Math 642:550 — Summer 2009
MTTh 6:00–8:30 PM Hill 425
Prof. Bumby

Supplement 3, The Cauchy-Binet formula

1. Introduction

The theorem that the determinant of a product of square matrices is the product of the determinants of the factors is **so memorable** that one is likely to lose sight of **the difficulty of its proof**. Here is one proof that is sketched in the textbook.

The elementary row operations used in Gaussian elimination are: (1) adding a multiple of one row to another; (2) multiplying a row by a nonzero constant; (3) interchanging two rows. Each of these may be accomplished by multiplying on the left by a suitable **elementary** matrix. Since these operations may be applied to any matrix, once we know that they are given by left multiplication, these matrices are given by applying those operations to the identity. The effect on the determinant in all cases is thus to multiply by the determinant of the matrix that performs the row operation by left multiplication. From this it follows (using the associative law for matrix multiplication) that any matrix A that can be written as a product of elementary matrices has the property that $\det(AB) = \det(A)\det(B)$ for all B. Gaussian elimination writes any **nonsingular** A as a product of elementary matrices, so this property holds for nonsingular matrices. To complete the proof, note that a singular matrix A is characterized by having a nontrivial **left nullspace**, and any vector in the left nullspace of A is also in the left nullspace of AB (the associative law again!). Thus, $\det(A) = 0$ implies $\det(AB) = 0$, completing the proof.

Similar ideas appears in the proof that $\det(A^{\mathsf{T}}) = \det(A)$. Here, the **equivalent** properties of having a nontrivial nullspace or a nontrivial left nullspace characterize singular matrices. For nonsingular matrices, the steps in Gaussian elimination may be summarized by the LU factorization of a matrix whose rows are a permutation of the given matrix. In this LU factorization, L has all diagonal entries equal to 1, so it is a product of the elementary matrices of the first type, but U may have arbitrary entries on its diagonal. If the matrix is nonsingular, these entries are all nonzero, so the **old** U is factored into a diagonal matrix D times another matrix, unfortunately also called U, with all diagonal entries equal to 1. This gives the LDU factorization in which L and U have symmetric roles, and which is uniquely determined for each **nonsingular** PA for which Gaussian elimination works with no further row interchanges. Then, $(PA)^{\mathsf{T}} = (LDU)^{\mathsf{T}} = U^{\mathsf{T}} D^{\mathsf{T}} L^{\mathsf{T}}$. Since $D^{\mathsf{T}} = D$, each factor has the same determinant as its transpose, so the same must be true of the product.

There is a more tedious proof that $\det(AB) = \det(A)\det(B)$, in the spirit of the use of linearity to obtain the full expansion of a determinant, that can be used to evaluate $\det(AB)$ more generally whenever A is an m by n matrix and B is an n by m matrix. The patience with the proof is rewarded with a stronger theorem. The expression in this theorem will reduce to zero if $m > n$, for then AB is certainly a singular matrix. The more general conclusion justifies the effort of examining that proof.

2. The Cauchy-Binet formula

We follow F. R. Gantmacher, *The Theory of Matrices*, Chelsea, 1990 (except for inverting the names of the creators of the formula to agree with present usage). The expressions that appear will initially be indexed by all functions ϕ from $\{1, \ldots, m\}$ to $\{1, \ldots, n\}$, but only those functions ψ for which $\psi(1) < \cdots < \psi(m)$ will appear in the final formula. For each such function, we use its values $\psi(1), \ldots, \psi(m)$ to select m columns of A to form an m by m matrix A_ψ, and the **corresponding rows** of B to form an m by m matrix B_ψ. Then, the desired formula is

$$\det(AB) = \sum_{\psi} \det(A_\psi)\det(B_\psi). \tag{1}$$

3. First part of the proof

Here, we explore the dependence on the first factor, building a formula by processing the m rows of A in order. The proof will be an induction on the number of rows that have already been processed, which we denote by k. The basis of the induction will be the case $k = 0$, in which we have only the **untouched** matrix AB.

In the induction step, from $k = j$ to $k = j + 1$, each term will split into n different terms, which will be identified with the value to be assigned to $\phi(j + 1)$ **for that term**. When the process is done, we will want each term to be matched with a function ϕ from $\{1, \ldots, m\}$ to $\{1, \ldots, n\}$. To reach such a conclusion, the **statement** of the inductive step includes a requirement that, after k steps, $\det(AB)$ is a sum of n^k terms indexed by the restriction of ϕ to $\{1, \ldots, n\}$. The form of the terms will appear in the proof. For the moment, we note that the terms will be a product of k entries of A times a determinant whose first k rows are taken from B and whose remaining rows are the original rows $k + 1$ through m of AB The description just given of the basis of the induction has this property with $k = 0$ where the one term corresponds to the empty function. That term must be the **empty product** of elements of A times $\det(AB)$ itself. A complete identification of the role of ϕ will be identified in the description of the induction step.

For the induction step, consider the dependence of $\det(AB)$ on row $j+1$. Our description of the terms shows that all terms have a factor that is a determinant with this row as row $j + 1$, so our analysis will apply to all such determinants. In the discussion of the defining properties of determinants, it was noted that this is a linear function. However, the rules of matrix multiplication tell us that row $j + 1$ of AB is $\sum_{l=1}^{n} a_{(j+1)\,l} B_l$, where B_l is row l of B. Thus $\det(AB)$ is a sum of n terms, each of which is $a_{(j+1)\,l}$ times the determinant of the matrix that results from replacing row $j + 1$ of AB by row l of B. Selecting the l^{th} term corresponds to restricting to functions with $\phi(j + 1) = l$, and the new factor coming from A in this term is $a_{(j+1)\,\phi(j+1)}$. This construction preserves the property that the term corresponding to ϕ consist of the product of all $\big(i, \phi(i)\big)$ entries of A and a matrix whose first k rows are the rows of B given by $\phi(i)$ for $1 \le i \le k$ with row l for $l > k$ being row l of AB. Thus, the elements of A are the $a_{i\,\phi(i)}$ for $1 \le i \le k$ and the rows of B are those indexed by $\phi(1)$ through $\phi(k)$.

When j terms have been processed, we have n^j terms, so after all rows has been processed in this manner, we have m^n terms indexed by functions ϕ each of which multiplies the single term

$$\prod_{i=1}^{m} a_{i\,\phi(i)}$$

formed from elements of A with a determinant built from the rows $\phi(i)$ for $i = 1, \ldots, m$ of B

4. Cancellation and permutation

If a matrix has two equal rows, its determinant is zero. Thus, if $\phi(i) = \phi(j)$ for some $i \neq j$, the term corresponding to ϕ is zero. Dropping these terms out of the sum restricts to one-to-one functions ϕ. In particular, the fact that $\det(AB) = 0$ if $m > n$ has now been proved, since all terms have been shown to be zero.

We also know that interchanging two rows of a matrix changes the sign of the determinant. Thus all ϕ with the same **set of values** $\{\phi(1), \ldots, \phi(m)\}$ have closely related contributions from B. **Sorting** these values expresses ϕ in the form $\sigma \circ \psi$, where ψ is an increasing function and σ is a permutation of the range of ψ.

We now collect those terms with the same ψ. This amounts to selecting the **set of rows** of B that we will use. Looking back at the full expansion, we see that these values also tell us which columns of A will be used. The increasing function ψ is just a **standard way to describe a set** of m elements selected from $\{1, \ldots, n\}$. Focusing on a single ψ, the terms corresponding to $\phi = \sigma \circ \psi$ now reduce to

$$\left(\sum_{\sigma} (-1)^{\operatorname{sgn}(\sigma)} a_{1\phi(1)} \cdots a_{m\phi(m)} \right) \det(B_\psi).$$

In this process, the contribution of B was limited to selecting sets of m rows and using the property that an interchange of rows of a determinant changes the sign. In particular, if B is zero except for a 1 in each $(\psi(i), i)$ position ($i = 1, \dots, m$), then AB is the submatrix A_ψ of A and the formula just found is the usual expansion of the determinant of this submatrix. This completes the proof of the formula. In this proof, we considered $\det(AB)$ as a function of a **variable matrix** B determined by the entries of A.

5. Conclusion, a special case.

We close with an interesting special case. Suppose $A = B^\mathsf{T}$. Then $\det(B^\mathsf{T}B)$ gives the **square** of the m dimensional volume of the parallelepiped in $\mathbb{R}^n$ with the columns of B as edges. This was proved in **Application 3 in Section 4.4 of the textbook** by a geometric argument inspired by the Gram-Schmidt process. In this proof, $B^\mathsf{T}B$ collect information about the **intrinsic geometry** of the parallelepiped. The entries are just the inner products of vectors giving the edges of the figure, so that the diagonal gives the squares of the length of the edges and the other entries allow the angles between the edges to be found. This information is preserved if the coordinates are changed to a system determined by the position of the figure in space. In this coordinate system, the volume is compared to that of a related figure in which the edges are mutually perpendicular.

On the other hand, for the Cauchy-Binet formula, $A_\psi = (B_\psi)^\mathsf{T}$, so $\det(A_\psi) = \det(B_\psi)$, and the formula reduces to

$$\det(B^\mathsf{T}B) = \sum_\psi \det(B_\psi)^2. \tag{2}$$

This says that the square of the volume is the sum of the squares of the volumes of the projections of the figure into all possible m dimensional **coordinate planes**.

The case $m = 1$ of this is the Pythagorean formula, and the case where $m = 2$ and $n = 3$ arises in showing the connection of the cross product of vectors with areas of plane figures in $\mathbb{R}^3$.

6. Exercises

A. Let

$$A = \begin{bmatrix} 1 & 1 & 0 & -1 & 2 \\ 2 & -1 & 0 & 1 & 4 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ 0 & 0 \\ -1 & 2 \\ -9 & 3 \\ 3 & -2 \end{bmatrix}. \tag{3}$$

Find $\det(A)$ by:
(i) expanding (3) and finding the determinant of the resulting expression for A; and
(ii) using the general Cauchy-Binet formula (1).

B. Find the area of the parallelogram in $\mathbb{R}^4$ whose sides are

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -1 \\ 1 \\ 1 \\ -1 \end{bmatrix}$$

by letting B be the 4 by 2 matrix with these columns, so the square of the area is $\det(B^\mathsf{T}B)$, and then:
(i) finding $B^\mathsf{T}B$ and evaluating its determinant; and
(ii) using formula (2).

Both calculations should be easy, and you should get the same answer.

C. Find the volume of the three dimensional parallelepiped in $\mathbb{R}^4$ whose sides are

$$\begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \qquad \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix} \qquad \begin{bmatrix} -1 \\ 0 \\ 0 \\ -1 \end{bmatrix}$$

by letting C be the 4 by 3 matrix with these columns, so the square of the volume is $\det(C^{\mathsf{T}}C)$, and then:

(i) finding $C^{\mathsf{T}}C$ and evaluating its determinant; and

(ii) using formula (2).

Again, both calculations should give the same answer.

End of Supplement