Math 642:550 — Summer 2003
MTTh 6:15–8:45 PM Hill 525
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**. 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 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 and 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

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

3. First part of the proof: dependence on first factor

We build a formula by processing the m rows of A in order. At the k^{th} step, each term will split into n different terms, which will be identified with the value of $\phi(k)$ **for that term**. When the process is done, each term will be matched with a function ϕ from $\{1,\ldots,m\}$ to $\{1,\ldots,n\}$.

We begin by looking at the dependence of $\det(AB)$ on its first row. One of the defining properties of determinants says that this must be a linear function. However, the rules of matrix multiplication tell us that the first row of AB is $\sum_{j=1}^{n} a_{1j} B_j$, where B_j is the j^{th} row of B. Thus $\det(AB)$ is a sum of n terms, each of which is a_{1j} times the determinant of the matrix that results from replacing the first row of AB by the jth row of B. Selecting the j^{th} term corresponds to restricting to functions with $\phi(1) = j$. In particular, the factor coming from A in this term is $a_{1\phi(1)}$.

Continuing in this fashion, when the first k rows have been processed, we have n^k terms described by $\phi(1), \ldots, \phi(k)$, consisting of $a_{1\phi(1)} \cdots a_{k\phi(k)}$ times the determinant of a matrix obtained from AB by replacing the first k rows by the rows of B given by the values of ϕ. More precisely, the j^{th} row is replaced by $B_{\phi(j)}$. To go the next step, one uses the expansion of the determinant by row $k+1$ in the same way the first row was used in the previous paragraph.

After the m^{th} row 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

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

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 the $(\psi(i), i)$ position for $i = 1, \ldots, m$, 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 multiplying B on the right by an upper triangular matrix M that (as in the Gram-Schmidt process) replaces the columns of B by vectors perpendicular to all previous columns. Multiplying A on the left by M^{T} gives $M^{\mathrm{T}}A = M^{\mathrm{T}}B^{\mathrm{T}} = (BM)^{\mathrm{T}}$, and $(BM)^{\mathrm{T}}(BM)$ is diagonal since the columns of BM are orthogonal. We can take M to be upper triangular with 1 in each diagonal position. Then the columns of BM span a figure of the same volume as the columns of B while $\det(M) = 1$ (the textbook chooses to normalize the columns of BM to have length 1, so that M must keep track of the lengths).

In this case, $A_\psi = (B_\psi)^{\mathrm{T}}$, so $\det(A_\psi) = \det(B_\psi)$. The formula reduces to

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

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 $\mathbb{R}^3$.

Another interpretation of this formula is that $B^{\mathrm{T}}B$ contains information about the **intrinsic** geometry of the figure: The diagonal gives the squares of the lengths of the sides and the other entries give inner products from which the cosines of angles between the sides can be found. Since the cosine is an **even function**, the distinction between an angle and its negative is lost, so this suppresses information about the orientation of the figure.

End of Supplement