Chapter 8

Matrices II: inverses

We have learnt how to add subtract and multiply matrices but we have not defined division. The reason is that in general it cannot always be defined. In this chapter, we shall explore when it can be. All matrices will be square and I will always denote an identity matrix of the appropriate size.

8.1 What is an inverse?

The simplest kind of linear equation is $ax = b$ where a and b are scalars. If $a \neq 0$ we can solve this by multiplying by a^{-1} on both sides to get $a^{-1}(ax) = a^{-1}b$. We now use associativity to get $(a^{-1}a)x = a^{-1}b$. Finally, $a^{-1}a = 1$ and so $1x = a^{-1}b$ and this gives $x = a^{-1}b$. The number a^{-1} is the multiplicative inverse of the non-zero number a. We now try to emulate this approach for the matrix equation

$$A\mathbf{x} = \mathbf{b}.$$

We suppose that there is a matrix B such that $BA = I$.

- Multiply on the left both sides of our equation $A\mathbf{x} = \mathbf{b}$ to get $B(A\mathbf{x}) = B\mathbf{b}$. Because order matters when you multiply matrices, which side you multiply on also matters.

- Use associativity of matrix mulitiplication to get $(BA)\mathbf{x} = B\mathbf{b}$.

- Now use our assumption that $BA = I$ to get $I\mathbf{x} = B\mathbf{b}$.

- Finally, we use the properties of the identity matrix to get $\mathbf{x} = B\mathbf{b}$.

We appear to have solved our equation, but we need to check it. We calculate $A(B\mathbf{b})$. By associativity this is $(AB)\mathbf{b}$. At this point we *also* have to assume that $AB = I$. This gives $I\mathbf{b} = \mathbf{b}$, as required. We conclude that in order to copy the method for solving a linear equation in one unknown, our coefficient matrix A must have the property that there is a matrix B such that

$$AB = I = BA.$$

We take this as the basis of the following definition.

A matrix A is said to be *invertible* if we can find a matrix B such that $AB = I = BA$. The matrix B we call it an *inverse* of A, and we say that the matrix A is *invertible*. Observe that A has to be square. A matrix that is not invertible is said to be *singular*.

Example 8.1.1. A real number r regarded as a 1×1 matrix is invertible if and only if it is non-zero, in which case an inverse is its reciprocal. Thus our definition of matrix inverse directly generalizes what we mean by the inverse of a number.

It's clear that if A is a zero matrix, then it can't be invertible just as in the case of real numbers. However, the next example shows that even if A is not a zero matrix, then it need not be invertible.

Example 8.1.2. Let A be the matrix

$$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$$

We shall show that there is no matrix B such that $AB = I = BA$. Let B be the matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

From $BA = I$ we get

$$a = 1 \text{ and } a = 0.$$

It's impossible to meet both these conditions at the same time and so B doesn't exist.

On the other hand here is an example of a matrix that is invertible.

Example 8.1.3. Let

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{pmatrix}$$

and

$$B = \begin{pmatrix} 1 & -2 & 5 \\ 0 & 1 & -4 \\ 0 & 0 & 1 \end{pmatrix}$$

Check that $AB = I = BA$. We deduce that A is invertible with inverse B.

As always, in passing from numbers to matrices things become more complicated. Before going any further, I need to clarify one point which will at least make our lives a little simpler.

Lemma 8.1.4. *Let A be invertible and suppose that B and C are matrices such that*

$$AB = I = BA \text{ and } AC = I = CA.$$

Then $B = C$.

Proof. Multiply $AB = I$ both sides on the left by C. Then $C(AB) = CI$. Now $CI = C$, because I is the identity matrix, and $C(AB) = (CA)B$ since matrix multiplication is associative. But $CA = I$ thus $(CA)B = IB = B$. It follows that $C = B$. $\square$

The above result tells us that *if* a matrix A is invertible *then* there is only one matrix B such that $AB = I = BA$. We call the matrix B *the* inverse of A. It is usually denoted by A^{-1}. It is important to remember that we can *only* write A^{-1} if we know that A is invertible. In the following, we describe some important properties of the inverse of a matrix.

Lemma 8.1.5.

1. *If A is invertible then A^{-1} is invertible and its inverse is A.*

2. *If A and B are both invertible and AB is defined then AB is invertible with inverse $B^{-1}A^{-1}$.*

3. *If $A_1, \ldots, A_n$ are all invertible and $A_1 \ldots A_n$ is defined then $A_1 \ldots A_n$ is invertible and its inverse is $A_n^{-1} \ldots A_1^{-1}$.*

Proof. (1) This is immediate from the equations $A^{-1}A = I = AA^{-1}$.

(2) Show that

$$AB(B^{-1}A^{-1}) = I = (B^{-1}A^{-1})AB.$$

(3) This follows from (2) above and induction. □

We shall deal with the practical computation of inverse later. Let me conclude this section by returning to my original motivation for introducing an inverse.

Theorem 8.1.6 (Matrix inverse method)**.** *A system of linear equations*

$$A\mathbf{x} = \mathbf{b}$$

in which A is invertible has the unique solution

$$\mathbf{x} = A^{-1}\mathbf{b}.$$

Proof. Observe that

$$A(A^{-1}\mathbf{b}) = (AA^{-1})\mathbf{b} = I\mathbf{b} = \mathbf{b}.$$

Thus $A^{-1}\mathbf{b}$ is *a* solution. It is unique because if $\mathbf{x}'$ is any solution then

$$A\mathbf{x}' = \mathbf{b}$$

giving

$$A^{-1}(A\mathbf{x}') = A^{-1}\mathbf{b}$$

and so

$$\mathbf{x}' = A^{-1}\mathbf{b}.$$

□

Example 8.1.7. We shall solve the following system of equations using the matrix inverse method

$$\begin{aligned} x + 2y &= 1 \\ 3x + y &= 2 \end{aligned}$$

Write the equations in matrix form.

$$\begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

Determine the inverse of the coefficient matrix. In this case, you can check that this is the following

$$A^{-1} = -\frac{1}{5}\begin{pmatrix} 1 & -2 \\ -3 & 1 \end{pmatrix}$$

Now we may solve the equations. From $A\mathbf{x} = \mathbf{b}$ we get that $\mathbf{x} = A^{-1}\mathbf{b}$. Thus in this case

$$\mathbf{x} = -\frac{1}{5}\begin{pmatrix} 1 & -2 \\ -3 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} \frac{3}{5} \\ \frac{1}{5} \end{pmatrix}$$

Thus $x = \frac{3}{5}$ and $y = \frac{1}{5}$. Finally, it is always a good idea to check the solutions. There are two (equivalent) ways of doing this. The first is to check by direct substitution

$$x + 2y = \frac{3}{5} + 2 \cdot \frac{1}{5} = 1$$

and

$$3x + y = 3 \cdot \frac{3}{5} + \frac{1}{5} = 2$$

Alternatively, you can check by matrix mutiplication

$$\begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix}\begin{pmatrix} \frac{3}{5} \\ \frac{1}{5} \end{pmatrix}$$

which gives

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

You can see that both calculations are, in fact, identical.

8.2 Determinants

The obvious questions that arise from the previous section are how do we decide whether a matrix is invertible or not and, if it is invertible, how do we compute its inverse? The material in this section is key to answering both of these questions. I shall define a number, called the determinant, that can be calculated from any *square* matrix. Unfortunately, the definition is unmotivated but it will justify itself by being extremely useful as well as having interesting properties.

Let A be a square matrix. We denote its *determinant* by $\det(A)$ or by replacing the round brackets of the matrix A with straight brackets. It is defined inductively: this means that I define an $n \times n$ determinant in terms of $(n-1) \times (n-1)$ determinants.

- The determinant of the 1×1 matrix $\begin{pmatrix} a \end{pmatrix}$ is a.
- The determinant of the 2×2 matrix

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

  denoted

  $$\begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

  is the number $ad - bc$.
- The determinant of the 3×3 matrix

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

  denoted

  $$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix}$$

  is the number

  $$a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix}$$

We could in fact define the determinant of any square matrix of whatever size in much the same way. However, we shall limit ourselves to calculating the determinants of 3×3 matrices at most. It's important to pay attention to the signs in the definition. You multiply alternately by plus one and minus one

$$+ \quad - \quad + \quad - \quad \ldots$$

More generally, the signs are found by computing $(-1)^{i+j}$ where i is the row of the element and j is the column. In the above definition, we are taking

$i = 1$ throughout. We have defined determinants by *expansion along the first row* but in fact you can expand them along any row and also along any column. The fact that the same answers arise however you expand the determinant is a first indication of their remarkable properties.

Examples 8.2.1.

1.
$$\begin{vmatrix} 2 & 3 \\ 4 & 5 \end{vmatrix} = 2 \times 5 - 3 \times 4 = -2.$$

2.
$$\begin{vmatrix} 1 & 2 & 1 \\ 3 & 1 & 0 \\ 2 & 0 & 1 \end{vmatrix} = 1 \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} - 2 \begin{vmatrix} 3 & 0 \\ 2 & 1 \end{vmatrix} + \begin{vmatrix} 3 & 1 \\ 2 & 0 \end{vmatrix} = -7$$

For our subsequent work, I want to single out two properties of determinants.

Theorem 8.2.2. *Let A and B be square matrices having the same size. Then*

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

Proof. The result is true in general, but I shall only prove it for 2×2 matrices. Let

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ and } B = \begin{pmatrix} e & f \\ g & h \end{pmatrix}$$

We prove directly that $\det(AB) = \det(A)\det(B)$. First

$$AB = \begin{pmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{pmatrix}$$

Thus

$$\det(AB) = (ae + bg)(cf + dh) - (af + bh)(ce + dg).$$

The first bracket multiplies out as

$$acef + adeh + bcgf + bdgh$$

and the second as

$$acef + adfg + bceh + bdgh.$$

Subtracting these two expressions we get

$$adeh + bcgf - adfg - bceh.$$

Now we calculate $\det(A)\det(B)$. This is just

$$(ad - bc)(eh - fg)$$

which multiplies out to give

$$adeh + bcfg - adfg - bceh.$$

Thus the two sides are equal, and we have proved the result. □

I shall mention one other property of determinants that we shall need when we come to study vectors and which will be useful in developing the theory of inverses. It can be proved in the 2×2 and 3×3 cases by direct verification.

Theorem 8.2.3. *Let A be a square matrix and let B be obtained from A by interchanging any two columns. Then* $\det(B) = -\det(A)$.

An important consequence of the above result is the following.

Proposition 8.2.4. *If two columns of a determinant are equal then the determinant is zero.*

Proof. Let A be a matrix with two columns equal. Then if we swap those two columns the matrix remains unchanged. Thus by Theorem 8.2.3, we have that $\det A = -\det A$. It follows that $\det A = 0$. □

I have defined determinants *algebraically* in this section. In Chapter 9, I shall also describe determinants *geometrically*.

Exercises 8.2

1. Compute the following determinants.

   (a)
$$\begin{vmatrix} 1 & -1 \\ 2 & 3 \end{vmatrix}$$

(b)
$$\begin{vmatrix} 3 & 2 \\ 6 & 4 \end{vmatrix}$$

(c)
$$\begin{vmatrix} 1 & -1 & 1 \\ 2 & 3 & 4 \\ 0 & 0 & 1 \end{vmatrix}$$

(d)
$$\begin{vmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 2 & 3 & 1 \end{vmatrix}$$

(e)
$$\begin{vmatrix} 2 & 2 & 2 \\ 1 & 0 & 5 \\ 100 & 200 & 300 \end{vmatrix}$$

(f)
$$\begin{vmatrix} 1 & 3 & 5 \\ 102 & 303 & 504 \\ 1000 & 3005 & 4999 \end{vmatrix}$$

(g)
$$\begin{vmatrix} 1 & 1 & 2 \\ 2 & 1 & 1 \\ 1 & 2 & 1 \end{vmatrix}$$

(h)
$$\begin{vmatrix} 15 & 16 & 17 \\ 18 & 19 & 20 \\ 21 & 22 & 23 \end{vmatrix}$$

2. Solve $\begin{vmatrix} 1-x & 4 \\ 2 & 3-x \end{vmatrix} = 0.$

3. Calculate
$$\begin{vmatrix} x & \cos x & \sin x \\ 1 & -\sin x & \cos x \\ 0 & -\cos x & -\sin x \end{vmatrix}$$

4. Prove that
$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = 0$$
if, and only if, one column is a scalar multiple of the other. Hint: consider two cases: $ad = bc \neq 0$ and $ad = bc = 0$ for this case you will need to consider various possibilities.

8.3 When is a matrix invertible?

Recall from Theorem 8.2.2 that
$$\det(AB) = \det(A)\det(B).$$
I use this property below to get a necessary condition for a matrix to be invertible.

Lemma 8.3.1. *If A is invertible then* $\det(A) \neq 0$.

Proof. By assumption, there is a matrix B such that $AB = I$. Take determinants of both side of the equation
$$AB = I$$
to get
$$\det(AB) = \det(I).$$
By the key property of determinants recalled above
$$\det(AB) = \det(A)\det(B)$$
and so
$$\det(A)\det(B) = \det(I).$$
But $\det(I) = 1$ and so
$$\det(A)\det(B) = 1.$$
In particular, $\det(A) \neq 0$. □

Are there any other properties that a matrix must satisfy in order to have an inverse? The answer is, surprisingly, no. Specifically, I shall prove that a square matrix A is invertible if, and only if, $\det A \neq 0$. This solves the theoretical question concerning inverses whereas the practical one of actually computing inverses is dealt with in the next section.

I start with a 2×2 matrix A where everything is easy to calculate. Let

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

We construct a new matrix as follows. Replace each entry a_{ij} of A by the element you get when you cross out the ith row and jth column. Thus we get

$$\begin{pmatrix} d & c \\ b & a \end{pmatrix}$$

We now use the following matrix of signs

$$\begin{pmatrix} + & - \\ - & + \end{pmatrix},$$

where the entry in row i and column j is the sign of $(-1)^{i+j}$, to get

$$\begin{pmatrix} d & -c \\ -b & a \end{pmatrix}$$

We now take the transpose of this matrix to get the matrix we call the *adjugate* of A

$$\operatorname{adj}(A) = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

The defining characteristic of the adjugate is that

$$A\operatorname{adj}(A) = \det(A)I = \operatorname{adj}(A)A$$

which can easily be checked. We deduce from the defining characteristic of the adjugate that if $\det(A) \neq 0$ then

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

We have therefore proved the following.

Proposition 8.3.2. *A 2×2 matrix is invertible if and only if its determinant is non-zero.*

Example 8.3.3. Let $A = \begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix}$. Determine if A is invertible and, if it is, find its inverse, and check the answer. We calculate $\det(A) = -5$. This is non-zero, and so A is invertible. We now form the adjugate of A:

$$\operatorname{adj}(A) = \begin{pmatrix} 1 & -2 \\ -3 & 1 \end{pmatrix}$$

Thus the inverse of A is

$$A^{-1} = -\frac{1}{5}\begin{pmatrix} 1 & -2 \\ -3 & 1 \end{pmatrix}$$

We now check that $AA^{-1} = I$ (to make sure that we haven't made any mistakes).

We now consider the general case. Here I will simply sketch out the argument. Let A be an $n \times n$ matrix with entries a_{ij}. We define its adjugate as the result of the following sequence of operations.

- Pick a particular row i and column j. If we cross out this row and column we get an $(n-1) \times (n-1)$ matrix which I shall denote by $M(A)_{ij}$. It is called a *submatrix* of the original matrix A.
- The determinant $\det(M(A)_{ij})$ is called the *minor* of the element a_{ij}.
- Finally, if we multiply $\det(M(A)_{ij})$ by the corresponding sign we get the *cofactor* $c_{ij} = (-1)^{i+j} \det(M(A)_{ij})$ of the element a_{ij}.
- If we replace each element a_{ij} by its cofactor, we get the matrix $C(A)$ of cofactors of A.
- The **transpose** of the matrix of cofactors $C(A)$, denoted $\operatorname{adj}(A)$, is called the *adjugate*[1] *matrix of* A. Thus the adjugate is the transpose of the matrix of signed minors.

The crucial property of the adjugate is described in the next result.

Theorem 8.3.4. *For any square matrix A, we have that*

$$A(\operatorname{adj}(A)) = \det(A)I = (\operatorname{adj}(A))A.$$

[1] This odd word comes from Latin and means 'yoked together'.

Proof. We have verified the above result in the case of 2×2 matrices. I shall now prove it in the case of 3×3 matrices by means of an argument that generalizes. Let $A = (a_{ij})$ and we write

$$B = \operatorname{adj}(A) = \begin{pmatrix} c_{11} & c_{21} & c_{31} \\ c_{12} & c_{22} & c_{32} \\ c_{13} & c_{23} & c_{33} \end{pmatrix}$$

We shall compute AB. We have that

$$(AB)_{11} = a_{11}c_{11} + a_{12}c_{12} + a_{13}c_{13} = \det A$$

by expanding the determinant along the top row. The next element is

$$(AB)_{12} = a_{11}c_{21} + a_{12}c_{22} + a_{13}c_{23}.$$

But this is the determinant of the matrix

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

which, having two rows equal, must be zero by Proposition 8.2.4. This pattern now continues with all the off-diagonal entries being zero for similar reasons and the diagonal entries all being the determinant. □

We may now prove the main theorem of this section.

Theorem 8.3.5 (Existence of inverses)**.** *Let A be a square matrix. Then A is invertible if and only if* $\det(A) \neq 0$. *When A is invertible, its inverse is given by*

$$A^{-1} = \frac{1}{\det(A)} \operatorname{adj}(A).$$

Proof. Let A be invertible. By our lemma above, $\det(A) \neq 0$ and so we can form the matrix

$$\frac{1}{\det(A)} \operatorname{adj}(A).$$

We now calculate

$$A \frac{1}{\det(A)} \operatorname{adj}(A) = \frac{1}{\det(A)} A \operatorname{adj}(A) = I$$

by our theorem above. Thus A has the advertized inverse.

Conversely, suppose that $\det(A) \neq 0$. Then again we can form the matrix

$$\frac{1}{\det(A)} \operatorname{adj}(A)$$

and verify that this is the inverse of A and so A is invertible. $\square$

Example 8.3.6. Let

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 0 & 1 \\ -1 & 1 & 2 \end{pmatrix}$$

We show that A is invertible and calculate its inverse. Calculate first its adjugate. The matrix of minors is

$$\begin{pmatrix} -1 & 5 & 2 \\ 1 & 5 & 3 \\ 2 & -5 & -4 \end{pmatrix}$$

The matrix of cofactors is

$$\begin{pmatrix} -1 & -5 & 2 \\ -1 & 5 & -3 \\ 2 & 5 & -4 \end{pmatrix}$$

The adjugate is the transpose of the matrix of cofactors

$$\begin{pmatrix} -1 & -1 & 2 \\ -5 & 5 & 5 \\ 2 & -3 & -4 \end{pmatrix}$$

We now calculate A times it adjugate. We get $-5I$. It follows that $\det A = -5$. The inverse of A is the adjugate with each entry divided by the determinant of A

$$A^{-1} = -\frac{1}{5} \begin{pmatrix} -1 & -1 & 2 \\ -5 & 5 & 5 \\ 2 & -3 & -4 \end{pmatrix}$$

The Moore-Penrose Inverse

We have proved that a square matrix has an inverse if, and only if, it has a non-zero determinant. For rectangular matrices, the existence of an inverse doesn't even come up for discussion. However, in later applications of matrix theory it is very convenient if every matrix have an 'inverse'. Let A be any matrix. We say that A^+ is its *Moore-Penrose inverse* if the following conditions hold:

1. $A = AA^+A$.
2. $A^+ = A^+AA^+$.
3. $(A^+A)^T = A^+A$.
4. $(AA^+)^T = AA^+$.

It is not obvious, but every matrix A has a Moore-Penrose inverse A^+ and, in fact, such an inverse is uniquely determined by the above four conditions. In the case, where A is invertible in the vanilla-sense, its Moore-Penrose inverse is just its inverse. But even singular matrices have Moore-Penrose inverse. You can check that the matrix defined below satisfies the four conditions above

$$\begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix}^+ = \begin{pmatrix} 0 \cdot 002 & 0 \cdot 006 \\ 0 \cdot 04 & 0 \cdot 12 \end{pmatrix}$$

The Moore-Penrose inverse can be used to find approximate solutuions to systems of linear equations that might otherwise have no solution.

Exercises 8.3

1. Use the adjugate method to compute the inverses of the following matrices. In each case, check that your solution works.

   (a) $\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$

   (b) $\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$

(c) $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$

(d) $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 0 & 1 \\ -1 & 1 & 2 \end{pmatrix}$

(e) $\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 3 \\ 1 & 2 & 4 \end{pmatrix}$

(f) $\begin{pmatrix} 2 & 2 & 1 \\ -2 & 1 & 2 \\ 1 & -2 & 2 \end{pmatrix}$

8.4 Computing inverses

The practical way to compute inverses is to use elementary row operations. I shall first describe the method and then I shall prove that it works. Let A be a square $n \times n$ matrix. We want to determine whether it is invertible and, if it is, we want to calculate its inverse. We shall do this at the same time and we shall not need to calculate a determinant. We write down a new kind of *augmented matrix* this time of the form $B = (A \mid I)$ where I is the $n \times n$ identity matrix. The first part of the algorithm is to carry out elementary row operations on B guided by A. Our goal is to convert A into a row echelon matrix. This will have zeros below the leading diagonal. We are interested in what entries lie on the leading diagonal. If one of them is zero we stop and say that A is not invertible. If all of them are 1 then the algorithm continues. We now use the 1's that lie on the leading diagonal to remove all element above each 1. Our original matrix B now has the following form $(I \mid A')$. I claim that $A' = A^{-1}$. I shall illustrate this method by means of an example.

Example 8.4.1. Let

$$A = \begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix}$$

We shall show that A is invertible and calculate its inverse. We first write down the augmented matrix

$$\left(\begin{array}{rrr|rrr} -1 & 1 & 1 & 1 & 0 & 0 \\ 1 & -1 & 1 & 0 & 1 & 0 \\ 1 & 1 & -1 & 0 & 0 & 1 \end{array}\right)$$

We now carry out a sequence of elementary row operations to get the following where the matrix to the left of the partition is an echelon matrix.

$$\left(\begin{array}{rrr|rrr} 1 & -1 & -1 & -1 & 0 & 0 \\ 0 & 1 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & \frac{1}{2} & \frac{1}{2} & 0 \end{array}\right)$$

The leading diagonal contains only 1's and so our original matrix *is* invertible. We now use these 1's to insert zeros above using elementary row operations.

$$\left(\begin{array}{rrr|rrr} 1 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & \frac{1}{2} & \frac{1}{2} & 0 \end{array}\right)$$

It follows that the inverse of A is

$$A^{-1} = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$$

At this point, it is always advisable to check that $A^{-1}A = I$ in fact rather than just in theory.

We now need to explain why this method works. An $n \times n$ matrix E is called an *elementary matrix* if it is obtained from the $n \times n$ identity matrix by means of a single elementary row operation.

Example 8.4.2. Let's find all the 2×2 elementary matrices. The first one is obtained by interchanging two rows and so is

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Next we obtain two matrices by multiplying each row by a non-zero scalar λ

$$\begin{pmatrix} \lambda & 0 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}$$

Finally, we obtain two matrices by adding a scalar multiple of one row to another row

$$\begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ \lambda & 1 \end{pmatrix}$$

There are now two key results we shall need.

Lemma 8.4.3.

1. *Let B be obtained from A by means of a single elementary row operation ρ. Thus $B = \rho(A)$. Let $E = \rho(I)$. Then $B = EA$.*

2. *Each elementary row matrix is invertible.*

Proof. (1) This has to be verified for each of the three types of elementary row operation. I shall deal with the third class of such operations: $R_j \leftarrow R_j + \lambda R_i$. Apply this elementary row operation to the $n \times n$ identity matrix I to get the matrix E. This agrees with the identity matrix everywhere *except* the jth row. There it has a λ in the ith column and, of course, a 1 in the jth column. We now calculate the effect of E on any suitable matrix A. Then EA will be the same as A except in the jth row. This will consist of the jth row of A to which λ times the ith row of A has been added.

(2) Let E be the elementary matrix that arises from the elementary row operation ρ. Thus $E = \rho(I)$. Let ρ' be the elementary row operation that undoes the effect of ρ. Thus $\rho\rho'$ and $\rho'\rho$ are both identity functions. Let $E' = \rho'(I)$. Then $E'E = \rho'(E) = \rho'(\rho(I)) = I$. Similarly, $EE' = I$. It follows that E is invertible with inverse E'. $\square$

Example 8.4.4. We give an example of 2×2 elementary matrices. Consider the elementary matrix

$$E = \begin{pmatrix} 1 & 0 \\ \lambda & 1 \end{pmatrix}$$

which is obtained from the 2×2 identity matrix by carrying out the elementary row operation $R_2 \leftarrow R_2 + \lambda R_1$. We now calculate the effect of this matrix when we multiply it into the following matrix

$$A = \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix}$$

and we get

$$EA = \begin{pmatrix} a & b & c \\ \lambda a + d & \lambda b + e & \lambda c + f \end{pmatrix}$$

But this matrix is what we would get if we applied the elementary row operation directly to the matrix A.

We may now prove that our the elementary row operation method for calculating inverses which we described above really works.

Proposition 8.4.5. *If $(I \mid B)$ can be obtained from $(A \mid I)$ by means of elementary row operations then A is invertible with inverse B.*

Proof. Let the elementary row operations that transform $(A \mid I)$ to $(I \mid B)$ be $\rho_1, \ldots, \rho_n$ in this order. Thus

$$(\rho_n \ldots \rho_1)(A) = I \text{ and } (\rho_n \ldots \rho_1)(I) = B.$$

Let E_i be the elementary matrix corresponding to the elementary row operation ρ_i. Then

$$(E_n \ldots E_1)A = I \text{ and } (E_n \ldots E_1)I = B.$$

Now the matrices E_i are invertible and so

$$A = (E_n \ldots E_1)^{-1} \text{ and } B = E_n \ldots E_1.$$

Thus B is the inverse of A as claimed. $\square$

Exercises 8.4

1. Use elementary row operations to compute the inverses of the following matrices. In each case, check that your solution works.

   (a) $\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$

   (b) $\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$

   (c) $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$

(d) $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 0 & 1 \\ -1 & 1 & 2 \end{pmatrix}$

(e) $\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 3 \\ 1 & 2 & 4 \end{pmatrix}$

(f) $\begin{pmatrix} 2 & 2 & 1 \\ -2 & 1 & 2 \\ 1 & -2 & 2 \end{pmatrix}$

8.5 The Cayley-Hamilton theorem

The goal of this section is to prove a major theorem about square matrices. It is true in general, although I shall only prove it for 2×2 matrices. It provides a first indication of the importance of certain polynomials in studying matrices.

Let A be a square matrix. We can therefore form the product AA which we write as A^2. When it comes to multiplying A by itself three times there are apparently two possibilities: $A(AA)$ and $(AA)A$. However, matrix multiplication is associative and so these two products are equal. We write this as A^3. In general $A^{n+1} = AA^n = A^nA$. We define $A^0 = I$, the identity matrix the same size as A. The usual properties of exponents hold

$$A^mA^n = A^{m+n} \text{ and } (A^m)^n = A^{mn}.$$

One important consequence is that powers of A commute so that

$$A^mA^n = A^nA^m.$$

We can form powers of matrices, multiply them by scalars and add them together. We can therefore form sums like

$$A^3 + 3A^2 + A + 4I.$$

In other words, we can substitute A in the polynomial

$$x^3 + 3x^2 + x + 4$$

remembering that $4 = 4x^0$ and so has to be replaced by $4I$.

Example 8.5.1. Let $f(x) = x^2 + x + 2$ and let

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

We calculate $f(A)$. Remember that $x^2 + x + 2$ is really $x^2 + x + 2x^0$. Replace x by A and so x^0 is replaced by A^0 which is I. We therefore get $A^2 + A + 2I$ and calculating gives

$$\begin{pmatrix} 5 & 2 \\ 2 & 3 \end{pmatrix}$$

It is important to remember that when a square matrix A is substituted into a polynomial, you *must* replace the constant term of the polynomial by the constant term times the identity matrix. The identity matrix you use will have the same size as A.

We now come to an important extension of what we mean by a root. If $f(x)$ is a polynomial and A is a square matrix, we say that A is a *matrix root* of $f(x)$ if $f(A)$ is the zero matrix.

Let A be a square $n \times n$ matrix. Define

$$\chi_A(x) = \det(A - xI).$$

Then $\chi_A(x)$ is a polynomial of degree n called the *characteristic polynomial* of A. It is worth observing that when $x = 0$ we get that $\chi_A(0) = \det(A)$, which is therefore the value of the constant term of the characteristic polynomial. The following theorem was proved by Cayley in the paper cited at the beginning of Chapter 7. In fact, he proved the 2×2 and 3×3 cases explicitly, and then wrote the following

> "...but I have not thought it necessary to undertake the labour of a formal proof of the theorem in the general case of a matrix of any degree."

Such were the days.

Theorem 8.5.2 (Cayley-Hamilton)**.** *Every square matrix is a root of its characteristic polynomial.*

Proof. I shall only prove this theorem in the 2×2 case, though it is true in general. Let

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

Then from the definition the characteristic polynomial is

$$\begin{vmatrix} a-x & b \\ c & d-x \end{vmatrix}$$

Thus

$$\chi_A(x) = x^2 - (a+d)x + (ad-bc).$$

We now calculate $\chi_A(A)$ which is just

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^2 - (a+d)\begin{pmatrix} a & b \\ c & d \end{pmatrix} + \begin{pmatrix} ad-bc & 0 \\ 0 & ad-bc \end{pmatrix}$$

This simplifies to

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

which proves the theorem in this case. The general proof uses the adjugate matrix of $A - xI$. $\square$

Example 8.5.3. Consider the matrix

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

Then the characteristic polynomial of A is

$$\chi_A(x) = \begin{vmatrix} 1-x & 1 \\ 1 & -x \end{vmatrix} = x^2 - x - 1.$$

An easy calculation shows that

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^2 - \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

There is one very nice application of this theorem.

Proposition 8.5.4. *Let A be an invertible $n \times n$ matrix. Then the inverse of A can be written as a polynomial in A of degree $n-1$.*

Proof. We may write $\chi_A(x) = f(x) + \det(A)$ where $f(x)$ is a polynomial with constant term zero. Thus $f(x) = xg(x)$ for some polynomial $g(x)$ of degree $n-1$. By the Cayley-Hamilton theorem, $0 = Ag(A) + \det(A)I$ Thus $Ag(A) = -\det(A)I$. Put $B = -\frac{1}{\det(A)}g(A)$. Then $AB = I$. But A and B commute since B is a polynomial in A. Thus $BA = I$. We have therefore proved that $A^{-1} = B$. $\square$

Example 8.5.5. Consider the matrix

$$A = \begin{pmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{pmatrix}$$

You should check that

$$\chi_A(x) = x^3 - 5x^2 + 8x - 4 = (x-1)(x-2)^2.$$

In order to demonstrate the Cayley-Hamilton theorem in this case, I could, of course, calculate

$$A^3 - 5A^2 + 8A - 4I$$

and check that I get the 3×3 zero matrix. But it is possible to simplify this calculation, which is onerous by hand, by using the factorization of the polynomial. We need then only calculate

$$(A - I)(A - 2I)^2.$$

You can readily check that this also yields the zero matrix. It is important to understand why this works. If you multiply out

$$(A - I)(A - 2I)^2$$

using the rules of matrix algebra you will indeed get

$$A^3 - 5A^2 + 8A - 4I.$$

The reason is that A commutes with all powers of itself and any scalar multiples of these powers. The identity matrix and any scalar multiple of the identity matrix commutes with all matrices. Thus the usual rules of algebra apply in this special case. Finally, let's look at how the inverse of A may be expressed as a polynomial in A. We have that

$$A^3 - 5A^2 + 8A - 4I = O.$$

Thus

$$I = \frac{1}{4}\left(A^3 - 5A^2 + 8A\right) = A\frac{1}{4}\left(A^2 - 5A + 8I\right).$$

It follows that

$$A^{-1} = \frac{1}{4}\left(A^2 - 5A + 8I\right).$$

If you calculate this explicitly, you will get

$$A^{-1} = \frac{1}{4}\begin{pmatrix} 2 & -2 & 1 \\ -2 & 2 & 1 \\ 0 & -4 & 1 \end{pmatrix}$$

The characteristic polynomial of a matrix plays a fundamental rôle in more advanced matrix theory. I want to briefly explain why. The nicest matrices are the diagonal matrices. A typical 3×3 diagonal matrix looks like this

$$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}$$

Why are they so nice? Here are a few examples of their nice properties. The determinant of A is just $\lambda_1\lambda_2\lambda_3$. If this is non-zero the inverse of A is just

$$\begin{pmatrix} \frac{1}{\lambda_1} & 0 & 0 \\ 0 & \frac{1}{\lambda_2} & 0 \\ 0 & 0 & \frac{1}{\lambda_3} \end{pmatrix}$$

Finally, calculating powers of A is very easy

$$A^n = \begin{pmatrix} \lambda_1^n & 0 & 0 \\ 0 & \lambda_2^n & 0 \\ 0 & 0 & \lambda_3^n \end{pmatrix}$$

Calculating powers of a matrix is an important calculation in many applications, but involves a lot of work in general. If, however, we could find someway of converting a matrix to a diagonal matrix then perhaps we could lighten the load. To this end, we say that a square matrix A is *diagonalizable* if there is an invertible matrix P and a diagonal matrix D such that

$$PAP^{-1} = D.$$

This means that

$$A = P^{-1}DP.$$

An easy calculation shows that

$$A^2 = P^{-1}D^2P$$

and so by induction we get that

$$A^n = P^{-1} D^n P.$$

Thus to calculate a large power of A, we simply calculate that power of D, which is trivial, and then carry out a couple of matrix multiplications. I should say that not all matrices are diagonalizable but many are. The whys and the wherefores must be left to another time. To get a bit more insight into what is going on, we need a small result.

Lemma 8.5.6. *Let A and B be $n \times n$ matrices and P an invertible matrix such that $B = PAP^{-1}$. Then $\chi_A(x) = \chi_B(x)$.*

Proof. By definition

$$\chi_B(x) = \det(B - xI) = \det(PAP^{-1} - xI)$$

and this is equal to

$$\det\left(P(A - xI)P^{-1}\right) = \det(P)\det(A - xI)\det(P^{-1}) = \chi_A(x).$$

$\square$

The roots of the characteristic polynomial of the matrix A are called the *eigenvalues* of A. If A is our matrix above then its characteristic polynomial is

$$\chi_A = (\lambda_1 - x)(\lambda_2 - x)(\lambda_3 - x)$$

and so its eigenvalues are precisely the numbers λ_1, λ_2 and λ_3.

We many now deduce the following. Suppose that B is an arbitrary 3×3 matrix. Then *if* it is diagonalisable *then* the diagonal matrix we get will have precisely the eigenvalues of A along the diagonal. Thus we may say in general that the eigenvalues of a matrix A determine the algebraic properties of A.

Example 8.5.7. Let

$$A = \begin{pmatrix} -3 & 5 \\ -2 & 4 \end{pmatrix}$$

Its characteristic polynomial is $(x-2)(x+1)$. Let

$$P = \begin{pmatrix} 1 & 5 \\ 1 & 2 \end{pmatrix}$$

You can check that P is invertible and that

$$P^{-1}AP$$

is the diagonal matrix with entries 2 and -1, the eigenvalues of A.

8.6 Determinants redux

Theorem 8.6.1. *Let A be any square matrix. Then*

$$\det(A^T) = \det(A).$$

Proof. The theorem is true in general, but I shall only prove it for 2×2 matrices. Let

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

We calculate $\det(A^T)$

$$\begin{vmatrix} a & c \\ b & d \end{vmatrix} = ad - cb = \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

as claimed. □

8.7 Complex numbers via matrices

Consider all matrices that have the following shape

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

where a and b are arbitrary real numbers. You should show first that the sum, difference and product of any two matrices having this shape is also a matrix of this shape. Rather remarkably matrix multiplication is commutative for matrices of this shape. Observe that the determinant of our matrix above is a^2+b^2. It follows that every non-zero matrix of the above shape is invertible. The inverse of the above matrix in the non-zero case is

$$\frac{1}{a^2+b^2}\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$$

and again has the same form. It follows that the set of all these matrices satisfies the axioms of high-school algebra. Define

$$\mathbf{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and

$$\mathbf{i} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

We may therefore write our matrices in the form

$$a\mathbf{1} + b\mathbf{i}.$$

Observe that

$$\mathbf{i}^2 = -\mathbf{1}.$$

It follows that our set of matrices can be regarded as the complex numbers in disguise.