Vectors, Matrices and Determinants

Simon C Benjamin

Chapter 4. Determinants

Here we'll introduce the determinant, and we'll use it to solve equations (via Cramer's Rule), and to find the inverse of a matrix.

4.1 What is the determinant?

The determinant is a simple number that can be derived from any square matrix. It is, in some ways, similar to the concept of magnitude which we used when thinking of vectors. However, the determinant can be negative.

4.1.1 Notation

We will denote the determinant of a matrix $\mathbf{A}$ using the modulus notation: a vertical line either side, as in $|\mathbf{A}|$. For a matrix written out explicitly with all its elements, we will write the determinant of that matrix by replacing the usual round brackets with vertical lines:

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 7 & 6 & 5 \\ 1 & 0 & -2 \end{pmatrix} \qquad |\mathbf{A}| = \begin{vmatrix} 1 & 2 & 3 \\ 7 & 6 & 5 \\ 1 & 0 & -2 \end{vmatrix}.$$

4.1.2 Definition

We will build up the definition of a determinant by first defining 1×1 and 2×2 determinants, and then expressing the determinants of larger matrices in terms of those.

Firstly, for a matrix with just one element (i.e. one row, one column) the determinant is simply equal to that sole element: $|\mathbf{A}| = A_{1,1}$ in that case.

The determinant of a two-by-two matrix is defined as follows:

$$\mathbf{G} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \qquad |\mathbf{G}| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} \equiv ad - bc.$$

In words, it is the product of the two elements on the diagonal, minus the product of the two off-diagonal elements.

Now, in order to properly define the determinant of some larger matrix, we need to introduce some new terminology. Suppose we have some $n \times n$ matrix $\mathbf{M}$. Focus on some element $M_{i,j}$, and delete the entire row **and** the entire column that contain this element. That is, delete row i and column

j completely. This will give us a new matrix which is $(n-1) \times (n-1)$ in size. Now, the determinant of this new, smaller matrix is called the **minor** of $M_{i,j}$. Example:

$$\text{If } \mathbf{M} = \begin{pmatrix} 1 & 2 & 3 \\ 7 & 6 & 5 \\ 1 & 0 & -2 \end{pmatrix} \text{ then } M_{2,3} = 5, \text{ and } \begin{vmatrix} 1 & 2 \\ 1 & 0 \end{vmatrix} \text{ is minor of } M_{2,3}.$$

Now consider the expression, $(-1)^{i+j}$. This of course means either $+1$ or -1 depending on the i, j indices: if $i + j$ is even, we get $+1$, otherwise -1.

When we multiply the minor of $M_{i,j}$ by this ± 1 factor, we obtain something called the **cofactor** of $M_{i,j}$. Continuing the above example,

$$(-1)^{(2+3)} \begin{vmatrix} 1 & 2 \\ 1 & 0 \end{vmatrix} = - \begin{vmatrix} 1 & 2 \\ 1 & 0 \end{vmatrix} \text{ is the cofactor of } M_{2,3}.$$

OK now we are ready to define the determinant of a general $n \times n$ matrix: choose any row or column, multiply each element by its cofactor, and add up the results. So, in the example of the 3×3 matrix $\mathbf{M}$ above, we would be smart to choose either column two, or row three, because they have a zero which will save us from working out the corresponding cofactor. Let's suppose we choose the third row:

$$\begin{aligned} |\mathbf{M}| &= \begin{vmatrix} 1 & 2 & 3 \\ 7 & 6 & 5 \\ 1 & 0 & -2 \end{vmatrix} \\ &= 1 \times (-1)^4 \begin{vmatrix} 2 & 3 \\ 6 & 5 \end{vmatrix} + 0 \times (-1)^5 \begin{vmatrix} 1 & 3 \\ 7 & 5 \end{vmatrix} + (-2) \times (-1)^6 \begin{vmatrix} 1 & 2 \\ 7 & 6 \end{vmatrix} \\ &= (10 - 18) + 0 + (-2)(6 - 14) \\ &= 8. \end{aligned}$$

Other examples

.

Now, if we need to use this rule to evaluate a 4×4 determinant, then first we will have to expand it into the sum of four 3×3 determinants, and then expand each of those into three 2×2 determinants. A lot of work! Fortunately, there are some basic rules that allow us to spot shortcuts.

- The determinant is ZERO if,
  a. All the elements of one row (or one column) are zero, or
  b. Two rows are identical, or two columns are identical, or
  c. Two rows are proportional, or two columns are proportional.
- If two rows, or two columns, are interchanged then the determinant changes sign.
- If each element of one row is multiplied by some number k, then the determinant is multiplied by k. Similarly for a column.
- The value of the determinant is unchanged if
  a. We transpose the matrix, making rows into columns, or
  b. We add onto each element of one row k times the corresponding element in some other row (and similarly for columns).

There is an exercise on the problem sheet to prove these rules. Using them we can sometimes spot when a determinant is going to evaluate to zero, without actually performing the calculation. And even if our determinant is not zero, we may be able to increase the number of zeros in the determinant and thus make the calculation easier.

Examples

4.2 Solving equations

Here we will look at two methods for solving a set of linear equations. The first, called row reduction, is the most practical for large sets of equations. The second approach, Cramer's rule, requires a lot of brute force calculation but it gives the solution as a single equation.

4.2.1 Method 1: Row reduction

Suppose we have a set of equations such as

$$\begin{aligned} a_1x + b_1y + c_1z &= k_1 \\ a_2x + b_2y + c_2z &= k_2 \\ a_3x + b_3y + c_3z &= k_3 \end{aligned}$$

where these constants a_i, b_i and c_i are some particular values we've been given. Then we write down a rectangular matrix as follows:

$$\mathbf{M} = \begin{pmatrix} a_1 & b_1 & c_1 & k_1 \\ a_2 & b_2 & c_2 & k_2 \\ a_3 & b_3 & c_3 & k_3 \end{pmatrix}.$$

We now perform a series of **row manipulations** which are any of these:

a. Interchange any two rows.

b. Multiply or divide any row by some constant.

c. Add or subtract any row from any other.

The aim is to the manipulate the matrix into the form

$$\mathbf{M_{new}} = \begin{pmatrix} d & e & f & g \\ 0 & h & i & j \\ 0 & 0 & l & m \end{pmatrix}.$$

Notice the elements below the diagonal are all zero — the matrix is said to be in echelon form. The elements on the diagonal are called "pivots" in this technique. Now the allowed row operations simply correspond to

a. Reordering the list of equations

b. Multiplying an equation by some number, on both the LHS and RHS.

c. Finding sum or difference of two equations: LHS1-LHS2=RHS1-RHS2.

So our altered matrix $\mathbf{M_{new}}$ is still a representation of three legitimate equations. But now we can simply read off the solution for z from the bottom line,

$$lz = m \quad \text{so} \quad z = \frac{m}{l}.$$

Then the next line up gives us y, since this line tells us $hy + iz = j$ and of course we now have z. Similarly, we finally get x. This method of using matrices is not covered further in the present course, but there is a good explanation in the book "Mathematical Methods in the Physical Sciences" by Mary Boas (Chp. 3 Section 2).

Row reduction example

4.2.2 Method 2: Cramer's rule

Using this rule involves a lot of computation, but the advantage is that it provides a one-step formula for the solutions. First we'll derive it for the simple case of two equations and two unknowns. Then we'll just state the general rule.

Suppose that we are given equations like,

$$\begin{aligned} a_1x + b_1y &= k_1 \\ a_2x + b_2y &= k_2 \end{aligned}$$

Let's note down how we would write these in matrix format (for later use!)

$$\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} k_1 \\ k_2 \end{pmatrix} \quad \text{or} \quad \mathbf{D}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} k_1 \\ k_2 \end{pmatrix} \quad \text{if} \quad \mathbf{D} = \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix}.$$

OK let's see one way of solving these equations working by normal methods. We could multiply the first equation by b_2, the second by b_1, and subtract to eliminate y, i.e.

$$(a_1b_2 - a_2b_1)x + (b_1b_2 - b_2b_1)y = k_1b_2 - k_2b_1$$

$$x = \frac{k_1b_2 - k_2b_1}{a_1b_2 - a_2b_1}$$

and similarly, we could have multiplied by the a coefficients to eliminate x,

$$y = \frac{a_1k_2 - a_2k_1}{a_1b_2 - a_2b_1}.$$

But we notice that we can express this using determinants:

$$x = \frac{\begin{vmatrix} k_1 & b_1 \\ k_2 & b_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}} \qquad y = \frac{\begin{vmatrix} a_1 & k_1 \\ a_2 & k_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}}.$$

We see that the denominator of both solutions is just $|\mathbf{D}|$, where $\mathbf{D}$ was the matrix containing the coefficients of x and y from above. What about the numerators? They are just the determinants of related matrices: in the case of the x solution, it is the matrix obtained by replacing a_1 and a_2 (the coefficients of x) with k_1 and k_2 instead. Similarly, in the solution for y we swap out the b terms.

Now, it turns out that a generalisation of this rule works for solving n equations in n unknowns.

- Write down the $n \times n$ matrix $\mathbf{D}$ containing the coefficients of the n unknowns. The pattern should be the same as we used above: each column is the set of coefficients of one of the unknowns.
- The solution for some particular unknown, call it u, is given by the ratio $u = \frac{|\mathbf{A}|}{|\mathbf{D}|}$, where matrix $\mathbf{A}$ is obtained from $\mathbf{D}$ by replacing the column of coefficients of u with the constants from the RHS of the equations.

Let's see the rule explicitly for a set of three equations.

$$\begin{aligned} a_1x + b_1y + c_1z &= k_1 \\ a_2x + b_2y + c_2z &= k_2 \\ a_3x + b_3y + c_3z &= k_3 \end{aligned}$$

Which we can write as

$$\mathbf{D}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} k_1 \\ k_2 \\ k_3 \end{pmatrix} \quad \text{where} \quad \mathbf{D} = \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}.$$

Here the blue highlighting is so we can see the pattern in the solutions:

$$x = \frac{\begin{vmatrix} k_1 & b_1 & c_1 \\ k_2 & b_2 & c_2 \\ k_3 & b_3 & c_3 \end{vmatrix}}{|\mathbf{D}|} \qquad y = \frac{\begin{vmatrix} a_1 & k_1 & c_1 \\ a_2 & k_2 & c_2 \\ a_3 & k_3 & c_3 \end{vmatrix}}{|\mathbf{D}|} \qquad z = \frac{\begin{vmatrix} a_1 & b_1 & k_1 \\ a_2 & b_2 & k_2 \\ a_3 & b_3 & k_3 \end{vmatrix}}{|\mathbf{D}|}$$

Example

4.3 Matrix inversion

We write the inverse of a matrix $\mathbf{M}$ as $\mathbf{M^{-1}}$. It is the matrix with the properties,

$$\mathbf{M^{-1}M} = \mathbf{MM^{-1}} = \mathbf{I}$$

where $\mathbf{I}$ is the identity matrix. We notice immediately that $\mathbf{M}$ must be square, otherwise it would not be the case that both of these multiplications give the same $\mathbf{I}$ (we would have a shape mismatch). If we think of $\mathbf{M}$ as a transformation, then $\mathbf{M^{-1}}$ is the reverse transformation that "undoes" the effect of $\mathbf{M}$.

We might ask, how can we find $\mathbf{M^{-1}}$ given a particular $\mathbf{M}$? And does the inverse always exist?

As with the case of equation solving, there are two methods we can try. One of them is needs less work but you have to think about the right 'moves'. The other is a 'brute force' method that gives you the answer as a fixed formula, but it needs lots of computation!

4.3.1 Row manipulation method for finding inverse

This is explained in Boas, Mathematical Methods for the Physical Sciences, Chp. 3 Section 6. Here we will briefly summarise. Suppose we want to find the inverse of

$$\mathbf{M} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

Where a_{ij} are particular numbers we've been given. Then we write out a 3×6 matrix, formed by the columns of $\mathbf{M}$ followed by the columns of the identity matrix. This is,

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & 1 & 0 & 0 \\ a_{21} & a_{22} & a_{23} & 0 & 1 & 0 \\ a_{31} & a_{32} & a_{33} & 0 & 0 & 1 \end{pmatrix}.$$

Now we apply the same row reduction rules that we saw earlier for equation solving. Our aim this time is to reach a form where the columns of the identity matrix are on the left:

$$\begin{pmatrix} 1 & 0 & 0 & a & b & c \\ 0 & 1 & 0 & d & e & f \\ 0 & 0 & 1 & g & h & i \end{pmatrix}.$$

At this point, the numbers in the last three columns are just the elements of our desired inverse matrix!

$$\mathbf{M^{-1}} = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}.$$

The same technique works for a matrix of any size (provided of course that the inverse exists, we'll come to that in a moment).

Example

4.3.2 Determinant method for finding inverse

The second method for finding the inverse involves computing a lot of determinants! But it has the merit that it is a one line formula.

$$\mathbf{M^{-1}} = \frac{1}{|\mathbf{M}|}\mathbf{C^T} \quad \text{where} \quad C_{i,j} = \text{ cofactor of } M_{i,j}.$$

So this is saying that we need to work out a new matrix $\mathbf{C}$, and each element of C is the cofactor of the corresponding element in $\mathbf{M}$. Then we take the

transpose of $\mathbf{C}$, divide the whole thing by the determinant of $\mathbf{M}$, and that's the inverse!

Recall that the cofactor of some element $M_{i,j}$ is the determinant of the matrix you get by deleting row i and column j from $\mathbf{M}$, times by a factor $(-1)^{i+j}$. So, if $\mathbf{M}$ is a 3×3 matrix, we will need to work out 9 cofactors, and each is a 2×2 determinant. Plus we need to work out the determinant of $\mathbf{M}$ itself. So that's ten determinants! This is not too bad, since 2×2 determinants are easy, but if we tried the same method to invert a 4×4 matrix, then it would be a **lot** of work.

Example

4.3.3 Existence of the inverse

The above approach for finding $\mathbf{M^{-1}}$ involved dividing by $|\mathbf{M}|$. This of course requires that $|\mathbf{M}| \neq 0$ or otherwise the definition would give us an infinity. And in fact, it turns out that this is a requirement for an inverse to exist: matrix $\mathbf{M}$ has an inverse if $\mathbf{M}$ is square and $|\mathbf{M}| \neq 0$.

4.3.4 Orthogonal matrices

There is a special family of matrices for which the inverse is easy to find: the **orthogonal** matrices. Now a matrix is orthogonal if,

- For every row: The sum of the squares of all the elements unity. Same for every column. ('Unity' is a fancy word for 'one').
- For every pair of rows: If we multiply each element by its partner and sum up (similar to dot product), the total is zero. Similarly for every pair of columns.

Now, such an orthogonal matrix has a beautifully simply propoerty: its inverse is its transpose! So,

$$\mathbf{M^{-1}} = \mathbf{M^{T}} \quad \text{if } \mathbf{M} \text{ is orthogonal.}$$

Of the matrices that we have met so far, the rotation matrices are orthogonal (check it!).
Example