NOTES on MATRIX MATHEMATICS
Professor Sager

Definitions: A **matrix** is a rectangular table of elements, denoted by $A = [a_{ij}]$, $i = 1, \ldots, m$ (rows) and $j = 1, \ldots, n$ (columns), a_{ij} = element in row i and column j
A **vector** is a matrix for which either m=1 or n=1 (one row or one column).
If m=1, the matrix is a **row vector**, denoted by u'.
If n=1, the matrix is a **column vector**, denoted by u.
If both m=1 and n=1, the matrix is a **scalar**.
If m=n, the matrix is **square**.

Definition: **A=B** if and only if A and B have the same number of rows and the same number of columns and all corresponding elements are the same: $a_{ij} = b_{ij}$, for all i and j.

Ex: $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$

Ex: $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \neq \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 7 \end{bmatrix}$

Ex: $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \neq \begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix}$

Addition and Subtraction of Matrices

Definition: A and B are **conformable for addition (or subtraction)** if A and B have the same number of rows and they have the same number of columns.

Adding (or subtracting) matrices makes no sense unless the matrices are conformable.

Definition: If A and B are conformable for addition (or subtraction), then **A + B** is the matrix formed by adding corresponding elements of A and B, and **A – B** is the matrix formed by subtracting corresponding elements of B from A.
Exception: If A is a scalar, then A + B is interpreted so that the addition makes sense. That is, the scalar A is expanded into a matrix of identical scalars that is conformable to B. Therefore, A + B is obtained by adding A to every element of B. Similarly, A – B is the matrix formed by adding A to the negative of each element of B.

Ex: $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} + \begin{bmatrix} -1 & 0 & 2 \\ 1 & -2 & 3 \end{bmatrix} = \begin{bmatrix} 0 & 2 & 5 \\ 5 & 3 & 9 \end{bmatrix}$

Ex: $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} - \begin{bmatrix} -1 & 0 & 2 \\ 1 & -2 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 2 & 1 \\ 3 & 7 & 3 \end{bmatrix}$

Ex: $[2] + \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \equiv \begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \end{bmatrix} + \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 3 & 4 & 5 \\ 6 & 7 & 8 \end{bmatrix}$

Ex: $[2] - \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \equiv \begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \end{bmatrix} - \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 \\ -2 & -3 & -4 \end{bmatrix}$

Multiplication of Matrices

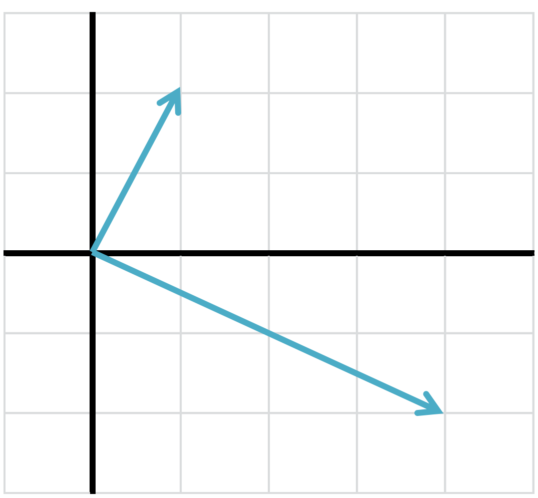
Definition: If A is a scalar, AB and BA are equal and each is the matrix formed by multiplying every element of B by A.

Ex: $[2]\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \end{bmatrix}$

Definition: If $u = (u_1, \ldots, u_n)$ and $v = (v_1, \ldots, v_n)$ are vectors with the same number of elements, then the **inner product** is $u' \cdot v = u_1 v_1 + \ldots + u_n v_n$.

Ex: $[1 \quad 2] \cdot [4 \quad -2] = 1 \cdot 4 + 2(-2) = 0$

Comment: The inner product of two vectors is zero if and only if the vectors are perpendicular (orthogonal). This is shown geometrically for the preceding example in the graph below. If these vectors were data on two different variables, say income and years of schooling for a sample of people, they would be uncorrelated. Statistical correlation is proportional to the inner product between two data vectors. There is no predictive relationship between uncorrelated variables.



Definition: A and B are **conformable for AB matrix multiplication** if the number of columns of A is the same as the number of rows of B.

Note: If A and B are conformable for AB multiplication, they may not be conformable for BA multiplication.

Definition: If A and B are conformable for AB multiplication, then **AB** is the matrix for which element i,j is the inner product of row i of A with column j of B.

Ex: $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 & -1 \\ 0 & -1 & 2 & 0 \\ -1 & 0 & 3 & 1 \end{bmatrix} = \begin{bmatrix} -2 & 0 & 13 & 2 \\ -2 & 3 & 28 & 2 \end{bmatrix}$

Ex: $\begin{bmatrix} 8 & 1 \\ -12 & 1 \end{bmatrix} \begin{bmatrix} P \\ Q \end{bmatrix} = \begin{bmatrix} 8P+Q \\ -12P+Q \end{bmatrix} = \begin{bmatrix} 120 \\ -80 \end{bmatrix}$

This represent a demand (top equation) and supply (bottom equation) relationship between price and quantity.

Ex: $\begin{bmatrix} -b & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} Y \\ C \end{bmatrix} = \begin{bmatrix} -bY+C \\ Y-C \end{bmatrix} = \begin{bmatrix} a \\ I_0 + G_0 \end{bmatrix}$

This represents a macroeconomic relationship between income and consumption (top equation) and between income and consumption, investment, and government spending (bottom equation).

Note: All systems of simultaneous linear equations can be expressed as $Ax = d$, where Ax is matrix multiplication of matrix A by vector x:

The system $\left\{ \begin{matrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = d_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = d_2 \\ \vdots \quad \vdots \quad \vdots \quad \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = d_n \end{matrix} \right\}$ is $\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix}$

Note: In general, AB ≠ BA, even if both multiplications are defined.

Ex: $\begin{bmatrix} 1 & 2 & 0 & -1 \\ 0 & -1 & 2 & 0 \\ -1 & 0 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ is not defined.

Ex: $\begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 2 & 2 \end{bmatrix}$

But

Ex: $\begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 3 \\ -1 & 5 \end{bmatrix}$

Division of matrices requires developing the concept of matrix inverse (later).

Geometric interpretation of vectors

- A vector $u = (u_1, \ldots, u_n)$ represents a point in n-dimensional space.
- The length of the vector is $\sqrt{u' \cdot u} = \sqrt{u_1^{\ 2} + \ldots + u_n^{\ 2}}$. This is the Pythagorean Theorem in n-dimensional space.
- The sum of two vectors is a point obtained by laying the vectors end-to-end.
- The inner product $u' \cdot v$ of two vectors is proportional to the cosine of the angle between *u* and *v*. If $u' \cdot v = 0$, then the two vectors are perpendicular (orthogonal).

Definition: A **linear combination of vectors** $v_1, v_2, \ldots, v_n$ is a vector formed by multiplying each vector by a scalar number and then adding the results: $u = c_1 v_1 + c_2 v_2 + \ldots + c_n v_n$, for some scalar constants $c_1, c_2, \ldots, c_n$.

Every vector $u = (u_1, \ldots, u_n)$ is a linear combination of the unit basis vectors $e_1 = (1,0,\ldots,0)$, $e_2 = (0,1,0,\ldots,0)$, …, $e_n = (0,\ldots,0,1)$: $u = u_1 e_1 + u_2 e_2 + \ldots + u_n e_n$.

Definition: The **span** of a set of vectors $\{v_1, v_2, \ldots, v_n\}$ is the set of all possible linear combinations of $\{v_1, v_2, \ldots, v_n\}$.

The span of the unit basis vectors $\{e_1, e_2, \ldots, e_n\}$ is the entire n-dimensional space $\Re^n$.

Ex: (1 0) and (0 1) span the plane. Every point (a b) in the plane is a linear combination of (1 0) and (0 1), namely (a b) = a(1 0) + b(0 1). In fact, with a little work, it can be shown that any two vectors span the plane, as long as the two vectors do not overlay each other.

Solving a system of linear equations simultaneously is a very important application of matrix algebra. For example, it is used in calculus to minimize/maximize multivariable quadratic functions and to calculate regression equations in statistics. It can be shown that a set of linear equations $Ax = d$ has one solution, no solution, or infinitely many solutions. If there is more than one solution, then there are infinitely many solutions.

Ex: $\begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 7 \\ 9 \end{bmatrix} \Rightarrow \begin{bmatrix} 3x_1 + 2x_2 = 7 \\ x_1 + 4x_2 = 9 \end{bmatrix} \Rightarrow x_1 = 1, \quad x_2 = 2$

Ex: $\begin{bmatrix} 3 & 2 \\ 6 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 7 \\ 9 \end{bmatrix} \Rightarrow \begin{bmatrix} 3x_1 + 2x_2 = 7 \\ 6x_1 + 4x_2 = 9 \end{bmatrix} \Rightarrow$ no solution

Ex: $\begin{bmatrix} 3 & 2 \\ 6 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} 3x_1 + 2x_2 = 0 \\ 6x_1 + 4x_2 = 0 \end{bmatrix} \Rightarrow$ infinitely many solutions

Whether or not the system of linear equations $Ax = d$ has a solution for *x* and whether the solution is unique depend on the properties of the column vectors of *A*. *Ax* is a linear combination of the column vectors of *A*. If the value of *x* is changed, the value of the linear combination *Ax* also changes. One (not very good) strategy for finding a solution to $Ax = d$ is to go through all possible values of *x*, and see if *d* is among the set of possible linear combinations *Ax* that result. This requires searching the *span* of A, since the span of the column vectors of A is defined to be the set of all possible linear combinations *Ax*. Thus if *d* is included among the span of the column vectors $a_1, a_2, \ldots, a_n$ of *A*, then for some numbers $x_1, \ldots, x_n$, we have $x_1 a_1 + x_2 a_2 + \ldots + x_n a_n = d$. That is, $Ax = d$. So if the column vectors $a_1, a_2, \ldots, a_n$ of *A* span $\Re^n$, then there must be a solution to $Ax = d$ for any *d*. There are several conditions on *A* that are equivalent to this and guarantee a solution. One of these conditions is the linear independence of the columns (or rows) of *A*.

Definition: The vectors $\{v_1, v_2, \ldots, v_n\}$ are **linearly independent** if the only solution for the set of equations $x_1 v_1 + x_2 v_2 + \ldots + x_n v_n = 0$ is $(x_1, x_2, \ldots, x_n) = (0,0, \ldots, 0)$. If there is a solution that is not all zeroes, then $\{v_1, v_2, \ldots, v_n\}$ are **linearly dependent**.

Note on the terminology: Linearly dependent vectors depend on each other. To see this, suppose $\{v_1, v_2, \ldots, v_n\}$ are linearly dependent. Then according to the definition, there is a set of x's, not all zero, that satisfy $x_1 v_1 + x_2 v_2 + \ldots + x_n v_n = 0$. Suppose $x_1 \neq 0$. Divide and rearrange to get $v_1 = -\frac{x_2}{x_1} v_2 - \ldots - \frac{x_n}{x_1} v_n$. This says that v_1 is a linear function of the other vectors, so v_1 depends linearly on the others.

Further note: To say that $x_1 v_1 + x_2 v_2 + \ldots + x_n v_n = 0$ is to say that $Vx = 0$, where *V* is the matrix that has $v_1, v_2, \cdots, v_n$ as its columns, i.e., $V = [v_1 \vdots v_2 \vdots \cdots \vdots v_n]$. Obviously, $Vx = 0$ has $x=0$ as a solution; the vectors $\{v_1, v_2, \ldots, v_n\}$ are linearly independent if $x=0$ is the only solution. So the term *linear independence*, which does not sound as though it is about solving systems of equations, is really about solving a special case of $Ax = d$: Linear independence says that the special case $Ax = 0$ has a unique solution for *x*.

Ex: $\begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} 3x_1 + 2x_2 = 0 \\ x_1 + 4x_2 = 0 \end{bmatrix} \Rightarrow x_1 = 0, \quad x_2 = 0$, so $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 4 \end{bmatrix}$ are linearly independent.

Ex: $\begin{bmatrix} 3 & 2 \\ 6 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} 3x_1 + 2x_2 = 0 \\ 6x_1 + 4x_2 = 0 \end{bmatrix} \Rightarrow$ infinitely many solutions, so $\begin{bmatrix} 3 \\ 6 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 4 \end{bmatrix}$ are linearly dependent. Note that $\begin{bmatrix} 3 \\ 6 \end{bmatrix} = 1.5 \begin{bmatrix} 2 \\ 4 \end{bmatrix}$, so $1 \begin{bmatrix} 3 \\ 6 \end{bmatrix} - 1.5 \begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ shows that these two vectors are linearly dependent by the definition.

Theorem: The vectors $\{v_1, v_2, \ldots, v_n\}$ are linearly independent if and only if they span $\Re^n$.

Proof: <omitted>

Matrix Algebra

Algebra of numbers:

Associative Law	
Addition	(a + b) + c = a + (b + c)
Multiplication	(ab)c = a(bc)
Commutative Law	
Addition	a + b = b + a
Multiplication	ab = ba
Distributive Law I	a(b + c) = ab + ac
Distributive Law II	(b + c)a = ba + ca

The above properties are all true whenever a, b, c are numbers. Which of these properties of numbers remain true for matrices?

Answer: All, except for commutative law for multiplication.

Ex (of noncommutativity of multiplication):

Let $u' = \begin{bmatrix} 3 & 2 \end{bmatrix}$, $v = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$. Then $u'v = 11$, but $vu' = \begin{bmatrix} 3 & 2 \\ 12 & 8 \end{bmatrix}$.

Let A = $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, B = $\begin{bmatrix} 0 & -1 \\ 6 & 7 \end{bmatrix}$. Then AB = $\begin{bmatrix} 12 & 13 \\ 24 & 25 \end{bmatrix}$, but BA = $\begin{bmatrix} -3 & -4 \\ 27 & 40 \end{bmatrix}$.

Identity and Zero

In number algebra, 0 is the identity for addition: a + 0 = a,
and 1 is the identity for multiplication: a1 = 1a = a.

The identity for matrix addition is a conformable matrix of zeroes:

$$A_{mxn} + 0_{mxn} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} + \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}.$$

The identity for matrix multiplication is a conformable square matrix with elements 1 on the main diagonal and 0 elsewhere:

$$A_{mxn}I_n = I_mA_{mxn} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}_{nxn} =$$

$$\begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}_{mxm} \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}.$$

Note: $I_n I_n = I_n$

Note: In number algebra, if ab = 0, then a = 0 or b = 0. This is not necessarily true in matrix algebra.

Ex: $\begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} -2 & 4 \\ 1 & -2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$.

Note: In number algebra, if cd = ce, then d = e as long as $c \neq 0$. But this is not necessarily true in matrix algebra.

Ex: $\begin{bmatrix} 2 & 3 \\ 6 & 9 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 5 & 8 \\ 15 & 24 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 6 & 9 \end{bmatrix}\begin{bmatrix} -2 & 1 \\ 3 & 2 \end{bmatrix}$. But $\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \neq \begin{bmatrix} -2 & 1 \\ 3 & 2 \end{bmatrix}$.

Matrix Transpose

Definition: The **transpose of A** is A', in which the rows and columns of A are interchanged. If A is mxn, then A' is nxm.

Ex: If A = $\begin{bmatrix} 3 & 8 & -9 \\ 1 & 0 & 4 \end{bmatrix}$, then $A' = \begin{bmatrix} 3 & 1 \\ 8 & 0 \\ -9 & 4 \end{bmatrix}$.

Properties of transpose:

(a) $(A')' = A$ [the transpose of the transpose is the original matrix]

Ex: If A = $\begin{bmatrix} 3 & 8 & -9 \\ 1 & 0 & 4 \end{bmatrix}$, then $A' = \begin{bmatrix} 3 & 1 \\ 8 & 0 \\ -9 & 4 \end{bmatrix}$, so $(A')' = \begin{bmatrix} 3 & 8 & -9 \\ 1 & 0 & 4 \end{bmatrix}$.

(b) $(A+B)' = A' + B'$ [the transpose of the sum is the sum of the transposes]

Ex: If A = $\begin{bmatrix} 3 & 8 & -9 \\ 1 & 0 & 4 \end{bmatrix}$ and B = $\begin{bmatrix} 1 & -2 & 6 \\ -3 & 1 & 0 \end{bmatrix}$, then $(A+B)' = \begin{bmatrix} 4 & 6 & -3 \\ -2 & 1 & 4 \end{bmatrix}' = \begin{bmatrix} 4 & -2 \\ 6 & 1 \\ -3 & 4 \end{bmatrix}$.

Also, $A' + B' = \begin{bmatrix} 3 & 1 \\ 8 & 0 \\ -9 & 4 \end{bmatrix} + \begin{bmatrix} 1 & -3 \\ -2 & 1 \\ 6 & 0 \end{bmatrix} = \begin{bmatrix} 4 & -2 \\ 6 & 1 \\ 3 & 4 \end{bmatrix}$.

(c) $(AB)' = B'A'$ [the transpose of the product is the product of the transposes, in reverse order]

Ex: If A = $\begin{bmatrix} 3 & 8 & -9 \\ 1 & 0 & 4 \end{bmatrix}$ and B = $\begin{bmatrix} 2 & 3 \\ -1 & 2 \\ 0 & 1 \end{bmatrix}$, then $(AB)' = \left(\begin{bmatrix} 3 & 8 & -9 \\ 1 & 0 & 4 \end{bmatrix}\begin{bmatrix} 2 & 3 \\ -1 & 2 \\ 0 & 1 \end{bmatrix}\right)' = \begin{bmatrix} -2 & 16 \\ 2 & 7 \end{bmatrix}' =$ $\begin{bmatrix} -2 & 2 \\ 16 & 7 \end{bmatrix}$. Also, $B'A' = \begin{bmatrix} 2 & 3 \\ -1 & 2 \\ 0 & 1 \end{bmatrix}'\begin{bmatrix} 3 & 8 & -9 \\ 1 & 0 & 4 \end{bmatrix}' = \begin{bmatrix} 2 & -1 & 0 \\ 3 & 2 & 1 \end{bmatrix}\begin{bmatrix} 3 & 1 \\ 8 & 0 \\ -9 & 4 \end{bmatrix} = \begin{bmatrix} -2 & 2 \\ 16 & 7 \end{bmatrix}$.

Matrix Inverse

A matrix may have an inverse only if it is square. But not every square matrix has an inverse.

<u>Definition</u>: Let A_{nxn} be a square matrix. If there is a square matrix B_{nxn} such that AB = I_n and BA = I_n, then A is said to be **nonsingular**, and B is said to be an **inverse** of A. We write B = $\mathbf{A^{-1}}$.

Ex: If A = $\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$ and B = $\frac{1}{6}\begin{bmatrix} 2 & -1 \\ 0 & 3 \end{bmatrix}$, then AB = $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ = BA. So B = A^{-1}.

<u>Properties of inverse</u>:

(a)

i. If A is nonsingular, then A^{-1} is nonsingular and $(A^{-1})^{-1} = A$.

ii. $I_n^{-1} = I_n$

iii. If A is nonsingular, then its inverse is unique.

(b) Suppose both A and B are nonsingular. Then AB is nonsingular and $(AB)^{-1} = B^{-1}A^{-1}$.

Ex: $\left(\begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}\right)^{-1} = \begin{bmatrix} -1 & 3 \\ -1 & 5 \end{bmatrix}^{-1} = -\frac{1}{2}\begin{bmatrix} 5 & -3 \\ 1 & -1 \end{bmatrix}$; and

$$\begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}^{-1}\begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\left(\frac{1}{-2}\right)\begin{bmatrix} 5 & -3 \\ -4 & 2 \end{bmatrix} = -\frac{1}{2}\begin{bmatrix} 5 & -3 \\ 1 & -1 \end{bmatrix}$$

(c) Suppose A is nonsingular. Then A' is nonsingular and $(A')^{-1} = (A^{-1})'$

Ex: $\left(\begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix}'\right)^{-1} = \left(\begin{bmatrix} 2 & 4 \\ 3 & 5 \end{bmatrix}\right)^{-1} = -\frac{1}{2}\begin{bmatrix} 5 & -4 \\ -3 & 2 \end{bmatrix}$; and

$$\left(\begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix}^{-1}\right)' = \left(-\frac{1}{2}\begin{bmatrix} 5 & -3 \\ -4 & 2 \end{bmatrix}\right)' = -\frac{1}{2}\begin{bmatrix} 5 & -4 \\ -3 & 2 \end{bmatrix}$$

<u>Importance of inverse</u>: If A is nonsingular, then the system of simultaneous linear equations $Ax = d$ can be solved uniquely by multiplying the matrix equations by A^{-1}:

$$\boldsymbol{A^{-1}Ax = A^{-1}d}, \text{ or } \boldsymbol{I_n x = A^{-1}d}, \text{ or } \boldsymbol{x = A^{-1}d}.$$

Two problems:

1) How can we determine if A^{-1} exists?
2) How can we find A^{-1} if it exists?

Equivalency Theorem: The following conditions are all equivalent:

1. $Ax = d$ has a unique solution for all d.
2. A^{-1} exists.
3. The columns of A are linearly independent.
4. The rows of A are linearly independent.
5. The columns of A span $\Re^n$.

Proof: <omitted>

This theorem provides several conditions that guarantee the existence of a solution for a set of simultaneous linear equations. But they are often difficult to apply in practice. We need an easier method to determine if a solution exists, and we still do not have a way to find A^{-1}. Both are provided by determinants.

There is another concept associated with matrices that provides an additional equivalent condition. Because you may see this concept in other places, it is presented here:

Definition: Let A_{mxn} be a matrix. The **rank of A** is the maximum number of linearly independent rows or columns of A. The rank cannot exceed the minimum of m and n. That is, suppose that $m > n$, and the columns of A are denoted by $\{v_1, v_2, \ldots, v_n\}$. The rank of A is the most numerous subset of $\{v_1, v_2, \ldots, v_n\}$ that is linearly independent. There are 2^n possible subsets of $\{v_1, v_2, \ldots, v_n\}$. All have $\leq n$ members, and some of them may be linearly independent. The one that is linearly independent and is most numerous determines the rank of A. If the whole set $\{v_1, v_2, \ldots, v_n\}$ is linearly independent, then A is said to have **full rank**.

The main fact about rank is that having full rank is another equivalency condition:
Theorem: A square matrix A_{nxn} has rank n if and only if A is nonsingular.

This is true because having full rank means that the columns are all linearly independent, which is condition 3 of the Equivalency Theorem (above). Conversely, if A is nonsingular, then the columns and the rows of A are all linearly independent by the Equivalency Theorem (conditions 3 and 4). Thus, the number of linearly independent columns and rows is n.

Determinants

Evaluation of Determinants

The determinant "determines" whether a matrix is nonsingular. The determinant is defined only for square matrices. The determinant is a number, not a matrix.

First, some special cases:

(a) The determinant of a scalar is the scalar.

(b) The determinant of a 2x2 matrix: $\begin{vmatrix} a & c \\ b & d \end{vmatrix} = ad - bc$.

Ex: $|-6|$ = -6.

Ex: $\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}$ = 1x1 – 0x0 = 1

Ex: $\begin{vmatrix} 1 & 2 \\ 2 & 4 \end{vmatrix}$ = 1x4 – 2x2 = 0

Ex: $\begin{vmatrix} 8 & 120 \\ -12 & -80 \end{vmatrix}$ = 8(-80) – (-12)(120) = 800

Note: A determinant is denoted by straight vertical lines. Do not confuse this with absolute value.

The determinant of a square matrix is defined in a round-about manner. Unfortunately, the definition does not provide a useful way to calculate the value of a determinant. Fortunately, there are useful ways to calculate the value of a determinant – including the most useful of all: Let a computer program do it!

Definition: Let $A_{nxn} = [a_{ij}]$. The **minor** of element a_{ij} is the determinant $|M_{ij}|$, where M_{ij} is the matrix formed by deleting row i and column j from A.

Ex: If A = $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$, then $|M_{32}|$ = $\begin{vmatrix} 1 & 3 \\ 4 & 6 \end{vmatrix}$ = -6 (calculated from special case [above] for 2x2 matrices).

Definition: The **cofactor** of element a_{ij} is $|C_{ij}|$ = $(-1)^{i+j}$ $|M_{ij}|$. (Note that the notation "$|C_{ij}|$" rather confusingly does not mean "determinant of C_{ij}". $|C_{ij}|$ will be either +1 times the determinant of the minor M_{ij}, or it will be –1 times the determinant of the minor M_{ij}.)

Ex: If A = $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$, then $|C_{32}|$ = $(-1)^{3+2}$ $\begin{vmatrix} 1 & 3 \\ 4 & 6 \end{vmatrix}$ = 6.

Definition: (Laplace Expansion) Let $A_{nxn} = [a_{ij}]$. The **determinant** of A is
$|A| = \sum_{j=1}^{n} a_{ij}|C_{ij}| = \sum_{i=1}^{n} a_{ij}|C_{ij}|$.

Notes on the definition:
(1) This definition of the determinant is called an "expansion" because one selects a row or column of the matrix and uses the values in the row or column to weight the corresponding cofactors. One "expands" the determinant around the row or column. The first summation in the definition ($\sum_{j=1}^{n} a_{ij}|C_{ij}|$) shows the expansion of $|A|$ around row i of A; the second summation ($\sum_{i=1}^{n} a_{ij}|C_{ij}|$) shows the expansion of $|A|$ around column j of A. We will accept as a fact that the expansions are the same for all rows and all columns – this is actually a theorem that we will not prove. *So we may use whichever row or column is convenient for expanding* $|A|$. (Note the last sentence. It provides the first leg of a practical basis for calculating determinants.)

(2) The determinant of A_{nxn} is defined in terms of cofactors. Observe that the cofactor for an element of A is a determinant of a lower-order matrix (n-1 x n-1), times +1 or -1. Thus a 3x3 matrix has a determinant defined in terms of 2x2 determinants (which we already know how to compute). So 2x2 determinants allow us to compute 3x3 determinants. Similarly, 4x4 determinants are defined in terms of 3x3 determinants. So knowing how to compute 3x3 determinants, which we know because we know how to compute 2x2 determinants, allows us to compute 4x4 determinants. And so forth. Thus, we can build up from lower-order determinants to calculate any desired level of determinant. This looks complicated – and it is. We do not actually want to compute determinants that way. But there will be some powerful ways to simplify it after we learn more about the properties of determinants.

Ex: Expand $\begin{vmatrix} 8 & 120 \\ -12 & -80 \end{vmatrix}$ around column 2: $\sum_{i=1}^{2} a_{i2}|C_{i2}|$ = 120$|C_{12}|$ - 80$|C_{22}|$ = -120$|M_{12}|$ - 80 $|M_{22}|$ = 120$(-1)^{1+2}$(-12) - 80$(-1)^{2+2}$(8) = 1440 -640 = 800

Ex: It is easiest to expand $\begin{vmatrix} 8 & 1 & 3 \\ 4 & 0 & 1 \\ 6 & 0 & 3 \end{vmatrix}$ around column 2 (why?): $\sum_{i=1}^{3} a_{i2}|C_{i2}|$ = 1 $|C_{12}|$ + 0 $|C_{22}|$ + 0 $|C_{32}|$ = 1$(-1)^{1+2}$$\begin{vmatrix} 4 & 1 \\ 6 & 3 \end{vmatrix}$ = -(4x3 – 6x1) = -6

Properties of Determinants

I. $|A| = |A'|$ (The determinant of a matrix is the same as the determinant of its transpose.)

II. Interchange of any two rows (or any two columns) of A changes the sign but not the absolute value of the determinant of A.

III. Multiplication of one row (or one column) of A by a constant *k* multiplies the value of the determinant by *k*.

IV. The addition (or subtraction) of a constant multiple of any row of A to another row does not change the value of the determinant. The addition (or subtraction) of a constant multiple of any column of A to another column does not change the value of the determinant.

V. If one row of A is a constant multiple of another row of A, then $|A| = 0$. If one column of A is a constant multiple of another column of A, then $|A| = 0$.

These properties can help in evaluating the determinant of A. How? By using them cleverly, we can reduce many entries in A to zeroes. Then the Laplace expansion will be easy.

Ex: (I) $\begin{vmatrix} 8 & 120 \\ -12 & -80 \end{vmatrix}$ = 800 and $\begin{vmatrix} 8 & -12 \\ 120 & -80 \end{vmatrix}$ = 8(-80) - 120(-12) = 800

Ex: (II) Interchange rows 1 and 2: $\begin{vmatrix} -12 & -80 \\ 8 & 120 \end{vmatrix}$ = -12(120) – (-80)(-8) = -800

Ex: (III) Divide column 2 by 10 (multiply by 0.1): $\begin{vmatrix} 8 & 12 \\ -12 & -8 \end{vmatrix}$ = 8(-8) - (-12)12 = 80

Ex: (IV) Add -15 x column 1 to column 2: $\begin{vmatrix} 8 & 120+(-15)8 \\ -12 & -80+(-15)(-12) \end{vmatrix} = \begin{vmatrix} 8 & 0 \\ -12 & 100 \end{vmatrix}$ = 8x100 - (-12)0 = 800

Ex: (V) $\begin{vmatrix} 1 & 2 \\ 2 & 4 \end{vmatrix}$ = 1x4 – 2x2 = 0

These properties can be combined to produce a systematic method for quickly calculating the determinant of any square matrix on a computer. Because of this, the following theorem gives us a practical method for determining when a system of simultaneous linear equations has a unique solution.

<u>Theorem</u>: Let A be nxn. $|A| \neq 0$ if and only if A is nonsingular.

Proof: <omitted>

<u>Note</u>: This provides a sixth condition for the Equivalency Theorem. Of all of the equivalency conditions, this is the easiest to verify.

Ex: Since $\begin{vmatrix} 1 & 2 \\ 2 & 4 \end{vmatrix}$ = 1x4 – 2x2 = 0, then $\begin{vmatrix} 1 & 2 \\ 2 & 4 \end{vmatrix}$ is singular. Thus, the system of simultaneous linear equations $\begin{bmatrix} x_1 + 2x_2 = 3 \\ 2x_1 + 4x_2 = 0 \end{bmatrix}$ (the same as $\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$) does not have a unique solution. (Actually, there is no solution at all: the equations are inconsistent). And the system $\begin{bmatrix} x_1 + 2x_2 = 0 \\ 2x_1 + 4x_2 = 0 \end{bmatrix}$ has infinitely many solutions (the two equations are "the same", and any (x_1, x_2) that satisfies $x_1 + 2x_2 = 0$ is a solution).

Ex: Does $\begin{bmatrix} 7x_1 - 3x_2 - 3x_3 = 7 \\ 2x_1 + 4x_2 + x_3 = 0 \\ -2x_2 - x_3 = 2 \end{bmatrix}$ have a unique solution? To find out, calculate $|A|$ = $\begin{vmatrix} 7 & -3 & -3 \\ 2 & 4 & 1 \\ 0 & -2 & -1 \end{vmatrix}$ = (subtract 2/7 x row 1 from row 2) $\begin{vmatrix} 7 & -3 & -3 \\ 0 & 34/7 & 13/7 \\ 0 & -2 & -1 \end{vmatrix}$ = (add 14/34 x row 2 to row 3) $\begin{vmatrix} 7 & -3 & -3 \\ 0 & 34/7 & 13/7 \\ 0 & 0 & -8/34 \end{vmatrix}$ = (expand by column 1) $7(-1)^{1+1}\begin{vmatrix} 34/7 & 13/7 \\ 0 & -8/34 \end{vmatrix}$ = 7[(34/7)(-8/34) – 0(13/7)] = -8 $\neq$ 0. So A is nonsingular, and therefore there is a unique solution by the Equivalency Theorem.

Here is an easily programmable method for quickly calculating any determinant. The reason that this works for any determinant is that all determinants can be reduced to upper triangular form by repeated application of matrix operation IV above – which leaves the value of the determinant unchanged.

Definition: A square matrix A = [a_{ij}] is **upper triangular** if every element below the main diagonal is zero ($a_{ij} = 0$ if i > j).

Theorem: If A_{nxn} is upper triangular, then $|A|$ is the product of the elements on the main diagonal ($|A| = \prod_{i=1}^{n} a_{ii}$).

Proof: <omitted>

Ex: $\begin{vmatrix} 7 & -3 & -3 \\ 0 & 34/7 & 13/7 \\ 0 & 0 & -8/34 \end{vmatrix}$ = 7 x 34/7 x (-8/34) = -8

Corollary: A square upper triangular matrix is singular if and only if at least one of its diagonal elements is zero.

Theorem: For every A_{nxn}, there is an upper triangular matrix A_{nxn}^* that can be obtained from A by repeated application of Property IV so that $|A| = |A^*|$.

Proof: <omitted>

Therefore, we can make every square matrix into an upper triangular matrix by repeatedly adding (or subtracting) an appropriate multiple of row 1 to the rows below it in order to "zero out" the elements in column 1 below the main diagonal. And then proceed to row 2 and add (or subtract) an appropriate multiple of row 2 to the rows below it in order to zero out the elements in column 2 below the main diagonal. Continue through all of the rows successively in this manner. At most (n-1)(n-2)/2 applications of property IV are needed. The determinant of the resulting matrix is the same as the determinant of the original matrix.

Ex: $|A| = \begin{vmatrix} 7 & -3 & -3 \\ 2 & 4 & 1 \\ 0 & -2 & -1 \end{vmatrix}$ = (subtract 2/7 x row 1 from row 2) $\begin{vmatrix} 7 & -3 & -3 \\ 0 & 34/7 & 13/7 \\ 0 & -2 & -1 \end{vmatrix}$ = (add 14/34 x row 2 to row 3) $\begin{vmatrix} 7 & -3 & -3 \\ 0 & 34/7 & 13/7 \\ 0 & 0 & -8/34 \end{vmatrix}$ = 7 * 34/7 *(-8/34) = -8.

We have now answered the question: When does A^{-1} exist? We have a computationally feasible procedure based on the determinant of A.

We turn now to the question: How to find A^{-1}? Again, the approach is roundabout.

Definition: **Expanding $|A|$ by alien cofactors** means to apply the Laplace expansion by using the elements of one row, but multiplying by the corresponding cofactors of another row – or to expand by the elements of one column, but multiplying by the corresponding cofactors of another column.

Theorem: (Property VI) The result of expanding $|A|$ by alien cofactors is zero.

Proof: <omitted>

Ex: $|A| = \begin{vmatrix} 7 & -3 & -3 \\ 2 & 4 & 1 \\ 0 & -2 & -1 \end{vmatrix}$ = (expand by row 1 using cofactors of row 2) $7\begin{vmatrix} -3 & -3 \\ -2 & -1 \end{vmatrix}$ -3$\begin{vmatrix} 7 & -3 \\ 0 & -1 \end{vmatrix}$ -3 $\begin{vmatrix} 7 & -3 \\ 0 & -2 \end{vmatrix}$ = 7(-3-6) –3(-7+0) –3(-14+0) = -63 +21 +42 = 0.

Definition: The **adjoint of A** is the transpose of the matrix of cofactors:

$$\text{Adj(A)} = \begin{bmatrix} |C_{11}| & |C_{12}| & \cdots & |C_{1n}| \\ |C_{21}| & |C_{22}| & \cdots & |C_{2n}| \\ \vdots & \vdots & \ddots & \vdots \\ |C_{n1}| & |C_{n2}| & \cdots & |C_{nn}| \end{bmatrix}' = \begin{bmatrix} |C_{11}| & |C_{21}| & \cdots & |C_{n1}| \\ |C_{12}| & |C_{22}| & \cdots & |C_{n2}| \\ \vdots & \vdots & \ddots & \vdots \\ |C_{1n}| & |C_{2n}| & \cdots & |C_{nn}| \end{bmatrix}$$

Theorem: If A^{-1} exists, then $A^{-1} = \dfrac{\text{adj}(A)}{|A|}$.

Proof: <omitted>

Ex: Let A = $\begin{bmatrix} 12 & 1 \\ -8 & 1 \end{bmatrix}$. $|A|$ = 20, so $A^{-1} = \dfrac{\text{adj}(A)}{|A|} = \dfrac{1}{20}\begin{bmatrix} |C_{11}| & |C_{21}| \\ |C_{12}| & |C_{22}| \end{bmatrix} =$

$$\frac{1}{20}\begin{bmatrix} (-1)^{1+1}1 & (-1)^{2+1}1 \\ (-1)^{1+2}(-8) & (-1)^{2+2}12 \end{bmatrix} = \frac{1}{20}\begin{bmatrix} 1 & -1 \\ 8 & 12 \end{bmatrix}$$

<u>Comment</u>. The approach to finding matrix inverses through co-factors and adjoints is not very practical. Here is a much more practical approach: Suppose that *A* is nonsingular. We seek to find a square matrix B such that *A B* = *I*, where *I* is the identity matrix. The idea is to apply operation IV (adding multiples of one column or row to another column or row) repeatedly to *A* and thereby reduce *A* to a diagonal matrix (not just upper triangular, but lower triangular as well). At the same time, apply the identical operations to *I* in the same order. Each operation IV is equivalent to a matrix multiplication. For example, subtracting 2/7 x row 1 from row 2 of *A* = $\begin{bmatrix} 7 & -3 & -3 \\ 2 & 4 & 1 \\ 0 & -2 & -1 \end{bmatrix}$ is equivalent to $\begin{bmatrix} 1 & 0 & 0 \\ -2/7 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 7 & -3 & -3 \\ 2 & 4 & 1 \\ 0 & -2 & -1 \end{bmatrix} = \begin{bmatrix} 7 & -3 & -3 \\ 0 & 34/7 & 13/7 \\ 0 & -2 & -1 \end{bmatrix}$. For every such simplifying multiplication of *A* by a matrix E_i, do the same to I. After *m* such steps, you will have $E_m E_{m-1} \cdots E_2 E_1 AB = E_m E_{m-1} \cdots E_2 E_1 I$ where $E_m E_{m-1} \cdots E_2 E_1 A = D$ is now diagonal. Finally, multiply both sides by D^{-1}, in which each diagonal element is the inverse of the diagonal of *D* (how do you know each diagonal element of *D* is nonzero?). You get $B = D^{-1} E_m E_{m-1} \cdots E_2 E_1 I$, in which the inverse of *A* is now on the right-hand side – the final result of applying to *I* all of those simplifications that you applied to *A*. After you have turned *A* into *I*, you have turned *I* into *A* inverse.

Cramer's Rule

Cramer's Rule: Suppose A_{nxn} is nonsingular. Then the unique solution $x = A^{-1}d$ for the system of simultaneous linear equations $Ax = d$ can be found as $\begin{Bmatrix} x_1 = |A_1|/|A| \\ x_2 = |A_2|/|A| \\ \vdots \\ x_n = |A_n|/|A| \end{Bmatrix}$, where A_i is obtained by substituting d for column i of A.

Note: Cramer's Rule solves the system of simultaneous linear equations one x_i at a time. Each x_i requires evaluating a new determinant ($|A_i|$). The matrix solution $x = A^{-1}d$ yields all solutions at once. But the matrix solution requires finding the inverse, which is hard, whereas Cramer's rule requires evaluating determinants, which is easier, but there may be many determinants to evaluate.

Ex: (Partial market equilibrium)
Let P be price and Q be quantity. Suppose the demand and supply equations are $\begin{Bmatrix} 8P + Q = 120 \\ -12P + Q = -80 \end{Bmatrix}$. Then $A = \begin{bmatrix} 8 & 1 \\ -12 & 1 \end{bmatrix}$, $A_1 = \begin{bmatrix} 120 & 1 \\ -80 & 1 \end{bmatrix}$, $A_2 = \begin{bmatrix} 8 & 120 \\ -12 & -80 \end{bmatrix}$.
So $|A| = 20$, $|A_1| = 200$, $|A_2| = 800$. Then $\begin{Bmatrix} P = |A_1|/|A| = 200/20 = 10 \\ Q = |A_2|/|A| = 800/20 = 40 \end{Bmatrix}$

Ex: (National income analysis)
Let C be national consumption, Y be national income, I be investment (given), and G be government expenditure (given). Suppose the model is $\begin{Bmatrix} C = a + bY \\ Y = C + I_0 + G_0 \end{Bmatrix}$ = $\begin{Bmatrix} -bY + C = a \\ Y - C = I_0 + G_0 \end{Bmatrix}$. Then $A = \begin{bmatrix} -b & 1 \\ 1 & -1 \end{bmatrix}$, $x = \begin{bmatrix} Y \\ C \end{bmatrix}$, $d = \begin{bmatrix} a \\ I_0 + G_0 \end{bmatrix}$. $A_1 = \begin{bmatrix} a & 1 \\ I_0 + G_0 & -1 \end{bmatrix}$, $A_2 = \begin{bmatrix} -b & a \\ 1 & I_0 + G_0 \end{bmatrix}$. So $|A| = b-1$, $|A_1| = -a - I_0 - G_0$. $|A_2| = -bI_0 - bG_0 - a$. Then
$\begin{Bmatrix} Y = |A_1|/|A| = (-a - I_0 - G_0)/(b-1) = (a + I_0 + G_0)/(1-b) \\ C = |A_2|/|A| = (-bI_0 - bG_0 - a)/(b-1) = (bI_0 + bG_0 + a)/(1-b) \end{Bmatrix}$, provided $b \neq 1$ (necessary for nonsingularity of A).

Ex: $\begin{Bmatrix} 7x_1 - x_2 - x_3 = 0 \\ 10x_1 - 2x_2 + x_3 = 8 \\ 6x_1 + 3x_2 - 2x_3 = 7 \end{Bmatrix}$. $|A| = \begin{vmatrix} 7 & -1 & -1 \\ 10 & -2 & 1 \\ 6 & 3 & -2 \end{vmatrix}$ = {Col3 = Col3 – Col2} $\begin{vmatrix} 7 & -1 & 0 \\ 10 & -2 & 3 \\ 6 & 3 & -5 \end{vmatrix}$ =

$\{Col1 = Col1 + 7xCol2\} \begin{vmatrix} 0 & -1 & 0 \\ -4 & -2 & 3 \\ 27 & 3 & -5 \end{vmatrix} = \{\text{expand Row1}\} \ -1(-1)^{1+2} \begin{vmatrix} -4 & 3 \\ 27 & -5 \end{vmatrix} = 20\text{-}81 = \text{-}61.$

$|A_1| = \begin{vmatrix} 0 & -1 & -1 \\ 8 & -2 & 1 \\ 7 & 3 & -2 \end{vmatrix} = \{Col2 = Col2 - Col3\} \begin{vmatrix} 0 & 0 & -1 \\ 8 & -3 & -1 \\ 7 & 5 & -2 \end{vmatrix} = \{\text{expand Row1}\} \ -1(-1)^{1+3} \begin{vmatrix} 8 & -3 \\ 7 & 5 \end{vmatrix}$

= -(40 – 21) = -61

$|A_2| = \begin{vmatrix} 7 & 0 & -1 \\ 10 & 8 & 1 \\ 6 & 7 & -2 \end{vmatrix} = \{Col1 = Col1 + 7xCol3\} \begin{vmatrix} 0 & 0 & -1 \\ 17 & 8 & 1 \\ -8 & 7 & -2 \end{vmatrix} = \{\text{expand Row1}\}$

$-1(-1)^{1+3} \begin{vmatrix} 17 & 8 \\ -8 & 7 \end{vmatrix}$ = -(119 + 64) = -183.

$|A_3| = \begin{vmatrix} 7 & -1 & 0 \\ 10 & -2 & 8 \\ 6 & 3 & 7 \end{vmatrix} = \{Col1 = Col1 + 7xCol2\} \begin{vmatrix} 0 & -1 & 0 \\ -4 & -2 & 8 \\ 27 & 3 & 7 \end{vmatrix} = \{\text{expand Row1}\}$

$-1(-1)^{1+2} \begin{vmatrix} -4 & 8 \\ 27 & 7 \end{vmatrix}$ = -28 – 216 = -244.

Thus, $\begin{Bmatrix} x_1 = |A_1|/|A| = -61/(-61) = 1 \\ x_2 = |A_2|/|A| = -183/(-61) = 3 \\ x_3 = |A_3|/|A| = -244/(-61) = 4 \end{Bmatrix}$.

Definition: The system of simultaneous linear equations $Ax = 0$ is said to be **homogeneous**.

Theorem: If A is nonsingular, then the unique solution for the homogeneous system of simultaneous linear equations $Ax = 0$ is $x = 0$.
Proof: <omitted>

Theorem: If A is singular, then there are infinitely many solutions for the homogeneous system of simultaneous linear equations $Ax = 0$.
Proof: <omitted>

Definition: The system of simultaneous linear equations $Ax = d$ is said to be **inconsistent** if d is not in the span of the columns of A.

If $Ax = d$ is inconsistent, then there is no solution, because there is no linear combination of the columns of A (Ax) that can produce d. If If $Ax = d$ is consistent, then there is at least one solution, because there must be a linear combination of the columns of A (Ax) that can produce d.

Complete classification of solutions for $Ax = d$:

- **If A is nonsingular and $d \neq 0$:** There is one and only one solution, and it can be calculated either by $x = A^{-1}d$ or by Cramer's Rule.
- **If A is nonsingular and $d = 0$:** There is one and only one solution, and it is $x = 0$.
- **If A is singular and $d \neq 0$:**
  (a) If the equations $Ax = d$ are consistent, then there are infinitely many solutions. But neither $x = A^{-1}d$ nor Cramer's Rule can be used to find them, because A^{-1} does not exist and $|A| = 0$.
  (b) If the equations $Ax = d$ are inconsistent, then there is no solution.
- **If A is singular and $d = 0$:** There are infinitely many solutions. One solution is $x = 0$. But neither $x = A^{-1}d$ nor Cramer's Rule can be used to find them, because A^{-1} does not exist and $|A| = 0$. In this case, the equations $Ax = 0$ cannot be inconsistent, since there is a linear combination ($A0$) that produces 0.

Eigenvalues and Eigenvectors

<u>Definition</u>: Suppose A is nxn and $Ax = \lambda x$ for some scalar λ and vector x. Then λ is called an **eigenvalue** (characteristic root, latent root) of A and x is called an **eigenvector** (characteristic vector, latent vector) of A.

An interpretation of eigenvalue and eigenvector: We know from the preceding discussion that the system of simultaneous linear equations $Ax = d$ has a solution if and only if d lies in the span of A. An eigenvector is a special kind of *d that is its own solution* (up to a constant of proportionality). That is, $\frac{1}{\lambda}Ad = d$. Such d's must be – and are – rare in the span of A. But when you find one, the entire set A of coefficients in the linear combination Ax can be replaced by a single number, λ, that represents the entire matrix A: $Ax = \lambda x$. Moreover, the magnitude of the eigenvalue must be related somehow to the size of A. This is easy to see from the observation that if $Ax = \lambda x$ then $(cA)x = (c\lambda)x$. That is, if you multiply all of the entries in the matrix A by the constant c, then the eigenvalue grows in proportion. The use of eigenvalues to assess the "size" of subspaces spanned by associated eigenvectors will be an important development in principal components analysis and factor analysis.

<u>Note</u>: If $Ax = \lambda x$ then $Ax - \lambda x = 0$, or $(A - \lambda)x = 0$, or $(A - \lambda I_n)x = 0$. $(A - \lambda I_n)$ is called the **characteristic matrix** of A. $(A - \lambda I_n)x = 0$ is a system of simultaneous linear equations. From the preceding discussion, we know that in order for this system to have a nonzero solution, $(A - \lambda I_n)$ must be singular. That is, $|A - \lambda I_n| = 0$. This observation provides a way to calculate the eigenvalues of A.

<u>Definition</u>: $|A - \lambda I_n| = 0$ is called the **characteristic equation** of A.

Upon expansion, $|A-\lambda I_n|=0$ yields a polynomial in λ of the n^{th} degree. Therefore, there will be n solutions for λ. That is, there are n eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$. For each eigenvalue λ_i, there is a corresponding eigenvector x_i: $(A-\lambda_i I_n)x_i = 0$. In fact there are infinitely many possible solutions for x_i, but they are all proportional to each other.

Ex: A = $\begin{bmatrix} 8 & 1 \\ -12 & 1 \end{bmatrix}$. $|A-\lambda I_2|$ = $\left|\begin{bmatrix} 8 & 1 \\ -12 & 1 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right|$ = $\begin{vmatrix} 8-\lambda & 1 \\ -12 & 1-\lambda \end{vmatrix}$ = $(8-\lambda)(1-\lambda) + 12$ = $\lambda^2 - 9\lambda + 20$. Set = 0 to solve for eigenvalues: $(\lambda - 5)(\lambda - 4) = 0$. So $\lambda_1 = 5$ and $\lambda_2 = 4$. (Eigenvalues are usually listed in order of decreasing value.) (Note also that 20 = $\lambda_1\lambda_2$ = |A| -- in general the determinant of a matrix equals the product of its eigenvalues.)

Ex: To get an eigenvector for $\lambda_1 = 5$, we solve $(A-5I_2)x_1 = 0$. That is,
$\left(\begin{bmatrix} 8 & 1 \\ -12 & 1 \end{bmatrix} - \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}\right)\begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix} = 0$, or $\begin{bmatrix} 3 & 1 \\ -12 & -4 \end{bmatrix}\begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix} = 0$. Note that the coefficient matrix is singular. So any $\begin{bmatrix} x_{11} & x_{21} \end{bmatrix}$ satisfying $3x_{11} + x_{21} = 0$ will work. It is standard practice to choose a solution with vector length = 1, that is $x_{11}^2 + x_{21}^2 = 1$. Hence, $x_{11} = 1/\sqrt{10}$ and $x_{21} = -3/\sqrt{10}$. (To get this, note that $x_{11} = 1$, $x_{21} = -3$ is a solution. And the length of this solution vector is $\sqrt{1^2 + (-3)^2} = \sqrt{10}$. Divide the solution by the length to get a solution vector with length 1.)

Ex: To get an eigenvector for $\lambda_2 = 4$, we solve $(A-4I_2)x_2 = 0$. That is,
$\left(\begin{bmatrix} 8 & 1 \\ -12 & 1 \end{bmatrix} - \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}\right)\begin{bmatrix} x_{12} \\ x_{22} \end{bmatrix} = 0$, or $\begin{bmatrix} 4 & 1 \\ -12 & -3 \end{bmatrix}\begin{bmatrix} x_{12} \\ x_{22} \end{bmatrix} = 0$. Note that the coefficient matrix is singular. So any $\begin{bmatrix} x_{12} & x_{22} \end{bmatrix}$ satisfying $4x_{12} + x_{22} = 0$ will work. It is standard practice to choose a solution with vector length = 1, that is $x_{12}^2 + x_{22}^2 = 1$. Hence, $x_{12} = 1/\sqrt{17}$ and $x_{22} = -4/\sqrt{17}$.

Ex: Suppose that $A = I_n$. To find the eigenvalues of I_n, solve $|I_n - \lambda I_n| = 0$, i.e., $(1-\lambda)^n |I_n| = 0$ or $(1-\lambda)^n = 0$. Hence, the n eigenvalues are all identical and equal to 1. In this case the eigenvectors are undetermined, because $(I_n - \lambda_i I_n)x_i = 0$ becomes $0x_i = 0$ for each eigenvalue. One could therefore conveniently take the eigenvectors to be the unit basis vectors $e_1, e_2, ..., e_n$.

Eigenvalues and eigenvectors are very important in multivariate analysis – especially for dimensionality reduction, data compression, and interpretation of hidden data factors. A few places where eigenvalues/eigenvectors are used:

- Principal components analysis in statistics
- Factor analysis in statistics
- Measure of multivariate variance in statistics
- Test for multicollinearity in regression
- Optimization (in calculus)

The Matrix Approach to Linear Regression

Ordinary Least Squares (OLS) involves fitting a linear equation to observed data.

The linear specification for the data in a regression model with k - 1predictors is

$$y_1 = \beta_1 + \beta_2 x_{12} + \beta_3 x_{13} + \cdots + \beta_k x_{1k} + \varepsilon_1$$
$$y_2 = \beta_1 + \beta_2 x_{22} + \beta_3 x_{23} + \cdots + \beta_k x_{2k} + \varepsilon_2$$
$$\vdots$$
$$y_n = \beta_1 + \beta_2 x_{n2} + \beta_3 x_{n3} + \cdots + \beta_k x_{nk} + \varepsilon_n$$

These n equations can be re-written in matrix notation as

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{12} & x_{13} & \cdots & x_{1k} \\ 1 & x_{22} & x_{23} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n2} & x_{n3} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Or more compactly as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

How to estimate the parameters $\boldsymbol{\beta}$?

From the linear specification, we can solve for the errors:

$$\varepsilon_1 = y_1 - \beta_1 - \beta_2 x_{12} - \beta_3 x_{13} - \cdots - \beta_k x_{1k}$$
$$\varepsilon_2 = y_2 - \beta_1 - \beta_2 x_{22} - \beta_3 x_{23} - \cdots - \beta_k x_{2k}$$
$$\vdots$$
$$\varepsilon_n = y_n - \beta_1 - \beta_2 x_{n2} - \beta_3 x_{n3} - \cdots - \beta_k x_{nk}$$

For least-squares estimation, we minimize the Error Sum of Squares ESS = $\sum \varepsilon_i^{\,2}$ = $\sum (y_i - \beta_1 - \beta_2 x_{i2} - \beta_3 x_{i3} - \cdots - \beta_k x_{ik})^2$. But $\sum \varepsilon_i^{\,2}$ = $\boldsymbol{\varepsilon}' \cdot \boldsymbol{\varepsilon}$ = $(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$, which displays ESS explicitly in terms of the quantities $\boldsymbol{\beta}$ that we are allowed to vary in the minimization. ESS is a 1x1 matrix. That is, ESS is a scalar. To find its minimum, we differentiate it partially with respect to each of the k β 's and set the k partial derivatives = 0. Collecting the k partial derivatives together in the form of a column vector and setting them equal to a column vector of zeroes, we have $2\mathbf{X}'\mathbf{y} - 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{0}$ or $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}$. This is the matrix representation of the Normal Equations.

The Normal Equations in matrix form: $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}$

If the matrix $\mathbf{X}'\mathbf{X}$ is non-singular (has an inverse), then we can solve the Normal Equations for the ESS-minimizing values of $\boldsymbol{\beta}$. This gives us the least-squares estimate of $\boldsymbol{\beta}$:

$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ or $\boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$.

The least-squares estimates of $\boldsymbol{\beta}$ in matrix form: $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$

If the matrix $\mathbf{X'X}$ is singular, then we have strict **multicollinearity**, which means there is no unique solution for OLS. (There are infinitely many solutions.) Even when $\mathbf{X'X}$ is non-singular, there can be a variety of problems if $\mathbf{X'X}$ is "almost" singular. Then we have **near multicollinearity**, resulting in inflated variances of the estimated coefficients, and other problems.

Note: Use of coded categorical predictor variables (like state of residence – e.g., TX, LA, AR, etc.) in $\mathbf{X}$ must be undertaken with care. The recommended approach for a categorical predictor variable is to create one **dummy variable** (coded 0 and 1) for each category. (Ex: One 0-1 column for TX, and one 0-1 column for LA, etc.). But if every category has a separate dummy column, then $\mathbf{X'X}$ is automatically singular, so there is a perfect multicollinearity and there is no unique least squares estimate. Therefore, one of the dummy columns is usually dropped, or additional conditions are imposed to force a unique solution. OLS with use of such coded categorical predictors is called **analysis of variance** (when all *x*-variables are 0-1) or **analysis of covariance** (when at least one *x*-variable is 0-1 and at least one is not).

Application to Simple Linear Regression

Suppose the simple linear regression model holds (only one predictor). Let us find the least squares estimates of $\boldsymbol{\beta}$ in matrix form. The linear specification for the simple linear regression model is:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

The least squares estimates are $\boldsymbol{\beta} = (\mathbf{X'X})^{-1}\mathbf{X'y}$, which I now calculate: Now $\mathbf{X'X}$ =

$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}$, and $\mathbf{X'y}$ $= \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$. So the Normal Equations are

$\mathbf{X'X\beta} = \mathbf{X'y}$ or $\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$ or $\begin{bmatrix} \beta_1 n + \beta_2 \sum x_i \\ \beta_1 \sum x_i + \beta_2 \sum x_i^2 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$. The inverse of $\mathbf{X'X}$ exists as long as $s_x^2 > 0$ and is $(\mathbf{X'X})^{-1} = \begin{bmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{bmatrix} \dfrac{1}{n(\sum x_i^2 - n\bar{x}^2)}$. [As an aside, it is worth noting that the determinant of $\mathbf{X'X}$ is proportional to the variance of $x_1, x_2, \ldots, x_n$.] So the least-squares estimates are

$$\hat{\boldsymbol{\beta}} = \frac{1}{n(\sum x_i^2 - n\bar{x}^2)} \begin{bmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{bmatrix} \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

After multiplying this out and simplifying, we find that it yields the same formulas for the estimates of intercept β_1 and slope β_2 that are given by direct differentiation in OLS, namely,

$$\hat{\beta}_1 = \bar{y} - \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^{\,2} - n\bar{x}^2}\,\bar{x} \text{ and } \hat{\beta}_2 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^{\,2} - n\bar{x}^2}$$

For regressions that involve more than 2 or 3 predictors, it is almost essential to use the matrix approach. Otherwise, the formulas become too unwieldy. In addition, the matrix approach increases the power of regression analysis by employing the tools of matrix theory.