LECTURE 2

Inverse and determinant of a product

A cautionary tale

We have all made a mistake like the following. Suppose we are given the equation

$$x^2 = x.$$

Then it is tempting to 'cancel out' a factor of x, and say that

$$x = 1.$$

This conclusion is wrong. When we 'cancelled out' the factor of x, we were actually dividing both sides of the equation by it. The problem is that x might be zero, in which case we are not allowed to divide by it. We should have rewritten our starting equation as

$$x^2 - x = 0,$$

from which it follows that

$$x(x-1) = 0$$

and hence $x = 0$ or $x = 1$.

In practice, we all do the last step by setting each factor to zero. A formal way to prove it would be: if $x \neq 0$, we can divide by x and then we see that $x = 1$. If $x \neq 1$, we can divide by $(x-1)$ and see that $x = 0$.

A similar mistake is commonly made when people work with matrices. Suppose I told you that

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

You might first be tempted to say that $\mathsf{B} = \mathsf{C}$. This is wrong. You might say 'Aha! I'm not falling for that trick again. Either $\mathsf{A} = 0$ or $\mathsf{B} = \mathsf{C}$'. This is also wrong. Take a look at the following example:

By $\mathsf{0}$, we have in mind the *null matrix*, which is the matrix full of zeroes.

Example 1. 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 $\mathsf{AB} = \mathsf{AC}$

It is straightforward to confirm that

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

So this is an example where $\mathsf{AB} = \mathsf{AC}$, but neither $\mathsf{A} = 0$ nor $\mathsf{B} = \mathsf{C}$. What is going on?

The mistake is that we're thinking of matrix multiplication in the same way that we think of multiplying numbers. You must not do this! Matrix multiplication does not obey the same rules.

In particular, **there is no such thing as dividing by a matrix**.

The only way to 'cancel out' matrices in equations is to multiply by inverse matrices. The complication arises because, as we have seen, inverse matrices do not always exist.

Let's look at $\mathsf{AB} = \mathsf{AC}$ more carefully. Suppose that A is an invertible matrix. Then we can multiply both equations on the left by the inverse of A:

Remember, as we saw in Lecture 1, if we multiply one side of an equation by something on the left, we must multiply the other side on the left as well. The order matters.

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

So if A^{-1} exists, it follows that $\mathsf{B} = \mathsf{C}$. However, if A is not invertible, we cannot draw that conclusion.

I will leave it to you in the problems to work out the condition under which $\mathsf{AB} = \mathsf{AC}$ implies that $\mathsf{A} = 0$.

The take-home messages from this are:

We cannot 'cancel out' matrices on both sides of an equation

and

We must check that inverse matrices exist, before we use them in proofs.

Homogeneous linear equations: redux number 2

This brings us back to our systems of homogeneous linear equations. For 3 equations in 3 unknowns, we have

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = 0 \tag{2.1}$$
$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = 0 \tag{2.2}$$
$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = 0. \tag{2.3}$$

As we saw yesterday, this can be written as

$$\mathsf{A}\mathbf{x} = \mathbf{0}$$

where $\mathbf{0}$ is the zero vector. What are the solutions of this system of equations?

From what we've just seen, we must not make the mistake of 'dividing by A' and arguing that $\mathbf{x} = \mathbf{0}$. We can only multiply both sides by the inverse of A, and then only if that inverse exists.

So, *if* A is invertible, we can say that

$$\mathsf{A}^{-1}\mathsf{A}\mathbf{x} = \mathsf{A}^{-1}\mathbf{0}$$

and therefore that $\mathbf{x} = \mathbf{0}$. This is the trivial solution, where all the unknowns are zero.

On the other hand, if A is not invertible, *the conclusion that* $\mathbf{x} = \mathbf{0}$ *does not follow*.

What's the condition for A being invertible? Well, it's that $\det \mathsf{A} \neq 0$. So this recovers exactly the same conclusion as we've seen (twice now) already: the system of homogeneous equations has only trivial solutions ($\mathbf{x} = \mathbf{0}$) if the determinant of the coefficient matrix is non-zero.

The rest of the lecture will derive some important results that we will need next week. It will also give us the opportunity to practice manipulating equations involving matrices. Don't think of all the proofs as things you must memorise: try to see the bigger picture of how to work with matrix equations. You'll get some practice yourself using the problems.

Determinant of a matrix product

Somewhat unusually, I'm going to give you the result without proving it. It turns out that

$$\boxed{\det(\mathsf{AB}) = \det \mathsf{A} \det \mathsf{B}. \qquad (2.4)}$$

There isn't a nice proof that I can show you, without going into a lot more detail than we need.

But I'm not going to give up without a fight. I'll try to make the result plausible. Remember from last Term that the modulus of the determinant is a *volume scaling factor*. Specifically, A transforms the $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ 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, i.e. transform by B then by A, it seems reasonable that the volume scaling factor for the overall transformation, AB, is the product of the individual scaling factors. Hence we would expect that

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

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 these two things together, the result of (2.4) is at least plausible, hopefully. An example might provide further comfort.

Example 2. Confirm (2.4) when

$$\det \mathsf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

and

$$\det \mathsf{B} = \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix}.$$

Straightforwardly,

$$\mathsf{AB} = \begin{pmatrix} 19 & 22 \\ 43 & 50 \end{pmatrix}.$$

Hence $\det \mathsf{A} = -2$, $\det \mathsf{B} = -2$, and $\det(\mathsf{AB}) = 4 = (-2) \cdot (-2)$.

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 $\det \mathsf{A} = 0$.

Inverse of a matrix product

This is an important result. If we transform a vector by B and then by A, what is the inverse transformation, if it exists?

It's clear that we won't get very far if A and B can't individually be inverted. So let's start by assuming that they can. Then, let

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

We said that B was invertible, so we can post-multiply both sides by B^{-1} to get

$$\mathsf{ABB}^{-1} = \mathsf{CB}^{-1}.$$

This simplifies to

$$\mathsf{A} = \mathsf{C}\mathsf{B}^{-1}.$$

Likewise, post-multiplying on the right by A^{-1} gives

$$\mathtt{I} = \mathsf{C}\mathsf{B}^{-1}\mathsf{A}^{-1}. \tag{2.5}$$

Now, we'd like to get that C over to the left. To do so, we note that C must be invertible, because

$$\det \mathsf{C} = \det(\mathsf{AB}) = (\det \mathsf{A})(\det \mathsf{B});$$

we already assumed that the determinants of A and B were non-zero, so it follows that $\det \mathsf{C}$ must also be non-zero. Hence, we can pre-multiply (2.5) by C^{-1} to get

$$\mathsf{C}^{-1} = \mathsf{B}^{-1}\mathsf{A}^{-1}.$$

Substituting in the defintion of C gives us

$$\boxed{(\mathsf{AB})^{-1} = \mathsf{B}^{-1}\mathsf{A}^{-1} \qquad \text{if } \mathsf{A} \text{ and } \mathsf{B} \text{ are invertible.}} \tag{2.6}$$

Left inverse equals right inverse

I'm including this, because it's something people often ask me. We said that the matrix inverse satisfies

$$\mathsf{A}^{-1}\mathsf{A} = \mathsf{A}\mathsf{A}^{-1} = \mathtt{I}.$$

Is this two conditions on the inverse matrix, or just one? It turns out that it's just one. Lets see why.

Let's assume that A is an invertible matrix that satisfies

$$\mathsf{AB} = \mathtt{I}. \tag{2.7}$$

By taking the determinant of both sides, we get

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

Using the rule of the previous section, (2.4), this implies that

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

We assumed that A is invertible, so $\det \mathsf{A} \neq 0$, and hence

$$\det \mathsf{B} = \frac{1}{\det \mathsf{A}} \neq 0.$$

So therefore B is also invertible, i.e. B^{-1} exists.

From here, it is relatively plain sailing. We can pre-multiply (2.7) by B

$$\mathsf{BAB} = \mathsf{B} \tag{2.8}$$

and then post-multiply by B^{-1}:

$$\mathsf{BABB}^{-1} = \mathsf{BB}^{-1}, \tag{2.9}$$

which simplifies to

$$\mathsf{BA} = \mathsf{I}. \tag{2.10}$$

So, if we identify B in (2.7) as the inverse of A by definition, then we've shown that

$$\text{If} \qquad \mathsf{A}\mathsf{A}^{-1} = \mathsf{I} \qquad \text{then} \qquad \mathsf{A}^{-1}\mathsf{A} = \mathsf{I}$$

The same inverse matrix works both ways round.

To prove the implication the other way, we simply set $\mathsf{A} = \mathsf{B}^{-1}$.

Notice that it was important to check that B^{-1} exists, before we post-multiply by it.

There is, at most, one inverse

We've checked that the same inverse matrix works on the left and the right. But maybe there is more than one inverse matrix that does this job. How do we know there's only the one that we constructed yesterday?

Suppose we found two matrices, B and C, that are both inverses of A. In other words,

$$\mathsf{AB} = \mathsf{BA} = \mathsf{I}$$

and

$$\mathsf{AC} = \mathsf{CA} = \mathsf{I}.$$

Then we can write that

$$\begin{aligned} \mathsf{B} &= \mathsf{BI} \\ &= \mathsf{B}(\mathsf{AC}) \\ &= (\mathsf{BA})\mathsf{C} \\ &= \mathsf{IC} \\ &= \mathsf{C}. \end{aligned}$$

We've shown that the two inverse matrices that we found are actually the same. Since the proof applies to *any* two inverse matrices that we find, the conclusion has to be that *there is only one possible inverse matrix*.

Therefore, the matrix we constructed in Lecture 1 must be this *unique* inverse matrix. It *exists* whenever $\det \mathsf{A} \neq 0$, and satisfies both the left inverse and right inverse properties.