20 Determinants and Inverses

In what follows, we will only be interested in linear maps that map from a space back to the same space. Once we have established a basis for a space, a linear map is uniquely determined by its action on the basis vectors. Thanks to the linearity properties, the mapping of all other vectors can be determined by writing them in terms of the basis vectors:

$$f(\mathbf{v}) = f\left(\sum_i v_i \mathbf{e}_i\right) = \sum_i v_i f(\mathbf{e}_i)\,. \tag{1}$$

The columns of the matrix that effects the transformation are the result of applying the transformation to the basis vectors. In two dimensions, this guarantees that rectangles will always be mapped to parallelograms. In three dimensions, cuboids are mapped to parallelepipeds, and this pattern generalises to any dimension.

20.1 Determinants

Consider the maps represented by the matrices

$$\begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \quad \begin{pmatrix} 1 & 2 \\ -1 & 3 \end{pmatrix} \quad \begin{pmatrix} 2 & 4 \\ 1 & 2 \end{pmatrix}\,. \tag{2}$$

In the first example, the basis vectors $(1, 0)$ and $(0, 1)$ are doubled and tripled in size, respectively. The square with area 1 defined by the basis vectors, then, is mapped to a rectangle of area $2 \times 3 = 6$. In fact, this means that any region of space will be mapped to a region with 6 times the area under this map. This scale factor is called the **determinant** of the matrix. To see how to calculate it in general, consider the second example above. Here, the basis vectors $(1, 0)$ and $(0, 1)$ are mapped to $\mathbf{x} = (1, -1)$ and $\mathbf{y} = (2, 3)$. This defines a parallelogram with sides of $|\mathbf{x}|$ and $|\mathbf{y}|$, with some angle θ between them. Recall that the area of a parallelogram is base $\times$ height or $|\mathbf{x}| \cdot |\mathbf{y}| \sin\theta$. Apart from a unit vector, this is essentially the cross product of the two vectors, $\mathbf{x}$ and $\mathbf{y}$. As such, we can deduce the general formula for the determinant from that of the cross product. For a two-dimensional matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\,, \tag{3}$$

the determinant is given by

$$\det(A) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc\,. \tag{4}$$

So the determinant of our second example is

$$\begin{vmatrix} 1 & 2 \\ -1 & 3 \end{vmatrix} = (1)(3) - (-1)(2) = 5\,, \tag{5}$$

and all regions will be scaled by a factor of 5 under this transformation.

Let's see what this means for the final example in Eq. 2. There the determinant is

$$\begin{vmatrix} 2 & 4 \\ 1 & 2 \end{vmatrix} = (2)(2) - (4)(1) = 0\,, \tag{6}$$

showing that all regions are mapped to regions with no area. This is because the basis vectors in this case are mapped to $(2, 1)$ and $(4, 2)$. Since these are parallel, this is a degenerate transformation. It maps a two-dimensional space on to a one-dimensional subspace: all parallelograms get squashed into a single line.

As a further example, consider the matrix

$$\begin{pmatrix} 2 & 3 \\ 3 & 4 \end{pmatrix} . \tag{7}$$

In this case, the determinant is -1, but what does it mean to have a negative scale factor? Note that it does not mean a reduction in size, as would, say $\det(A) = 1/2$. To see how to intepret this, it will help to consider the difference between the matrices

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \text{ and } \begin{pmatrix} \cos\frac{\pi}{2} & -\sin\frac{\pi}{2} \\ \sin\frac{\pi}{2} & \cos\frac{\pi}{2} \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} , \tag{8}$$

with determinants -1 and 1, respectively.

The second of these is a rotation by $\pi/2$, moving the $\widehat{\mathbf{i}}$ basis vector to the positive y axis and the $\widehat{\mathbf{j}}$ vector to the negative x axis. Similarly, it takes a square defined by these vectors and rotates it about the origin. The first matrix does something very different: just like the second, it moves $\widehat{\mathbf{i}}$ to the positive y axis, but unlike the second, it moves $\widehat{\mathbf{i}}$ to the **positive** x axis. It has **swapped** the positions of the two basis vectors. Notice that this leaves the square defined by $\widehat{\mathbf{i}}$ and $\widehat{\mathbf{j}}$ in the same place, but we have inverted it. If we were to label the corners of the square, there would be no way to map the old square to the new without leaving the two-dimensional space. The transformation with the positive determinant preserves the orientation of the square, but the negative determinant does not.

20.2 Combining Maps

Since the determinant of a map is just the overall scale factor of the transformation, we can immediately deduce two important properties of the determinant. Firstly, the determinant cannot depend on the particular basis chosen. So two different matrices that represent the same linear map in different coordinate systems must have the same determinant. Secondly, the overall scale factor for a composite map must be the product of the scale factors for the individual maps. That is, for a composite linear map $h = g \circ f$, we have $\det(h) = \det(g) \times \det(f)$. Recall that a composite map has a matrix representation that is the product of the matrices for the individual maps. So we must have

$$\det(AB) = \det(A)\det(B)\,. \tag{9}$$

It is not obvious from the formula for the determinant that this must always be true, but it follows easily from the definition.

20.3 Inverse Matrices

When mapping from a space to the same space, generally we can undo any transformation by reversing it. The composite transformation in this case is the trivial "do nothing" transformation. This tells us that the matrices effecting the transformations must obey

$$BA = I \tag{10}$$

where I is the identity matrix. We say that B is the **inverse** of A and denote it $B = A^{-1}$. Notice that the concept of an inverse matrix does not have any sensible meaning for a matrix that is not square. From the definition, it should also be clear that the inverse is unique: given a transformation, there is only one way to undo it.

There is a type of transformation that cannot be undone, and which therefore must correspond to a matrix that cannot be inverted. If a transformation collapses two-dimensional regions to a one-dimensional line, then infinitely many initial points are mapped to each final point. In this case, given a final point, we cannot say where it came from, and so cannot undo the transformation. Recall that these transformations are those with a determinant of 0. So just as we cannot define a multiplicative inverse of 0 in the real numbers, for matrices, we cannot define an inverse for those with a determinant of 0.

Calculating inverse matrices can be a tedious task in general, but in the case of 2×2 matrices, it is straightforward. Consider the product

$$\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} da - bc & db - bd \\ -ca + ac & -cb + ad \end{pmatrix} = \begin{pmatrix} ad - bc & 0 \\ 0 & ad - bc \end{pmatrix}. \tag{11}$$

This is just $\det(A)I$, so we have

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{\det(A)}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}, \tag{12}$$

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

20.4 Higher Dimensions

The determinant of a 3×3 matrix is the scale factor for volumes rather than areas. A negative determinant still has the same interpretation of failing to preserve the relative orientation of the basis vectors, so now switching any two changes the sign. A zero determinant still means collapsing the space on to a space of smaller dimension, though this can now mean collapsing to a plane as well as to a line. Calculating the determinant of a 3×3 matrix is a little less straightforward than for 2×2 but there is still a memorable formula for it:

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a(ei - fh) - b(di - fg) + c(dh - eg)\,. \tag{13}$$

To see that this makes some sense as a generalisation of the 2×2 case, consider the matrix

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & a & b \\ 0 & c & d \end{pmatrix}\,. \tag{14}$$

This matrix performs some general transformation of the y-z plane, but its only effect on the x coordinate is to double it. It makes sense in this case that the determinant would be

$$2 \times \det \begin{pmatrix} a & b \\ c & d \end{pmatrix}\,. \tag{15}$$

Finding the inverse of a 3×3 or larger matrix is best done through one of various algorithms rather than using a formula. That said, we will give the general formula for completeness. For 3×3 matrices, we have

$$A^{-1} = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}^{-1} = \frac{1}{\det(A)} \begin{pmatrix} ei - fh & ch - bi & bf - ce \\ fg - di & ai - cg & cd - af \\ dh - eg & bg - ah & ae - bd \end{pmatrix}\,. \tag{16}$$

The general pattern followed here (and which works for any size square matrix) is:

1. Replace each entry of the matrix with the determinant of its **minor**: the $(n-1) \times (n-1)$ matrix left when that entry's row and column are deleted.
2. Multiply each entry by ± 1 in a "checkerboard" pattern with the top left entry positive.
3. Transpose the result.
4. Divide by the determinant.