LECTURE 6

Successive linear transformations: matrix multiplication

The matrices R_x, R_y and R_z rotate a vector by an angle θ around the x-, y- and z-axes, respectively.

$$\mathsf{R}_x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix} \qquad \mathsf{R}_y = \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix} \qquad \mathsf{R}_z = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

A question arises: suppose we rotate around one axis and then another. What have we done overall? This brings us on to the subject of matrix multiplication.

Matrix multiplication

Let A and B be two transformation matrices. We can apply B to transform a vector $\mathbf{v}$ into $\mathbf{v}'$:

$$\mathbf{v}' = \mathsf{B}\mathbf{v}. \tag{6.1}$$

But then, this transformed vector can *itself* be transformed by A into a third vector, $\mathbf{v}''$:

$$\mathbf{v}'' = \mathsf{A}\mathbf{v}'. \tag{6.2}$$

Overall, we have

$$\mathbf{v}'' = \mathsf{A}\left(\mathsf{B}\mathbf{v}\right).$$

We will show that the overall transformation can be performed by a single matrix, C, calculated as a *matrix-matrix product* from A and B.

Remember from last lecture that (6.1) is just shorthand for

$$v_i' = \sum_{j=1}^{3} b_{ij} v_j. \tag{6.3}$$

[see (5.2)]. And likewise, (6.2) means

$$v_i'' = \sum_{j=1}^{3} a_{ij} v_j'. \tag{6.4}$$

To work out how to go from $\mathbf{v}$ to $\mathbf{v}''$ directly, we just substitute (6.3) into (6.4). To do this, though, we first need to manipulate the subscripts a little.

Note first that the summation index, j, in (6.3) is a *dummy index*. It does not matter which letter we use, since its job is only to be replaced by the numbers that we are summing over. So (6.3) is exactly the same as

$$v'_i = \sum_{k=1}^{3} b_{ik} v_k. \tag{6.5}$$

Next, we are using i to specify the component we are interested in. We could obviously use j instead, in which case it must be that

$$v'_j = \sum_{k=1}^{3} b_{jk} v_k. \tag{6.6}$$

And now, we can substitute this expression into (6.4), to get

$$v''_i = \sum_{j=1}^{3} a_{ij} \left(\sum_{k=1}^{3} b_{jk} v_k \right). \tag{6.7}$$

Our final task is to reorder the summations. First, we can put the a_{ij} inside the brackets. This is because $a(x_1 + x_2 + \cdots) = (ax_1 + ax_2 + \cdots)$:

$$v''_i = \sum_{j=1}^{3} \left(\sum_{k=1}^{3} a_{ij} b_{jk} v_k \right). \tag{6.8}$$

Next, it does not matter which order we sum things up in. So (6.8) becomes

$$v''_i = \sum_{k=1}^{3} \left(\sum_{j=1}^{3} a_{ij} b_{jk} v_k \right). \tag{6.9}$$

Finally, v_k is the same for all terms in the j sum. So we can take it out as a constant:

$$v''_i = \sum_{k=1}^{3} \left(\sum_{j=1}^{3} a_{ij} b_{jk} \right) v_k. \tag{6.10}$$

The key observation is that (6.10) now has the form

$$v''_i = \sum_{k=1}^{3} c_{ik} v_k \tag{6.11}$$

with

$$c_{ik} = \sum_{j=1}^{3} a_{ij} b_{jk}. \tag{6.12}$$

Equation (6.11) says that $\mathbf{v}''$ *can be obtained as a single linear transformation of* $\mathbf{v}$, by multiplying by a matrix C. And (6.12) *tells us how to calculate* C from A and B.

We say that C is the matrix-matrix product of A and B, which we denote in the obvious way as

$$\mathsf{C} = \mathsf{AB}.$$

Example 1. Let A and B be the matrices that rotate around the y- and z-axes, respectively, each by an angle $\pi/2$.

$$\mathsf{A} = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \qquad \mathsf{B} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Suppose $\mathbf{v}$ is transformed by B and then by A. Find the matrix that does the transformation in one step.

This is just an application of (6.12). We need to find the 9 elements of matrix C. Ordinarily, one would probably start with c_{11}. For ease of explaining the method, though, let's look at an element with different row and column indices, e.g. c_{21}.

Equation (6.12) gives us

$$c_{21} = \sum_{j=1}^{3} a_{2j} b_{j1}. \tag{6.13}$$

Notice that c_{21} depends on the elements a_{2j}. In other words, we need to look at elements from row 2 of A. Likewise, we need to look at elements from column 1 of B.

The formula says: take the jth element from row 2 of A. Multiply it by the jth element from column 1 of B. Do this for all j and sum everything up.

Put even more simply: *we take the dot product of row* 2 *of* A *and column* 1 *of* B.

$$\begin{pmatrix} 0 & 0 & -1 \\ \boxed{0 \quad 1 \quad 0} \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \boxed{\begin{matrix} 0 \\ 1 \\ 0 \end{matrix}} & \begin{matrix} -1 \\ 0 \\ 0 \end{matrix} & \begin{matrix} 0 \\ 0 \\ 1 \end{matrix} \end{pmatrix} = \begin{pmatrix} . & . & . \\ \boxed{0 \times 0 + 1 \times 1 + 0 \times 0} & . & . \\ . & . & . \end{pmatrix}.$$

Repeating this procedure for all the other 8 elements, we obtain

$$\mathsf{C} = \mathsf{AB} = \begin{pmatrix} 0 & 0 & -1 \\ 1 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix}$$

Example 2. Repeat Example 1, but for the case where $\mathbf{v}$ is transformed first by A and then by B.

In this case, we are saying that

$$\mathbf{v}'' = \mathsf{B}\,(\mathsf{A}\mathbf{v})\,,$$

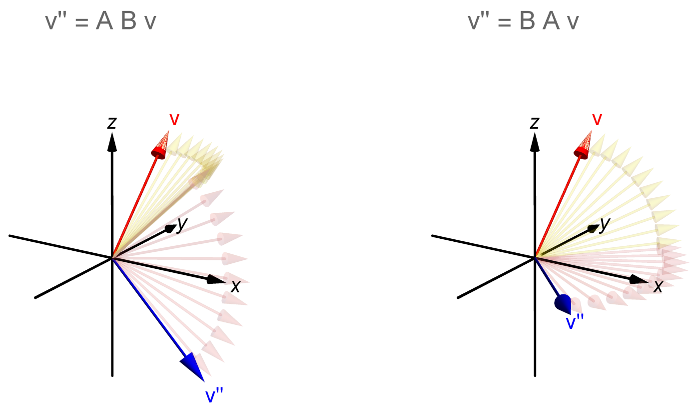
so we need the matrix product BA rather than AB. We find, after some calculation,

$$\mathsf{BA} = \begin{pmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ 1 & 0 & 0 \end{pmatrix}$$

The key observation is that $\mathsf{AB} \neq \mathsf{BA}$. *It matters which way round we multiply the matrices!*

The terminology here is that matrix multiplication is, in general, *non-commutative*. In *some* cases, it turns out that $\mathsf{AB} = \mathsf{BA}$, but this does not hold generally for all matrices.

To visualise why AB and BA are generally different, look at the the figures below. They show the effect of rotating by $\pi/2$ around two successive axes, as in Examples 1 and 2. The left figure rotates around the z-axis first, then the y-axis. The right figure does it the other way round. It is clear that the order of the two transformations makes a difference.



More than two transformations

The arguments made above can be generalised to any number of successive transformations. For example, if A, B and C describe three transformations, such that

$$\mathbf{v}''' = \mathsf{C}\,(\mathsf{B}\,(\mathsf{A}\mathbf{v}))\,,$$

then one can go from $\mathbf{v}$ to $\mathbf{v}'''$ in a single step using some matrix D. This matrix is the product

$$\mathsf{D} = \mathsf{CBA}.$$

Notice that we have dropped all the brackets. We can calculate D in two ways, and they give us the same result:

$$\mathsf{D} = \mathsf{C}\,(\mathsf{BA}) = (\mathsf{CB})\,\mathsf{A}.$$

The technical term for is *associativity*. Matrix multiplication is associative. However, notice that we must keep the order the same from left to right: it would be a mistake to calculate D as $\mathsf{B}(\mathsf{CA})$, for example.

Determinant of a product

It is helpful, at this point on our introductory tour of matrices, to ask the question: what is the determinant of a product of matrices?

The answer can be proven using the method of *elementary row operations*. But there is a way to justify it using what we already know.

Remember that $|\det \mathsf{A}|$ is a volume scaling factor. Specifically, A transforms the **i**, **j** and **k** basis vectors to $\mathsf{A}\mathbf{i}$ etc., which define a parallelepiped. We saw that $|\det \mathsf{A}|$ is the volume of this parallelepiped, relative to the volume of the cube defined by the original basis vectors.

If we consider two successive transformations, then the overall volume scaling factor should just be the product of the scaling factors for each transformation separately, i.e.

$$|\det(\mathsf{AB})| = |\det \mathsf{A}||\det \mathsf{B}|.$$

And the sign of the determinant tells us whether the transformation flips from a right- to a left-handed basis. If neither or both of A and B flip the handedness then, overall, AB will keep the basis right-handed and $\det(\mathsf{AB})$ will be positive. Otherwise, $\det(\mathsf{AB})$ will be negative. This means that the sign of $\det(\mathsf{AB})$ is the product of the signs of $\det \mathsf{A}$ and $\det \mathsf{B}$.

Putting this together, we find that

$$\boxed{\det(\mathsf{AB}) = \det \mathsf{A} \det \mathsf{B}.} \qquad (6.14)$$

Example 3. Confirm (6.14) for the matrices of Example 1.

It follows in a few lines that

$$\det \mathsf{A} = 1$$

$$\det \mathsf{B} = 1$$

$$\det(\mathsf{AB}) = \det \mathsf{C} = 1$$

Hence (rather trivially in this case), (6.14) holds.

The inverse of a transformation

In the previous lecture, we saw that a matrix transformation can only be inverted (i.e. reversed) if the determinant of the matrix is non-zero.

The notation we use for the inverse matrix is

$$\mathsf{A}^{-1} \text{ is the inverse of } \mathsf{A}.$$

When $\det \mathsf{A} \neq 0$, an $n \times n$ matrix A has an $n \times n$ inverse matrix that satisfies

$$\mathsf{A}^{-1}\mathsf{A} = \mathsf{A}\mathsf{A}^{-1} = \mathtt{I}_n \qquad (6.15)$$

In fact, if $\mathsf{A}^{-1}\mathsf{A} = \mathtt{I}_n$, then it can be *proven* that $\mathsf{A}\mathsf{A}^{-1} = \mathtt{I}_n$ also.

Example 4. If

$$\mathsf{A} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

confirm that the following matrix is the inverse of A:

$$\mathsf{A}^{-1} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

We check that this inverse satisfies (6.15):

$$\mathsf{A}^{-1}\mathsf{A} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathtt{I}_2.$$

$$\mathsf{A}\mathsf{A}^{-1} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathtt{I}_2.$$

The inverse matrix can be computed in several ways. One is left as an example for the problem sheet. Another involves the method of Gaussian elimination.

We can combine the result for $\det(\mathsf{AB})$ with the property of the inverse matrix, to obtain another way to understand the condition for invertibility. We have

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

But we also know that

$$\det(\mathsf{A}^{-1}\mathsf{A}) = \det \mathtt{I}_\mathsf{n} = 1$$

(write out an identity matrix and take its determinant, if this is not immediately obvious). Combining these two results gives

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

Clearly this is impossible if $\det \mathsf{A} = 0$. Hence, as before, we conclude that A^{-1} cannot exist if $\det \mathsf{A} = 0$.

Beware!

The fact that matrix inverses do not always exist can lead to mistakes being made. People often assume (without realising it) that a matrix is invertible, when in fact it is not.

Here's an example. What does the following imply, if all the matrices are $n \times n$?

$$\mathsf{AB} = \mathsf{AC}.$$

It is very tempting to say that this implies

$$\mathsf{B} = \mathsf{C}.$$

If we were a bit more awake, we might realise that A could be the *null matrix* $\mathsf{0}_n$ (an $n \times n$ matrix full of zeros). So we might say that

$$\mathsf{B} = \mathsf{C} \qquad \text{or} \qquad \mathsf{A} = \mathsf{0}_n.$$

Both these answers are wrong, unfortunately. To turn $\mathsf{AB} = \mathsf{AC}$ into $\mathsf{B} = \mathsf{C}$, we need to multiply both sides (on the left) by the inverse of A:

We use here the fairly obvious result that multiplying a matrix by the identity matrix does not change it, i.e. $\mathtt{I}_n\mathsf{B} = \mathsf{B}$, etc..

$$\begin{aligned} \mathsf{AB} &= \mathsf{AC} \\ \mathsf{A}^{-1}\mathsf{AB} &= \mathsf{A}^{-1}\mathsf{AC} \\ \mathtt{I}_n\mathsf{B} &= \mathtt{I}_n\mathsf{C} \\ \mathsf{B} &= \mathsf{C} \end{aligned}$$

If A^{-1} does not exist, we cannot do that! To convince you that this really is an issue, take a look at the following

Example 5. If

$$\mathsf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \qquad \mathsf{B} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \qquad \mathsf{C} = \begin{pmatrix} 2 & 4 \\ 2 & 2 \end{pmatrix}$$

show that $\mathsf{AB} = \mathsf{BC}$, even though $\mathsf{A} \neq 0_n$ and $\mathsf{B} \neq \mathsf{C}$.

It is straightforward to show that

$$\mathsf{AB} = \mathsf{AC} = \begin{pmatrix} 4 & 6 \\ 4 & 6 \end{pmatrix}.$$

So if all we know is that $\mathsf{AB} = \mathsf{AC}$, we most definitely *cannot* conclude that $\mathsf{B} = \mathsf{C}$ or $\mathsf{A} = 0_n$.

In some sense, a matrix potentially lacking an inverse is the matrix generalisation of a number potentially being zero. When dealing with the algebraic equation

$$ab = ac$$

we (should) know that we cannot divide both sides by a unless we know that a is non-zero. When dealing with matrices, we must now be careful to check whether inverses exist.