Matrix Algebra & Calculus Review

A matrix is a rectangular array of numbers (real numbers for the purposes of most econometrics)

$$\underset{n \times k}{X} = [x_{ik}] = X_{ik} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ x_{n1} & \cdots & \cdots & x_{nk} \end{bmatrix}$$

Matrix X has dimensions $n \times k$. That is, n rows and k columns. Also, $X \in \mathbb{R}^{n \times k}$ In terms of data, the rows will often be
observations and the columns will be variables (also called features).
If $n = k$ we have a square matrix.

Matrices are composed of vectors, which are ordered sets of numbers arranged either in a row or column.

Symmetric matrix: One in which $x_{ik} = x_{ki}$ $\forall\ i, k$

Diagonal matrix: One with zeroes everywhore but the main diagonal.

Scalar matrix: A Diagonal matrix where all the elements in the main diagonal are the same.

Identity Matrix: A Diagonal matrix where the main diag. has all ones.

Zero or Null matrix: A matrix with zeroes everywhere. Works the same way as a univariate zero.

Some Basic Algebra

Equality: Two matrices, X and Y, are equal if they have the same dimensions and all elements are the same

$$X = Y \text{ iff } x_{ik} = y_{ik} \quad \forall i,k$$

Transposition: The transpose of matrix X, denoted X^T or X', is obtained by "flipping" X over its diagonal.

$$(A^T)_{ik} = A_{ki}$$

If X is symmetric, Then $X^T = X$ and so $(X^T)^T = X$

Addition: Holds only if the matrices have the same dimensions. Let $X, Y \in \mathbb{R}^{n \times k}$

$$X + Y = [x_{ik} + y_{ik}]$$

$$X - Y = [x_{ik} - y_{ik}]$$

Addition is Commutative and Associative

$$X + Y = Y + X$$

$$(X + Y) + C = X + (Y + C)$$

$$(X + Y)^T = X^T + Y^T$$

Multiplication: Holds only if the left matrix has the same number of columns as the right one has rows.

Let $X \in \mathbb{R}^{n \times k}$ and $Y \in \mathbb{R}^{k \times m}$

$$X \cdot Y = x_i^T y_k = C$$

The ik^{th} element of C, $c_{ik} \in C$, is the inner product (or dot product) of row i in X and column k in Y.

The inner product returns a scalar,

and it's defined for two vectors as:

Let $x \in \mathbb{R}^{1 \times k}$ and $y \in \mathbb{R}^{k \times 1}$

$$x^T y = x_1 y_1 + x_2 y_2 + \cdots + x_k y_k$$

The dot product is Commutative

$$x^T y = y^T x$$

But Matrix multiplication is often not.

$$XY \neq YX \quad \text{, generally}$$

Scalar multiplication simply multiplies each element in the matrix by the scalar

Let $c \in \mathbb{R}$ and $X \in \mathbb{R}^{n \times k}$

$$cX = [c \cdot x_{ik}]$$

In this class, we will often multiply a matrix times a vector. For example, a linear system problem is defined as

$$Ab = c$$

where $A \in \mathbb{R}^{n \times k}$, $b \in \mathbb{R}^{k \times 1}$, $c \in \mathbb{R}^{n \times 1}$

Ex.

$$\begin{bmatrix} 4 & 2 & 1 \\ 2 & 6 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 5 \\ 4 \\ 1 \end{bmatrix}$$

We wont to find b.

$$4a + 2b + c = 5$$

...

$$2a + 6b + c = 4$$

$$a + b + 0 = 1$$

OR

$$a\begin{bmatrix}4\\2\\1\end{bmatrix} + b\begin{bmatrix}2\\6\\1\end{bmatrix} + c\begin{bmatrix}1\\1\\0\end{bmatrix} = \begin{bmatrix}5\\4\\1\end{bmatrix}$$

So, the solution vector C is a linear combination of the data matrix A and unknown vectors.

There are many ways to find the unknowns, OLS is just one more. We will study how it works next class. But keep this LSP in mind.

Matrix multiplication is Associative and Distributive

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

$$A(B+C) = AB + AC$$

Also, transposition has useful properties for proofs

$$(AB)^T = B^T A^T$$

$$(ABC)^T = C^T B^T A^T$$

Remember it by thinking of a Carousell.

Exercise: Denote $\sum_{i=1}^{n} x_i$ in matrix form

- $i^T x$, define dimensions

Exercise: Now denote the sample mean

$\frac{1}{n} \sum x_i$

- $\frac{1}{n} i^T x$

Exercise: Denote Σx_i^2 in matrix form

- $x^T x$

Idempotent matrix: Very common in Regression This is a matrix who's square yields the original matrix

$$M^2 = M$$

It can be used to compute deviations from the mean.

$i \in \mathbb{R}^{n \times 1}$ $\quad \bar{x} \in \mathbb{R}$ $\quad x \in \mathbb{R}^{n \times 1}$

$$i \bar{x} = i \frac{1}{n} i^T x = \frac{1}{n} i i^T x = \begin{bmatrix} \bar{x} \\ \bar{x} \\ \vdots \\ \bar{x} \end{bmatrix}$$

$\frac{1}{n} i i^T$ is an $n \times n$ matrix where every element equals $\frac{1}{n}$. Then,

$$[x - i\bar{x}] = [x - \frac{1}{n} i i^T x] = \begin{bmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{bmatrix}$$

is the set of deviations. Since $x = Ix$

$$[x - \frac{1}{n} i i^T x] = [Ix - \frac{1}{n} i i^T x] = [I - \frac{1}{n} i i^T] x$$

$$= M^0 x$$

M^0 has $(1 - \frac{1}{n})$ in its diagonal and $-\frac{1}{n}$ everywhere else.

We can use M^0 to compute squared deviations

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = \left(\sum_{i=1}^{n} x_i^2 \right) - n\bar{x}^2$$

In Matrix form,

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = (x - \bar{x}i)^T (x - xi)$$

$$= (M^0 x)^T (M^0 x)$$

$$= x^T M^{0T} M^0 x$$

M^0 is symmetric and idempotent

$M^0 \cdot M^0 = M^0$

Quadratic forms: Useful in distinguishing maxima from minima, and checking 2nd order properties like concavity | convexity

$$q = \sum_{i=1}^{n} \sum_{k=1}^{n} x_i x_k a_{ik} = x^T A x$$

where A is symmetric. One practical example is a production function with two inputs x_1, x_2

$$y = \theta + \lambda_1 x_1 + \lambda_2 x_2 + a_{11} x_1^2 + a_{22} x_2^2 + a_{12} x_1 x_2$$

$$y = \theta + \lambda^T x + x^T A x$$

If $x^T A x > (<)\ 0$ for all non-zero x then it is positive (negative) definite

We will talk about matrix inversion when we discuss OLS.

Matrix Calculus

Scalar wrt Vector: Let $y = f(x)$, $x \in \mathbb{R}^{1 \times K}$

We compute the gradient

$$\nabla f(x) = \frac{\partial f(x)}{\partial x} = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_K} \end{bmatrix} = g$$

which yields a vector of partial derivatives

As a general rule of thumb, the shape of the derivatieve (vector, scalar, matrix) is determined by the shape of the denominator of the derivative.

Hessian: A square and symmetric matrix of 2^{nd} order derivatives. The properties of the Hessian are useful for checking

conditions of a fuction like the OLS estimator.

$$
H = \begin{bmatrix} \dfrac{\partial^2 f(x)}{\partial x_1 \partial x_1} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_1 \partial x_K} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f(x)}{\partial x_K \partial x_1} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_K \partial x_K} \end{bmatrix}
$$

Each column in H is the derivative of gradient g with respect to the corresponding variable in x^T. That is,

$$
\begin{aligned}
H &= \begin{bmatrix} \dfrac{\partial(\partial f(x)/\partial x)}{x_1} & \cdots & \dfrac{\partial(\partial f(x)/\partial x)}{x_K} \end{bmatrix} \\
&= \frac{\partial(\partial f(x)/\partial x)}{\partial(x_1 \cdots x_K)} = \frac{\partial(\partial f(x)/\partial x)}{\partial x^T} \\
&= \frac{\partial^2 f(x)}{\partial x \, \partial x^T}
\end{aligned}
$$

Dot Product : We con construct a linear function as $y = a^T x = \sum_{i=1}^{n} a_i x_i$

Since a is a vector of constants/weights/

coefficients, you should expect that the derivative of the dot product wrt x yields a

$$\frac{\partial (a^T x)}{\partial x} = a$$

If we have a Matrix, rather than a vector, of coefficients A then

$$\frac{\partial A x}{\partial x} = A^T$$

Quadratic form: Recall we can write

this as $x^T A x = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j a_{ij}$

Ex. $A = \begin{bmatrix} 1 & 3 \\ 3 & 4 \end{bmatrix} \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$$x^T A x = \sum_{i=1}^{2} \sum_{j=1}^{2} a_{ij} x_i x_j$$

$$= 1 x_1 x_1 + 3 x_1 x_2 + 3 x_2 x_1 + 4 x_2 x_2$$

$$= x_1^2 + 4 x_2^2 + 6 x_1 x_2$$

$$\frac{\partial x^T A x}{\partial x} = \begin{bmatrix} 2x_1 + 6x_2 \\ 8x_2 + 6x_1 \end{bmatrix} = \begin{bmatrix} 2 & 6 \\ 6 & 8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$= 2Ax$ (General result if A is symmetric)

If A not symmetric

$$\frac{\partial x^T A x}{} = (A + A^T) x$$

∂x

wrt the coefficients

$$\frac{\partial x^T A x}{\partial A} = x x^T$$