Notes On Matrix Algebra

James L. Powell
Department of Economics
University of California, Berkeley

Fall 2008

For the usual linear equation

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + ... + \beta_K X_{iK} + \varepsilon_i$$

and the corresponding least squares estimators and related statistics, it is extremely useful to have a notation system that allows us to avoid almost all of the subscripts, summation signs, and "..." fillers when talking about the properties of the statistics. Linear algebra, with its system of **matrix notation**, was invented for this purpose; it gives a convenient way to discuss solutions of multiple linear equations (like the "normal equations" for the least squares estimator).

Matrices and their Shapes

A **matrix** is a table of numbers or symbols representing numbers (like variables, random variables, functions, etc.). This table must be full – that is, all rows and columns must contain an element. Typically matrices are denoted by capital roman letters; the shape of a matrix is the number of rows and columns it has. So if a matrix A has K rows and L columns (with K and L both positive integers), we say A is a $(K \times L)$ matrix, and it looks like

$$A = \begin{bmatrix} a_{11} & a_{12} & ... & a_{1L} \\ a_{21} & a_{22} & ... & a_{2L} \\ ... & ... & ... & ... \\ a_{K1} & a_{k2} & ... & a_{KL} \end{bmatrix},$$

which we sometimes express as

$$A = [a_{ij}], \ i = 1, ..., K, \ j = 1, ..., L.$$

For the elements a_{ij} of A, the first subscript always represents the row number and the second the column number.

Matrix algebra treats rows and columns differently, and often it is useful to interchange the rows and columns. The **transpose** of a matrix A, denoted A' (or sometimes A^T, when the "prime" superscript is being reserved for derivatives), is defined as the matrix A with rows and columns swapped. For example, if

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix},$$

then

$$A' = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}.$$

More generally, if $A = [a_{ij}]$ for $i = 1, ..., K$ and $j = 1, ..., L$, then $A' = [a_{ji}]$ for $j = 1, ..., L$ and $i = 1, ..., J$.

A matrix that is unchanged if its rows and columns are interchanged – that is, a matrix that is the same as its transpose – is called a **symmetric** matrix. If a matrix is symmetric, i.e., $A = A'$, then it has to be **square,** meaning its number of rows and columns is equal. There are plenty of square matrices that aren't symmetric, but symmetric matrices have to be square.

Matrices with only one row or one column are called **vectors** – either *row vectors* or *column vectors*, respectively. So a "column vector of dimension K" is a $(K \times 1)$ matrix, like

$$\beta \equiv \begin{bmatrix} \beta_1 \\ \beta_2 \\ ... \\ \beta_K \end{bmatrix},$$

which is the (column) vector of unknown regression coefficients $\beta_1, ..., \beta_K$. It is my habit (not universally observed, even by me) to write all vectors as *column* vectors, and denote row vectors using a transpose – for example, for the i^{th} observation, the K-dimensional vector of explanatory variables is

$$X_i \equiv \begin{bmatrix} 1 \\ X_{i2} \\ ... \\ X_{iK} \end{bmatrix},$$

which can be expressed as a row vector as

$$X_i' = \begin{bmatrix} 1 & X_{i2} & ... & X_{iK} \end{bmatrix}.$$

In linear algebra, vectors are usually denoted by lower case Roman letters, like x or y, but we'll usually use capital Roman or Greek letters to denote vectors of random variables or parameters.

Matrix addition and multiplication

Given these definitions of vectors and matrices, **matrix algebra** is a set of rules for addition, subtraction, multiplication, and division (of a sort) of these symbolic objects. Two matrices A and B are **equal**, $A = B$, when they have the same shape (number of rows and columns) and the same elements, i.e., $a_{ij} = b_{ij}$ if $A \equiv [a_{ij}]$ and $B \equiv [b_{ij}]$. And if they are the same shape, the **sum** of A and B is defined to be the matrix C that contains the sums of the elements, i.e.

$$C = A + B$$

has elements

$$c_{ij} = a_{ij} + b_{ij},$$

with $C = [c_{ij}] = [a_{ij} + b_{ij}]$.

If α is any real number (termed a **scalar**, to distinguish it from a matrix), we can define **scalar multiplication** of α and A as the matrix which contains each element of A multiplied by α, i.e.

$$\alpha A = [\alpha \cdot a_{ij}].$$

This gives a definition of "subtraction" of two matrices of the same shape:

$$A - B = A + (-1)B = [a_{ij} - b_{ij}].$$

The key concept for matrix algebra is **matrix multiplication.** Unlike ordinary multiplication of numbers, two matrices cannot always be multiplied together, and the order of multiplication is important.

The **matrix product** $C = AB$ is only *defined* if the *number of columns of* A *equals the number of rows of* B, i.e., A is $(K \times L)$ and B is $(L \times M)$. If so, A and B are said to be "conformable", and the product $C = AB$ is a $(K \times M)$ matrix with elements

$$c_{im} = \sum_{j=1}^{L} a_{ij} b_{jm},\ i = 1, ..., K,\ m = 1, ..., M.$$

One of the simplest special cases of matrix multiplication is premultiplication of a column vector by a row vector – that is, if x and y are column vectors of the same length, then $x'y$ is a number (a (1×1) matrix). This is usually called the **inner product** or "dot product" of x and y. For example, if the column vector of unknown regression coefficients in the multiple regression model is written again as a column vector,

$$\beta \equiv \begin{bmatrix} \beta_1 \\ \beta_2 \\ ... \\ \beta_K \end{bmatrix},$$

and the corresponding row vector of regressors for the i^{th} observation is defined as above,

$$X_i' \equiv \begin{bmatrix} 1 & X_{i2} & ... & X_{iK} \end{bmatrix},$$

then the inner product of X_i and β is defined as

$$\begin{aligned} X_i'\beta &\equiv 1 \cdot \beta_1 + (X_{i2} \cdot \beta_2) + ... + (X_{iK} \cdot \beta_k) \\ &= \sum_{j=1}^{K} X_{ij}\beta_j, \end{aligned}$$

and we can write the structural equation for the multiple regression model as

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + ... + \beta_K X_{iK} + \varepsilon_i \\ &\equiv X_i'\beta + \varepsilon_i, \end{aligned}$$

which at least hides one set of subscripts (i.e., the second subscript on the regressors denoting variable number).

In general, multiplication of two matrices involves taking inner products of the rows of the first matrix with the columns of the second. That is, the element in row i and column j of the matrix $C = AB$ is the inner product of the i^{th} row of A with the j^{th} row of B, which requires the rows of A to be the same length as the columns of B.

We now have enough rules to express the multiple regression model (or at least the equation) in matrix form with *no subscripts*. Define Y and ε to be the $(N \times 1)$ (column) vectors of all the individual values of the dependent variable Y_i and error terms ε_i,

$$Y \equiv \begin{bmatrix} Y_1 \\ Y_2 \\ ... \\ Y_N \end{bmatrix},\ \varepsilon \equiv \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ ... \\ \varepsilon_N \end{bmatrix},$$

and let X be the $(N \times K)$ matrix whose columns have all the individual values of each of the explanatory variables,

$$X \equiv \begin{bmatrix} 1 & X_{12} & ... & X_{1K} \\ 1 & X_{22} & ... & X_{2K} \\ ... & ... & ... & ... \\ 1 & X_{N2} & ... & X_{NK} \end{bmatrix}.$$

Then, with the previous definition of the $(K \times 1)$ vector β of regression coefficients, the multiple regression equation

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + ... + \beta_K X_{iK} + \varepsilon_i \qquad \text{for} \qquad i = 1, ..., N$$

can be written more elegantly as

$$Y = X\beta + \varepsilon.$$

The i^{th} row of $X\beta$ has the inner product $X_i'\beta$ of regression coefficients and regressors for the i^{th} observation.

Matrix Inversion

So now we have addition, subtraction, and multiplication (actually, two kinds of multiplication) defined for matrices. What about division? In linear algebra, "division" is important for the solutions of systems of linear equations, like trying to find the values of $z_1, ..., z_K$ that solve

$$\begin{aligned} c_1 &= a_{11}z_1 + ... + a_{1K}z_{k,} \\ c_2 &= a_{21}z_1 + ... + a_{2K}z_{k,} \\ &\cdots \\ c_K &= a_{K1}z_1 + ... + a_{KK}z_{k,} \end{aligned}$$

where the $\{c_i\}$ and the $\{a_{ij}\}$ coefficients are assumed known. In matrix form, this system of equations can be written as

$$c = Az,$$

where c and z are (column) vectors of the $\{c_i\}$ and $\{z_i\}$ and the matrix A has elements $\{a_{ij}\}$. For the simplest case of $K = 1$ – for which c and A are just (1×1) matrices, i.e., numbers – the solution for this single equation would be

$$z^* = \frac{c}{A},$$

provided $A \neq 0$ (otherwise, the equation either has no solutions or infinitely many, depending upon whether $c \neq 0$).