Matrix Algebra, Class Notes (part 1)

by

Hrishikesh D. Vinod

1 Sum, Product and Transpose of Matrices.

If a_{ij} with i= $1, 2, \cdots$, m and j= $1, 2, \cdots$, n are a set of real numbers arranged in m rows and n columns we have an m by n matrix A= (a_{ij}). For example, when m= 2 and n= 3 we have a 2×3 matrix A defined as:

$$\mathrm{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} = \begin{bmatrix} 15 & -12 & 53 \\ 25 & 22 & -2 \end{bmatrix} \tag{1}$$

In the above example if we had three rows instead of two, it would have been a square matrix. A square matrix has as many rows as columns, or m= n. Vectors are matrices with only one column. Ordinary numbers can also be thought to be matrices of dimension 1×1, and are called scalars. Similar to most authors we use only column vectors, unless stated otherwise. Both vectors and matrices follow some simple rules of algebra.

(1) **Sum.** If A= (a_{ij}) and B= (b_{ij}) then C= (c_{ij}) is defined by $c_{ij} = a_{ij} + b_{ij}$. This is defined only when both A and B are of exactly the same dimension, and obtained by simply adding the corresponding terms. **Exercise** verify that A+B is a matrix of all zeros when A is given by (1) and

$$\mathrm{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix} = \begin{bmatrix} -15 & 12 & -53 \\ -25 & -22 & 2 \end{bmatrix} \tag{2}$$

(2) **Product.** If A= (a_{ij}) is m×n and B= (b_{ij}) has i= $1, 2, \cdots$, n and j= $1, 2, \cdots$, p then the product matrix C= (c_{ij}) is of dimension m×p, and is obtained by "row-column" multiplication as follows. $c_{ij} = \Sigma_{t=1}^{n} a_{it}b_{tj}$, where i= $1, 2, \cdots$, m and j= $1, 2, \cdots$, p. Note that a_{it} represent the elements of i-th row and b_{tj} represent elements of j-th column. It is obvious that the above product can be defined only when the range of t is the same for both matrices A and B, that is, A should have as many columns as B has rows. An easy way to remember this is that for our "row-column" multiplication, "column-rows" must

match. **Exercise 1**: Verify that the product AB does not exist or is not defined for A and B defined by (1) and (2). Since A has 3 columns B must have exactly 3 rows and any number of columns for the product AB to exist. **Exercise 2**: Verify that a product of two column vectors D and E does not exist if we write it as DE. We have to write the first vector D as a row vector D$'$, and then write D$'E$. **Exercise 3** Let the vector x= $\left[\begin{smallmatrix}11\\2\end{smallmatrix}\right]$ be a column vector. Verify that the sum of squares of its elements s= $11^2 + 2^2 = 125$ is a scalar, and that s= $x'x$, where x$'$ denotes row vector [11 2].

In general we do **not** expect $AB = BA$. In fact BA may not even exist even when AB is well defined. In the rare case when $AB = BA$ we say that matrices A and B commute.

(3) **Multiplication of a matrix by a scalar element**. If h is an element and A is a matrix as

above $hA = Ah = C$ defined by C= (c_{ij}), where $c_{ij} = h\ a_{ij}$.

(4) **Multiplication of a matrix by a vector**. Let A be m$\times n$ as above, and x= (x_1, x_2, $\cdots$, x_n) be a 1$\times n$ row vector. Now the multiplication Ax is not defined. To see this, simply treat multiplication of a matrix by a vector as a product of two matrices above. On the other hand, if x is a column vector $n \times 1$, then Ax= c is an m $\times$ 1 column vector.

(5) **Rules for Sum and Product of more than two matrices.** (i) **Associativity.** A+$B + C = A+$ (B+C) = $(A + B)$ +C; or ABC = A(BC) = (AB)C. (ii) **Distributivity:** A(B+C) = $AB + AC$. (B+C)$A = BA + CA$. (iii) **Identity:** If $O_{m,n}$ is an m$\times n$ matrix of all zeroes, then A+$O_{m,n} = A$

If I_n is an n$\times n$ matrix (δ_{ij}), where $\delta_{ij} = 1$ when i= j and $\delta_{ij} = 0$ otherwise, it is called the identity matrix. Verify that the nonzero elements of I are only along it main diagonal, hence it is a diagonal matrix. A$I_n = A$. Note that cancellation and simplification of matrices in complicated matrix expressions is usually obtained by making some expressions equal to the identity matrix.

(6) **Transpose** of a matrix is usually denoted by a prime, and is obtained by interchanging rows and columns. A$' = (a_{ji})$. The transpose of a transpose gives the original matrix. If a matrix equals its transpose, it is called a **symmetric** matrix. **Exercise 1:** Find the transpose of A in (1) and verify that the product A$'B$ where B is from (2) is well defined. The transpose A$'$ has 3 columns matching the 3 rows of B, giving the column-row match mentioned above.

A useful rule for the transpose of a product of two matrices AB (not necessarily square matrices) is: (AB)$' = B'A'$, where the order of multiplication reverses. **Exercise 2:** Verify that the transpose of (A$'B) = B'A$, and that when A is from (1) and B is from (2) B$'A$ is also well defined. The transpose B$'$ has 2 columns, matching the 2 rows of A, giving the necessary column-row match mentioned above.

2 Determinant of a Square Matrix and Singularity

Let the matrix A be a square matrix A= (a_{ij}) with i and j = 1,2,$\cdots$, n in this section. The determinant is a scalar number. When n= 2 we have a simple calculation of the determinant as follows:

$$\text{A} = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{22}, \tag{1}$$

where one multiplies the elements along the diagonal (going from top left corner to bottom right corner) and subtracts a similar multiplication of diagonals going from bottom to top.

When n= 3 we need to expand it in terms of three 2×2 determinants and the first row as follows:

$$|A| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}|A_{11}| - a_{12}|A_{12}| + a_{13}|A_{13}|, \tag{2}$$

where the upper case A with subscripts is a new notation. A_{ij} denotes a submatrix formed by erasing i-th row and j-th column from the original A. For example,

$A_{13} = \begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$. This is also called a minor. $|A_{ij}|$ in (2) denotes the determinant of the 2×2 matrix.

Exercise: Verify the expansion of $|A|$ of (2) upon substituting for determinants of minors: $|A| = a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32}c + a_{12}a_{23}a_{31} - a_{12}a_{21}a_{33} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}$.

An easy way to compute a 3×3 determinant is to carefully write its numbers as a 3×5 grid with the two extra columns created by simply repeating the first two columns. Note that this expanded matrix has three diagonals top to bottom and three diagonals bottom to top. Now by analogy with the 2×2 matrix multiply out and add the elements along the three diagonals (going from top left corner to bottom right corner) and subtract a similar multiplication of three diagonals going from bottom to top. Your exercise is to verify that this 'easy' method yields exactly the same six terms as the ones listed above. For example,

$$det \begin{vmatrix} a & d & g \\ b & e & h \\ c & f & j \end{vmatrix}$$

$$\begin{pmatrix} a & d & g & a & d \\ b & e & h & b & e \\ c & f & j & c & f \end{pmatrix}$$

gives the solution as $aej + dhc + gbf - ceg - fha - jbd$.

Cofactor: Signed determinants of minors A_{ij} are called cofactors. Denote

$$c_{ij} = (-1)^{i+j}|A_{ij}| \tag{3}$$

In general, for any n one can expand the determinant of A, denoted by $|A|$ or det(A) in terms of cofactors as

det(A)= $\sum\limits_{j=1}^{n} a_{ij}c_{ij}$ for any one row from out of i= $1, 2, \cdots n$.

$$= \sum_{i=1}^{n} a_{ij}c_{ij} \text{ for any one column from out of j} = 1, 2, \cdots n. \tag{4}$$

An alternative definition of a determinant is

$$\text{det(A)} = \Sigma_\sigma sgn(\sigma)\Pi_{i=1}^{n}\ a_{i,\sigma(i)} \tag{5}$$

where σ runs over all permutations of integers 1,2, $\cdots$, n and the function sgn(σ) is sometimes called signature function. Here it is 1 or -1 depending on whether a permutation is odd or even. For example, if n= 2, the even permutation of { 1,2} is { 1,2} , whereby $\sigma(1) = 1$ and $\sigma(2) = 2$. Now the odd permutation of { 1,2} is { 2,1} having $\sigma(1) = 2$ and $\sigma(2) = 1$. **Exercise:** Verify that the determinant of a simple 2×2 matrix A defined by (5) remains $a_{11}a_{22} - a_{12}a_{21}$.

2.1 Properties of Determinants

Following **properties of determinants** are useful, and may be verified as an exercise:
1. det(A)= $det(A')$, where A$'$ is the transpose of A.
2. det(AB)= $det(A)det(B)$ for the determinant of a product of two matrices. Unfortunately, for the determinant of a sum of two matrices there is no simple equality or inequality: det(A+$B) \neq det(A)$ + det(B).

3. If B is obtained from A by interchanging a pair of rows (or columns), then det(B)= $-det(A)$
4. If B is obtained from A by multiplying the elements of a row (or column), by constant k, then det(B)= k det(A). **Exercise:** How is this property of determinants different from matrices? If A is a matrix kA means multiply each element of A by k, but in computing k det(A) only one row or column is multiplied by k. Show that this follows from the expansions in (4). Also, show that det(kA)= $k^n det(A)$, whence det$(-A) = (-1)^n det(A)$, by choosing k= -1.
5. If B is obtained from A by multiplying the elements of i-th row (or column), of A by constant k, and adding the result of j-th row of A then det(B)= det(A).
6. Determinant of a diagonal matrix is simply the product of diagonal entries.
7. Determinant is a product of eigenvalues. det(A)= $\Pi_{i=1}^{n} \lambda_i(A)$, where the eigenvalues are defined in Section 4 below.

Zero Determinant and Singularity.
When is the det(A)= 0 ? (1) If two rows of A are identical (2)If two columns of A are identical, (3)If a row or column has all zeros. (4) One eigenvalue is zero.
Nonsingularity: Matrix A is nonsingular if det(A)$\neq$ 0.

3 The rank and trace of a matrix

An T$\times p$ matrix X is said to be of rank p if the dimension of the largest nonsingular square submatrix is p. (Recall that nonsingular means nonzero determinant, and X is often the matrix of regressors with T observations). The idea of the rank is related to linear independence as follow. We have

$$\text{rank(A)} = min[\text{ Row rank(A), Column rank(A)}], \tag{1}$$

where **Row rank** is the largest number of linearly independent rows, and where
Column rank is the largest number of linearly independent columns. In the following example,

$$\text{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} = \begin{bmatrix} 15 & 30 & 53 \\ 25 & 50 & -2 \end{bmatrix} \tag{1b}$$

Note that $a_{12} = 2a_{11}$ and $a_{22} = 2a_{21}$. Hence the first two columns are linearly dependent, and there are only two linearly independent columns. The column rank is only 2. Recall that a set of vectors a_1, $a_2 \cdots, a_n$ is linearly dependent if a set of scalars c_i exists which are not all zero and which satisfy $\sum_{i=1}^{n} c_i a_i = 0$. In (1) we can let $a_1 = [\,{}^{15}_{25}]$ and $a_2 = [\,{}^{30}_{50}]$ and choose $c_1 = 2$ and $c_2 = -1$ to verify that the summation is indeed zero. **Exercise:** Show that for a square nonsingular matrix, the rank= *row* rank = *column* rank.

3.1 Properties of the Rank of a Matrix

Following **properties of rank** are useful:
1) rank(X) $\leq$ min(T,p), which says that the rank of a matrix is no greater than the smaller of the two dimensions− of rows (T) and columns(p). In Statistics and Econometrics texts one often encounters the expression that "the matrix X of regressors is assumed to be of **full** (column) rank." Since the number of observations in a regression problem should exceed the number of variables, T> p should hold, which means min(T,p)= p, whence rank(X) $\leq$ p. If rank(X)= p, the rank is the largest it can be, and hence we say that X is of full rank.
2) rank(A)= $rank(A')$.
3) rank(A+ B)$\leq$ rank(A) + rank(B) .
4) Sylvester's Law: If A is m$\times n$ and B is n$\times q$ defined over a field
rank(A)+$rank(B)$ $-n$ $\leq$ rank(AB) $\leq$ min [rank(A), rank(B)]
5) If B is nonsingular rank(AB)= rank(BA) $= rank(A)$
6) If A is a matrix of real numbers, rank(A)= $rank(A'A) = rank(A'A)$.
Assuming the matrix products AB, BC and ABC are well defined, we have
rank(AB)+rank(BC) $\leq$ rank(B)+rank(ABC).
7) The rank equals the number of nonzero eigenvalues(defined in Section 5) of a matrix. With the advent of computer programs for eigenvalue computation, it is sometimes easier to determine the rank by merely counting the number of nonzero eigenvalues.

3.2 Trace of a Matrix

Trace is simply the summation of the diagonal elements. It is also the sum of eigenvalues.
Tr(A)= $\sum\limits_{i=1}^{n} a_{ii} = \sum\limits_{i=1}^{n} \lambda_i(A)$

Also note that Tr(A+B) $= Tr(A)$ + Tr(B), and Tr(AB)= $Tr(BA)$. It is customary to use the latter result to simplify matrix expressions.

4 Matrix Inverse, Partitioned Matrices and their Inverse

Let A_{ij} denote an (n–1)$\times$(n-1) submatrix (called a minor) obtained by deleting i-th row and j-th column of an n$\times n$ matrix A= (a_{ij}). The cofactor of A is defined in equation (3) of Section 2 above as C= (c_{ij}), where i,j the element is $c_{ij} = (-1)^{i+j} det(A_{ij})$. The adjoint of

A is defined as Adj(A)= $C' = (c_{ji})$, which involves interchanging the rows and columns of th matrix of cofactors$-$*indicated* by the subscript ji instead of the usual ij. The operation of interchange is the transpose. (See item 5 of Section 1).

4.1 Matrix Inverse

Matrix inverse is defined for only square matrices and denoted by a superscript -1. If A= (a_{ij}) with i,j= $1, 2, \cdots, n$ then its inverse

$$A^{-1} = (a^{ij}) = \text{Adj(A)}/\text{det(A)} = (-1)^{i+j} det(A_{ji})/det(A) \tag{1}$$

where the element at the i,j location of A^{-1} is denoted by the superscript a^{ij} , and where A_{ji} is the submatrix of A obtained by eliminating j-th row and i-th column. Since the denominator of the inverse has its determinant, the inverse of a matrix is not defined unless the matrix is nonsingular, i.e., has nonzero determinant. Otherwise, one has the problem of dividing by a zero.

Inverse of a product of two matrices: $(\text{AB})^{-1} = B^{-1}A^{-1}$, where the order is reversed, which is similar to the transpose of a product of two or more matrices. However, do not apply a similar rule to the inverse of a sum. In general, $(\text{A}+B)^{-1} \neq A^{-1} + B^{-1}$. In fact even if A^{-1} and B^{-1} exist their sum may not even have an inverse. For example, if B= $-A$ the sum A$+B$ is the null matrix, which does not have an inverse, since one cannot divide by a zero.

4.2 Solution of a set of Linear Equations S$\beta = y$

Let S be a 2×2 matrix defined below and let $\beta = (\beta_1, \ \beta_2)'$, be the vector of unknown quantities to be computed (solved for) and let y= $(19, 14)'$ be the right hand known column vector of constants. Upon solving the system of two equations, verify that the following two equations have $\beta_1 = 2$ and $\beta_2 = 3$ as the solutions written as $\beta = S^{-1}y$.

$$5 \ \beta_1 + 3 \ \beta_2 = 19 \tag{2}$$

$$4 \ \beta_1 + 2 \ \beta_2 = 14 \tag{3}$$

Note that since $\beta = (2, 3)'$ is a solution.

$$\text{S}= \begin{bmatrix} 5 & 3 \\ 4 & 2 \end{bmatrix}, \text{Cofactor(S)}= \begin{bmatrix} 2 & -4 \\ -3 & 5 \end{bmatrix}, \text{det(S)}= 10 - 12 = -2$$

$$\text{Adj(S)}= \begin{bmatrix} 2 & -3 \\ -4 & 5 \end{bmatrix}, S^{-1} = \begin{bmatrix} -1 & 1.5 \\ 2 & -2.5 \end{bmatrix}, S^{-1}y = \begin{bmatrix} -1 & 1.5 \\ 2 & -2.5 \end{bmatrix} \left[\begin{smallmatrix} 19 \\ 14 \end{smallmatrix}\right] = \left[\begin{smallmatrix} -19+21 \\ 38-35 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 2 \\ 3 \end{smallmatrix}\right]$$

4.3 Algebra of Partitioned Matrices

A= $\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$, B= $\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$, and C= $\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$, are all m$\times n$ matrices, and where

the submatrices have appropriate and comparable dimensions: A_{ij} is $m_i \times n_j$ for i and j= $1, 2$. Now m= $m_1 + m_2$, n= $n_1 + n_2$. Such partitions of A, B and C are called conformable. Now we have the sum defined by

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} + \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} A_{11} + B_{11} & A_{12} + B_{12} \\ A_{21} + B_{21} & A_{22} + B_{22} \end{bmatrix} \tag{4}$$

The product is defined by:

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix}. \tag{5}$$

The submatrices are called blocks and if $A_{12} = 0$ (the null matrix), A is called a lower block triangular matrix. If $A_{21} = 0$ it is called upper block triangular. If both $A_{12} = 0$ and $A_{21} = 0$ it is called a block diagonal matrix.

The determinant of a partitioned matrix is not intuitive, and needs to be learned. There are two answers for the above A:

det(A)= $det(A_{11})det(A_{22} - A_{21}A_{11}^{-1}\ A_{12}) = det(A_{22})det(A_{11} - A_{12}A_{22}^{-1}\ A_{21}).$ (6)

As an aid to memory, think of the submatrices as scalars, i.e., 1$\times$1 matrices. Now the det(A)= $A_{11}A_{22} - A_{21}A_{12} = A_{11}(\ A_{22} - [A_{21}A_{12}]/A_{11}\ )$, and the above formula (6) merely replaces the A_{11} in the denominator with A_{11}^{-1} conformably placed in the middle to make sure that the matrix multiplication is well defined.

Exercise: Verify that if A is block diagonal, det(A)= $det(A_{11})det(A_{22})$, and furthermore that the same relation holds for upper or lower triangular A.

4.4 Computation of Elements of the Inverse of a Partitioned Matrix

The inverse B of a square partitioned matrix A is also not intuitive. It can be evaluated from the basic definition of any inverse:

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$$

which assumes that the matrix B= A^{-1}. The last equation leads to four matrix equations where the first is I= $A_{11}B_{11} + A_{12}B_{21}$, based on equating the expressions in the 1,1 position of the last two matrices. Now these four equations can be solved for the four elements of B are respectively: $B_{11} = (A_{11} - A_{12}A_{22}^{-1}\ A_{21})^{-1}$, $B_{12} = -A_{11}^{-1}A_{12}(A_{22} - A_{21}A_{11}^{-1}\ A_{12})$,
$B_{21} = -A_{22}^{-1}A_{21}(A_{11} - A_{12}A_{22}^{-1}\ A_{21})^{-1}$, and $B_{22} = (A_{22} - A_{21}A_{11}^{-1}\ A_{12})$.

Thus the partitioned matrix elements A_{ij} can be used directly in these formulas to obtain the elements B_{ij} of its inverse.

5 Characteristic Polynomial, Eigenvalues and Eigenvectors.

There are many notions in Econometrics, such as multicollinearity which are better understood with the help of characteristic polynomials, eigenvalues and eigenvectors. Given a matrix A, it is interesting to note that one can define a characteristic polynomial in λ as:

$$f(\lambda) = det(A - \lambda I), \tag{1}$$

where A$-\lambda I$ is sometimes called the characteristic matrix of A. How does a determinant become a polynomial? The best way to see this is with an example of a symmetric 2×2 matrix

$$\text{A}-\lambda I = \begin{bmatrix} 1 & 0.7 \\ 0.7 & 1 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1-\lambda & 0.7 \\ 0.7 & 1-\lambda \end{bmatrix} \tag{2}$$

Now the determinant det(A$-\lambda I$) $= (1-\lambda)^2 - 0.7^2$, which is seen to be a quadratic (polynomial) in λ. Note that the highest power of λ is 2 which is also the dimension of A. For n dimensional matrix A we would have an n-th degree polynomial in λ.

5.1 Eigenvalues

The eigenvalues are also called characteristic roots, proper values, latent roots, etc. These have fundamental importance in understanding the properties of a matrix. For arbitrary square matrices, eigenvalues are complex roots of the characteristic polynomial. After all, not all polynomials have real roots. For example, the polynomial $\lambda^2 + 1 = 0$ has roots $\lambda_1 =$

i and $\lambda_2 = -i$ where i denotes $\sqrt{-1}$. The eigenvalues are denoted by λ_i where i$= 1, 2, \cdots$, n are in non-increasing order of absolute values. For complex numbers, the absolute value is defined as the square root of the product of the number and its complex conjugate. The complex conjugate of i is $-i$, their product is 1, with square root also 1. Thus $|i| = 1$ and $|-i| = 1$, and the polynomial $\lambda^2 + 1 = (\lambda - \lambda_1)(\lambda - \lambda_2)$ holds true. As in the above example, the roots of a polynomial need not be distinct.

By the fundamental theorem of algebra, an n-th degree polynomial defined over the field of complex numbers has n roots. Hence we have to count each eigenvalue with its proper multiplicity when we write $|\lambda_1(A)| \geq |\lambda_2(A)| \geq , \cdots , \geq |\lambda_{n-1}(A)| \geq |\lambda_n(A)|$. If eigenvalues are all real, we distinguish between positive and negative eigenvalues when ordering them by

$$\lambda_1(A) \geq \lambda_2(A) \geq , \cdots , \geq \lambda_{n-1}(A) \geq \lambda_n(A). \tag{3}$$

For the cases of interest in statistical applications we usually consider eigenvalues of only those matrices which can be proved to have only real eigenvalues. Exercise: Find the eigenvalues of a symmetric matrix and verify that they are real numbers.

5.2 Eigenvectors

Eigenvector z of an n$\times n$ matrix A is of dimension n$\times$ 1. These are defined by the relationship: Az$= \lambda z$, which can be written as Az$= \lambda I z$. Now moving Iz to the left hand side, we have (A$-\lambda I)z = 0$. In the above example where n$= 2$ this relation is a system of two equations when the matrices are explicitly written out. For the 2$\times$2 symmetric matrix above the two equations are:

$$(1 - \lambda)\ z_1 + 0.7\ z_2 = 0 \tag{4}$$

$$0.7\ z_1 + (1 - \lambda) z_2 = 0 \tag{5}$$

Note that a so-called trivial solution is to choose z$= 0$, that is both elements of z are zero. We ignore this valid solution because it is not interesting or useful. This system of n equations is called degenerate, and it has no solution unless additional restrictions are imposed. The vector z has two elements z_1 and z_2 to be solved from two equations, then why is it that it is degenerate ? The problem is that both equations yield the same solution. The two equations are not linearly independent. Exercise: Impose the additional restriction that the z vector must lie on the unit circle, that is $z_1^2 + z_2^2 = 1$. Now show that there are two ways of solving the system of two equations. The choice $\lambda_1 = 1{+}0.7$ and $\lambda_2 = 1{-}0.7$ yields the two solutions. Associated with each of the two solutions there are two eigenvectors which satisfy the defining equations Az$= \lambda_1 I z$ and Az$= \lambda_2 I z$.

The characteristic equation (1) for the above example (n= 2) is
det(A$-\lambda I$) $= 0 = \Pi_{i=1}^{n}(\lambda - \lambda_i)$. (6)

A remarkable result known as **Cayley-Hamilton** theorem states that the matrix A satisfies its own characteristic equation in the sense that if we replace λ by A, we still have

$$0= \Pi_{i=1}^{n}(A - \lambda_i I) = (A - 1.7\text{ I})(\text{A}-0.3\text{ I})= \begin{bmatrix} -0.7 & 0.7 \\ 0.7 & -0.7 \end{bmatrix} \begin{bmatrix} 0.7 & 0.7 \\ 0.7 & 0.7 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \qquad (7)$$

6 Orthogonal Vectors and Matrices

Geometrically, orthogonal means perpendicular. Two nonnull vectors a and b are perpendicular to each other if the angle θ between them is 90 degrees or $\pi/2$ radians. Geometrically, the vector a is represented as line from the origin to the point A (say) whose coordinates are a_1 to a_n in an n-dimensional Euclidian space. Similarly the vector b is represented by a line from the origin to a point B with coordinates b_1 to b_n. From analytic Geometry, the line joining the points A and B is represented by $|a - b|$, and we know the following fact about the angle θ.
$\cos\,\theta = \frac{|a|^2+|b|^2-|a-b|^2}{2|a||b|}$, (1)

where $|a|$ is the Euclidian length of the vector a, that is $|a| = \sqrt{(}\Sigma_{i=1}^{n}\; a_i^2)$, and similarly for b. The length of the line joining the points A and B is
$|a - b| = \sqrt{(}\Sigma_{i=1}^{n}\; [a_i - b_i]^2) = \sqrt{(}\Sigma_{i=1}^{n}\; [a_i^2 + b_i^2 - 2\; a_i b_i])$. (2)

The vectors are perpendicular or orthogonal if $\theta = \pi/2$, and $\cos\,\theta = 0$. Thus the left side of the above equation for $\cos\,\theta$ is zero if the numerator is zero, since the denominator is definitely nonzero. Thus we require that $|a|^2 +\; |b|^2 = |a-b|^2$, which amounts to requiring the cross-product term to be zero, or $-\Sigma_{i=1}^{n} 2a_i b_i = 0$, that is $\Sigma_{i=1}^{n} a_i b_i = 0 = a'b$, using the prime to denote the transpose of the column vector a. The reader should clearly understand the representation of summation $\Sigma_{i=1}^{n} a_i b_i$ by the simpler a$'b$. In particular if b$\equiv a$ one has the squared Euclidian length $|a|^2 = a'a$ is the sum of squares of the elements of the vector a. If the Euclidian length of both a and b is unity, and a$'b = 0$, that is the vectors are orthogonal to each other they are called orthonormal vectors. For example let the two vectors defining the usual two dimensional axes be $e_1 = (1,0)$ and $e_2 = (0,1)$ respectively. Now $e_1' e_2 = (0+0) = 0$, that is, $e_1 and\; e_2$ are orthogonal. Since $|e_1|^2 = e_1' e_1 = 1$ and similarly $|e_2| = 1$ means they are orthonormal also.

A remarkable result called **Gram-Schmidt orthogonalization** is a process which assures that any given m linearly independent vectors can be transformed into a set of m orthonormal vectors by a set of linear equations.

6.1 Orthogonal Matrices

The distinction between orthogonal and orthonormal is usually not made for matrices. A matrix A is said to be orthogonal if its transpose equals its inverse. $A' = A^{-1}$, which also means that $A'A = I$.

The determinant of orthogonal matrix is either $+1$ or -1, and its eigenvalues are also $+1$ or -1. To make these results plausible, recall that det(AB)= $det(A)det(B)$. Hence applying this to the relation $A'A = I$ for orthogonal A we have det(A)det(A$'$) $= det(I)$. Now det(I)= 1 is obvious.

7 Idempotent Matrices

In ordinary algebra the familiar numbers whose square is itself are: 1 or 0. Any integer power of 1 is 1. In matrix algebra, the Identity matrix I plays the role of the number 1, and I certainly has the property $I^n = I$. There are other nonnull matrices which are not identity and yet have this property, namely $A^2 = A$. Hence a new name "idempotent" is needed to describe them. For example, the hat matrix is defined as $\hat{H} = X(X'X)^{-1}X'$. Writing $\widehat{H}^2 = \hat{H}\hat{H} = X(X'X)^{-1}X'X(X'X)^{-1}X' = X(X'X)^{-1}X' = \hat{H}$ means that $\hat{H}$ matrix is idempotent. Verify that $\widehat{H}^n = \hat{H}$ for any integer power n. If eigenvalues are $\lambda_i(\hat{H})$, the relation $\widehat{H}^n = \hat{H}$ means that $\lambda_i(\widehat{H}^n) = \lambda_i(\hat{H}) = [\lambda_i(\hat{H})\ ]^n$. Since the only numbers whose n-th power is itself are unity and zero, it is plausible that the eigenvalues of an idempotent matrix are 1 or 0.

Recall that the rank of a matrix equals the number of nonzero eigenvalues and that the trace of matrix is the sum of eigenvalues. Using these two results it is clear that for idempotent matrices G whose eigenvalues are 0 or 1 we have the interesting relation trace(G)= $rank(G)$.

8 Quadratic and Bilinear Forms

If x is an n$\times$1 vector and A is an n$\times n$ matrix the expression $x'Ax$ is of dimension 1$\times$1 or a scalar. A scalar does not mean that it is just one term in an expression, only that when it is evaluated it is a number. For example, when n= 2, and A= (a_{ij}) note that:

$$x' \text{ A x} = (x_1, x_2)\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = a_{11}x_1^2 + a_{12}x_1x_2 + a_{21}x_2x_1 + a_{22}x_2^2 \tag{1}$$

is a quadratic expression. Hence x$'Ax$ is called a quadratic form, which is a scalar and has an expression having 4 terms. Note that the largest power of any element of x is 2 even for the case when A is a 3×3 matrix. When the a_{ij} and x_i are replaced by numbers, it is clear that the quadratic form is just a number, that is called a scalar to be distinguished from a matrix. A scalar cannot be a matrix, but a matrix can be a scalar.

8.1 Positive Negative and Other Definite Quadratic Forms

Just as a number can be negative, zero or positive, a quadratic form also can be negative, zero or positive. Since the sign of zero can be negative or positive, it is customary in matrix algebra to say that a quadratic form is negative or positive definite, meaning that x$'Ax < 0$ and x$'Ax > 0$ respectively. The expression "definite" reminds us that it is not zero. If x$'Ax$ ≥ 0 it is called non-negative definite (nnd); and if $x'Ax \leq 0$ it is called non-positive definite (npd). Remember that the matrix A must be a square matrix for it to be the matrix of a quadratic form, but it **need not be a symmetric matrix** (i.e. A$' = A$ is unnecessary). If one has a symmetric matrix, the cross product terms can be merged. For example, the symmetry of A means that $a_{12} = a_{21}$ in the above 2×2 illustration:

$a_{11}x_1^2 + a_{12}x_1x_2 + a_{21}x_2x_1 + a_{22}x_2^2 = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2$.

If one has an asymmetric matrix A and still wishes to simplify the quadratic form this way, one can redefine the matrix of the quadratic form as B= $(1/2)(A + A')$ and use x$'Bx$ as the quadratic form, which can be proved to be equal to x$'Ax$.

How to determine whether a given quadratic form is positive definite or non negative definite?

A simple practical method of determining whether a given quadratic form A is positive definite in the modern era of computers is to find its eigenvalues, and concentrate on the smallest eigenvalue. If the smallest eigenvalue $\lambda_{min}(A) > 0$, is strictly positive, the quadratic form x$'Ax$ is said to be positive definite. Similarly, if the smallest eigenvalue $\lambda_{min}(A) \geq 0$, which can be zero, the quadratic form x$'Ax$ is said to be non-negative definite.

In terms of determinants, there is a sequence of determinants which should be checked to be positive definite for the quadratic form to be positive definite. We check:

$$a_{11} > 0, \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0 \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} > 0, \cdots, \tag{3}$$

where we have indicated the principal determinants of orders $= 1, 2, 3$. To see the intuitive reason for these relations consider n= 2 and the quadratic form

$$\text{Q} = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 \tag{4}$$

Complete the square by adding and subtracting $(a_{12}/a_{11})^2x_2^2$. Now write

$$\text{Q} = a_{11}[x_1 + (a_{12}/a_{11})x_2]^2 + [a_{22} - (a_{12}^2/a_{11})]x_2^2. \tag{5}$$

Observe that Q$>$ 0 requires that $a_{11} > 0$ from the first term, and $a_{11}a_{22} - a_{12}^2 > 0$ from the second term. The second term is positive if and only if the 2×2 determinant above is positive.

8.2 Bilinear Form

In the above discussion of quadratic forms, if we define y as an n×1 vector, x$'Ay$ is called a bilinear form. This is also a scalar.

Exercise: Let e_k denote a column vector of all zeros except for the k-th element, which is unity. The individual elements a_{ij} of the matrix A are obviously scalar. Show that $a_{ij} = e_i'Ae_j$, a bilinear form.

9 Further Study of Eigenvalues and Eigenvectors

There are two fundamental results in the analysis of real symmetric matrices.

(1) **The eigenvalues of a real symmetric matrix are real.** This is proved by using the fact that if the eigenvalues were complex numbers, they must come in conjugate pairs. Ax$= \lambda x$ and A$\bar{x} = \bar{\lambda}\bar{x}$, where the bar indicates the conjugate complex number. Premultiply Ax$= \lambda x$ by the $\bar{x}'$ vector to yield $\bar{x}'Ax = \lambda\bar{x}'x$, ($\lambda$ is a scalar). Now premultiply A$\bar{x} = \bar{\lambda}\bar{x}$ by x$'$ to yield x$'A\bar{x} = \bar{\lambda}x'\bar{x}$. Now the quadratic forms on the left sides of these relations must be equal to each other by the symmetry of A, i.e., $\bar{x}'Ax = x'A\bar{x}$. Hence we have $(\lambda - \bar{\lambda})x'\bar{x} = 0$ which can only be true if $\lambda = \bar{\lambda}$, a contradiction. The complex conjugate cannot be equal to the original.

(2) **Eigenvectors associated with distinct eigenvalues are orthogonal to each other**.

The proof involves an argument similar to the one above, except that instead of $\bar{\lambda}$ we use a distinct root μ say and instead of $\bar{x}$ we use the vector y. The relation $(\lambda - \mu)x'y = 0$, now implies that $x'y = 0$ since the roots λ and μ are assumed to be distinct. Hence the two eigenvectors x and y must be orthogonal.

9.1 Reduction of a Real Symmetric Matrix to the Diagonal Form

Let A be a real symmetric matrix of order $n \times n$ having distinct eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ ordered in a sequence from the largest $\lambda_1 = max_i \ \{\lambda_i, i = 1, 2, \cdots n\}$ to the smallest $\lambda_n = min_i\{\lambda_i, i = 1, 2, \ \cdots n\}$. Denote by G the $n \times n$ orthogonal matrix of corresponding eigenvectors G= $[x_{.1}, x_{.2}, \cdots, x_{.n}]$, where the dot notation is used to remind us that the first subscript is suppressed and each $x_{.i}$ is an $n \times 1$ column vector. Orthogonality of G is verified by the property: $G'G = GG'$. Recall the fundamental relation defining i-th eigenvalue and eigenvector $Ax_{.i} = \lambda_i x_{.i}$, in terms of the dot notation. Denote by $\Lambda = diag(\lambda_1, \lambda_2, \cdots, \lambda_n)$, a diagonal matrix containing the eigenvalues in the non-increasing order. Combining all eigenvalues and eigenvectors we write $AG = G\Lambda$. (Exercise: verify that $G\Lambda \neq \Lambda G$, and that the latter will not work on the RHS). The relation $G'AG = G'G\Lambda = \Lambda$ is called eigenvalue-eigenvector decomposition of the real symmetric matrix A. For example, let

$$A = \begin{bmatrix} 1 & 0.7 \\ 0.7 & 1 \end{bmatrix}, \ G = \begin{bmatrix} w & w \\ -w & w \end{bmatrix}, \text{ where } w = 1/\sqrt{2}, \text{ and } \Lambda = \begin{bmatrix} 1.7 & 0 \\ 0 & 0.3 \end{bmatrix} \tag{1}$$

One can easily verify that $G'G = I = GG'$, and $G'AG = \Lambda$ by direct multiplication. In the dot notation, note that $x_{.1} = \begin{bmatrix} w \\ -w \end{bmatrix}$ and $x_{.2} = \begin{bmatrix} w \\ w \end{bmatrix}$ and G= $[x_{.1} \ , x_{.2}]$. Let us use (1) to verify $A = G\Lambda G'$.

$$G\Lambda = \begin{bmatrix} w & w \\ -w & w \end{bmatrix} \begin{bmatrix} 1.7 & 0 \\ 0 & 0.3 \end{bmatrix} = \begin{bmatrix} 1.7w & 0.3w \\ -1.7w & 0.3w \end{bmatrix}$$

$$G\Lambda G' = \begin{bmatrix} 1.7w^2 + 0.3w^2 & -1.7w^2 + 0.3w^2 \\ -1.7w^2 + 0.3w^2 & 1.7w^2 + 0.3w^2 \end{bmatrix} =$$

$$\begin{bmatrix} 2w^2 & -1.4w^2 \\ -1.4w^2 & 2w^2 \end{bmatrix} = \begin{bmatrix} 1 & 0.7 \\ 0.7 & 1 \end{bmatrix} = A \tag{2}$$

since $w^2 = (1/2)$ by the definition of w. This completes the verification of the eigenvalue eigenvector decomposition in the simple 2×2 case.

9.2 Simultaneous Reduction of two Matrices to a Diagonal Form:

Let G be an n×n orthogonal matrix of eigenvectors as above, which reduces the n×n symmetric matrix A to the diagonal form in the sense that G$'AG = \Lambda$. Now, AB= BA is a necessary and sufficient condition that G also reduces another n×n symmetric matrix B to a diagonal form with the eigenvalues of B denoted by ϕ_i along the diagonal of Φ, in the sense that GBG$' = \Phi$, where $\Phi = diag(\phi_i,$ i$= 1, 2, \cdots,$ n). In other words, if and only if A and B commute, there is an orthogonal matrix G with common eigenvectors which simultaneously diagonalizes both matrices.

In the above reduction we insisted on G being an orthogonal matrix, with common eigenvectors. For some purposes it is not necessary to have an orthogonal matrix, but one would like to relax the commutativity requirement. Then we have the following result: Given A which is real symmetric, nonsingular and positive definite, and B which is real symmetric, then there exists a non-singular matrix S which diagonalizes A to the identity matrix and simultaneously diagonalizes B, such that S$'AS = I$ and S$'BS = \Phi$, where $\Phi = diag(\phi_i,$ i$= 1, 2, \cdots,$ n).

10 Kronecker Product of Matrices

Kronecker product of matrices is mainly a notational device. We define the Kronecker product by the notation C= $A \otimes$ B, where A is T×n, B is m×p and C is a large matrix of dimension Tm×np with elements defined by C= (a_{ij} B), where each term a_{ij} B is actually a_{ij} times the whole m×p matrix.

For example, if A = (a_{11}, a_{12}, a_{13}) with T= 1 and n= 3, and if B= $\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$ we have

$$C= \left[a_{11} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}, a_{12} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}, a_{13} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \right]$$

$$= \begin{bmatrix} a_{11}b_{11} & a_{11}b_{12} & a_{12}b_{11} & a_{12}b_{12} & a_{13}b_{11} & a_{13}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} & a_{12}b_{21} & a_{12}b_{22} & a_{13}b_{21} & a_{13}b_{22} \end{bmatrix}.$$

In the following, let A, B, C and D be arbitrary matrices. The Kronecker product satisfies the following kinds of linearity and distributive laws.
(i) A $\otimes$ (aB) = a(A $\otimes$ B), a being a scalar;

(ii) (A + B) $\otimes$ C = A $\otimes C$ + B $\otimes$ C, A and B being of the same order.
(iii) A $\otimes$ (B+C) = A $\otimes$ B + A $\otimes$ C, B and C being of the same order;
(iv) A $\otimes$ (B $\otimes$ C) = (A $\otimes$ B) $\otimes$ C;
(v) (A $\otimes$ B)$'$ = A$'$ $\otimes$ B$'$; (The order does not reverse here as in simple matrix multiplication)
(vi) (A $\otimes$ B)(C $\otimes$ D) = AC $\otimes$ BD;
(vii) (A $\otimes$ B)$^{-1}$ = $A^{-1} \otimes B^{-1}$. Again, the order does not reverse here as in the inverse of simple matrix multiplication. Obviously we assume that A and B are square non-singular matrices.
(viii) tr(A $\otimes$ B) = (tr A)(trB), A and B being square matrices;
(ix) det(A $\otimes$ B) = (det A)^m(det B)^n, A and B being n$\times n$ and m$\times m$ matrices;

Eigenvalues of Kronecker Product equals the product of eigenvalues

If A and B are square matrices of dimension m and n respectively, then C= $A \otimes$ B is mn dimensional matrix. If the eigenvalues of A are denoted by $\lambda_i(A)$ i= $1, 2, \cdots, m$ and those of B are denoted by $\lambda_j(B)$ j= $1, 2, \cdots, n$ the mn eigenvalues of C are simply a product of the eigenvalues ($\lambda_i(A)$ $\lambda_j(B)$) with i= $1, 2, \cdots, m$ and j= $1, 2, \cdots, n$.

Eigenvectors of Kronecker Products contain stacked products of elements of corresponding individual eigenvectors

Using the above notation, the eigenvectors of C have k= $i(j) = 1, 2, \cdots mn$ elements each.

Let the t-th element of i-th eigenvector of A be denoted by α_{ti} with t= $1, 2, \cdots m$.

Also, let the s-th element of j-th eigenvector of B be denoted by β_{sj} with $s = 1, 2, \cdots n$. The elements in the k-th eigenvector of C equal

($\alpha_{1i}\beta_{sj}$ with $s = 1, 2, \cdots n; \alpha_{2i}\beta_{sj}$ with $s = 1, 2, \cdots, n; \cdots; \alpha_{mi}\beta_{sj}$ with $s = 1, 2, \cdots n$).