Queen Mary
University of London

Maths & Stats Pre-Sessional

Derivatives

Lecturer: Claudio Vallar
School of Economics and Finance

Basics of Derivatives and Matrix Algebra

In this session:

- Derivative and differentiation for functions of one variable
- Optimization of functions of one variable
- Matrices
- System of linear equations

For more extensive reading, refer to Chapters 5 to 9 of Hoy, M., Livernois, J., McKenna, C., Rees, R., and Stengos, R. (2011). Mathematics for Economics, MIT Press, 3rd Edition

What is a derivative?

- The derivative of a function of a real variable measures the sensitivity of change of a quantity (a function or dependent variable) which is determined by another quantity (the independent variable).

- The process of finding a derivative is called differentiation.
- The reverse process is called integration.
- Differentiation is also known as the process to find rate of change.
- Derivative tells us the slope of the function at any point.

What is a derivative?

Given a function $y = f(x)$, as $y = 3x^2$

Question: What is the marginal impact of x on y? In other words, what is the marginal variation of y due to a marginal variation of x?

Answer: this impact is what we call derivative and can be calculated as:

$$f'(x) = \frac{dy}{dx} = 6x$$

It tells us that, when x changes by 1, y changes by 1 times 6.

What is a derivative?

The geometric meaning of the derivative:

$$f'(x) = \frac{df(x)}{dx}$$

Is the slope of the tangent line $y = f(x)$ at x

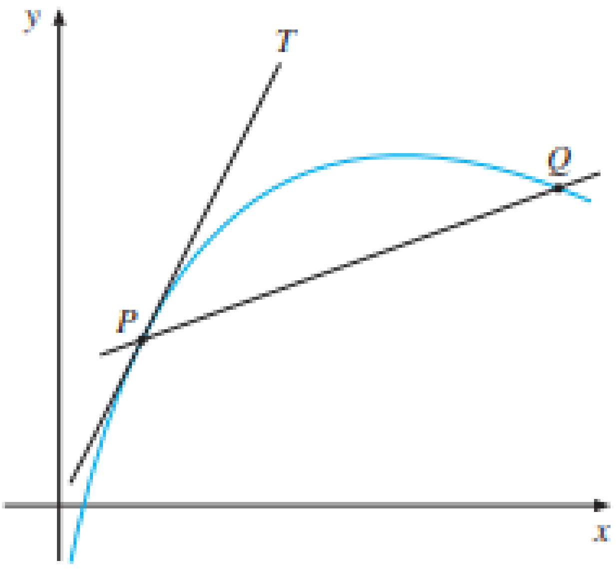


Figure 1

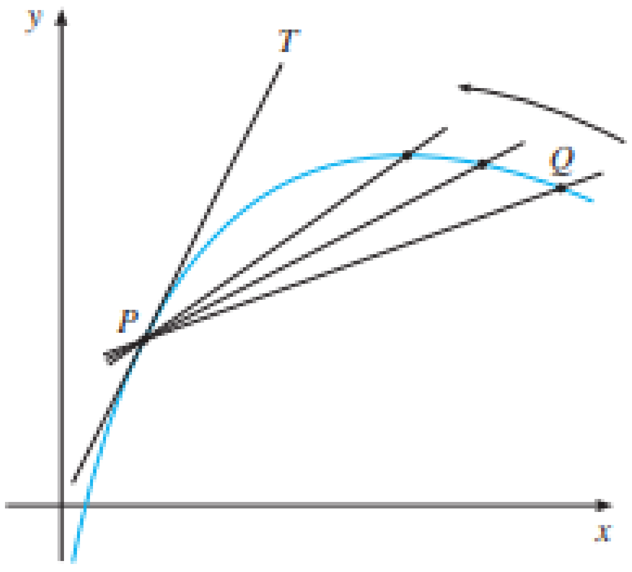


Figure 2

What is a derivative?

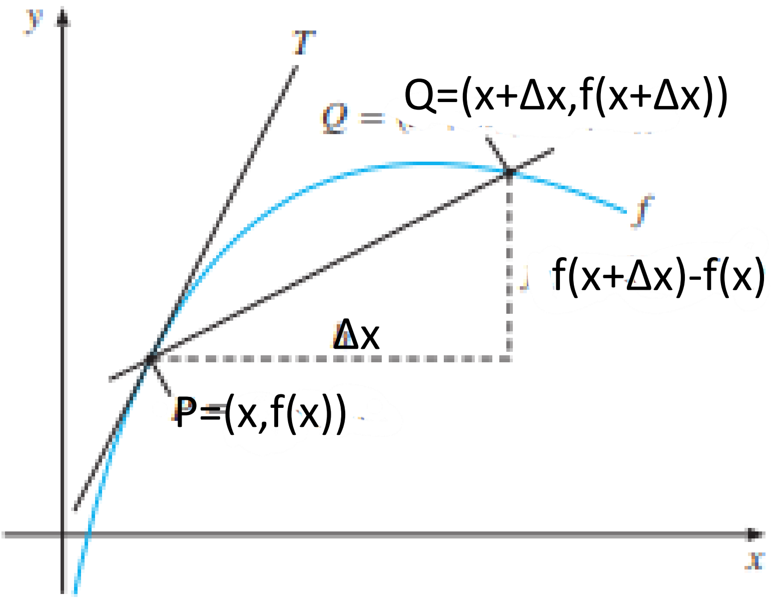


Figure 3

Let's look for the slope at P.

Notice that the Secant line through P and Q has slope:

$$\frac{f(x+\Delta x)-f(x)}{(x+\Delta x)-x}=\frac{f(x+\Delta x)-f(x)}{\Delta x}$$

We can approximate the Tangent line through P by moving Q towards P, decreasing Δx. In the limit as Δx, we get the tangent line through P with slope:

$$\lim_{\Delta x\to 0}\frac{f(x+\Delta x)-f(x)}{\Delta x}$$

What is a derivative?

- Given a function $y = f(x)$, the derivative is defined as

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

or from

$$f'(x) = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h}$$

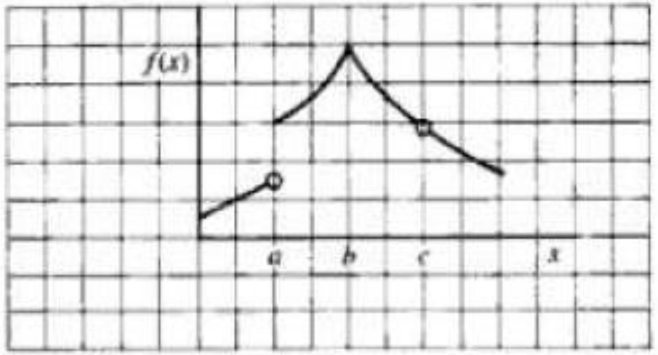
Where $f'(x)$ is read "the first derivative of f with respect of x" or "f prime of x".

The derivative of a function is itself a function which measures both the slope of the tangent line at a given point and the rate of change of the original function $f(x)$ at a given point

Differentiability and Continuity

A function is **differentiable** at a point if the derivative exists (may be taken) at the point. To be differentiable at a point, a function must be:

- Be **continuous** at that point
- Have a **unique tangent** at that point.



In the figure above $f(x)$ is not differentiable at a and c because gaps exists in the function at those points (so the function is not continuous). $f(x)$ is not differentiable in point b because at a cusp (sharp point) an infinite number of tangent lines can be drawn.

Rules of Differentiation

Differentiation is the process of finding the derivative of a function.

This process involves applying few basic rules or formulas to a given function.

Note that in the rules of differentiation of the function $y = f(x)$, other functions such as $g(x)$ and $h(x)$ are commonly used, where g and h are both unspecified functions of x

Rules of Differentiation

The rules of differentiation:

- *Constant function rule*:

  If $f(x) = c$, where c is a constant, then $f'(x)$=0.

- *Linear function rule*:

  If $f(x) = \mathrm{m}x + b$, with m and b constants, then $f'(x) = m$.

- *Power function rule*:

  If $f(x)$=ax^n, where a is a constant and n is any real number, then $f'(x) = \mathrm{a} \cdot n \cdot x^{(n-1)}$

  if $f(x) = \frac{1}{x^n}$, then $f'(x) = -\frac{n}{x^{n+1}}$

Rules of Differentiation

- *Sum / Difference rule:*

  If $f(x) = g(x) \pm h(x)$, then $f'(x) = g'(x) \pm h'(x)$.

- *Product rule:*

  If $f(x) = g(x)h(x)$, then $f'(x) = g'(x)h(x) + g(x)h'(x)$.

- *Quotient rule:*

  If $f(x) = \frac{g(x)}{h(x)}$, then $f'(x) = \frac{g'(x)h(x) - g(x)h'(x)}{h(x)^2}$.

- *Generalised Power rule:*

  If $f(x) = g(x)^n$, then $f'(x) = n \cdot g(x)^{(n-1)}$

Rules of Differentiation

- *Chain rule:*

  If $y = f(g(x))$, then $y' = f'(g(x)) \cdot g'(x)$.

- *Exponential rule:*

  If $f(x) = e^x$, then $f'(x) = e^x$.

  If $f(x) = a^x$, then $f'(x) = a^x \cdot \log(a)$.

- *Logarithm rule:*

  If $f(x) = \ln(x)$, then $f'(x) = \frac{1}{n}$

Rules of Differentiation

Given the following functions $f(x)$, find their respective derivatives $f'(x)$:

a) $f(x) = 2x^3 + 4x^2$

b) $f(x) = 4x^3(2x - 1)$

c) $f(x) = (2x + 4)^2$

d) $f(x) = \frac{x+1}{x+2}$

e) $f(x) = e^x(x + 1)$

Rules of Differentiation

Given the following functions $f(x)$, find their respective derivatives $f'(x)$:

a) $f(x) = 2x^3 + 4x^2$

$$\Rightarrow \quad f'^{(x)} = 2 \cdot 3x^{3-1} + 4 \cdot 2x^{2-1} = 6x^2 + 8x$$

Rules applied:

$$f(x) = g(x) \pm h(x) \Rightarrow \; f'(x) = g'(x) \pm h'(x)$$

$$f(x)=ax^n \Rightarrow \; f'(x) = \mathrm{a} \cdot n \cdot x^{(n-1)}$$

Rules of Differentiation

$b)\quad f(x) = 4x^3(2x-1)$

$\Rightarrow\quad f'^{(x)} = 4\cdot 3x^{3-1}(2x-1) + 4x^3\cdot 2\ = 12x^2(2x-1) + 8x^3 = 24x^3 - 12x^2 + 8x^3 = 32x^3 - 12x^2$

Rules applied:

$f(x) = g(x)h(x) \Rightarrow f'(x) = g'(x)h(x) + g(x)h'(x)$

$f(x)=ax^n \Rightarrow\ f'(x) = \mathrm{a}\cdot n\cdot x^{(n-1)}$

$f(x) = \mathrm{m}x + b \Rightarrow\ f'\,(x) = m$

Higher - Order Derivatives

$f'(x)$ is also called the first-order derivative of $f(x)$.

If we take derivative of $f'(x)$, we obtain the second-order derivative of $f(x)$:

$$f''(x) = \frac{d}{dx} f'(x) = \frac{d^2y}{dx^2}.$$

Example: $f(x) = x^2 \quad \Rightarrow \quad f''(x) = \frac{d}{dx}(2x) = 2.$

The second-order derivative measures the slope and the rate of change of the first derivatives.

Higher-order derivatives are found by applying the rules of differentiation to lower-order derivatives.

Concavity and Convexity

A positive second derivative at $x = a$ denotes the function $f(x)$ is convex at x; so, the graph of the function lies above its tangent line. A negative second derivative at $x = a$ denotes the function $f(x)$ is concave at x; so, the graph of the function lies below its tangent line.

The sign of the first derivative is irrelevant for concavity.

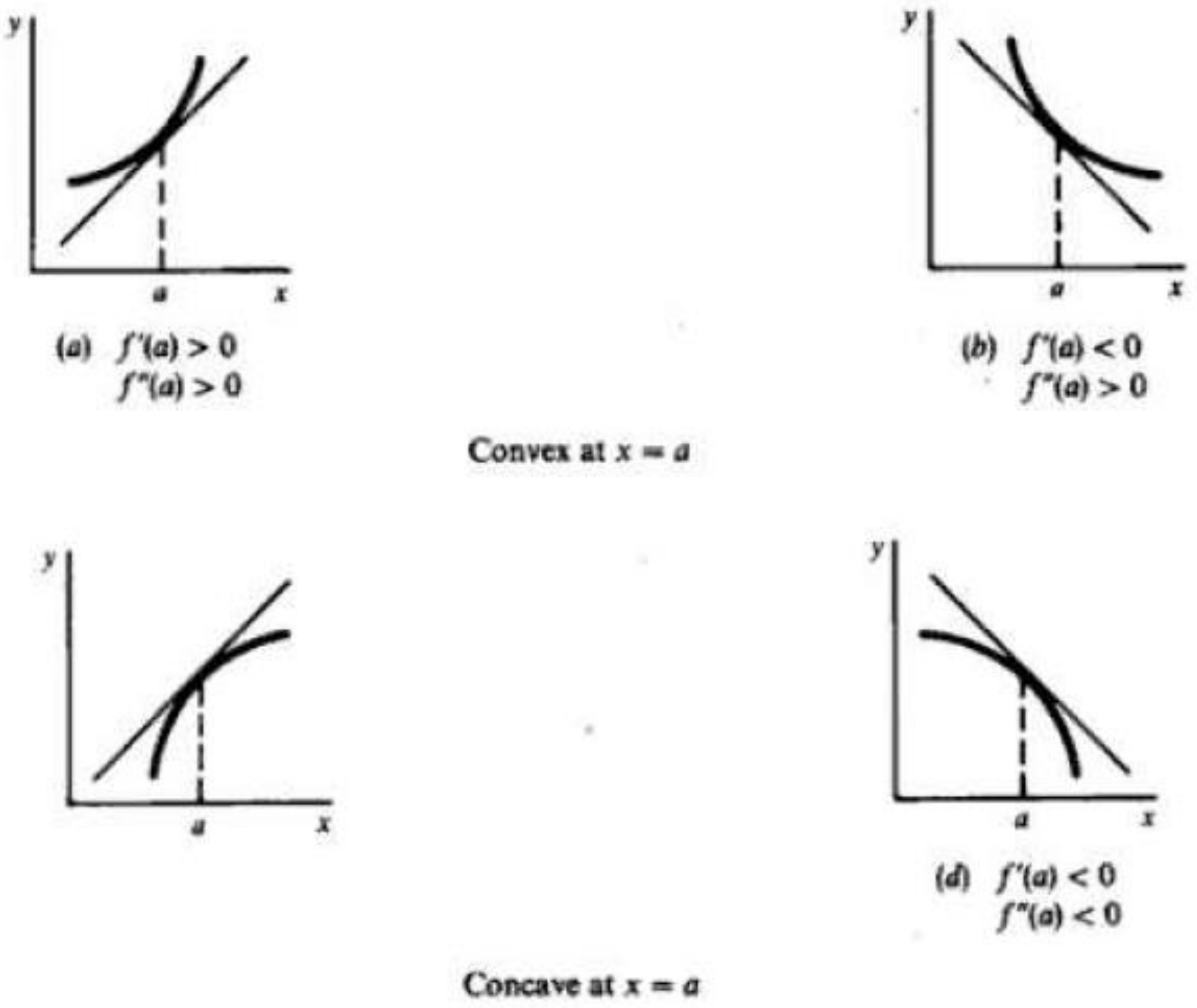
$$f''(a) > 0: \quad f(x)\ \boldsymbol{is\ convex}\ \ at\ x = a$$

$$f''(a) < 0: \quad f(x)\ \boldsymbol{is\ concave}\ at\ x = a$$

If $f''(a) > 0$ for all x in the domain, $f(x)$ is strictly convex.

If $f''(a) < 0$ for all x in the domain, $f(x)$ is strictly concave.

Concavity and Convexity



Relative Extrema

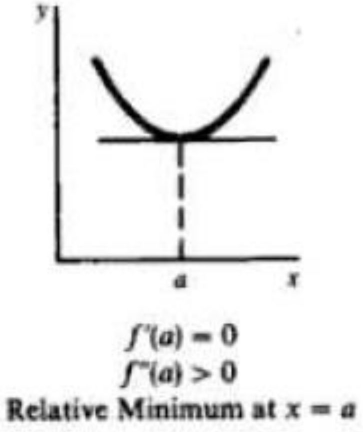
A *relative extremum* is a point at which a function is at relative maximum or minimum. To be at a relative maximum or minimum at a point a, the function must be a relative plateau, i.e. neither increasing or decreasing at a. if the function is neither increasing or decreasing at a, the first derivatives of the function at a must be equal to zero or undefined. A point in the domain of the function where the derivative equals zero is called a **critical point or stationary point**.

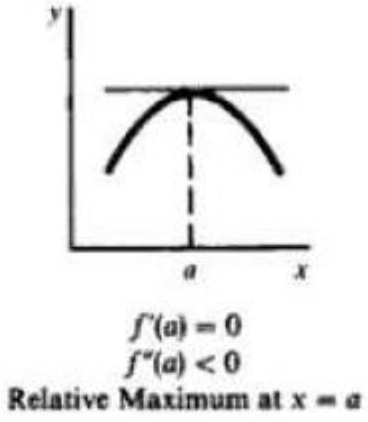
For functions that are differentiable:

$$f'(a) = 0 \; and \; f''(a) > 0 \;\; then \; \boldsymbol{relative\ minimum} \; at \; x = a$$

$$f'(a) = 0 \; and \; f''(a) < 0 \;\; then \; \boldsymbol{relative\ maximum} \; at \; x = a$$

Relative Extrema





Assuming $f'(a) = 0$:

- If $f''(a)$ is positive, it indicates that the function is convex and the graph of the function lies completely above its tangent line at $x = a$, the function is at relative minimum at $x = a$.
- If $f''(a)$ is negative, it indicates that the function is concave and the graph of the function lies completely below its tangent line at $x = a$, the function is at relative maximum at $x = a$.

Optimization of Functions

Optimization is the process of finding the relative maximum or minimum of a function.

1. Take the first derivative, set it equal to zero, and solve for the critical point(s). This step represents a necessary condition known as the **first-order condition (FOC)**. It identifies all the points at which the function is neither increasing or decreasing, but at a plateau. All such point are candidates for a possible relative maximum or minimum.

Optimization of Functions

2. Take the second derivative evaluate it at the critical point(s) and check the sign(s). If at a critical point a:
   - If $f''(a) > 0$, the function is convex at a, and hence at a relative minimum.
   - If $f''(a) < 0$, the function is concave at a, and hence at a relative maximum.
   - If $f''(a) = 0$, the test is inconclusive.

Assuming the necessary first-order condition is met, this step represents a sufficient condition and it is known as **second-order derivative (SOC).**

Note that of the function is strictly concave (convex), there will be only one maximum (minimum) called a *global maximum (minimum).*

Optimization of Functions

Example:

Optimize $f(x) = 7x^2 + 122x + 54$

a) Find the critical points by taking the first derivatives, setting it equal to zero, and solving for x:

$$f'(x) = 14x + 122 = 0$$

$$x = -{}^{122}/_{14} = -8 \ \textit{critical value}$$

b) Take the second derivative, evaluate it at the critical value, and check the sign for a relative maximum or minimum:

$$f''(x) = 14$$

The second derivative is positive, so the point $x = -8$ is a relative minimum.

Since the second derivative is always greater than zero, $f(x)$ is strictly convex and $f(x)$ has a global minimum at $x = -8$

Application in Economics

Example:

A firm's short-run production function is given by : Q=(30−L)L

where Q denotes the quantity of output and L is the number of workers.

Find the size of the workforce that maximises output.

Solution:

- We can write: $Q = 30L - L^2$, $so\ FOC$: $f'(L) = \frac{\partial Q}{\partial L} = 30 - 2L$
- Set $30 - 2L = 0$ to find the critical point. We have $L = \frac{30}{2} = L = 15$
- Then we find SOC: $f''(L) = \frac{\partial^2 Q}{\partial L^2} = -2 < 0.$ This indicated a maximum point.
- The size of workforce that maximises output is 15

Partial Derivatives

So far, the study of derivative was limited to functions of a single independent variable such as $y = f(x)$. Many economic activities, however, involve functions of more that one independent variable, e.g. $z = f(x, y)$. This functions is defined as a function of two independent variables.

To measure the effect of a change in a single independent variable (x or y) on the dependent variable (z) in a multivariate function, the **partial derivative** is needed.

Partial Derivatives with respect to one of the independent variable follows the same rule as the ordinary differentiation while the other independent variables are treated as constant.

Maths & Stats Pre-Sessional

Matrix Algebra

Lecturer: Claudio Vallar
School of Economics and Finance

Matrices - Definition

A **matrix** is an array of numbers

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}.$$

- Let's call this matrix A.
- A is an $m \times n$ matrix, or it has size $m \times n$ and sometimes we denote it as $A_{m\times n}$.
- This means A has m rows and n columns.
- An element a_{ij} of A is the number at the ith row and jth column in A.

Definitions

Definitions:

- A matrix that has the same number of rows and columns is called a **square matrix**.

- Any square matrix that has only nonzero entries on the main diagonal and zeros everywhere else is known as a **diagonal matrix**.

- A special case of a diagonal matrix is the **identity matrix**, where all diagonal entries are 1 and off-diagonal entries are zero. Example:

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- A **Null Matrix** is a matrix in which the elements are all zero.

Definitions

Definitions:

- A **Symmetric Matrix** is a matrix in which the elements opposed to the principal diagonal are the same:

$$\begin{pmatrix} -1 & 2 & -3 \\ 2 & 0 & 5 \\ -3 & 5 & 1 \end{pmatrix}$$

- A **Transpose matrix** is obtained by changing the columns in rows. A Symmetric Matrix coincides with its Transpose.

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \; then \; A^T = \begin{pmatrix} \mathbf{1} & 4 & 7 \\ 2 & \mathbf{5} & 8 \\ 3 & 6 & \mathbf{9} \end{pmatrix}$$

Basic Matrix Operations

Matrix addition:

If A and B have the same size, then the (i, j)th element in $A + B$ is the sum of (i, j)th element in A and (i, j)th element in B.

Example.

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 3 & -1 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} 1+3 & 0-1 \\ 0+2 & 1+4 \end{pmatrix} = \begin{pmatrix} 4 & -1 \\ 2 & 5 \end{pmatrix}.$$

Basic Matrix Operations

Scalar multiplication:

If c is a number (scalar), then the (i, j)th element of cA is c times the (i, j)th element of A.

Example.

$$2\begin{pmatrix}1 & -1\\ 2 & 0\end{pmatrix} = \begin{pmatrix}2 & -2\\ 4 & 0\end{pmatrix}.$$

Basic Matrix Operations

Matrix multiplication:

if the number of columns of A is equal to the number of rows of B, then the (i,j)th element of $C = AB$ is the dot product of the ith row of A and the jth column of B.

Example. We have $A_{2\times 3}$ and $B_{3\times 2}$.

$$C = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{pmatrix},$$

Then $c_{11} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \end{pmatrix} \begin{pmatrix} b_{11} \\ b_{21} \\ b_{31} \end{pmatrix}$

$$= a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31}.$$

Basic Matrix Operations

$$C = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{pmatrix},$$

Then $c_{12} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \end{pmatrix} \begin{pmatrix} b_{12} \\ b_{22} \\ b_{32} \end{pmatrix}$

$$= a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32}.$$

Similarly for c_{21} and c_{22}.

In general, $A_{m \times n} B_{n \times p} = C_{m \times p}$

Basic Matrix Operations: Examples

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

$$A+B = \begin{bmatrix} 9 & 12 & 13 \\ 8 & 8 & 6 \\ 11 & 14 & 12 \end{bmatrix}$$

$$A^{(2*3)} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & 4 \end{bmatrix} \text{ and } B^{(3*4)}\begin{bmatrix} 0 & 1 & -1 & 2 \\ 1 & 3 & 2 & 5 \\ -1 & -1 & 0 & 1 \end{bmatrix}$$

$$\Longrightarrow C^{(2*4)}\begin{bmatrix} 0+2-3 & 1+6-3 & -1+4+0 & 2+10+3 \\ 0-1-4 & 0-3-4 & 0-2-0 & -5-4 \end{bmatrix}$$

Determinant of a Matrix

- It is possible to associate to each matrix a number called **Determinant** of the Matrix.

- Recall that the Determinant is a single number - scalar - and it is defined only for square matrices.

- If the Determinant of a matrix is zero, the matrix is termed **Singular**, which is a matrix in which there exists linear dependence between at least two rows or columns.

- If the Determinant is different from zero, the matrix is **Nonsingular** and all rows and columns are linearly independent.

Rank of a Matrix

The Rank of a matrix is the maximum number of linearly independent rows or columns present in the matrix. It is, in other words, the maximum number of non-null vectors extracted from the matrix.

- if $rank(A) = \min(m, n)$ $\Rightarrow$ the Matrix is singular and it has maximum rank;
- if $rank(A) < \min(m, n)$ $\Rightarrow$ the Matrix is non-singular and contains linearly dependent vectors;

$$D^{(3*3)} = \begin{bmatrix} -2 & -1 & 1 \\ 0 & 1 & -1 \\ 4 & 3 & -7 \end{bmatrix} \Longrightarrow \det|D| = 0$$

Inverse Matrix

The **inverse matrix** A^{-1} of a square matrix A of order n is the matrix that satisfies the condition that

$$AA^{-1} = A^{-1}A = I_n$$

Where I_n is the identity matrix of order n.

- First, we need to check whether it is a square matrix because *square matrices only can have the inverse*.
- Second, we need to make sure that the determinant is non-zero, because *only non-singular matrices can have the inverse*.

Maths & Stats Pre-Sessional

Matrix Algebra : System of Linear Equations

Lecturer: Claudio Vallar
School of Economics and Finance

System of Linear Equations

A system of m linear equations in n unknowns $x_1, x_2, \ldots, x_n$ is a set of m equations of the form

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

The a_{ij}s are the **coefficients** of the system.

System of Linear Equations

We rewrite the system of equations in matrix form:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

$$\Downarrow \qquad \Downarrow \qquad \Downarrow$$

$$\boldsymbol{A} \qquad \boldsymbol{x} = \boldsymbol{b}$$

That is $\boldsymbol{Ax} = \boldsymbol{b}$, a compact way of writing the system

Multiple Regression Model and the Matrix form

Consider the model:

$$y_t = \beta_1 + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + u_t, \quad t = 1, \ldots, T$$

For each t:

$$\begin{aligned} y_1 &= \beta_1 + \beta_2 x_{21} + \ldots + \beta_k x_{k1} + u_1 \\ y_2 &= \beta_2 + \beta_2 x_{22} + \ldots + \beta_k x_{k2} + u_2 \\ \vdots &= \vdots \\ y_T &= \beta_1 + \beta_2 x_{2T} + \ldots + \beta_k x_{kT} + u_T \end{aligned}$$

We can express this more conveniently and compactly in matrix form

Multiple Regression Model and the Matrix form

The matrix form:

$$\underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix}}_{\substack{\mathbf{y} \\ [T\times 1]}} = \underbrace{\begin{pmatrix} 1 & x_{21} & \dots & x_{k1} \\ 1 & x_{22} & \dots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{2T} & \dots & x_{kT} \end{pmatrix}}_{\substack{\mathbf{X} \\ [T\times k]}} \underbrace{\begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}}_{\substack{\boldsymbol{\beta} \\ [k\times 1]}} + \underbrace{\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{pmatrix}}_{\substack{\mathbf{u} \\ [T\times 1]}}$$

$$\underset{[T\times 1]}{\mathbf{y}} = \underset{[T\times 1]}{\mathbf{X}\boldsymbol{\beta}} + \underset{[T\times 1]}{\mathbf{u}}$$