642:550, Summer 2004, Supplement 3

The Cauchy-Binet formula

Summer 2004

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**. One proof uses Gaussian elimination to write a matrix as a product of elementary matrices and exploits the fact that left multiplication by an elementary matrix gives a row operation whose effect on the determinant is known. This means that $\det(AX)$ must be a constant multiple of $\det(X)$. Applying this to the special case where X is the identity shows that the multiplier must be $\det(A)$.

There is a more tedious proof, in the spirit of the use of linearity to obtain the full expansion of a determinant, that can be used to evaluate $\det(AB)$ when 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.

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, by 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 $\{1,\ldots,k\}$. The description just given of the basis of the induction has this property. A further property that must be included in the statement will be identified in the description of the induction step.

For the induction step, consider the dependence of $\det(AB)$ on row $j+1$. 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 the l^{th} row 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 the lth row of B. Selecting the l^{th} term corresponds to restricting to functions with $\phi(j+1) = l$, and the factor coming from A in this term is $a_{(j+1)\,\phi(j+1)}$. The additional property preserved by this construction is that the terms consist of the product of (a specific set of) k entries of A and a matrix whose first k rows are (specific) rows of B with row l for $l > k$ being row l of AB.

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 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 ϕ of the form $\sigma \circ \psi$ now reduce to

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

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 ($i = 1, \ldots, m$), 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 $A = B^{\mathrm{T}}$. Then $\det(B^{\mathrm{T}}B)$ 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** by a geometric argument inspired by the Gram-Schmidt process. In this proof, $B^{\mathrm{T}}B$ 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 alow 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, $A_{\psi} = (B_{\psi})^{\mathrm{T}}$, so $\det(A_{\psi}) = \det(B_{\psi})$, and the formula reduces to

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

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. 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 volume is $\det(B^{\mathrm{T}}B)$, and then:

(i) finding $B^{\mathrm{T}}B$ and evaluating its determinant; and

(ii) using formula (2).

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

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

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

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

(i) finding $B^{\mathrm{T}}B$ and evaluating its determinant; and

(ii) using formula (2).

Again, both calculations should give the same answer.

End of Supplement