DETERMINANTS, COFACTORS, AND INVERSES

1 1.0 GENERAL

Determinants originally were encountered in the solution of simultaneous linear equations, so we will use that perspective here. A general statement of this problem is: if the N by N matrix A and the N-vector y, both consisting of real or complex elements, are known, solve the N equations

$$A \cdot \underline{x} = \underline{y} \tag{1.1}$$

for the N elements of the vector $\underline{x}$. In the paragraphs that follow, the essentials of solving simultaneous linear equations will be examined.

The determinant is the key quantity and will be explored first. This will be used as a basis for finding the properties and computation of the solutions for the elements of the vector $\underline{x}$.

In solving for an element of $\underline{x}$, individual equations involving a single element of $\underline{y}$, or rows of the equation as given above, are scaled and added to other rows to eliminate dependence on elements of $\underline{x}$, until only a single element of $\underline{x}$ occurs in a row; then the equation for this element may be solved. Systematically eliminating the first element from the second and lower rows, the second element from the third and lower rows, etc. allows element *N* of $\underline{x}$ to be found first. Then element *N-1* can be found, etc. This process is called Gaussian elimination, and is discussed in more detail later.

The solution for each element of the vector $\underline{x}$ is in the form of the ratio of polynomials in the elements of the matrix *A* and the vector $\underline{y}$. The denominator is always the same polynomial in the elements of the matrix *A* and is called the determinant. Each term has *N* factors, one from each row and one from each column.

In the process of performing Gaussian elimination, two fundamental properties of the determinant have been noted. These are the axiomatic properties:

(a) The value of the determinant of a diagonal matrix is the product of the values along the main diagonal.

(b) The value of the determinant of a matrix is unchanged when any row is multiplied by a constant and added to any other row.

2 EULER'S EXPANSION

Euler's expansion of a determinant is an expansion of each of the terms in the determinant in terms of the elements of the rows. We define it as the sum

$$|A| = \sum_{p_1, p_2, p_3 \cdots p_N = 1}^{N} \varepsilon_{p_1 p_2 p_3 \cdots p_N} \cdot a_{1,p_1} \cdot a_{2,p_2} \cdot a_{3,p_3} \ldots a_{N,p_N} \tag{2.1}$$

where $\varepsilon_{p_1 p_2 p_3 \cdots p_N}$ is +1, -1, or zero, depending on the indices p_1, p_2, etc. This dependence is expressed by the following rules:

(a) If any two arguments p_i and p_j are equal, $\varepsilon_{p_1 p_2 p_3 \cdots p_N}$ is zero.

(b) All subscripts of $\varepsilon_{p_1 p_2 p_3 \cdots p_N}$ are distinct if it is nonzero, by (a) above. The arguments can therefore be reordered to obtain the monotonic sequence *(1,2,3,...N)* by interchanging adjacent arguments (for example, by using a bubble sort procedure). If the number of interchanges to produce the monotonic sequence is even, the value of $\varepsilon_{p_1 p_2 p_3 \cdots p_N}$ is +1. If the number of interchanges is odd, the value *s* is -1.

There are *N!* terms in the summation. We will show that Euler's expansion has all the properties of the determinant, and thus is equal to the determinant.

From this explanation, it is seen that is $\varepsilon_{p_1 p_2 p_3 \cdots p_N}$ equal to the determinant of a matrix formed by constructing each row *n* by placing a 1 in column p_n and zeros in the other columns. The determinant of a matrix formed by replacing row *i* with zeros except for a 1 at column *j* is called the *cofactor of element* $a_{i,j}$ because it represents the only term in the determinant that has a factor of $a_{i,j}$. The quantity $\varepsilon_{p_1 p_2 p_3 \cdots p_N}$ is called the completely antisymmetric tensor of rank N.

The sum given by Euler's expansion can be shown to be equal to the value of the determinant by showing that the two axiomatic properties apply. Since the Euler's expansion for a matrix defines a unique value for the determinant, this serves as a proof that the value of the determinant of a matrix is uniquely defined by the two axiomatic properties. Restating this as a theorem:

THEOREM: The unique value of the determinant of a square matrix of real or complex numbers is given by Euler's expansion.

PROOF: The fact that the value of the sum of a diagonal matrix is the product of the terms on the diagonal is immediately obvious. To complete the proof, we only need to show that the value of the sum is unchanged by adding one row, multiplied by a constant c, to another row. Suppose row i is multiplied by c and added to row j, where j is not i. The sum is

$$\begin{aligned} |A| &= \sum_{p_1, p_2, p_3 \cdots p_N = 1}^{N} \varepsilon_{p_1 p_2 p_3 \cdots p_N} \cdot a_{1,p_1} \ldots \left(a_{j,p_j} + c \cdot a_{i,p_i} \right) \ldots a_{N,p_N} \\ &= |A| + c \cdot \sum_{p_1, p_2, p_3 \cdots p_N = 1}^{N} \varepsilon_{p_1 p_2 p_3 \cdots p_N} \cdot a_{1,p_1} \ldots a_{i,p_i} \ldots a_{N,p_N} \\ &= |A|. \end{aligned} \tag{2.2}$$

The sum in the second line is the determinant of a matrix with two identical rows. In order to avoid circular logic, it is necessary to avoid properties of the determinant not shown at this point, and to show that this sum is zero using only the form given immediately above for the sum. This is done by noting that, for each distinct pair of indices (p_i,p_j), the same term in the summation appears for $\varepsilon_{p_1 p_2 p_3 \cdots p_N}$ equal to +1 and to -1 for the term where the indices are reversed, (p_j,p_i). This is because, if the indices are r rows apart, r interchanges are required to move p_i up to row p_j, and $r-1$ interchanges are required to move p_j down to the initial position of p_j. This is a total of $2 \cdot r-1$ interchanges, which is always an odd number, so that the sign of $\varepsilon_{p_1 p_2 p_3 \cdots p_N}$ is reversed. This completes the proof.

Euler's expansion can be used to show that the axiomatic properties apply to columns as well as rows as follows. The diagonal matrix property applies directly. From Euler's expansion, multiplying column p_i by c and adding to column p_j gives the sum

$$\begin{aligned} C &= \sum_{p_1, p_2, p_3 \cdots p_N = 1}^{N} \varepsilon_{p_1 p_2 p_3 \cdots p_N} \cdot a_{1,p_1} \ldots \left(a_{j,p_j} + c \cdot a_{j,p_i} \right) \ldots a_{N,p_N} \\ &= |A| + c \cdot \sum_{p_1, p_2, p_3 \cdots p_N = 1}^{N} \varepsilon_{p_1 p_2 p_3 \cdots p_N} \cdot a_{1,p_1} \ldots a_{j,p_i} \ldots a_{N,p_N} \end{aligned} \tag{2.3}$$

The second sum in the last form given for C does not contain any elements of the matrix A corresponding to column p_j. It is the determinant of the matrix A modified by having column p_j replaced by column p_i. Since the sign of $\varepsilon_{p_1 p_2 p_3 \cdots p_N}$ is reversed when the arguments p_i and p_j are reversed in the subscript list of $\varepsilon_{p_1 p_2 p_3 \cdots p_N}$, but the remainder of the term in the sum remains the same, the value of the sum is zero. Restating this result as a theorem,

THEOREM: Any operation on the rows of a matrix which leaves the value of the determinant unchanged can also be applied to the columns and also leaves the value of the determinant unchanged.

PROOF: The proof follows from the preceding discussion. Since the axiomatic properties defining the determinant are the same for columns as for rows, and since all operations on the rows which leave the determinant unchanged follow from the axiomatic properties, these operations also leave the determinant unchanged when applied to columns. This completes the proof.

2.1 Minors and Cofactors

Euler's expansion can be used to show the dependence of the determinant on a single element. This is most simply expressed as

$$\frac{\partial |A|}{\partial a_{i,j}} = \sum_{\substack{p_1 \cdots p_{i-1}, p_{i+1} \cdots p_N = 1 \\ p_j = j}}^{N} \varepsilon_{p_1 p_2 p_3 \cdots p_N} \cdot a_{1,p_1} \ldots 1 \ldots a_{N,p_N} \tag{2.4}$$

where the only allowable value of p_i in the sum is j, and the element of the matrix A in row i is +1. This is clearly the determinant of the matrix A, modified by setting to zero all elements in row *i* and column *j* except the element in position *(i,j)*, which is +1. Comparison with the minor about this same position, defined as the determinant of the matrix *A* with row *i* and column *j* removed,

$$\text{minor}(i,j) = \sum_{p_1 \cdots p_{i-1}, p_{i+1} \cdots p_N = 1}^{N} \varepsilon_{p_1 \cdots p_{i-1}, p_{i+1} \cdots p_N} \cdot a_{1,p_1} \ldots a_{i-1,p_{i-1}} \cdot a_{i+1,p_{i+1}} \ldots a_{N,p_N} \tag{2.5}$$

it is clear that the determinants are the same, except possibly for sign. Comparing the two sign functions in the sums, it can be seen that

(a) The sign functions are the same if $j=i$, because setting the argument *i* equal to *i* reduces the *N* argument sign function to the *N-1* argument sign function.

(b) If *j* differs from *i* by 1 (or any odd number), an odd number of interchanges must be made in the arguments of the cofactor's sign function to reduce it to the case where $j=i$, so that it is the negative of the sign function for the minor.

(c) Conversely, if *j* differs from i by any even number, an even number of interchanges must be made in the arguments of the cofactor's sign function to reduce it to the case where $j=i$, so that it is the same as the sign function for the minor.

Thus, the cofactor of element *(i,j)* is equal to value of the minor about *(i,j)* when $i+j$ is even, and the cofactor of element *(i,j)* is the negative of the minor about *(i,j)* when $i+j$ is odd.

2.2 Expansion by Cofactors: Laplace's Development

From the previous discussion, it can be seen immediately that the sum across any row of the matrix A of the cofactors of that row times the elements in that row is equal to the value of the determinant:

$$|A| = \sum_{j=1}^{N} c\,\text{of}\left(a_{i,j}\right) \cdot a_{i,j} \tag{2.6}$$

By the symmetry discussed above, the determinant can also be expressed as a similar sum over any column as well. This equality is also called expansion by minors.

2.3 The Inverse

The inverse of a matrix is easily verified to be given by

$$A^{-1} = \left[c_{i,j}\right] \tag{2.7}$$

where

$$c_{i,j} = \frac{\text{cof}\left(a_{j,i}\right)}{|A|} \tag{2.8}$$

The form given for the inverse follows from an analysis of the vector

$$\underline{b} = \left[\operatorname{cof}\left(a_{j,k} \right) \right] \cdot \underline{a}_k = \operatorname{Cof}\left(A^T \right) \cdot \underline{a}_k \tag{2.9}$$

where a_k is row k of the matrix A. For any row i, the element of $\underline{b}$ is equal to a determinant of a matrix in which every row is equal to the corresponding row of A except row i, which is replaced by row k. Thus, every row of $\underline{b}$ is zero, except row k, which is equal to |A|. From this, it is evident that

$$A^{-1} = \frac{\operatorname{Cof}\left(A^T \right)}{|A|}. \tag{2.10}$$

When $|A|$ is not zero, the inverse exists and is given by

$$A^{-1} = \frac{\operatorname{Cof}\left(A^T \right)}{|A|} \tag{2.11}$$

which is the proposed definition of the inverse.

2.4 Cramer's Rule

When solving the N equations in N unknowns

$$A \cdot \underline{x} = \underline{y} \tag{2.12}$$

for x, the solution is obviously

$$\underline{y} = A^{-1} \cdot \underline{x} = \frac{\operatorname{Cof}\left(A^T \right) \cdot \underline{x}}{|A|}. \tag{2.13}$$

From the discussion above on expansion by minors, row i of the product Cof$(A^T)\cdot x$ is equal to the determinant of a matrix made up by replacing column i of the matrix A with the vector x:

$$y_i = \frac{\sum_{j=1}^{N} \operatorname{cof}\left(a_{j,i} \right) \cdot x_j}{|A|}. \tag{2.14}$$

This can be useful when N is small, or when only a few elements of the vector $\underline{y}$ are to be found.

3 GAUSSIAN ELIMINATION

The process of diagonalizing a matrix by adding scaled rows to other rows is called Gaussian elimination. The first column below the top row can be annihilated (reduced to zeros) as follows. The elements of the matrix A are represented by

$$A = \left[a_{i,j} \right] \tag{3.1}$$

where the first subscript refers to the row and the second refers to the column. Axiom (b) is used by multiplying the first row by $-(a_{i,1}/a_{1,1})$ and adding to each row i below the first. The result is a matrix which has all zeros in the first column, except for the element $a_{1,1}$, which is unchanged from that of the original matrix.

The operation of annihilating the first column of a matrix A by Gaussian elimination is equivalent to left-multiplying the matrix by an upper triangular matrix G_1 which has ones on the main diagonal and zeros elsewhere except across the top row, which contains the ratios $-(a_{i,1}/a_{1,1})$ in column i.

The second column of the modified matrix can be annihilated below $a_{2,2}$ by repeating the process, except that the first row is ignored this time and the second row is scaled and added to the other rows. This process is repeated for the third column and other columns until the result is a triangular matrix. The annihilation of each column is equivalent to left-multiplying by an upper triangular matrix which has ones on the main diagonal.

For the purposes of programming Gaussian elimination as an algorithm or to tie up loose ends for a rigorous proof, the special case of encountering a zero element on the main diagonal during the course of the progression of the algorithm must be considered. Suppose that, after annihilating column $j-1$, element $a_{j,j}$ is zero. When this does occur, any row below row j which contains a nonzero value for the element in column j can be added to row j to obtain a new row which contains a nonzero element on the main diagonal. If there are none, Gaussian elimination continues with the next row because annihilation of column j is unnecessary.

Once an upper triangular matrix is obtained, the columns above the main diagonal can be eliminated in the same way, this time beginning with the bottom row. Notice that eliminating the upper triangular portion does not change the values of the diagonal elements during the procedure.

If one of the diagonal elements $a_{j,j}$ is zero, the entire row j of the matrix is zero once columns $j+1$ through N have been annihilated, so that the value of the determinant is zero, as is shown in property (6) below.

4 ELEMENTARY PROPERTIES OF THE DETERMINANT

From the two axiomatic properties (a) and (b) above, and from Euler's expansion, several properties of the determinant of a matrix immediately follow:

(1) The value of the determinant of a triangular matrix is the product of the values along the main diagonal. This is because a triangular matrix can be converted to a diagonal matrix by using (b) above without changing the elements on the main diagonal.

(2) The property (b) applies to columns also, as was shown in the discussion of Euler's expansion.

(3) Since the properties determining the value of a determinant are symmetrical for rows and columns, the value of the determinant of the transpose of a matrix is the same as the determinant of the matrix.

(4) Since (b) and (2) are exactly equivalent to left-multiplying and right multiplying a matrix by a triangular matrix with ones on the main diagonal, it also follows that multiplying a matrix by triangular matrices whose main diagonal elements are all ones results in a matrix whose determinant is equal to the determinant of the original matrix.

(5) Multiplying any row or column by a constant results in a matrix whose determinant is the determinant of the original matrix multiplied by the same constant; this follows immediately from Euler's expansion.

(6) A determinant of a matrix with a row (or column) of all zeros, or which is equal or proportional to another row (or column) is zero, since this determinant is equal to another determinant times zero.

(7) The determinant of the product of two matrices A*B is the product of their determinants. This is because the matrix A can be converted to an upper (or lower) triangular form by Gaussian elimination by rows, which is equivalent to left multiplication by a triangular matrix with ones on the main diagonal, while the matrix B can be converted to upper (or lower) triangular form by columns. The values of the diagonal elements of the product of the two triangular matrices At and Bt are given as the products of the corresponding diagonal elements of At and Bt.

(8) Interchanging any two rows (or columns) of a matrix toggles the arithmetic sign of the value of its determinant. This can be seen because it is equivalent to left (or right) multiplication by an identity matrix which has the same columns (or rows) interchanged. This matrix is easily converted to a triangular matrix which has a -1 on the main diagonal.

5 A GEOMETRICAL INTERPRETATION OF THE DETERMINANT

The determinant of a matrix can be interpreted in terms of its row (or column) vectors in N-space:

The value of the determinant of an N by N matrix is the volume of the parallelepiped defined by a vertex at the origin and subtended by its row vectors in N-space.

This is shown in geometric terms as follows. First, we must agree that the volume of the parallelepiped is, when all the sides are orthogonal, the product of the lengths of the sides. When the sides are not orthogonal, the volume of the parallelepiped is equal to the volume of another parallelepiped which is obtained from the first parallelepiped by moving its vertices in a direction parallel to one of its subtending vectors. This is illustrated for a parallelogram:

```
   _________         __________
  /        /   -->  |          |
 /________/         |__________|
```

The vertical sides were "squared up" with the horizontal sides by adding components parallel to the horizontal sides, and the area of the rectangle is equal to the volume of the parallelogram.

For a general matrix, this process can proceed as follows. Starting with the bottom row aN, the next lowest row aN-1 can be squared up with this vector by adding to it a vector parallel to aN. After this operation, row N-1 is

$$\underline{a}_{N-1} = \frac{\underline{a}_N^T \cdot \underline{a}_{N-1}}{\underline{a}_N^T \cdot \underline{a}_N} \cdot \underline{a}_N \tag{5.1}$$

which can be seen to be orthogonal to a_N. This operation does not change the height or depth of the parallelepiped with respect to the vector a_N or any facet bordered by the vector a_N. Since this operation is an example of (a) above, the determinant is not changed.

This operation is identical to left-multiplying the original matrix by an upper triangular matrix with ones on the main diagonal and zeros everywhere else except for the element at row N-1 and column N, which is equal to

$$v_{N-1,N} = -\frac{\underline{a}_N^T \cdot \underline{a}_{N-1}}{\underline{a}_N^T \cdot \underline{a}_N}. \tag{5.2}$$

The row third from the bottom can be squared up with the bottom two rows by subtracting the components along them in the same way. The matrix V now contains nonzero elements in row N-2 above the main diagonal. The elements in row N-2 are found as the ratio of dot products exactly as in the first example. Continuing in this way results in a matrix whose rows are orthogonal, and the definition of all the elements of a matrix V which is upper triangular with ones on the main diagonal.

This process is known as Gram-Schmidt orthogonalization. Representation of Gram-Schmidt orthogonalization by rows as left-multiplication by an upper triangular matrix with ones on the main diagonal is a concept that will be used later in the formulation of special Kalman filter techniques.

At this point, we represent the operation as

$$A = V \cdot A_0 \tag{5.3}$$

where V is upper triangular with ones on the main diagonal, and A_0 consists of rows which are orthogonal. We can factor A_0 into a diagonal matrix D, the values of the diagonal elements being given by the lengths of the rows of A_0, and the matrix A_0 with its rows rescaled so that they all are of length one:

$$A = V \cdot D \cdot A_{0s} \tag{5.4}$$

It is easy to show that

$$A_{0s} \cdot A_{0s}^T = I \tag{5.5}$$

so that the value of the determinant of Aos is plus or minus one and the value of the determinant of A is given by the product of the elements of D, which is the volume of the parallelepiped subtended by the origin and the row vectors of A_0. Since we carefully ensured that the volume subtended by the row vectors of A_0 was the same as that subtended by the row vectors of A, we have now proved the theorem.

There is one loose end; the arithmetic sign of the determinant has not been addressed. Since the determinant of A_{0s} appears on both sides of

$$A \cdot A_{os}^T = V \cdot D \cdot A_{0s} \cdot A_{0s}^T \tag{5.6}$$

then what we have is that if the value of the determinant of A is negative, then the value of the determinant of A_{0s} is -1.

The rows of A A_{0s} are orthogonal unit vectors. Considering A_{0s} as a transformation, the operation of A_{0s} on a vector $\underline{b}$,

$$\underline{b}' = A_{0s} \cdot \underline{b} \tag{5.7}$$

b' = Aos*b

we see that each element of $\underline{b}'$ is the projection of $\underline{b}$ onto each of the unit vectors given by the rows of A_{0s}. This means that if the rows of A_{0s} are considered as the axes of a rotated Cartesian coordinate system, then the elements $\underline{b}'$ are the components of $\underline{b}$ in that coordinate system. We know that the determinant of any such matrix will necessarily be plus or minus one. At this point, we will make an arbitrary definition:

DEFINITION: A coordinate system is right-handed if the determinant of a matrix T, made up of its unit vectors used as rows in the order, is positive. If the determinant of T is negative, the coordinate system is left-handed.

If the square matrix T consists of row vectors which are of magnitude one and which are mutually orthogonal, left-multiplication of column vectors by T produces a result which is the representation of the vector in the rotated Cartesian coordinate system whose axes are represented by the rows of T.

The definition of right-handed versus left-handed applies to oblique coordinate systems, since the process of Gram-Schmidt orthogonalization does not reverse the direction of any row except along the direction of previously defined orthogonal basis vectors.