Mathematical Economics (ECON 471)
Lecture 2
Revision of Matrix Algebra

Teng Wah Leo

The need to use linear algebra and matrices in economics is principally borne out of systems of equations when dealing in multiple sectors within an economy in macroeconomics, or the various possible outcomes that an economy or an individual could realize in both macro− and micro− economics, and most commonly in econometrics. We will now revise some key ideas and concepts you may have done, as well as learn some new ones.

1 Basics in Matrix Algebra

1.1 System of Linear Equations & their Relationship to Matrix Algebra

Firstly, a general system of k linear equations each with r common variables is just,

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + & \cdots + a_{1r}x_r = & b_1 \\
a_{21}x_1 + a_{22}x_2 + & \cdots + a_{2r}x_r = & b_2 \\
\vdots \qquad & \qquad \vdots & \vdots \\
a_{k1}x_1 + a_{k2}x_2 + & \cdots + a_{kr}x_r = & b_k
\end{aligned}$$

This system can always be written in matrix form,

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1r} \\ a_{21} & a_{22} & \dots & a_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \dots & a_{kr} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_r \end{bmatrix}$$

which in term can be written even more succinctly as,

$$\mathbf{Ax} = \mathbf{b}$$

If this is not familiar to you. It will be shortly. The matrix $\mathbf{A}$ is referred to as the **coefficient matrix**. When the vector $\mathbf{b}$ is a column of zeros, the system of equation is said to be **homogeneous**. Such systems of equation has at the least one solution, when all the x's are zeros.

Beginning with terminology with reference to the above, a matrix is said to have k **leading zeros** if the first k elements in that are zeros, and the first nonzero element is the $(k+1)^{\text{th}}$ element of that row. A matrix is said to be in **row echelon form** if each succeeding row (from the first row) has more leading zeros than the last. An example you should be familiar with is the **identity matrix**,

$$\begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

Likewise, the **zero matrix**,

$$\begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}$$

is in row echelon form. The first nonzero element of a matrix in row echelon form is called the **pivot**. Finally, a **reduced row echelon form** matrix is one in which each pivot is 1, and every column containing a pivot has no other nonzero element, and examples of which are the identity and zero matrix.

The terminologies and definitions above are important in understanding whether a system of equations has a solution, and if it does, whether the solution is unique or infinitely many. One of the most important ideas, which you may recall, is the **rank** of a matrix and it refers to the number of nonzero rows in a row echelon form matrix. It is a fact that the rank of a matrix $\mathbf{A}$, **rankA** is less than or equal to the number of rows and/or columns of $\mathbf{A}$. The matrix equation or the system of equations it represents has a unique solution if the coefficient matrix is **nonsingular**, where such a matrix has equal columns and rows (**a square matrix**) and are equal to the rank of the matrix. This corresponds to the idea or fact that uniqueness in the solution of a system of equation occurs when you have as many equations as you have variables, which corresponds with the number of rows and columns. You should read your text for a lengthier discussion.

1.2 Matrix Algebra

Let $\mathbf{A}$ and $\mathbf{B}$ be two $k \times r$ matrices with typical elements a_{ij} and b_{ij} respectively, where $i \in \{1, 2, \ldots, k\}$ and $j \in \{1, 2, \ldots, r\}$. The sum of the two matrices $\mathbf{A} + \mathbf{B}$ is then,

$$\begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \ldots & a_{1r} + b_{1r} \\ a_{21} + b_{21} & a_{22} + b_{22} & \ldots & a_{2r} + b2r \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} + b_{k1} & a_{k2} + b_{k2} & \ldots & a_{kr} + b_{kr} \end{bmatrix}$$

Similarly, $\mathbf{A} - \mathbf{B}$, we get,

$$\begin{bmatrix} a_{11} - b_{11} & a_{12} - b_{12} & \ldots & a_{1r} - b_{1r} \\ a_{21} - b_{21} & a_{22} - b_{22} & \ldots & a_{2r} - b2r \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} - b_{k1} & a_{k2} - b_{k2} & \ldots & a_{kr} - b_{kr} \end{bmatrix}$$

Letting α be a scalar, then $\alpha\mathbf{A}$

$$\begin{bmatrix} \alpha a_{11} & \alpha a_{12} & \ldots & \alpha a_{1r} \\ \alpha a_{21} & \alpha a_{22} & \ldots & \alpha a_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha a_{k1} & \alpha a_{k2} & \ldots & \alpha a_{kr} \end{bmatrix}$$

The multiplication of two matrices are however marginally more complicated. Let $\mathbf{C}$ be a $r \times k$ matrix with typical element c_{ij} where $i \in \{1, 2, \ldots, k\}$ and $j \in \{1, 2, \ldots, r\}$. Then $\mathbf{A}.\mathbf{C}$ is,

$$\begin{bmatrix} \sum_{q=1}^{r} a_{1q}c_{q1} & \sum_{q=1}^{r} a_{1q}c_{q2} & \cdots & \sum_{q=1}^{r} a_{1q}c_{qr} \\ \sum_{q=1}^{r} a_{2q}c_{q1} & \sum_{q=1}^{r} a_{2q}c_{q2} & \cdots & \sum_{q=1}^{r} a_{2q}c_{qr} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{q=1}^{r} a_{kq}c_{q1} & \sum_{q=1}^{r} a_{kq}c_{q2} & \cdots & \sum_{q=1}^{r} a_{kq}c_{qr} \end{bmatrix}$$

Note further that

$$\begin{aligned} \mathbf{A}.\mathbf{I}_{r \times r} &= \mathbf{A} \\ \mathbf{I}_{k \times k}.\mathbf{A} &= \mathbf{A} \end{aligned}$$

where as before $\mathbf{A}$ is a $k \times r$ matrix.

You should recall the following laws of Matrix Algebra as well,

1. **Associative Laws**:

$$\begin{aligned} (\mathbf{A} + \mathbf{B}) + \mathbf{C} &= \mathbf{A} + (\mathbf{B} + \mathbf{C}) \\ (\mathbf{AB})\mathbf{C} &= \mathbf{A}(\mathbf{BC}) \end{aligned}$$

2. **Commutative Law for Additions**:

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$$

3. **Distributive Laws**:

$$\begin{aligned} \mathbf{A}(\mathbf{B} + \mathbf{C}) &= \mathbf{AB} + \mathbf{AC} \\ (\mathbf{A} + \mathbf{B})\mathbf{C} &= \mathbf{AC} + \mathbf{BC} \end{aligned}$$

An important and commonly used operation in matrix algebra is the **transpose** which is simply the interchanging of the rows and columns of a matrix. The transpose of a matrix is denoted with a superscript T, such as $\mathbf{A}^T$, or sometimes simply with a prime, such as $\mathbf{A}'$. To be precise,

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1r} \\ a_{21} & a_{22} & \dots & a_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \dots & a_{kr} \end{bmatrix}^T = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{k1} \\ a_{12} & a_{22} & \dots & a_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1r} & a_{2r} & \dots & a_{kr} \end{bmatrix}$$

The use of the transpose operator follows these rules,

1. $(\mathbf{A}+\mathbf{B})^{\mathbf{T}} = \mathbf{A}^{\mathbf{T}} + \mathbf{B}^{\mathbf{T}}$
2. $(\mathbf{A}-\mathbf{B})^{\mathbf{T}} = \mathbf{A}^{\mathbf{T}} - \mathbf{B}^{\mathbf{T}}$
3. $(\mathbf{A}^{\mathbf{T}})^{\mathbf{T}} = A$
4. $(\mathbf{rA})^{\mathbf{T}} = \mathbf{rA}^{\mathbf{T}}$
5. $(\mathbf{AC})^{\mathbf{T}} = \mathbf{C}^{\mathbf{T}}\mathbf{A}^{\mathbf{T}}$

where both matrix $\mathbf{A}$ and $\mathbf{B}$ are $(k \times r)$ matrices, $\mathbf{C}$ is a $(r \times k)$ matrix and r is a scalar.

To complete our discussion on matrices, we will note here several kinds of special matrices,

1. **Square Matrix** is a matrix with the same number of rows and columns.
2. **Diagonal Matrix** is a matrix with nonzero elements on the diagonal, and zeros on the off diagonal,

$$\begin{bmatrix} d_{11} & 0 & \dots & 0 \\ 0 & d_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_{kk} \end{bmatrix}$$

3. **Upper Triangular Matrix** is a matrix (usually square) with zero element on the lower triangle off the diagonal.

$$\begin{bmatrix} d_{11} & d_{12} & \dots & d_{1k} \\ 0 & d_{22} & \dots & d_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_{kk} \end{bmatrix}$$

4. **Lower Triangular Matrix** is a matrix (usually square) with zero elements on the upper triangle above the diagonal.

$$\begin{bmatrix} d_{11} & 0 & \dots & 0 \\ d_{21} & d_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ d_{k1} & d_{k2} & \dots & d_{kk} \end{bmatrix}$$

5. **Symmetric Matrix** is a matrix where the upper and lower triangle has the same elements so that $\mathbf{A}^T = \mathbf{A}$.

$$\begin{bmatrix} d_{11} & d_{12} & d_{13} \\ d_{12} & d_{22} & d_{23} \\ d_{13} & d_{23} & d_{33} \end{bmatrix}$$

6. **Idempotent Matrix** is a matrix with the following property, $\mathbf{B}.\mathbf{B} = \mathbf{B}$. An example being $\mathbf{B} = \mathbf{I}$ or

$$\begin{bmatrix} d_{11} & -d_{11} \\ d_{12} & -d_{12} \end{bmatrix}$$

7. **Permutation Matrix** is a square matrix of zeros and ones in which each column and row has only one element which is a one. An example is the identity matrix.

8. **Nonsingular Matrix** is a square matrix whose rank is equal to the number of rows/columns. When the coefficient matrix is a nonsingular matrix, the system of equations has a unique solution.

Let $\mathbf{D}$ be a square matrix, then $\mathbf{D}^{-1}$ is the inverse of $\mathbf{D}$ if $\mathbf{D}.\mathbf{D^{-1}} = \mathbf{D^{-1}}.\mathbf{D} = \mathbf{I}$. If such an inverse matrix exists, we say that $\mathbf{D}$ is **invertible**. Further, there can be only one inverse for an invertible matrix. However, for a non-square matrix, say $\mathbf{A}$ with dimension $k \times r$, it is possible for it to have two different inverses. A matrix is the **left inverse** of $\mathbf{A}$, call it $\mathbf{B}$ with dimension $r \times k$, if $\mathbf{B}.\mathbf{A} = \mathbf{I}$. On the other hand, for a matrix $\mathbf{C}$ with dimension $r \times k$, it is the **right inverse** if $\mathbf{A}.\mathbf{C} = \mathbf{I}$. If a matrix $\mathbf{A}$ has a right ($\mathbf{B}$) and left ($\mathbf{C}$) inverse, then it is invertible and $\mathbf{B} = \mathbf{C} = \mathbf{A^{-1}}$. Finally, if an $k \times k$ matrix $\mathbf{A}$ is invertible $\Rightarrow$ it is nonsingular. In that case, the solution to the system of linear equation we had expressed prior is,

$$\begin{aligned}\mathbf{Ax} &= \mathbf{b} \\ \mathbf{A^{-1}Ax} &= \mathbf{A^{-1}b} \\ \mathbf{x} &= \mathbf{A^{-1}b}\end{aligned}$$

And similarly, if a square matrix $\mathbf{A}$ is nonsingular $\Rightarrow$ invertible. Additional properties of invertible matrices are as follows. Letting $\mathbf{A}$ and $\mathbf{B}$ be square matrices,

1. $(\mathbf{A}^{-1})^{-1} = A$
2. $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$
3. If both matrices are invertible, $(\mathbf{AB})^{-1} = \mathbf{A}^{-1}\mathbf{B}^{-1}$
4. If $\mathbf{A}$ is invertible, for any integer q, $(\mathbf{A}^q)^{-1} = (\mathbf{A}^{-1})^q = \mathbf{A}^{-q}$
5. For any integer p and q, $\mathbf{A^p A^q} = \mathbf{A^{p+q}}$
6. For any scalar s, where $s \neq 0$, $(s\mathbf{A})^{-1} = (1/s)\mathbf{A}^{-1}$

The actual calculation of an inverse is formulaic, and tedious as the dimension of the matrix increases. Before we can calculate the inverse of a matrix, we have to define the **determinant** and the **cofactor** of a matrix. Let $\mathbf{A}$ be a $k \times k$ square matrix, with typical element a_{ij}, where $i, j \in \{1, 2, \ldots, k\}$. The definition of a determinant is achieved inductively. For a scalar, a, the determinant of it is defined as $\det(a) = a$. For a 2×2 matrix,

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

the determinant is just,

$$\begin{aligned}\det(\mathbf{A}) &= a_{11}a_{22} - a_{12}a_{21} \\ &= a_{11}\det(a_{22}) - a_{12}\det(a_{21})\end{aligned}$$

More generally, for the $k \times k$ matrix case, let $\mathbf{A}_{ij}$ denote the $(k-1) \times (k-1)$ submatrix we obtain by deleting the i^{th} row and the j^{th} column from the matrix $\mathbf{A}$. We can then define the $(i, j)^{\text{th}}$ **minor** of $\mathbf{A}$ as,

$$m_{ij} \equiv \det(A_{ij})$$

and it is a scalar. Further, the **cofactor** of $\mathbf{A}$ is defined as,

$$c_{ij} \equiv (-1)^{i+j} m_{ij}$$

It should be clear the the sign of the cofactor is positive if $i + j$ is even, and negative otherwise. Then for a 3×3 matrix, the determinant is just

$$\begin{aligned} \det(\mathbf{A}) &= a_1 1 c_{11} + a_{12} c_{12} + a_{13} c_{13} \\ &= a_1 1 m_{11} - a_{12} m_{12} + a_{13} m_{13} \end{aligned}$$

For some matrices, the determinant is quite easy to calculate. For instance the determinant of a lower, upper, and diagonal matrix is just the product of its diagonal elements. The importance of the determinant is that if it is nonzero, then the matrix is then invertible/nonsingular, otherwise it is not. If you recall, it also figures in the actual calculation of the inverse of a matrix.

Before we define the calculation of an inverse, we need another special matrix, the **adjoint** matrix which is defined for a $k \times k$ matrix $\mathbf{A}$, written as $\operatorname{adj} A$, as the $k \times k$ matrix derived from matrix $\mathbf{A}$ where the $(i, j)^{\text{th}}$ element is the the cofactor c_{ji} of $\mathbf{A}$ (note the switch in indices for the cofactor). Then the inverse of $\mathbf{A}$ is,

$$\mathbf{A}^{-1} = \frac{1}{\det(A)} . \operatorname{adj} \mathbf{A}$$

To complete the discussion here, note the following properties of the determinant of $\mathbf{A}$ and another $k \times k$ matrix $\mathbf{B}$,

1. $\det \mathbf{A}^T = \det \mathbf{A}$
2. $\det(\mathbf{A}.\mathbf{B}) = \det \mathbf{A} . \det \mathbf{B}$, and
3. in general $\det((A + B)) \neq \det \mathbf{A} + \det \mathbf{B}$

The method is not difficult, albeit is quite tedious for large dimensional matrices. Fortunately for you and I, with increased computational power, these calculations can be easily performed by your personal computer.

2 Euclidean Spaces & Linear Independence

With the basics of matrices out of the way, we can begin to relate to vectors and the spaces they occupy. This is important when we are dealing with multiple variable calculus where in effect we are generalizing our basic two dimensional models of your intermediate economics models. But before we can explore multiple variable calculus and the various types of constrained optimization, we will explore the **Euclidean Spaces**.

In drawing a relationship between two variables in our earlier discussion, the variables are all located on $\mathbf{R}^1$. But you will recall, in your intermediate microeconomics class for instance, the choices made by the consumers are a couplet, and consequently, it is located on the **Cartesian plane** or alternatively known as the **Euclidean 2−Space**, written as $\mathbf{R}^2$. Of course when we incorporate more variables, the space they exist in expands. For example, in consider three variables, we would be in **Euclidean 3−Space**, written as $\mathbf{R}^3$, and so on and so forth (above 3, we would be hard pressed to depict it diagrammatically though).

Even if we cannot draw them, it does not prevent us from understand what happens in that space. It may be particularly useful in economics to understand how different combination of choices over k number of goods for example that an individual makes, in which case we would get $k-$tuples of numbers denoting those choices. Although these numbers denote points commonly, we can likewise think of them as giving us "directions" or displacements which is more useful when we are doing calculus. Then each vector can be thought of as a **directed line segment**. Further, you would notice the the $k-$ tuples can also be thought of and represented as a single row or single column matrix, which is why it was necessary to begin on earlier discussion in matrix algebra. Because of that, the rules for matrix operation applies here as well. As a convention, you might have already noticed, matrices are typically represent by bold upper case letters, and vectors are represented by bold lower case letters.

Let p and q to two points in $\mathbf{R}^n$. The line segment formed by them is written as $\overline{\mathbf{pq}}$, the vector from p to q is $\overrightarrow{pq}$, and the length of this line segment is $||\overline{pq}||$. If you recall, the distance between this two points $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ and $\mathbf{q} = (q_1, q_2, \ldots, q_n)$ in $\mathbf{R}^n$ is just,

$$||\overline{pq}|| = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_n - q_n)^2}$$

Letting r be a scalar value, if we multiply this to the vector, the length of the length segment would increase by r times if $r > 0$, and shrink r times if r is negative. That is for a vector $\mathbf{v}$, $||r\mathbf{v}|| = |r|.||\mathbf{v}||$, for $\mathbf{v}$ in $\mathbf{R}^n$. A direction vector is a useful idea that is derived from the extension/shrinkage, which is the **unit vector** that "points" in the same direction as $\mathbf{v}$, but has length of 1. To do so, all we need to do is to divide the length of the vector by itself, which is very intuitive. Then a unit vector of $\mathbf{v}$, call it $\mathbf{w}$,

$$\mathbf{w} = r.||\mathbf{v}|| = \frac{1}{||\mathbf{v}||}.||\mathbf{v}||$$

Now we can examine some of the properties and operations using vectors. An inner/dot product of two vectors $\mathbf{p}$ and $\mathbf{q}$ as above is defined as,

$$\mathbf{p}.\mathbf{q} = p_1q_1 + p_2q_2 + \cdots + p_nq_n$$

The inner product has the following properties for three vectors $\mathbf{p}$, $\mathbf{q}$ and $\mathbf{w}$ all in $\mathbf{R}^n$, and scalar r,

1. $\mathbf{p}.\mathbf{q} = \mathbf{q}.\mathbf{p}$
2. $\mathbf{p}(\mathbf{q} + \mathbf{w}) = \mathbf{p}.\mathbf{q} + \mathbf{p}.\mathbf{w}$
3. $\mathbf{p}.(r\mathbf{q}) = r(\mathbf{p}.\mathbf{q}) = (r\mathbf{p}).\mathbf{q}$
4. $\mathbf{p}.\mathbf{p} \geq 0$
5. $\mathbf{p}.\mathbf{p} = 0 \Rightarrow \mathbf{p} = \mathbf{0}$
6. $(\mathbf{p} + \mathbf{q}).(\mathbf{p}.\mathbf{q}) = \mathbf{p}.\mathbf{p} + 2(\mathbf{p}.\mathbf{q}) + \mathbf{q}.\mathbf{q}$

Note that $\mathbf{0}$ is the origin of $\mathbf{R}^n$, or a vector of zeros. Next, for θ which is the angle between two vectors $\mathbf{p}$ and $\mathbf{q}$, then

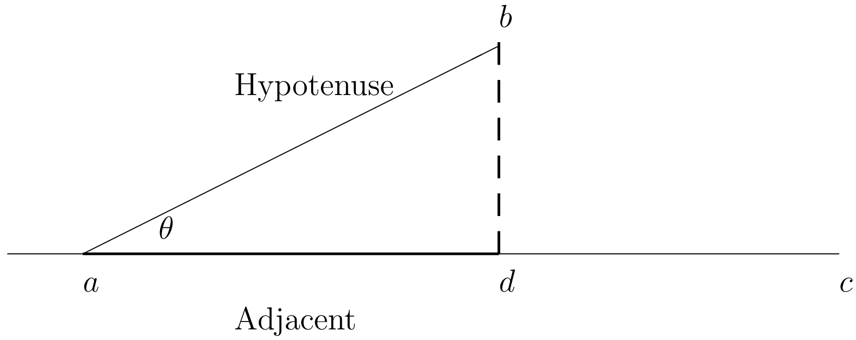
$$\mathbf{p}.\mathbf{q} = ||\mathbf{p}||||\mathbf{q}|| \cos\theta$$

So that given the properties of cos, then

1. if $\mathbf{p}.\mathbf{q} > 0$, then the angle θ is acute
2. if $\mathbf{p}.\mathbf{q} < 0$, then the angle θ is obtuse
3. if $\mathbf{p}.\mathbf{q} = 0$, then the angle θ is at right angle

The last is important in econometrics, and when two vectors are at right angles to each other, we say they are **orthogonal**. If this is not completely obvious to you, first note that, cos is just the length of the adjacent line divided by the length of the hypotenuse of a right angle triangle. When the angle formed by the hypotenuse and adjacent line

Figure 1: $\cos\theta = ||ad||/||ab||$



segments is acute, the values of $\cos(\theta)$ ranges between zero and one, where $\cos(0°) = 1$ and $\cos(90°) = 0$, as θ increases beyond $90°$, the angle formed tends towards -1, and is -1 when $\theta = 180°$.

In addition, an important rule is the **triangle inequality** which is,

$$||\mathbf{p} + \mathbf{q}|| \leq ||\mathbf{p}|| + ||\mathbf{q}||$$

The import of this inequality is that a substantial number of proofs requires this fact. To proof the above, first note that

$$\begin{aligned} \frac{\mathbf{p}.\mathbf{q}}{||\mathbf{p}||.||\mathbf{q}||} &= \cos(\theta) \leq 1 \\ \Rightarrow \mathbf{p}.\mathbf{q} &\leq ||\mathbf{p}||.||\mathbf{q}|| \\ \Rightarrow ||\mathbf{p}||^2 + 2\mathbf{p}.\mathbf{q} + ||\mathbf{q}||^2 &\leq ||\mathbf{p}||^2 + 2||\mathbf{p}||.||\mathbf{q}|| + ||\mathbf{q}||^2 \end{aligned}$$

But based on the definition of a length, and that the dot product is commutative,

$$\begin{aligned}
\mathbf{p}.\mathbf{p} + 2\mathbf{p}.\mathbf{q} + \mathbf{q}.\mathbf{q} &\leq ||\mathbf{p}||^2 + 2||\mathbf{p}||.||\mathbf{q}|| + ||\mathbf{q}||^2 \\
\Rightarrow \mathbf{p}.\mathbf{p} + 2\mathbf{p}.\mathbf{q} + \mathbf{q}\mathbf{q} &\leq ||\mathbf{p}||^2 + 2||\mathbf{p}||.||\mathbf{q}|| + ||\mathbf{q}||^2 \\
\Rightarrow (\mathbf{p} + \mathbf{q}).(\mathbf{p} + \mathbf{q}) &\leq (||\mathbf{p}|| + ||\mathbf{q}||)^2 \\
\Rightarrow ||(\mathbf{p} + \mathbf{q})||^2 &\leq (||\mathbf{p}|| + ||\mathbf{q}||)^2
\end{aligned}$$

And finally since the power function to the degree of 2 is monotonic, the triangle inequality is derived.

Further,

$$||\mathbf{p} - \mathbf{q}|| \geq |||\mathbf{p}|| - ||\mathbf{q}|||$$

You can read you text book for the simple proof. Note that $||\mathbf{p}||$ for a vector $\mathbf{p}$ is known as the Euclidean length (length from the origin) if it abides by the following properties,

1. $||\mathbf{p}|| \geq 0$ and is equal to 0 only when $\mathbf{p} = \mathbf{0}$
2. $||r\mathbf{p}|| = |r|||\mathbf{p}||$
3. the triangle inequality

It is then also called a **norm**.

With the above, we can now define a plane and begin to see some semblance of usefulness of these mathematical truths. The fundamental object is the line, which since high school has been,

$$y = mx + b$$

In the case of the Euclidean $2-$space, generally a point on a line can be written at $(p_1(t), p_2(t))$ where t is a scalar parameter that varies the coordinates on the line. Then a line in $\mathbf{R}^2$ can be determined by two things, a particular point the line has to pass through $\mathbf{p}_0$, and the direction $\mathbf{d}$. Then the equation for the line is $\mathbf{p}(t) = \mathbf{p}_0 + t\mathbf{d}$. As the dimension increases, of course the coordinates has to increase, so that the coordinates of a point on a line is a triplet and so on. But the idea of describing a line remains the same. You can also describe a line segment given two points, as you well know. In Euclidean space, the line segment is just,

$$l(\mathbf{p}, \mathbf{q}) = \{(1 - \alpha)\mathbf{p} + \alpha\mathbf{q} : 0 \leq \alpha \leq 1\}$$

In other words, just a **convex combination** of the two points.

What is more important to you is the plane which as you know is two dimensional as opposed to the line which is uni-dimensional. Where in the line, we needed only one parameter to map out the line, because the plane is in two dimension, we naturally need two parameters. Let $\mathcal{P}$ be a plane in $\mathbf{R}^3$, or the Euclidean 3−space. Further let $\mathbf{p}$ and $\mathbf{q}$ be two vectors pointing in differing directions on the plane $\mathcal{P}$. When vectors are pointing in differing directions, we say that they are **linearly independent**. Then any **linear combination** as opposed to **convex combition** must lie on the plane $\mathcal{P}$, such as $\mathbf{x} = \alpha\mathbf{p} + \beta\mathbf{q}$, where

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \alpha p_1 & + & \beta q_1 \\ \alpha p_2 & + & \beta q_2 \\ \alpha p_3 & + & \beta q_3 \end{bmatrix}$$

and this describes a plane that passes through the origin. Note that both $\alpha, \beta \in \mathbf{R}^1$ Of course you could have a plane that does not pass through the origin, in which case, you should guess that we would need a vector that transforms the plane away from the origin, and you would be correct. Then such a plane can be describe by $\mathbf{x} = \mathbf{d} + \alpha\mathbf{p} + \beta\mathbf{q}$ where $\mathbf{d}$ is a point on the plane $\mathcal{P}$.

The next question is given three vectors, which is the minimum number of points from which to obtain a plane (the idea is similar to two points to describe a line), how we can describe a plane that contains all three points. Let those three points be $\mathbf{p}$, $\mathbf{q}$ and $\mathbf{r}$. We could pick either of the vectors as a "key" vector, let it be $\mathbf{r}$. Then $\mathbf{p} - \mathbf{r}$ and $\mathbf{q} - \mathbf{r}$ are displacement vectors from $\mathbf{r}$. So we can write the plane as being described by,

$$\begin{aligned} x &= \mathbf{r} + \alpha(\mathbf{p} - \mathbf{r}) + \beta(\mathbf{q} - \mathbf{r}) \\ &= (1 - \alpha - \beta)\mathbf{r} + \alpha\mathbf{p} + \beta\mathbf{q} \end{aligned}$$

From the above, you can gather that more generally, we can always describe a plane using three vectors, in the following form,

$$x = \alpha\mathbf{p} + \beta\mathbf{q} + \gamma\mathbf{r}$$

where the coefficients α, β and γ sums to one. What would the equation describe if all the coefficients are nonnegative? You would obtain a triangle with the vectors as vertices in $\mathbf{R}^3$. The coefficients are known as the **barycentric coordinates**.

Just as there is an equation that describes a line, there is an equation that can describe a plane. In the description of a straight line, we needed a point the line would pass through, and the slope/inclination of the line. Similarly, for a plane, we need to provide an inclination, which is done through the **normal vector** which, in the case of planes is perpendicular to the plane. Let $\mathbf{n} = (n_1, n_2, n_3)$ be such a normal vector to plane $\mathcal{P}$. Let the plane pass through a point $\mathbf{p}_0 = (x_0, y_0, z_0)$, and let $\mathbf{p} = (x, y, z)$ be an arbitrary point on the plane. First note that the vector $\mathbf{p} - \mathbf{p_0}$ is a vector that is on the plane $\mathcal{P}$. Since it is on the plane, it must be perpendicular to $\mathbf{n}$, implying

$$\begin{aligned} \mathbf{n}.(\mathbf{p} - \mathbf{p}_0) &= 0 \\ \Rightarrow (n_1, n_2, n_3).(x - x_0, y - y_0, z - z_0) &= 0 \\ \Rightarrow n_1(x - x_0) + n_2(y - y_0) + n_3(z - z_0) &= 0 \end{aligned}$$

which is what is commonly referred to as the **point-normal equation** of a plane. It is also written as,

$$\alpha x + \beta y + \gamma z = \delta$$

where δ compared to the former equation is just $n_1 x_0 + n_2 y_0 + n_3 z_0$. In the latter form, the normal vector is just (α, β, γ), and the points $(\frac{\delta}{\alpha}, 0, 0)$, $(0, \frac{\delta}{\beta}, 0)$, and $(0, 0, \frac{\delta}{\gamma})$ are all on the plane.

These spaces as captured by the plane are simply just sets of points, and more generally, including lines, they are sets of points that can be described by a single equation. These spaces are generally known as **hyperplanes**. Generalizing the equations we have discussed, a hyperplane in $\mathbf{R}^n$ can always be written as,

$$a_1 p_1 + a_2 p_2 + \cdots + a_n p_n = d$$

Note that the normal vector here to the hyperplane is $\mathbf{a} = (a_1, a_2, \ldots, a_n)$.

So why is this useful? First off, think about your budget constraint in intermediate microeconomics. For two goods, x_1 and x_2, with respective prices of p_1 and p_2, and individual income of y, we write the budget constraint as,

$$y = p_1 x_1 + p_2 x_2$$

Does it look relevant now? How about in a world where the individual have n goods to choose from, each with their own respective price, p_i, where $i \in \{1, 2, \ldots, n\}$, so that the budget constraint is,

$$
\begin{aligned}
y &= p_1 x_1 + p_2 x_2 + \cdots + p_n x_n \\
&= \mathbf{p}\mathbf{x}
\end{aligned}
$$

where $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ and $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. Now it should be obvious. The normal vector is just the vector of prices of the hyperplane describing the boundary of an individuals budget set! Further, each of the points $(y/p_1, 0, \ldots, 0)$, $(0, y/p_2, 0, \ldots, 0)$, $\ldots$, $(0, 0, \ldots, y/p_n)$ are all on the hyperplane.

You would likewise see it in econometrics, particularly when dealing with statistical ideas. For one, notice that in the description of an experiment, when the probabilities of all outcomes must sum to one, then the number vector is just a vector of ones! With this, you would be able to visualize, at least in your mind, as you apply mathematical techniques to economic theory, and empirical methods what you are really doing geometrically.