MATRIX ALGEBRA REVIEW

(

PRELIMINARIES

A matrix is a way of organizing information.

It is a rectangular array of elements arranged in rows and columns. For example, the following matrix A has m rows and n columns.

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}$$

All elements can be identified by a typical element a_{ij}, where i=1,2,…,m denotes rows and j=1,2,…,n denotes columns.

A matrix is of order (or dimension) m by n (also denoted as (m x n)).
A matrix that has a single column is called a column vector.
A matrix that has a single row is called a row vector.

TRANSPOSE

The **transpose** of a matrix or vector is formed by interchanging the rows and the columns. A matrix of order (m x n) becomes of order (n x m) when transposed.

For example, if a (2 x 3) matrix is defined by

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$$

Then the transpose of A, denoted by A', is now (3 x 2)

$$A' = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{bmatrix}$$

- $(A')' = A$
- $(kA)' = kA'$, where k is a scalar.

SYMMETRIC MATRIX

When $A' = A$, the matrix is called **symmetric**. That is, a symmetric matrix is a square matrix, in that it has the same number of rows as it has columns, and the off-diagonal elements are symmetric (i.e. $a_{ij} = a_{ji}$ *for all i and j*).

For example,

$$A = \begin{bmatrix} 4 & 5 & -3 \\ 5 & 7 & 2 \\ -3 & 2 & 10 \end{bmatrix}$$

A special case is the **identity matrix**, which has 1's on the diagonal positions and 0's on the off-diagonal positions.

$$I = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

The identity matrix is a **diagonal matrix**, which can be denoted by $diag(a_1, a_2, ..., a_n)$, where a_i is the i^{th} element on the diagonal position and zeros occur elsewhere. So, we can write the identity matrix as $I = diag(1,1,...,1)$.

ADDITION AND SUBTRACTION

Matrices can be added and subtracted as long as they are of the same dimension. The addition of matrix A and matrix B is the addition of the corresponding elements of A and B. So, $C = A + B$ implies that $c_{ij} = a_{ij} + b_{ij}$ for all i and j.

For example, if

$$A = \begin{bmatrix} 2 & -3 \\ 6 & 10 \end{bmatrix} \qquad B = \begin{bmatrix} 0 & 6 \\ 5 & -8 \end{bmatrix}$$

Then

$$C = \begin{bmatrix} 2 & 3 \\ 11 & 2 \end{bmatrix}$$

- $A \pm B = B \pm A$
- $(A \pm B) \pm C = A \pm (B \pm C)$
- $(A \pm B)' = A' \pm B'$

MULTIPLICATION

If k is a scalar and A is a matrix, then the product of k times A is called scalar multiplication. The product is k times each element of A. That is, if $B = kA$, then $b_{ij} = ka_{ij}$ for all i and j.

In the case of multiplying two matrices, such as $C = AB$, where neither A nor B are scalars, it must be the case that

the number of columns of A = the number of rows of B

So, if A is of dimension (m x p) and B of dimension (p x n), then the product, C, will be of order (m x n) whose ij^{th} element is defined as

$$c_{ij} = \sum_{k=1}^{p} a_{ik} b_{kj}$$

In words, the ij^{th} element of the product matrix is found by multiplying the elements of the i^{th} row of A, the first matrix, by the corresponding elements of the j^{th} column of B, the second matrix, and summing the resulting product. For this to hold, the number of columns in the first matrix must equal the number of rows in the second.

For example,

$$F = AD = \begin{bmatrix} 6 & 8 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} 3 & -8 & 1 \\ 9 & 2 & 5 \end{bmatrix}$$

$$= \begin{bmatrix} 6*3+8*9 & 6*(-8)+8*2 & 6*1+8*5 \\ (-2)*3+4*9 & (-2)*(-8)+4*2 & (-2)*1+4*5 \end{bmatrix}$$

$$= \begin{bmatrix} 90 & -32 & 46 \\ 30 & 24 & 18 \end{bmatrix}$$

- A (m x 1) column vector multiplied by a (1 x n) row vector becomes an (m x n) matrix.
- A (1 x m) row vector multiplied by a (m x 1) column vector becomes a scalar.
- In general, $AB \neq BA$.
- But, $kA = Ak$ if k is a scalar and A is a matrix.
- And, $AI = IA$ if A is a matrix and I is the identity matrix and conformable for multiplication.

The product of a row vector and a column vector of the same dimension is called the **inner product** (also called the dot product), its value is the sum of products of the components of the vectors. For example, if j is a (T x 1) vector with elements 1, then the inner product, j'j, is equal to a constant T.

Note: two vectors are **orthogonal** if their inner product is zero.

- $A(B+C) = AB + AC$.
- $(A+B)C = AC + BC$.

- $A(BC) = (AB)C$.

A matrix with elements all zero is called a **null matrix**.

- $(AB)' = B'A'$.
- $(ABC)' = C'B'A'$.

<u>TRACE OF A SQUARE MATRIX</u>

The **trace of a square matrix** A, denoted by tr(A), is defined to be the sum of its diagonal elements.

$$tr(A) = a_{11} + a_{22} + a_{33} + ... + a_{nn}$$

- $tr(A) = A$, if A is a scalar.
- $tr(A') = tr(A)$, because A is square.
- $tr(kA) = k \cdot tr(A)$, where k is a scalar.
- $tr(I_n) = n$, the trace of an identity matrix is its dimension.
- $tr(A \pm B) = tr(A) \pm tr(B)$.
- $tr(AB) = tr(BA)$.
- $tr(AA') = tr(A'A) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2$.

<u>DETERMINANT OF A SQUARE MATRIX</u>

The **determinant of a square matrix** A, denoted by det(A) or $|A|$, is a uniquely defined scalar number associated with the matrix.

i) for a single element matrix (a scalar, $A = a_{11}$), det(A) = a_{11}.

ii) in the (2 x 2) case,

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

the determinant is defined to be the difference of two terms as follows,

$$|A| = a_{11}a_{22} - a_{12}a_{21}$$

which is obtained by multiplying the two elements in the principal diagonal of A and then subtracting the product of the two off-diagonal elements.

iii) in the (3 x 3) case,

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

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

iv) for general cases, we start first by defining the **minor** of element a_{ij} as the determinant of the submatrix of A that arises when the i[th] row and the j[th] column are deleted and is usually denoted as $|A_{ij}|$. The **cofactor** of the element a_{ij} is $c_{ij} = (-1)^{i+j}|A_{ij}|$. Finally, the determinant of an n x n matrix can be defined as

$$|A| = \sum_{j=1}^{n} a_{ij}c_{ij} \quad \textit{for any row } i = 1,2,...,n.$$

$$= \sum_{i=1}^{n} a_{ij}c_{ij} \quad \textit{for any column } j = 1,2,...,n.$$

- $|A'| = |A|$
- $\begin{vmatrix} a & kc \\ b & kd \end{vmatrix} = \begin{vmatrix} ka & c \\ kb & d \end{vmatrix} = k\begin{vmatrix} a & c \\ b & d \end{vmatrix}$
- $|kA| = k^n|A|$, for scalar k and n x n matrix A.
- If any row (or column) of a matrix is a multiple of any other row (or column) then the determinant is zero, e.g.
  $\begin{vmatrix} a & ka \\ b & kb \end{vmatrix} = k\begin{vmatrix} a & a \\ b & b \end{vmatrix} = k(ab - ab) = 0$
- If A is a diagonal matrix of order n, then $|A| = a_{11}a_{22}\cdots a_{nn}$
- If A and B are square matrices of the same order, then $|AB| = |A||B|$.
- In general, $|A + B| \neq |A| + |B|$

RANK OF A MATRIX AND LINEAR DEPENDENCY

Rank and linear dependency are key concepts for econometrics. The rank of any (m x n) matrix can be defined (i.e., the matrix does not need to be square, as was the case for the determinant and trace) and is inherently linked to the invertibility of the matrix.

The **rank** of a matrix A is equal to the dimension of the largest square submatrix of A that has a nonzero determinant. A matrix is said to be of **rank r** if and only if it has at least one submatrix of order r with a nonzero determinant but has no submatrices of order greater than r with nonzero determinants.

For example, the matrix

$$A = \begin{bmatrix} 4 & 5 & 2 & 14 \\ 3 & 9 & 6 & 21 \\ 8 & 10 & 7 & 28 \\ 1 & 2 & 9 & 5 \end{bmatrix}$$

has rank 3 because $|A| = 0$, but $\begin{vmatrix} 4 & 5 & 2 \\ 3 & 9 & 6 \\ 8 & 10 & 7 \end{vmatrix} = 63 \neq 0$

That is, the largest submatrix of A whose determinant is not zero is of order 3.

The concept of rank also can be viewed in terms of linear dependency. A set of vectors is said to be **linearly dependent** if there is a nontrivial combination (i.e., at least one coefficient in the combination must be nonzero) of the vectors that is equal to the zero vector. More precisely, denote n columns of the matrix A as $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$. This set of these vectors is **linearly dependent** if and only if there exists a set of scalars $\{c_1, c_2, \ldots, c_n\}$, at least one of which is not zero, such that $c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \ldots + c_n\mathbf{a}_n = \mathbf{0}$.

In the above example, the columns of the matrix A are linearly dependent because,

$$1\begin{bmatrix} 4 \\ 3 \\ 8 \\ 1 \end{bmatrix} + 2\begin{bmatrix} 5 \\ 9 \\ 10 \\ 2 \end{bmatrix} + 0\begin{bmatrix} 2 \\ 6 \\ 7 \\ 9 \end{bmatrix} - 1\begin{bmatrix} 14 \\ 21 \\ 28 \\ 5 \end{bmatrix} = \mathbf{0}$$

If a set of vectors is not linearly dependent, then it is **linearly independent**. Also, any subset of a linearly independent set of vectors is linearly independent.

In the above example, the first three columns of A are linearly independent, as are the first two columns of A. That is, we cannot find a set of scalars (with at least one nonzero) such that the linear combination of scalars and columns equals the zero vector.

The connection between linear dependency and the rank of a matrix is as follows: **the rank of a matrix A may be defined as the maximum number of linearly independent columns or rows of A.**

In other words, the maximum number of linearly independent columns is equal to the maximum number of linearly independent rows, each being equal to the rank of the matrix. If the maximum number of linearly independent columns (or rows) is equal to the number of columns, then the matrix has **full column rank**. Additionally, if the maximum number of linearly independent rows (or columns) is equal to the number of rows, then the matrix has **full row rank**. When a square matrix A does not have full column/row rank, then its determinant is zero and the matrix is said to be **singular**. When a square matrix A has full row/column rank, its determinant is not zero, and the matrix is said to be **nonsingular** (and therefore invertible).

- Full rank (nonsingular matrix) $\Leftrightarrow$ $|\mathrm{A}| \neq 0$ $\Leftrightarrow$ A is invertible.

Furthermore, the maximum number of linearly independent (m x 1) vectors is m. For example, consider the following set of two linearly independent vectors,

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

If there is a third vector,

$$b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

where b_1 *and* b_2 can be any numbers, then the three unknown scalars $c_1, c_2,$ *and* c_3 can always be found by solving the following set of equations,

$$c_1\begin{bmatrix} 1 \\ 2 \end{bmatrix} + c_2\begin{bmatrix} 3 \\ 4 \end{bmatrix} + c_3\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

In other words, the addition of *any* third vector will result in a (2 x 3) matrix that is not of full rank and therefore not invertible.

Generally speaking, this is because any set of m linearly independent (m x 1) vectors are said to **span** m-dimensional space. This means, by definition, that any (m x 1) vector can be represented as a linear combination of the m vectors that span the space. The set of m vectors therefore is also said to form a **basis** for m-dimensional space.

- $rank(I_n) = n$
- $rank(kA) = rank(A)$, where k is a nonzero constant
- $rank(A') = rank(A)$
- If A is an (m x n) matrix, then $rank(A) \le \min\{m, n\}$.
- If A and B are matrices, then $rank(AB) \le \min\{rank(A), rank(B)\}$.
- **If A is an (n x n) matrix, then $rank(A) = n$ if and only if A is nonsingular; $rank(A) < n$ if and only if A is singular.**

There are operations on the rows/columns of a matrix that leave its rank unchanged:

- Multiplication of a row/column of a matrix by a nonzero constant.
- Addition of a scalar multiple of one row/column to another row/column.
- Interchanging two rows/columns.

INVERSE OF A MATRIX

The **inverse of a nonsingular (n x n) matrix** A is another (n x n) matrix, denoted by A^{-1}, that satisfies the following equalities: $A^{-1}A = AA^{-1} = I$. The inverse of a nonsingular (n x n) matrix is unique.

The inverse of a matrix A in terms of its elements can be obtained from the following formula:

- $A^{-1} = \dfrac{C'}{|A|}$ *where* $C' = [c_{ij}]'$ *and* $c_{ij} = (-1)^{i+j}|A_{ij}|$

Note that C' is the transpose of the matrix of cofactors of A as defined in the section on determinants. C' is also called the **adjoint** of A.

For example, let

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

det(A) = -2 and the cofactors are $c_{11} = 4, c_{22} = 1, c_{12} = -3, c_{21} = -2$. So, the inverse is calculated as,

$$A^{-1} = \frac{1}{-2}\begin{bmatrix} 4 & -3 \\ -2 & 1 \end{bmatrix}' = \begin{bmatrix} -2 & 1 \\ 1.5 & -0.5 \end{bmatrix}.$$

- $I^{-1} = I$
- $(A^{-1})^{-1} = A$
- $(A')^{-1} = (A^{-1})'$
- If A is nonsingular, then A^{-1} is nonsingular.
- If A and B are nonsingular, then $(AB)^{-1} = B^{-1}A^{-1}$.

SOLUTIONS FOR SYSTEMS OF SIMULTANEOUS LINEAR EQUATIONS

Consider the following system of linear equations: $Ax = b$ where A is a (m x n) matrix of known coefficients, x is a (n x 1) vector of unknown variables, and b is a (m x 1) vector of known coefficients.

We want to find the conditions under which: 1) the system has no solution, 2) the system has infinitely many solutions, 3) the system has a unique solution. Define the matrix A|b as the augmented matrix of A. This is just the matrix A with the b vector attached on the end. The dimension of A|b is therefore (m x (n+1)).

Succinctly put, the conditions for the three types of solutions are as follows. (Note: there are numerous ways of characterizing the solutions, but we will stick to the simplest representation):

1. The system has *no solution* if rank(A|b) > rank(A).
2. The system has *infinitely many solutions* if rank(A|b) = rank(A) and rank(A) < n.
3. The system has *a unique solution* if rank(A|b) = rank(A) and rank(A) = n.

Let's look at examples for each case.

Case 1: No Solution

Intuition: if rank(A|b) > rank(A), then b is not in the space spanned by A; so b cannot be represented as a linear combination of A; so there is no x that solves (Ax = b); so there is no solution.

Consider the system,

$$\begin{bmatrix} 2 & 3 \\ 4 & 6 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 8 \\ 9 \end{bmatrix} \qquad or \qquad \begin{aligned} 2x_1 + 3x_2 &= 8 \\ 4x_1 + 6x_2 &= 9 \end{aligned}$$

$$\begin{vmatrix} 2 & 3 \\ 4 & 6 \end{vmatrix} = 0 \quad \Rightarrow \quad \text{singular}$$

$$rank\begin{bmatrix} 2 & 3 \\ 4 & 6 \end{bmatrix} = 1$$

$$rank\begin{bmatrix} 2 & 3 & 8 \\ 4 & 6 & 9 \end{bmatrix} = 2 \quad \Rightarrow \quad rank(A\,|\,b) > rank(A)$$

If we attempt to solve for x_1 in the first equation and substitute the result into the second equation, the resulting equality is $16 = 9$, which is a contradiction.

<u>Case 2: Infinitely Many Solutions</u>
Intuition: if rank(A|b) = rank(A), then b is in the space spanned by A; so b can be represented as a linear combination of A; so there exists an x that solves (Ax = b). But because rank(A) < n, there are more variables than equations. This gives us "free variables" and therefore multiple solutions, one associated with each choice of values for the free variables.

Consider the following system of equations

$$\begin{bmatrix} 2 & 4 \\ 3 & 6 \\ 4 & 8 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 8 \\ 12 \\ 16 \end{bmatrix} \qquad or \qquad \begin{aligned} 2x_1 + 4x_2 &= 8 \\ 3x_1 + 6x_2 &= 12 \\ 4x_1 + 8x_2 &= 16 \end{aligned}$$

$$rank\begin{bmatrix} 2 & 4 \\ 3 & 6 \\ 4 & 8 \end{bmatrix} = 1$$

$$rank\begin{bmatrix} 2 & 4 & 8 \\ 3 & 6 & 12 \\ 4 & 8 & 16 \end{bmatrix} = 1$$

In this case, rank(A|b) = rank(A), but the rank is less than the number of unknown variables (n). Also notice that each equation is just some linear combination of the other two, so we really have only one equation and two unknowns. There are infinitely many solutions that can solve this system, including (4 0)', (2 1)', (0 2)'.

Case 3: Unique Solution
Intuition: if rank(A|b) = rank(A), then b is in the space spanned by A; so b can be represented as a linear combination of A; so there exists an x that solves (Ax = b). Because rank(A) = n, there are equal numbers of variables and equations. This gives us no "free variables" and therefore a single solution.

Consider the following system,

$$\begin{bmatrix} 2 & 3 \\ 3 & 5 \\ 4 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 7 \\ 11 \\ 14 \end{bmatrix} \qquad or \qquad \begin{aligned} 2x_1 + 3x_2 &= 7 \\ 3x_1 + 5x_2 &= 11 \\ 4x_1 + 6x_2 &= 14 \end{aligned}$$

$$rank \begin{bmatrix} 2 & 3 \\ 3 & 5 \\ 4 & 6 \end{bmatrix} = 2$$

$$rank \begin{bmatrix} 2 & 3 & 7 \\ 3 & 5 & 11 \\ 4 & 6 & 14 \end{bmatrix} = 2 \quad because \quad \begin{vmatrix} 2 & 3 & 7 \\ 3 & 5 & 11 \\ 4 & 6 & 14 \end{vmatrix} = 0 \quad and \quad \begin{vmatrix} 2 & 3 \\ 3 & 5 \end{vmatrix} = 1 \neq 0$$

So, rank(A|b) = rank(A) = 2 = n < m. There is full column rank, and the system can be uniquely solved. In fact, any two independent equations can be used to solve for the x's. The solution is $x_1 = 2,\ x_2 = 1$.

In econometrics, we often deal with square matrices, so the following is important for us:

- **If A is a square matrix (m = n) and nonsingular, then $x = A^{-1}b$ is the unique solution.**

KRONECKER PRODUCT
Let A be an (M x N) matrix and B be a (K x L) matrix. Then the **Kronecker** product (or direct product) of A and B, written as $A \otimes B$, is defined as the (MK x NL) matrix

$$C = A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1N}B \\ a_{21}B & a_{22}B & \cdots & a_{2N}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1}B & a_{M2}B & \cdots & a_{MN}B \end{bmatrix}$$

For example if

$$A = \begin{bmatrix} 1 & 3 \\ 2 & 0 \end{bmatrix} \quad and \quad B = \begin{bmatrix} 2 & 2 & 0 \\ 1 & 0 & 3 \end{bmatrix}$$

Their Kronecker product is

$$A \otimes B = \begin{bmatrix} 1\begin{bmatrix} 2 & 2 & 0 \\ 1 & 0 & 3 \end{bmatrix} & 3\begin{bmatrix} 2 & 2 & 0 \\ 1 & 0 & 3 \end{bmatrix} \\ 2\begin{bmatrix} 2 & 2 & 0 \\ 1 & 0 & 3 \end{bmatrix} & 0\begin{bmatrix} 2 & 2 & 0 \\ 1 & 0 & 3 \end{bmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 2 & 0 & 6 & 6 & 0 \\ 1 & 0 & 3 & 3 & 0 & 9 \\ 4 & 4 & 0 & 0 & 0 & 0 \\ 2 & 0 & 6 & 0 & 0 & 0 \end{bmatrix}$$

Note that

$$B \otimes A = \begin{bmatrix} 2 & 6 & 2 & 6 & 0 & 0 \\ 4 & 0 & 4 & 0 & 0 & 0 \\ 1 & 3 & 0 & 0 & 3 & 9 \\ 2 & 0 & 0 & 0 & 6 & 0 \end{bmatrix}$$

- $A \otimes B \neq B \otimes A$,
- $(A \otimes B)' = A' \otimes B'$
- $(A \otimes B)(C \otimes D) = AC \otimes BD$
- $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$
- $A \otimes (B + C) = A \otimes B + A \otimes C$

VECTOR AND MATRIX DIFFERENTIATION

In least squares and maximum likelihood estimation, we need to take derivatives of the objective function with respect to a vector of parameters.

Let a function relating y, a scalar, to a set of variables $x_1, x_2, \ldots, x_n$ be $y = f(x_1, x_2, \ldots, x_n)$ or $y = f(\mathbf{x})$, where **x** is an (n x 1) column vector. (Notice that **x** is in bold to indicate a vector.)

The **gradient** of y is the derivatives of y with respect to each element of x as follows

$$\frac{\partial y}{\partial \mathbf{x}} = \begin{bmatrix} \partial y / \partial x_1 \\ \partial y / \partial x_2 \\ \vdots \\ \partial y / \partial x_n \end{bmatrix}$$

Notice the matrix of derivatives of y is a column vector because y is differentiated with respect to x, an (n x 1) column vector.

The same operations can be extended to derivatives of an (m x n) matrix X, such as

$$\frac{\partial y}{\partial X} = \begin{bmatrix} \partial y/\partial x_{11} & \partial y/\partial x_{12} & \cdots & \partial y/\partial x_{1n} \\ \partial y/\partial x_{21} & \partial y/\partial x_{22} & \cdots & \partial y/\partial x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \partial y/\partial x_{m1} & \partial y/\partial x_{m2} & \cdots & \partial y/\partial x_{mn} \end{bmatrix}$$

Notice in this case, the matrix of derivatives is an (m x n) matrix (the same dimension as X).

If, instead, y is an (m x 1) column vector of $y_i, i = 1,2,...,m$ and x is a (n x 1) column vector of $x_j, j = 1,2,...,n$, then the first derivatives of y with respect to x can be represented as an (m x n) matrix, called the **Jacobian** matrix of y with respect to x':

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}'} = \left[\frac{\partial y_i}{\partial x_j}\right] = \begin{bmatrix} \partial y_1/\partial x_1 & \partial y_1/\partial x_2 & \cdots & \partial y_1/\partial x_n \\ \partial y_2/\partial x_1 & \partial y_2/\partial x_2 & \cdots & \partial y_2/\partial x_n \\ \vdots & \vdots & \ddots & \vdots \\ \partial y_m/\partial x_1 & \partial y_m/\partial x_2 & \cdots & \partial y_m/\partial x_n \end{bmatrix}$$

<u>Aside</u>: when differentiating vectors and matrices, note the dimensions of the independent variable (y) and the dependent variables (x). These will determine if the differentiation will entail the transpose of a matrix. In the above example, the first column of the resulting (m x n) matrix is the derivative of the vector of $y_i, i = 1,2,...,m$ with respect to the first x_1. The second column is the derivative with respect to x_2 and so on. Also note that the first row is the derivative of y_1 with respect to the vector x' (a (1 x n) row vector). Therefore because x is a column vector, we need to transpose it to represent the derivative of the m observations of y (down the column) with respect to the n unknown x variables (across the row). The y vector does not need to be transposed because y is represented along the column of the resulting Jacobian matrix.

If we turn back to the scalar case of y, the second derivatives of y with respect to the column vector **x** are defined as follows.

$$\frac{\partial^2 \mathrm{y}}{\partial \mathbf{x} \partial \mathbf{x}'} = \left[\frac{\partial^2 y}{\partial x_i \partial x_j}\right] = \begin{bmatrix} \partial^2 y/\partial x_1^2 & \partial^2 y/\partial x_1 \partial x_2 & \cdots & \partial^2 y/\partial x_1 \partial x_n \\ \partial^2 y/\partial x_2 \partial x_1 & \partial^2 y/\partial x_2^2 & \cdots & \partial^2 y/\partial x_2 \partial x_n \\ \vdots & \vdots & \ddots & \vdots \\ \partial^2 y/\partial x_n \partial x_1 & \partial^2 y/\partial x_n \partial x_2 & \cdots & \partial^2 y/\partial x_n^2 \end{bmatrix}$$

This matrix is symmetric and is called the **Hessian** matrix of y.

Note that the Hessian matrix is just the second derivative of the gradient with respect to the x vector. We need to transpose the x vector when taking the second derivative because for the Hessian, we are taking the derivative of the gradient (a vector) with respect to each x variable. So, the first column is the gradient differentiated with respect to x_1, and the second column is the gradient differentiated with respect to x_2 and so on. So, we need to differentiate the gradient with respect to x' to order these derivatives across the rows of the resulting matrix.

Based on the previous definitions, the rules of derivatives in matrix notation can be established for reference. Consider the following function $z = \mathbf{c}'\mathbf{x}$, where **c** is a (n x 1) vector and does not depend on **x**, and **x** is an (n x 1) vector, and z is a scalar. Then

$$\frac{\partial z}{\partial \mathbf{x}} = \frac{\partial \mathbf{c}'\mathbf{x}}{\partial \mathbf{x}} = \begin{bmatrix} \partial z/\partial x_1 \\ \partial z/\partial x_2 \\ \vdots \\ \partial z/\partial x_n \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \mathbf{c}$$

If $z = C'\mathbf{x}$, where C is an (n x n) matrix and **x** is an (n x 1) vector, then

$$\frac{\partial z'}{\partial \mathbf{x}} = \frac{\partial \mathbf{x}'\mathrm{C}}{\partial \mathbf{x}} = \left(\mathbf{c_1} \quad \mathbf{c_2} \quad \cdots \quad \mathbf{c_n}\right) = \mathrm{C}$$

where $\mathbf{c}_{\mathrm{i}}$ is the i$^{\text{th}}$ column (remember **c** is a vector) of C.

The following formula for the **quadratic form** $z = \mathbf{x}'A\mathbf{x}$ is also useful (<u>for any (n x n) matrix A</u>),

$\frac{\partial z}{\partial \mathbf{x}} = \frac{\partial \mathbf{x}'\mathrm{A}\mathbf{x}}{\partial \mathbf{x}} = A'\mathbf{x} + \mathrm{A}\mathbf{x} = (\mathrm{A}' + \mathrm{A})\mathbf{x}$. The proof of this result is given in the appendix.

If A is a <u>symmetric</u> matrix (A = A'), then

$$\frac{\partial \mathbf{x}'\mathrm{A}\mathbf{x}}{\partial \mathbf{x}} = 2A\mathbf{x}$$

For the second derivatives for any square matrix A,

$$\frac{\partial^2 (\mathbf{x}'\mathrm{A}\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}'} = A + \mathrm{A}'$$

and if A = A' (if A is symmetric), then

$$\frac{\partial^2 (\mathbf{x}'\mathrm{A}\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}'} = 2\mathrm{A}$$

Some other rules (x is a scalar, unless noted otherwise):

- $\frac{\partial \mathbf{x}'\mathrm{B}\mathbf{y}}{\partial \mathrm{B}} = \mathbf{x}\mathbf{y}'$, where **x** and **y** are (n x 1) column vectors and B is an (n x n) matrix
- $\frac{\partial tr(A)}{\partial \mathrm{A}} = I$

- $\dfrac{\partial|A|}{\partial \mathrm{A}} = |A|(A')^{-1}$
- $\dfrac{\partial \ln|A|}{\partial \mathrm{A}} = (A')^{-1}$
- $\dfrac{\partial AB}{\partial \mathrm{x}} = A\left(\dfrac{\partial \mathrm{B}}{\partial \mathrm{x}}\right) + \left(\dfrac{\partial A}{\partial \mathrm{x}}\right)B$
- $\dfrac{\partial A^{-1}}{\partial \mathrm{x}} = A^{-1}\left(\dfrac{\partial A}{\partial x}\right)A^{-1}$

Since this review was by no means complete, if you want to learn more about matrix algebra, the following are good references:

Anton, Howard (1994), *Elementary Linear Algebra*, 7th edition, New York: John Wiley & Sons.

The math behind it all. Check out chapters 1, 2, 5.6.

Judge, George G., R. Carter Hill, William E. Griffiths, Helmut Lutkepohl, and Tsoung-Chao Lee (1988), *Introduction to the Theory and Practice of Econometrics*, 2nd Edition, New York: John Wiley & Sons, Appendix A.

These notes follow the Appendix fairly closely.

Leon, Steven J. (1994), *Linear Algebra with Applications*, 4th edition, New Jersey: Prentice Hall.

Simon, Carl P. and Lawrence Blume (1994), *Mathematics for Economists*, New York: W.W. Norton.

Look at chapters 6 – 9, & 26.

APPENDIX

Claim: $\dfrac{\partial z}{\partial \mathbf{x}} = \dfrac{\partial \mathbf{x}'\mathrm{A}\mathbf{x}}{\partial \mathbf{x}} = A'\mathbf{x} + \mathrm{A}\mathbf{x} = (\mathrm{A}' + \mathrm{A})\mathbf{x}$

Proof:

Write out the quadratic form for an (n x n) matrix A,

$$z = \mathbf{x}'A\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}x_1 + a_{21}x_2 + \ldots + a_{n1}x_n & a_{12}x_1 + a_{22}x_2 + \ldots + a_{n2}x_n & \cdots & a_{1n}x_1 + a_{2n}x_2 + \ldots + a_{nn}x_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$= \left[x_1(a_{11}x_1 + a_{21}x_2 + \ldots + a_{n1}x_n) + x_2(a_{12}x_1 + a_{22}x_2 + \ldots + a_{n2}x_n) + \cdots + x_n(a_{1n}x_1 + a_{2n}x_2 + \ldots + a_{nn}x_n) \right]$$

$$= \begin{bmatrix} a_{11}x_1^2 + a_{22}x_2^2 + \ldots + a_{nn}x_n^2 + (a_{12} + a_{21})x_1x_2 + (a_{13} + a_{31})x_1x_3 + (a_{14} + a_{41})x_1x_4 + \ldots + (a_{1n} + a_{n1})x_1x_n \\ + (a_{23} + a_{32})x_2x_3 + (a_{24} + a_{42})x_2x_4 + \ldots + (a_{2n} + a_{n2})x_2x_n + \ldots + (a_{n,n-1} + a_{n-1,n})x_nx_{n-1} \end{bmatrix}$$

Now differentiate this with respect to the vector x,

$$\frac{\partial z}{\partial \mathbf{x}} = \begin{bmatrix} 2a_{11}x_1 + (a_{21} + a_{12})x_2 + (a_{31} + a_{13})x_3 + \ldots + (a_{n1} + a_{1n})x_n \\ (a_{12} + a_{21})x_1 + 2a_{22}x_2 + (a_{32} + a_{23})x_3 + \ldots + (a_{n2} + a_{2n})x_n \\ \vdots \\ (a_{1n} + a_{n1})x_1 + (a_{2n} + a_{n2})x_2 + (a_{3n} + a_{n3})x_3 + \ldots + 2a_{nn}x_n \end{bmatrix}$$

But this can be rewritten as,

$$\frac{\partial z}{\partial \mathbf{x}} = \begin{bmatrix} 2a_{11} & (a_{21} + a_{12}) & (a_{31} + a_{13}) & \cdots & (a_{n1} + a_{1n}) \\ (a_{12} + a_{21}) & 2a_{22} & (a_{32} + a_{23}) & \ldots & (a_{n2} + a_{2n}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (a_{1n} + a_{n1}) & (a_{2n} + a_{n2}) & (a_{3n} + a_{n3}) & \cdots & 2a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

or

$$\frac{\partial z}{\partial \mathbf{x}} = (A' + A)\mathbf{x}$$

If A is symmetric, then $a_{ij} = a_{ji} \quad for \quad all \quad i, j$, so

$$\frac{\partial z}{\partial \mathbf{x}} = 2A\mathbf{x}$$

This also holds if n = n +1, so, by induction, the result holds for any (n x n) matrix.