopiped*. Its a shape like a box without right angles between the sides.



All the faces are parallelograms, and opposite faces are parallel to one another. If we think of the origin as one corner, the 3 sides emanating from there could represent three vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ in space.

On the other hand if we start with 3 vectors $\mathbf{u} = u_1\mathbf{i} + u_2\mathbf{j} + u_3\mathbf{k}$, $\mathbf{v} = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}$ and $\mathbf{w} = w_1\mathbf{i} + w_2\mathbf{j} + w_3\mathbf{k}$, then we can draw them all as arrows starting from the origin. Then we can make a parallelopiped using those as three of the edges.

The link with determinants is

$$\det\begin{bmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{bmatrix} = \pm(\text{volume of the parallelopiped}) \tag{1}$$

The simplest way to explain why this fact is true involves the idea of the cross product of vectors in space. We'll come to that later, but we will use the facts abour areas and volumes in some simple examples first.

6.5.1 Example. Find the volume of the parallelopiped in space where one corner is at $(0,0,0)$ and the 3 vectors

$$\begin{aligned}\mathbf{u} &= 4\mathbf{i} - 2\mathbf{j} + \mathbf{k}\\ \mathbf{v} &= 5\mathbf{i} + \mathbf{j} - \mathbf{k}\\ \mathbf{w} &= \mathbf{i} - \mathbf{j} + \mathbf{k}\end{aligned}$$

are parallel to three of the edges.

Answer is

$$\det\begin{bmatrix}4 & -2 & 1\\ 5 & 1 & -1\\ 1 & -1 & 1\end{bmatrix} = \pm(\text{volume})$$

So the absolute value of the determinant gives the volume.

$$\begin{aligned}\det\begin{bmatrix}4 & -2 & 1\\ 5 & 1 & -1\\ 1 & -1 & 1\end{bmatrix} &= 4\det\begin{bmatrix}1 & -\frac{1}{2} & \frac{1}{4}\\ 5 & 1 & -1\\ 1 & -1 & 1\end{bmatrix}\\ &= 4\det\begin{bmatrix}1 & -\frac{1}{2} & \frac{1}{4}\\ 0 & \frac{7}{2} & -\frac{9}{4}\\ 0 & -\frac{1}{2} & \frac{3}{4}\end{bmatrix}\\ &= 4\left(\frac{7}{2}\right)\det\begin{bmatrix}1 & -\frac{1}{2} & \frac{1}{4}\\ 0 & 1 & -\frac{9}{14}\\ 0 & -\frac{1}{2} & \frac{3}{4}\end{bmatrix}\\ &= 14\det\begin{bmatrix}1 & -\frac{1}{2} & \frac{1}{4}\\ 0 & 1 & -\frac{9}{14}\\ 0 & 0 & \frac{3}{4} - \frac{9}{28}\end{bmatrix}\\ &= 14\left(-\frac{3}{7}\right) = -6\end{aligned}$$

So the volume is 6.

6.6 Cross products

This is something that makes sense in three dimensions only. There is no really similar product of two vectors in higher dimensions. In this respect it is different from things we have seen before. For example, we might have started with dot products in 2 dimensions, then extended the notion to 3 dimensions, and later realised that the formula we had for dot products has an obvious extension to 4 dimensions $\mathbb{R}^4$, to $\mathbb{R}^5$ and to every $\mathbb{R}^n$. We just need to extend the formula in a rather easy way. Most of the other formulae we had also extend to $\mathbb{R}^n$ with no real bother. Cross products are different. (By the way, this topic is in §3.4 of Anton & Rorres, as is the material above about areas and volumes.)

The easiest way to remember the formula for the *cross product* $\mathbf{v} \times \mathbf{w}$ of two vectors

$$\begin{aligned}\mathbf{v} &= v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k},\\ \mathbf{w} &= w_1\mathbf{i} + w_2\mathbf{j} + w_3\mathbf{k},\end{aligned}$$

in space is to use a kind of "determinant" formula

$$\mathbf{v} \times \mathbf{w} = \det \begin{bmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{bmatrix}$$

Now we have not allowed determinants of matrices where the entries are not scalars, and we should not allow it. However, in this particular case, we can get away with it if we interpret the determinant as what we would get by a cofactor expansion along the first row. So, in a more correct statement, the definition of the cross product is

$$\begin{aligned}\mathbf{v} \times \mathbf{w} &= \mathbf{i}\det\begin{bmatrix} v_2 & v_3 \\ w_2 & w_3 \end{bmatrix} - \mathbf{j}\det\begin{bmatrix} v_1 & v_3 \\ w_1 & w_3 \end{bmatrix} + \mathbf{k}\det\begin{bmatrix} v_1 & v_2 \\ w_1 & w_2 \end{bmatrix}\\ &= (v_2w_3 - v_3w_2)\mathbf{i} - (v_1w_3 - v_3w_1)\mathbf{j} + (v_1w_2 - v_2w_2)\mathbf{k}\\ &= (v_2w_3 - v_3w_2)\mathbf{i} + (v_3w_1 - v_1w_3)\mathbf{j} + (v_1w_2 - v_2w_2)\mathbf{k}\end{aligned}$$

There is actually a pattern[1] to this last formula and so it is not quite impossible to remember. But the (slightly suspect) determinant formula is easier to recall, I think.

6.7 Properties of cross products (in $\mathbb{R}^3$)

(i) $\mathbf{v} \times \mathbf{w}$ is a vector in space.

(ii) $\mathbf{w} \times \mathbf{v} = -\mathbf{v} \times \mathbf{w}$

Proof. This is not hard. Amounts to a property of determinants. Switching two rows changes the determinant by a factor -1. □

(iii) $\mathbf{v} \times \mathbf{w}$ is perpendicular to both $\mathbf{v}$ and $\mathbf{w}$.

[1] The first component of $\mathbf{v} \times \mathbf{w}$ depends on the components of $\mathbf{v}$ and $\mathbf{w}$ other than the first. Starting with $v_2w_3 - v_3w_2$ we can get to the next component by adding 1 to the subscripts and interpreting $3 + 1$ as 1. Or think in terms of cycling the subscripts around $1 \to 2 \to 3 \to 1$ to get the next component. You still have to remember the first one.

Proof.

$$
\begin{aligned}
&\mathbf{v}\cdot(\mathbf{v}\times\mathbf{w})\\
&= (v_1\mathbf{i}+v_2\mathbf{j}+v_3\mathbf{k})\cdot\left(\mathbf{i}\det\begin{bmatrix} v_2 & v_3\\ w_2 & w_3\end{bmatrix} - \mathbf{j}\det\begin{bmatrix} v_1 & v_3\\ w_1 & w_3\end{bmatrix} + \mathbf{k}\det\begin{bmatrix} v_1 & v_2\\ w_1 & w_2\end{bmatrix}\right)\\
&= v_1\det\begin{bmatrix} v_2 & v_3\\ w_2 & w_3\end{bmatrix} - v_2\det\begin{bmatrix} v_1 & v_3\\ w_1 & w_3\end{bmatrix} + v_3\det\begin{bmatrix} v_1 & v_2\\ w_1 & w_2\end{bmatrix}\\
&= \det\begin{bmatrix} v_1 & v_2 & v_3\\ v_1 & v_2 & v_3\\ w_1 & w_2 & w_3\end{bmatrix}\\
&= 0
\end{aligned}
$$

because there are two equal rows in this determinant.

So $\mathbf{v}\perp\mathbf{v}\times\mathbf{w}$.

To show $\mathbf{w}\perp\mathbf{v}\times\mathbf{w}$, we can either repeat a similar calculation or we can use $\mathbf{v}\times\mathbf{w} = -\mathbf{w}\times\mathbf{v}\perp\mathbf{w}$. □

(iv) $\|\mathbf{v}\times\mathbf{w}\| = \|\mathbf{v}\|\|\mathbf{w}\|\sin\theta\|$ where $\theta =$ the angle between $\mathbf{v}$ and $\mathbf{w}$.

Proof. The proof for this is a calculation using

$$\cos\theta = \frac{\mathbf{v}\cdot\mathbf{w}}{\|\mathbf{v}\|\|\mathbf{w}\|} = \frac{v_1w_1+v_2w_2+v_3w_3}{\|\mathbf{v}\|\|\mathbf{w}\|}$$

$\sin^2\theta = 1-\cos^2\theta$ and multiplying out

$$\|\mathbf{v}\|\|\mathbf{w}\|^2 = (v_2w_3-v_3w_2)^2+(v_3w_1-v_1w_3)^2+(v_1w_2-v_2w_2)^2$$

to show it is the same as

$$\|\mathbf{v}\|^2\|\mathbf{w}\|^2\sin^2\theta = \|\mathbf{v}\|^2\|\mathbf{w}\|^2 - \|\mathbf{v}\|^2\|\mathbf{w}\|^2\cos^2\theta = \|\mathbf{v}\|^2\|\mathbf{w}\|^2-(\mathbf{v}\cdot\mathbf{w})^2$$

It is not real hard to do the required algebra, but a bit messy. □

(v) Now that we know in a geoemtrical way the length of $\mathbf{v}\times\mathbf{w}$, and we also know that it is a vector perpendicular to both $\mathbf{v}$ and $\mathbf{w}$, we can try to describe cross products in a geometrical way.

If the angle θ between $\mathbf{v}$ and $\mathbf{w}$ is not 0 and not π, then the vectors $\mathbf{v}$ and $\mathbf{w}$ are not in the same direction and also not in exactly opposite directions. So as long as $0<\theta<\pi$, then we can say that there is one plane through the origin parallel to both $\mathbf{v}$ and $\mathbf{w}$ (or containing both vectors if we draw them from the origin). The cross product is then in one of the two normal directions to that plane.

If $\theta = 0$ or $\theta = \pi$, there is no one plane containing $\mathbf{v}$ and $\mathbf{w}$, but in these cases $\sin\theta = 0$ and so we know $\mathbf{v} \times \mathbf{w} = \mathbf{0}$.

In the case $0 < \theta < \pi$, we can describe the cross product up to one of the two normal directions to the plane. The question then is to say which direction it is in. If we can identify the top (or upwards) side of the plane somehow, is the cross product pointing up or down? And if the plane is vertical? The anwer to this depends on having the axes fixed in such a way that the direction of $\mathbf{i}, \mathbf{j}$, and $\mathbf{k}$ obey a 'right-hand rule'. This can be described in terms of the directions of the index finger, first finger and thumb on your right hand if you hold them perpendicular to one another. Another way is to place a cokscrem (an ordinary right-handed corkscrew) along the vertical axis and twist the screw from the x-axis towards the y-axis. It should travel in the direction of the positive z-axis.

For two vectors $\mathbf{v}$ and $\mathbf{w}$, the direction of $\mathbf{v} \times \mathbf{w}$ is described by a right-hand rule. Imaging a corkscrew placed so it is perpendicular to the plane of $\mathbf{v}$ and $\mathbf{w}$. Turn the screw from $\mathbf{v}$ towards $\mathbf{w}$ and the direction it travels is the same as the direction of $\mathbf{v} \times \mathbf{w}$.

(vi) There are some algebraic properties of the cross product that are as you would expect for products:

$$\begin{aligned}
\mathbf{u} \times (\mathbf{v} + \mathbf{w}) &= \mathbf{u} \times \mathbf{v} + \mathbf{u} \times \mathbf{w} \\
(\mathbf{u} + \mathbf{v}) \times \mathbf{w} &= \mathbf{u} \times \mathbf{w} + \mathbf{v} \times \mathbf{w} \\
(k\mathbf{v}) \times \mathbf{w} &= k(\mathbf{v} \times \mathbf{w}) \\
&= \mathbf{v} \times (k\mathbf{w})
\end{aligned}$$

for any vectors $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$ and any scalar $k \in \mathbb{R}$. (But recall that the order matters since $\mathbf{v} \times \mathbf{w} = -\mathbf{w} \times \mathbf{v}$.)

These properties are quite easy to check out.

(vii) Vector triple products

$$\mathbf{u} \times (\mathbf{v} \times \mathbf{w})$$

make sense, but that is usually not the same as $(\mathbf{u} \times \mathbf{v}) \times \mathbf{w}$.

For example

$$\begin{aligned}
\mathbf{i} \times (\mathbf{i} \times \mathbf{j}) &= \mathbf{i} \times \mathbf{j} \\
&= -\mathbf{j} \\
(\mathbf{i} \times \mathbf{i}) \times \mathbf{j} &= \mathbf{0} \times \mathbf{j} \\
&= \mathbf{0}
\end{aligned}$$

6.7.1 Example. Find the equation of the plane that goes through $(1, 2, 3)$, $(3, 1, 2)$ and $(2, 3, 1)$.
Solution: We did have a way of doing this problem before, using equations that have to be satisfied by the coefficients a, b, c and d in the equation of the plane $ax + by + cz = d$ (see §3.12). Here is another approach using cross products.

Let $P = (1, 2, 3)$, $Q = (3, 1, 2)$ and $R = (2, 3, 1)$ and use the same letters for the position vectors $\mathbf{P}$, $\mathbf{Q}$ and $\mathbf{R}$. Then we can notice that

$$\begin{aligned}
\vec{PQ} &= \mathbf{Q} - \mathbf{P} \\
&= (3\mathbf{i} + \mathbf{j} + 2\mathbf{k}) - (\mathbf{i} + 2\mathbf{j} + 3\mathbf{k}) \\
&= 2\mathbf{i} - \mathbf{j} + \mathbf{k} \\
\vec{PR} &= \mathbf{R} - \mathbf{P} \\
&= (2\mathbf{i} + 3\mathbf{j} + \mathbf{k}) - (\mathbf{i} + 2\mathbf{j} + 3\mathbf{k}) \\
&= \mathbf{i} + \mathbf{j} - 2\mathbf{k}
\end{aligned}$$

are two vectors that are in the plane we want (or parallel to it). So their cross product must be normal to the plane:

$$\begin{aligned}
\vec{PQ} \times \vec{PR} &= \det \begin{bmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 2 & -1 & 1 \\ 1 & 1 & -2 \end{bmatrix} \\
&= \mathbf{i} \det \begin{bmatrix} -1 & 1 \\ 1 & -2 \end{bmatrix} - \mathbf{j} \det \begin{bmatrix} 2 & 1 \\ 1 & -2 \end{bmatrix} + \mathbf{k} \det \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} \\
&= 3\mathbf{i} + 3\mathbf{j} + 3\mathbf{k}
\end{aligned}$$

So the plane we want has an equation

$$3x + 3y + 3z = \text{ const.}$$

and we can plug in any of the points P, Q or R to see that the constant has to be 18. Thus the equation of the plane is

$$3x + 3y + 3z = 18$$

or rather this is one possible equation. We can multiply or divide this equation by any nonzero number and still have an equation for the plane. A tidier-looking equations is

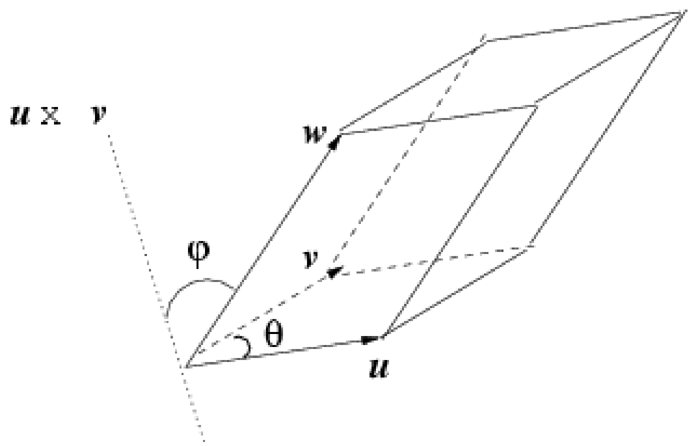
$$x + y + z = 6$$

(In retrospect maybe we could have guessed the equation, because the 3 points P, Q and R had the same coordinates permuted around. But the method we used would work for any 3 points, as long as they did not lie in a line.)

6.8 Volume of a parallelopiped

We'll now give a reason (with the help of cross products) why 3×3 determinants can be interpreted as volumes of parallepipeds. That is we will justify equation (1) now.

As before we consider the parallelopiped determined by the 3 vectors $\mathbf{u} = u_1\mathbf{i} + u_2\mathbf{j} + u_3\mathbf{k}$, $\mathbf{v} = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}$ and $\mathbf{w} = w_1\mathbf{i} + w_2\mathbf{j} + w_3\mathbf{k}$.



We take θ for the angle between $\mathbf{u}$ and $\mathbf{v}$ and look at the face spanned by these vectors as the 'base' of the parallelopiped. The volume is given by a similar formula to the one for the area of a parallelogram

$$\text{volume (parallelopiped)} = \text{ area of base } \times \text{ perpendicular height}$$

Geometrically we know

$$\text{area of base } = \|\mathbf{u}\|\|\mathbf{v}\|\sin\theta$$

where θ is the angle between $\mathbf{u}$ and $\mathbf{v}$. But we can also recognise this as the length of the cross product and so

$$\text{area of base } = \|\mathbf{u} \times \mathbf{v}\|$$

The direction of the cross product $\mathbf{u} \times \mathbf{v}$ is perpendicular to the base, and in our picture it is in the 'upwards' direction, though it could be going the opposite way. Take ϕ for the angle between $\mathbf{w}$ and the direction of $\mathbf{u} \times \mathbf{v}$. So the pependicular height is

$$\text{perpendicular height } = \pm\|\mathbf{w}\|\cos\phi$$

(where the minus sign would be needed if $\cos\phi < 0$ and the cross product was in the 'downwards' direction from the base). We usualy compute angles between vectors using the dot product and so we should look at

$$\begin{aligned} \mathbf{w} \cdot (\mathbf{u} \times \mathbf{v}) &= \|\mathbf{u} \times \mathbf{v}\|\|\mathbf{w}\|\cos\phi \\ &= \text{ area of base } \times (\pm(\text{ perpendicular height })) \\ &= \pm \text{ volume (parallelopiped)} \end{aligned}$$

We can compute the expression $\mathbf{w} \cdot (\mathbf{u} \times \mathbf{v})$ (which is sometimes called a *scalar triple product*) and show it comes out as a determinant. This calculation should remind you of the proof that

$\mathbf{v} \cdot (\mathbf{v} \times \mathbf{w}) = 0$.

$$
\begin{aligned}
&\mathbf{w} \cdot (\mathbf{u} \times \mathbf{v}) \\
&= (w_1\mathbf{i} + w_2\mathbf{j} + w_3\mathbf{k}) \cdot \left(\mathbf{i} \det \begin{bmatrix} u_2 & u_3 \\ v_2 & v_3 \end{bmatrix} - \mathbf{j} \det \begin{bmatrix} u_1 & v_3 \\ v_1 & v_3 \end{bmatrix} + \mathbf{k} \det \begin{bmatrix} u_1 & u_2 \\ v_1 & v_2 \end{bmatrix} \right) \\
&= w_1 \det \begin{bmatrix} u_2 & y_3 \\ v_2 & v_3 \end{bmatrix} - w_2 \det \begin{bmatrix} u_1 & u_3 \\ v_1 & v_3 \end{bmatrix} + w_3 \det \begin{bmatrix} u_1 & u_2 \\ v_1 & v_2 \end{bmatrix} \\
&= \det \begin{bmatrix} w_1 & w_2 & w_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{bmatrix} \\
&= -\det \begin{bmatrix} u_1 & u_2 & u_3 \\ w_1 & w_2 & w_3 \\ v_1 & v_2 & v_3 \end{bmatrix} \\
&= \det \begin{bmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{bmatrix}
\end{aligned}
$$

So we have shown now that the determinant if $\pm$ the volume (and we also showed that a scalar triple product is a determinant, though that makes sense only for vectors in 3 dimensions).

In a way it is true that determinants of $n \times n$ matrices give some sort of n-dimensional volume. We've proved it in 2 dimensions (parallelogram case) and in 3 dimensions (parallelopiped), but the idea of an n-dimensional volume is beyond us in this course.

6.9 Using determinants to find equations

In this section we present a kind of application of determinants. Not exactly a scientific application, as we will just be looking for equations. To illustrate the method, we'll start with a complicated way of finding the equation of a line through 2 points in the plane, then go on to the equation of a circle through 3 specified points, and finally look at the equation of a plane in space through 3 given points. In most of these examples we can already do them another way, and the determinant approach is kind of neat from the point of view of a mathematician. Not so really practical. Something like this is explained in §11.1 of Anton & Rorres.

Line through 2 points in the plane. If we are given two points $P = (p_1, p_2)$ and $Q = (q_1, q_2)$ in the plane we already know several ways to find the line going through them.

If the points are not on a vertical line (which would have to have equation $x = p_1$ in the case $p_1 = q_1$) then the line has a slope

$$m = \frac{q_2 - q_1}{p_2 - p1}$$

and then it has to have an equatiion like $y = mx + c$. We can find c now that we know m by plugging in either P or Q (into $y = mx + c$). Maybe even slicker is to write down

$$y - p_2 = m(x - p_1)$$

as the equation, or even

$$y - p_2 = \left(\frac{q_2 - q_1}{p_2 - p1}\right)(x - p_1)$$

So we certainly don't need determinants to do this, but we are just using it as a warm-up example. The equation of a line is either $x = c$ if it is vertical or $y = mx + c$ if it has a slope. To cover both cases at once we can write

$$ax + by = c$$

or

$$ax + by - c = 0$$

The coefficients a, b and c are the things you need to know to write down the equation. We don't want them all to be zero as $0 = 0$ is not the equation of a line. We want a, b and c not all zero.

If we want the line to go through P and Q, we get two equations to be satisfied by a, b and c:

$$\begin{aligned} ap_1 + bp_2 - c &= 0 \\ aq_1 + bq_2 - c &= 0 \end{aligned}$$

or

$$\begin{aligned} p_1a + p_2b - c &= 0 \\ q_a1 + q_2b - c &= 0 \end{aligned}$$

Since we have 2 equations for the 3 unknowns a, b and c (the poinst P and Q are supposed to be ones we know, and so p_1, p_2, q_1 and q_2 are all things we know) there are either going to be infinitely many solutions or none. We don't have none since $a = b = c = 0$ is a solution. So the equations are consistent. If we were to solve the equations by using (say) Gauss-Jordan elimination on

$$\left[\begin{array}{cccccc} p_1 & p_2 & -1 & : & 0 \\ q_1 & q_2 & -1 & : & 0 \end{array}\right]$$

we would sureley end up with free variables and so infinitely many solutions. Taking a nonzero value for the free variable, we would get a nonzero solution for a, b and c, so an equation for our line.

If we were to add in a third point $R = (r_1, r_2)$ we would get 3 equations rather than just 2

$$\begin{aligned} r_1a + r_2b - c &= 0 \\ p_1a + p_2b - c &= 0 \\ q_a1 + q_2b - c &= 0 \end{aligned}$$

and we don't expect more that one solution $a = b = c = 0$ when we have 3 equations in 3 unknowns.

If we think more carefully, we see that the only time we have other solutions is when there is actually a line through R, P and Q, that is only when R is on the line through P and Q.

Using dertminants we can then say what the condtion is for the above 3 equations to have a solution other than $a = b = c = 0$. It is the same as requiring the matrix

$$\begin{bmatrix} r_1 & r_2 & -1 \\ p_1 & p_2 & -1 \\ q_1 & q_2 & -1 \end{bmatrix}$$

to have no inverse (see Theorem 5.13.2 (b)) and by Theorem 6.4.2 above, that is the same as

$$\det \begin{bmatrix} r_1 & r_2 & -1 \\ p_1 & p_2 & -1 \\ q_1 & q_2 & -1 \end{bmatrix} = 0$$

We are now going to treat R as an unknown or variable point on the line through P and Q. This determinant gives an equation that has to be satisfied by $R = (r_1, r_2)$. To emphasise the different rôle for the point R (varibale) and the points P adn Q (fixed) we'll switch to writing $R = (x, y)$. Finally we get that

$$\det \begin{bmatrix} x & y & -1 \\ p_1 & p_2 & -1 \\ q_1 & q_2 & -1 \end{bmatrix} = 0$$

gives the equation of the line through $P = (p_1, p_2)$ and $Q = (q_1, q_2)$.

6.9.1 Example. Find the equation of the line through $P = (1, 2)$ and $Q = (4, 9)$ using determinants.

Solution: The equation is

$$\det \begin{bmatrix} x & y & -1 \\ 1 & 2 & -1 \\ 4 & 9 & -1 \end{bmatrix} = 0$$

That works out as

$$x \det \begin{bmatrix} 2 & -1 \\ 9 & -1 \end{bmatrix} - y \det \begin{bmatrix} 1 & -1 \\ 4 & -1 \end{bmatrix} + (-1) \det \begin{bmatrix} 1 & 2 \\ 4 & 9 \end{bmatrix} = 0$$

or

$$7x - 3y - 1 = 0$$

Circle through 3 points in the plane. If we start with 3 points in the plane, there is usually one circle through the 3 points. There is a possibility we could pick 3 points in a line and then there would be no regular circle, rather a line, but we'll proceed without worrying about the case of collinear points.

First some words about the equations of circles. The circle centred at (x_0, y_0) with radisu $r > 0$ has equation

$$(x - x_0)^2 + (y - y_0)^2 = r^2$$

but if you multiply it out it is

$$x^2 + y^2 - 2x_0 x - 2y_0 y + (x_0^2 + y_0^2 - r^2) = 0.$$

Taking account of the fact that we can multiply the equation across by any nonzero number and still have the same equation, we'll take the equation to have the form

$$a(x^2 + y^2) + bx + cy + d = 0.$$

A genuine circle has $a \neq 0$ and there is another condition because the radius squared has to be positive, but we can think about that a little later. For now, recall that the unknown coefficients in the equation are the 4 numbers a, b, c, d and they should not all be 0.

Now say we are given 3 points $P = (p_1, p_2)$, $Q = (q_1, q_2)$ and $R = (r_2, r_2)$ on the circle. This gives us 3 equations that the 4 unknowns a, b, c, d have to satisfy:

$$\left\{ \begin{array}{rcl} a(p_1^2 + p_2^2) + bp_1 + cp_2 + d & = & 0 \\ a(q_1^2 + q_2^2) + bq_1 + cq_2 + d & = & 0 \\ a(r_1^2 + r_2^2) + br_1 + cr_2 + d & = & 0 \end{array} \right.$$

Since there are only 3 equations we have to get a nonzero solution. One approach would be to write out the equations as an augmented matrix like

$$\left[\begin{array}{cccccc} p_1^2 + p_2^2 & p_1 & p_2 & 1 & : & 0 \\ q_1^2 + q_2^2 & q_1 & q_2 & 1 & : & 0 \\ r_1^2 + r_2^2 & r_1 & r_2 & 1 & : & 0 \end{array} \right]$$

and find a nonzero solution after using Gauss-Jordan elimination to find solution in terms of a free variable.

The method using determinants is to think of a fourth point (x, y) and add in the condtion that this point should be on the circle. Then we have 4 equations

$$\left\{ \begin{array}{ccccccccc} a(x^2 + y^2) & + & bx & + & cy & + & d & = & 0 \\ a(p_1^2 + p_2^2) & + & bp_1 & + & cp_2 & + & d & = & 0 \\ a(q_1^2 + q_2^2) & + & bq_1 & + & cq_2 & + & d & = & 0 \\ a(r_1^2 + r_2^2) & + & br_1 & + & cr_2 & + & d & = & 0 \end{array} \right.$$

Unless the point (x, y) happens to be on the circle through P, Q and R, there will be no a, b, c and d satisfying all 4 equations. Usually the matrix of coefficients

$$\left[\begin{array}{cccc} x^2 + y^2 & x & y & 1 \\ p_1^2 + p_2^2 & p_1 & p_2 & 1 \\ q_1^2 + q_2^2 & q_1 & q_2 & 1 \\ r_1^2 + r_2^2 & r_1 & r_2 & 1 \end{array} \right]$$

will be invertible and that is the same as there being on the zero solution. Invertible means nonzero determinant.

But then, the equation

$$\det \begin{bmatrix} x^2+y^2 & x & y & 1 \\ p_1^2+p_2^2 & p_1 & p_2 & 1 \\ q_1^2+q_2^2 & q_1 & q_2 & 1 \\ r_1^2+r_2^2 & r_1 & r_2 & 1 \end{bmatrix} = 0$$

is exactly the condtion for (x, y) to be one of the points on the circle through P, Q and R.

Expanding out the determinant we see why we have something that looks like the equation of a circle:

$$\begin{aligned} (x^2+y^2)\det \begin{bmatrix} p_1 & p_2 & 1 \\ q_1 & q_2 & 1 \\ r_1 & r_2 & 1 \end{bmatrix} - x \det \begin{bmatrix} p_1^2+p_2^2 & p_2 & 1 \\ q_1^2+q_2^2 & q_2 & 1 \\ r_1^2+r_2^2 & r_2 & 1 \end{bmatrix} \\ +y \det \begin{bmatrix} p_1^2+p_2^2 & p_1 & 1 \\ q_1^2+q_2^2 & q_1 & 1 \\ r_1^2+r_2^2 & r_1 & 1 \end{bmatrix} - \det \begin{bmatrix} p_1^2+p_2^2 & p_1 & p_2 \\ q_1^2+q_2^2 & q_1 & q_2 \\ r_1^2+r_2^2 & r_1 & r_2 \end{bmatrix} &= 0 \end{aligned}$$

Plane through 3 points in space. The idea is again very similar, maybe marginally more complicated because it is in 3 dimensions.

Say we have 3 points $P = (p_1, p_2, p_3)$, $Q = (q_1, q_2, q_3)$ and $R = (r_1, r_2, r_3)$ and we want to know the equation of the plane in $\mathbb{R}^3$ that contains all 3 points. If the 3 points are collinear, there will be many such planes containing the line, but we will ignore this situation.

One thing to note is that we already know 2 ways to solve this. The most recent way used cross products (see Example 6.7.1). Now yet another way!

We write an equation for our plane

$$ax + by + cz = d$$

but it will be more convenient for us to have it

$$ax + by + cz - d = 0.$$

We should have a, b, c not all zero for a real plane.

Using the 3 points P, Q, R and also another unknown point (x, y, z) on the plane we find 4 equations that a, b, c and d must satisfy

$$\left\{ \begin{array}{ccccccccc} ax & + & by & + & cz & - & d & = & 0 \\ ap_1 & + & bp_2 & + & cp_3 & - & d & = & 0 \\ aq_1 & + & bq_2 & + & cq_3 & - & d & = & 0 \\ ar_1 & + & br_2 & + & cr_3 & - & d & = & 0 \end{array} \right.$$

and then the condtion for there to be a nonzero solution is

$$\det \begin{bmatrix} x & y & z & -1 \\ p_1 & p_2 & p_3 & -1 \\ q_1 & q_2 & q_3 & -1 \\ r_1 & r_2 & r_3 & -1 \end{bmatrix} = 0$$

This is the eqaution for the points (x, y, z) on the plane.

6.9.2 Example. Find the equation of the plane that goes through $(1, 2, 3)$, $(3, 1, 2)$ and $(2, 3, 1)$.

Solution: Let $P = (1, 2, 3)$, $Q = (3, 1, 2)$ and $R = (2, 3, 1)$ and the above determinant equation to get

$$\det \begin{bmatrix} x & y & z & -1 \\ 1 & 2 & 3 & -1 \\ 3 & 1 & 2 & -1 \\ 2 & 3 & 1 & -1 \end{bmatrix} = 0$$

This expands to

$$x \det \begin{bmatrix} 2 & 3 & -1 \\ 1 & 2 & -1 \\ 3 & 1 & -1 \end{bmatrix} - y \det \begin{bmatrix} 1 & 3 & -1 \\ 3 & 2 & -1 \\ 2 & 1 & -1 \end{bmatrix} + z \det \begin{bmatrix} 1 & 2 & -1 \\ 3 & 1 & -1 \\ 2 & 3 & -1 \end{bmatrix} - (-1) \det \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \\ 2 & 3 & 1 \end{bmatrix} = 0$$

After some effort, this works out as

$$3x + 3y + 3z - 18 = 0,$$

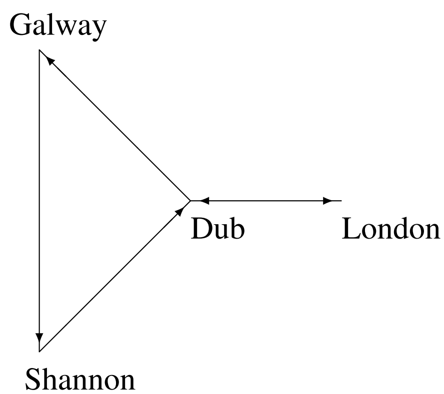
which we could maybe tidy up as $x + y + z = 6$. (Same anwser as last time!)

6.10 Application of matrices to graphs

This topic is perhaps out of place here. We might have fitted it in earlier as we will use matrix multiplication only, nothing about determinants. [This topic is dealt with in §11.7 of Anton & Rorres.]

Graph theory is a subject that is somehow abstract, but at the same time rather close to applications. Mathematically a graph is something that has vertices (also known as nodes) and edges (also called paths) joining some of the nodes to some others. Fairly simple minded examples would be an intercity rail network (nodes would be stations and the edges would correspond to the existence of a direct line from one station to another), or an airline route network, or an ancestral tree graph, or a telecommunications network. For our situation, we will take an example something like an airline network (joining different airports by direct flights), but we will take account of the fact that some airports might not be connected by flights that go direct in both directions.

Here is a route network for a start-up airline that has two routes it flies. One goes Dublin $\to$ London and London $\to$ Dublin, while another route makes a round trip Dublin $\to$ Galway $\to$ Shannon $\to$ Dublin.



The arrows on the edges mean that this is an example of a *directed graph*. Here we allow one-way edges and bi-drectional edges between nodes (or vertices) of the graph, which we draw by indicating arrows.

To get the *vertex matrix* for a graph like this, we first number or order the vertices, for instance

Dub	1
London	2
Galway	3
Shannon	4

and then we make a matrix, a 4×4 matrix in this case since there are 4 vertices, according to the following rules. The entries of the matrix are either 0 or 1. All diagonal entries are 0. The (i, j) entry is 1 if there is a direct edge from vertex i to vertex j (in that direction).

So in our example graph, the vertex matrix is

$$M = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

For instance the first row in $0\,1\,1\,0$ because there is a direct link $1 \to 2$ and $1 \to 3$, but no direct link $1 \to 4$ (Dublin to Galway is not directly linked).

There are other ways to use matrices to deal with graphs. For example, you could try to deal with graphs where there are lengths associated with the edges. The lengths might be the toll payable along a route, or the distance. And sometimes people put these numbers into a matrix instead of the zeros and ones we have used (which say there is or there is not a link). Just to say graph theory is an extensive subject and we won't be doing much about it here.

Here is one result that makes a connection to matrix muuliplication.

6.10.1 Theorem. *If M is the vertex matrix of a directed graph, then the entries of M^2 give the numbers of 2-step (or 2-hop) connections.*

More precisely, the (i, j) entry of M^2 gives the number of ways to go from vertex i to vertex j with exactly 2 steps (or exactly one intermediate vertex).

Similarly M^3 gives the number of 3-step connections, and so on for higher powers of M.

In our example

$$M^2 = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

The diagonal 1's in the matrix correspond to the fact that there is a round trip Dublin $\to$ London $\to$ Dublin (or $1 \to 2 \to 1$) and also London $\to$ Dublin $\to$ London. The 1 in the top right corresponds to the connection Dublin $\to$ Galway $\to$ Shannon.

If we add $M + M^2$ we get nonzero entries in every place where there is a conenction in 1 or 2 steps

$$M + M^2 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

and the zeros off the diagonal there correspond to the connections that can't be made in either a direct connection of a 2-hop connection (which is Gaway $\to$ London and London $\to$ Shannon in our example).

Although we don't see it here the numbers in the matrix M^2 can be bigger than 1 if there are two routes available using 2-hops.

TO BE checked