# Linear Algebra:
# Applications to Econometric Theory

Imperial College London
Department of Finance
Hormoz Ramian

2018

# Contents

**A note:** This text is intended for first year graduate students in business schools with various backgrounds. The main focus is to introduce relevant linear algebraic tools that are useful to econometrics and multivariate statistical inference. The text primarily uses an intuitive perspective to briefly introduce notations, definitions and applications whilst also adopting a reasonable amount of rigour.

Correspondence: h.ramian@imperial.ac.uk

# Chapter 1

# Vectors

## 1.1 Notations

Vectors are compact notations to represent collection of numbers, characteristics or mathematically quantifiable concepts. Suppose there are $n$ commodities in a market, we can show collection of trade quantities $(q_1, q_2, \dots, q_n)$, and their prices $(p_1, p_2, \dots, p_n)$ in the following way:

$$\boldsymbol{q} = \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{bmatrix} \qquad \boldsymbol{p} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix} \tag{1.1.1}$$

In some texts "round" brackets, $(\dots)$ are used instead of square brackets, $[\dots]$. In this notation, the bold typeface is commonly used to distinguish a vector from an individual number or *scalar*. For instance, the collection of trade quantities and prices can be shown with $\boldsymbol{q}$ and $\boldsymbol{p}$, respectively. Each individual scalar within a vector is referred to as an *element* or a *co-ordinate* of vector such as $q_i$ with $i = 1, \dots, n$ shows the traded quantity of $i$th commodity.

There are two basic properties that define size or dimension of a vector: number of rows and columns. In our example, each of $\boldsymbol{q}$ and $\boldsymbol{p}$ is an $n \times 1$ vector. Since a vector is always either a single column or a single row, it can be referred to as a size $n$ column vector or size $n$ row vector, that is a simplification to avoid using $n \times 1$ or $1 \times n$. By convention, any size $n$ (or $n$-dimensional) vector is assumed to be a column vector unless otherwise is mentioned.

Each element within a vector is a number or variable that belongs to a certain domain such as real numbers $\mathbb{R}$. In our example, we assume that all prices and quantities of each of the $n$ commodities is a non-negative number, or more formally $q_i \in \mathbb{R}_+$ and $p_i \in \mathbb{R}_+$ where $\mathbb{R}_+$ refers to zero and all positive real numbers. An $n$-dimensional column vector or row vector that represents $n$ real-valued elements, belongs to an $n$-dimensional real numbers space or $\mathbb{R}^n$ and is referred to as a real-vector. For example, $\boldsymbol{q} \in \mathbb{R}^n_+$ and $\boldsymbol{p} \in \mathbb{R}^n_+$ that is to say there exists $n$ non-negative real numbers that represent quantities and prices. The $n$-dimensional space that a vector spans is referred to as **vector space** or alternatively **$n$-dimensional Euclidean space**.

Elements in vectors can be re-arranged to construct new vectors. In this context, vectors $\boldsymbol{q}$ and $\boldsymbol{p}$ summarize quantity and price information, however, one may need to re-arrange the same information to construct vectors for characteristics of each commodity. Suppose there are only two commodities ($n = 2$) then we can define new 2-dimensional vectors $\boldsymbol{x} = \binom{q_1}{p_1}$ and $\boldsymbol{y} = \binom{q_2}{p_2}$ to collect trade quantity and price for the first and second commodities:

| | $\boldsymbol{q}$ | $\boldsymbol{p}$ |
|---|---|---|
| $\boldsymbol{x}$ | $q_1$ | $p_1$ |
| $\boldsymbol{y}$ | $q_2$ | $p_2$ |

The geometrical representation of vectors is a good way to generalize upon scalars, however, only vectors with two or three elements can be illustrated in this fashion and higher dimensional vectors remain abstract. In the previous example, when $q_1 = 1$, $q_2 = 3$, $p_1 = 2$ and $p_2 = 1$ we can illustrate the geometrical representation with the following diagram (left):

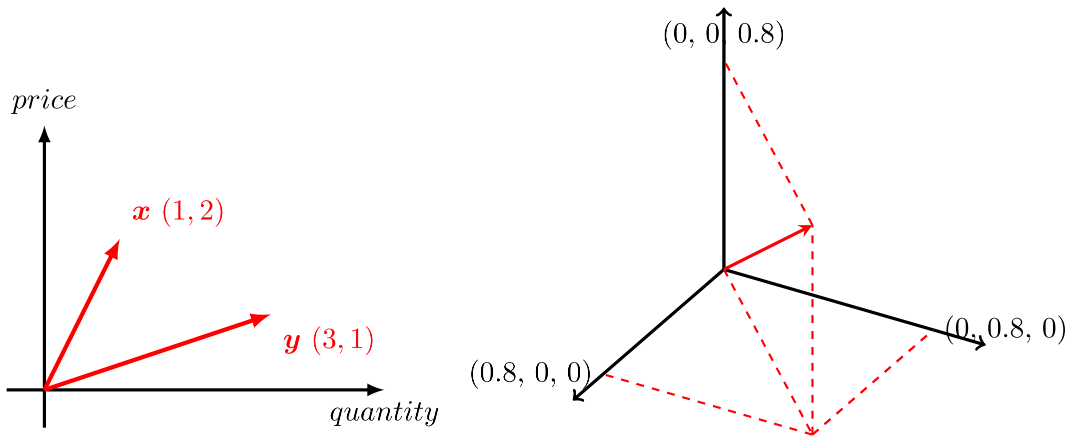



**Figure 1.1:** The left figure illustrates two 2-dimensional vectors. The first element of each vector measures the quantity and the second element measures price on the horizontal and vertical axes, respectively. The figure on the right shows a more general case where an additional characteristic, e.g. currency, is depicted on the third axis (the figure on the right shows only one vector in $\mathbb{R}^3_+$ space.)

## 1.2 Axioms

Basic vector operations are very similar to scalar operations. Before discussing these operations, let's introduce the following two vector. A **null vector** is an $n$-dimensional column (or row) vector with every element equal to zero shown by $\mathbf{0}_n$ or $\boldsymbol{O}_n$. A **one vector** is an $n$-dimensional column vector with every element equal to one shown by either of the following notations, $\mathbf{1}_n$, or $\boldsymbol{\iota}_n$. Similar to scalars, two arbitrary $n$-dimensional column vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are equal when the following holds:

$$\boldsymbol{x} = \boldsymbol{y} \quad \text{iff} \quad x_i = y_i \tag{1.2.1}$$

for[1] all of their corresponding elements ($\forall_i$). Note that both $\boldsymbol{x}$ and $\boldsymbol{y}$ must have same dimensions $n$ and be column (row) vectors. Two arbitrary $n$-dimensional column vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ can be compared with an inequality:

$$\boldsymbol{X} > \boldsymbol{Y} \quad \text{iff} \quad x_i > y_i \tag{1.2.2}$$

[1] 'iff' refers to the statement 'if and only if', that indicates a statement, e.g. $\boldsymbol{x} = \boldsymbol{y}$, is true only if the condition $x_i = y_i$ is true. It also refers to the reverse logic, that is to say when a statement is true, then the condition must be true. In particular, iff is a two directional conditional logical operator, and applies a more strict logic relative to simple 'if'.

for all of their elements. However, if strict inequality does not hold even only for one of their elements, then $\boldsymbol{x}$ and $\boldsymbol{y}$ cannot be compared with an inequality. Lastly, two arbitrary $n$-dimensional column vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ can be compared with a weak inequality:

$$\boldsymbol{x} \geq \boldsymbol{y} \text{ iff } x_i \geq y_i \tag{1.2.3}$$

for all of their elements. However, if weak inequality does not hold only for one of their elements, then $\boldsymbol{x}$ and $\boldsymbol{y}$ cannot be compared with a weak inequality. Let $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ denote arbitrary $n$-dimensional vectors, and $r$ and $s$ be any arbitrary real-valued scalars, then we can express vector axioms in the following way:

1. Commutativity

$$\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{y} + \boldsymbol{x} \tag{1.2.4}$$

2. Associativity of vector addition

$$(\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{z} = \boldsymbol{x} + (\boldsymbol{y} + \boldsymbol{z}) \tag{1.2.5}$$

3. Existence of "*additive inverse*"

$$\boldsymbol{x} + (-\boldsymbol{x}) = \mathbf{0}_n \tag{1.2.6}$$

4. Associativity of scalar multiplication

$$r \times (s \times \boldsymbol{x}) = (r \times s) \times \boldsymbol{x} \tag{1.2.7}$$

5. Distributivity of scalar sums

$$(r + s) \times \boldsymbol{x} = r \times \boldsymbol{x} + s \times \boldsymbol{x} \tag{1.2.8}$$

6. Distributivity of vector sums

$$r \times (\boldsymbol{x} + \boldsymbol{y}) = r \times \boldsymbol{x} + r \times \boldsymbol{y} \tag{1.2.9}$$

## 1.3 Addition and Subtraction

Adding two or more $n$-dimensional vectors is simply conducted by adding their respective elements. In general, for any two real-valued vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ where,

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad \boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad \boldsymbol{x} + \boldsymbol{y} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix} \qquad \boldsymbol{x} - \boldsymbol{y} = \begin{bmatrix} x_1 + (-y_1) \\ x_2 + (-y_2) \\ \vdots \\ x_n + (-y_n) \end{bmatrix} \tag{1.3.1}$$

the resulting $\boldsymbol{x} + \boldsymbol{y}$ is an n-dimensional vector. Subtraction is conducted by summing over the additive inverse $\boldsymbol{x} + (-\boldsymbol{y})$.

## 1.4 Transpose and Product

A column (row) vector can be converted to a row (column) by considering the **transpose** of the vector. The transpose of an $n$-dimensional column (row) vector $\boldsymbol{x}$ is an $n$-dimensional row (column) vector $\boldsymbol{x}^T$. Formally, the transpose operator swaps dimensions of a vector where the new transposed transformation has its number of rows (columns) equal to the number columns (rows) of the original

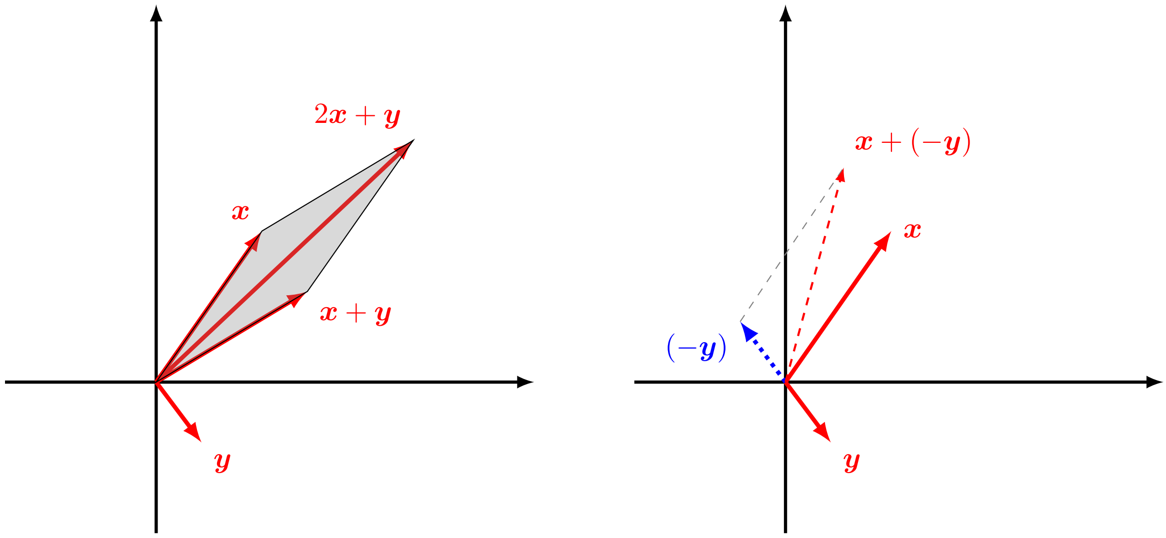



**Figure 1.2:** The figure on the left shows addition of two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ and addition of scaled vector $2\boldsymbol{x}$ and $\boldsymbol{y}$, in a 2-dimensional space. The figure on the right shows subtraction of two vectors $\boldsymbol{x} - \boldsymbol{y}$ or addition of $\boldsymbol{x} + (-\boldsymbol{y})$.

vector. The notation varies across different texts and some authors also use[2] $\boldsymbol{x}^\top$ or $\boldsymbol{x}'$:

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad \boldsymbol{x}' = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \tag{1.4.1}$$

Vectors can be multiplied in various fashions. The simplest method to multiply two vectors is referred to as the inner product of two vectors[3]:

$$z = \boldsymbol{x}'\boldsymbol{y} \tag{1.4.2}$$

where the right-hand-side is a product of an $n$-dimensional row vector by an $n$-dimensional column vector. The inner product of two arbitrary vectors is always a *scalar*, that is why in equation (1.4.2), the left-hand-side variable $z$ is not boldface, in fact, it represents the following:

$$z = \boldsymbol{x}'\boldsymbol{y} = \sum_{i=1}^{n} x_i \times y_i \tag{1.4.3}$$

or equivalently, in vector notation:

$$z = \boldsymbol{x}'\boldsymbol{y} = \begin{bmatrix} x_1 \times y_1 + x_2 \times y_2 + \ldots + x_n \times y_n \end{bmatrix} \tag{1.4.4}$$

Perhaps the most important property of the inner product is **conformity** property which requires two inner dimensions of two vectors to be identical:

$$z = \underset{(1\times n)}{\boldsymbol{x}'} \underset{(n\times 1)}{\boldsymbol{y}} \tag{1.4.5}$$

[2]Note that one should avoid confusing the latter notation $\boldsymbol{x}'$ with differentiation.

[3]Inner product is also referred to as dot product where $\boldsymbol{x}'\boldsymbol{y} = \boldsymbol{x}.\boldsymbol{y}$, (when we use the dot operator, then the transpose is not used). Furthermore, some authors use angle brackets notation $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ which is similar to dot or inner product. Both angle brackets and dot product notations are less common in econometrics texts.

where in this statement the second dimension of the first term, and the first dimension of the second term, $(1 \times n)$ $(n \times 1)$ are identical. This is a requirement for inner product of any two arbitrary vectors. Recall that we denoted trade quantities and prices with $\boldsymbol{q}$ and $\boldsymbol{p}$. We can express, for example, expenditure on commodity $i$ by multiplying two scalars $q_i \times p_i$. One can also express total expenditure on all commodities by constructing the inner product of quantity and price vectors $\boldsymbol{q}'\boldsymbol{p}$. We always need to transpose the first vector, if both vectors are column vectors, to obtain their inner product or simply sum of their element-wise products. Note that $\boldsymbol{q}'\boldsymbol{p} \neq \boldsymbol{q}\boldsymbol{p}'$ as the right-hand-side is called an **outer product** which will be introduced in the next chapter.

## 1.5 Euclidean Norm

A vector norm or magnitude is simply the length of a vector. In a two-dimensional case, this refers to Pythagoras theorem but extends its definition to higher dimensions. The Euclidean norm of an arbitrary vector $\boldsymbol{x}$ is defined by:

$$||\boldsymbol{x}|| = \sqrt{\sum_{i=1}^{n} x_i^2} \tag{1.5.1}$$

which is a scalar[4]. The operator $||\dots||$ is similar to absolute value $|\dots|$ that works for scalars. In fact, the Euclidean norm of any real-valued vector is always a non-negative scalar as this can be trivially shown in equation (1.5.1), with sum of some squared terms which ensures that the norm is always non-negative. Suppose that $\boldsymbol{x}$ is a two-dimensional vector, $\boldsymbol{x} = [x_1 \quad x_2]'$ ($[x_1 \quad x_2]'$ is a column vector), then the Euclidean norm $||\boldsymbol{x}||$ is:

$$||\boldsymbol{x}|| = \sqrt{x_1^2 + x_2^2} \tag{1.5.2}$$

which is simply the hypotenuse of a right-triangle. One can write the Euclidean norm definition using the inner product of vector $\boldsymbol{x}$ by itself:

$$||\boldsymbol{x}|| = \sqrt{\boldsymbol{x}'\boldsymbol{x}} \tag{1.5.3}$$

where the term under square root is a scalar and is equal to sum of element-wise squared elements of vector $\boldsymbol{x}$.

## 1.6 Orthogonality

Two arbitrary vectors are orthogonal if their inner product is equal to zero. Formally, orthogonality implies that $\boldsymbol{x}'\boldsymbol{y} = 0$ or more compactly shown with $\boldsymbol{x} \perp \boldsymbol{y}$. If two vectors are orthogonal, then any *scaled transformation* of them is also orthogonal to each other. Suppose $\lambda_1$ and $\lambda_2$ are two real-valued scalars, then two transformed vectors $\lambda_1\boldsymbol{x}$ and $\lambda_2\boldsymbol{y}$ are also orthogonal ($\lambda_1\boldsymbol{x} \perp \lambda_2\boldsymbol{y}$) iff $\boldsymbol{x} \perp \boldsymbol{y}$.

[4] Other notations $||\boldsymbol{x}||$ or $||\boldsymbol{x}||_2$ or $||\boldsymbol{x}||^2$ show the same concept.

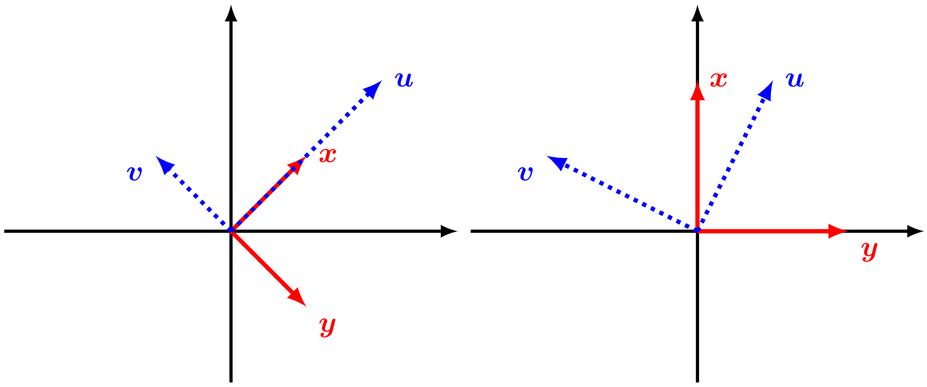



**Figure 1.3:** This figure illustrates two examples of orthogonal vectors in a 2-dimensional space. Vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are orthogonal because they are at right-angles (or perpendicular, to one another). Similarly, vectors $\boldsymbol{u} = 2\boldsymbol{x}$ and $\boldsymbol{v} = -\boldsymbol{y}$ are orthogonal because the original vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, from which they are linearly constructed, are orthogonal.

## 1.7 Linear Independence

A set of $m$ individual $n$-dimensional vectors $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m\}$ are linearly independent, if and only if none of the vectors in the set can be expresses as a linear combination of the rest. If at least one vector $\boldsymbol{x}_i$ $(i = 1, \ldots, m)$ can be expressed as a linear combination of the rest, then the set is linearly dependent:

$$\sum_{i=1}^{m} \lambda_i \boldsymbol{x}_i \quad = \quad \lambda_1 \boldsymbol{x}_1 + \lambda_2 \boldsymbol{x}_2 + \ldots \lambda_m \boldsymbol{x}_m \quad = \quad \mathbf{0} \quad \text{for some} \quad \lambda_j \neq 0 \tag{1.7.1}$$

otherwise the set is linearly independent. Intuitively, linear independence indicates that one vector can be constructed using linear transformation of another vector. For example, suppose $\boldsymbol{x} = [1 \;\; 2]'$ and $\boldsymbol{y} = [3 \;\; 6]$ then one can express vector $\boldsymbol{x}$ in terms of vector $\boldsymbol{y}$ such that $\boldsymbol{x} = 3\boldsymbol{y}$, thus satisfying the sum $\boldsymbol{x} - 3\boldsymbol{y} = \mathbf{0}$.

## 1.8 Functions of Vectors: $N$-to-$N$ and $N$-to-1 Mappings

Vector-valued functions, vector functions or functions of vectors is a generalization of functions in univariate calculus. In particular, a vector function represents certain operations of vector variables. For example, for any arbitrary real-valued $n$-dimensional vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ and a scalar $\lambda$, we can define the following functions of vectors:

$$\underset{(n\times 1)}{f(\boldsymbol{x})} \quad = \quad \mathbf{1}_n + \lambda.\boldsymbol{x} \tag{1.8.1}$$

$$\underset{(n\times 1)}{g(\boldsymbol{x}, \boldsymbol{y})} \quad = \quad \frac{\lambda_1}{\lambda_1 + \lambda_2}\boldsymbol{x} + \frac{\lambda_2}{\lambda_1 + \lambda_2}\boldsymbol{y} \tag{1.8.2}$$

$$\underset{(1\times 1)}{h} \quad = \quad ||\boldsymbol{x}||_2^{-1}\boldsymbol{x} + ||\boldsymbol{y}||_2^{-1}\boldsymbol{y} \tag{1.8.3}$$

where vector functions, $f(\boldsymbol{x})$, $g(\boldsymbol{x},\boldsymbol{y})$ and $h(\boldsymbol{x},\boldsymbol{y})$ are all $n$-dimensional column vectors themselves. We can re-write in the following way:

$$f(\boldsymbol{x}) = \begin{bmatrix} 1+\lambda x_1 \\ 1+\lambda x_2 \\ \vdots \\ 1+\lambda x_n \end{bmatrix} \qquad g(\boldsymbol{x},\boldsymbol{y}) = \begin{bmatrix} x_1+\lambda y_1 \\ x_2+\lambda y_2 \\ \vdots \\ x_n+\lambda y_n \end{bmatrix} \qquad h(\boldsymbol{x},\boldsymbol{y}) = \begin{bmatrix} \lambda + x_1 + \sqrt{\sum_{i=1}^n x_i^2} y_1 \\ \lambda + x_2 + \sqrt{\sum_{i=1}^n x_i^2} y_2 \\ \vdots \\ \lambda + x_n + \sqrt{\sum_{i=1}^n x_i^2} y_n \end{bmatrix}$$

An important aspect of functions of vectors is that a resulting function inherits dimensions of its vector arguments. Formally, an arbitrary vector function can be expressed by:

$$\{f \,|\, f : \mathbb{R}^n \to \mathbb{R}^n\} \tag{1.8.4}$$

that is to say both the comprising vectors (function domain) and the resulting vector function value (function range) belong to an $n$-dimensional vector space. It is common in econometrics to also have a *scalar* function that has *vector* arguments,

$$\{f \,|\, f : \mathbb{R}^n \to \mathbb{R}\} \tag{1.8.5}$$

This indicates that vector arguments map into a univariate vector or a scalar. For example, the Euclidean norm is a function from an $n$-dimensional vector to a single number that represents length of the vector. In the example at the beginning of this chapter, quantities and prices represented by $\boldsymbol{q}$ and $\boldsymbol{p}$ are vectors but the total expenditure that can be constructed using their inner product is a scalar.

Vector notation offers a succinct way to work with multivariate statistical models. Most statistical operators such as expected value, variance, etc. are functions that can be easily generalized to use vectors. Let an $n$-dimensional column vector $\boldsymbol{x} = [x_1 \; x_2 \; \dots \; x_n]'$ represent observed values of a certain variable $X \sim N(\mu, \sigma^2)$, where in this setting $n$ is the sample size or number of observations. Applying the expectation operator to vector $\boldsymbol{x}$, treating each element as a random variable results in the following vector:

$$\mathbb{E}[\boldsymbol{x}] = \begin{bmatrix} \mathbb{E}[x_1] \\ \mathbb{E}[x_2] \\ \vdots \\ \mathbb{E}[x_n] \end{bmatrix} = \begin{bmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{bmatrix} = \mu.\mathbf{1}_n = \boldsymbol{\mu}$$

In this context, the expectation of a vector whose elements represent random variables, is itself a vector with each element equal to population mean of the random variable. Other properties of the expectation operator also carries over, for example, $\mathbb{E}[s\boldsymbol{x}] = s.\boldsymbol{\mu}$ and $\mathbb{E}[\mathbf{1}_n + r\boldsymbol{x}] = \mathbf{1}_n + r.\boldsymbol{\mu}$ where $s$ and $r$ are arbitrary real-valued scalars.

## 1.9 Vectors of Vectors

Before proceeding to the next chapter, let's consider a generalization of vectors. A vector is a collection of elements but each element can also be itself a vector. Suppose $\boldsymbol{X}$ is an $n$-dimensional column vector:

$$\boldsymbol{X} \quad = \quad \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_n \end{bmatrix}$$

where each individual element $\boldsymbol{x}_i$ is a $k$-dimensional row vector:

$$\boldsymbol{x}_i \quad = \quad \begin{bmatrix} x_{i,1} & x_{i,2} & \dots & x_{i,k} \end{bmatrix}$$

where each $x_{i,j}$ with $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, k$ is a scalar. Vector $\boldsymbol{X}$ is an $n$-dimensional column vector in terms of each of its vector-elements $\boldsymbol{x}_i$, but another way to describe $\boldsymbol{X}$ is to say it is an $n \times k$ rectangular array or a matrix. We will develop this topic in the next chapter.

## 1.10 Exercises

**Question 1:** Let $\boldsymbol{x} = (1, 2)'$ and $\boldsymbol{y} = (2, 1)'$:
(1.1) Find $\boldsymbol{v} = \boldsymbol{x} + \boldsymbol{y}$
(1.2) Find $\boldsymbol{u} = \boldsymbol{x} - \boldsymbol{y}$
(1.3) Draw a diagram of $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{u}$ and $\boldsymbol{v}$
(1.4) Find norms of $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{u}$ and $\boldsymbol{v}$
(1.5) When can the (Euclidean) norm of a vector be zero or negative?
(1.6) Find the inner product of $\boldsymbol{x}$ and $\boldsymbol{y}$
(1.7) Are $\boldsymbol{x}$ and $\boldsymbol{y}$ linearly independent or dependent?

**Question 2:** Let $\boldsymbol{x}_1 = (1, 1, 1)'$ and $\boldsymbol{x}_2 = (1, 2, 0)'$ (note that subscripts 1 and 2 are indexes which distinguish two different vectors, and do not refer to size of the matrices):
(2.1) Can a weak inequality between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ be established?
(2.2) Are $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ linearly independent or dependent?
(2.3) Is the function of vectors $\boldsymbol{u}(\boldsymbol{x}_1) = \mathbf{1}_{3\times1} + 2\boldsymbol{x}_1$ a linear or nonlinear transformation of $\boldsymbol{x}_1$? What is the dimension of $\boldsymbol{u}$?
(2.4) What is the dimension of $\boldsymbol{v}$ if $\boldsymbol{v}(\boldsymbol{x}_1, \boldsymbol{x}_2) = ||\boldsymbol{x}_1||.||\boldsymbol{x}_2|| + ||\boldsymbol{x}'_1.\boldsymbol{x}_2||$? Is it an $N$-to-1 mapping or an $N$-to-$N$ mapping? How do you depict $\boldsymbol{v}$ on a Cartesian space?
(2.5) Is the function of vector $\boldsymbol{y}(\boldsymbol{x}_1, \boldsymbol{x}_2) = b_0\mathbf{1}_{3\times1} + b_1\boldsymbol{x}_1 + b_2\boldsymbol{x}_2$ a linear transformation of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$? what is the dimension of $\boldsymbol{y}$? what is the numerical value of vector $\boldsymbol{y}$?
(2.6) What is the univariate analogue of $\boldsymbol{x}'\mathbf{1}_{3\times1}$?

**Question 3:** Suppose,

$$\boldsymbol{x} \quad = \quad \begin{bmatrix} 1 & 2 & x & 2 & 1 \end{bmatrix}' \qquad \boldsymbol{y} \quad = \quad \begin{bmatrix} x & 1 & 0 & 1 & x \end{bmatrix}'$$

(3.1) Find functions $\lambda(x) = \boldsymbol{x}'\boldsymbol{y}$ and $\boldsymbol{\lambda}(x) = \boldsymbol{x} + \boldsymbol{y}$.
(3.2) What condition on $x$ is required such that $\boldsymbol{x}$ and $\boldsymbol{y}$ are orthogonal, if any?
(3.3) Find $x$ such that $||\boldsymbol{y}|| = 3$.
(3.4) Find $||\boldsymbol{\lambda}(x)||$. Under what values for $x$, $||\boldsymbol{\lambda}(x)|| > \sqrt{3}x$.

**Question 4:** Suppose,

$$\boldsymbol{x} = \begin{bmatrix} 1 & x & y & 2 \end{bmatrix}' \qquad \boldsymbol{y} = \begin{bmatrix} x & 1 & 1 & 1 \end{bmatrix}'$$

(4.1) What condition on $x$ and $y$ does imply $\boldsymbol{x} \perp \boldsymbol{y}$?

(4.2) Find the following norms: $||\boldsymbol{x}||_2^{-1}\boldsymbol{x}$ and $||\boldsymbol{y}||_2^{-1}\boldsymbol{y}$.

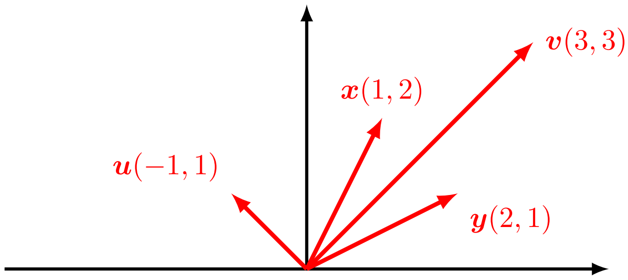
## 1.11 Solution to Exercises

**Question 1:**

(1.1) Find $\boldsymbol{v} = (1,2) + (2,1) = (3,3)'$

(1.2) Find $\boldsymbol{u} = (1,2) - (2,1) = (-1,1)'$

(1.3) Draw a diagram of $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{u}$ and $\boldsymbol{v}$




(1.4) $||\boldsymbol{x}|| = \sqrt{5}$, $||\boldsymbol{y}|| = \sqrt{5}$ and $||\boldsymbol{u}|| = \sqrt{2}$ and $||\boldsymbol{v}|| = \sqrt{3^2 + 3^2} = \sqrt{18}$

(1.5) Euclidean norm is always a non-negative number (for real-valued vectors). It is zero iff the vector is a null vector.

(1.6) $\boldsymbol{x}'.\boldsymbol{y} = 1 \times 2 + 2 \times 1 = 4$

(1.7) Vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are linearly independent because there is no such combination $\boldsymbol{x} = \lambda \boldsymbol{y}$, for $\lambda \neq 0$, can be established.

**Question 2:**

(2.1) No, because neither of the vectors is element-wise weakly greater or less than the other one.

(2.2) Vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are linearly independent since $\boldsymbol{x} = \lambda \boldsymbol{y}$, for $\lambda \neq 0$ cannot be established.

(2.3) Function $\boldsymbol{u}(\boldsymbol{x}_1)$ is a linear transformation of $\boldsymbol{x}_1$, and dimensions coincide with those of $\boldsymbol{x}_1$.

(2.4) Function $\boldsymbol{v}$ is an $N$-to-1 mapping because the right-hand-side is sum of scalars (norm of any size vectors is a scalar operator). The graphical illustration would be only one single point on the real line:

$$\begin{aligned} \boldsymbol{y}(\boldsymbol{x}_1, \boldsymbol{x}_2) &= \sqrt{1^2 + 1^2 + 1^2} \times \sqrt{1^2 + 2^2 + 0^2} + \sqrt{(1 \times 1)^2 + (1 \times 2)^2 + (1 \times 0)^2} \\ &= 2\sqrt{5} \end{aligned}$$

(2.5) Function $\boldsymbol{y}(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is a linear transformation of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. Dimensions are $3 \times 1$ and the numerical value of elements of this vector are defined in terms of parameters $(b_0 + b_1 + b_2,\ b_0 + b_1 + 2b_2,\ b_0 + b_1)'$.

(2.6) Univariate analogue refers to operation that yields the same results but without vector/matrix operations. Vector operation $\boldsymbol{x}'\mathbf{1}_{3\times 1}$ is an inner product (giving sum of element-wise product of all elements from both vectors). In this question, we have a special case of inner product because one of the vectors is a one vector (e.g. $\mathbf{1}_{3\times 1}$). Particularly, this implies that the inner product of any vector $\boldsymbol{x}$ by a $\mathbf{1}$ vector only sums up the individual elements of vector $\boldsymbol{x}$. This operation can also be written using a sum:

$$\sum_{i=1}^{n} x_i = 1 + 1 + 1$$

# Chapter 2

# Matrices

Matrices are widely used in many branches of mathematics, statistics and econometrics. Similar to vectors, matrices simplify multivariate notations and offer a more convenient way to handle mathematical operations. Unlike vectors, with actual numerical values and a unique definition, a matrix has multiple definitions depending on the context that it is being used in. A matrix can be defined as a collection of vectors, a collection of $n$-dimensional points in an $m$-dimensional space, or just a rectangular array of numbers. In econometrics, we use matrices as a generalization of *vectors*, where each vector represents collection of sample data, observations or random variables. This chapter introduces important matrices with common use in econometrics and relevant matrix operations.

## 2.1 Zero, One

Matrices generalize vector notations, particularly, a matrix is an $n \times k$ array of scalars, vectors or even smaller matrices. For example,

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \tag{2.1.1}$$

where $\boldsymbol{X}$ is a $2 \times 2$ matrix, with 2 rows and 2 columns ($n = 2$ and $k = 2$). Similar to vectors, we use boldface notation to distinguish matrices from scalars and also depending on authors, different brackets (e.g. round, square, etc.) may be used to illustrate a matrix. Each individual item within a matrix is called an element of the matrix and we use *both* row and column identifiers to distinguish them, such as element $x_{12}$ that sits on the first row and under the second columns. In general there is no restriction on the dimension of a matrix but it is required that a matrix must be a rectangular array, that is to say there must be an entry for every element (e.g. zero). The general notation for a matrix uses $\boldsymbol{X} = \{x_{i,j}\}$ where $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, k$:

$$\begin{aligned} \boldsymbol{X} &= \{x_{i,j}\} \\ &= \begin{bmatrix} x_{11} & x_{21} & \ldots & x_{1k} \\ x_{21} & x_{22} & \ldots & x_{2k} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{nk} \end{bmatrix} \in \mathbb{R}^{n \times k} \end{aligned}$$

An arbitrary real-valued matrix $\boldsymbol{X}$ belongs to $\mathbb{R}^{n\times k}$ space which is known as **matrix span**. This $n\times k$-dimensional space is the minimum space needed to express this matrix. Consider the following example which is a $(4\times 4)$ array and is constructed using $\boldsymbol{X}$ $(2\times 2)$:

$$\begin{aligned}\boldsymbol{Y} &= \begin{bmatrix}\boldsymbol{X} & \boldsymbol{X}\\ \boldsymbol{X} & \boldsymbol{X}\end{bmatrix}\\ &= \begin{bmatrix}x_{11} & x_{12} & x_{11} & x_{12}\\ x_{21} & x_{22} & x_{21} & x_{22}\\ x_{11} & x_{12} & x_{11} & x_{12}\\ x_{21} & x_{22} & x_{21} & x_{22}\end{bmatrix}\end{aligned}$$

which is a matrix of repeated matrices $\boldsymbol{X}$. Similar to vectors, we have the following basic but important matrices: a **zero** or **null** matrix is an $n\times m$ matrix which has zeros for all of its elements and is denotes by $\mathbf{0}_{n\times m}$ or $\mathbf{O}_{n\times m}$. A **one** matrix is an $n\times m$ matrix which has ones for all of its elements and is denotes by $\mathbf{1}_{n\times m}$. Note that matrices are two-dimensional arrays, however, if the number rows and columns are equal, then we can show them using only of one their dimensions and mentioning that the matrix is a square matrix, for example, $\mathbf{0}_n$ is an $n\times n$ zero matrix or $\mathbf{1}_n$ is an $n\times n$ one matrix. One should avoid confusing this notation with vectors, as this notation uses similar boldface and only one index. Generally, the notation of vectors and matrices is interchangeably used and the only way to distinguish them is to use indeces or follow context-dependent explanations about their dimensions.

| Variable | Vector Notation | Matrix Notation |
|---|---|---|
| Scalar | $1\times 1$ vector | $1\times 1$ matrix |
| — | $n$-dimensional vector | $n\times 1$ or $1\times n$ matrix |

**Table 2.1:** Relationship between a scalar, a vector and a matrix.

A scalar is also a $1\times 1$ vector which has only one element. A scalar also fits into the definition of a matrix, that is a $1\times 1$ matrix. Similarly, a vector is an $n\times 1$ or $1\times n$ matrix. This is a useful summary to remember because when a scalar or vector fits into the definition of matrix, it means that matrix operations can be applied to scalars and vectors.

## 2.2 Square, Diagonal, and Identity

A square matrix is a square array of elements with equal number of rows and columns. In general, matrix $\boldsymbol{X}$ is a square matrix if it has dimensions $n\times n$. Square matrices are a special case of rectangular matrices with additional useful properties. Before discussing these properties, let's define the following two important matrices. A **diagonal** matrix is a *square*, $n\times n$ matrix which has zeros for all of its elements except the main diagonal. The main diagonal of a matrix refers to elements within a matrix with equal row and column identifiers[1]. For example,

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & 0 & 0 \\ 0 & x_{22} & 0 \\ 0 & 0 & x_{33} \end{bmatrix} \tag{2.2.1}$$

is a diagonal matrix (must be square). Note that a diagonal matrix includes only one non-zero element on each row (or column), thus we can see this as diagonal vector within a square matrix. In the previous example:

$$\boldsymbol{x} = \begin{bmatrix} x_{11} & x_{22} & x_{33} \end{bmatrix}' \tag{2.2.2}$$

is a $3 \times 1$ column vector collecting elements on the main diagonal of matrix $\boldsymbol{X}$. The general notation to represent a diagonal matrix is:

$$\boldsymbol{x} = \text{diag}(\boldsymbol{X}) \tag{2.2.3}$$

where $\boldsymbol{x}$ can be an arbitrary $n$-dimensional vector. Intuitively, the diag(.) operator takes a vector argument and returns a square diagonal matrix. The notion of a diagonal matrix can be extended to a more generalized form known as **block diagonal**:

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{A} & \mathbf{0}_n & \dots & \mathbf{0}_n \\ \mathbf{0}_n & \boldsymbol{A} & \dots & \mathbf{0}_n \\ \vdots & \vdots & & \vdots \\ \mathbf{0}_n & \mathbf{0}_n & \dots & \boldsymbol{A} \end{bmatrix} \tag{2.2.4}$$

where $\boldsymbol{A}$ is an $n \times n$ square matrix, and $\mathbf{0}_n$ is an $n \times n$ zero matrix. Matrix $\boldsymbol{X}$ is a block diagonal matrix because it has zeros everywhere except on its main block-diagonal with each block itself being an $n \times n$ matrix. If we assume that matrix $\boldsymbol{X}$ has $m$ block elements along its diagonal, then its dimensions in terms of scalars is $(m \times n) \times (m \times n)$. An **identity** matrix is a *square* and *diagonal* $n \times n$ matrix which has zeros for all of its elements, and 1 for its main diagonal and is denoted by $\mathbf{I}_n$. Identity matrix is the analog of number 1.00 in univariate calculus, which we discuss in the matrix multiplication section.

Lastly, a **tridiagonal** matrix is a square matrix that has non-zero elements on the main diagonal, the first diagonal below it, and the first diagonal above the main diagonal.

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{A} & \mathbf{0}_n & \dots & \mathbf{0}_n \\ \boldsymbol{A} & \boldsymbol{A} & \boldsymbol{A} & \dots & \mathbf{0}_n \\ \mathbf{0}_n & \boldsymbol{A} & \boldsymbol{A} & \dots & \mathbf{0}_n \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{0}_n & \mathbf{0}_n & \mathbf{0}_n & \dots & \boldsymbol{A} \end{bmatrix} \tag{2.2.5}$$

## 2.3 Transpose, Symmetry and Trace

The transpose operation of matrix switches row and column identifiers of each element of a matrix. The transpose can be applied to any matrix, regardless of its dimensions (square or non-square).

[1]The term main refers to elements within a matrix with $i = j$ (e.g. $[\searrow]$), but we also can have the opposite definition antidiagonal (sometimes counter diagonal, secondary diagonal, trailing diagonal or minor diagonal) that refers to elements on the diagonal from top-right to bottom-left (e.g. $[\swarrow]$).

In particular, an $n \times k$ matrix $\boldsymbol{X}$ can be transposed to a $k \times n$ matrix $\boldsymbol{X}'$ (or $\boldsymbol{X}^\top$ or $\boldsymbol{X}^T$). For example,

$$\boldsymbol{X} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \text{ then } \boldsymbol{X}' = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix} \tag{2.3.1}$$

when we apply the transpose operator to a *square* matrix, then resulting matrix is also a square matrix. More interestingly, a matrix is a **symmetric** matrix if the original and the transposed transformation are identical (only square matrices can have this property). More formally,

$$\boldsymbol{X} = \boldsymbol{X}' = (\boldsymbol{X}')' \tag{2.3.2}$$

We can immediately verify that a square zero matrix, a square one matrix and an identity matrix (always square by definition) are symmetric matrices, because:

$$\begin{aligned} \mathbf{0}_n &= \mathbf{0}'_n \\ \mathbf{1}_n &= \mathbf{1}'_n \\ \mathbf{I} &= \mathbf{I}' \end{aligned}$$

For example,

$$\boldsymbol{X} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \text{ then } \boldsymbol{X}' = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \tag{2.3.3}$$

Another important matrix operator is the **trace** operator which sums the elements on the main diagonal of a matrix. The term diagonal implies that the matrix must be a *square* matrix which is a requirement before one applies the trace operator. More formally, for any arbitrary square $n \times n$ matrix $\boldsymbol{X}$:

$$tr(\boldsymbol{X}) = \sum_{i=1}^{n} x_{i,i} \tag{2.3.4}$$

For example, $tr(\mathbf{1}_n) = n$ or $tr(\mathbf{I}_n) = n$ because an $n$-dimensional square one matrix or an $n$-dimensional identity matrix has only ones on their main diagonal hence its trace is equal to sum of ones, $n$-times. For example,

$$\boldsymbol{X} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \text{ then } tr(\boldsymbol{X}) = 1 + 4 \tag{2.3.5}$$

In general, trace operator has the following two important properties:

$$tr(\boldsymbol{X} + \boldsymbol{Y}) = tr(\boldsymbol{X}) + tr(\boldsymbol{Y}) \tag{2.3.6}$$

and for any real-valued scalar $\lambda$:

$$tr(\lambda.\boldsymbol{X}) = \lambda.tr(\boldsymbol{X}) \tag{2.3.7}$$

Trace is neutral to transpose:

$$tr(\boldsymbol{X}) = tr(\boldsymbol{X}') \tag{2.3.8}$$

## 2.4 Linear Dependence, Linear Combinations and Elementary Matrix Operations

Let $\boldsymbol{X}$ denote an arbitrary $n \times k$ matrix with $k < n$. It is sometimes possible to express one of the rows (columns) of $\boldsymbol{X}$ in terms of one or a linear combination[2] of other rows (columns) of $\boldsymbol{X}$. For example,

$$\boldsymbol{X} = \begin{array}{c c} & \begin{array}{cc} C_1 & C_2 \end{array} \\ \begin{array}{c} R_1 \\ R_2 \end{array} & \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \end{array} \tag{2.4.1}$$

where $R_1$ and $R_2$ are row labels and $C_1$ and $C_2$ are column labels. It is possible to express $C_2 = 2C_1$ or $R_2 = 2R_1$. Consider the following example:

$$\boldsymbol{X} = \begin{array}{c c} & \begin{array}{ccc} C_1 & C_2 & C_3 \end{array} \\ \begin{array}{c} R_1 \\ R_2 \\ R_3 \end{array} & \begin{pmatrix} 1 & 2 & 4 \\ 2 & 4 & 8 \\ 3 & 6 & 12 \end{pmatrix} \end{array} \tag{2.4.2}$$

Similarly, we can express linear combinations $C_3 = 4C_1$ or $C_3 = 2C_1 + C_2$ or $C_3 = 3C_2 - 2C_1$. We can also establish additional relationships by multiplying (e.g. $C_1 \times C_2 - ...$) but note that this is not a linear combination of columns and we only focus on possible linear combinations. If a linear combination can be established, then rows (columns) are **linearly dependant**. Consider the following $3 \times 3$ identity matrix:

$$\boldsymbol{I}_3 = \begin{array}{c c} & \begin{array}{ccc} C_1 & C_2 & C_3 \end{array} \\ \begin{array}{c} R_1 \\ R_2 \\ R_3 \end{array} & \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{array} \tag{2.4.3}$$

in this example, it is not possible to express any of rows (columns) in terms of a linear combination of other rows (columns). If no linear combination can be established, then rows (columns) are **linearly independent**. A row (column) or a linear combination of rows (columns) of a matrix can be added to (subtracted from) any other rows (columns) within that matrix. Such operations change the original matrix but leave the span of the matrix intact. For example:

$$\boldsymbol{X} = \begin{array}{c c} & \begin{array}{ccc} C_1 & C_2 & C_3 \end{array} \\ \begin{array}{c} R_1 \\ R_2 \\ R_3 \end{array} & \begin{pmatrix} 1 & 2 & 4 \\ 2 & 4 & 8 \\ 3 & 6 & 12 \end{pmatrix} \end{array} \tag{2.4.4}$$

can be re-written in the following way:

$$\begin{array}{c c} & \begin{array}{ccc} C_1 & C_2 & C_3 \end{array} \\ \begin{array}{c} R_1 \\ R_2 \\ R_3 \end{array} & \begin{pmatrix} 1 & 2+1 & 4 \\ 2 & 4+2 & 8 \\ 3 & 6+3 & 12 \end{pmatrix} \end{array} \tag{2.4.5}$$

[2]A linear combination refers to operations that only include adding or subtracting constants or variables, but not multiplying or dividing.

where the first column is added to the second column, similarly:

$$\begin{array}{c} \\ R_1 \\ R_2 \\ R_3 \end{array} \begin{array}{c} \begin{array}{ccc} C_1 & C_2 & C_3 \end{array} \\ \begin{pmatrix} 1 & 2+1 & 4 \\ 2+1 & 4+2+2+1 & 8+4 \\ 3 & 6+3 & 12 \end{pmatrix} \end{array} \tag{2.4.6}$$

where the first row is added to the second row. Furthermore, a row (column) can be switched with another row (columns). If linear combinations can be established between rows (columns), then one or more of the matrix rows (columns) can be expressed in terms of others. For example if we start with the following matrix:

$$\boldsymbol{X} = \begin{array}{c} \\ R_1 \\ R_2 \\ R_3 \end{array} \begin{array}{c} \begin{array}{ccc} C_1 & C_2 & C_3 \end{array} \\ \begin{pmatrix} 1 & 2 & 4 \\ 2 & 4 & 8 \\ 3 & 6 & 12 \end{pmatrix} \end{array} \tag{2.4.7}$$

and add $-\frac{1}{2}C_2$ to $C_1$, then:

$$\boldsymbol{X} = \begin{array}{c} \\ R_1 \\ R_2 \\ R_3 \end{array} \begin{array}{c} \begin{array}{ccc} C_1 & C_2 & C_3 \end{array} \\ \begin{pmatrix} 1-\frac{1}{2}\times 2 & 2 & 4 \\ 2-\frac{1}{2}\times 4 & 4 & 8 \\ 3-\frac{1}{2}\times 6 & 6 & 12 \end{pmatrix} \end{array} \rightarrow \begin{pmatrix} 0 & 2 & 4 \\ 0 & 4 & 8 \\ 0 & 6 & 12 \end{pmatrix} \rightarrow \begin{pmatrix} 2 & 4 \\ 4 & 8 \\ 6 & 12 \end{pmatrix} \tag{2.4.8}$$

where in the first step, $C_1$ reduced to zeros, as a result of elementary operation $-\frac{1}{2}C_2$, and in the second step, the resulting zero column was eliminated from the matrix. In fact, the original $3 \times 3$ matrix $\boldsymbol{X}$ and the resulting $3 \times 2$ matrix are not identical but they both belong to the same space. This property is an important feature since it defines the minimum matrix span which has important application to econometrics. In the previous example, we can take an additional step and apply another elementary matrix operation:

$$\begin{pmatrix} 2 & 4 \\ 4 & 8 \\ 6 & 12 \end{pmatrix} \rightarrow \begin{pmatrix} 2 & 4-2\times 2 \\ 4 & 8-2\times 4 \\ 6 & 12-2\times 6 \end{pmatrix} \rightarrow \begin{pmatrix} 2 & 0 \\ 4 & 0 \\ 6 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix} \tag{2.4.9}$$

One can take two further steps to express the first row of the very last result in terms of second row ($R_1 = \frac{1}{2}R_2$), and then repeat a similar operation with the third row. This eventually reduces the matrix to a scalar. On the contrary, an identity matrix:

$$\boldsymbol{I}_3 = \begin{array}{c} \\ R_1 \\ R_2 \\ R_3 \end{array} \begin{array}{c} \begin{array}{ccc} C_1 & C_2 & C_3 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{array} \tag{2.4.10}$$

can never be reduced to any smaller matrix because there exists no linear combination between rows or columns that can result in one of the entire rows or columns equal to zero.

As a corollary, any matrix with linear dependence between its rows (columns) can be reduced to a smaller matrix. This process through elementary matrix operations can continue until the stage when the matrix can no longer be reduced. Linear (in)dependence property is an important

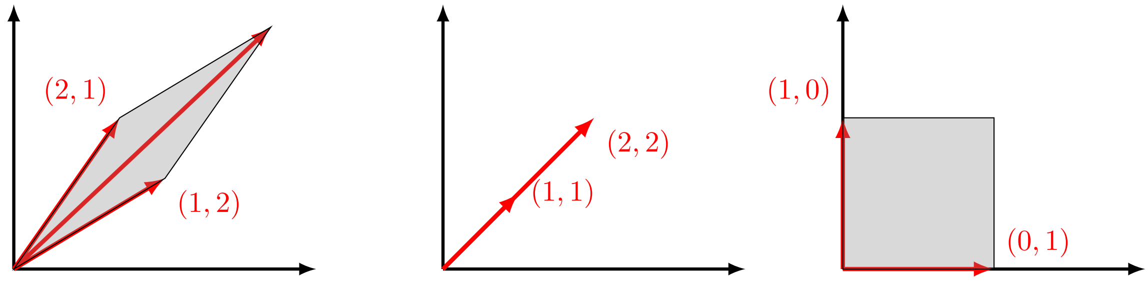



**Figure 2.1**

property that determines concepts such as collinearity in econometrics or existence of an inverse when solving the least squares method. We elaborate on this point in the next section.

## 2.5 Rank, Determinant and Singularity

The maximum number of linearly independent rows or columns within a matrix is called the **rank** of a matrix. Intuitively, if an $n \times n$ matrix for which we can establish no linear combination among its rows or columns is called rank $n$ matrix. There are many ways that the rank of a matrix can be found. Elementary matrix operations is the first way to find rank of a matrix through successive operations to reduce a matrix to its minimum size possible. However, this is a tedious process and can be very costly for large matrices. An alternative way to find rank of a matrix is to find the determinant of a matrix. The determinant is an abstract concept but let's first begin with a graphical illustration of the following $2 \times 2$ matrices in a 2-dimensional space:

$$\boldsymbol{X} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = \left[\begin{bmatrix} 2 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix}\right] \qquad \boldsymbol{Y} = \left[\begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix}\right] \qquad \boldsymbol{Z} = \left[\begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right]$$

we express matrix $\boldsymbol{X}$ as collection of two vectors, which are depicted in the following figure. Graphically the shaded area occupied by the parallelogram in the figure shows the determinant of $\boldsymbol{X}$. If vectors overlap, then the area occupied by the parallelogram vanishes (Figure (2.1), middle panel), and conversely, if the vectors for a perpendicular angle (e.g. matrix $\boldsymbol{Z}$) with each other then area occupied by the parallelogram becomes a square. Matrices $\boldsymbol{X}$ and $\boldsymbol{Z}$ cannot be reduce to any smaller matrix, hence no linear combination between their rows (columns) can be established (all rows and columns are linearly *independent*), and subsequently they have non-zero determinant. On the contrary, matrix $\boldsymbol{Y}$ can be reduced to a small matrix through elementary matrix operations because, for instance, the second row can be defined as a factor of the first columns, hence one can establish linear combination between its rows or columns (linearly *dependent*).

The **determinant** of matrix (only square) is a measure to distinguish linear independence of rows or columns of thr matrix. More formally, the determinant of a $2 \times 2$ matrix $\boldsymbol{X}$:

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{2,1} & x_{22} \end{bmatrix} \tag{2.5.1}$$

is defined as:

$$\begin{aligned} det(\boldsymbol{X}) &\equiv \begin{vmatrix} x_{11} & x_{12} \\ x_{2,1} & x_{22} \end{vmatrix} \\ &= x_{11} \times x_{22} - x_{12} \times x_{21} \end{aligned}$$

where $det(\boldsymbol{X})$ can also be shown with $|\boldsymbol{X}|$. The determinant of a $3 \times 3$ matrix is defined as:

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}$$

$$det(\boldsymbol{X}) = (-1)^{1+1} x_{11} \times \begin{vmatrix} x_{2,2} & x_{2,3} \\ x_{3,2} & x_{3,3} \end{vmatrix} + (-1)^{1+2} x_{12} \times \begin{vmatrix} x_{2,1} & x_{2,3} \\ x_{3,1} & x_{3,3} \end{vmatrix} + (-1)^{1+3} x_{13} \times \begin{vmatrix} x_{2,1} & x_{2,2} \\ x_{3,1} & x_{3,2} \end{vmatrix}$$

as shown in the last equation, determinant of a $3 \times 3$ matrix can be expressed in terms of three $2 \times 2$ determinant multiplied by an element of the $3 \times 3$ matrix.

There is an important relationship between rank and determinant of a matrix. When an $n \times n$ matrix cannot be reduced to a smaller matrix, then it has $rank = n$ and its determinant is non-zero. When an $n \times n$ matrix can be reduced to a smaller matrix, $rank < n$ and determinant is zero. Once the matrix is reduced until the stage that no further reduction through finding linear combination of rows or columns is possible, then the determinant is non-zero. In fact, finding the determinant is much easier than finding possible linear combinations since a non-zero determinant simply indicates that no linear combination can be established. An $n \times n$ matrix with non-zero determinant is also referred to as a **full rank** or **non-singular** matrix. Otherwise, when the matrix is not full rank it is referred to as a **singular** matrix. Determinant of larger matrices (e.g. larger than $3 \times 3$) can also be found by expressing the determinant in terms of determinant of smaller matrices, however, this is a numerically tedious task and usually done by using a software.

## 2.6 Equality, Addition and Subtraction

Similar to vectors, two matrices are equal (given that they are of the same size) if and only if all of their corresponding elements are element-wise equal. Also we can add or subtract two matrices, by applying add or subtract to their corresponding elements. In general, we can establish the following axioms for matrices:

1. Commutativity

$$\boldsymbol{X} + \boldsymbol{Y} = \boldsymbol{Y} + \boldsymbol{X} \tag{2.6.1}$$

2. Associativity of vector addition

$$(\boldsymbol{X} + \boldsymbol{Y}) + \boldsymbol{Z} = \boldsymbol{X} + (\boldsymbol{Y} + \boldsymbol{Z}) \tag{2.6.2}$$

3. Existence of "*additive inverse*"

$$\boldsymbol{X} + (-\boldsymbol{X}) = \boldsymbol{0}_n \tag{2.6.3}$$

4. Associativity of scalar multiplication

$$r \times (s \times \boldsymbol{X}) = (r \times s) \times \boldsymbol{X} \tag{2.6.4}$$

5. Distributivity of scalar sums

$$(r + s) \times \boldsymbol{X} = r \times \boldsymbol{X} + s \times \boldsymbol{X} \tag{2.6.5}$$

6. Distributivity of vector sums

$$r \times (\boldsymbol{X} + \boldsymbol{Y}) = r \times \boldsymbol{X} + r \times \boldsymbol{Y} \tag{2.6.6}$$

## 2.7 Matrix Multiplication

There are many different ways to multiply two matrices. Each method has different application and simplifies operations in a certain way. This section introduces three matrix multiplication methods that are commonly used in econometrics.

### 2.7.1 General Matrix Product and Conformity

Two matrices can be (general product operation) multiplied only if they satisfy the conformity property. In particular, conformity property requires the inner dimensions of two matrices to be identical. More formally, a $k \times n$ dimensional matrix $\boldsymbol{Y}$ can be pre-multiplied by an $n \times k$ dimensional matrix $\boldsymbol{Z}$:

$$\underset{n\times n}{\boldsymbol{X}} = \underset{(n\times k)}{\boldsymbol{Y}}\underset{(k\times n)}{\boldsymbol{Z}} \tag{2.7.1}$$

because their inner dimensions are identical. When $\boldsymbol{X}$ is an $n \times n$ square matrix, then:

$$\boldsymbol{X} = \boldsymbol{I}_n\boldsymbol{X} = \boldsymbol{X}\boldsymbol{I}_n \tag{2.7.2}$$

where the identity matrix is multiplicatively neutral (pre-multiplying or post-multiplying). However, this is not the case in general for two arbitrary $n \times m$ matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$:

$$\boldsymbol{X}'\boldsymbol{Y} \neq \boldsymbol{X}\boldsymbol{Y}' \tag{2.7.3}$$

a simple verification is to work out the dimensions of the left hand side and the right hand side, where $\boldsymbol{X}'$ is $m \times n$ and $\boldsymbol{Y}$ is $n \times m$ hence $\boldsymbol{X}'\boldsymbol{Y}$ is $m \times m$, but by the same logic $\boldsymbol{X}\boldsymbol{Y}'$ is $n \times n$ which is a contradition. Before proceeding to the next section, let's mention the general matrix product properties. First,

$$(\boldsymbol{X}_1\boldsymbol{X}_2)\boldsymbol{X}_3 = \boldsymbol{X}_1(\boldsymbol{X}_2\boldsymbol{X}_3) \tag{2.7.4}$$

The trace also has the following useful properties:

$$tr(\boldsymbol{X}_1\boldsymbol{X}_2\boldsymbol{X}_3) = tr(\boldsymbol{X}_2\boldsymbol{X}_3\boldsymbol{X}_1) = tr(\boldsymbol{X}_3\boldsymbol{X}_1\boldsymbol{X}_2) = tr(\boldsymbol{X}_1\boldsymbol{X}_2\boldsymbol{X}_3) \tag{2.7.5}$$

but,

$$tr(\boldsymbol{X}_1\boldsymbol{X}_2\boldsymbol{X}_3) \neq tr(\boldsymbol{X}_1\boldsymbol{X}_3\boldsymbol{X}_2) \tag{2.7.6}$$

Note that the lemma to handle trace of multiplicative forms is to follow a *circular rule*. For example, we can reshuffle multiplicative arguments of trace operator as long as their ordering is maintained (Figure (2.2)).

Another important property of multiplicative forms is the following (assuming that conformity property satisfies):

$$(\boldsymbol{X}\boldsymbol{Y}\boldsymbol{Z})' = \boldsymbol{Z}'\boldsymbol{Y}'\boldsymbol{X}' \tag{2.7.7}$$

Let's now assume that $\boldsymbol{X}$ is $n \times k$, $\boldsymbol{Y}$ is $k \times m$ and $\boldsymbol{Z}$ is $m \times s$. Then size of the left-hand-side product, before transpose, is, $n \times s = (n \times k)(k \times m)(m \times s)$, which becomes $s \times n$ after applying the transpose. Now let's consider the dimension of the right-hand-side, which is given by (after transposing each individual element), $s \times n = (s \times m)(m \times k)(k \times n)$. Lastly, when a text discusses matrix product, it always refers to the general product by default. This can also be distinguished from other types of matrix products as the following methods such as Kronecker product uses "$\otimes$"

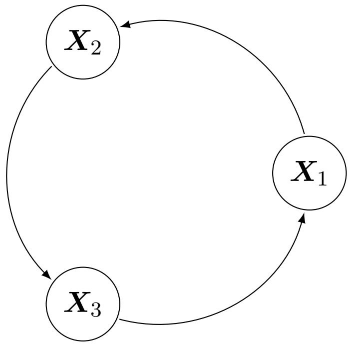



**Figure 2.2:** This diagram describes the possible ordering of arguments of trace operator (with three arguments). As illustrated, arguments $\boldsymbol{X}_1\boldsymbol{X}_2\boldsymbol{X}_3$ or for short, 123 can be circularly exchanged by removing the first term and adding it at the end of the last term, e.g. 231 or 312 or 312 but not 132.

(e.g $\boldsymbol{A} \otimes \boldsymbol{B}$) and Hadamard product uses "$\circ$" (e.g $\boldsymbol{A} \circ \boldsymbol{B}$), instead of a dot or nothing (e.g $\boldsymbol{AB}$), to label the multiplication operator.

### 2.7.2 Kronecker Product

Another method to construct a product from two matrices is called the **Kronecker** product. Unlike the previous matrix operation, the Kronecker has no requirement and it can be applied to any arbitrary matrices of different dimensions. In particular, the Kronecker multiplies each element of the pre-multiplying matrix by the entire post-multiplying matrix using the following rule for a pre-multiplying $n \times k$ matrix by a post-multiplying $s \times m$ matrix:

$$\boldsymbol{X} \otimes \boldsymbol{Y} = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \otimes \begin{bmatrix} y_{11} & y_{21} & \dots & y_{1m} \\ y_{21} & y_{22} & \dots & y_{2m} \\ \vdots & \vdots & & \vdots \\ y_{s1} & y_{s2} & \dots & y_{sm} \end{bmatrix} = \begin{bmatrix} x_{11}\boldsymbol{Y} & x_{21}\boldsymbol{Y} & \dots & x_{1k}\boldsymbol{Y} \\ x_{21}\boldsymbol{Y} & x_{22}\boldsymbol{Y} & \dots & x_{2k}\boldsymbol{Y} \\ \vdots & \vdots & & \vdots \\ x_{n1}\boldsymbol{Y} & x_{n2}\boldsymbol{Y} & \dots & x_{nk}\boldsymbol{Y} \end{bmatrix}$$

where $\otimes$ is the Kronecker product operation. This product has limited applications but it can be a very helpful method when working with certain multivariate econometric methods such Vector Autoregressive Regression (VAR).

### 2.7.3 Hadamard Product

The last matrix operation that this section introduces is the Hadamard product which is simply an element-wise product of two matrices. This product requires both the pre- and post-multiplying matrices to be of the exact size. This is a stronger conformity condition relative to the requirement mention in the general matrix product section.

$$\boldsymbol{X} \circ \boldsymbol{Y} = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \circ \begin{bmatrix} y_{11} & y_{21} & \dots & y_{1k} \\ y_{21} & y_{22} & \dots & y_{2k} \\ \vdots & \vdots & & \vdots \\ y_{n1} & y_{n2} & \dots & y_{nk} \end{bmatrix} = \begin{bmatrix} x_{11}y_{11} & x_{21}y_{21} & \dots & x_{1k}y_{1k} \\ x_{21}y_{21} & x_{22}y_{22} & \dots & x_{2k}y_{2k} \\ \vdots & \vdots & & \vdots \\ x_{n1}y_{n1} & x_{n2}y_{n2} & \dots & x_{nk}y_{nk} \end{bmatrix}$$

## 2.8 Inversion

Matrix inversion is an analog to division in univariate calculus. One may notice that we can add, subtract or multiply matrices but not divide them by other matrices. Matrix inversion is an important operation which is only applicable to *square* and *full rank* matrices. When a square matrix $\boldsymbol{X}$ is full rank then it is referred to as an invertible matrix, where the inverse is defined as:

$$\boldsymbol{X}^{-1} = \frac{1}{\det(\boldsymbol{X})}\mathrm{adj}(\boldsymbol{X}) \tag{2.8.1}$$

where adj($\boldsymbol{X}$) is the adjoint (or adjugate) matrix and for a $2 \times 2$ matrix $\boldsymbol{X}$:

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \tag{2.8.2}$$

is define as:

$$\mathrm{adj}(\boldsymbol{X}) = \begin{bmatrix} x_{22} & -x_{12} \\ -x_{21} & x_{11} \end{bmatrix} \tag{2.8.3}$$

and for a $3 \times 3$ matrix defined as:

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}$$

$$\mathrm{adj}(\boldsymbol{X}) = \begin{bmatrix} + \begin{vmatrix} x_{22} & x_{23} \\ x_{32} & x_{33} \end{vmatrix} & - \begin{vmatrix} x_{21} & x_{23} \\ x_{31} & x_{33} \end{vmatrix} & + \begin{vmatrix} x_{21} & x_{22} \\ x_{31} & x_{32} \end{vmatrix} \\ - \begin{vmatrix} x_{12} & x_{13} \\ x_{32} & x_{33} \end{vmatrix} & + \begin{vmatrix} x_{11} & x_{13} \\ x_{31} & x_{33} \end{vmatrix} & - \begin{vmatrix} x_{11} & x_{12} \\ x_{31} & x_{32} \end{vmatrix} \\ + \begin{vmatrix} x_{12} & x_{13} \\ x_{22} & x_{23} \end{vmatrix} & - \begin{vmatrix} x_{11} & x_{13} \\ x_{21} & x_{23} \end{vmatrix} & + \begin{vmatrix} x_{11} & x_{12} \\ x_{21} & x_{32} \end{vmatrix} \end{bmatrix}$$

where each individual element of the adjoint matrix is itself a determinant of four elements. The sign of each element, in general, is driven by $(-1)^{i+j}$ where $i$ is the row and $j$ is the column identifiers of the element inside the matrix. For example, the first element sits at the first row and first column with $i = 1$ and $j = 1$ hence $(-1)^2 = 1$. The determinant plays an important role in matrix inversion operation. The determinant appears in the denominator which immediately tells us that it cannot be equal to zero. When the determinant is zero, however, it reveals that the original matrix is not full rank and therefore inversion is not possible. Inverse matrices have the following important properties:

$$\boldsymbol{X}\boldsymbol{X}^{-1} = \boldsymbol{I}_n \tag{2.8.4}$$

also

$$(\boldsymbol{X}\boldsymbol{Y})^{-1} = \boldsymbol{Y}^{-1}\boldsymbol{X}^{-1} \tag{2.8.5}$$

if $\boldsymbol{X}$ is a diagonal matrix, such as:

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & 0 & 0 \\ 0 & x_{22} & 0 \\ 0 & 0 & x_{33} \end{bmatrix} \tag{2.8.6}$$

then

$$\boldsymbol{X}^{-1} = \begin{bmatrix} x_{11}^{-1} & 0 & 0 \\ 0 & x_{22}^{-1} & 0 \\ 0 & 0 & x_{33}^{-1} \end{bmatrix} \tag{2.8.7}$$

This result for diagonal matrix holds for any full rank matrix of any size (square with $n = 1, 2, \ldots$). Further, if $\boldsymbol{X}$ is non-symmetric, then:

$$(\boldsymbol{X}^{-1})' = (\boldsymbol{X}')^{-1} \tag{2.8.8}$$

which simplifies to the following if $\boldsymbol{X}$ is symmetric:

$$\boldsymbol{X}^{-1} = (\boldsymbol{X}^{-1})' = (\boldsymbol{X}')^{-1} \tag{2.8.9}$$

## 2.9 Idempotent, Projection

A square (not necessarily symmetric) matrix is called an **idempotent** matrix when its (general) product with itself results in the original matrix. Let $\boldsymbol{M}$ be an idempotent matrix, then:

$$\boldsymbol{M}^2 = \boldsymbol{M} \tag{2.9.1}$$

When an idempotent matrix $\boldsymbol{M}$ is subtracted from the identity matrix $\boldsymbol{I}_n$, the result is also idempotent:

$$\boldsymbol{I}_n - \boldsymbol{M} = (\boldsymbol{I}_n - \boldsymbol{M})(\boldsymbol{I}_n - \boldsymbol{M}) \tag{2.9.2}$$

This is because:

$$\begin{aligned} \boldsymbol{I}_n - \boldsymbol{M} &= \boldsymbol{I}_n\boldsymbol{I}_n - \boldsymbol{M} - \boldsymbol{M} + \boldsymbol{M}^2 \\ &= \boldsymbol{I}_n - 2\boldsymbol{M} + \boldsymbol{M}^2 \\ &= \boldsymbol{I}_n - 2\boldsymbol{M} + \boldsymbol{M} \\ &= \boldsymbol{I}_n - \boldsymbol{M} \end{aligned}$$

Idempotency property has vast applications to econometrics. Consider the following Ordinary Least Squares (OLS) regression:

$$min_{\beta} \quad (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) \tag{2.9.3}$$

where $\boldsymbol{Y}$ is an $n \times 1$ vector of dependent variable, $\boldsymbol{X}$ is an $n \times k$ matrix of independent variable and $\boldsymbol{\beta}$ is a $k \times 1$ vector of regression coefficients. We wish to minimize the the matrix product $(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$ by choosing the regression coefficients $\boldsymbol{\beta}$. To this end, we need to differentiate[3] the objective function with respect to regression coefficients. The resulting solution to this minimization problem is:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \tag{2.9.4}$$

[3] Section (3.2) discusses matrix differential calculus but we only use the final result in this section.

Before discussing the application of idempotent matrix, let's analyze this algebraic form. The last equation is constructed by pre-multiplying an inverse form that is itself product of matrix $\boldsymbol{X}$ by itself. In particular, $\boldsymbol{X}'\boldsymbol{X}$ is a $k \times k$ square matrix. One should first ensure that this square matrix is full rank because the inverse form $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ is defined only if the determinant term $|\boldsymbol{X}'\boldsymbol{X}| \neq 0$. This implies that there should exist $k$ linearly independent vectors within the product form $\boldsymbol{X}'\boldsymbol{X}$. If the determinant is zero, then $\boldsymbol{X}'\boldsymbol{X}$ is not invertible and the solution to the minimization problem is not unique or does not exist. If, however, $\boldsymbol{X}'\boldsymbol{X}$ is full rank (non-zero determinant) then $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ is defined. The second part of the solution involves the term $\boldsymbol{X}'\boldsymbol{Y}$ that is a $k \times 1$ vector, thus the product $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$ that has the dimensions $(k \times n)(n \times k)(k \times n)(n \times 1)$ satisfies the matrix product conformity requirement (note that adjacent dimensions are identical) and the overall product is a $k \times 1$ vector.

Now let's consider the role of idempotent matrix in OLS. The fitted regression's error term can be written in the following way:

$$\widehat{\boldsymbol{e}} \;=\; \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} \tag{2.9.5}$$

that represents the difference between the fitted value $\boldsymbol{X}\widehat{\boldsymbol{\beta}}$ and the dependent variable $\boldsymbol{Y}$. Substituting for $\widehat{\boldsymbol{\beta}}$ from the solution in equation (2.9.4) gives:

$$\widehat{\boldsymbol{e}} \;=\; \boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \tag{2.9.6}$$

Similar to univariate calculus, we can factorize a term. However, this operation breaks into two separate cases: pre-factorizing and post-factorizing. Pre-factorizing (Post-factorizing) is the operation when a variable is factorized from left (right), and is an important distinction to maintain matrix product conformity. For example, we can post-factorize $n \times 1$ vector $\boldsymbol{Y}$ only from right:

$$\begin{aligned} \widehat{\boldsymbol{e}} &= \boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \\ &= (\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{Y} \end{aligned}$$

one can verify this operation by post-multiplying $\boldsymbol{Y}$ to the term in parenthesis, where the product of $\boldsymbol{I}_n\boldsymbol{Y}$ gives $\boldsymbol{Y}$. As suspected, the term in the parenthesis $(\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')$ is an idempotent matrix because the product $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ is an idempotent matrix. To verify this statement, let $\boldsymbol{M} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$, then we can check if $\boldsymbol{M}$ satisfies properties of an idempotent matrix by pre- or post-multiplying it by itself:

$$\boldsymbol{M}^2 \;=\; \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' \;\; \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' \tag{2.9.7}$$

We observe that $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}$ is indeed two pair-wise products that are inverse of each other, hence $\boldsymbol{I}_n = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}$ which simplifies equation (2.9.8) to:

$$\begin{aligned} \boldsymbol{M}^2 &= \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' \\ &= \boldsymbol{M} \end{aligned}$$

which verifies that $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ is an idempotent matrix and also $\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ is also an idempotent matrix. The term $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ is also known as a **projection** matrix which plays a central role econometrics. A projection matrix is a special case of an idempotent matrix or more precisely, a projection matrix is a square, symmetric and idempotent matrix. A projection matrix is commonly denoted by $\boldsymbol{P}$ and because it inherits properties of idempotent matrices, then:

$$\boldsymbol{P}^2 \;=\; \boldsymbol{P} \;=\; \boldsymbol{P}'\boldsymbol{P} \;=\; \boldsymbol{P}\boldsymbol{P}' \;=\; \boldsymbol{P}' \tag{2.9.8}$$

also $tr(\boldsymbol{P}) = rank(\boldsymbol{P})$.

## 2.10 Quadratic Form

This section builds upon general matrix product[4] to extend linear algebra to terms that are essentially *quadratic*. Specifically, a *quadratic form* is a quadratic function of the elements in an $n \times 1$ matrix (vector). Such forms arise frequently in econometrics and applied statistical models. Intuitively, a quadratic form is generalization of $(x_1 + x_2)^2 = x_1^2 + x_1x_2 + x_2x_1 + x_2^2$ where it is assumed that all coefficients of variables are equal to one. In a more general case,

$$\begin{aligned} F(x_1, x_2) &= a_{11}x_1^2 + a_{12}x_1x_2 + a_{21}x_2x_1 + a_{22}x_2^2 \\ &= \sum_{i=1}^{2}\sum_{j=1}^{2} a_{ij}x_ix_j \end{aligned}$$

where the arguments of the left-hand-side function is a scalar $N$-to-1 mapping function of vectors and can be written with vector notation $\boldsymbol{x}$ $(n \times 1)$ for its arguments:

$$\begin{aligned} F(\boldsymbol{x}) &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j \\ &= a_{11}x_1^2 + a_{12}x_1x_2 + \ldots + a_{1n}x_1x_n + \\ &\quad a_{21}x_2x_1 + a_{22}x_2^2 + \ldots + a_{2nx_2x_n} + \\ &\quad \ldots \\ &\quad a_{n1}x_nx_1 + a_{n2}x_nx_2 + \ldots + a_{nn}x_n^2 \end{aligned} \tag{2.10.1}$$

Note that the value of $F(\boldsymbol{x})$ varies when $\boldsymbol{x}$ change. The main premise of linear algebra is its compactness advantage which is absent in the previous equation. However, equation (2.10.1) can be re-written using the following notation. Let,

$$c_i = \sum_j a_{ij}x_j \tag{2.10.2}$$

or in vector notation,

$$\boldsymbol{c} = \boldsymbol{Ax} \tag{2.10.3}$$

where $\boldsymbol{A} = \{a_{ij}\}$, for $i = 1, 2, \ldots, n$, hence:

$$\begin{aligned} F(\boldsymbol{x}) &= \sum_{i=1}^{2}\sum_{j=1}^{2} a_{ij}x_ix_j \\ &= \sum_{i=1}^{n}\left\{a_{ij}\sum_{j=1}^{n} x_i\right\} \\ &= \sum_{i=1}^{n} x_ic_i \\ &= \boldsymbol{x}'\boldsymbol{c} \\ &= \boldsymbol{x}'\boldsymbol{Ax} \end{aligned}$$

[4] Henceforth, matrix product

where the compact form $F(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ is referred to as the quadratic form of $\boldsymbol{x}$. We verify that $F(\boldsymbol{x})$ is indeed a scalar by simplifying the dimensions of right hand side that is to say $(1\times n)(n\times n)(n\times 1)$ that collapses to $1\times 1$. Interestingly, the quadratic form $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ is equal to its transpose. This can be seen through applying the transpose operation and property (2.7.7), $(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x})' = \boldsymbol{x}'\boldsymbol{A}'\boldsymbol{x}$ which implies that matrix $\boldsymbol{A}$ is symmetric. This is a non-trivial property but let's analyze it a bit further to see why matrix $\boldsymbol{A}$ is symmetric. We can write $2F(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{A}'\boldsymbol{x}$, then pre- and post-factorizing gives:

$$2F(\boldsymbol{x}) = \boldsymbol{x}'(\boldsymbol{A}+\boldsymbol{A}')\boldsymbol{x}$$

let $\boldsymbol{B} = \frac{1}{2}(\boldsymbol{A}+\boldsymbol{A}')$ and re-write:

$$\begin{aligned} 2F(\boldsymbol{x}) &= \boldsymbol{x}'(\boldsymbol{A}+\boldsymbol{A}')\boldsymbol{x} \\ &= \boldsymbol{x}'(2\boldsymbol{B})\boldsymbol{x} \end{aligned}$$

hence:

$$F(\boldsymbol{x}) = \boldsymbol{x}'(\boldsymbol{B})\boldsymbol{x}$$

the key is that the matrix $\boldsymbol{B} = \frac{1}{2}(\boldsymbol{A}+\boldsymbol{A}')$ is always symmetric$\square$. That is to say any quadratic form with a coefficient matrix, can be re-expressed with a symmetric coefficient matrix. An application to econometrics is that the following two quadratic forms are indeed identical:

$$\boldsymbol{x}'\begin{bmatrix} 4 & 2 \\ 4 & 6 \end{bmatrix}\boldsymbol{x}, \quad \boldsymbol{x}'\begin{bmatrix} 4 & 3 \\ 3 & 6 \end{bmatrix}\boldsymbol{x} \tag{2.10.4}$$

because

$$\frac{1}{2}\times\left(\begin{bmatrix} 4 & 2 \\ 4 & 6 \end{bmatrix} + \begin{bmatrix} 4 & 4 \\ 2 & 6 \end{bmatrix}\right) = \begin{bmatrix} 4 & 3 \\ 3 & 6 \end{bmatrix} \tag{2.10.5}$$

## 2.11 Definiteness, Indefiniteness

Consider a general quadratic form $F(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ for an arbitrary real-valued symmetric coefficient matrix $\boldsymbol{A}$. This section investigates the *sign* of $F(\boldsymbol{x})$ given any possible choice of values for elements within $\boldsymbol{x}$:

1. If $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0$, for all $\boldsymbol{x} \neq \boldsymbol{0}_n$, the quadratic form and matrix $\boldsymbol{A}$ are referred to as *positive definite* (pd).
2. If $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \geq 0$, for all $\boldsymbol{x} \neq \boldsymbol{0}_n$, the quadratic form and matrix $\boldsymbol{A}$ are referred to as *positive semi-definite* (psd).
3. If $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} < 0$, for all $\boldsymbol{x} \neq \boldsymbol{0}_n$, the quadratic form and matrix $\boldsymbol{A}$ are referred to as *negative definite* (nd).
4. If $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \leq 0$, for all $\boldsymbol{x} \neq \boldsymbol{0}_n$, the quadratic form and matrix $\boldsymbol{A}$ are referred to as *negative semi-definite* (nsd).
5. If $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$, can be positive or negative, the quadratic form and matrix $\boldsymbol{A}$ are referred to as *indefinite*.

Let's restrict the scope to $2\times 2$ matrices. A quadratic form is,

1. p.d. ($\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0$, for all $\boldsymbol{x} \neq \boldsymbol{0}_n$) iff $a_{11} > 0$ and $det(\boldsymbol{A}) > 0$
2. p.s.d. ($\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \geq 0$, for all $\boldsymbol{x} \neq \boldsymbol{0}_n$) iff $a_{11} \geq 0$ and $det(\boldsymbol{A}) \geq 0$
3. n.d. ($\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} < 0$, for all $\boldsymbol{x} \neq \boldsymbol{0}_n$) iff $a_{11} < 0$ and $det(\boldsymbol{A}) > 0$

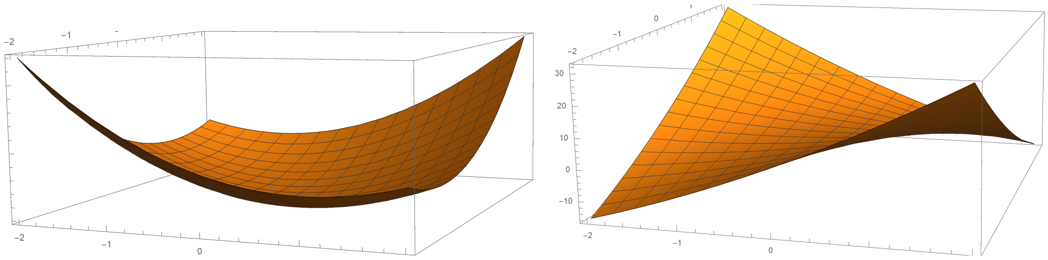


**Figure 2.3:** Figure on the left shows a pd quadratic form since the surface tends to increase from all edges and hence the entire surface value is above the zero plane. Figure on the right is an indefinite quadratic form since the surface edges tend to increase from two corners but tend to decrease from the other two corners. Hence part of the surface sits above the zero plane but another part fall below.

4. n.s.d. ($\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \leq 0$, for all $\boldsymbol{x} \neq \mathbf{0}_n$) iff $a_{11} \leq 0$ and $det(\boldsymbol{A}) \geq 0$
5. indefinite iff $det(\boldsymbol{A}) < 0$

## 2.12 Exercises

**Question 1:** Let

$$\boldsymbol{X} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \boldsymbol{Y} = \begin{bmatrix} 5 & 4 & 3 \\ 1 & 1 & 1 \\ 2 & 2 & 2 \\ 0 & 0 & 1 \end{bmatrix}$$

(1.1) What are the transpose transformations $\boldsymbol{X}'$ and $\boldsymbol{Y}'$?
(1.2) Is matrix $\boldsymbol{X}'$ symmetric?
(1.3) In how many ways can $\boldsymbol{X}$ and $\boldsymbol{Y}$ (or their transpose transformations) can be multiplied (general multiplication)?
(1.4) How many linearly independent vectors can you find within $\boldsymbol{X}$?
(1.5) What is the determinant of $\boldsymbol{X}$, $det(\boldsymbol{X})$?
(1.6) Find the rank of $\boldsymbol{X}$, $rank(\boldsymbol{X})$?
(1.7) Find the rank of $\boldsymbol{X}\boldsymbol{Y}'$?
(1.8) What is the univariate analogue of $\mathbf{1}'_{3\times1}\boldsymbol{Y}'\mathbf{1}_{4\times1}$?
(1.9) Find the norm of $\boldsymbol{X}$?
(1.10) Is $\boldsymbol{X}$ invertible, why?
(1.11) Is $\boldsymbol{X}'\boldsymbol{X}$ invertible, why?
(1.12) Is $\boldsymbol{X}$ a singular matrix?

**Question 2:** Matrix inversion and rank:
(2.1) Find the inverse form of $\boldsymbol{I}_2$
(2.2) Find the inverse form of $\mathbf{1}_{2\times2}$
(2.3) What is the rank of $\mathbf{0}_{n\times n}$?

**Question 3:** Let

$$X = \begin{bmatrix} 2 & -2 & -4 \\ -1 & 3 & 4 \\ 1 & -2 & -3 \end{bmatrix}$$

(3.1) Find the rank and determinant of $\boldsymbol{X}$.
(3.2) Is $\boldsymbol{X}$ an idempotent matrix?

**Question 4:** Let

$$X = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad Y = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

(4.1) What is $(\boldsymbol{X}'\boldsymbol{X})^{-1}$.
(4.2) Find $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$.
(4.3) Denote $\boldsymbol{M} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ and find $\boldsymbol{MY}$.
(4.4) Find $\boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$.
(4.5) Find $(\boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y})'(\boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y})$.

**Question 5:** Let

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} \quad Y = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

(5.1) What is $(\boldsymbol{X}'\boldsymbol{X})^{-1}$.
(5.2) Find $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$.
(5.3) Denote $\boldsymbol{M} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ and find $\boldsymbol{MY}$.
(5.4) Find $\boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$
(5.5) Find $(\boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y})'(\boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y})$.
(5.6) Compare scalars found in (5.5) and (4.5) and comment.

**Question 6:** Let

$$X = \begin{bmatrix} 4x & 0 & 0 \\ x & -y & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

(6.1) Find $det(x)$.
(6.2) What condition is required to ensure $\boldsymbol{X}^{-1}$ exists?
(6.3) How does the answer to previous parts change if $x = y$?
(6.4) What condition is required to ensure $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ exists?
(6.5) Find $rank(\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')$ when $x = y = 1$.
(6.6) Find $tr(\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')$ and $det(\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')$when $x = y = 1$.

**Question 7:** Let

$$\boldsymbol{X} = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 0 & 2 \\ 0 & 2 & 1 \end{bmatrix} \quad \boldsymbol{Y} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 4 & 1 \\ 1 & 3 & 0 \end{bmatrix} \quad \boldsymbol{Z} = \begin{bmatrix} z & 0 & 1 \\ 1 & 3 & 0 \\ 1 & 4z & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

(7.1) Find if $\boldsymbol{X}'\boldsymbol{X}$ and $\boldsymbol{Y}'\boldsymbol{Y}$ are positive (semi-) definite, negative (semi-) definite or indefinite.
(7.2) Show that $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ an idempotent matrix.
(7.3) What condition on $z$ is required such that $\boldsymbol{Z}'\boldsymbol{Z}$ is positive definite?

**Question 8:** Let $\Omega$ be an $N \times N$ positive definite (PD) matrix that is $b'\Omega b > 0$ (for any non-null $N \times 1$ vector $b$). Define a $T \times N$ matrix $\Psi$ with $rank(\Psi) = N$. Show that $\Psi'\Psi$ is a PD matrix.

## 2.13 Solution to Exercises

**Question 1:**

(1.1) Transpose transformations $\boldsymbol{X}'$ and $\boldsymbol{Y}'$ are:

$$\boldsymbol{X}' = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \boldsymbol{Y}' = \begin{bmatrix} 5 & 1 & 2 & 0 \\ 4 & 1 & 2 & 0 \\ 3 & 1 & 2 & 1 \end{bmatrix}$$

(1.2) Matrix $\boldsymbol{X}'$ is symmetric because $\boldsymbol{X} = \boldsymbol{X}'$.

(1.3) $\boldsymbol{X}$ is $3 \times 3$ and $\boldsymbol{Y}$ is $4 \times 3$, noting the conformity property, we can have $\boldsymbol{YX}$, $\boldsymbol{XY}'$ and $\boldsymbol{X}'\boldsymbol{Y}'$.

(1.4) None, since the determinant is equal to $2 = 2 \times 1 \times 1$ which is non-zero.

(1.5) $det(\boldsymbol{X}) = 3$.

(1.6) $rank(\boldsymbol{X}) = 3$ because this matrix has non-zero determinant (it is full-rank or non-singular), hence there exists three linearly independent column (row) vectors within the matrix.

(1.7) Second and third rows of $\boldsymbol{Y}$ are linearly dependent. After eliminating one of vectors of $\boldsymbol{Y}$, there exists only three linearly independent vectors within $\boldsymbol{Y}$ which indicates $rank(\boldsymbol{Y}) = 3$. The product $\boldsymbol{YX}$ is $3 \times 3$ and $rank(\boldsymbol{YX}) = 3$ since the product of two matrices, each with $rank = 3$ amounts to a $rank = 3$ matrix.

(1.8) Re-write $\mathbf{1}_{3\times1}\boldsymbol{Y}'\mathbf{1}_{4\times1}$ operation without using vectors/matrices operations. The special point about this is that for matrices we have to pre- and post-multiply matrix $\boldsymbol{Y}'$ with vectors of ones (e.g. $\mathbf{1}'_{3\times1}$ and $\mathbf{1}_{4\times1}$):

$$\begin{aligned} \mathbf{1}'_{3\times1}\boldsymbol{Y}'\mathbf{1}_{4\times1} &= \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 5 & 1 & 2 & 0 \\ 4 & 1 & 2 & 0 \\ 3 & 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 5+1+2+0 \\ 4+1+2+0 \\ 3+1+2+1 \end{bmatrix} \end{aligned}$$

where from the first line to the second, we post-multiplied only $\boldsymbol{Y}'\mathbf{1}_{4\times1}$. As we can see:

$$\begin{bmatrix} 5+1+2+0 \\ 4+1+2+0 \\ 3+1+2+1 \end{bmatrix}$$

is, first, a $3 \times 1$ vector with each element equal to sum of row elements of $\boldsymbol{Y}'$. If we apply the pre-multiplication $\mathbf{1}'_{3\times1}\,(\boldsymbol{Y}'\mathbf{1}_{4\times1})$:

$$\begin{aligned} \mathbf{1}'_{3\times1}\boldsymbol{Y}'\mathbf{1}_{4\times1} &= \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 5+1+2+0 \\ 4+1+2+0 \\ 3+1+2+1 \end{bmatrix} \\ &= \begin{bmatrix} 5+1+2+0+4+1+2+0+3+1+2+1 \end{bmatrix} \end{aligned}$$

Last line is only the sum of all elements of the original matrix $\boldsymbol{Y}$, which we can re-do using the sum operation:

$$\sum_{i=1}^{4}\sum_{j=1}^{3} x_{x_{ij}} = 5+1+2+0+4+1+2+0+3+1+2+1$$

Note that we need to sum over rows, and then sum over columns (or vice versa), that is why we have the double sum operator.

(1.9) There are two ways of working out the norm of a matrix, if the matrix is a diagonal matrix (which is also square because it is diagonal). The quadratic form under the square root $\boldsymbol{X'X}$ is:

$$\begin{aligned} \boldsymbol{X'X} &= \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 2^2 & 0 & 0 \\ 0 & 1^2 & 0 \\ 0 & 0 & 1^2 \end{bmatrix} \end{aligned}$$

and then:

$$\begin{aligned} \sqrt{tr(\boldsymbol{X'X})} &= \left[ tr\left( \begin{bmatrix} 2^2 & 0 & 0 \\ 0 & 1^2 & 0 \\ 0 & 0 & 1^2 \end{bmatrix} \right) \right]^{\frac{1}{2}} \\ &= \sqrt{2^2+1^2+1^2} \end{aligned}$$

which is equivalent to $||\boldsymbol{X}||_2 = \sqrt{2^2+1^2+1^2}$.

(1.10) Yes, because $det(\boldsymbol{X}) \neq 0$.

(1.11) Yes, because $\boldsymbol{X'X}$ is full rank and therefore invertible.

(1.12) $\boldsymbol{X}$ is non-singular because $det(\boldsymbol{X}) \neq 0$.

**Question 2:**

(2.1) The inverse form of $\boldsymbol{I}_2$ is $\boldsymbol{I}_2$ because it is an identity matrix.

(2.2) The inverse form of $\mathbf{1}_{2\times 2}$ is not defined because matrix one, has all elements equal to one which makes all of its columns (rows) linearly dependent and thus it is not full rank, and non-invertible.

(2.3) The null matrix is by definition a rank-zero matrix: $rank(\mathbf{0}_{n\times n}) = 0$

**Question 3:**

(3.1) $det(\boldsymbol{X}) = 0$, and $rank(\boldsymbol{X}) = 2$ (a singular matrix or not full rank). The maximum rank possible for this matrix is 3, however, since the determinant is zero, there exists at least one linear combination among rows or columns (in this case $C_3 = C_2 - 2C_1$). After establishing this linear combination, no further linear combination is possible and rank becomes 2, as there are only two linearly independent vectors within $\boldsymbol{X}$, $rank(\boldsymbol{X}) = 2$

(3.2) Yes, because $\boldsymbol{XX} = \boldsymbol{X}$.

**Question 4:** Let

$$\boldsymbol{X} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad \boldsymbol{Y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

(4.1) $(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$.

(4.2) $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} = (3, -1, -1)'$.

(4.3) $\boldsymbol{MY} = (3, 2, 1)'$.

(4.4) $\boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} = (0, 0, 0)'$.

(4.5) $(0, 0, 0)'(0, 0, 0) = 0$ (scalar).

# Chapter 3

# Matrix Differential Calculus

This chapter incorporates differential calculus into matrix algebra and econometrics. Differentiation is an essential operation in econometric methods (such as ordinary least squares, maximum likelihood or generalized method of moments) which involve estimation of a parameter that is researcher's object of interest. Differential calculus serves as a search method to find suitable parameter values.

The following three assumptions hold throughout this chapter. First, note that continuity is a necessary condition before differentiation can be applied and it is always assumed to be satisfied in this chapter unless mentioned otherwise. Second, similar to previous chapters, we restrict the scope to real-valued scalars, vectors, matrices and functions of such objects. Lastly, when we apply differentiation to product rules, it only focuses on general matrix multiplication.

## 3.1 Gradient and Derivative Vectors

Consider the following three notations: a univariate function $f(.)$ (scalar-valued, $1 \times 1$), vector function $\boldsymbol{f}(.)$ (vector-valued, e.g. $n \times 1$), and matrix function $\boldsymbol{F}(.)$ (e.g. $n \times p$). Also, let's establish the following: $x$ is a scalar argument, $\boldsymbol{x} = (x_1, x_2, \dots, x_m)'$ is an $m \times 1$ vector-argument, and $\boldsymbol{X} = \{x_{ij}\}_{i=1,2,\dots,m}^{j=1,2,\dots,q}$ is an $m \times q$ matrix-argument. Table (3.1) illustrates three functional forms $f$, $\boldsymbol{f}$ and $\boldsymbol{F}$ where each of them can take input arguments of different dimensions $x$, $\boldsymbol{x}$ and $\boldsymbol{X}$.

| | | Arguments | | |
|---|---|---|---|---|
| Function | Size | Scalar $(x_{1\times1})$ | Vector $(\boldsymbol{x}_{m\times1})$ | Matrix $(\boldsymbol{X}_{m\times q})$ |
| $f(.)$ | $1 \times 1$ | $x^2$ | $\boldsymbol{x'x}$ | $\mathbf{1}_{1\times m}\boldsymbol{X}\mathbf{1}_{q\times1}$ |
| $\boldsymbol{f}(.)$ | $n \times 1$ | $x^2\mathbf{1}_{n\times1}$ | $\mathbf{1}_{n\times m}\boldsymbol{x}\mathbf{1}_{1\times1}$ | $\mathbf{1}_{n\times m}\boldsymbol{X}\mathbf{1}_{q\times1}$ |
| $\boldsymbol{F}(.)$ | $n \times p$ | $x^2\mathbf{1}_{n\times p}$ | $\mathbf{1}_{n\times m}\boldsymbol{x}\mathbf{1}_{1\times p}$ | $\mathbf{1}_{n\times m}\boldsymbol{X}\mathbf{1}_{q\times p}$ |

**Table 3.1:** There is no restriction on the size of arguments of functions, and the important factor is the functional form itself that determines the overall dimension of function.

The norm operator $||.||$ is always a scalar value (regardless of argument's dimension) and can be attributed to $f(.)$. Inner product, or quadratic form of two vectors is also always a scalar and can be attributed to $f(.)$, Similarly, trace, $tr(.)$ of a square and $det(.)$ of a (non-singular) matrix is a scalar function.

Dimensions of functions and their arguments suggest that there exists nine possibilities to differentiate a function in matrix algebraic context. The simplest case, which is differentiating a scalar function with respect to its scalar argument coincides with conventional univariate calculus differentiation which is a trivial case in matrix context. The second case, however, falls outside univariate calculus and requires differentiation to be extended to matrix context.

**Scalar Function of Vector Arguments:** This section focuses on a scalar-valued function $f(\boldsymbol{x})$ that depends on elements in its vector argument $\boldsymbol{x}$. We assume that $\boldsymbol{x}$ is at least a 2-dimensional vector (or $m$-dimensional with $m \geq 2$) indicating that the scalar function is a bi-variate or multivariate function and therefore it has *partial* derivatives with respect to each of its individual elements. The *partial derivative* vector of a function $f : \mathbb{R}^m \to \mathbb{R}$ or $f(\boldsymbol{x})$ that maps an $m$-dimensional (column) vector argument $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)'$ to a scalar value is given by:

$$\begin{aligned} Df(\boldsymbol{x}) &\equiv \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}'} \\ &= \left(\frac{\partial f(\boldsymbol{x})}{\partial x_1}, \frac{\partial f(\boldsymbol{x})}{\partial x_2}, \ldots, \frac{\partial f(\boldsymbol{x})}{\partial x_m}\right) \end{aligned} \tag{3.1.1}$$

which is an $1 \times m$ (row) vector. The gradient vector of $f(\boldsymbol{x})$ is expressed as below:

$$\nabla f(\boldsymbol{x}) \ = \ \boldsymbol{g}(\boldsymbol{x}) \ = \ \boldsymbol{\nabla}(\boldsymbol{x}) \tag{3.1.2}$$

This notation can be summarized in the following way:

$$\nabla f(\boldsymbol{x}) \ = \ (Df(\boldsymbol{x}))' \ = \ (D_{\boldsymbol{x}} f(\boldsymbol{x}))' \ = \ \boldsymbol{g}(\boldsymbol{x}) \ = \ \boldsymbol{\nabla}(\boldsymbol{x}) \tag{3.1.3}$$

where different authors may use one of the notations[1] in the equation above[2]. The gradient vector is an $m \times 1$ (column) vector. Intuitively, the scalar function value, $f(\boldsymbol{x})$, varies when each of the individual arguments $x_i$, $i = 1, 2, \ldots, m$ changes, and each individual gradient is expressed by the respective elements in $\nabla f(\boldsymbol{x})$ vector.

Differentiation of scalar-valued functions of vectors benefits from the following lemma.

**Lemma 1:** Suppose $\boldsymbol{a}$ is $m \times n$, $\boldsymbol{b}$ is $n \times q$ and $\boldsymbol{c}$ is $q \times r$, then:

$$(\boldsymbol{abc})' \ = \ \boldsymbol{c}'\boldsymbol{b}'\boldsymbol{a}' \tag{3.1.4}$$

is an $m \times r$ matrix. In general,

$$\boldsymbol{abc} \ \neq \ \boldsymbol{c}'\boldsymbol{b}'\boldsymbol{a}' \tag{3.1.5}$$

but if this expression is a scalar, with $m = 1$ and $r = 1$, then $\boldsymbol{a}$ is $1 \times n$ and $\boldsymbol{c}$ is $p \times 1$:

$$(\boldsymbol{abc})' \ = \ \boldsymbol{abc} \ = \ \boldsymbol{c}'\boldsymbol{b}'\boldsymbol{a}' \tag{3.1.6}$$

Consider the following inner product of two $m$-dimensional vectors, $\boldsymbol{a} = (a_1, a_2, \ldots, a_m)'$ and $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)'$:

$$f(\boldsymbol{x}; \boldsymbol{a}) \ = \ \boldsymbol{a}'\boldsymbol{x} \ = \ \sum_{i=1}^{m} a_i x_i \tag{3.1.7}$$

[1]Matrix differential calculus is vastly used across different disciplines, and it is common to see various notations because one letter or operator may be reserved in a discipline for a certain use.

[2]Operator, or letter, $\nabla$ is pronounced, 'del' or 'nabla'.

individual derivatives are $\partial f/\partial x_i = a_i$, and then:

$$\begin{aligned}
Df(\boldsymbol{x};\boldsymbol{a}) &= \frac{\partial \boldsymbol{a}'\boldsymbol{x}}{\partial \boldsymbol{x}'}\\
&= \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_m}\right)\\
&= (a_1, a_2, \ldots, a_m)\\
&= \boldsymbol{a}'
\end{aligned}$$

which a row vector and the gradient vector is:

$$\begin{aligned}
\nabla f(\boldsymbol{x};\boldsymbol{a}) &= (Df(\boldsymbol{x}))'\\
&= \frac{\partial \boldsymbol{a}'\boldsymbol{x}}{\partial \boldsymbol{x}}\\
&= \boldsymbol{a}
\end{aligned}$$

which is a column vector. Suppose also that $\boldsymbol{y}$ is $m \times 1$ and $\boldsymbol{A}$ is a non-symmetric $m \times m$ matrix, the gradient column vector of function $f(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{y}'\boldsymbol{A}\boldsymbol{x}$ with respect to $\boldsymbol{x}$ and $\boldsymbol{y}$ are:

$$\begin{aligned}
\nabla_{\boldsymbol{x}} f(\boldsymbol{x},\boldsymbol{y}) &= \frac{\partial \boldsymbol{y}'\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}}\\
&= \boldsymbol{y}'\boldsymbol{A} \qquad (3.1.8)
\end{aligned}$$

$$\begin{aligned}
\nabla_{\boldsymbol{y}} f(\boldsymbol{x},\boldsymbol{y}) &= \frac{\partial \boldsymbol{y}'\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{y}}\\
&= \boldsymbol{x}'\boldsymbol{A}' \qquad (3.1.9)
\end{aligned}$$

Now assume that $\boldsymbol{y} = \boldsymbol{x}$, that is to say $f(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ is a quadratic form in $\boldsymbol{x}$. The gradient vector with respect to $\boldsymbol{x}$ now involves differentiating $f(\boldsymbol{x})$ twice, because $\boldsymbol{x}$ appears twice in the quadratic form. This can simply be done by treating each $\boldsymbol{x}$ in the quadratic form as a separate variable (equations (3.1.8) and (3.1.9)):

$$\begin{aligned}
\frac{\partial \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} &= \boldsymbol{x}'\boldsymbol{A} + \boldsymbol{x}'\boldsymbol{A}'\\
&= \boldsymbol{x}'(\boldsymbol{A} + \boldsymbol{A}')
\end{aligned}$$

If $\boldsymbol{A}$ is symmetric, then $\boldsymbol{A} = \boldsymbol{A}'$:

$$f(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \qquad (3.1.10)$$

$$\frac{\partial \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = 2\boldsymbol{x}'\boldsymbol{A} \qquad (3.1.11)$$

Before proceeding to the next section, let's establish the following properties:

$$\nabla_x(f(\boldsymbol{x}) + h(\boldsymbol{x})) = \nabla_x f(\boldsymbol{x}) + \nabla_x h(\boldsymbol{x}) \qquad (3.1.12)$$

$$\nabla_x(\lambda f(\boldsymbol{x})) = \lambda \nabla_x f(\boldsymbol{x}) \qquad (3.1.13)$$

where $\lambda$ is constant scalar.

**Vector Function of Vector Arguments:** This section focuses on an $n \times 1$ vector-valued function $\boldsymbol{f}(\boldsymbol{x})$ that depends on elements in its vector argument $\boldsymbol{x}$, that is an $m$-dimensional vector ($m \geq 2$). Specifically, $\boldsymbol{f}(\boldsymbol{x})$ is defined to be $f\mathbb{R}^m \to \mathbb{R}^n$. Intuitively, $\boldsymbol{f}(\boldsymbol{x})$ collect $n$ individual function $\{f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \ldots, f_n(\boldsymbol{x})\}$ and where each individual function is multivariate function with $n$ partial derivatives. The array of all partial derivatives is an $n \times m$ matrix and is called the **jacobian** matrix:

$$\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}'} = D\boldsymbol{f}(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f_1}{\partial \boldsymbol{x}'} \\ \frac{\partial f_2}{\partial \boldsymbol{x}'} \\ \vdots \\ \frac{\partial f_n}{\partial \boldsymbol{x}'} \end{bmatrix} = \begin{bmatrix} Df_1(\boldsymbol{x}) \\ Df_2(\boldsymbol{x}) \\ \vdots \\ Df_n(\boldsymbol{x}) \end{bmatrix} \tag{3.1.14}$$

which is an $n \times m$ matrix with elements $\left\{\frac{\partial f_i}{\partial x_j}\right\}$, $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$. The *gradient* is the following $m \times n$ matrix:

$$\nabla \boldsymbol{f}(\boldsymbol{x}) = (D\boldsymbol{f}(\boldsymbol{x}))' \tag{3.1.15}$$

## 3.2 Ordinary Least Squares

Recall that in the previous chapter, equation (2.9.4) was the solution to Ordinary Least Square (OLS) problem. The most suitable value for OLS regression parameter vector $\widehat{\boldsymbol{\beta}}$ is driven by differentiation of the OLS objective function:

$$\min_{\beta} \quad (\boldsymbol{Y} - \boldsymbol{X\beta})'(\boldsymbol{Y} - \boldsymbol{X\beta}) \tag{3.2.1}$$

Simplifying the objective function:

$$\begin{aligned} \min_{\beta} \quad & (\boldsymbol{Y}' - \boldsymbol{\beta}'\boldsymbol{X}')(\boldsymbol{Y} - \boldsymbol{X\beta}) \\ & \boldsymbol{Y}'\boldsymbol{Y} - \boldsymbol{Y}'\boldsymbol{X\beta} - \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{Y} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X\beta} \end{aligned}$$

which is a $1 \times 1$ scalar value equal to sum of squared regression residuals. Note that $-\boldsymbol{Y}'\boldsymbol{X\beta}$ and $-\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{Y}$ are in fact equal (using Lemma 1) and can be written as:

$$\min_{\beta} \quad \boldsymbol{Y}'\boldsymbol{Y} - 2\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{Y} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X\beta}$$

Differentiating with respect to $\boldsymbol{\beta}$ gives:

$$\frac{\partial(\boldsymbol{Y}' - \boldsymbol{\beta}'\boldsymbol{X}')'(\boldsymbol{Y} - \boldsymbol{X\beta})}{\partial \boldsymbol{\beta}} = -2\boldsymbol{X}'\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{X\beta} \tag{3.2.2}$$

This is the first order condition (FOC) of the OLS problem which is also referred to as system of normal equations. In order to solve for the most suitable value of the regression parameters, the FOC must be set equal to zero to find a flat tangent hyperplane to the objective function where the value of the objective function is globally minimized:

$$0 = -2\boldsymbol{X}'\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}}$$

Re-arrange:

$$\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{y}$$

and therefore:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \tag{3.2.3}$$

We can step an additional step and take derivative of (3.2.2) to examine second order conditions (SOC) of the least squares problem. The SOC together with results in section (2.11) can determine if the solution to the least squares (3.2.3) is a minimizer.

$$\partial\frac{-2\boldsymbol{X}'\boldsymbol{y}+2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}}{\partial\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{X} \tag{3.2.4}$$

which is a quadratic form thus $det(\boldsymbol{X}'\boldsymbol{X}) \geq 0$. However, we require that $det(\boldsymbol{X}'\boldsymbol{X}) \neq 0$ since the term $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ in equation (3.2.3) is only defined when $\boldsymbol{X}'\boldsymbol{X}$ is non-singular and therefore the SOC results in a positive-definite quadratic form that confirms that:

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\beta} \ (\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}) \tag{3.2.5}$$

# Summary

Consider an $n \times 1$ vector $\boldsymbol{x}$, and $n \times 1$ vector $\boldsymbol{y}$ then:

(3.1) The inner product $\boldsymbol{x}'\boldsymbol{y} = 0$ iff both vectors are orthogonal

(3.2) The magnitude (length) or Euclidean norm of $\boldsymbol{x}$ is:

$$||\boldsymbol{x}|| = \sqrt{\boldsymbol{x}'\boldsymbol{x}} \tag{3.2.6}$$

(3.3) For any matrix $\boldsymbol{A}$, the Euclidean norm is:

$$||\boldsymbol{A}|| = \sqrt{tr(\boldsymbol{A}'\boldsymbol{A})} = \sqrt{tr(\boldsymbol{A}\boldsymbol{A}')} \tag{3.2.7}$$

(3.4) A square matrix is symmetric if $\boldsymbol{A} = \boldsymbol{A}'$

(3.5) In general, $(\boldsymbol{AB})' = \boldsymbol{B}'\boldsymbol{A}'$

(3.6) In general, $(\boldsymbol{X}_1\boldsymbol{X}_2)\boldsymbol{X}_3 = \boldsymbol{X}_1(\boldsymbol{X}_2\boldsymbol{X}_3)$ but $\boldsymbol{AB} \neq \boldsymbol{BA}$

(3.7) If $\boldsymbol{AB}$ is defined, then $tr(\boldsymbol{AB}) = tr(\boldsymbol{BA})$, also $tr(\boldsymbol{ABC}) = tr(\boldsymbol{CBA})$

(3.8) Only square matrix can be considered for inversion. The square matrix must be full-rank or non-singular for the inverse to exist.

(3.9) If the inverse form exists $\boldsymbol{A}^{-1}$ then $\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}_n$

(3.10) If $\boldsymbol{A}$ is non-singular, then $(\boldsymbol{A}^{-1})' = (\boldsymbol{A}')^{-1}$

(3.11) If both $\boldsymbol{A}$ and $\boldsymbol{B}$ are non-singular, then $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$

(3.12) An idempotent matrix is a square matrix satisfying $\boldsymbol{MM} = \boldsymbol{M}$ with $tr\boldsymbol{M} = rank(\boldsymbol{M})$