ECON 441: INTRODUCTION TO MATHEMATICAL ECONOMICS

Review Notes

Div Bhagia

Based on Chiang, Alpha C, and Wainwright K. (2005), Fundamental Methods of Mathematical Economics: 4th edition

Contents

1 Linear Algebra

1.1 Introduction to Matrices

A **matrix** is a rectangular array of numbers, parameters, or vectors.

Example. $A = \begin{bmatrix} 2 & 3 & 1 \\ -1 & 4 & 6 \end{bmatrix}$

Dimensions of matrix:

- Number of rows (m)

- Number of columns (n)

A matrix with m rows and n columns is referred to as an $m \times n$ matrix.

It's common to denote the dimensions of a matrix as a subscript. So we can write the 2×3 matrix A in the example above as:

$$A = \begin{bmatrix} 2 & 3 & 1 \\ -1 & 4 & 6 \end{bmatrix}_{2 \times 3}$$

More generally a $m \times n$ matrix can be written as:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix}$$

Or more compactly as:

$$A = [a_{ij}] \quad i = 1, 2, \dots, m; j = 1, 2, \dots, n$$

A **square matrix** is matrix that has equal number of rows and columns. For example,

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}_{3\times3}$$

Two matrices are **equal** if all their elements are identical. So $A = B$ if and only if $a_{ij} = b_{ij}$ for all i, j. *Example.*

$$\begin{bmatrix} 1 & 8 \\ 4 & -1 \end{bmatrix} \neq \begin{bmatrix} 1 & 8 \\ 4 & 2 \end{bmatrix}$$

1.2   Matrix Operations

1.2.1   Addition and Subtraction

To add two matrices, we need to add each element of one matrix to the corresponding element in the other matrix. Similarly, to take the difference between two matrices, we subtract each element of one matrix from the corresponding element in the other matrix. For this to work, we need both matrices to have the same dimension.

For two matrices of the same dimension, A and B, we have

$$A + B = [a_{ij} + b_{ij}] \qquad A - B = [a_{ij} - b_{ij}]$$

In summary:

- Addition and subtraction operations are performed *element-by-element*

- Only matrices with the same dimension can be added or subtracted.

1.2.2   Scalar Multiplication

To multiply a matrix A by a scalar λ, we need to multiply each element of A by λ. So

$$\lambda A = [\lambda a_{ij}]$$

Example. Given, $A = \begin{bmatrix} 2 & 3 \\ 4 & -6 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 8 \\ -2 & 3 \end{bmatrix}$, we can find:

a) $A + B = \begin{bmatrix} 3 & 11 \\ 2 & -3 \end{bmatrix}$

b) $A - B = \begin{bmatrix} 1 & -5 \\ 6 & -9 \end{bmatrix}$

c) $2A = \begin{bmatrix} 4 & 6 \\ 8 & -12 \end{bmatrix}$

d) $A - 2B = \begin{bmatrix} 2 & 3 \\ 4 & -6 \end{bmatrix} - \begin{bmatrix} 2 & 16 \\ -4 & 6 \end{bmatrix} = \begin{bmatrix} 0 & -13 \\ 8 & -12 \end{bmatrix}$

1.2.3 Matrix Multiplication

We can only multiply two matrices $A_{m \times n}$ and $B_{p \times q}$ to get AB if $n = p$. In other words, we can only multiply two matrices, if the number of columns in the first matrix is equal to the number of rows in the second matrix.

Example. $A = \begin{bmatrix} 2 & 3 & 1 \\ 4 & -6 & -2 \end{bmatrix}_{2 \times 3}$ $B = \begin{bmatrix} 1 & 8 \\ -2 & 3 \end{bmatrix}_{2 \times 2}$

Here, we cannot find AB, but can find BA.

Note: The order of multiplication matters, so even when both products are possible, AB is not necessarily equal to BA. In fact, the two will only be equal under special circumstances.

If we have two matrices $A_{m \times n}$ and $B_{n \times p}$ and we can multiply these to get $C = AB$. Then the dimension of C will be $m \times p$. To find C we take the ith row of A and the jth row of B to get the ij element of C as follows:

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \ldots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj}$$

Example.

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}_{2 \times 3} \qquad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{32} \\ b_{21} & b_{32} \end{bmatrix}_{3 \times 2}$$

$$C = AB = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \end{bmatrix}_{2\times2}$$

1.3   Vectors

Matrices with only one column are called **column vectors**.

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Matrices with only one row are called **row vectors**.

$$x' = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}$$

Inner product of two vectors each with n elements is given by:

$$u \cdot v = u_1 v_1 + u_2 v_2 + \dots + u_n v_n = \sum_{i=1}^{n} u_i v_i$$

Example.

$$u = \begin{bmatrix} 1 \\ 5 \\ 2 \end{bmatrix} \qquad v = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$$

$$u \cdot v = 1.2 + 5.1 + 2.3 = 2 + 5 + 6 = 13$$

The rows and columns of a matrix are row and column vectors, respectively.

1.3.1   Linear Dependence

A set of vectors is said to be **linearly dependent** if and only if any one of them can be expressed as a linear combination of the remaining vectors.

Example.

$$v_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad v_2 = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

Here v_1 and v_2 are linearly dependent because $v_2 = 2v_1$.

Formally, a set of m-vectors $v_1, v_2, ..., v_n$ is linearly dependent if and only if there exists a set of scaler $k_1, k_2, ..., k_n$ (not all zero) such that:

$$\sum_{i=1}^{n} k_i v_i = 0 \quad (m \times 1)$$

1.4   Identity and Null Matrix

An **identity matrix** is a square matrix that has all the elements in its *principal diagonal* equal to 1 and all other elements equal to 0.

A 2×2 identity matrix:

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

A 3×3 identity matrix:

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

Identity matrix behaves like 1 for matrices:

$$AI = IA = A$$

In particular, pre- or post-multiplying a matrix by the identity matrix returns the original matrix.

A matrix is an **idempotent** matrix if it remains unchanged when multiplied by itself any number of times. That is A is idempotent if and only if $A = A^k$.

A **null** matrix is a matrix with all elements 0.

Example.

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

- $A + 0 = A$

- $A0 = 0$

1.5　Transpose and Inverse of a Matrix

Transpose of a matrix A is denoted by A' and is obtained by interchanging rows and columns of A.

Example.

$$A = \begin{bmatrix} 2 & 3 & 1 \\ 4 & -6 & 2 \end{bmatrix} \qquad A' = \begin{bmatrix} 2 & 4 \\ 3 & -6 \\ 1 & 2 \end{bmatrix}$$

- A matrix A is said to be *symmetric* if $A' = A$

- A matrix A is said to be *skew-symmetric* if $A' = -A$

- A matrix A is said to be *orthogonal* if $A'A = I$

Properties of transposes:
$$(A')' = A$$
$$(A + B)' = A' + B'$$
$$(AB)' = B'A'$$

For a *square* matrix A, it's **inverse** A^{-1} is defined as:

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

Properties of inverses:
$$\left(A^{-1}\right)^{-1} = A$$
$$(AB)^{-1} = B^{-1}A^{-1}$$
$$(A')^{-1} = \left(A^{-1}\right)'$$

1.6   Conditions for Nonsingularity

If a matrix's inverse exists, it's called a **nonsingular** matrix.

Squareness is a *necessary* condition for an inverse to exist, but not a *sufficient* condition. The sufficient condition for nonsingularity of a matrix is if all the rows or (equivalently) columns of this matrix are linearly independent.

Rank of a matrix is defined as the maximum number of linearly independent rows or (equivalently) columns.

Example.

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

Rows and columns of A are linearly independent, so A is nonsingular and has full rank 2. The second row of B can be written as two times the first row of B, so the rank of B is 1 and B is a singular matrix.

We can check for linear independence and find the rank of a matrix by converting the matrix to its **echelon** form.

For a matrix in echelon form:

- First row: all elements *can be* non-zero

- Second row: first element 0

- Third row: first two elements 0

$$\vdots$$

- Last row: first $m - 1$ elements zero

Echelon form of a 2×2 matrix.

$$A = \begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{bmatrix}$$

Echelon form of a 3×3 matrix.

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{bmatrix}$$

Valid operations to convert to echelon form:

- Interchange any two rows

- Multiplication (or division) of a row by a scalar $k \neq 0$

- Addition of a (or k times of a) row to another

How to convert a 3×3 matrix to its echelon form?

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Step 1: Try to make $a_{31} = 0$

$$A_1 = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{bmatrix}$$

Step 2: Try to make $a_{21} = 0$

$$A_2 = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{bmatrix}$$

Step 3: Try to make $a_{32} = 0$

$$A_3 = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{bmatrix}$$

In summary, a square matrix is nonsingular i.e. its inverse exists if and only if all its rows (or equivalently columns) are linearly independent. We can also rephrase this as — a square matrix is nonsingular if it is full rank.

1.7 Determinants

Determinant $|A|$ is a unique scalar associated with a *square* matrix A. Before we can write down the determinant for an $n \times n$ matrix, we need to define minors and cofactors.

The **minor** of the element a_{ij}, denoted by $|M_{ij}|$ is obtained by deleting the ith row and jth column of the matrix and taking the determinant of the resulting matrix.

Whereas, **cofactor** $|C_{ij}|$ is defined as:

$$|C_{ij}| = (-1)^{i+j}|M_{ij}|$$

Example.

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Minor for the element a_{12}:

$$|M_{12}| = \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix}$$

Cofactor for the element a_{12}:

$$|C_{12}| = (-1)^{(1+2)}|M_{12}| = -|M_{12}|$$

Determinant for an $n \times n$ matrix is given by:

$$|A| = \sum_{i=1}^{n} a_{ij}|C_{ij}| = \sum_{j=1}^{n} a_{ij}|C_{ij}|$$

The first expression corresponds to expanding with respect to the ith row, while the

second expression is the expression for the determinant when expanding with respect to the jth column. In practice, we can expand with respect to any row or column.

Using the above formula we can see that the determinant for a 2×2 matrix:

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

is given by:

$$|A| = a_{11}a_{22} - a_{12}a_{21}$$

Similarly, the determinant of a 3×3 matrix is given by:

$$|A| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

$$= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

Determinant criteria for nonsingularity: The determinant of singular matrices is equal to 0.

Example.

$$\begin{vmatrix} 4 & 0 & -1 \\ 2 & 1 & -7 \\ 3 & 3 & 9 \end{vmatrix} = 4 \begin{vmatrix} 1 & -7 \\ 3 & 9 \end{vmatrix} - 0 \begin{vmatrix} 2 & -7 \\ 3 & 9 \end{vmatrix} - 1 \begin{vmatrix} 2 & 1 \\ 3 & 3 \end{vmatrix}$$

$$= 4(9 + 21) - 0(18 + 21) - 1(6 - 3)$$

$$= 120 - 0 - 3 = 117$$

Properties of determinants:

1. $|A| = |A'|$

2. Interchanging rows or columns will alter the sign but not the value

3. Multiplication of any one row (or one column) by a scalar k will change the value of the determinant k-fold

4. The addition (subtraction) of a multiple of any row (or column) to (from) another row (or column) will leave the determinant unaltered

5. If one row (or column) is a multiple of another row (or column), the value of the determinant will be zero.

1.8   Matrix Inversion

To find the inverse of a nonsingular matrix A take the transpose of its cofactor matrix $C = [|C_{ij}|]$ to find the adjoint of A and divide it by the determinant of A.

$$A^{-1} = \frac{1}{|A|} adj A$$

Adjoint of a nonsingular $n \times n$ matrix

$$adj A = C' = \begin{bmatrix} |C_{11}| & |C_{21}| & \dots & |C_{n1}| \\ |C_{12}| & |C_{22}| & \dots & |C_{n2}| \\ \vdots & \vdots & \dots & \vdots \\ |C_{1n}| & |C_{2n}| & \dots & |C_{nn}| \end{bmatrix}$$

Example.

$$A = \begin{bmatrix} 4 & -2 & 1 \\ 7 & 3 & 0 \\ 2 & 0 & 1 \end{bmatrix}$$

We can write the cofactors of A as follows:

$$|C_{11}| = \begin{vmatrix} 3 & 0 \\ 0 & 1 \end{vmatrix} = 3 \qquad |C_{12}| = -\begin{vmatrix} 7 & 0 \\ 2 & 1 \end{vmatrix} = -7 \qquad |C_{13}| = \begin{vmatrix} 7 & 3 \\ 2 & 0 \end{vmatrix} = -6$$

$$|C_{21}| = -\begin{vmatrix} -2 & 1 \\ 0 & 1 \end{vmatrix} = 2 \qquad |C_{22}| = \begin{vmatrix} 4 & 1 \\ 2 & 1 \end{vmatrix} = 2 \qquad |C_{23}| = -\begin{vmatrix} 4 & -2 \\ 2 & 0 \end{vmatrix} = -4$$

$$|C_{31}| = \begin{vmatrix} -2 & 1 \\ 3 & 0 \end{vmatrix} = -3 \qquad |C_{32}| = -\begin{vmatrix} 4 & 1 \\ 7 & 0 \end{vmatrix} = 7 \qquad |C_{33}| = \begin{vmatrix} 4 & -2 \\ 7 & 3 \end{vmatrix} = 26$$

$$\text{adj } A = \begin{bmatrix} |C_{11}| & |C_{21}| & |C_{31}| \\ |C_{12}| & |C_{22}| & |C_{32}| \\ |C_{13}| & |C_{23}| & |C_{33}| \end{bmatrix} = \begin{bmatrix} 3 & 2 & -3 \\ -7 & 2 & 7 \\ -6 & -4 & 26 \end{bmatrix}$$

Now to find the determinant of $|A|$, say we expand using the second row:

$$A = a_{21}|C_{21}| + a_{22}|C_{22}| + a_{23}|C_{23}| = 7 \times 2 + 3 \times 2 + 0 \times -4 = 20$$

Note that expanding by any row or column will give us the same answer. For example, expanding by the third column:

$$A = a_{13}|C_{13}| + a_{23}|C_{23}| + a_{33}|C_{33}| = 1 \times -6 + 0 \times -4 + 1 \times 26 = 20$$

1.9   Solving a system of linear equations

All the linear algebra we have learned so far can be used to solve a system of linear equations.

For example, we have the following two equations:

$$q + 2p = 100$$
$$q - 3p = 20$$

We can write this as:
$$Ax = b$$

where

$$A = \begin{bmatrix} 1 & 2 \\ 1 & -3 \end{bmatrix} \quad x = \begin{bmatrix} q \\ p \end{bmatrix} \quad b = \begin{bmatrix} 100 \\ 20 \end{bmatrix}$$

If we pre-multiply the equation $Ax = b$ by A^{-1}, we get

$$A^{-1}Ax = A^{-1}b \implies x^* = A^{-1}b$$

1.9.1 Cramer's Rule

Cramer's rule basically gives a slightly more efficient way of solving a system of equations than finding the inverse. This method is still based on matrix inversion. In particular, the kth element of x can be solved by:

$$x_k^* = \frac{|A_k|}{|A|}$$

where A_k is a matrix formed by exchanging kth column of A by b.

1.9.2 Homogeneous-equation system

A homogeneous equation system is given by

$$Ax = 0$$

For this system of equations, if A is nonsingular, $x^* = A^{-1}0 = 0$. If A is singular there can be infinite number of solutions (this is true for any system of equations).

2 Comparative Statics

2.1 Difference Quotient and the Derivative

We are often interested in how some variable y changes in response to another variable x where:

$$y = f(x)$$

To see how y changes with x, we can think about how the value of $f(x)$ changes when say x goes from x_0 to x_1. Denote change in the value of a variable by Δ. Then $\Delta x = x_1 - x_0$. When x changes from an initial value x_0 to a new value $x_0 + \Delta x$, the value of the function $y = f(x)$ changes from $f(x_0)$ to $f(x_0 + \Delta x)$. The change in y per unit of change in x can be represented by the difference quotient:

$$\frac{\Delta y}{\Delta x} = \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}.$$

Example. For the function: $y = f(x) = x^2 + 20$

$$
\begin{aligned}
\frac{\Delta y}{\Delta x} &= \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} \\
&= \frac{(x_0 + \Delta x)^2 + 20 - (x_0^2 + 20)}{\Delta x} \\
&= \frac{x_0^2 + (\Delta x)^2 + 2x_0 \Delta x - x_0^2}{\Delta x} \\
&= 2x_0 + \Delta x
\end{aligned}
$$

We are usually interested in minuscule changes from x_0. The *derivative* of a function is defined as:

$$\frac{\partial y}{\partial x} = f'(x) = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$$

17

So for the example above:

$$\frac{dy}{dx} = f'(x_0) = \lim_{\Delta x \to 0} 2x_0 + \Delta x = 2x_0$$

Note that both the following definitions of the derivative are equivalent:

$$(1). \quad \frac{dy}{dx} = f'(x_0) = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

$$(2). \quad \frac{dy}{dx} = f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

Example. Let's find the derivative of $y = f(x) = x^2 + 20$ again, but now using the second version of the definition.

$$\begin{aligned}
\frac{dy}{dx} &= \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} \\
&= \lim_{x \to x_0} \frac{x^2 + 20 - x_0^2 - 20}{x - x_0} \\
&= \lim_{x \to x_0} \frac{x^2 - x_0^2}{x - x_0} \\
&= \lim_{x \to x_0} x + x_0 = 2x_0
\end{aligned}$$

The derivative is a function and in this example, it depends on the initial value x_0. More generally, we can write the above derivative as $2x$ instead of $2x_0$.

Graphically, the derivative is the slope of the tangent line for a curve.

2.2   Concept of a Limit

We say that L is the *limit* of $f(x)$ at a, i.e.

$$\lim_{x \to a} f(x) = L$$

if $f(x)$ approaches L as x approaches a from any direction. Note that x is not set exactly equal to a but x gets very very close to a.

It is not necessary that the limit of a function exists at every point. For the limit to exist at a point we need the function to approach the same value from both directions. In particular, for the limit of $f(x)$ at a to exist, we need the left-side and right-side limits of f to be equal at a.
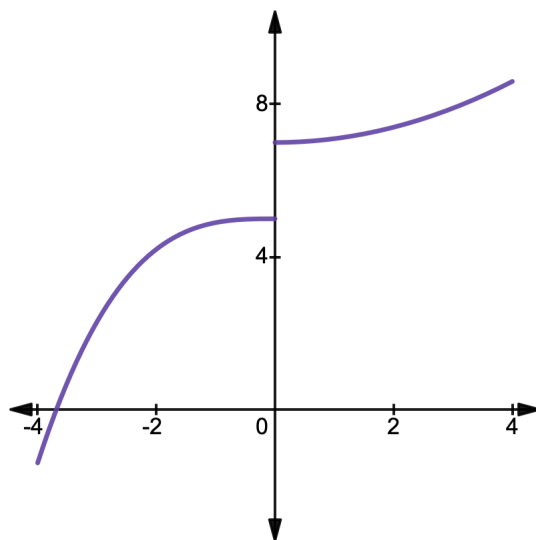
Left-side limit: If x approaches a from the left side:

$$\lim_{x \to a^-} f(x)$$

Right-side limit: If x approaches a from the right side:

$$\lim_{x \to a^+} f(x)$$

Only when both left-side and right-side limits have a common finite value, we say that the limit exists. In the example below the limit of the function does not exist at 0 as the left-side limit at 0 is not equal to the right-side limit.

Example. Let's calculate the limit of the following function at $x = 1$.

$$f(x) = \frac{4(x-1)}{|x-1|}$$

To calculate the left-side limit plug in a value of x just a tiny bit below 1, e.g. 0.99. Similarly, to calculate the right-side limit plug in a value of x just a tiny bit above 1, such

as 1.01. Then we can find that:

$$\lim_{x \to 1^-} f(x) = -4$$

$$\lim_{x \to 1^+} f(x) = 4$$

Since the left-side and right-side limit are not equal to each other, we will say that the limit of this function does not exist at 1. The graph of this function is presented below:
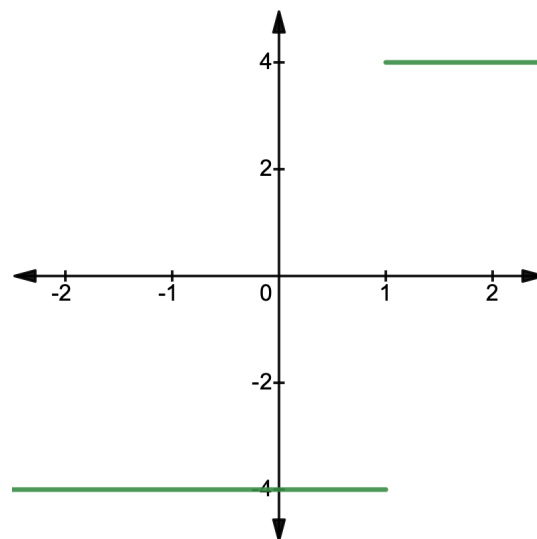
2.3   Continuity of a Function

A function $y = f(x)$ is said to be continuous at a if the limit of $f(x)$ at a exists and is equal to the value of the function at a i.e.,

$$\lim_{x \to a} f(x) = f(a)$$

Example. For the function,

$$y = \begin{cases} 2x + 1 & \text{if } x < 1 \\ 2 & \text{if } x = 1 \\ 2x + 1 & \text{if } x > 1 \end{cases}$$

The limit for this function exists at 1 as both the left-side limit and the right-side limit

at 1 for this function is 3. But this function is not continuous because it takes value 2 when $x = 1$.

Continuity vs differentiability:

$f'(x_0)$ exists if the following limit exists:

$$f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

A function $y = f(x)$ is continuous at x_0 if

$$\lim_{x \to x_0} f(x) = f(x_0)$$

From the two definitions, we can see that continuity is a necessary condition for differentiability, but it is not a sufficient condition. In other words, if f is not continuous, it implies that f is not differentiable as well. But if f is continuous, f could either be differentiable or not. For example, the function $y = |x|$ (presented below) is continuous but not differentiable
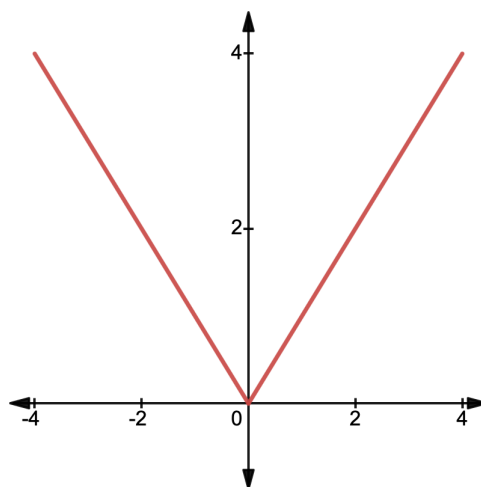
2.4   Partial Derivatives

For a function of several variables:

$$y = f(x_1, x_2, \cdots, x_n)$$

If x_1 changes by Δx_1 but all other variables remain constant:

$$\frac{\Delta y}{\Delta x_1} = \frac{f(x_1 + \Delta x_1, x_2, \cdots, x_n) - f(x_1, x_2, \cdots, x_n)}{\Delta x_1}$$

Partial derivative of y with respect to x_i is defined as:

$$\frac{\partial y}{\partial x_i} = f_i = \lim_{\Delta x_i \to 0} \frac{\Delta y}{\Delta x_i}$$

Note that here, we use ∂ to differentiate partial derivatives from total derivatives. In particular,

$$\frac{\partial f}{\partial x_i} = \frac{df}{dx_i}\bigg|_{\text{other variables are constant}}$$

Example. Given the production function $Q = AK^\alpha L^{1-\alpha}$.
Marginal product of capital (MPK):

$$\frac{\partial Q}{\partial K} = Q_K = \alpha A K^{\alpha-1} L^{1-\alpha} = \frac{\alpha Q}{K}$$

Marginal product of labor (MPL):

$$\frac{\partial Q}{\partial L} = Q_L = \alpha A K^\alpha L^{-\alpha} = \frac{(1-\alpha)Q}{L}$$

The *gradient* of a function is defined as the vector of all partial derivatives of the function:

$$\nabla f(x_1, x_2, \cdots, x_n) = [f_1, f_2, \cdots, f_n]'$$

With many functions, we can define the Jacobian matrix. Consider n differentiable functions in n variables that are not necessarily linear.

$$y_1 = f^1(x_1, x_2, \cdots, x_n)$$
$$y_2 = f^2(x_1, x_2, \cdots, x_n)$$
$$\cdots$$
$$y_n = f^n(x_1, x_2, \cdots, x_n)$$

The Jacobian matrix is defined

$$J = \frac{\partial(y_1, y_2, \cdots, y_n)}{\partial(x_1, x_2, \cdots, x_n)} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial y_n}{\partial x_1} & \frac{\partial y_n}{\partial x_2} & \cdots & \frac{\partial y_n}{\partial x_n} \end{bmatrix}$$

The n functions $f^1, f^2, \ldots f^n$ are functionally (linear or nonlinear) dependent if and only if the jacobian $|J| = 0$ for all values of $x_1, x_2, \cdots, x_n$.

Differentials

Note that,

$$\Delta y \equiv \left[\frac{\Delta y}{\Delta x}\right] \Delta x$$

Then for infinitesimal changes we can write,

$$dy \equiv \left[\frac{dy}{dx}\right] dx \quad \text{or} \quad dy = f'(x)dx$$

We call dy and dx differentials of y and x, respectively. We can think of $f'(x)$ as a ratio of two quantities dy and dx.

Elasticity of function is defined as:

$$\varepsilon = \frac{\text{Percentage change in y}}{\text{Percentage change in x}} = \frac{dy/y}{dx/x}$$

We can calculate the elasticity by taking the derivative of y with respect to x as follows:

$$\varepsilon = \frac{dy}{dx} \cdot \frac{x}{y}$$

- $\varepsilon > 1$, elastic

- $\varepsilon = 1$, unit elasticity

- $\varepsilon < 1$, inelastic

Example. For the consumption function $C = a + bY$, the elasticity is $\varepsilon = bY/(a + bY)$.

For a function of n variables, we can define the *total differential*. For the function, $y = f(x_1, x_2, \cdots, x_n)$, the total differential is given by:

$$df = \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2 + \cdots + \frac{\partial f}{\partial x_n} dx_n = \sum_{i=1}^{n} f_i dx_i$$

Example. Consider the savings function $S = S(Y, i)$ where S is savings, Y is national income, and i is the interest rate. Total differential is given by:

$$dS = \frac{\partial S}{\partial Y} dY + \frac{\partial S}{\partial i} di$$

The *total derivative* of the f with respect to x_1 can be obtained by dividing the total differential df by dx_1 as follows:

$$\frac{df}{dx_1} = \frac{\partial f}{\partial x_1} + \frac{\partial f}{\partial x_2} \frac{dx_2}{dx_1} + \cdots + \frac{\partial f}{\partial x_n} \frac{dx_n}{dx_1} = f_1 + f_2 \cdot \frac{dx_2}{dx_1} + \cdots + f_n \cdot \frac{dx_n}{dx_1}$$

In the case where x_j does not depend on x_1, the term corresponding to x_j will not enter the total differential as $dx_j/dx_1 = f_j = 0$. For example, if none of the other variables depend on x_1, the total derivative will be equal to the partial derivative.

$$\frac{df}{dx_1} = \frac{\partial f}{\partial x_1} = f_1$$

Example 1. Suppose we have $y = f(x_1, x_2)$ and $x_2 = g(x_1)$. In this case the total derivative of y with respect to x is given by:

$$\frac{df}{dx_1} = f_1 + f_2 \cdot g'(x_1)$$

Example 2. Given $y = f(x_1, x_2, w)$, $x_1 = g(w)$, and $x_2 = h(w)$. The total derivative of f with respect w is given by:

$$\frac{df}{dw} = f_1 \cdot g'(w) + f_2 \cdot h'(w) + f_3$$

In this example, if instead we had $y = f(x_1, x_2)$, we could still use the definition of total derivative and just plug-in $f_3 = 0$ as in this case the function does not directly depend on w.

Implicit Function Theorem

We can write an *explicit* function $y = f(x_1, x_2, \cdots, x_n)$, as an implicit function:

$$F(y, x_1, x_2, \cdots, x_n) = y - f(x_1, x_2, \cdots, x_n) = 0$$

Example. For the explicit function $y = f(x_1, x_2) = x_1 + 3x_2 - 2$, we can find the implicit function $F(y, x_1, x_2) = y - x_1 - 3x_2 + 2$.

However, it is not always true that for every explicit function an implicit function ex-

ists. The conditions when an implicit function exists are given by the Implicit Function Theorem stated below for the two variable case.

Implicit Function Theorem

Given, $F(y, x) = 0$, if the following conditions are met:

- F_y and F_x are continuous, and

- at some point x_0, y_0 where $F(x_0, y_0) = 0$, F_y is non-zero

Then in a neighborhood around x_0, an implicit function exists. Moreover, this function is continuous and has continuous partial derivatives.

Also, when an implicit function exists, we can find its derivatives from the derivatives of the explicit function. Taking the total differential of $F(y, x_1, x_2, \cdots, x_n) = 0$,

$$F_y dy + F_1 dx_1 + \cdots + F_n dx_n = 0$$

Now suppose only y and x_1 are allowed to vary, then $F_2 = F_3 = \ldots = F_n = 0$ and we have:

$$F_y dy + F_1 dx_1 = 0 \quad \rightarrow \quad \frac{\partial y}{\partial x_1} = -\frac{F_1}{F_y}.$$

In the simple case where the given equation is $F(y, x) = 0$, the rule gives

$$\frac{dy}{dx} = -\frac{F_x}{F_y}.$$

3 Exponential and Logarithmic Functions

Reference: Chiang, Alpha C, and Wainwright K. (2005), Fundamental Methods of Mathematical Economics: 4th edition, Chapter 10

Exponential and logarithmic functions have important applications in economics. So we will discuss some basic properties and derivatives for these functions.

3.1 Exponential Functions

The exponential or power function can be represented as:

$$y = f(t) = b^t \quad (b > 1)$$

where b denotes a fixed base of the exponent. Note that y is a function of t, i.e. the power of b. A more generalized version can be written as:

$$y = ab^{ct}$$

When the base is a special mathematical constant called Euler's number $e = 2.71828...$, i.e.

$$y = ae^{rt}$$

it is referred to as the natural exponential function (though common to just refer to it as the exponential function as most other exponential functions don't show up as often). It is also common to write it as:

$$y = a \exp(rt)$$

Where did this number e come from? It can be shown that:

$$e \equiv \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n$$

Jacob Bernoulli discovered this constant in 1683 while studying a question about compound interest.

Suppose you invest A dollars at a 5% interest rate that compounds yearly. After one year you will have:

$$A\,(1 + 0.05)$$

After t years:

$$A\,(1 + 0.05)^t$$

Now rewrite the expression above as:

$$A\left(\underbrace{\left(1 + \frac{1}{1/0.05}\right)^{1/0.05}}_{\approx e}\right)^{0.05t} \approx Ae^{0.05t}$$

More generally, if A is the initial principal, r is the interest rate, and t is the investment horizon, then the principal at t:

$$A_t \approx A_0 e^{rt}$$

3.2   Logarithmic Functions

Since the exponential function is a monotonic function, its inverse exists. The inverse of the exponential function is called the log or logarithmic function.

For the exponential function:

$$y = b^t \rightarrow log_b(y) = t$$

where the expression on the right-hand side is referred to as the log of y to base b. For the natural exponential function:

$$y = e^t \rightarrow \log_e y = ln(y)$$
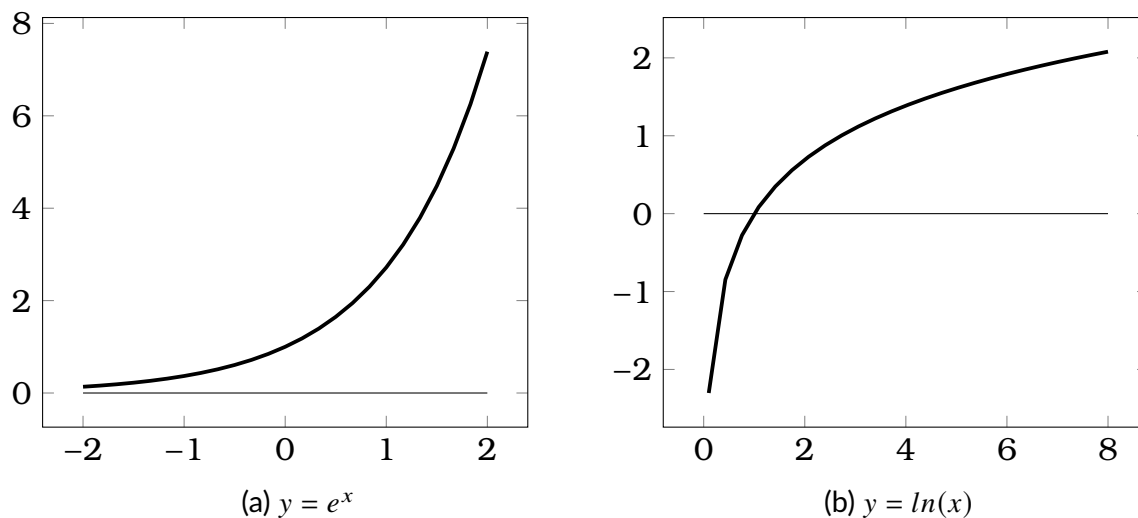
(a) $y = e^x$

(b) $y = ln(x)$

Figure 3.1: Plots for exponential and logarithmic functions

Since the natural log is employed quite often, we can write it without the base as ln. If you say log of x, without mentioning the base, it is sort of assumed that you are talking about the natural log.

For example, since $2^3 = 8$. So we can write $\log_3 8 = 2$.

Since $y = b^t \iff t = \log_b y$, we can write

$$b^{\log_b y} = b^t = y$$

Figure 1 presents plots for $y = e^x = exp(x)$ and $y = ln(x)$. From the plots, we can see that both log and exponential functions are strictly monotonically increasing. In addition,

- $y = 1, log(y) = 0$

- $0 < y < 1, \log y < 0$

- $\log y \to \infty$ as $y \to \infty$

- $\log y \to -\infty$ as $y \to 0$

3.3   Rules for Logarithmic Functions

- $\ln(uv) = \ln u + \ln v$

- $\ln(u/v) = \ln u - \ln v$

- $\ln u^a = a \ln u$

To see why the above rules hold, let $\ln u = s$ and $\ln v = t$. Then, $u = e^s$ and $v = e^t$. In which case,

$$uv = e^s e^t = e^{s+t} \rightarrow \ln uv = s + t = \ln u + \ln v$$

Similarly,

$$u/v = e^s/e^t = e^{s-t} \rightarrow \ln(u/v) = s - t = \ln u - \ln v$$

Finally for the last rule,

$$u^a = (e^s)^a = e^{as} \rightarrow \ln u^a = as = a \ln u$$

3.4   Derivatives of Exponential and Logarithmic Functions

Derivative of the exponential function:

$$y = b^t \quad \rightarrow \quad \frac{dy}{dt} = b^t \ln b$$

Derivative of the generalized exponential function:

$$y = e^{f(t)} \quad \rightarrow \quad \frac{dy}{dt} = f'(t)e^{f(t)}$$

Derivative of the log function:

$$\frac{d}{dt} \ln f(t) = \frac{f'(t)}{f(t)}$$

Let's find the derivatives for the following functions:

1. $f(t) = e^t$, $f'(t) = e^t$

2. $f(t) = \ln t$, $f'(t) = 1/t$

3. $f(t) = ae^{rt}$, $f'(t) = are^{rt}$

4. $f(t) = e^{-t}$, $f'(t) = -e^{-t}$

5. $f(t) = \ln at$, $f'(t) = a/at = 1/t$

6. $f(t) = \ln t^c$, $f'(t) = ct^{c-1}/t^c = c/t$

4 Optimization

Optimization is the process of finding the best solution to a problem, given a set of constraints and objectives. In economics, optimization plays a central role in understanding how individuals, firms, and governments make decisions to allocate scarce resources efficiently. In the context of functions, optimization involves finding a function's maximum or minimum value.

Global vs. Local Extrema

- Local maximum: A point where the function attains its *highest* value within a neighborhood.

- Local minimum: A point where the function attains its *lowest* value within a neighborhood.

- Global (or absolute) maximum: A point where the function attains its *highest* value over the entire domain.

- Global (or absolute) minimum: A point where the function attains its *lowest* value over the entire domain.
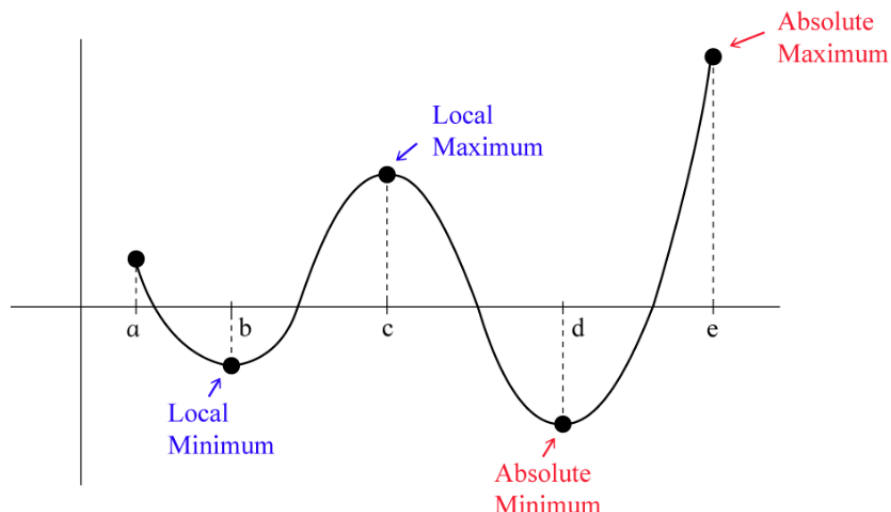
Note that all global extrema are also considered local extrema. Furthermore, there can be multiple local and global extrema for a function. For example, consider the constant function $f(x) = 2$. In this case, all points in the function's domain are global maxima and global minima. There are also functions that have no extremum points, such as $f(x) = x$, which continuously increases as x increases and continuously decreases as x decreases, without reaching any peak or valley.

4.1 Unconstrained Optimization with One Variable

We will start with the simple case of finding a maximum or a minimum of a single-variable function. First, we will cover the first and second-order conditions for identifying the local maxima or minima of a function. We will then briefly discuss the conditions necessary for identifying global extrema.

4.1.1 First and Second-Order Conditions

As illustrated in the figure above, our objective is to locate the peaks or valleys of a function. Since the derivative captures the rate of change of a function, the derivative of the function at peaks and valleys should be equal to zero. This is because the function is relatively flat in the small neighborhood of the peak or valley.

We define a critical point as a point where the function's derivative is equal to zero. All local maxima and minima of a function, if any, must occur at critical points where the derivative of the function is equal to zero.

The *necessary* condition for x^* to be a maximum or minimum point of a continuously differentiable function $f(x)$ is:

$$f'(x^*) = 0$$

Here, x^* is called a critical/stationary point.

Example. Let's find all the critical points for the function $f(x) = x^2 - 24x + 36$. We start by taking the derivative of $f(x)$:

$$f'(x) = 2x - 24$$

Next, we set $f'(x)$ equal to zero and solve for x. $2x - 24 = 0 \rightarrow x = 12$. Therefore, the critical point of $f(x)$ is $x^* = 12$.

Note that it is necessary that at any maxima or minima, the derivative of the function is 0. To see this, consider a point c where $f'(c) > 0$, so the function is increasing at point c. Then the value of the function at a point just to the left of c will be smaller than $f(c)$, so c cannot be a local minimum. Similarly, the value of the function at a point just to the right of c will be greater than $f(c)$, so c cannot be a local maximum.

While being a critical point is a necessary condition, it is not a sufficient condition for maxima or minima. This means that at any maximum or minimum point, the derivative of the function must be equal to zero. However, simply having a zero derivative at a point is not enough to conclude that the point is a maximum or a minimum. In some cases, the function may have an *inflection* point or a flat spot, as shown in the following figure.
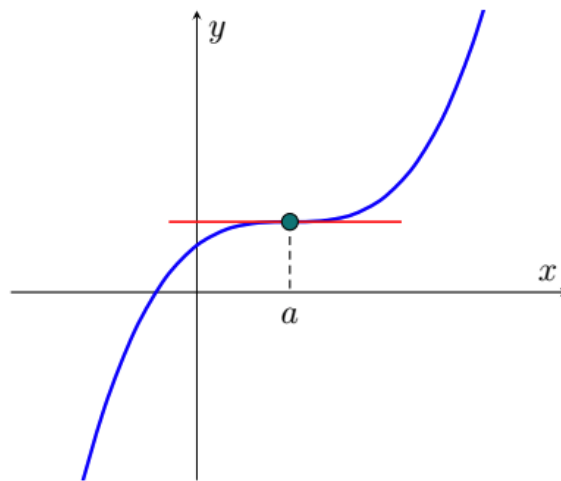
Furthermore, even if we have identified a critical point that is not an inflection point, we still cannot tell whether it is a maximum or a minimum point. To determine whether the point is indeed a maximum, a minimum, or neither, we need to analyze the behavior of the function in the neighborhood of the critical point. For instance, when we are at a peak, it should be the case that the function was increasing on the left side of the critical point x^* and decreasing on the right side. In other words, the derivative of the function was positive on the left side of x^* and negative on the right side. This approach is called the first derivative test.

First Derivative Test

Suppose $f'(x_0) = 0$, then x^* is a

- maximum if $f'(x)$ goes from $+$ to $-$ in the immediate neighborhood of x_0

- minimum if $f'(x)$ goes from $-$ to $+$ in the immediate neighborhood of x_0

- not an extreme point if $f'(x)$ has the same sign in the immediate neighborhood of x_0
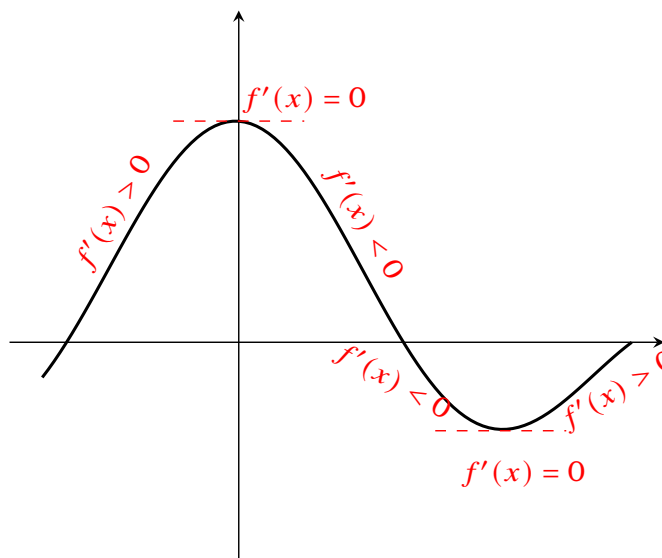
Example. For the function $f(x) = x^2 - 24x + 36$, with the derivative $f'(x) = 2x - 24$, the critical point is $x^* = 12$. Consider a point just below 12, say $x = 11.99$. The derivative at this point is $f'(11.99) = 2(11.99) - 24 = -0.02 < 0$. Now consider a point just above 12, say $x = 12.01$. The derivative at this point is $f'(12.01) = 2(12.01) - 24 = 0.02 > 0$. So the function is decreasing as x approaches x^* and then starts increasing beyond x^*. So we can conclude that the critical point $x^* = 12$ corresponds to a valley or a local minimum.

Instead of checking whether the derivative goes from positive to negative, we can also just look at the sign of the second derivative at the critical point, which is often easier. This is called the second derivative test.

Second Derivative Test

Suppose $f'(x_0) = 0$, then x^* is a

   1. a maximum if $f''(x^*) < 0$

   2. a minimum if $f''(x^*) > 0$


The second derivative of a function is the derivative of the first derivative, and, hence it measures the rate of change of the first derivative. Consider a critical point x^*. When $f''(x^*) < 0$, it means that $f'(x)$ is decreasing in the neighborhood of x^*. Since $f'(x^*) = 0$ and $f'(x)$ is decreasing in the neighborhood of x^*, it must be that the sign of $f'(x)$

changes from positive to negative at x^*, indicating a local maximum. Similarly, when $f''(x^*) > 0$, it means that $f'(x)$ is increasing at x^*, and it must be that $f'(x)$ changes from negative to positive at x^*, indicating a local minimum.

Example. For the function $f(x) = x^2 - 24x + 36$, with the derivative $f'(x) = 2x - 24$, the critical point is $x^* = 12$. The second derivative is given by, $f''(x) = 2$, which is positive at all points, including at $x = 12$. So from the second derivative test, we can conclude that 12 is a local minimum.

Finally, note that if $f'(x^*) = 0$ and $f''(x^*) > 0$ or $f''(x^*) < 0$, then we have a sufficient condition for determining whether x^* is a maximum or minimum point. However, this condition is not necessary, as there are cases where $f''(x^*) = 0$, and yet x^* is a maximum or minimum point. An example of such a case is the function $f(x) = x^4$, with $f'(x) = 4x^3$ and $f''(x) = 12x^2$. The critical point for this function is at 0 and $f''(0) = 0$, yet 0 is the minimum point. The graph of this function is presented below.
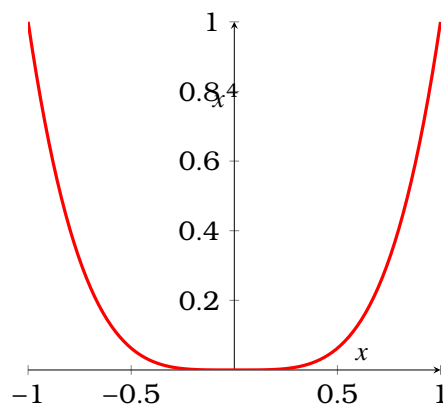
Alternatively, we can say that $f'(x^*) = 0$ and $f''(x^*) \geq 0$ or $f''(x^*) \leq 0$ is the necessary condition for determining whether x^* is a maximum or minimum point.

The following table summarizes the first and second-order conditions for local maximum and minimum.

Condition	Maximum	Minimum
First-order necessary	$f'(x) = 0$	$f'(x) = 0$
Second-order necessary [†]	$f''(x) \leq 0$	$f''(x) \geq 0$
Second-order sufficient [†]	$f''(x) < 0$	$f''(x) > 0$

[†] Applicable only after the first-order necessary condition has been satisfied.

4.1.2   Conditions for Global Extrema

The conditions for global extrema are similar to those for local extrema, but now we need to look at the sign of $f''(x)$ over the entire domain of the function and not just locally at x^*. In particular, if $f''(x) < 0$ for all x in the domain of f, then any critical point of f is a unique global maximum. This is because $f'(x)$ is decreasing over the entire domain of x, and can only be 0 at a unique critical point where it goes from positive to negative, implying that it is the unique global maximum.

At this point, it is useful to define concave and convex functions.

- A function $f(x)$ is said to be *concave* if $f''(x) \leq 0$ for all x.

- A function $f(x)$ is said to be *convex* if $f''(x) \geq 0$ for all x.

- A function $f(x)$ is said to be *strictly concave* if $f''(x) < 0$ for all x.

- A function $f(x)$ is said to be *strictly convex* if $f''(x) > 0$ for all x.

So now, we can write the conditions for global extrema in terms of concave and convex functions.

- If a function is (strictly) concave, any critical point will give us a (unique) global maximum.

- If a function is (strictly) convex, any critical point will give us a (unique) global minimum.

Since concave and convex functions allow flat regions, it is possible for there to be

multiple extrema. This was the case in our example with $f(x) = 2$, which is both a concave and a convex function as $f''(x) = 0$.

4.2   Unconstrained Optimization with Multiple Variables

In economics, we often encounter situations where we need to maximize a function that involves more than one variable. For example, we may want to determine the optimal quantity of two or more inputs, such as labor and capital, that a firm should use to produce a given level of output.

To solve for the optimal values of multiple input variables, we need to take the partial derivative of the output function with respect to each input variable and set them equal to zero. This is the first-order condition (FOC) for multivariate functions.

For a function of two variables $f(x, y)$, the FOC is given by:

$$f_x = \frac{\partial f}{\partial x} = 0, \qquad f_y = \frac{\partial f}{\partial y} = 0$$

Here, f_x and f_y are the partial derivatives of f with respect to x and y, respectively. In particular, f_x captures, how f changes as x changes while holding y constant. So when calculating the partial derivative with respect to one variable, we treat the other variable as a constant.

Example. Let's find all the critical points for the function

$$f(x, y) = x + y^2 - x^2 y$$

Note that, $f_x = 1 - 2xy$ and $f_y = 2y - x^2$. Setting $f_x = 0$, we get $xy = 0.5$ and setting $f_y = 0$, we get $y = 0.5x^2$. Plugging $y = 0.5x^2$ in $xy = 0.5$, we get at $x = 1$ and $y = 0.5$. So the critical point for this function is $(1, 0.5)$.

Example. Consider the profit function $\pi(K, L) = Q(K, L) - rK - wL$, where $Q(K, L)$ is the production function that specifies the maximum quantity of output that can be produced using K units of capital and L units of labor. Here, r and w are the prices of capital and labor, respectively.

The partial derivative of π with respect to L is given by:

$$\pi_L = \frac{\partial \pi}{\partial L} = Q_L - w = 0$$

Here, Q_L is the partial derivative of the production function with respect to labor, which measures the marginal product of labor (MPL) or the increase in output resulting from an additional unit of labor. Similarly, Q_K is the partial derivative of the production function with respect to capital, which measures the marginal product of capital (MPK).

The partial derivative of π with respect to K is:

$$\pi_K = \frac{\partial \pi}{\partial K} = Q_K - r = 0$$

At the critical point where both FOCs are satisfied, the firm will choose inputs so that the price of each input is equal to its marginal product. (Note that Q_K and Q_L are functions of K and L.)

For multivariate functions, the second-order conditions involve calculating all the (own and cross) second partial derivatives of the function. We will skip details on these conditions here. However, it is important to note that the concept of concavity and convexity can be extended to multivariate functions. In particular, we can define a concave (convex) function as a function whose graph lies below (above) any line segment connecting any two points on the graph. (See the slides for Lecture 12 for more details.) Moreover, as before, any critical point of a multivariate (strictly) concave/convex function represents a (unique) global maximum/minimum.

4.3   Constrained Optimization

Lots of problems in economics involve optimizing a function subject to one or more constraints. For example, a firm may want to maximize its profits subject to production constraints, or a consumer may want to maximize their utility subject to a budget constraint. To solve such problems, we use the method of Lagrange multipliers.

Suppose we want to maximize (or minimize) a function $f(x, y)$ subject to a constraint $g(x, y) = c$, where c is a constant. The method of Lagrange multipliers involves constructing a new function, known as the Lagrangian, that combines the objective func-

tion and the constraint function as follows:

$$L(x, y, \lambda) = f(x, y) + \lambda[c - g(x, y)]$$

Here, λ is the Lagrange multiplier. To find the optimal values of x and y, we need to take partial derivatives of the Lagrangian with respect to each variable and the Lagrange multiplier and set them equal to zero. This gives us the first-order conditions:

$$L_x(x, y, \lambda) = \frac{\partial L}{\partial x} = f_x(x, y) - \lambda \cdot g_x(x, y) = 0$$

$$L_y(x, y, \lambda) = \frac{\partial L}{\partial y} = f_x(x, y) - \lambda \cdot g_x(x, y) = 0$$

$$L_\lambda(x, y, \lambda) = \frac{\partial L}{\partial \lambda} = c - g(x, y) = 0$$

Here, f_k and g_k represent the partial derivative with respect to x_k of the objective function and the constraint, respectively.

Intuitively, the Lagrange multiplier method works because it allows us to take into account the constraint when optimizing the function. By introducing the Lagrange multiplier, we are effectively adding a penalty term to the objective function that accounts for the constraint. The Lagrange multiplier λ has an economic interpretation as the shadow price of the constraint. It measures the change in the objective function resulting from a small change in the constraint function. Thus, the Lagrange multiplier provides us with information about the value of relaxing or tightening the constraint.

Example. We want to maximize utility given by $U(x, y) = xy$ subject to the budget constraint $4x + y = 12$.

We start by setting up the Lagrangian:

$$L(x, y, \lambda) = xy + \lambda(12 - 4x - y)$$

First-order conditions:

$$\frac{\partial L}{\partial x} = y - 4\lambda = 0 \tag{4.1}$$

$$\frac{\partial L}{\partial y} = x - \lambda = 0 \tag{4.2}$$

$$\frac{\partial L}{\partial \lambda} = 12 - 4x - y = 0 \tag{4.3}$$

From equations (1) and (2), we have $y = 4\lambda$ and $x = \lambda$, which implies that $y = 4x$. Plugging this in equation (3):

$$12 - 4x - 4x = 0 \rightarrow x^* = 1.5, y^* = 6, \lambda^* = 1.5$$

Interpretation of λ^*: If income increases by a dollar, utility increases by 1.5.

The method of Lagrange multipliers can be extended to problems with more than two variables and constraints. For example, consider the problem of maximizing a function $f(x, y, z)$ subject to two constraints $g(x, y, z) = c$ and $h(x, y, z) = m$, where c and m are constants. To solve this problem, we will need to introduce two Lagrange multipliers and then construct the Lagrangian function as follows:

$$L(x, y, z, \lambda, \mu) = f(x, y, z) + \lambda[c - g(x, y, z)] + \mu[m - h(x, y, z)]$$

To find the optimal values of x, y, and z that maximize $f(x, y, z)$ subject to the two constraints, we can take the partial derivatives of the Lagrangian with respect to x, y, z, λ, and μ, and set them equal to zero.

Global Extrema with Constraints

The Lagrange multiplier method transforms the constrained optimization problem of maximizing $f(x, y)$ subject to $g(x, y)$ into an unconstrained optimization problem of maximizing $L(x, y, \lambda)$. We know that in the case of unconstrained optimization, any critical point of a concave function represents a global maximum. In the case of constrained optimization, if both $f(x, y)$ and $g(x, y)$ are concave, then $L(x, y, \lambda)$ will also be concave, and any critical point of $L(x, y, \lambda)$ will represent the global maximum of $f(x, y)$ subject to $g(x, y)$. However, requiring functions to be concave is a strong condition. We can instead use the weaker condition that $f(x, y)$ is quasiconcave, and the constraint set, $C = \{(x, y) : g(x, y) = c\}$, is a convex set to ensure that we are identifying the global maximum. (See the slides for Lecture 12 for definitions.)

The stationary point $(x_1^*, x_2^*, ..., x_n^*)$ of the Lagrangian corresponding to the problem of optimizing $f(x_1, x_2, .., x_n)$ subject to a constraint $g(x_1, x_2, ..., x_n) = c$ is a global maximum if:

1. $f(x_1, x_2, ..., x_n)$ is quasiconcave

2. The constraint set is convex

4.4   Envelope Theorem

In economics, we are often interested in understanding how the value of a function changes in response to certain *parameters*. For example, we may be interested in how a consumer's utility changes in response to changes in prices, or how a firm's profits change in response to changes in wages. When analyzing how a function's value changes in response to changes in parameters, we need to consider both the direct impact of changing the parameters on the function's value, as well as the indirect impact of changing the parameters on the optimal inputs, which in turn affects the function's value. The Envelope Theorem tells us that we do not need to worry about the indirect effect of changing parameters on the optimal inputs — we can simply focus on the direct effect.

Envelope Theorem: Unconstrained Maximization
Consider the following maximization problem:

$$\max_{x} \quad f(x, \alpha)$$

Here, α is a parameter. Since the optimal input may depend on α, let's denote it by $x^*(\alpha)$. Now let's define the *maximum value function*:

$$V(\alpha) = f(x^*(\alpha), \alpha)$$

The envelope theorem states that,

$$V_\alpha = \frac{\partial V}{\partial \alpha} = f_\alpha^*$$

where f_α^* is the partial derivative of f with respect to x at x^*

It is actually straightforward to see why this is the case. Note, that when we differentiate V with respect to α,

$$V_\alpha = f_x^* \cdot \frac{\partial x^*}{\partial \alpha} + f_\alpha^*$$

Here, f_x^* is the partial derivative of f with respect to x at x^*. By first order conditions for optimization, $f_x^* = 0$, and hence $V_\alpha = f_\alpha^*$.

Example. Say we want to choose labor input L to maximize profit

$$\max_L \quad \pi(L, w) = \ln L - wL$$

Here, w is the wage rate. Writing the first-order condition:

$$\pi_L = \frac{1}{L} - w = 0 \rightarrow L^* = \frac{1}{w}$$

Since L^* depends on w, we can write it as $L^*(w)$.

Now, we are interested in how optimal profit changes with respect to changes in wages. To answer this question, we first write down the maximum value function, which here depends on w.

$$V(w) = \pi(L^*(w), w) = \ln L^*(w) - wL^*(w)$$

Now to see how $V(w)$ varies with w we need the derivative of V with respect to w. The envelope theorem says that this derivative is given by

$$V'(w) = \pi_w(L^*(w), w) = -L^*(w) = -\frac{1}{w}$$

We can also verify this by plugging in $L^*(w) = 1/w$ in the expression for $V(w)$ and then differentiating with respect to w.

$$V(w) = \ln L^*(w) - wL^*(w) = \ln(1/w) - 1 = \ln 1 - \ln w - 1 \rightarrow V'(w) = -\frac{1}{w}$$

The envelope theorem for constrained optimization problems is stated below.

Envelope Theorem: Constrained Maximization

Consider the following constrained optimization problem

$$\max_{x,y} \quad f(x, y; \alpha) \quad \text{s.t.} \quad G(x, y; \alpha) = 0$$

Lagrangian function:

$$L(x, y, \lambda; \alpha) = f(x, y; \alpha) + \lambda G(x, y; \alpha)$$

As before, define the value function $V(\alpha) = f(x^*, y^*; \alpha)$. Then the envelope theorem states,

$$\frac{\partial V}{\partial \alpha} = \frac{\partial L(x^*, y^*, \lambda^*; \alpha)}{\partial \alpha}$$

The above theorem directly implies interpretation of the lagrange multiplier. Say max $f(x, y)$ subject to $g(x, y) = c$ Can think of the constraint as:

$$G(x, y, c) = c - g(x, y)$$

So here c is the parameter α. Now what happens due to change in c to the value function is given by

$$\frac{\partial V}{\partial \alpha} = \frac{\partial L(x^*, y^*, \lambda^*; \alpha)}{\partial \alpha} = \lambda^*$$

Example. We want to maximize the utility function $U(x_1, x_2) = x_1 x_2$ subject to the budget constraint $p_1 x_1 + p_2 x_2 = m$, where m is the total income. We are interested in how the total utility changes due to a change in price of good 1.

Setting up the Lagrangian:

$$L(x_1, x_2, \lambda) = x_1 x_2 + \lambda(m - p_1 x_1 - p_2 x_2)$$

First-order conditions:

$$L_1 = x_2 - \lambda p_1 = 0 \tag{4.4}$$
$$L_2 = x_1 - \lambda p_2 = 0 \tag{4.5}$$
$$L_\lambda = m - p_1 x_1 - p_2 x_2 = 0 \tag{4.6}$$

Equations (4) and (5) imply that $x_2/x_1 = p_1/p_2$. Plugging $x_2 = p_1 x_1/p_2$ in equation (6):

$$m - p_1 x_1 - p_2 x_2 = m - p_1 x_1 - p_1 x_1 \rightarrow x_1^* = \frac{m}{2p_1}, x_2^* = \frac{m}{2p_2}, \lambda^* = \frac{m}{2p_1 p_2}$$

Denote $L^* = L(x_1^*, x_2^*, \lambda^*)$. Then according to the envelope theorem how optimal utility changes with respect to p_1 is given by

$$\frac{\partial L^*}{\partial p_1} = -\lambda^* x_1^* = -\frac{m^2}{4p_1^2 p_2}$$

We can verify that this is true as:

$$V(p_1, p_2, m) = x_1^* x_2^* = \frac{m^2}{4p_1 p_2} \rightarrow \frac{\partial V}{\partial p_1} = -\frac{m^2}{4p_1^2 p_2}$$

5 Optimization Examples

Example 1: Utility Maximization

Consider the following utility maximization problem:

$$\max_{\{x_1, x_2\}} \quad x_1^\alpha x_2^\beta \quad s.t. \quad p_1 x_1 + p_2 x_2 = m$$

Here, p_1 and p_2 are the prices of goods 1 and 2, and m is the total income available to spend on the two goods. To find the critical points, we start by setting up the Lagrange function:

$$\mathcal{L}(x_1, x_2, \lambda) = x_1^\alpha x_2^\beta + \lambda(m - p_1 x_1 - p_2 x_2)$$

The first-order conditions are given by:

$$\frac{\partial \mathcal{L}}{\partial x_1} = \alpha x_1^{\alpha-1} x_2^\beta - \lambda p_1 = 0 \tag{5.1}$$

$$\frac{\partial \mathcal{L}}{\partial x_1} = \beta x_1^\alpha x_2^{\beta-1} - \lambda p_2 = 0 \tag{5.2}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = m - p_1 x_1 - p_2 x_2 = 0 \tag{5.3}$$

From equations (1) and (2), we have $\alpha x_1^{\alpha-1} x_2^\beta = \lambda p_1$ and $\beta x_1^\alpha x_2^{\beta-1} = \lambda p_2$, dividing the first expression by the second, we get:

$$\frac{\alpha x_1^{\alpha-1} x_2^\beta}{\beta x_1^\alpha x_2^{\beta-1}} = \frac{\lambda p_1}{\lambda p_2}$$

Simplifying this expression further:

$$\frac{\alpha x_2^\beta x_2^{1-\beta}}{\beta x_1^\alpha x_1^{1-\alpha}} = \frac{p_1}{p_2} \longrightarrow \frac{\alpha x_2}{\beta x_1} = \frac{p_1}{p_2}$$

Then we can write that,

$$x_2 = \frac{\beta}{\alpha} \cdot \frac{p_1}{p_2} \cdot x_1$$

Plugging the above expression for x_2 in equation (3):

$$m - p_1 x_1 - p_2 x_2 = m - p_1 x_1 - \frac{\beta}{\alpha} \cdot p_1 x_1 = m - p_1 x_1 \left(1 + \frac{\beta}{\alpha}\right) = 0$$

So we can find,

$$x_1^* = \frac{\alpha}{\alpha + \beta} \frac{m}{p_1}$$

Plugging back x_1 in the expression for x_2, we can find:

$$x_2^* = \frac{\beta}{\alpha + \beta} \frac{m}{p_2}$$

Example 2: Cost Minimization

Consider the following cost minimization problem where p is the price of capital, and w is the price of labor. Quantity is constrained at $\bar{Q}$.

$$\max_{\{K,L\}} \quad pK + wL \quad s.t. \quad Q(K, L) = \bar{Q}$$

Lagrange function:

$$\mathcal{L}(x_1, x_2, \lambda) = pK + wL + \lambda(\bar{Q} - Q(K, L))$$

First-order conditions:

$$\frac{\partial \mathcal{L}}{\partial K} = p - \lambda Q_K(K, L) = 0$$
$$\frac{\partial \mathcal{L}}{\partial L} = w - \lambda Q_L(K, L) = 0$$
$$\frac{\partial \mathcal{L}}{\partial \lambda} = \bar{Q} - Q(K, L)$$

Note that here $Q_K = \frac{\partial Q}{\partial K}$ and $Q_L = \frac{\partial Q}{\partial L}$.

So optimal K and L satisfy:

$$(1). \quad \frac{Q_K(K^*, L^*)}{Q_L(K^*, L^*)} = \frac{p}{w}, \qquad (2). \quad Q(K^*, L^*) = \bar{Q}$$

Example 3: Inter-Temporal Utility Maximization

You are given the following inter-temporal utility function:

$$U = U(c_1, c_2) = \ln c_1 + \beta \ln c_2 \tag{5.4}$$

where c_1 and c_2 is consumption in period 1 and 2, respectively. $0 < \beta < 1$ is the rate at which you discount the future and it measures your impatient. You earn income $y_1 > 0$ in period 1 and income $y_2 > 0$ in period 2. Any of the income you save s in period 1 earns interest $r > 0$. So,

$$c_1 + s = y_1, \qquad c_2 = y_2 + (1 + r)s$$

Combining these constraints:

$$c_1 + \frac{1}{1 + r} c_2 = y_1 + \frac{1}{1 + r} y_2$$

Let the present-discounted income be denoted by m, such that:

$$m = y_1 + \frac{1}{1 + r} y_2$$

You want to choose c_1 and c_2 to maximize utility $U(c_1, c_2)$ in equation (4) subject to the constraint:

$$c_1 + \frac{1}{1 + r} c_2 = m \tag{5.5}$$

Lagrangian function:

$$L(c_1, c_2, \lambda) = \ln c_1 + \beta \ln c_2 + \lambda \left(m - c_1 - \frac{1}{1 + r} c_2 \right)$$

First order conditions:

$$\frac{\partial L}{\partial c_1} = \frac{1}{c_1^*} - \lambda^* = 0 \tag{5.6}$$

$$\frac{\partial L}{\partial c_2} = \frac{\beta}{c_2^*} - \frac{\lambda^*}{1+r} = 0 \tag{5.7}$$

$$\frac{\partial L}{\partial \lambda} = m - c_1^* - \frac{1}{1+r}c_2^* = 0 \tag{5.8}$$

Note that from (6), $\lambda^* = 1/c_1^*$, plugging this in (7), we get:

$$c_2^* = \beta(1+r)c_1^*$$

Plugging this expression for c_2^* in (8):

$$c_1^* + \frac{1}{1+r}\beta(1+r)c_1^* = m \rightarrow c_1^* = \frac{m}{1+\beta}$$

In which case,

$$c_2^* = \beta(1+r)c_1^* = \frac{\beta m(1+r)}{1+\beta}$$

Finally, since $\lambda^* = 1/c_1^*$,

$$\lambda^* = \frac{1}{c_1^*} = \frac{1+\beta}{m}$$

Now suppose we are interested in knowing how utility changes due to changes in total income. By the envelope theorem:

$$\frac{\partial U^*}{\partial m} = \frac{\partial L^*}{\partial m} = \lambda^* = \frac{1+\beta}{m}$$

Similarly, how utility changes due to a change in interest rate would be given by:

$$\frac{\partial U^*}{\partial r} = \frac{\partial L^*}{\partial r} = \lambda^* \cdot \frac{1}{(1+r)^2} \cdot c_2^* = \frac{\beta}{1+r}$$