

Matrix algebra for beginners, Part I
matrices, determinants, inverses

Jeremy Gunawardena

Department of Systems Biology
Harvard Medical School
200 Longwood Avenue, Cambridge, MA 02115, USA
jeremy@hms.harvard.edu

3 January 2006

Contents

1	Introduction	1
2	Systems of linear equations	1
3	Matrices and matrix multiplication	2
4	Matrices and complex numbers	5
5	Can we use matrices to solve linear equations?	6
6	Determinants and the inverse matrix	7
7	Solving systems of linear equations	9
8	Properties of determinants	10
9	Gaussian elimination	11

1 Introduction

This is a Part I of an introduction to the matrix algebra needed for the Harvard Systems Biology 101 graduate course. Molecular systems are inherently many dimensional—there are usually many molecular players in any biological system—and linear algebra is a fundamental tool for thinking about many dimensional systems. It is also widely used in other areas of biology and science.

I will describe the main concepts needed for the course—determinants, matrix inverses, eigenvalues and eigenvectors—and try to explain where the concepts come from, why they are important and how they are used. If you know some of the material already, you may find the treatment here quite slow. There are mostly no proofs but there are worked examples in low dimensions. New concepts appear in italics when they are introduced or defined and there is an index of important items at the end. There are many textbooks on matrix algebra and you should refer to one of these for more details, if you need them.

Thanks to Matt Thomson for spotting various bugs. Any remaining errors are my responsibility. Let me know if you come across any or have any comments.

2 Systems of linear equations

Matrices first arose from trying to solve systems of linear equations. Such problems go back to the very earliest recorded instances of mathematical activity. A Babylonian tablet from around 300 BC states the following problem¹:

There are two fields whose total area is 1800 square yards. One produces grain at the rate of 2/3 of a bushel per square yard while the other produces grain at the rate of 1/2 a bushel per square yard. If the total yield is 1100 bushels, what is the size of each field?

If we let x and y stand for the areas of the two fields in square yards, then the problem amounts to saying that

$$\begin{aligned}x + y &= 1800 \\ 2x/3 + y/2 &= 1100.\end{aligned}\tag{1}$$

This is a system of two *linear equations* in two unknowns. The *linear* refers to the fact that the unknown quantities appear just as x and y , not as $1/x$ or y^3 . Equations with the latter terms are *nonlinear* and their study forms part of a different branch of mathematics, called *algebraic geometry*. Generally speaking, it is much harder to say anything about nonlinear equations. However, linear equations are a different matter: we know a great deal about them. You will, of course, have seen examples like (1) before and will know how to solve them. (So what is the answer?). Let us consider a more general problem (this is the kind of thing mathematicians love to do) in which we do not know exactly what the coefficients are (ie: 1, 2/3, 1/2, 1800, 1100):

$$\begin{aligned}ax + by &= u \\ cx + dy &= v,\end{aligned}\tag{2}$$

and suppose, just to keep things simple, that none of the numbers a , b , c or d are 0. You should be able to solve this too so let us just recall how to do it. If we multiply the first equation by (c/a) , which we can do because $a \neq 0$, and subtract the second, we find that

$$(cb/a)y - dy = cu/a - v.$$

¹For an informative account of the history of matrices and determinants, see http://www-groups.dcs.st-and.ac.uk/history/HistTopics/Matrices_and_determinants.html.

Provided $(ad - bc) \neq 0$, we can divide across and find that

$$y = \frac{av - cu}{ad - bc}. \quad (3)$$

Similarly, we find that

$$x = \frac{ud - bv}{ad - bc}. \quad (4)$$

The quantity $(ad - bc)$, which we did not notice in the Babylonian example above, turns out to be quite important. It is a *determinant*. If it is non-zero, then the system of equations (2) always has a unique solution: the determinant determines whether a solution exists, hence the name. You might check that it is indeed non-zero for example (1). If the determinant is zero, the situation gets more interesting, which is the mathematician's way of saying that it gets a lot more complicated. Depending on u, v , the system may have no solution at all or it may have many solutions. You should be able to find some examples to convince yourself of these assertions.

One of the benefits of looking at a more general problem, like (2) instead of (1), is that you often learn something, like the importance of determinants, that was hard to see in the more concrete problem. Let us take this a step further (generalisation being an obsessive trait among mathematicians). How would you solve a system of 3 equations with 3 unknowns,

$$\begin{aligned} ax + by + cz &= u \\ dx + ey + fz &= v \\ gx + hy + iz &= w, \end{aligned} \quad (5)$$

or, more generally still, a system of n equations with n unknowns?

You think this is going too far? In 1800, when Carl Friedrich Gauss was trying to calculate the orbit of the asteroid Pallas, he came up against a system of 6 linear equations in 6 unknowns. (Astronomy, like biology, also has lots of moving parts.) Gauss was a great mathematician—perhaps the greatest—and one of his very minor accomplishments was to work out a systematic version of the technique we used above for solving linear systems of arbitrary size. It is called *Gaussian elimination* in his honour. However, it was later discovered that the “*Nine Chapters of the Mathematical Art*”, a handbook of practical mathematics (surveying, rates of exchange, fair distribution of goods, etc) written in China around the 3rd century BC, uses the same method on simple examples. Gauss made the method into what we would now call an *algorithm*: a systematic procedure that can be applied to any system of equations. We will learn more about Gaussian elimination in §9 below.

The modern way to solve a system of linear equations is to transform the problem from one about numbers and ordinary algebra into one about matrices and matrix algebra. This turns out to be a very powerful idea but we will first need to know some basic facts about matrices before we can understand how they help to solve linear equations.

3 Matrices and matrix multiplication

A *matrix* is any rectangular array of numbers. If the array has n rows and m columns, then it is an $n \times m$ matrix. The numbers n and m are called the *dimensions* of the matrix. We will usually denote matrices with capital letters, like A , B , etc, although we will sometimes use lower case letters for one dimensional matrices (ie: $1 \times m$ or $n \times 1$ matrices). One dimensional matrices are often called *vectors*, as in *row vector* for a $n \times 1$ matrix or *column vector* for a $1 \times m$ matrix but we are going to use the word “vector” to refer to something different in Part II. We will use the notation A_{ij} to refer to the number in the i -th row and j -th column. For instance, we can extract the numerical coefficients from the system of linear equations in (5) and represent them in the matrix

$$A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}. \quad (6)$$

It is conventional to use brackets (either round or square) to delineate matrices when you write them down as rectangular arrays. With our notation, $A_{23} = f$ and $A_{32} = h$.

The first known use of the matrix idea appears in the “*The Nine Chapters of the Mathematical Art*”, the 3rd century BC Chinese text mentioned above. The word *matrix* itself was coined by the British mathematician James Joseph Sylvester in 1850. Matrices first arose from specific problems like (1). It took nearly two thousand years before mathematicians realised that they could gain an enormous amount by *abstracting* away from specific examples and treating matrices as objects in their own right, just as we will do here. The first fully abstract definition of a matrix was given by Sylvester’s friend and collaborator, Arthur Cayley, in his 1858 book, “*A memoir on the theory of matrices*”. Abstraction was a radical step at the time but became one of the key guiding principles of 20th century mathematics. Sylvester, by the way, spent a lot of time in America. In his 60s, he became Professor of Mathematics at Johns Hopkins University and founded America’s first mathematics journal, *The American Journal of Mathematics*.

There are a number of useful operations on matrices. Some of them are pretty obvious. For instance, you can add any two $n \times m$ matrices by simply adding the corresponding entries. We will use $A + B$ to denote the sum of matrices formed in this way:

$$(A + B)_{ij} = A_{ij} + B_{ij} .$$

Addition of matrices obeys all the formulae that you are familiar with for addition of numbers. A list of these are given in Figure 2. You can also multiply a matrix by a number by simply multiplying each entry of the matrix by the number. If λ is a number and A is an $n \times m$ matrix, then we denote the result of such multiplication by λA , where

$$(\lambda A)_{ij} = \lambda A_{ij} . \tag{7}$$

Multiplication by a number also satisfies the usual properties of number multiplication and a list of these can also be found in Figure 2. All of this should be fairly obvious and easy. What about products of matrices? You might think, at first sight, that the “obvious” product is to just multiply the corresponding entries. You can indeed define a product like this—it is called the *Hadamard product*—but this turns out not to be very productive mathematically. The *matrix matrix product* is a much stranger beast, at first sight.

If you have an $n \times k$ matrix, A , and a $k \times m$ matrix, B , then you can *matrix multiply* them together to form an $n \times m$ matrix denoted AB . (We sometimes use $A.B$ for the matrix product if that helps to make formulae clearer.) The matrix product is one of the most fundamental matrix operations and it is important to understand how it works in detail. It may seem unnatural at first sight and we will learn where it comes from later but, for the moment, it is best to treat it as something new to learn and just get used to it. The first thing to remember is how the matrix dimensions work.

You can only multiply two matrices together if the number of columns of the first equals the number of rows of the second.

Note two consequences of this. Just because you can form the matrix product AB does not mean that you can form the product BA . Indeed, you should be able to see that the products AB and BA only both make sense when A and B are *square matrices*: they have the same number of rows as columns. (This is an early warning that reversing the order of multiplication can make a difference; see (9) below.) You can always multiply any two square matrices of the same dimension, in any order. We will mostly be working with square matrices but, as we will see in a moment, it can be helpful to use non-square matrices even when working with square ones.

To explain how matrix multiplication works, we are going to first do it in the special case when $n = m = 1$. In this case we have a $1 \times k$ matrix, A , multiplied by a $k \times 1$ matrix, B . According to

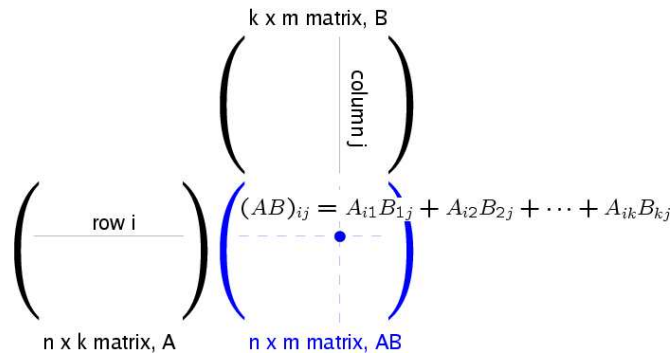


Figure 1: Matrix multiplication. A , on the left, is a $n \times k$ matrix. B , on the top, is a $k \times m$ matrix. Their product, AB , is a $n \times m$ matrix shown on the bottom right (in blue). The ij -th element of AB is given by matrix multiplying the i -th row of A by the j -th column of B according to the formula for multiplying 1 dimensional matrices in (8).

the rule for dimensions, the result should be a 1×1 matrix. This has just has one entry. What is the entry? You get it by multiplying corresponding terms together and adding the results:

$$(A_{11} \ A_{12} \ \dots \ A_{1k}) \begin{pmatrix} B_{11} \\ B_{21} \\ \vdots \\ B_{k1} \end{pmatrix} = (A_{11}B_{11} + A_{12}B_{21} + \dots + A_{1k}B_{k1}) \quad (8)$$

Once you know how to multiply one dimensional matrices, it is easy to multiply any two matrices. If A is an $n \times k$ matrix and B is a $k \times m$ matrix, then the ij -th element of AB is given by taking the i row of A , which is a $1 \times k$ matrix, and the j -th column of B , which is a $k \times 1$ matrix, and multiplying them together just as in (8). Schematically, this looks as shown in Figure 1. It can be helpful to arrange the matrices in this way if you are multiplying matrices by hand. You can try this out on the following example of a 2×3 matrix multiplied by a 3×2 matrix to give a 2×2 matrix.

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \begin{pmatrix} -2 & 2 \\ 1 & 0 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 6 & 5 \\ 1 & 5 \end{pmatrix}$$

There is one very important property of matrix multiplication that it is best to see early on. Consider the calculation below, in which two square matrices are multiplied in a different order

$$\begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} 3 & -1 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 5 & 5 \\ 5 & 5 \end{pmatrix} \quad \begin{pmatrix} 3 & -1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 7 \\ 7 & -1 \end{pmatrix}.$$

We see from this that

$$\boxed{\text{matrix multiplication is not commutative.}} \quad (9)$$

This is one of the major differences between matrix multiplication and number multiplication. You can get yourself into all kinds of trouble if you forget it.

You might well ask where such an apparently bizarre product came from. What is the motivation behind it? That is a good question. A partial answer comes from the following observation. (We will give a more complete answer later.) Once we have matrix multiplication, we can use it to rewrite the system of equations (5) as an equation about matrices. You should be able to check that (5) is

identical to the matrix equation

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} u \\ v \\ w \end{pmatrix} \quad (10)$$

in which a 3×3 matrix is multiplied by a 3×1 matrix to give another 3×1 matrix. You should also be able to see that if we had a system of n linear equations in n unknowns then we could do the same thing and write it as a matrix equation

$$Ax = u \quad (11)$$

where A is the $n \times n$ matrix of coefficients and we have used x to denote the $n \times 1$ matrix of unknown quantities (ie: the x , y and z in (10)) and u to denote the $n \times 1$ matrix of constant quantities (ie: the u , v and w in (10)). (We are clearly going to have to use a different notation for the unknown quantities and the constants when we look at n dimensional matrices but we will get to that later.) Now that we can express systems of linear equations in matrix notation, we can start to think about how we might solve the single matrix equation (11) rather than a system of n equations.

4 Matrices and complex numbers

Before dealing with solutions of linear equations, let us take a look at some matrix products that will be useful to us later. You should be familiar with complex numbers, of the form $a + ib$, where a and b are ordinary numbers and i is the so-called “imaginary” square root of -1 . Complex numbers are important, among other things, because they allow us to solve equations which we cannot solve in the ordinary numbers. The simplest such equation is $x^2 + 1 = 0$, which has no solutions in ordinary numbers but has $\pm i$ as its two solutions in the complex numbers. What is rather more amazing is that if you take any polynomial equation, such as

$$a_m x^m + a_{m-1} x^{m-1} + \cdots + a_1 x + a_0 = 0,$$

even one whose coefficients, a_0, \cdots, a_m , are complex numbers, then this polynomial has a solution over the complex numbers. This is the *fundamental theorem of algebra*. The first more-or-less correct proof of it was given by Gauss in his doctoral dissertation.

What have complex numbers got to do with matrices? We can associate to any complex number the matrix

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}. \quad (12)$$

Notice that, given any matrix of this form, we can always construct the corresponding complex number, $a + ib$, and these two processes, which convert back and forth between complex numbers and matrices, are inverse to each other. It follows that matrices of the form in (12) and complex numbers are in *one-to-one correspondence* with each other. What does this correspondence do to the algebraic operations on both sides? Well, if you take two complex numbers, $a + ib$ and $c + id$, then the matrix of their sum $(a + c) + i(b + d)$, is the sum of their matrices. What about multiplication? Using the rule for matrix multiplication, we find that

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} \cdot \begin{pmatrix} c & -d \\ d & c \end{pmatrix} = \begin{pmatrix} ac - bd & -(ad + bc) \\ ad + bc & ac - bd \end{pmatrix}.$$

This is exactly the matrix of the product of the complex numbers:

$$(a + ib)(c + id) = (ac - bd) + i(ad + bc).$$

We see from this that the subset of matrices having the form in (12) are the complex numbers in disguise, as it were. This may confirm your suspicion that the matrix product is infernally

complicated! A one-to-one correspondence between two algebraic structures, which preserves the operations on both sides, is called an *isomorphism*. We can as well work with either structure. As we will see later on, this can be very useful. Some of the deepest results in mathematics reveal the existence of unexpected isomorphisms.

5 Can we use matrices to solve linear equations?

After that diversion, let us get back to linear equations. If you saw the equation $ax = u$, where a , x and u were all numbers, you would know immediately how to solve for x . Provided $a \neq 0$, you would multiply both sides of the equation by $1/a$, to get $x = u/a$.

Perhaps we can do the same thing with matrices? (Analogy is often a powerful guide in mathematics.) It turns out that we can and that is what we are going to work through in the next few sections. Before starting, however, it helps to pause for a bit and look at what we have just done with numbers more closely. It may seem that this is a waste of time because it is so easy with numbers but we often take the familiar for granted and by looking closely we will learn some things that will help us understand what we need when we move from numbers to matrices.

The quantity $1/a$ is the *multiplicative inverse* of a . That is a long winded way of saying that its the (unique) number with the property that when multiplied by a , it gives 1. (Why is it unique?) Another way to write the multiplicative inverse is a^{-1} ; it always exists, provided only that $a \neq 0$. (As you know, dividing by zero is frowned upon. Why? If 0^{-1} existed as a number, then we could take the equation $0 \times 2 = 0 \times 3$, which is clearly true since both sides equal 0, and multiply through by 0^{-1} , giving $2 = 3$. Since this is rather embarrassing, it had better be the case that 0 has no multiplicative inverse.) Let us go through the solution of the simple equation $ax = u$ again using the notation a^{-1} and this time I will be careful with the order of multiplication. If we multiply the left hand side by a^{-1} we get

$$a^{-1}(ax) = (a^{-1}a)x = 1x = x,$$

while if we multiply the right hand side we get just

$$a^{-1}u.$$

Hence, $x = a^{-1}u$. Notice that there are couple of things going on behind the scenes. We needed to use the property that multiplication of numbers is associative (in other words, $a(bc) = (ab)c$) and that multiplying any number by 1 leaves it unchanged.

If we want to use the same kind of reasoning for the matrix equation (11), we shall have to answer three questions.

1. **What is the matrix equivalent of the number 1?**
2. **Is matrix multiplication associative?**
3. **Does a matrix have a multiplicative inverse?**

If we can find a satisfactory answer to all of these, then we can solve the matrix equation (11) just as we did the number equation $ax = u$.

From this point, we are mostly going to deal with square matrices. Remember that we can always multiply two square matrices of the same dimension.

The matrix equivalent of the number 1 is the called the *identity matrix*. The $n \times n$ identity matrix is denoted I_n but we often simplify this to just I when the dimension is clear from the context. All

the diagonal entries of I_n are 1 and all the off-diagonal entries are 0, so that I_4 looks like

$$I_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

You should be able to convince yourself very quickly, using the rule for matrix multiplication in Figure 1, that if A is any $n \times n$ matrix, then

$$A.I_n = I_n.A = A.$$

so that I_n really does behave like the number 1 for square matrices of dimension n .

Here is a little problem to test your understanding of matrix multiplication. How would you interchange two columns (or two rows) of a matrix just by matrix multiplication? More specifically, let $A(i \leftrightarrow j)$ denote the matrix A with the columns i and j interchanged. Show that $A(i \leftrightarrow j) = A.I(i \leftrightarrow j)$. What happens if you want to interchange rows?

The second question asks, given three $n \times n$ matrices A , B and C , whether $A.(B.C) = (A.B).C$? Well, you can just write down each of the products in terms of the individual entries A_{ij} , B_{ij} and C_{ij} using the formula in Figure 1 and work it out. The answer turns out to be “yes”, although the calculation is quite tedious. In fact, it is just as tedious to show that associativity holds for any three matrices which can be multiplied together. If A is an $n \times k$ matrix, B a $k \times j$ matrix and C is a $j \times m$ matrix, then

$$A.(B.C) = (A.B).C \tag{13}$$

Although you can show that the matrix product is associative this way, it is not a particularly edifying proof. Mathematicians are always a bit suspicious of results that follow from long, boring calculations because they suspect they are missing something. A result as simple as (13) ought to be true for simple reasons, not complex ones. This kind of aesthetic uneasiness has often led to new mathematics. In this case, there is a simple reason why matrix multiplication is associative and it is the same reason that explains why the matrix product itself is “natural”. However, we have to change our perspective quite a bit, from algebra to geometry, to see things in this light. We will get to it in Part II.

Matrix multiplication obeys several rules that are similar to those that hold for number multiplication (with the exception, of course, of commutativity). These are listed in Figure 2.

The third question asks about the multiplicative inverse of a matrix. (From now on, we will just say “inverse”.) The inverse of the matrix A is a matrix, which we will call, by analogy with numbers, A^{-1} , such that $A^{-1}.A = A.A^{-1} = I$. This only makes sense for square matrices (why?). Notice that because matrix multiplication is not commutative, it is important to make sure that both $A^{-1}.A$ and $A.A^{-1}$ give the same result. If the inverse exists, it must be unique (why?).

It turns out that the inverse matrix is closely related to the determinant. This should not really be a surprise, given what we found out above for example (2). We are going to define the inverse matrix and the determinant of a matrix through the same basic formula (16) below.

6 Determinants and the inverse matrix

You have already worked out the determinant of a 2×2 matrix. If we extract the 2×2 coefficient matrix from example (2)

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{14}$$

then its determinant is the quantity we came across before, $ad - bc$. What is the determinant of an arbitrary $n \times n$ matrix? If A is a square matrix, we will denote its determinant $\det A$. (Strictly speaking, the determinant is a function from matrices to numbers, so we should put brackets around the argument, like $\det(A)$. However, this can make formulae look awfully messy at times and I try not to use too many brackets, unless, for some reason, it is important to clarify which matrix the determinant applies to.)

The easiest way to explain the determinant is to do it by induction (or what computer scientists would call *recursion*). That is, I will show how you the determinant of an $n \times n$ matrix can be calculated if you know how to calculate the determinant of an $(n - 1) \times (n - 1)$ matrix. We need to start this off somewhere, so here is the determinant of a 1×1 matrix:

$$\det(a) = a . \tag{15}$$

Notice that 1×1 matrices are equivalent to numbers but there is a notational distinction between the matrix (a) and the number a .

Now suppose you have an $n \times n$ matrix A . We are going to form an associated $n \times n$ matrix, called the *adjugate* of A , denoted $\text{adj } A$. (Once again, we should probably write $\text{adj}(A)$ but, as above, we try and avoid excess brackets when we can.) To define the adjugate we need a notation that allows us to select a submatrix from A . The $(n - 1) \times (n - 1)$ submatrix $A_{(ij)}$ is obtained from A by omitting the i -th row and the j -th column of A . You have to be careful about the small difference between $A_{(ij)}$ and A_{ij} : the former is a matrix which is one size smaller than A ; the latter is a number. We can now define the adjugate. The ij -th entry of the adjugate matrix, $\text{adj } A$, is the product of the determinant of the submatrix $A_{(ji)}$ (NOTE THE REVERSAL OF ROWS AND COLUMNS) and the sign factor $(-1)^{i+j}$. In other words,

$$(\text{adj } A)_{ij} = (-1)^{i+j} \det A_{(ji)}$$

It is easier to see what is going on here by working through an example. Consider the 2×2 matrix in (14). The 11 element of $\text{adj } A$ is given by omitting the first row and first column, which gives the matrix (d) , taking its determinant, which is d , and then multiplying by $(-1)^2 = 1$. The 12 element is given by omitting the second row and first column, which gives the matrix (b) , taking its determinant, which is b , and multiplying by $(-1)^3 = -1$. Proceeding in this way, we find that the adjugate of (14) is

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

What is the big deal? I still have not explained why the adjugate helps you to work out $\det A$. Well, you can check for yourself with this example that

$$A \cdot (\text{adj } A) = (\text{adj } A) \cdot A = \begin{pmatrix} ad - bc & 0 \\ 0 & ad - bc \end{pmatrix} .$$

The determinant of the 2×2 matrix has reappeared, even though we only used determinants of 1×1 matrices in the calculation. This turns out to be a general result. If A is any $n \times n$ matrix then

$$\boxed{A \cdot (\text{adj } A) = (\text{adj } A) \cdot A = (\det A) I_n} \tag{16}$$

Notice that the first two expressions are matrix products while the last expression is multiplication of the matrix I_n by the number $\det A$, as in (7). Formula (16) is very useful and tells us many things. For a start, it explains how to calculate determinants. If you look at individual entries of $(\det A) I_n$, you can recover a set of formulae that were first worked out by the French mathematician, Pierre Simon (Marquis de) Laplace, in the 18th century. They are still known as Laplace expansions in his honour. We will not write these all down but this is what you get if you look at the 11 entry of $(\det A) I_n$ and expand out the formula $A \cdot (\text{adj } A) = (\det A) I_n$ from (16):

$$\det A = A_{11} \det A_{(11)} - A_{12} \det A_{(12)} + A_{13} \det A_{(13)} - \dots + (-1)^{1+n} A_{1n} \det A_{(1n)} . \tag{17}$$

You can try writing down some of the other Laplace expansions yourself. Notice that if you pick one of the off diagonal entries of $(\det A)I_n$ you get a relation among the determinants of the submatrices $A_{(ij)}$.

Formula (17) provides a recursive algorithm for computing $\det(A)$. You can use it to work out the determinant of the 3×3 matrix (6). You should find it to be

$$a(ei - fh) - b(di - fg) + c(dh - eg). \quad (18)$$

It can be fairly unhealthy calculating determinants of large matrices this way. It can sometimes be done but it can also get extremely tedious. The better method is to use some form of Gaussian elimination, which we will learn how to do in §9. The determinant has many wonderful properties, which are collected together in §8.

Formula (16) also tells us the inverse of A . If $\det A \neq 0$, then it should be clear from the definition of an inverse that

$$\boxed{A^{-1} = \left(\frac{1}{\det(A)} \right) \text{adj } A} \quad (19)$$

If $\det(A) = 0$ then A does not have a multiplicative inverse. That is another important difference between matrices and numbers. The only number which does not have a multiplicative inverse is 0 but there are lots of non-zero matrices that do not have a multiplicative inverse (try and write down a couple).

So now you know how to work out the matrix inverse. The same caveat as for determinants holds for the inverse: (19) can be helpful to understand the inverse and its properties but Gaussian elimination is a better way of computing the inverse matrix if you have to invert a large matrix by hand.

7 Solving systems of linear equations

We have now got all the ingredients we need to solve a system of n linear equations in n unknowns. Suppose that you have expressed this system, as we did in (10), in the form of a single matrix equation, $Ax = u$, where the unknown quantities are x_1, x_2, \dots, x_n and the constant quantities are u_1, u_2, \dots, u_n . In this case the $n \times 1$ matrices x and u are given by

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}.$$

The first thing to do is to work out $\det A$. You could use the Laplace expansion in (17), Gaussian elimination as in §9 or even get MATLAB to calculate it for you. If $\det A = 0$ then the system may have no solution. (There is a lot more that could be said about this case but it takes more work and is not particularly useful for us.) If $\det A \neq 0$ then, you can form the inverse matrix given in (19) and follow the strategy that works for numbers by multiplying through by the inverse matrix and using associativity to simplify the algebra. You should find that

$$x = \left(\frac{1}{\det A} \right) (\text{adj } A)u. \quad (20)$$

The only problem with this is the adjugate, which is quite complicated to calculate. There is a clever way of avoiding this which was worked out by the Swiss mathematician Gabriel Cramer in the 18th

century and is known as Cramer's rule to this day. You can get a hint of the idea by going back to our bare hands solution of the 2×2 system in (2) and noticing that, although the denominator of the solutions (3) and (4) is the determinant, just as we would expect from (20), the numerator also looks like a determinant. The question is, of what matrix is it the determinant? Cramer found a neat answer to this. Suppose you want to know the solution for the i -th unknown, x_i . Replace the i -th column of the matrix A by the $n \times 1$ matrix u . Let us call this matrix $A \overset{i}{\leftarrow} u$. Cramer noticed that (in our language—he knew nothing of the adjugate) $(\text{adj } A)u = \det(A \overset{i}{\leftarrow} u)$. This follows from a Laplace expansion. (Can you see why?) Cramer's rule then states that:

$$x_i = \frac{\det(A \overset{i}{\leftarrow} u)}{\det A}.$$

Let us try this out on example (2) using the old notation. If we replace the first column of A by u and the second column of A by v we get, respectively, the two 2×2 matrices

$$\begin{pmatrix} u & b \\ v & d \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} a & u \\ c & v \end{pmatrix}.$$

Their determinants are, respectively, $ud - bv$ and $av - cu$. According to Cramer's rule, the solutions of (2) should be

$$x = \frac{ud - bv}{ad - bc} \quad y = \frac{av - cu}{ad - bc},$$

which is exactly what we found for (2). You now know about inverse matrices and how to solve systems of linear equations, at least when $\det A \neq 0$.

8 Properties of determinants

The most important property of the determinant is that it is a multiplicative function on matrices. In other words,

$$\boxed{\det AB = (\det A)(\det B)} \tag{21}$$

This is fairly amazing and it is something you need to remember. You should at least check the 2×2 case for yourself to satisfy yourself that it works. The proof in higher dimensions is one of the hardest in elementary matrix algebra. We will get a better idea of why (21) is true when we look at matrices from a geometric perspective in Part II.

Just because (21) is true, do not make the mistake of thinking that the determinant is also additive! However, it does obey several simple rules that are helpful in calculating it.

Rule (1) If matrix B is obtained from A by multiplying each entry in one column by λ (which is not the same thing as λA) then $\det B = \lambda(\det A)$. For example,

$$\det \begin{pmatrix} \lambda a & b \\ \lambda c & d \end{pmatrix} = \lambda ad - b\lambda c = \lambda(ad - bc).$$

You should see from this that if A is an $n \times n$ matrix then $\det(\lambda A) = \lambda^n \det A$. Also, if one column of A consists entirely of zeros, then this rule tells us that $\det A = 0$. (Why?)

Rule (2) Although the determinant is not additive, it obeys a restricted form of additivity on individual rows or columns. If u and v are any two $n \times 1$ matrices, then, recalling the notation we used in §7, $\det(A \overset{i}{\leftarrow} (u + v)) = \det(A \overset{i}{\leftarrow} u) + \det(A \overset{i}{\leftarrow} v)$. For example,

$$\det \begin{pmatrix} a & u_1 + v_1 \\ c & u_2 + v_2 \end{pmatrix} = a(u_2 + v_2) - (u_1 + v_1)c = (au_2 - u_1c) + (av_2 - v_1c).$$

Rule (3) Exchanging any two columns changes the sign of the determinant. For example,

$$\det \begin{pmatrix} b & a \\ d & c \end{pmatrix} = -(ad - bc).$$

Notice that if two columns are identical, then this rule implies that the determinant must be 0.

Rule (4) The determinant of the identity matrix is 1: $\det I_n = 1$.

Identical results hold for rows in place of columns. These rules are not too hard to prove. The first and second follow from the appropriate Laplace expansions. The third can be easily understood if you assume the multiplicative property of the determinant. If you worked out the problem in §5 about interchanging two columns of a matrix by matrix multiplication then you know that $A(i \leftrightarrow j) = A.I(i \leftrightarrow j)$. It follows from (21) that $\det A(i \leftrightarrow j) = (\det A)(\det I(i \leftrightarrow j))$. It is easy to show, using Laplace expansions, that $\det I(i \leftrightarrow j) = -1$, which proves the third rule. The fourth rule is a special case of a more general result that is also quite helpful. A matrix is said to be upper (respectively, lower) triangular if all the entries below (respectively, above) the diagonal are 0. The identity matrix is both upper and lower triangular.

Rule (5) The determinant of an upper or lower triangular matrix is the product of the diagonal entries. So, for instance,

$$\det \begin{pmatrix} a & 0 & 0 \\ b & d & 0 \\ c & e & f \end{pmatrix} = adf$$

This rule also follows quite easily by using a Laplace expansion. Rules (1)-(4) characterise the determinant. In other words, the determinant is the unique function from matrices to numbers that satisfies Rules (1)-(4) above.

You can test your understanding of these rules by showing that if you add any multiple of the entries in one column to any of the other columns of a matrix, then the determinant of the matrix does not change (similarly for rows). This can be useful for calculating determinants and we will use it in §9 when we study Gaussian elimination. So now you know something about determinants. The simple formula that we stumbled across when solving (2) gets a lot more complicated in higher dimensions!

9 Gaussian elimination

Although the adjugate formula (16) is a convenient way to define the determinant and the inverse matrix and to understand their properties, it is not a good way to calculate either as soon as the matrix gets large (which means more than three dimensional). Gaussian elimination is the preferred technique and this is such a basic tool in matrix calculations that you need to know something about it. Recall that although Gauss was the first to make it into a systematic procedure, the idea was known to the unknown Chinese mathematicians of the 3rd century BC who compiled “*The Nine Chapters of the Mathematical Art*”. It is essentially the same technique that we used to solve the system of linear equations in (2).

We will perform Gaussian elimination on a matrix A in two stages. In the first stage we will reduce A to upper triangular form. This will allow us to calculate its determinant. If the determinant is non-zero then we will continue the procedure to reduce A to diagonal form. That is, to a matrix whose only non-zero entries are on the main diagonal. This will allow us to work out the inverse matrix.

Gaussian elimination consists of repeatedly applying three simple operations to A , all of which are derived from the determinantal rules that we learnt in §8. We will apply these rules to rows, although they could equally well be applied to columns. Suppose that the first $k - 1$ columns of A have been reduced to upper triangular form.

Step 1 Take the next column. If $A_{kk} = 0$ look at all the entries in column k lying below A_{kk} . If all of them are zero then column k is already in upper triangular form and you should move to the next column. If there is some non-zero entry on row l , where $l > k$, then interchange rows k and l . (We will keep calling the matrix that results from changes like this A , although, of course, it is no longer the matrix A we started with.) A now has $A_{kk} \neq 0$, so we can proceed to the next step.

Step 2 Use A_{kk} to triangularise column k . Suppose there is some non-zero entry in column k lying below A_{kk} . That is, for some $l > k$, $A_{lk} \neq 0$. We are going to change row l by subtracting from it row k multiplied through by A_{lk}/A_{kk} . (Multiplying through by a number means that each entry on row k is multiplied by the number.) The effect is to make $A_{lk} = 0$. (If you think about it for a moment, you will see that this is exactly the trick that we used to solve the linear equations in (2).) Proceed in this way until all the entries lying below A_{kk} are zero. Column k is now in upper triangular form and you can proceed to the next column.

At the end of the procedure, A will be in upper triangular form. Why is that useful? Well, the two operations that we have repeatedly used are interchanging two rows and adding a multiple of one row to another. Both of these operations are equivalent to pre-multiplying A by a suitable matrix. We worked out the matrix needed for the interchange operation in §8. At least, we worked it out for columns, while here we are doing it for rows. The principle is identical but instead of post-multiplying A , we pre-multiply A . Note that, as above, the determinant of the pre-multiplication matrix for interchanging two rows is always -1 . As for the other operation, if you did the exercise at the end of the previous section then you already know that this operation does not change the determinant but let us look at it from the viewpoint of matrix multiplication. Suppose you want to add λ times row 1 to row 3 of a 4×4 matrix A . You should be able to see quickly that this is equivalent to pre-multiplying A by

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \lambda & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (22)$$

This matrix is lower triangular so Rule (5) in §8 tells us that its determinant is 1. It follows that what we have done to A by Gaussian elimination is to form the matrix product $W_1 W_2 \cdots W_p A$ where the matrices W_i are either interchange matrices, whose determinant is -1 , or matrices like (22), whose determinant is $+1$. But now remember formula (21) which tell us that the determinant is a multiplicative function. It follows that the determinant of the matrix we started with is the determinant of the matrix we ended with, multiplied by $(-1)^u$, where u is the number of times we interchanged rows. The matrix we ended with is upper triangular and we know from rule (5) in §8 that the determinant of an upper-triangular matrix is the product of its diagonal entries. Hence, you can readily work out the determinant of A —its the determinant of the upper triangular matrix that you ended up with, up to a sign, and the sign can be worked out by keeping track of the number of times you interchanged rows in bringing A to upper triangular form.

Let us try this out on the example below. The notation should be self-explanatory. Since $A_{11} = 1$, we do not need to interchange any rows. We first use A_{11} to make $A_{21} = 0$

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 2 \\ 3 & 5 & 5 \end{pmatrix} \xrightarrow{R2 \Rightarrow R2 - 2 * R1} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -1 & -4 \\ 3 & 5 & 5 \end{pmatrix},$$

and then use it to make $A_{31} = 0$,

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & -1 & -4 \\ 3 & 5 & 5 \end{pmatrix} \xrightarrow{R3 \Rightarrow R3 - 3R1} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -1 & -4 \\ 0 & -1 & -4 \end{pmatrix}.$$

At this point column 1 is in upper triangular form so we turn to column 2. $A_{22} = -1$, so, once again, we do not need to do any row interchanges. Since $A_{32} \neq 0$, we use A_{22} to make it so,

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & -1 & -4 \\ 0 & -1 & -4 \end{pmatrix} \xrightarrow{R3 \Rightarrow R3 - R2} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -1 & -4 \\ 0 & 0 & 0 \end{pmatrix},$$

which leaves the matrix in upper triangular form. Since $A_{33} = 0$, we conclude that our original matrix had a determinant of 0. You might like to check this using formula (18).

If $\det A \neq 0$, unlike this example, then we know that A^{-1} exists. We can work out A^{-1} by taking the procedure above one stage further. Assume that we have brought our matrix to upper triangular form and that all the diagonal entries are non-zero.

Step 3 Since each A_{kk} is non-zero, we can use each in turn to set to zero all the entries lying above the diagonal. The matrix is now in diagonal form.

Step 4 Multiply each row k by $1/A_{kk}$. This brings the matrix to the identity.

What good is this? We seem to have lost the matrix we started with and be no wiser as to its inverse. To see how to recover the inverse, note that we have used an additional operation in Step 4 but this operation also amounts to pre-multiplying A by a suitable matrix. You should be able to work out that this is a simple diagonal matrix. (Note that its determinant is no longer ± 1 but we no longer need that at this point.) The net result is that we have constructed the identity matrix through the series of matrix multiplications, $W_1 W_2 \cdots W_q A$, where the W_i are either interchange matrices, lower triangular matrices of the form (22) or diagonal matrices needed for row multiplications. In other words,

$$(W_1 W_2 \cdots W_q) A = I,$$

from which it follows that $W_1 W_2 \cdots W_q = A^{-1}$. We were building up the inverse matrix all along. But how do we keep track of the W_i matrices? There is a very neat trick for doing this, which allows you to get away with not worrying about the W_i at all. Instead of doing the row operations on just A we do them on the $n \times 2n$ matrix which consists of A in the first n columns and the identity matrix in the last n columns. Let us denote this matrix, in a non-standard way, by $A | I$, to show the two parts of it. If you now go through all the row operations in Steps 1 to 4 needed to reduce A to the identity matrix but apply them to $A | I$ instead of just to A , the effect is equivalent to $W_1 W_2 \cdots W_q A | W_1 W_2 \cdots W_q I$ (this should be clear from the rule for matrix multiplication). But, because I is the identity matrix, $W_1 W_2 \cdots W_q I = W_1 W_2 \cdots W_q$, so what we end up with in the last n columns of $A | I$ is exactly the inverse matrix that we want.

$$A + B = B + A$$

$$A + (B + C) = (A + B) + C$$

$$A + 0 = 0 + A = A$$

$$A + -A = -A + A = 0$$

$$\lambda(\mu A) = (\lambda\mu)A$$

$$\lambda(A + B) = \lambda A + \lambda B$$

$$1A = A \quad (-1)A = -A$$

Matrix multiplication is not commutative

$$A.(B.C) = (A.B).C$$

$$A.I = I.A = A$$

$$A^{-1}.A = A.A^{-1} = I \quad \text{provided } \det A \neq 0$$

$$A.(B + C) = A.B + A.C$$

$$(B + C).A = B.A + C.A$$

$$(\lambda A).B = A.(\lambda B) = \lambda(A.B)$$

Figure 2: Properties of the three matrix operations, $A+B$, λA and $A.B$. A, B, C are matrices, which must be of the same dimensions when summed and of the appropriate dimensions when multiplied, and λ, μ are numbers. $-A$ denotes the additive inverse in which all the entries of A are replaced by their negatives. A^{-1} denotes the multiplicative inverse, as defined in (19), which only exists when $\det A \neq 0$. I denotes the identity matrix, which must always be square and of the appropriate dimension for any matrix multiplications to be valid.

Index

- abstracting, 3
- adjugate, 8, 10
- algebraic geometry, 1
- algorithm, 2
- associativity, 7

- complex numbers, 5
- Cramer's rule, 10

- determinant, 2, 8, 10, 12
 - 2×2 , 7
 - 3×3 , 9
 - multiplicative function, 10
 - simple rules, 10
 - triangular matrix, 11

- fundamental theorem of algebra, 5

- Gaussian elimination, 2, 9, 11

- Hadamard product, 3

- inverse, 6
- isomorphism, 6

- Laplace expansions, 8
- linear equations, 1, 9

- matrix, 2, 3
 - dimensions, 2
 - identity, 6
 - inverse, 7, 9, 13
 - product, 3, 4
 - square, 3
 - triangular, 11, 12

- nonlinear, 1

- one-to-one correspondence, 5

- recursion, 8

- vector
 - column vector, 2
 - row vector, 2