MTH6140 **Linear Algebra II**

Notes 4 **1st November 2010**

---

4 Determinants

The determinant is a function defined on square matrices; its value is a scalar. It has some very important properties: perhaps most important is the fact that a matrix is invertible if and only if its determinant is not equal to zero.

We denote the determinant function by det, so that $\det(A)$ is the determinant of A. For a matrix written out as an array, the determinant is denoted by replacing the square brackets by vertical bars:

$$\det\begin{bmatrix}1 & 2\\ 3 & 4\end{bmatrix} = \begin{vmatrix}1 & 2\\ 3 & 4\end{vmatrix}.$$

4.1 Definition of determinant

You have met determinants in earlier courses, and you know the formula for the determinant of a 2×2 or 3×3 matrix:

$$\begin{vmatrix}a & b\\ c & d\end{vmatrix} = ad-bc, \qquad \begin{vmatrix}a & b & c\\ d & e & f\\ g & h & i\end{vmatrix} = aei+bfg+cdh-afh-bdi-ceg.$$

Our first job is to define the determinant for square matrices of any size. We do this in an "axiomatic" manner:

Definition 4.1 A function D defined on $n\times n$ matrices is a *determinant* if it satisfies the following three conditions:

(D1) For each $1\le i\le n$, the function D is a linear function of the ith column: this means that, if A and A' are two matrices which agree everywhere except the ith column, and if A'' is the matrix whose ith column is c times the ith column of A plus c' times the ith column of A', but agreeing with A and A' everywhere else, then

$$D(A'') = c\,D(A)+c'\,D(A').$$

(D2) If A has two equal columns, then $D(A) = 0$.

(D3) $D(I_n) = 1$, where I_n is the $n \times n$ identity matrix.

We show the following result:

Theorem 4.1 There is a unique determinant function on $n \times n$ matrices, for any n.

Proof First, we show that applying elementary column operations to A has a well-defined effect on $D(A)$.

(a) If B is obtained from A by adding c times the jth column to the ith, then $D(B) = D(A)$.

(b) If B is obtained from A by multiplying the ith column by a non-zero scalar c, then $D(B) = cD(A)$.

(c) If B is obtained from A by interchanging two columns, then $D(B) = -D(A)$.

For (a), let A' be the matrix which agrees with A in all columns except the ith, which is equal to the jth column of A. By rule (D2), $D(A') = 0$. By rule (D1),

$$D(B) = D(A) + cD(A') = D(A).$$

Part (b) follows immediately from rule (D1).

To prove part (c), we observe that we can interchange the ith and jth columns by the following sequence of operations:

- add the ith column to the jth;
- multiply the ith column by -1;
- add the jth column to the ith;
- subtract the ith column from the jth.

In symbols,

$$(c_i, c_j) \mapsto (c_i, c_j + c_i) \mapsto (-c_i, c_j + c_i) \mapsto (c_j, c_j + c_i) \mapsto (c_j, c_i).$$

The first, third and fourth steps don't change the value of D, while the second multiplies it by -1.

Now we take the matrix A and apply elementary column operations to it, keeping track of the factors by which D gets multiplied according to rules (a)–(c). The overall effect is to multiply $D(A)$ by a certain non-zero scalar c, depending on the operations.

- If A is invertible, then we can reduce A to the identity (see Corollary 2.8), so that $cD(A) = D(I) = 1$ (by (D3)), whence $D(A) = c^{-1}$.
- If A is not invertible, then its column rank is less than n. So the columns of A are linearly dependent, and one column can be written as a linear combination of the others. Applying axiom (D1), we see that $D(A)$ is a linear combination of values $D(A')$, where A' are matrices with two equal columns; so $D(A') = 0$ for all such A', whence $D(A) = 0$.

This proves that the determinant function, if it exists, is unique. We show its existence in the next section, by giving a couple of formulae for it.

Given the uniqueness of the determinant function, we now denote it by $\det(A)$ instead of $D(A)$. The proof of the theorem shows an important corollary:

Corollary 4.2 A square matrix is invertible if and only if $\det(A) \neq 0$.

Proof See the case division at the end of the proof of the theorem.

One of the most important properties of the determinant is the following.

Theorem 4.3 If A and B are $n \times n$ matrices over K, then $\det(AB) = \det(A)\det(B)$.

Proof Suppose first that B is not invertible. Then $\det(B) = 0$. Also, AB is not invertible. (For, suppose that $(AB)^{-1} = X$, so that $XAB = I$. Then XA is the inverse of B.) So $\det(AB) = 0$, and the theorem is true.

In the other case, B is invertible, so we can apply a sequence of elementary column operations to B to get to the identity (by Corollary 2.8). The effect of these operations is to multiply the determinant by a non-zero factor c (depending on the operations), so that $c\det(B) = \det(I) = 1$, or $c = (\det(B))^{-1}$. Now these operations are represented by elementary matrices; so we see that $BQ = I$, where Q is a product of elementary matrices (see Lemma 2.2).

If we apply the same sequence of elementary operations to AB, we end up with the matrix $(AB)Q = A(BQ) = AI = A$. The determinant is multiplied by the same factor, so we find that $c\det(AB) = \det(A)$. Since $c = (\det(B))^{-1}$, this implies that $\det(AB) = \det(A)\det(B)$, as required.

Finally, we have defined determinants using columns, but we could have used rows instead:

Proposition 4.4 The determinant is the unique function D of $n \times n$ matrices which satisfies the conditions

(D1′) for each $1 \le i \le n$, the function D is a linear function of the ith row;

(D2′) if two rows of A are equal , then $D(A) = 0$;

(D3′) $D(I_n) = 1$.

The proof of uniqueness is almost identical to that for columns. To see that $D(A) = \det(A)$: if A is not invertible, then $D(A) = \det(A) = 0$; but if A is invertible, then it is a product of elementary matrices (which can represent either row or column operations), and the determinant is the product of the factors associated with these operations.

Corollary 4.5 If $A^\top$ denotes the transpose of A, then $\det(A^\top) = \det(A)$.

For, if D denotes the "determinant" computed by row operations, then $\det(A) = D(A) = \det(A^\top)$, since row operations on A correspond to column operations on $A^\top$.

4.2 Calculating determinants

We now give a couple of formulae for the determinant. This finishes the job we left open in the proof of the last theorem, namely, showing that a determinant function actually exists!

The first formula involves some background notation (see also the additional sheet Permutations, available from the module website).

Definition 4.2 A *permutation* of $\{1,\dots,n\}$ is a bijection from the set $\{1,\dots,n\}$ to itself. The *symmetric group* S_n consists of all permutations of the set $\{1,\dots,n\}$. (There are $n!$ such permutations.) For any permutation $\pi \in S_n$, there is a number $\operatorname{sign}(\pi) = \pm 1$, computed as follows: write π as a product of disjoint cycles; if there are k cycles (including cycles of length 1), then $\operatorname{sign}(\pi) = (-1)^{n-k}$. A *transposition* is a permutation which interchanges two symbols and leaves all the others fixed. Thus, if τ is a transposition, then $\operatorname{sign}(\tau) = -1$.

The last fact holds because a transposition has one cycle of size 2 and $n-2$ cycles of size 1, so $n-1$ altogether; so $\operatorname{sign}(\tau) = (-1)^{n-(n-1)} = -1$.

We need one more fact about signs: if π is any permutation and τ is a transposition, then $\operatorname{sign}(\pi\tau) = -\operatorname{sign}(\pi)$, where $\pi\tau$ denotes the composition of π and τ (apply first τ, then π).

Definition 4.3 Let A be an $n \times n$ matrix over K. The *determinant* of A is defined by the formula

$$\det(A) = \sum_{\pi \in S_n} \operatorname{sign}(\pi) A_{1\pi(1)} A_{2\pi(2)} \cdots A_{n\pi(n)}.$$

Proof In order to show that this is a good definition, we need to verify that it satisfies our three rules (D1)–(D3).

(D1) According to the definition, $\det(A)$ is a sum of $n!$ terms. Each term, apart from a sign, is the product of n elements, one from each row and column. If we look at a particular column, say the ith, it is clear that each product is a linear function of that column; so the same is true for the determinant.

(D2) Suppose that the ith and jth columns of A are equal. Let τ be the transposition which interchanges i and j and leaves the other symbols fixed. Then $\pi(\tau(i)) = \pi(j)$ and $\pi(\tau(j)) = \pi(i)$, whereas $\pi(\tau(k)) = \pi(k)$ for $k \neq i, j$. Because the elements in the ith and jth columns of A are the same, we see that the products $A_{1\pi(1)}A_{2\pi(2)}\cdots A_{n\pi(n)}$ and $A_{1\pi\tau(1)}A_{2\pi\tau(2)}\cdots A_{n\pi\tau(n)}$ are equal. But $\mathrm{sign}(\pi\tau) = -\mathrm{sign}(\pi)$. So the corresponding terms in the formula for the determinant cancel one another. The elements of S_n can be divided up into $n!/2$ pairs of the form $\{\pi, \pi\tau\}$. As we have seen, each pair of terms in the formula cancel out. We conclude that $\det(A) = 0$. Thus (D2) holds.

(D3) If $A = I_n$, then the only permutation π which contributes to the sum is the identity permutation ι: for any other permutation π satisfies $\pi(i) \neq i$ for some i, so that $A_{i\pi(i)} = 0$. The sign of ι is $+1$, and all the terms $A_{i\iota(i)} = A_{ii}$ are equal to 1; so $\det(A) = 1$, as required.

This gives us a nice mathematical formula for the determinant of a matrix. Unfortunately, it is a terrible formula in practice, since it involves working out $n!$ terms, each a product of matrix entries, and adding them up with $+$ and $-$ signs. For n of moderate size, this will take a very long time! (For example, $10! = 3628800$.)

Here is a second formula, which is also theoretically important but very inefficient in practice.

Definition 4.4 Let A be an $n \times n$ matrix. For $1 \leq i, j \leq n$, we define the (i, j) *minor* of A to be the $(n-1) \times (n-1)$ matrix obtained by deleting the ith row and jth column of A. Now we define the (i, j) *cofactor* of A to be $(-1)^{i+j}$ times the determinant of the (i, j) minor. (These signs have a chessboard pattern, starting with sign $+$ in the top left corner.) We denote the (i, j) cofactor of A by $K_{ij}(A)$. Finally, the *adjugate* of A is the $n \times n$ matrix $\mathrm{Adj}(A)$ whose (i, j) entry is the (j, i) cofactor $K_{ji}(A)$ of A. (Note the transposition!)

Theorem 4.6 (a) For $1 \leq j \leq n$, we have

$$\det(A) = \sum_{i=1}^{n} A_{ij}K_{ij}(A).$$

(b) For $1 \leq i \leq n$, we have

$$\det(A) = \sum_{j=1}^{n} A_{ij} K_{ij}(A).$$

This theorem says that, if we take any column or row of A, multiply each element by the corresponding cofactor, and add the results, we get the determinant of A.

Example 4.1 Using a cofactor expansion along the first column, we see that

$$\begin{aligned}
\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{vmatrix} &= \begin{vmatrix} 5 & 6 \\ 8 & 10 \end{vmatrix} - 4 \begin{vmatrix} 2 & 3 \\ 8 & 10 \end{vmatrix} + 7 \begin{vmatrix} 2 & 3 \\ 5 & 6 \end{vmatrix} \\
&= (5 \cdot 10 - 6 \cdot 8) - 4(2 \cdot 10 - 3 \cdot 8) + 7(2 \cdot 6 - 3 \cdot 5) \\
&= 2 + 16 - 21 \\
&= -3
\end{aligned}$$

using the standard formula for a 2×2 determinant.

Proof We prove (a); the proof for (b) is a simple modification, using rows instead of columns. Let $D(A)$ be the function defined by the right-hand side of (a) in the theorem, using the jth column of A. We verify rules (D1)–(D3).

(D1) It is clear that $D(A)$ is a linear function of the jth column. For $k \neq j$, the cofactors are linear functions of the kth column (since they are determinants), and so $D(A)$ is linear.

(D2) If the kth and lth columns of A are equal (where k and l are different from j), then each cofactor is the determinant of a matrix with two equal columns, and so is zero. The harder case is when the jth column is equal to another, say the kth. Using induction, each cofactor can be expressed as a sum of elements of the kth column times $(n-2) \times (n-2)$ determinants. In the resulting sum, it is easy to see that each such determinant occurs twice with opposite signs and multiplied by the same factor. So the terms all cancel.

(D3) Suppose that $A = I$. The only non-zero cofactor in the jth column is $K_{jj}(I)$, which is equal to $(-1)^{j+j} \det(I_{n-1}) = 1$. So $D(I) = 1$.

By the main theorem, the expression $D(A)$ is equal to $\det(A)$.

At first sight, this looks like a simple formula for the determinant, since it is just the sum of n terms, rather than $n!$ as in the first case. But each term is an $(n-1) \times (n-1)$ determinant. Working down the chain we find that this method is just as labour-intensive as the other one.

But the cofactor expansion has further nice properties:

Theorem 4.7 For any $n \times n$ matrix A, we have

$$A \cdot \mathrm{Adj}(A) = \mathrm{Adj}(A) \cdot A = \det(A) \cdot I.$$

Proof We calculate the matrix product. Recall that the (i, j) entry of $\mathrm{Adj}(A)$ is $K_{ji}(A)$.

Now the (i,i) entry of the product $A \cdot \mathrm{Adj}(A)$ is

$$\sum_{k=1}^{n} A_{ik}(\mathrm{Adj}(A))_{ki} = \sum_{k=1}^{n} A_{ik}K_{ik}(A) = \det(A),$$

by the cofactor expansion. On the other hand, if $i \neq j$, then the (i, j) entry of the product is

$$\sum_{k=1}^{n} A_{ik}(\mathrm{Adj}(A))_{kj} = \sum_{k=1}^{n} A_{ik}K_{jk}(A).$$

This last expression is the cofactor expansion of the matrix A' which is the same of A except for the jth row, which has been replaced by the ith row of A. (Note that changing the jth row of a matrix has no effect on the cofactors of elements in this row.) So the sum is $\det(A')$. But A' has two equal rows, so its determinant is zero.

Thus $A \cdot \mathrm{Adj}(A)$ has entries $\det(A)$ on the diagonal and 0 everywhere else; so it is equal to $\det(A) \cdot I$.

The proof for the product the other way around is the same, using columns instead of rows.

Corollary 4.8 If the $n \times n$ matrix A is invertible, then its inverse is equal to

$$(\det(A))^{-1}\,\mathrm{Adj}(A).$$

So how can you work out a determinant efficiently? The best method in practice is to use elementary operations.

Apply elementary operations to the matrix, keeping track of the factor by which the determinant is multiplied by each operation. If you want, you can reduce all the way to the identity, and then use the fact that $\det(I) = 1$. Often it is simpler to stop at an earlier stage when you can recognise what the determinant is. For example, if the matrix A has diagonal entries $a_1, \ldots, a_n$, and all off-diagonal entries are zero, then $\det(A)$ is just the product $a_1 \cdots a_n$.

Example 4.2 Let

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix}.$$

Subtracting twice the first column from the second, and three times the second column from the third (these operations don't change the determinant) gives

$$\begin{bmatrix} 1 & 0 & 0 \\ 4 & -3 & -6 \\ 7 & -6 & -11 \end{bmatrix}.$$

Now the cofactor expansion along the first row gives

$$\det(A) = \begin{vmatrix} -3 & -6 \\ -6 & -11 \end{vmatrix} = 33 - 36 = -3.$$

(At the last step, it is easiest to use the formula for the determinant of a 2×2 matrix rather than do any further reduction.)

4.3 The Cayley–Hamilton Theorem

Since we can add and multiply matrices, we can substitute them into a polynomial. For example, if

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

then the result of substituting A into the polynomial $x^2 - 3x + 2$ is

$$A^2 - 3A + 2I = \begin{bmatrix} -2 & 3 \\ -6 & 7 \end{bmatrix} + \begin{bmatrix} 0 & -3 \\ 6 & -9 \end{bmatrix} + \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

We say that the matrix A *satisfies the equation* $x^2 - 3x + 2 = 0$. (Notice that for the constant term 2 we substituted $2I$.)

It turns out that, for every $n \times n$ matrix A, we can calculate a polynomial equation of degree n satisfied by A.

Definition 4.5 Let A be a $n \times n$ matrix. The *characteristic polynomial* of A is the polynomial

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

This is a polynomial in x of degree n.

For example, if

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

then

$$c_A(x) = \begin{vmatrix} x & -1 \\ 2 & x-3 \end{vmatrix} = x(x-3) + 2 = x^2 - 3x + 2.$$

Indeed, it turns out that this is the polynomial we want in general:

Theorem 4.9 (Cayley–Hamilton Theorem) Let A be an $n \times n$ matrix with characteristic polynomial $c_A(x)$. Then $c_A(A) = O$.

Example 4.3 Let us just check the theorem for 2×2 matrices. If

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

then

$$c_A(x) = \begin{vmatrix} x-a & -b \\ -c & x-d \end{vmatrix} = x^2 - (a+d)x + (ad-bc),$$

and so

$$c_A(A) = \begin{bmatrix} a^2+bc & ab+bd \\ ac+cd & bc+d^2 \end{bmatrix} - (a+d)\begin{bmatrix} a & b \\ c & d \end{bmatrix} + (ad-bc)\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = O,$$

after a small amount of calculation.

Proof We use the theorem

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

In place of A, we put the matrix $xI - A$ into this formula:

$$(xI - A)\operatorname{Adj}(xI - A) = \det(xI - A)I = c_A(x)I.$$

Now it is very tempting just to substitute $x = A$ into this formula: on the right we have $c_A(A)I = c_A(A)$, while on the left there is a factor $AI - A = O$. Unfortunately this is not valid; it is important to see why. The matrix $\operatorname{Adj}(xI - A)$ is an $n \times n$ matrix whose entries are determinants of $(n-1) \times (n-1)$ matrices with entries involving x. So the entries of $\operatorname{Adj}(xI - A)$ are polynomials in x, and if we try to substitute A for x the size of the matrix will be changed!

Instead, we argue as follows. As we have said, $\operatorname{Adj}(xI - A)$ is a matrix whose entries are polynomials, so we can write it as a sum of powers of x times matrices, that is, as a polynomial whose coefficients are matrices. For example,

$$\begin{bmatrix} x^2+1 & 2x \\ 3x-4 & x+2 \end{bmatrix} = x^2\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + x\begin{bmatrix} 0 & 2 \\ 3 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ -4 & 2 \end{bmatrix}.$$

The entries in $\operatorname{Adj}(xI - A)$ are $(n-1) \times (n-1)$ determinants, so the highest power of x that can arise is x^{n-1}. So we can write

$$\operatorname{Adj}(xI - A) = x^{n-1}B_{n-1} + x^{n-2}B_{n-2} + \cdots + xB_1 + B_0,$$

for suitable $n \times n$ matrices $B_0, \ldots, B_{n-1}$. Hence

$$\begin{aligned} c_A(x)I &= (xI - A)\,\mathrm{Adj}(xI - A) \\ &= (xI - A)(x^{n-1}B_{n-1} + x^{n-2}B_{n-2} + \cdots + xB_1 + B_0) \\ &= x^n B_{n-1} + x^{n-1}(-AB_{n-1} + B_{n-2}) + \cdots + x(-AB_1 + B_0) - AB_0. \end{aligned}$$

So, if we let

$$c_A(x) = x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0,$$

then we read off that

$$\begin{array}{rcccl} & & B_{n-1} & = & I, \\ -AB_{n-1} & + & B_{n-2} & = & c_{n-1}I, \\ & & \cdots & & \\ -AB_1 & + & B_0 & = & c_1 I, \\ -AB_0 & & & = & c_0 I. \end{array}$$

We take this system of equations, and multiply the first by A^n, the second by A^{n-1}, ..., and the last by $A^0 = I$. What happens? On the left, all the terms cancel in pairs: we have

$$A^n B_{n-1} + A^{n-1}(-AB_{n-1} + B_{n-2}) + \cdots + A(-AB_1 + B_0) + I(-AB_0) = O.$$

On the right, we have

$$A^n + c_{n-1}A^{n-1} + \cdots + c_1 A + c_0 I = c_A(A).$$

So $c_A(A) = O$, as claimed.