Math Camp Notes

Jose L. Casco

Applied Economics

This document gives a brief summary of the topics covered in the math camp course. It is NOT intended to be exhaustive. Many of the proofs are simply sketched and others are omitted entirely. There are probably many typos too.[1]

1 Introduction to First-Order Logic, Quantifiers and Methods for Proofs

In this section we give an introduction to basic first-order logic, elementary methods for constructing proofs and some important considerations concerning notational convention.

Table 1: Common Logic Symbols

Symbol	Definition
$\wedge$	and
$\vee$	or
$\neg$	not
$\forall$	for all
$\exists$	there exist
$\nexists$	there does not exist
$\Longrightarrow$	implies
$\Longleftrightarrow$	implies and is implied by (if and only if or iff)
$\in$	an element of
$\subseteq$	is contained in
$\subset$	is strictly contained in

$\mathbb{N}$: Set of natural numbers. i.e. $\mathbb{N} = \{1, 2, 3, \ldots\}$.
$\mathbb{Z}$: Set of integers. i.e. $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$.
$\mathbb{Q}$: Set of rational numbers. i.e. $\mathbb{Q} = \{q/p : \ p \neq 0, p, q \in \mathbb{Z}\}$.
$\mathbb{R}$: Set of real numbers. i.e. the set includes $\mathbb{N}$, $\mathbb{Q}$ and π,e, etc.
$\mathbb{C}$: Set of complx numbers. i.e. $a + bi$, where $a, b \in \mathbb{R}$ and $i = \sqrt{-1}$.

Clearly from this definitions we can observe that: $\mathbb{N} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$. In the present course and in the microeconomic theory sequence you are going to work

[1]This notes are constructed from the following sources: Konishi lecture notes for the Math Review of Applied Economics, Phelan lectures notes for the Math Review of Economics and Coggins lectures notes for Math Vaccination for Applied Economics. Additionally, E. OK (2007), Simon and Blume (1994), W. Rudin (1976), O. Guler (2010), G. Debreu (1987), A. Takayama (1985), SLP (1989).

mainly with real numbers, therefore we will no examine the complex numbers or complex analysis.

Also, it useful for our analysis to define the extended (augmented) real number system, which is the set $\mathbb{R}$ together with the elements $-\infty$ and $+\infty$.

1.1 Logical operations and truth tables

Our treatment of first-order logic will be slightly superficial. For our purposes it will suffice to think of a set as an arbitrary collection of objects and a formulae as a mathematical statement that may be true of false. For instance, the collection of even integers is a set and the statements $'2 = 5+3'$ and $'4 = 2+2'$ are formulae. We are going to see the logical operations on formulae. Such operations are defined by the truth values assumed for all possible assignments of truth values for their constituent parts. For instance, if φ and ψ are any two mathematical formulae, we define $\varphi \wedge \psi$ to be true if and only if both φ and ψ and are true. The following table denes other fundamental logical operations on formulae.

Table 2: Truth Table

φ	ψ	$\varphi \wedge \psi$	$\varphi \vee \psi$	$\varphi \Longrightarrow \psi$	$\varphi \Longleftrightarrow \psi$	$\neg\varphi$
T	T	T	T	T	T	F
T	F	F	T	F	F	F
F	T	F	T	T	F	T
F	F	F	F	T	T	T

The above table defines the logical operations of conjunction, disjunction, implies, equivalence and negation as they are used in mathematics. The truth values of more complicated logical formulae may be evaluated by repeatedly applying the above definitions.

Consider the following two logical statements:
A =“Lives in Minnesota"
B =“Lives in USA"

The expression $[A \Longrightarrow B]$ is read “A implies B". It has at least three alternative interpretations, all equivalent and all true if $[A \Longrightarrow B]$ is true: “A is sufficient for B", “B is necessary for A", and “if A, then B". Clearly, we can acknowledge that the statement $[A \Longrightarrow B]$ is a true statement: one cannot live in Minnesota without living in USA.

For any statement, whether true or false, is accompanied by a family of related statements that are connected to the original statement logically. These statements are:

Original Statement:	$A \Longrightarrow B$
Contrapositive:	$\neg B \Longrightarrow \neg A$
Converse:	$B \Longrightarrow A$
Inverse:	$\neg A \Longrightarrow \neg B$

The truth or falsehood of a statement and the truth or falsehood of its contrapositive always agree. The two statements are logically equivalent That is, we may write:

$$[A \implies B] \iff [\neg B \implies \neg A]$$

This statement is read "[A implies B] is true if and only if [not B implies not A] is true." For our example, it is true that living in Minnesota is sufficient for living in USA: $[A \implies B]$ is true. Thus, it is also true that not living in USA is sufficient for not living in Minnesota. Is the statement $B \implies A$ true? No, because one can live in USA without living in Minnesota. However, a statement's converse and its inverse are logically equivalent. If one is true the other must be true.

Now, let's understand when two statements or properties are implied by each other. Consider the following:

$C =$ "Stands 1 meter tall"

$D =$ "Stands 100 centimeters tall"

Clearly, we observe that C can be true only if D is true, and vice versa. Thus, it is a true statement. In general, the expression $[C \iff D]$is read, "C if and only if D" or "C iff D".

Example 1.1. Construct the truth table for $[\varphi \longrightarrow \psi] \vee [\psi \longrightarrow \varphi]$, and show that it is a tautology (a statement that is always true).

Table 3: Truth Table for $[\varphi \longrightarrow \psi] \vee [\psi \longrightarrow \varphi]$

φ	ψ	$\varphi \longrightarrow \psi$	$\psi \longrightarrow \varphi$	$[\varphi \longrightarrow \psi] \vee [\psi \longrightarrow \varphi]$
T	T	T	T	T
T	F	F	T	T
F	T	T	F	T
F	F	T	T	T

Since we observe that the expression $[\varphi \longrightarrow \psi] \vee [\psi \longrightarrow \varphi]$ have a T in all the situations, we can conclude that it is a tautology.

1.2 Quantifiers

If A is any set (collection of objects), the expression "$\forall x \in A$" is shorthand for "for all elements in A". The variable x is here a dummy variable; the expression would be unchanged if we replaced x with any other variable. If A is any set, then the expression "$\exists x \in A$" is shorthand for "there exists an element in the set A". If we omit specification of the set A, then $\forall x$ and $\exists x$ are understood to range over "the universe of all sets" (just think of this as 'everything' for the time being) or a given set that is understood from the context (such as the real numbers).

More generally, if $P(x)$ is a logical statement that depends upon the variable x (such as 'x is an even number') then "$\forall x P(x)$" and "$\exists x P(x)$" stand for "all objects satisfy P" and "there exists an object satisfying P ", respectively. The truth values of the above cannot be defined via truth tables, since such a table would have to exhaust all possible objects.

It is very important that students develop the ability to understand and manipulate logical expressions involving quantifiers. In particular, one must understand how to negate such expressions, and understand that the order of quantifiers is important.

For instance, if $P(x)$ is a logical statement depending upon the variable x, then denying the statement $\forall x P(x)$ is equivalent to asserting $\exists x \neg P(x)$. Similarly, $\neg \exists x P(x)$ is equivalent to $\forall x \neg P(x)$.

None of the above comments require us to assume that the logical expression P depends upon only one variable. As such, we may qualify logical statements with multiple quantifiers.

Example 1.2. For instance, $P(x, y)$ might stand for the statement "$x > y$", where x and y are understood to be real numbers. To take two examples, we see that $P(6, 2)$ is true and $P(2, 4)$ is false. Now let us compare and contrast the expressions:

$$\forall x \exists y\, P(x, y) \text{ and } \exists y \forall x\, P(x, y).$$

The first asserts that given any real number x one may find a real number y greater than x. The second statements asserts that there exists a (fixed) real number y that is strictly less than all other real numbers x. The first statement is clearly true whilst the second statement is clearly false. We thus arrive at the elementary but important observation that the order of quantifiers matters.

Example 1.3. Let's translate the following statement into first order logic:

There exists a unique natural number that is not the successor (i.e. one more than) any other natural number.

$\exists n \in \mathbb{N}\, (\neg\, (\exists m \in \mathbb{N} :\ n = m + 1) \wedge \forall y \in \mathbb{N}\, (\neg\, (\exists z \in \mathbb{N} :\ y = z + 1)) \implies y = n).$

1.3 Methods of proof

Formal mathematical results e.g. theorems, lemmas, propositions, and the like begin with a set of assumptions, definitions, axioms, and other statements that are taken to be true. The proof of a result proceeds from this foundation, taking a series of logical steps that are supported by sound mathematical reasoning. We have in general four ways of delivering a proof:

(1) Deductive reasoning (direct proofs)

The first method is straightforward, and consists of a chain of logic. Find a sequence of accepted axioms, definitions and theorems of the forms A_i, $i = 1, \ldots, n$ such that:

$$A = A_1,\ B = A_n \text{ and } A_i \implies A_{i+1} \text{ for all } i = 1, \ldots, n-1$$

Therefore, we have that: $A = A_1 \implies A_2 \implies \ldots \implies A_n = B$

Example 1.4. Statement A: Let n be an even integer. Let p be an integer. Statement B: np is an even integer. We want to prove $A \Longrightarrow B$.

Proof. By denition, n is an even integer if and only if there exists some integer q such that $n = 2q$. Thus, $np = (2q)p$. By associative property of multiplication, we can write $np = 2(qp)$. Now, it is important to prove that qp is also an integer. But, this is trivially true. Thus, by denition, np is an even integer. $\square$

(2) Proofs by contrapositive (direct proofs)

Recall that contrapositive of a true proposition is always true. Given the proposition P of the form $A \Longrightarrow B$, the contrapositive of P is a proposition $\neg B \Longrightarrow \neg A$. Thus proving one of the above statements is equivalent to proving the other.

(3) Proofs by contradiction (indirect proofs)

This method is essentially an extension of the second method. If the statement A is true and the statement B is not true, then we should NOT find a chain of arguments such that :$\neg B \Longrightarrow \neg A$. For otherwise, it would prove $A \Longrightarrow B$ by contrapositive. This is called a method of proof by contradiction, because we seek a case that the statement B is not true leads to a contradiction.

Example 1.5. Statement: The square root of 2 is irrational.

Proof. By way of contradiction (BWOC), suppose that $\sqrt{2}$ were rational. Then there exist two integers, m and n, that contain no common factors, with $\sqrt{2} = m/n$ or $2 = (m/n)^2$. But then $2n^2 = m^2$, so m^2 is even because it is twice n^2. If m^2 is even, though, m is even so m^2 must be divisible by 4, which means $m^2/2$ is even. Thus, n^2 is even and we know that m and n are both divisible by 2, contradicting the claim that m and n contain no common factors. This completes the proof. $\square$

(4) Inductive reasoning (called mathematical induction)

The fourth method is called mathematical induction. We need this method often when the propositions of interest are indexed by a set of natural numbers. The principle of induction works as follows:

Step 1. Prove that statement $P(1)$ is true

Step 2. Pick an arbitrary natural number k. Prove that, whenever $P(k)$ is true, then $P(k+1)$ is true. Because we have proved $P(1)$ is true in Step 1, by the result of Step 2, $P(2)$ is also true. Because $P(2)$ is true, $P(3)$ is true, and so on. Thus, since propositions are indexed by a set of numbers, an at most countable set, we can safely infer that $P(k)$ is true for any $k \in \mathbb{N}$.

Example 1.6. Statement: For an integer $n \in \mathbb{N}$, we have $\sum_{i=1}^{n} i = n(n+1)/2$.

Proof. For any $n \in \mathbb{N}$, let $P(n)$ be the statement:

$P(n) = 1 + 2 + \ldots + n = \frac{n(n+1)}{2}$

First, we need to check whether $P(1)$ is true, by computing:

$1 = \frac{1(1+1)}{2} = \frac{2}{2} = 1$

Thus, $P(1)$ is true. Now, we assume that $P(k)$ is true for $k > 1$.
We must show that $P(k+1)$ is true.
Because by assumption that $P(k)$ is true, we know that:
$P(k) = 1 + 2 + \ldots + k = \frac{k(k+1)}{2}$
Now, add $k+1$ to both sides to get:
$P(k+1) = 1+2+\ldots+k+k+1 = \frac{k(k+1)}{2}+k+1 = (\frac{k}{2}+1)(k+1) = \frac{(k+2)(k+1)}{2}$
Therefore, we conclude that $P(k+1)$ is true. This completes the proof. $\square$

2 Set theory

We shall largely follow the book of Ok. In the words of Georg Cantor, a set is "a Many which allows itself to be thought of as a One". We will not pursue a rigourous axiomatic treatment of set theory. For our purposes it will suffice to think of a set as any collection of objects. Although this is unsatisfactory for a set theorist, the objects one encounters in a typical graduate sequence of mathematical economics are relatively concrete, and will give rise to no difficulties.

Definition 2.1. A set S is a collection of objects, which are called members of this set.

We will usually denote sets by upper-case letters A, B, C, etc. The objects lying within a given set are referred to as its "members", or "elements". We shall write $x \in A$ for "the object x is an element of the set A". For instance, $2 \in \{2,4\}$ and $3 \notin \{2,4\}$. Two sets are regarded as identical if and only if their members coincide. Formally, for any two sets A and B, we have $A = B$ if and only if $\forall x(x \in A \iff x \in B)$. If every element of a set B also lies in a set A, we call B a subset of A and write $B \subseteq A$. Note the difference between the symbols $\in$ and $\subseteq$. For instance, we have $2 \in \{2,4\}$ and $\{2\} \subseteq \{2,4\}$ but not $2 \subset \{2,4\}$since the real number 2 does not contain any elements.

We may specify sets in several ways. The most natural way is to simply list of the elements as we did above. Thus the set containing the integers between 1 and 5 may be denoted $\{1,2,3,4,5\}$. We may also specify sets by means of logical propositions. For instance, if $\mathbb{Z}$ denotes the set of integers (positive and negative), then the set of even integers may be written $\{2n : n \in \mathbb{Z}\}$. The unique set with no elements is called the empty set and is denoted by $\emptyset$. Note that for any set A, $\emptyset \subseteq A$. Given a set A we may form the set of all subsets of A, termed the power set of A.

Definition 2.2. Given a set A the set $P(A) = \{B : B \subseteq A\}$ is named the power set of A.

The power set of a set A is often denoted as 2^A. If $A = \{a,b\}$, then $P(A) = \{\{\emptyset\},\{a\},\{b\},\{a,b\}\}$.

Note that sets do not come equipped with an order structure, and so for instance, $\{x,y\} = \{y,x\}$.

Note also the distinction between a A and the set containing A as its sole member. In particular, $\{A\} \neq A$.

2.1 Elementary operations on sets

Let's discuss the basic operations on sets. Suppose that A and B are two sets. Then,

Definition 2.3. Given two sets A and B in U, their union, denoted $A \cup B$, is the set $A \cup B = \{x \in U : x \in A \, or \, x \in B\}$.

Definition 2.4. Definition: Given two sets A and B in U, their intersection, denoted $A \cap B$, is the set $A \cap B = \{x \in U : x \in A \, and \, x \in B\}$

Definition 2.5. Definition: Given a set S in U, the complement of S in U is the set $S^c = \{x \in U : x \notin S\}$

Definition 2.6. The set with no elements, denoted by $\emptyset$, is the empty set. Note that this set can depend on the universal set U.

Definition 2.7. The two sets A and B are equivalent, denoted $A = B$, if for all $x \in U$ we have that $[x \in A] \iff [x \in B]$.

Definition 2.8. Given two sets A and B in U, their difference, denoted $A \backslash B$, is the set $A \backslash B = \{x \in U : x \in A \, and \, x \notin B\}$

Definition 2.9. Given two sets A and B in U, then B is a subset of A, denoted $B \subset A$, if $[x \in B] \Longrightarrow [x \in A]$. B is proper subset of A if $B \subset A$ and $B \neq A$.

Definition 2.10. Given two sets A and B in U are disjoint if there is no $x \in U$ with $[x \in A]$ and $[x \in B]$, or if $A \cap B = \emptyset$.

Definition 2.11. Given a set $S \subset U$, a partition of S is a collection of disjoint sets $A_1, A_2, \ldots, A_k$ such that $\bigcup_{i=1}^{k} A_i = S$.

Here, disjointness means that for all i and j, $A_i \cap A_j = \emptyset$.
Now, let's see some properties of elementary operations on sets.
(1) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
(2) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

(3) $A \cap (A \cup B) = A$
(4) $A \cup (B \cap C) = A$
(5) Given the set X, $X \backslash \emptyset = X$ and $X \backslash X = \emptyset$
(6) Given the set X, $A \subset B \Longrightarrow X \backslash B \subset X \backslash A$
(7) Given the set X, $A \cap (X \backslash A) = \emptyset$

Proof. Let's proof property (1):
We have that:
$x \in A \cap (B \cup C) \Longleftrightarrow \{x \in A \, and \, (x \in B \, or \, x \in C)\}$
$\Longleftrightarrow \{(x \in A \, and \, x \in B) \, or \, (x \in A \, and \, x \in C)\}$
$\Longleftrightarrow \{(x \in A \cap B) \, or \, (x \in A \cap C)\}$
$\Longleftrightarrow x \in (A \cap B) \cup (A \cap C)$ □

Theorem 2.1. *(de Morgan): Suppose that t*
here is a countable family of sets $A_1, A_2, \ldots$. We have the following formulae:
(i) For any set B, $(\bigcup_{i=1}^{\infty} A_i) \cap B = \bigcup_{i=1}^{\infty} (A_i \cap B)$ and $(\bigcap_{i=1}^{\infty} A_i) \cup B = \bigcap_{i=1}^{\infty} (A_i \cup B)$.
(ii) $(\bigcup_{i=1}^{\infty} A_i)^c = \bigcap_{i=1}^{\infty} A_i^c$ and $(\bigcap_{i=1}^{\infty} A_i)^c = \bigcup_{i=1}^{\infty} A_i^c$.

In order to understand this theorem let see an example of the first expression of (ii). Consider the following three sets, all subsets of the real numbers: $A_1 = [0, 1]$, $A_2 = [0.5, 3]$ and $A_3 = [4, 5]$. The union of these three sets is $\cup_i = [0, 3] \cup [4, 5]$; the complement of this union is:

$$(\bigcup_{i=1}^{3} A_i)^c = (-\infty, 0) \cup (3, 4) \cup (5, \infty).$$

The complements of the three sets are:
$A_1^c = (-\infty, 0) \cup (1, \infty)$
$A_1^c = (-\infty, 5) \cup (3, \infty)$
$A_1^c = (-\infty, 4) \cup (5, \infty)$
The intersection of the three complements is:
$\bigcap_{i=1}^{3} A_i^c = (-\infty, 0) \cup (3, 4) \cup [5, \infty)$
which is what we expected.

Definition 2.12. Given two sets A and B in U, the cartesian product, denoted $A \times B$, is the set $A \times B = \{(a, b) : \, a \in A \, and \, b \in B\}$.

Example 2.1. If we have the following sets, $A = [1, 2]$ and $B = [0.5, 1]$, then their Cartesian product is $A \times B = \big\{\big(a, b \in \mathbb{R}^2 : \, a \in [1, 2] \, \, and \, b \in [0.5, 1]\big)\big\}$.

2.2 Relations

Definition 2.13. Given a set X , a relation R on X is simply a subset $R \subseteq X^2 = X \times X$.

Two given entities x and y may be "related" in many ways. Consider the following list:
$x = y$
$x < y$
x is "better than" y
x is "further than" y
x is "less damp than" y
Generically, let the notation xRy mean that x is related to y under the relation R.

We say that R is:
- Reflexive if $(x, x) \in R$ for all $x \in X$;
- Symmetric if $(x, y) \in R$ whenever $(y, x) \in R$;
- Antisymmetric if $(x, y) \in R$ and $(y, x) \in R$ together imply $x = y$.
- Transitive if $(x, y) \in R$ and $(y, z) \in R$ imply $(x, z) \in R$ for all $x, y, z \in X$.
- Complete if for all $x, y \in X$, either $(x, y) \in R$ or $(y, x) \in R$.

Definition 2.14. A relation R is an equivalence relation if it is transitive, symmetric and complete. Such a relation is often denoted by the symbol $\sim$.

Given an equivalence relation $\sim$ and an element $x \in X$, we define the equivalence of containing x to be

$$[x]_\sim = \{y \in X : x \sim y\}$$

Some examples:

Example 2.2. Example: Consider the relation R on $\mathbb{R}^2$ given by $((x1, x2), (y1, y2)) \in R$ iff $\sqrt{x_1^2 + x_2^2} = \sqrt{y_1^2 + y_2^2}$. Then R is an equivalence relation and equivalence classes are circles (or the origin).

Example 2.3. Example: Let X be any set and consider the relation R on $P(X)$ given by $(A, B) \in R$ iff $A \cap B \neq \emptyset$. Then R is not an equivalence relation as it is not transitive.

Definition 2.15. A set $A \subset X$ whose members are all related to each other by the equivalence relation is called an equivalence class.

Definition 2.16. Given any two sets $A, B \subseteq R$ we define the set sum (sometimes called Minkowski sum) and set difference as follows:
$A + B = \{a + b : a \in A, b \in B\}$
$A - B = \{a - b : a \in A, b \in B\}$

We emphasise that the above addition and subtraction operations are only defined on sets equipped with an additive structure. In contrast, the operations of intersection, union and complementation introduced are well-defined on arbitrary sets.

Example 2.4. Lets have two triangles: $A = \{(2,0), (1,1), (1,-1)\}$ and $B = \{(1,0), (2,1), (2,-1)\}$. Calculate the Minkowski sum.
$A + B = \{(3,0), (4,1), (4,-1), (2,1), (3,2), (3,0), (2,-1), (3,0), (3,-2)\}$

Remark 2.1. Be sure to note the difference between the operations $\setminus$ and $-$ on sets. For instance, $\mathbb{Q}\setminus\mathbb{Q} = \emptyset$ but $\mathbb{Q} - \mathbb{Q} = \mathbb{Q}$.

2.3 Orders

We now provide a few more definitions that extend our discussion of relations. Given a set X and a relation $R \subseteq X^2$, we adopt the following conventions:

- R is a preorder if it is transitive and reflexive;
- R is a partial order if it is transitive, reflexive and antisymmetric; and
- R is a linear order if it is transitive, reflexive, antisymmetric and complete.

To remind ourselves that we are dealing with preorders and not arbitrary relations, we often denote an arbitrary preorder by $\succeq$ instead of R. Given such a preorder on a set X , we define the asymmetric part of $\succeq$, denoted $\succ$ as follows:

$$x \succ y \text{ iff } x \succeq y \wedge (\neg y \succeq x)$$

The symmetric part of $\succeq$, denoted by $\sim$ is defined by:

$$x \sim y \text{ iff } x \succeq y \wedge y \succeq x$$

Given a preorder $\succeq$ and a subset $A \subseteq X$ an element x is $\succeq$-maximal in A if there does not exist $y \in A$ such that $y \succ x$, and is a $\succeq$-maximum in A if$x \succeq y$ for all $y \in A$ (minimal elements and minima are defined similarly).

Usually in many textbooks you are going to encounter just the definition of an order set. Therefore, it is important to have a perspective of this concept.

Definition 2.17. (Order): Let A be a set. An order on A is a relation, denoted by $<$, with the following properties:

(i) For all $x, y \in A$, one and only one of the following statements is true: $x < y$, $x = y$, $x > y$ (often called, "complete")

(ii) For all $x, y, z \in A$, if $x > y$ and $y > z$, then $x > z$ (often called "transitive").

Definition 2.18. An ordered set is a set A in which an order is defined.

Example 2.5. (Preferences) A complete preorder $\succeq$ is locally nonsatiated on a set X if for all $x \in X$ and all $\epsilon > 0$, $\exists y \in X$ such that $||x - y|| < \epsilon$ and $y \succ x$.

Definition 2.19. (Upper and lower bounds): Let S be an ordered set and let $E \subseteq S$. If there exists a $\beta \in S$ such that $x \leq \beta$ for every $x \in E$, we say that E is bounded above, and call β an upper bound of E. A lower bound is similarly defined (with $\geq$ instead of $\leq$).

Definition 2.20. (Least upper bound and greatest lower bound): Let S be an ordered set, $E \subset S$, and E is bounded above. Then, we say that α is the least upper bound (or supremum) of E if it has the following properties:
(i) α is an upper bound of E (i.e. for all $x \in E$, $x \leq \alpha$).
(ii) If $\gamma < \alpha$, γ is not an upper bound of E.

Usually, the least upper bound is denoted as:

$$\alpha = sup\, E$$

The greatest lower bound (or infimum), of a set E which is bounded below is defined in the same manner. Usually, the greatest lower bound is denoted as:

$$\alpha = inf\, E$$

Example 2.6. Let $E_1 = [0, 10)$ and $E_2 = [0, 10]$, $E_1, E_2 \subset \mathbb{R}$. What are the $\sup E_1$and $\sup E_2$?
Ans: We have that $\sup E_1 = \sup E_2 = 10$, however $10 \notin E_1$ and $10 \notin E_2$.
Let $E = \{1/n : n = 1, 2, 3, \ldots\} \subset \mathbb{R}$ What are the $\sup E$and $\inf E$?
Ans: We have that $\sup E = 1$ and $\inf E = 0$.

Definition 2.21. (Convex Sets): A set $U \subseteq \mathbb{R}^n$ is convex if and only if for all $x, y \in U$ and $\lambda \in [0, 1]$, $\lambda x + (1 - \lambda) \in U$.

Example 2.7. Consider the interval $[a, b] \subset \mathbb{R}$. We want to show that this set is a convex set.

Let pick two arbitrary elements $x, x^{'} \in [a, b]$. The we need to show that: $\lambda x + (1 - \lambda) \in [a, b]$ for all $\lambda \in [0, 1]$.

Select an arbitrary λ in $[0, 1]$. Since $x, x^{'} \in [a, b]$, then $x, x^{'} \leq b$. Then, as we have that $\lambda \in [0, 1]$ it follows that $\lambda x + (1 - \lambda) \leq b$. Similarly, we obtain that $\lambda x + (1 - \lambda) \geq a$. Therefore, we have that $\lambda x + (1 - \lambda) \in [a, b]$ for all arbitrary $x, x^{'}\, and\, \lambda$.

3 Functions

Definition 3.1. For any two sets A and B a function $f : A \longrightarrow B$ is simply a map that assigns to every element $x \in A$ a unique element $f(x) \in B$.

The set A is called the domain and the set B is called the codomain. The set $\{y \in B : \exists x \in A, f(x) = y\}$ is called the range. For example the function $f : \mathbb{R} \longrightarrow \mathbb{R}$ defined by $f(x) = x^2$ for all $x \in \mathbb{R}$ has domain $\mathbb{R}$, codomain $\mathbb{R}$ and range $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$. The set of all functions from a set A to a set B is denoted B^A. Note that this motivates the above notation for the power set of a set.

Definition 3.2. A function $f : A \longrightarrow B$ is called injective (or one-to-one) if there does not exist a pair $x, y \in A$ such that $x \neq y$ and $f(x) = f(y)$.

That is, one and only one member of A maps into any given member of B.

Definition 3.3. A function $f : A \longrightarrow B$ is called surjective (or onto) if for all $z \in B$ there exists $x \in A$ such that $f(x) = z$.

That is, every member of A is mapped into by some member of B.

Definition 3.4. A function $f : A \longrightarrow B$ is called a bijection if it is both injective and surjective.

Theorem 3.1. *If $f : A \longrightarrow B$ is a bijection, then there exists a function $f^{-1} : B \longrightarrow A$, called the inverse function of f, such that:*

$$f\left(f^{-1}(b)\right) = b \ \forall b \in B \text{ and}$$

$$f^{-1}(f(a)) = a \ \forall a \in A$$

Example 3.1. If $f : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ is given by $y = f(x) = x^2$ then $f^{-1}(y) = \sqrt{y}$ is the inverse function, where $f^{-1} : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$.

Definition 3.5. Consider two functions $f : A \longrightarrow B$ and $g : B \longrightarrow C$. The composition $g \circ f : A \longrightarrow C$ is the function $g \circ f(a) = g(f(a))$.

Example 3.2. Let $f : [0, 1] \longrightarrow \mathbb{R}$ be defined by $f(x) = 3x$ and let $g : \mathbb{R} \longrightarrow \mathbb{R}$ be defined by $g(y) = y^2 + \sqrt{y}$. Then the composite fucntion is $g(f(x)) = 9x^2 + \sqrt{3x}$.

If a set A has a finite number of elements, then we write $|A|$ for the number of its elements. The quantity $|A|$ is termed the cardinality of the set A. The notion of size may be extended to infinite sets, but as we shall see shortly, this involves some subtlety. Two sets (not necessarily finite) A and B have the same cardinality if there exists a bijection between them. We now come to the first observation with real content.

Definition 3.6. If there exists a one-to-one mapping from A to B, then we say that A and B have the same cardinality or A and B are equivalent.

Example 3.3. The set of integers has the same cardinality as the set of even integers, as $n \longrightarrow 2n$ is a bijection.

Example 3.4. The set of real numbers in the open interval $(0, 1)$ has the same cardinality as the set of real numbers, as $x \longrightarrow \arctan x$ is a bijection.

Definition 3.7. (Finite, infinite, countable sets): For any set A,

(i) A is finite if A has the same cardinality as the set $\{1, 2, \ldots, n\}$ for some finite $n \in \mathbb{N}$.

(ii) A is infinite if A is not finite.

(iii) A is countable (enumerable/denumerable) if A has the same cardinality as $\mathbb{N}$.

(iv) A is uncountable if A is neither finite nor countable.

(v) A is at most countable if A is finite or countable.

A set A is specially called countably infinite if $A \sim \mathbb{N}$.

Example 3.5. $\mathbb{Q}$ and $\mathbb{Z}$ are countably infinite sets. $\mathbb{R}$ is uncountable. Every (infinite) subset of a countable set is countable.

Theorem 3.2. *(Cantor). There exists no bijection between the natural numbers and the real numbers.*

Proof. The diagonal derivation. □

Theorem 3.2 is a very important piece of miscellaneous information. It shows that not all infinite sets are the same 'size' and that there are strictly 'more' reals than naturals. Given any two sets A and B let us write $A \longrightarrow B$ if there exists an injection from A to B. The following theorem demonstrates that the above relation extends the notion of 'less than or equal to' to infinite cardinals.

Theorem 3.3. *3.2 (Schroeder-Bernstein). For any two sets A and B there exists a bijection between A and B if and only if there exist injective functions $f : A \longrightarrow B$ and $g : B \longrightarrow A$ between the sets A and B.*

It is possible to restate this theorem with the notion of cardinality. This means that if $|A| \leq |B|$ and $|B| \leq |A|$, then $|A| = |B|$.

Proof. The only if direction is obvious. For the converse assertion, let us suppose that $f : A \longrightarrow B$ and $g : B \longrightarrow A$ are two injections. We wish to construct a bijection $h : A \longrightarrow B$. We omit this as it is a bit complicated. □

Definition 3.8. (Concave Function): For a set $A \subset \mathbb{R}^n$, a function $f : A \longrightarrow \mathbb{R}$ is concave if for all $x_1, x_2 \in A$ and $\lambda \in [0, 1]$, we have that: $f(\lambda x_1 + (1 - \lambda) x_2) \geq \lambda f(x) + (1 - \lambda) x_2$. The function f is strictly concave if the inequality is strict for $\lambda \in (0, 1)$.

Definition 3.9. (Covex Function): For a set $A \subset \mathbb{R}^n$, a function $f : A \longrightarrow \mathbb{R}$ is convex if for all $x_1, x_2 \in A$ and $\lambda \in [0, 1]$, we have that: $f(\lambda x_1 + (1 - \lambda) x_2) \leq \lambda f(x) + (1 - \lambda) x_2$. The function f is strictly convex if the inequality is strict for $\lambda \in (0, 1)$.

Remark 3.1. Note that the notion of a convex function is quite dierent than the notion of a convex set.

4 Basic Real Analysis and Topology

Definition 4.1. (Sequence): A sequence is a function $f : \mathbb{Z} \longrightarrow A$, for any set A. We denote the sequence by $\{x_i\}_{i=1}^{n}$. The elements x_i are called the terms of the sequence.

4.1 Distance and limit of sequences

In mathematics and economics, we need a rigorous definition of distance in order to avoid ambiguity, as we usually deal with abstract spaces such as the space of real numbers.

Definition 4.2. (Metric and metric spaces): A metric space is an ordered pair (X, d), in where the set X is endowed with a metric (distance) d, such that for any $p, q, r \in X$, we have that the following holds:
(i) $d(p, q) > 0$ if $p \neq q$ and $d(p, q) = 0$ if $p = q$ (non-negative and identity)
(ii) $d(p, q) = d(q, p)$ (symmetry)
(iii) $d(p, q) \leq d(p, r) + (r, q)$ (triangle inequality)

We call any function with these properties a distance function or a metric. In $\mathbb{R}$, the absolute value is the natural choice for a metric. That is, $d(x, y) = |x-y|$ for any $x, y \in \mathbb{R}$.

We can easily apply the same to the Euclidean norm in $\mathbb{R}^n$, defined by:

$$d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

Since we have already defined the concept of distance, we are able to formally discuss the limit of a sequence.

Definition 4.3. (Limit of a real sequence) Let $\{x_n\}$ be a sequence of real numbers. The sequence is said to converge if there exists a real number r such that for every (small) $\epsilon > 0$ there is an integer N such that $\forall n \geq N$, we have that $|x_n - r| < \epsilon$. We call r the limit of this sequence.

This denition can be extended to any metric space.

Definition 4.4. (Limit of sequence) Let $\{x_n\}$ be a sequence in a metric sapce X. The sequence is said to converge if there exists a $p \in X$ such that for every (small) $\epsilon > 0$, there is an integer N such that $\forall n \geq N$, we have that $d(x_n - p < \epsilon)$.

Formally, a real sequence is a function $f : \mathbb{N} \longrightarrow \mathbb{R}$, although we will usually denote sequences by $(x_1, x_2, x_3, \ldots)$ or $\{x_n\}$.

Definition 4.5. A sequence $\{x_n\}$ is Cauchy if for all $\epsilon > 0$ there exists a N such that $|x_n - x_m| < \epsilon$ whenever $n, m > N$.

In first-order logic we have that a Cauchy sequence is defined as:

$$\forall \epsilon > 0 \exists N \in \mathbb{N} \, (n, m > N \Longrightarrow |x_n - x_m| < \epsilon)$$

Once again, it is worth emphasising that this statement is very dierent from:

$$\exists N \forall \epsilon > 0 \in \mathbb{N} \, (n, m > N \Longrightarrow |x_n - x_m| < \epsilon)$$

which can only hold if the sequence $(x_1, x_2, x_3, \ldots)$ is ultimately constant.

Unless otherwise noted, we will focus on real sequences. Nevertheless, as we have defined above, we can straightforwardly extend our denitions/theorems to metric spaces. The definitions aforementioned mainly says that if the sequence $\{x_n\}$ converges, $\{x_n\}$ should become arbitrarily close to a point, as n goes to infinity. Usually, we can also write this as:

$$\lim_{n \to \infty} x_n = r$$

If $\{x_n\}$ does not converge, then we say it diverges. There are two kinds of divergent sequences:

- The sequence decreases or increases without bound. e.g.: $x_n = 2^n = 2, 4, 8, \ldots$
- The sequence oscillates (have cyclical behavior). e.g.:$x_n = (-1)^n \left(\frac{n}{n+1}\right) = -\frac{1}{2}, \frac{2}{3}, -\frac{3}{4}, \ldots$

In the second example, the sequence jumps back and forth and get arbitrarily close to either -1 or 1. We call these points, accumulation points.

Definition 4.6. Let r be a real number. We call r an accumulation point of a real sequence $\{x_n\}$ if for any $\epsilon > 0$, the set $\{x_i : |x_i - r| < \epsilon\}$ contains infinitely many elements of $\{x_n\}$.

Therefore, there can be multiple accumulation points in any sequence, but there should be at most one limit in any sequence.

Definition 4.7. (Subsequence) Given a sequence $\{x_n\}$, consider a strictly increasing sequence $\{n_k\}$ of positive integers (i.e. $n_1 < n_2 < n_3, \ldots$). Then, the sequence x_{n_i} is called a subsequence of x_n. If the subsequence converges, its limit is called a subsequential limit of $\{x_n\}$

Theorem 4.1. *Suppose x_n and y_n are real sequences, and* $\lim x_n = x$ *and* $\lim y_n = y$. *Then, the following are true:*
(i) $\lim (x_n + y_n) = x + y$
(ii) $\lim cx_n = cx$ *for any number* c
(iii) $\lim x_n y_n = xy$
(iv) $\lim \frac{1}{x_n} = \frac{1}{x}$, $x_n \neq 0$ *and* $x \neq 0$.

Proof. We are going to prove only (i) here. Note first that we want to prove that, for every $\epsilon > 0$, there exists $N_0 > 0$ such that $\forall n \geq N_0$, $|\,(x_n + y_n) - (x - y)\,| < \epsilon$. So, first pick any arbitrarily small $\epsilon > 0$ and fix it. Now, we know that x_n and y_n both converge to x and y, respectively. So, by definition, for a number '$\epsilon^{'} = \frac{\epsilon}{2}$, there exist numbers N_1 and N_2, respectively, such that:

$$\forall n \geq N_{1,} |x_n - x| < \epsilon^{'} = \tfrac{\epsilon}{2}$$
$$\forall n \geq N_{2,} |y_n - y| < \epsilon^{'} = \tfrac{\epsilon}{2}$$

Then, letting $N_0 = \max\{N_1, N_2\}$,we have:

$$\forall n \geq N_{0,} |x_n - x| < \epsilon^{'} = \tfrac{\epsilon}{2}$$
$$\forall n \geq N_{0,} |y_n - y| < \epsilon^{'} = \tfrac{\epsilon}{2}$$

Therefore, for $\epsilon > 0$, we have found an integer N_0 such that $\forall n \geq N_0$:

$|\,(x_n + y_n) - (x - y)\,| = |x_n - x + y_n - y| \leq |x_n - x| + |y_n - y| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$

Because $\epsilon > 0$ is chosen arbitrarily , we can do the same operation for every possible $\epsilon > 0$. ☐

The limit of the sequence concept aforementioned can be extended to the multi-dimensional space $\mathbb{R}^n$ (the Euclidean space). The application is similar to what we do in $\mathbb{R}$, however in $\mathbb{R}^n$ we use the Euclidean norm $\|\cdot\|$ instead of the absolute value $|\cdot|$. It is also important to understand that as we are in a multi-dimensional space, a sequence can move in many different directions in $\mathbb{R}^n$. For example, in $\mathbb{R}^2$, we might have a sequence $x_n = \left(\frac{1}{n}, n^2\right)$. This sequence has the following pattern: $\left\{(1,1), \left(\frac{1}{2}, 4\right), \left(\frac{1}{3}, 9\right), \ldots\right\}$. This is obviously not convergent.

Theorem 4.2. *Suppose $\{x_n\}$ is a sequence in $\mathbb{R}^n$. $\{x_n\}$ converges to $r = (r_1, \ldots, r_n)$ if and only if its all components $x_{j,n}$ converge to r_j , $j = 1, \ldots, n$*

Proof. We skip the proof. Make sure you understand the proof in p.262-263 in S&B, especially the only if part. ☐

Now, it is important to understand that in any metric space, we can define the concepts of open set, closed set, compact set, etc.

4.2 Point set topology on the real line and $\mathbb{R}^n$

For any pair $a, b \in \mathbb{R}$ we first dene open and closed intervals, both infinite and finite:

$$[a, b] = \{x \in \mathbb{R} :\ a \leq x \leq b\}$$

$$[a, b) = \{x \in \mathbb{R} :\ a \leq x < b\}$$

$$(a, b] = \{x \in \mathbb{R} :\ a < x \leq b\}$$

$$(a, b) = \{x \in \mathbb{R} :\ a < x < b\}$$

A subset $O \subseteq \mathbb{R}$ of the real line is open if for all $x \in O$ there exists $\epsilon > 0$ such that $(x - \epsilon, x + \epsilon) \subseteq O$. A point x is a point of closure of the set $A \subseteq \mathbb{R}$ if every open set containing x contains a point in A. The closure of a subset of the real line is the set of all its points of closure. A subset of the real line is closed if it is equal to its closure.

We can expand this perspective to any metric space.

Definition 4.8. (Neighborhood): For any metric space X, a neighborhood of a point x is a set $N_r(x)$ consisting of the all points y such that $d(x, y) < r$, for some $r > 0$. That is: $N_r(x) = \{y \in X : d(x, y) < r\}$ for some $r > 0$.

Definition 4.9. (Interior): For any set $E \subset X$, a point x is an interior point of E if there exists a neighborhood N of x such that $N \subset E$. We denote the set of all interior points of E as $int(E)$.

Definition 4.10. (Open set): A set $E \subset X$ is said to be open if every point $x \in E$ is an interior point of E, i.e., E is open if and only if

$\forall x \in E \exists r > 0,\ N_r(x) \subset E.$

Example 4.1. An open interval $(1, 2)$ is an open set in $\mathbb{R}$. To see why, pick any x in $(1, 2)$. So, by denition, $1 < x < 2$ and $x \neq 1$ and $x \neq 2$. Thus, $|1 - x| = r_1 > 0$ and $|x - 2| = r_2 > 0$. Then, let $r = \frac{1}{2}\min\{r_1, r_2\}$. Clearly, the set $N_r(x) = \{y : y - x < r\}$ is contained in $(1, 2)$.

Definition 4.11. A set $F \subseteq X$ is said to be closed if and only if its complement F^c is open. Also, a set $F \subseteq X$ is said to be closed if and only if for every convergent sequence $\{x_n\}$contained in F, its limit x is also contained in F.

Definition 4.12. (Limit points): A point x is a limit point of E if every neighborhood $N(x)$ of x contains a point $y \neq x$ such that $y \in E$.

Definition 4.13. (Closure): A closure of a set $E \subseteq X$ is the intersection of all closed sets containing E (i.e. the 'smallest' closed set containing E). We denote the closure of E, clE or $\bar{E}$.

Definition 4.14. (Bounded set): A set $E \subseteq X$ is said to be bounded if there is a (finite) real number M and a point y such that $d(x, y) < M$ for all points x in E.

Definition 4.15. (Open cover): An open cover of a set $E \subseteq X$ is a collection $\{G_\alpha\}$ of open subsets such that $E \subseteq \bigcup_\alpha G_\alpha$.

Definition 4.16. (Compact sets): A set $E \subseteq X$ is said to be compact if, for every open cover $\{G_\alpha\}$ of E, there exists a finite collection such that: $E \subset \bigcup_{i=1}^{n} G_{\alpha_i}$.

Theorem 4.3. *(Heine-Borel). Let $A \subseteq \mathbb{R}$ be closed and bounded. Then every cover of A has a finite subcover (A is compact).*

Proof. First suppose that $A\,[a,b]$ for some $a, b \in \mathbb{R}$ and let $\mathcal{F}$ be an open cover of A. Define E to be the set of all $x \in [a,b]$ such that $[a,x]$ may be covered with finitely many elements of $\mathcal{F}$. Consider the supremum of E. It must be b. Then take complements. □

Theorem 4.4. *(Nested Sets). Every nested sequence of non-empty closed, bounded subsets of the real line has a non-empty intersection.*

Proof. If the conclusion were false, the complements of the given sets would form an open cover of the original closed and bounded with no finite subcover, contradicting Heine- Borel. □

Theorem 4.5. *(i) Any (finite or infinite) union of open sets is open. (ii) The finite intersection of open sets is open.*

Proof. (i) Let $G = \bigcup_\alpha G_\alpha$. Pick any x in G. Then, by definition of a union, x must belong to some set G_α, for some α. Since each G_α is open, we can find a neighborhood $N_r(x)$ such that $N_r(x) \subset G_\alpha$. Thus, all points in $N_r(x)$ are contained in G_α. But, then, by definition, all points of $N_r(x)$ must be contained in G, as all points of G_α are contained in G.

(ii) Consider a fnite collection of open sets $G_1, \ldots, G_n$. Let $G = \bigcap_{i=1}^{n} G_i$. Suppose x is in G. (Note here that G may be empty. An empty set is open and closed by definition). Then, x must be contained in each and every G_i. Since each G_i is open, there exists a neighborhood $N_{r_i}(x)$ such that $N_{r_i}(x) \subset G_i$, for each i. Put, $r = min\,\{r_{1,\ldots,}r_n\}$. Then, the neighborhood $N_r(x)$ must be contained in each and every G_i, so that $N_r(x) \subset G$. Thus, G is open. □

Example 4.2. For part (ii) of the previous theorem, the finiteness of the collection is essential. To see why, consider the following infinite collection of open sets: $G_n = \left(-\frac{1}{n}, \frac{1}{n}\right)$, $n = 1, 2, \ldots$. Each G_n is an open set in $\mathbb{R}$. Put $G = \bigcap_n G_n$. If n stops at some finite number N, then, G is clearly an open set $\left(-\frac{1}{N}, \frac{1}{N}\right)$. However, if $n \longrightarrow \infty$, then G consists of only one point, i.e. 0. (Note that the open interval $\left(-\frac{1}{N}, \frac{1}{N}\right)$ does not include the endpoints $-\frac{1}{N}, \frac{1}{N}$, so this is the set of points between $-\frac{1}{N}$ and $\frac{1}{N}$ excluding the endpoints. Thus, $\lim_{n \longrightarrow \infty} \frac{1}{N}$

does not become actually 0, the set $\lim_{n\longrightarrow\infty}\left(-\frac{1}{N},\frac{1}{N}\right)$ becomes a singleton set $\{0\}$. Therefore, it is not an open set in $\mathbb{R}$.

Theorem 4.6. *(Bolzano-Weierstrass). Every bounded sequence of real numbers contains a convergent subsequence.*

Proof. Closure of the tail of the sequence forms a nested sequence of closed sets. Then inductively choose an appropriate subsequence. □

Proposition 4.1. *A real sequence converges if and only if it is Cauchy.*

Proof. The fact that convergent sequence are Cauchy follows from $|x_n - x_m| \leq |x_n - x| + |x_m - x|$. Conversely, we show that Cauchy sequences are bounded, invoke Bolzano-Weierstrass to get existence of a convergent subsequence, and then use this to get full convergence, using $|x_n - x_m| \leq |x_n - x_{n_k}| + |x_{n_k} - x|$. □

Now, let's show an example to understand the concepts of limit superior and limit inferior.

Example 4.3. Consider the function $f:\mathbb{R}_{++}\longrightarrow\mathbb{R}$ given by $f(x)=\sin\left(\frac{1}{x}\right)$. Then $\limsup_{x\longrightarrow 0} f(x) = 1$ and $\liminf_{x\longrightarrow 0} f(x) = -1$.

Definition 4.17. (Limit superior and inferior) Given a sequence $\{x_n\}_{n=1}^{\infty}$, we define limit inferior as $\liminf_{n\longrightarrow\infty} x_n = \lim_{n\longrightarrow\infty}\left(\inf_{m\geq n} x_m\right)$ and $\limsup_{n\longrightarrow\infty} x_n = \lim_{n\longrightarrow\infty}\left(\sup_{m\geq n} x_m\right)$ as limit superior, provided that they exists as a possibly extended-valued real number.

4.3 Continuity

In the real line we can define continuity as:

Definition 4.18. A function $f:\mathbb{R}\longrightarrow\mathbb{R}$ is continuous at a point $x\in\mathbb{R}$ if and only if $\forall\epsilon>0\ \exists\delta>0$ such that: $|x-y|<\delta \Longrightarrow |f(x)-f(y)|<\epsilon$.

Proposition 4.2. *A function $f:\mathbb{R}\longrightarrow\mathbb{R}$ is continuous at a point $x\in\mathbb{R}$ if and only if $|f(x_n)\longrightarrow f(x)|$ for all sequences $\{x_n\}_{n=1}^{\infty}$ such that $x_n\longrightarrow x$.*

Proposition 4.3. *A function $f:\mathbb{R}\longrightarrow\mathbb{R}$ is continuous at a point $x\in\mathbb{R}$ if and only if the limit inferior and limit superior coincide there.*

Proposition 4.4. *A real-valued function on the real line is continuous if and only if preimages of open sets are always open.*

Theorem 4.7. *(Extreme value theorem). Continuous functions on closed and bounded sets assume their suprema and infima.*

Proof. First show boundedness of the function using the Heine-Borel theorem, then consider $|f(x) - \sup f|^{-1}$ if supremum is not attained. □

Theorem 4.8. *(Intermediate value theorem). Let f be a continuous function on a closed interval $[a, b]$ such that $f(a) < f(b)$. Then for any $z \in (f(a), f(b))$ there exists a point $c \in [a, b]$ such that $f(c) = z$.*

Proof. Keep bisecting the intervals until we arrive at the desired point and use the nested set theorem. □

Proposition 4.5. *A continuous, real-valued function on a closed, bounded set of real numbers is uniformly continuous.*

Example 4.4. For example, the following function is not continuous at $x_0 = 0$:

$$f(x) = \begin{cases} 2x\, if & x > 0 \\ x^2 + 1\, if & x \leq 0 \end{cases}$$

To see this, pick $\varepsilon = 1/2 > 0$. At $x_0 = 0$, $f(0) = 0 + 1 = 1$. But, pick any point to the right of x_0 $(i.e\, x > 0)$. Then, no matter how close we make this x to 0, the value of $f(x)$ goes only to 0 and never gets to 1. This can be seen very easily in a graph. But, clearly, the following function is continuous:

$$f(x) = \begin{cases} 2x\, if & x > 0 \\ x^2\, if & x \leq 0 \end{cases}$$

Although the concept is intuitively clear, determining whether or not a function is continuous is sometimes very difficult. There are two kinds of discontinuities. The first kind is called "jump discontinuity" and is easy to detect. The second kind is more difficult, because even the right-hand or left-hand limit of f cannot be defined.

Theorem 4.9. Let $X \subset \mathbb{R}^n$. Let f and g be real-valued continuous functions from $X \subset \mathbb{R}^n$. Then,

$(i)\, f(x) + g(x)$

$(ii)\, f(x)\, g(x)$

$(iii)\, f(x)/g(x)$ $(where\, g(x) \neq 0)$

are also continuous on X.

Theorem 4.10. Let X, Y, Z be metric spaces and $E \subset X$. If $f : E \to Y$ and $g : f(E) \subset Y \to Z$ are continuous, then the composite function $h = (g \circ f)$ of these two functions, defined as:

$h(x) = g((f(x)))$ $for\, x \in E$

is also continuous. The function $h(x)$ is called a composite function. This is an intuitively trivial but very important result.

5 One-variable Calculus

Theorem 5.1. *(Rolle's). Let f be a continuously dierentiable function defined on an interval $[a, b]$ satisfying $f(a) = f(b) = 0$. Then there exists a point $x \in (a, b)$ such that $f'(x) = 0$.*

Proof. This is similar in flavour to the intermediate value theorem applied to the derivative. □

Definition 5.1. (Differentiability): Let $X \subset \mathbb{R}$. A function $f : X \to \mathbb{R}$ is said to be differentiable at $x = x_0$ if and only if the limit (1) exists. The function is differentiable on an open set X if and only if it is differentiable at every point on X.

Note that, for the derivative to exist, h must be able to take both negative and positive values (that is, it can approach from above and below). Thus, if a function is dened on a closed interval $[a, b]$, it is not differentiable at the endpoints $x = a$ and $x = b$, by definition. Formally, we have the following:

Definition 5.2. A function $f : X \to \mathbb{R}$ is differentiable at $x = x_0$ if and only if both the right-hand derivative,

$$f'_+(x_0) = \lim_{h \searrow 0} \frac{f(x_0+h)-f(x_0)}{h}$$

and the left-hand derivative,

$$f'_-(x_0) = \lim_{h \nearrow 0} \frac{f(x_0+h)-f(x_0)}{h}$$

exist and are equal to each other.

Theorem 5.2. Let f be dened on $[a, b]$. If f is differentiable at $x \in [a, b]$, then it is continuous at x.

Proof. Let $f'(x_0) = A$. By denition, for every $\varepsilon > 0$, there exist $\delta > 0$ such that,

$| h - 0 | < \delta \Rightarrow | \frac{f(x+h)-f(x)}{h} - A | < \varepsilon$

So, let $\varepsilon = 1$ and δ_1 be the δ that makes this statement holds for $\varepsilon = 1$. Note that we can rewrite the right-hand side:

$| f(x+h) - f(x) - Ah | = | \frac{f(x+h)-f(x)}{h} h - Ah | \leq | \frac{f(x+h)-f(x)}{h} h - A | | h | < 1 \cdot | h |$

Moreover, we can also manipulate $| f(x+h) - f(x) - Ah |$ to get:

$$| f(x+h) - f(x) | \leq | f(x+h) - f(x) - f(x) - Ah | + | Ah |$$

Or,

$$| f(x+h) - f(x) | - | Ah | \leq | f(x+h) - f(x) - Ah |$$

Combining these two, we have:

$$\therefore | f(x+h) - f(x) | < | h | + | A || h |$$

Now, for each $\varepsilon > 0$, take $\delta = min \left\{ \frac{\varepsilon}{1+|A|}, \delta_1 \right\}$. Then, the following must hold:

$| h | < \delta \Rightarrow | f(x+h) - f(x) | < | h | + | A || h | = | h | (1 + | A |) < \frac{\varepsilon}{1+|A|} (1 + | A |) = \varepsilon$

Now, since$h$ is arbitrary, rewrite $x+h=x^{'}$. Then, this exactly coincides with the definition of continuity. □

Example 5.1. Lets compute the derivative of $f(x)=x^2$ at $x=3$, using the denition.

$f^{'}(3)=\lim\frac{f(3+h)-f(3)}{h}$
$=\lim\frac{f(3+h)^2-3^2}{h}$
$=\lim\frac{9+6h+h^2-9}{h}$
$=\lim\frac{h(6+h)}{h}$
$=\lim 6+h$

On the other hand, we know that $f^{'}(x)=2x$, so that $f^{'}(3)=6$.

Theorem 5.3. Suppose that $f$and$g$ are dened on a,b and are differentiable at $x\in(a,b)$. Then, $f+g$, fg, and f/g are differentiable at x, and

$(i)\,(f+g)^{'}(x)=f^{'}(z)+g^{'}(x)$
$(ii)\,(fg)(x)=f^{'}(x)\,g(x)+f(x)\,g^{'}(x)$(Product Rule)
$(iii)\,(f/g)^{'}(x)\,\left[f^{'}(x)\,g(x)-f(x)\,g^{'}(x)\right]/g^2(x)$(Quotient Rule)

Moreover, let k be an arbitrary constant in $\mathbb{R}$. Then,

$(iv)\,\left(x^k\right)^{'}=kx^{k-1}$

Finally, for any a in $\mathbb{R}$,

$(v)\,(a^x)^{'}=a^x log(a)$

(i) Recall derivatives of some special functions (Try to prove them at home):

$(e^x)^{'}=e^x$
$(logx)^{'}=\frac{1}{x}$
$(sin\,x)^{'}=cos\,x$
$(cos\,x)^{'}=-sin\,x$

(ii)

$\left(2x^k log(x)\right)=\left(2x^k\right)(log(x))+\left(2x^k\right)^{'}(log(x))$
$=2kx^{k-1}log(x)+2x^k(1/x)$
$=2kx^{k-1}log(x)\,2x^{k-1}$
$=2kx^{k-1}log(x)+2x^{k-1}$
$=2x^{k-1}(klog(x)+1)$

(iii)

$(e^{ax}/x)^{'}=\frac{(e^{ax})^{'}x-e^{ax}(x)^{'}}{x^2}$
$=\frac{ae^{ax}x-e^{ax}}{x^2}$
$=e^{ax}\left(\frac{ax-1}{x^2}\right)$

Theorem 5.4. (Chain Rule): Suppose that $f:[a;b]\to\mathbb{R}$ is differentiable on (a,b) and $g:\,Y\to\mathbb{R}$ is also differentiable on Y (where $f([a,b])\subset Y$). Then, the composite function $h:[a,b]\longrightarrow\mathbb{R}$, defined by:

$h(x)=g(f(x))\;\;for\,x\in(a,b)$

is also differentiable on (a,b) and its derivative is given by:

$h^{'}(x)=g^{'}(f(x))\,f^{'}(x)\;\;for\,x\in(a,b)$

Example 5.2. (i) For example, if $f(x) = y = x^2$ and $g(y) = z = 3y + 1$, then, the composite of two functions is $h(x) = g(f(x)) = 3(x^2) + 1 = 3x^2 + 1$. We can derive the derivative of this function either directly or by using Chain Rule. It is easy to see $h^{'}(x) = 6x$. What if we use Chain Rule? $h^{'}(x) = g^{'}(f(x)) f^{'}(x) = (3) \cdot (2x) = 6x$.

(ii) Consider the function $h(x) = (x^2 + 2x + 1)^3$. We can consider this as a composite of two functions: $y = x^2 + 2x + 1$ and $z = y^3$. Therefore, $z^{'} = 3y^2$ and $y^{'} = 2x + 2$. Thus, $h^{'}(x) = 3(x^2 + 2x + 1)^2(2x + 2)$.

There are several other theorems that are sometimes useful in economic analysis. I will state them without proof.

Theorem 5.5. (Mean Value Theorem): If f is a real-valued continuous function on $[a, b]$ and is differentiable on (a, b), then there exist a point $x \in [a, b]$ such that:

$$f^{'}(x) = \frac{f(b) - f(a)}{b - a}$$

Theorem 5.6. (Intermediate Value Theorem): Let f be a real-valued continuous function on $[a, b]$. IF $f(a) < f(b)$ and if c is a number such that $f(a) < c < f(b)$, then there exists a point $x \in (a, b)$ such that:

$$f(x) = c$$

This theorem is often used to prove the existence of a competitive equilibrium. For example, let p be the price of a good and $D(p)$, $S(p)$ be demand and supply for that good. If D and S are continuous functions and $D\left(p^{'}\right) - S\left(p^{'}\right) > 0$ for some $p^{'}$ and $D\left(p^{''}\right) - S\left(p^{''}\right) < 0$, then there exists a market clearing price $p \in \left(p^{'}, p^{''}\right)$ such that $D(p) - S(p) = 0$.

Definition 5.3. (Second Derivative): Let $X \subset \mathbb{R}$. Suppose that a function $f : X \to \mathbb{R}$ is differentiable on X. (Thus, the derivative function: $f^{'} : X \to \mathbb{R}$is well-defined). The second derivative of the function $f^{'}$at $x = x_0$ is given by:

$$\frac{df^{'}}{dx}(x_0) = f^{''}(x_0) = \lim_{h \to 0} \frac{f^{'}(x_0 + h) - f^{'}(x_0)}{h}$$

The function f is said to be twice-differentiable at x_0 if and only if this derivative exists. The higher-order derivatives are dened similarly:

Definition 5.4. (Higher-Order Derivative): Let $X \subset \mathbb{R}$. Suppose that a function $f : X \to \mathbb{R}$ is n-times differentiable on X. The $(n + 1)$-th derivative of the function $f(n)$ at $x = x0$ is given by:

$$\frac{df^{(n)}}{dx}(x_0) = f^{(n+1)}(x_0) = \lim_{h \to 0} \frac{f^{(n)}(x_0 + h) - f^{(n)}(x_0)}{h}$$

We also say that a function f is continuously differentiable if its first-derivative exists and is continuous. Continuous differentiability is not the same as continuous and differentiable functions. The latter simply means that they are continuous and differentiable. As we know, the higher-order derivative rules are exactly the same as the rst derivative rules. We denote a class of continuously differentiable functions as C^1, twice continuously differentiable functions as C^2, and so on.

Example 5.3. The first, second, and third derivative of the function $f(x) = x^n$ is:

$f^{'}(x) = nx^{n-1}$
$f^{''}(x) = n(n-1)x^{(n-1)-1} = n(n-1)x^{n-2}$
$f^{'''}(x) = n(n-1)(n-2)x^{(n-2)-1} = n(n-1)(n-2)x^{n-3}$

Theorem 5.7. *(Taylor). Let f be an n-times continuously diferentiable function defined on an interval $[a,b]$. Then for some $x \in [a,b]$ we have:*

$$f(b) = f(a) + f^{'}(a)(b-a) + \ldots + \frac{f^{(n)}(x)}{n!}(b-a)^n.$$

Proof. For fixed n, denfie the $(n-1)$th degree Taylor polynomial on $[a,b]$ by

$$P_{n-1}(t) = f(a) + f^{'}(a)(t-a) + \ldots + \frac{f^{(n-1)}(x)}{(n-1)!}(t-a)^{n-1}$$

and note that if we write $g_{n-1}(t) = P_{n-1}(t) - f(t)$ then $g_{n-1}^{(k)}(a) = 0$ for $k = 0, 1, \ldots n-1$. We now choose K such that if $g_n(t) = P_{n-1}(t) - f(t) + K(t-1)^n/n!$, then g_n has the properties:

$g_{n-1}^{(k)}(a) = 0$ for $k = 0, 1, \ldots, n-1$
$g_n(b) = 0$.

Applying Rolle's theorem n times we obtain an $x \in [a,b]$ such that $g_n^n(x) = 0$, which implies, $K = f^{(n)}(x)$ as desired. □

Note that we are not yet in a position to introduce analytic functions.

We now establish two basic properties of derivatives: the product rule and the chain rule. We derive these for one-dimensional functions only.

Proposition 5.1. *(Product rule). Let $f, g : [a,b] \longrightarrow \mathbb{R}$ be two real-valued dierentiable functions. Then the function $k(x) = f(x)g(x)$ is differentiable and $k^{'}(x) = f^{'}(x)g(x) + f(x)g^{'}(x)$.*

Proof. This is immediate from the definition of derivative and basic properties of limits:

$k^{'}(x) = \lim\limits_{h \longrightarrow 0} \frac{k(x+h)-k(x)}{h}$
$k^{'}(x) = \lim\limits_{h \longrightarrow 0} \frac{f(x+h)g(x+h)-f(x)g(x)}{h}$
$k^{'}(x) = \lim\limits_{h \longrightarrow 0} \frac{f(x+h)g(x+h)-f(x)g(x)}{h} + \lim\limits_{h \longrightarrow 0} \frac{f(x)g(x+h)+g(x)f(x+h)-f(x)g(x+h)-f-g(x)f(x+h)}{h}$
$k^{'}(x) = \lim\limits_{h \longrightarrow 0} \frac{f(x+h)g(x)-f(x)g(x)}{h} + \lim\limits_{h \longrightarrow 0} \frac{f(x)g(x+h)-f(x)g(x+h)+g(x+h)f(x+h)-g(x)f(x+h)}{h}$
$k^{'}(x) = f^{'}(x)g(x) + \lim\limits_{h \longrightarrow 0} \frac{g(x+h)-g(x)}{h} \lim\limits_{h \longrightarrow 0} f(x+h)$
$k^{'}(x) = f^{'}(x)g(x) + f(x)g^{'}(x)$
as desired. □

The following observation may be obtained from Proposition 5.1 by replacing g with its reciprocal.

Corollary 5.1. *(Quotient rule). Let $f, g : [a, b] \longrightarrow \mathbb{R}$ be two real-valued differentiable functions such that g vanishes nowhere on $[a, b]$. Then the function $k(x) = f(x)/g(x)$ is differentiable and*

$$k'(x) = \frac{f'(x)g(x)+f(x)g'(x)}{[g(x)]^2}$$

The following is sometimes referred to as l'Hopital's rule. It reinforces the idea that differentiable functions behave locally like affine functions and is very useful for evaluating limits. Perhaps first note the trivial case:

$$\lim_{x \longrightarrow a} \frac{Ax-Aa}{Bx-Ba} = \frac{A}{B} \lim_{x \longrightarrow a} \frac{x-a}{x-a} = \frac{A}{B}$$

Theorem 5.8. *(l'Hopital). Let f and g be dierentiable, real-valued functions on the interval (a, b) (where a and b might be infinite) and that $g'(x) \neq 0$ on (a, b). If*

$$\lim_{x \longrightarrow a} \frac{f'(x)}{g'(x)} = A$$

and either: $f(x)$, $g(x) \longrightarrow 0$

or $g(x) \longrightarrow \infty$ as $x \longrightarrow a$

then: $\lim_{x \longrightarrow a} \frac{f(x)}{g(x)} = A$

Remark 5.1. The above is stated asymmetrically, but this is only for convenience.

Proof. The above idea that "differentiable functions behave locally like affine functions" may be made precise using the mean value theorem. We simply treat by cases. First suppose that $A < +\infty$. Then for every pair q, r such that $A < r < q$ by $\lim_{x \longrightarrow a} \frac{f'(x)}{g'(x)}$ there exists $c > a$ such that $\frac{f'(x)}{g'(x)} < r$ for all $x \in (a, c)$. The crucial observation is that by the mean value theorem, for any $x, y \in (a, c)$ with $x < y$ there exists $t \in (x, y)$ such that:

$$\frac{f(x)-f(y)}{g(x)-g(y)} = \frac{f'(t)}{g'(t)} < r < q$$

Now, if $\lim_{x \longrightarrow a} f(x) = 0$ and $\lim_{x \longrightarrow a} g(x) = 0$ holds then we send $x \longrightarrow a$ and obtain:

$$\lim_{x \longrightarrow 0} \frac{f(x)-f(y)}{g(x)-g(y)} = \frac{-f(y)}{-g(y)} \leq r < q$$

Since $y \in (a, c)$ is arbitrary, we obtain $\lim_{x \longrightarrow a} \frac{f(x)}{g(x)} \leq r$ for every $r > A$. On the other hand, if $\lim_{x \longrightarrow a} g(x) = +\infty$ holds then for fixed y satisfying $\frac{f(x)-f(y)}{g(x)-g(y)} = \frac{f'(t)}{g'(t)} < r < q$ we know that for x close to a, $g(x) - g(y) > 0$, and so by multiplying by $[g(x) - g(y)]/g(x)$ we get:

$$\frac{f(x)}{g(x)} < r - r\frac{g(y)}{g(x)} + \frac{f(y)}{g(x)}$$

which shows that there exists $c_2 > a$ such that $f(x)/g(x) < q$ for all $x \in (a, c_2)$. This gives one side of theorem 5.3. The other side is similar. □

Example 5.4. The following are easy consequences of l'Hopital's rule:

$$\lim_{x \longrightarrow 0} \frac{\sin x}{x} = \lim_{x \longrightarrow 0} \frac{\cos x}{1} = 1$$

$$\lim_{\sigma \longrightarrow 0} \frac{x^{\sigma}-1}{\sigma} = \lim_{\sigma \longrightarrow 0} \frac{x^{\sigma} \ln x}{1} = lnx$$

Theorem 5.9. *(Inverse Function Theorem): Let $f : X \longrightarrow \mathbb{R}$ be a C^1 function. Suppose that $f^x(x) \neq 0$ for all x in some interval $I \subset X$, then:*

(i) well-defined inverse f^{-1} exists on I

(ii) f^{-1} is also continuously differentiable on $f(I)$

(iii) for all $y \in f(I)$, we have: $\left(f^{-1}\right)^{'}(y) = 1/f^{'}\left(f^{-1}(y)\right)$

5.1 Critical points, local minima, local maxima

Theorem 5.10. *Suppose that $f : X \longrightarrow \mathbb{R}$ is continuously differentiable at $x = x_0$. Then, the following are true:*

(i) if $f^{'}(x_0) > 0$, there exists a (small) open interval I such that f is (strictly) increasing on I

(ii) if $f^{'}(x_0) < 0$, there exists a (small) open interval I such that f is (strictly) decreasing on I.

Definition 5.5. (Monotonic functions): Suppose that $f : [a, b] \longrightarrow \mathbb{R}$. Then:

(i) f is said to be (weakly) increasing if for any $x, y \in [a, b]$,
$x < y \Longrightarrow f(x) \leq f(y)$

(ii) f is said to be (strictly) increasing if for any $x, y \in [a, b]$,
$x < y \Longrightarrow f(x) < f(y)$

(iii) f is said to be (weakly) decreasing if for any $x, y \in [a, b]$,
$x < y \Longrightarrow f(x) \geq f(y)$

(iv) f is said to be (strictly) decreasing if for any $x, y \in [a, b]$,
$x < y \Longrightarrow f(x) > f(y)$

Corollary 5.2. *Suppose that $f : [a, b] \longrightarrow \mathbb{R}$ is continuously differentiable on (a, b). Then:*

(i) if $f^{'}(x) > 0$ for all $x \in (a, b)$, f is (strictly) increasing on (a, b)

(ii) if $f^{'}(x) < 0$ for all $x \in (a, b)$, f is (strictly) decreasing on (a, b)

(iii) if f is (strictly) increasing on (a, b), $f^{'}(x) \geq 0$ for all $x \in (a, b)$

(iv) if f is (strictly) decreasing on (a, b), $f^{'}(x) \leq 0$ for all $x \in (a, b)$

Definition 5.6. (Critical Points) Suppose that $f : [a, b] \longrightarrow \mathbb{R}$ is continuously differentiable on (a, b). Then, a point $x \in (a, b)$ is called a critical point of f if $f^{'}(x) = 0$.

These critical points can be characterized into local max, local min, or degenerate critical (inflection) points.

Definition 5.7. (Maxima and minima): Let $f : X \longrightarrow \mathbb{R}$. Then:

(i) a point $x^* \in X$ is said to be a global maximum of f on X if $f(x^*) \geq f(x))$ for all $x \in X$

(ii) a point $x^* \in X$ is said to be a local maximum of f on X if there exists an open interval I with $x \subset I$ such that $f(x^*) \geq f(x))$ for all $x \in I \cap X$

For minima, simply change the direction of the inequality ($\leq$).

Theorem 5.11. *(Necessary Condition for Local Min/Max): Suppose that $f : [a, b] \longrightarrow \mathbb{R}$ is differentiable on (a, b). If f has a local maximum or minimum at a point $x \in (a, b)$, then $f^{'}(x) = 0$.*

Additionally, we also have the following sufficient conditions.

Theorem 5.12. *(Sufficient Conditions): Suppose that $f : [a, b] \longrightarrow \mathbb{R}$ is twice-differentiable on (a, b).*
(i) if $f^{'}(x) = 0$ and $f^{''}(x) < 0$, then x is a local maximum of f
(ii) if $f^{'}(x) = 0$ and $f^{''}(x) > 0$, then x is a local minimum of f
(iii) if $f^{'}(x) = 0$ and $f^{''} = 0$, then x may be either a local max, local min, or inflection point.

Example 5.5. We have:
(i) $f(x) = x^2$. Then, $f^{'}(x) = 2x$ and $f^{''}(x) = 2 > 0$. Therefore, $x = 0$ is a min.
(ii) $f(x) = x^3$. Then, $f^{'}(x) = 3x^2$ and $f^{''}(x) = 6x$. Therefore, $f^{''}(0) = 0$, so it is not a min or max, it is an inflection point.

Definition 5.8. (Global max/min): Suppose that $f : I \longrightarrow \mathbb{R}$ is twice-differentiable and I is a connected interval. Suppose further that:
(i) x is a local maximum of f
(ii) x is the only critical point of f on I. Then, x is the global maximum (minimum) of f.

Theorem 5.13. (Global max/min): Suppose that $f : [a, b] \rightarrow \mathbb{R}$ is twice-differentiable on (a, b).
(i) If $f^{''}$ is never zero on (a, b), then f has at most one critical point in (a, b).
(ii) This critical point is a global minimum if $f^{''} > 0$ and global maximum if $f^{''} < 0$.

Of course, a max or min of a function can occur at an endpoint of the domain. It is very important to pay attention to the domain of a function. The max/min of a function, whether global or local, is a concept relative to its domain. The maxima or minima at an endpoint of the domain is called boundary maxima or boundary minima.

Lastly, we have the following important theorem. It is this theorem that we are safe (at least in most economic applications) to assume the existence of a global maximum or minimum. This theorem is more general and applies to more than one variable cases.

Theorem 5.14. (Weierstrasss Theorem): Let $C \subset \mathbb{R}$ be a compact set (i.e. closed and bounded set). Let $f : C \rightarrow \mathbb{R}$ be a continuous function. Then, there exist points x_{min}, x_{max} such that*:*
(i) $f(x) \geq f(x_{min}) \, \forall x \in C$
(ii) $f(x) \geq f(x_{max}) \, \forall x \in C$
These points are obviously a global minimum and a global maximum, by definition.

What does this mean in one-variable case? A compact set in $\mathbb{R}$ is simply a closed interval of the form $[a, b]$. Thus, it states that on the closed interval $[a, b]$, there are always a global max and a global min, as long as the function is continuous. So, if f is also differentiable, then we have the simplest way of calculating a global max.

6 Calculus of Several Variables

Preliminaries

In economics, we often consider multivariate calculus, because in the real world we have many input and output goods. It is rarely the case that functions of our interest, such as utility functions or production functions, have only one variable as an argument. Thus, we often analyze functions $f : X \to \mathbb{R}$ where $X \subset \mathbb{R}^n$ instead of $X \subset \mathbb{R}$.

Example 6.1. (i) A consumer in the real world would be aware of finitely many goods, say n, for her possible consumption. She would choose a consumption bundle $x = (x_1, x_2, \ldots, x_n)$. Thus, she would solve a maximization problem:

$$max\, U(x_1, x_2, \ldots, x_n) \quad s.t \sum_{i=1}^{n} p_i x_i \le M$$

(ii) A production of a good y may require three (primary) inputs x_1, x_2, x_3:

$$y = Ax_1^{\alpha 1} x_2^{\alpha 2} x_3^{\alpha 3}$$

(iii) An aggregate demand may exhibit some substitution effects between two goods, so that:

$$x_i = A - p_1^{\alpha} + p_2^{\beta}$$

Typical functional forms for utility and production functions:
(i) Linear/log-linear functions: $u = ax_1 + bx_2$; $u = a\, logx_1 + b\, logx2$;
(ii) Cobb-Douglas functions: $u = Ax_1^{\alpha 1} x_2^{\alpha 2}$;
(iii) Leontif functions: $u = min\{x_1, x_2\}$;
(iv) Constant-elasticity-of-substitution functions: $u = A(c_1 x_1^{-\alpha} + c_2 x_2^{-\alpha})^{-b/a}$.
In addition, we may sometimes study the functions $f : X \to \mathbb{R}^m$where $X \subset \mathbb{R}^n$
.

$$f(x) = \begin{bmatrix} f_1(x_1, x_2, \ldots, x_n) \\ f_2(x_1, x_2, \ldots, x_n) \\ \vdots \\ f_m(x_1, x_2, \ldots, x_n) \end{bmatrix}$$

Example 6.2. Two different output goods y_1, y_2 may require inputs of three goods $x1, x2x3$ with different technologies:

$y_1 = ax_1 + bx_1 + cx_3$
$y_2 = A\sqrt{x_1 x_2 x_3}$

We may want to analyze them jointly, instead of separately one-by-one.

It is always a good idea to draw a graph of functions to understand them. However, it is impossible to represent functions of more than two arguments and sometimes drawing in a three- dimensional space does not help our intuition. So, economists often use two-dimensional diagrams to represent a function of two-variables.

Definition 6.1. (**Level Sets**): Suppose $f : X \to \mathbb{R}$ where $X \subset \mathbb{R}^n$. Then, the level set for a point $a \in f(x)$ is the set $L \subset X$ such that:
$$L(a) = \{x \in X : f(x) = a\}$$

Example 6.3. (i) Suppose $f(x, y) = x^2 + y^2$. Then, for $f(x, y) = 5$, lets nd the level set. By definition,
$$L(5) = \left\{(x, y) : x^2 + y^2 = 5\right\}$$

We know from the property of right-angled triangles, it is a circle in the (x, y)-plane whose radius is $\sqrt{5}$. If level sets can be drawn as curves, we sometimes call "level curves".

Definition 6.2. A function $f : X \subset \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is called a linear operator or linear transformation if f preserves the following properties:
(i) $\forall x, y \in X$,we have $f(x + y) = f(x) + f(y)$ and
(ii) $\forall x \in X$, $\forall r \in \mathbb{R}$ we have $f(rx) = rf(x)$

Theorem 6.1. *Let $f : X \subset \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be a linear operator. Then, there exists a $(m \times n)$-matrix A such that $f(x) = Ax$ for all $x \in X$*

Definition 6.3. (Quadratic Forms): A quadratic form on $\mathbb{R}^n$ is a real-valued function $Q : X \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ such that:
$$Q(x) = \sum_{i \le j \le n} a_{ij} x_i x_j$$

In matrix representation we have that: $Q(x) = x^t Ax$ where A is a unique symmetric matrix.

Now, we shall begin by considering real-valued function defined on an open subset of finite-dimensional Euclidean space. Let $U \subseteq \mathbb{R}^n$ and $f : U \longrightarrow \mathbb{R}$. For each $i = 1, \ldots, n$ we write $e_i = (0, \ldots, 1, \ldots 0)$ for the vector with 1 in the ith component and 0 elsewhere. In this notation each $d \in \mathbb{R}^n$ may be written:

$$d = d_1 e_1 + \ldots + d_n e_n$$

Definition 6.4. (Partial and directional derivatives). For each $i = 1, \ldots, n$ the ith partial derivative of f at a point $x \in U$ is defined to be:

$$\frac{\partial f}{\partial x_i}(x) = \lim_{t \longrightarrow 0^+} \frac{f(x + te_i) - f(x)}{t}$$

More generally, if $d \in \mathbb{R}^n$ is a vector with $\|d\| = 1$ we write:

$$f^{'}(x; d) = \lim_{t \longrightarrow 0^+} \frac{f(x + td) - f(x)}{t}$$

for the directional derivative of f at x in the direction d.

Once we leave the real line there are two distinct notions of dierentiability with which we are concerned.

Definition 6.5. (Gateaux dierentiability). A function $f : U \longrightarrow \mathbb{R}$ Gateaux differentiable at a point $x \in U$ if the directional derivatives $f^{'}(x; d)$ exist for all directions and $f^{'}(x; \cdot)$ is linear.

And also,

Definition 6.6. (Frechet differentiability). A function $f : U \longrightarrow \mathbb{R}$ is Frechet differentiable at a point $x \in U$ if there exists a linear map $l : \mathbb{R}^n \longrightarrow \mathbb{R}$ such that:

$$\lim_{h \longrightarrow 0^+} \frac{f(x+h)-f(x)-<l,h>}{t} = 0$$

Example 6.4. Consider the function $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ given by

$$f(x,y) = \frac{xy^2}{x^2+y^4}$$

if $(x,y) \neq (0,0)$, and $f(0,0) = 0$.

Show that f has directional derivatives in all directions at the origin. Show that f is not Gateaux differentiable at the origin. Show that f is not continuous at the origin, even though it is continuous when restricted to straight lines through the origin.

Substituting $d = (d_1, d_2)$ and $x = 0$ into the definition of directional derivative we evaluate the limit:

$$f'(0;d) = \lim_{t \longrightarrow 0^+} \frac{f(td)-f(0)}{t}$$

$$f'(0;d) = \lim_{t \longrightarrow 0^+} \frac{1}{t} \frac{t^3 d_1 d_2^2}{t^2 d_1^2 + t^4 d_2^4}$$

$$f'(0;d) = \lim_{t \longrightarrow 0^+} \frac{d_1 d_2^2}{d_1^2 + t^2 d_2^4} = \frac{d_2^2}{d_1}$$

which is clearly not linear in d, which gives the first two assertions. To establish the third, consider a sequence of points $((x_n, y_n))_{n=1}^{\infty}$ approaching the origin satisfying $x_n = y_n^2$ for all $n \geq 1$.

Example 6.5. Consider the function $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ given by

$$f(x,y) = \frac{2y\exp\left(-x^{-2}\right)}{\exp(-2x^{-2})+y^2}$$

if $(x,y) \neq (0,0)$, and $f(0,0) = 0$. Show that f is Gateaux differentiable at the origin, but not continuous there.

Proceeding as above we fix $d = (d_1, d_2)$ with $d_1 \neq 0$ and evaluate the limit:

$$f'(0;d) = \lim_{t \longrightarrow 0^+} \frac{f(td)-f(0)}{t}$$

$$f'(0;d) = \lim_{t \longrightarrow 0^+} \frac{1}{t} \frac{2td_2\exp\left(-t^{-2}d_1^{-2}\right)}{\exp\left(-2t^{-2}d_1^{-2}\right)+t^2d_2^2}$$

$$f'(0;d) = \lim_{t \longrightarrow 0^+} \frac{2d_2\exp\left(-t^{-2}d_1^{-2}\right)}{\exp\left(-2t^{-2}d_1^{-2}\right)+t^2d_2^2} = 0$$

which is linear in d. Note, however, that f is not continuous at the origin since it blows up if we approach the origin along the curve $y = \exp\left(-x^{-2}\right)$, along which the function is given by $\frac{2y^2}{2y^2} = 1$ which tends to a non-zero limit as $y \longrightarrow 0$.

Remark 6.1. Note that if the partial derivatives exist and are "well-behaved", then Gateaux differentiability implies Frechet. This requires a weak form of the mean value theorem.

Definition 6.7. (Second-Order Partial Derivative): Let $X \subset \mathbb{R}^n$. Suppose that the derivative function $f_{x_i} : X \longrightarrow \mathbb{R}$ is well-defined at a point x. Then, for each variable x_j, the second-order partial derivative at a point x is given by:

$$\frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)(x) = f_{x_i x_j}(x) = \lim_{t \longrightarrow 0} \frac{f_{x_i}(x+td) - f_{x_i}(x)}{t}$$

Higher-order partial derivatives are defined similarly.

Example 6.6. Let's compute the second-order partials of the following function:

$f(x,y) = 2x^3 + 5xy + 5y^2$.

First let's obtain the first-order partial derivatives:

$\frac{\partial f(x,y)}{\partial x} = 6x^2 + 5y$

$\frac{\partial f(x,y)}{\partial y} = 5x + 10y$

Thus, the second-order partials can be computed by applying the same rule again:

$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial\left(6x^2+5y\right)}{\partial x} = 12x$

$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial\left(6x^2+5y\right)}{\partial y} = 5$

$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial(5x+10y)}{\partial x} = 5$

$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial(5x+10y)}{\partial y} = 10$

These partial derivatives can be arrayed into an $(n \times n)$-matrix. The matrix is called the Hessian matrix of f and is written:

$$H = D^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \frac{\partial^2 f}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

Theorem 6.2. *(Young's Theorem): Suppose that $f : X \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ is twice-differentiable on an open set $U \subset X$. Then, for all $x \in U$, we have:*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x) \text{ for each } i, j = 1, \ldots, n$$

This implies that the Hessian of f is symmetric.

Definition 6.8. (Differentiability in $\mathbb{R}^n$): Let $X \subset \mathbb{R}^n$. A function $f : X \longrightarrow \mathbb{R}^n$ is differentiable at a point x if there exists a linear operator $A : \mathbb{R}^n \longrightarrow \mathbb{R}$ such that:

$$\lim_{h \longrightarrow 0} \frac{|f(x+h) - f(x) - Ah|}{\|h\|} = 0$$

We write $f^{'}(x) = A$. Note that A is a vector. We call $f^{'}(x)$ the total derivative or Jacobian derivative of f and write:

$$J = D^1 f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{pmatrix}$$

When we want to represent it as the column matrix and think of it as a vector, then we refer to it as the gradient vector of f and write:

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

When we have a function $F : \mathbb{R}^n \longrightarrow \mathbb{R}^m$:

$$F(x_1, \ldots, x_n) = \begin{pmatrix} f_1(x_1, \ldots, x_n) \\ f_2(x_1, \ldots, x_n) \\ \vdots \\ f_m(x_1, \ldots, x_n) \end{pmatrix}$$

its Jacobian matrix is:

$$J = D^1 F(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

Theorem 6.3. *(Differentiability): Suppose a function $f : X \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ is differentiable at a point x. Then, all of its partial derivatives $\frac{\partial f}{\partial x_i}$ exist at x, for $i = 1, \ldots, n$.*

Definition 6.9. (Total Differentials): Let $f : X \subset \mathbb{R}^n \longrightarrow \mathbb{R}$. The following approximation of a change in f is called the total differential of f:

$$df = \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2 + \ldots + \frac{\partial f}{\partial x_n} dx_n$$

Theorem 6.4. *(Taylor's Theorem in $\mathbb{R}^n$): Suppose a function $f : X \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ is twice continuously differentiable on an open set $U \subset X$ and its third total derivative exists on U. Then, for any point $a, b \in U$, we have:*

$$f(b) = f(a) + D^1 f(a)(b-a) + \frac{1}{2!}(b-a) D^2 f(a)(b-a) + O\left(\|b-a\|^2\right)$$

where $O(\cdot)$ indicates that this term goes to zero faster than $\|b-a\|^2$ (i.e. $\frac{O\left(\|b-a\|^2\right)}{\|b-a\|^2} \longrightarrow 0$ as $b \longrightarrow a$.

Example 6.7. Let's find the linear and the quadratic approximations around $x = 1$ for:

$$h(x) = \frac{x^\alpha - x^\beta}{x^\alpha + x^\beta}$$

where $\alpha > \beta > 0$.
The linear approximation around $x = 1$ is:

$$h(x) \approx h(1) + h^{'}(1)(x-1)$$

and
$h(1) = 0$
$h^{'}(x) = \frac{1}{(x^\alpha + x^\beta)}\left[\left(\alpha x^{\alpha-1} - \beta x^{\beta-1}\right)\left(x^\alpha + x^\beta\right) - \left(x^\alpha - x^\beta\right)\left(\alpha x^{\alpha-1} - \beta x^{\beta-1}\right)\right]$
$h^{'}(1) = \frac{\alpha-\beta}{4}$
Thus,

$$h(x) \approx \frac{\alpha-\beta}{4}(x-1)$$

Now, the quadratic approximation around $x = 1$ is:

$$h(x) \approx h(1) + h^{'}(1)(x-1)\frac{1}{2}h^{''}(1)(x-1)^2$$

and
$h^{''}(x) = -2(\alpha-\beta)x^{\alpha+\beta-1}\left(x^\alpha + x^\beta\right)^{-3} + (\alpha-\beta)\left(x^\alpha + x^\beta\right)^{-2}(\alpha+\beta-1)x^{\alpha+\beta-2}$
$h^{''}(1) = \frac{1}{4}(\alpha-\beta)(\alpha+\beta-1)$
Thus,

$$h(x) \approx \frac{\alpha-\beta}{4}(x-1) + \frac{1}{8}(\alpha-\beta)(\alpha+\beta-1)(x-1)^2$$

6.1 Implicit Functions

Consider the function

$$y = f(x_1, \ldots, x_n)$$

we can rewrite this function as follows:

$$g(x_1, \ldots, x_n, y) = y - f(x_1, \ldots, x_n) = 0$$

Clearly , we know that g maps from $\mathbb{R}^{n+1}$ to $\mathbb{R}$ whereas f maps from $\mathbb{R}^n$ to $\mathbb{R}$.

We now move on to the implicit function theorem. Informally, the implicit function theorem states that if, at a zero of a function f on a domain $U \times V \subseteq \mathbb{R}^n \times \mathbb{R}^m$, the partial derivative with respect to the second m variables is non-singular, then that variable may be locally written as a dierentiable function of the first n variables, and gives an explicit formula for the derivative.

Theorem 6.5. *(Implicit function). Let U and V be open subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, and $f : U \times V \longrightarrow \mathbb{R}$ a continuously differentiable function. If for some $(x_0, y_0) \in U \times V$, $D_y f(x_0, y_0)$ is non-singular and $f(x_0, y_0) = 0$ then there exist open sets $U_0 \subseteq U$ and $V_0 \subseteq V$ such that for some continuously differentiable function $y : U_0 \longrightarrow V_0$ such that for all $(x, y) \in U_0 \times V_0$ we have $f(x_0, y_0) = 0$ if and only if $y = y(x)$. Further, we have the explicit formula:*

$$Dy(x_0) = -[D_y f(x_0, y_0)]^{-1} D_x f(x_0, y_0)$$

for the derivative of y at x_0.

Proof. For convenience we shall assume throughout that $(x_0, y_0) = (0, 0)$. First note that by continuity DF_y remains invertible within a given neighbourhood of (x_0, y_0). Further (this is a bit stronger), note that by the continuity of the partial derivatives of f at $(x_0, y_0) = (0, 0)$ there exist neighbourhoods $U_0 \subset U$ and $V_0 \subset V$ such that for all $x \in U_0$ and $y_1, \ldots, y_n \in V_0$ the matrix:

$$\begin{bmatrix} \nabla f_1(x, y_1) \\ \vdots \\ \nabla f_m(x, y_m) \end{bmatrix}$$

is invertible. Note that this matrix is not simply the Jacobian of $f(x_0, \cdot)$ at $y \in V_0$, since we allow each of the rows to vary independently (this is crucial for our application of the mean value theorem). Now we may show that for each $x \in U_0$ there is at most one $y \in V_0$ satisfying $f(x, y) = 0$, for in particular if for some $x \in U_0$ there exists $y, z \in V_0$ with $f(x, y) = f(x, z)$ then by the weak mean-value theorem for multivariable functions for each $i = 1, \ldots, m$ there exists y_i on the line segment (y, z) such that:

$$0 = f_i(x, y) = \langle \nabla f_i(x, y_i), y - z \rangle$$

which implies that $y = z$ since the matrix is invertible (note that the extra generality mentioned above is necessary since there is no reason to think that the y_i will be equal).

Now, define the function $F : U_0 \times V_0 \longrightarrow \mathbb{R}$ by

$$F(x, y) = \sum_{i=1}^{m} [f_i(x, y)]^2$$

It is clear that there exists $r > 0$ such that $B_r(0) \subseteq V_0$ and a neighbourhood $U_1 \subseteq U_0$ such that for all $x \in U_1$ the function $F(x_0, \cdot)$ achieves a minimum in the interior of $B_r(0)$. At such a point, denoted $y(x)$, we have by the first order conditions for optimality that:

$$DF_y(x, y(x)) = 2D_y f(x, y(x)) f(x, y(x)) = 0 \in \mathbb{R}^m$$

where we have used the multivariable chain rule. Since $D_y f(x, y(x))$ is invertible, we find $f(x, y(x)) = 0$. It remains to show that y is continuous, then differentiable, and derive its derivative.

We now simply let $\triangle x$ denote a small change in x, $\triangle y = y(x + \triangle x)$ denote the induced change in y, and use the multivariable Taylor's theorem to write:

$$0 = f(x + \triangle x, y(x + \triangle x)) - f(x, y(x))$$

$$0 = Df_x(x + y(x)) \triangle x + Df_y(x + y(x)) \triangle y + o((\triangle x, \triangle y))$$

Rearranging and sending x to zero, we obtain the expression of interest. □

An easier version for a single variable case can be stated:

Theorem 6.6. *(Implicit Function Theorem): Let $g : X \subset \mathbb{R}^2 \longrightarrow \mathbb{R}$ be continuously differentiable function in the neighborhood U of (x_0, y_0). Suppose $g(x_0, y_0) = c$ and $g_y(x_0, y_0) \neq 0$. Then, there exists a C^1 function $y = f(x)$ defined on U such that:*
(i) $g(x, f(x)) = c$ for all $x \in U$
(ii) $y_0 = f(x_0)$
(iii) $\frac{dy}{dx}(x_0) = -\frac{g_x(x_0,y_0)}{g_y(x_0,y_0)}$

Example 6.8. Consider the function $g(x, y) = x^2 - \alpha xy + y^4$. Find $\frac{dy}{dx}$ at the point $(3, 3)$ for $\alpha = 3$.
First,
$g_y(3, 3) = -\alpha x + 3y^2 = -9 + 27 = 18 \neq 0$
$g_x(3, 3) = 2x - \alpha y = 6 - 9 = -3$
Using the implicit function theorem:
$\frac{dy}{dx}(3, 3) = -\frac{g_x(3,3)}{g_y(3,3)} = -\frac{(-3)}{18} = \frac{1}{6}$

7 Linear Algebra

A system of linear equations:

$$x - 2y = 8$$

$$3x + y = 3$$

Can be rewritten in a matrix form as:

$$\begin{pmatrix} 1 & -2 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 8 \\ 3 \end{pmatrix}$$

In general, we can represent a system of linear equations as a matrix form:

$$\underset{(m\times n)}{A} \underset{(n\times 1)}{x} = \underset{(m\times 1)}{b}$$

The matrix A is called the coefficient matrix of the system. We often create the augmented matrix $\hat{A}$ by adding the column vector corresponding to the right-hand side.

$$\hat{A} = (A|b) = \begin{pmatrix} 1 & -2 & | & 8 \\ 3 & 1 & | & 3 \end{pmatrix}$$

7.1 Elementary Row Operations

The following elementary row operations do not change the property of a linear system (i.e. it would result in the equivalent system):
(i) Interchange two rows of a matrix:

$$\begin{pmatrix} 1 & -2 & | & 8 \\ 3 & 1 & | & 3 \end{pmatrix} \Longleftrightarrow \begin{pmatrix} 3 & 1 & | & 3 \\ 1 & -2 & | & 8 \end{pmatrix}$$

(ii) Adding one row to another:

$$\begin{pmatrix} 1 & -2 & | & 8 \\ 3 & 1 & | & 3 \end{pmatrix} \Longleftrightarrow \begin{pmatrix} 1+3 & -2+1 & | & 8+3 \\ 3 & 1 & | & 3 \end{pmatrix}$$

(iii) Multiply through each row by a nonzero number:

$$\begin{pmatrix} 1 & -2 & | & 8 \\ 3 & 1 & | & 3 \end{pmatrix} \Longleftrightarrow \begin{pmatrix} a & -2a & | & 8a \\ 3 & 1 & | & 3 \end{pmatrix}$$

7.2 Row Echelon Form

The purpose of performing row operations is to create a matrix of the following form:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & | & b_1 \\ 0 & a_{22} & a_{23} & | & b_2 \\ 0 & 0 & a_{33} & | & b_3 \end{pmatrix}$$

This matrix form is called row echelon form.

Definition 7.1. (Leading Zero & Row Echelon Form): A row of a matrix is said to have k leading zeros if the first k elements of the row are all zeros and the $(k+1)$-th element is nonzero. A matrix is said to be in row echelon form if each row of the matrix has more leading zeros than the row preceding it.

7.3 Rank and Solutions to A Linear System

Definition 7.2. The rank of a matrix A is the number of nonzero rows in its (reduced) echelon form. We write $rank(A)$.

You probably wonder if the echelon form is unique. In fact, the echelon form is not unique, it can have many di�erent entries in the echelon form. However, the number of nonzero rows in its echelon form is unique and does not depend on how we compute the echelon form.

Definition 7.3. (Reduced Echelon Form): A matrix is said to be in reduced echelon form if each row of the matrix in its echelon form has one in its pivot position and each column containing the pivot has no other nonzero entries.

Example 7.1. The following matrix is in reduced echelon form:

$$\begin{pmatrix} 1 & a & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & a \\ 0 & 0 & 0 & 1 & a \end{pmatrix}$$

Example 7.2. Consider creating a reduced echelon form of the following:

$$\begin{pmatrix} 0 & 2 \\ 5 & 3 \end{pmatrix} \overset{add\ 2nd\ to\ 1st}{\Longleftrightarrow} \begin{pmatrix} 5 & 5 \\ 5 & 3 \end{pmatrix} \overset{substract\ 1st\ from\ 2nd}{\Longleftrightarrow} \begin{pmatrix} 5 & 5 \\ 0 & -2 \end{pmatrix}$$

$$\overset{divide\ 1st\ by\ 5}{\Longleftrightarrow} \begin{pmatrix} 1 & 1 \\ 0 & -2 \end{pmatrix} \overset{divide\ 2nd\ by\ -2}{\Longleftrightarrow} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \overset{divide\ 1st\ by\ 5}{\Longleftrightarrow} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Therefore, the rank of the matrix is 2.

Fact 7.1. *Let A be a $(m \times n)$-matrix. Then,*
(i) $rank\,(A) \leq m$, the number of rows of A
(ii) $rank\,(A) \leq n$, the number of columns of A.

The corollary to this fact is the following: Let A be a $(m \times n)$-matrix and B be a $(m \times l)$-matrix. Then:

$$rank\,(AB) = \min\{rank\,(A)\,, rank\,(B)\}$$

Definition 7.4. (Homogenous System): An equation with a constant term = 0 is called a homogenous equation. A system of linear equations $Ax = b$ with b =0 is said to be a homogenous system.

Theorem 7.1. *Consider a system of linear equations of the form: $Ax = b$*
(a) When $m < n$ (i.e. the number of equations is less than the number of variables):
(i) For every b, $Ax = b$ has either 0 or infinitely many solutions.
(ii) If $rank(A) = m$, then $Ax = b$ has innitely many solutions for every b.
(b) When $m > n$ (i.e. the number of equations is more than the number of variables):
(i) $Ax = 0$ has one or innitely many solutions.
(ii) For every b, $Ax = b$ has either $0, 1,$ or innitely many solutions.
(iii) If $rank(A) = n$, then $Ax = b$ has 0 or 1 solution for every b.
(c) When $m = n$ (i.e. the number of equations is the same as that of variables):
(i) $Ax = 0$ has one or innitely many solutions.
(ii) For every b, $Ax = b$ has either $0, 1,$ or innitely many solutions.
(iii) If $rank(A) = n = m$, then $Ax = b$ has exactly 1 solution for every b.

You can construct examples to check why all these cases.

Theorem 7.2. *(Linear Implicit Function Theorem): Consider a system of linear equations, $Ax = b$ with $m < n$. Consider a partition of unknown variables:*

$$x = \begin{pmatrix} x^{'} \\ x^{''} \end{pmatrix} = \begin{matrix} \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix} & \textit{Endogenous} \\ \begin{bmatrix} x_{k+1} \\ \vdots \\ x_n \end{bmatrix} & \textit{Exogenous} \end{matrix}$$

Then, the system has a unique solution for every choice of $x^{''} \in \mathbb{R}^{n-k}$ if and only if (i) $k = m$ and (ii) the corresponding coefficient matrix (for the endogenous variables):

$$rank \begin{pmatrix} a_{11} & \dots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{k1} & \dots & a_{kk} \end{pmatrix} = k = m$$

7.4 Basic Rules of Matrix Algebra

Let's consider a $(m \times n)$-matrix, we have an array of data assorted in m rows and n columns. It is common to index each entry of the data, as follows:

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}$$

7.4.1 Addition and Subtraction

Addition and subtraction of matrices is defined only when they are of the same size. We simply add and subtract element-by-element.

$$A = \begin{pmatrix} 2 & 3 \\ 1 & 1 \\ 0 & 2 \end{pmatrix}; B = \begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 5 & 2 \end{pmatrix}$$

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

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

7.4.2 Scalar Multiplication

For any $r \in \mathbb{R}$ and any $A \in M_{m \times n}(\mathbb{R})$, we can dene a scalar multiplication:

$$rA = r \begin{pmatrix} 2 & 3 \\ 1 & 1 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 2r & 3r \\ r & r \\ 0 & 2r \end{pmatrix}$$

7.4.3 Matrix Multiplication

We can define the matrix product AB if and only if the number of columns of A is equal to the number of rows of B. That is, the product is dened for any combination of k, m, n such that:

$$\underset{(k \times m)}{A} \underset{(m \times n)}{B}$$

So, we have:

$$\begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} aA + bC & aB + bD \\ cA + dC & cB + dD \\ eA + fC & eB + fD \end{pmatrix}$$

Because we have a collection of (i, j)-elements in the resulting matrix, where the index i coming from the first matrix and j coming from the second matrix, we have the $(k \times n)$-matrix as a result of matrix multiplication.

$$\underset{(k\times m)}{A}\underset{(m\times n)}{B} = \underset{(k\times n)}{C}$$

In general, the (i, j)-element of the resulting matrix is written as:

$$c_{ij} = \begin{pmatrix} a_{i1} & \cdots & a_{im} \end{pmatrix} \begin{pmatrix} b_{1j} \\ \vdots \\ b_{mj} \end{pmatrix} = a_{i1}b_{1j} + \ldots + a_{1m}b_{mj} = \sum_{h=1}^{m} a_{1h}b_{hj}$$

7.4.4 Laws of Matrix Algebra

We have the following laws whenever these operations are well-defined:
(i) Associative Law: $(A + B) + C = A + (B + C)$ and $(AB)C = A(BC)$
(ii) Commutative Law of Addition: $A + B = B + A$
(iii) Distributive Law: $A(B + C) = AB + AC$ and $(A + B)C = AC + BC$
These are exactly the same laws as we had in case of real numbers. But, there is one important law which real numbers satisfy but matrices dont. It is the commutative law for multiplication.

7.5 Important Matrices

Definition 7.5. (Identify Matrix): The identity matrix is an $(n \times n)$-matrix with entries satisfying:

$$a_{ij} = \begin{cases} 1 & if \quad i = j \\ 0 & if \quad i \neq j \end{cases}$$

Thus,

$$I = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{pmatrix}$$

The identify matrix has the following property: For any $(k \times n)$-matrix A: $AI = A$ and for any $(n \times k)$-matrix B, $IB = B$. Therefore, if A is a $(n \times n)$-matrix, we have that: $AI = IA = A$.

Definition 7.6. (Transpose): The transpose of a $(m \times n)$-matrix A is the $(n \times m)$-matrix obtained by interchanging the rows and columns of A. That is, the transpose, denoted $A^\intercal$, is a matrix such that: $a'_{ji} = a_{ij}$, where a_{ij} is the (i, j)-element of A and a'_{ji} is the (j, i)-element of $A^\intercal$.

Theorem 7.3. *(Transpose Rules): Let A, B be arbitrary matrices. We have the following rules whenever the operations are well-defined:*
(i) $(A \pm B)^\intercal = A^\intercal \pm B^\intercal$
(ii) $(A^\intercal)^\intercal = A$
(iii) $(rA)^\intercal = rA^\intercal$
(iv) $(AB)^\intercal = B^\intercal A^\intercal$

Definition 7.7. (Square, Diagonal, Triangular Matrices): Any $(n \times n)$-matrix is called a square matrix. Diagonal matrix is a square matrix in which all non-diagonal entries are zero. Upper-triangular matrix is a matrix (usually square) in which all entries below the diagonal are zero. Lower-triangular matrix is a matrix (usually square) in which all entries above the diagonal are zero.

$$D = \begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix} U = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix} L = \begin{pmatrix} a_{11} & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

Definition 7.8. (Symmetric matrix): Symmetric matrix is a square matrix A such that $A = A^\intercal$.

$$S = \begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix}$$

Definition 7.9. (Idempotent matrix): Idempotent matrix is a square matrix A such that $AA = A$.

$$MM = \begin{pmatrix} 5 & -5 \\ 4 & -4 \end{pmatrix} \begin{pmatrix} 5 & -5 \\ 4 & -4 \end{pmatrix} = \begin{pmatrix} 5 & -5 \\ 4 & -4 \end{pmatrix}$$

Definition 7.10. (Permutation matrix): A square matrix of zeros and ones in which each row and each column contains exactly one 1. It is called permutationmatrix, because it permutes entries of a matrix.

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Definition 7.11. (Nonsingular matrix): Nonsingular matrix is a square matrix of full rank (i.e. $rank(A) = n$ when A is a $(n \times n)$-matrix).

Now let'ss define an important matrix. Formally, an inverse element of an real number a is defined as a number b such that: $ab = ba = 1$. Now in terms of matrices, an appropriate definition would be: $AB = BA = I$.

Definition 7.12. (Inverse Matrices): Let A be a $(n \times n)$ square matrix. The matrix $B \in M_{n\times n}(\mathbb{R})$ is said to be an inverse of A if $AB = BA = I_{n\times n}$.

When the inverse exists, we say that A is invertible. We write $B = A^{-1}$.

Theorem 7.4. *(Uniqueness of Inverse): Any square matrix A can have at most one inverse.*

Proof. Suppose that B and C are both inverses of A. Then, $C = CI = C(AB) = (CA)B = IB = B$. So, B must be necessarily equal to C. □

Theorem 7.5. *(Equivalence): For any square matrix A, the following statements are equivalent:*
(i) A is invertible
(ii) Every system of linear equations $Ax = b$ has a unique solution for all $b \in \mathbb{R}^n$
(iii) A is nonsingular
(iv) A has a full rank.

You can check this properties.

Theorem 7.6. *Let A and B be invertible square matrices. Then:*
(i) $\left(A^{-1}\right)^{-1} = A$
(ii) $\left(A^{\intercal}\right)^{-1} = \left(A^{-1}\right)^{\intercal}$
(iii) AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$

Proof. (i) is obvious. To see (ii), postmultiply both sides by $A^{\intercal}$:
$LHS = \left(A^{\intercal}\right)^{-1} A^{\intercal} = I$
$RHS = \left(A^{-1}\right)^{\intercal} A^{\intercal} = \left(AA^{-1}\right)^{\intercal} = I^{\intercal} = I$
To see (iii), by denition, if we find a matrix C such that:
$C\,(AB) = (AB)\,C = I$
then C is an inverse of AB and we can write $C = (AB)^{-1}$.
Let $C = B^{-1}A^{-1}$. Then, we have:
$B^{-1}A^{-1}(AB) = (AB)\,B^{-1}A^{-1} = I$ □

Definition 7.13. (Partitioned Matrix): Any matrix A can be partitioned into submatrices.

For example, a(4×6) -matrix can be partitioned into:

$$A = \left(\begin{array}{ccc|c|cc} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} \end{array}\right)$$

which can be written as a (2×3)-matrix of submatrices:

$$A = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{pmatrix}$$

This is called a partitioned matrix of A.
Finally, let's discuss how to actually compute an inverse of a matrix A.

7.6 Determinants

Definition 7.14. (Nonsingular matrix): A square matrix is nonsingular if and only if its determinant is nonzero.

Now, from the equivalence theorem, we know that A is invertible if and only if its determinant is nonzero. So, lets work backward to and determinants from this theorem. Suppose A is a scalar, i.e. $A = a \in M_1(\mathbb{R})$. We know that a is invertible (i.e. $\frac{1}{a}$ exists) if and only if $a \neq 0$. So, from this, we define that:

$$\det(a) = a$$

Definition 7.15. (Minor, Cofactor): Let A be an $(n \times n)$-matrix. Let $\hat{A}_{ij}$ be the $(n-1) \times (n-1)$-submatrix obtained by deleting row i and column j from A. Then, the scalar $M_{ij} = \det\left(\hat{A}_{ij}\right)$is called the (i,j)-th minor of A. The sign-adjusted scalar $C_{ij} = (-1)^{i+j} M_{ij} = (-1)^{i+j} \det\left(\hat{A}_{ij}\right)$ is called the (i,j)-th cofactor of A.

Give these definitions, we have that:

Definition 7.16. (Determinant): The determinant of a $(n \times n)$-matrix is a scalar given by:

$\det(A) = a_{11}C_{11} + a_{12}C_{12} + \ldots + a_{1n}C_{1n}$

$\det(A) = a_{11}M_{11} - a_{12}M_{12} + \ldots + (-1)^{1+n} a_{1n}M_{1n}$

$\det(A) = a_{11}\det\left(\hat{A}_{11}\right) - a_{12}\det\left(\hat{A}_{12}\right) + \ldots + (-1)^{1+n} a_{1n}\det\left(\hat{A}_{1n}\right)$

As you probably know, there are many ways of computing a determinant. This denition uses the first row, but we can use any row or any collumn $k = 1, 2, \ldots, n$ compute it:

$\det(A) = a_{k1}C_{k1} + a_{k2}C_{k2} + \ldots + a_{kn}C_{kn}$ for any $k = 1, 2, \ldots, n$

or

$\det(A) = a_{1k}C_{1k} + a_{2k}C_{2k} + \ldots + a_{nk}C_{nk}$ for any $k = 1, 2, \ldots, n$

Theorem 7.7. *The determinant of a lower-triangular, upper-triangular, or diagonal matrix is simply the product of its diagonal entries. That is:*

$$\det\begin{pmatrix} a_{11} & 0 & \ldots & 0 \\ 0 & a_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a_{nn} \end{pmatrix} = \det\begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ 0 & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a_{nn} \end{pmatrix} =$$

$$\det\begin{pmatrix} a_{11} & 0 & \ldots & 0 \\ a_{21} & a_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{pmatrix} = a_{11}a_{22}\ldots a_{nn}$$

Proof. First, note that it is enough to prove it for upper-triangular and lower-triangular matrices, because diagonal matrices are a special case of triangular matrices. To see how the proof works, let's just for simplicity illustrate it for a (3×3)-matrix. Lets compute:

$$\det \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix} = a_{11} \det \begin{pmatrix} a_{22} & a_{23} \\ 0 & a_{33} \end{pmatrix} - 0 \det \begin{pmatrix} a_{12} & a_{13} \\ 0 & a_{33} \end{pmatrix} +$$

$$0 \det \begin{pmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{pmatrix} = a_{11} \left(a_{22} a_{33} - 0 a_{23} \right) = a_{11} a_{22} a_{33}$$

We can easily see that this logic works for any $(n \times n))$ upper-triangular matrix. For a lower- triangular matrix, pick the first row, instead of the first column, to compute its determinant. In order to extend the proof for an $(n \times n)$, apply mathematical induction. □

Theorem 7.8. *For any* $A, B \in M_{n \times n}(\mathbb{R})$

(i) $\det(A^\intercal) = \det(A)$

(ii) $\det(Ab) = \det(A)\det(B)$

(iii) $\det(A + B) \neq \det(A) + \det(B)$*, in general*

We need one more important definition.

Definition 7.17. (Adjoint): Recall that, for any $(n \times n)$-matrix A, the (i, j)-th cofactor of A is $C_{ij} = (-1)^{i+j} \det \left(\hat{A}_{ij} \right)$ where $\hat{A}_{ij}$ is obtained by deleting the row i and the row j of A. Then, the adjoint of A is defined to be:

$$adj(A) = \begin{pmatrix} C_{11} & C_{12} & \dots & C_{1n} \\ C_{12} & C_{22} & \dots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & \dots & C_{nn} \end{pmatrix}$$

That is, $adj(A)$ is a $(n \times n)$-matrix whose (i, j)-element is C_{ji}, the (j, i)-th cofactor of A.

7.7 Inverse and Cramer's Rule

Suppose that A is invertible. Create an augmented matrix $\hat{A}$ such that:

$$\hat{A} = \left(\underset{n \times n}{A} \,|\, \underset{n \times n}{I} \right)$$

If there exists an inverse, then we can premultiply this by the inverse to get:

$$A^{-1} \hat{A} = A^{-1} \left(\underset{n \times n}{A} \,|\, \underset{n \times n}{I} \right) = \left(\underset{n \times n}{I} \,|\, \underset{n \times n}{A^{-1}} \right)$$

This means that, if we can find a matrix that conduct elementary row operations such that it convert A to I, we can find an inverse of A. In actual computation, we apply elementary row operations consecutively.

We also have the following theorems:

Theorem 7.9. *(Inverse): Let* $A \in M_{n \times n}(\mathbb{R})$ *be a nonsingular matrix. Then*

$$A^{-1} = \frac{1}{\det(A)} adj\,(A)$$

Theorem 7.10. *(Cramers Rule): Let $A \in M_{n\times n}(\mathbb{R})$ be a nonsingular matrix. The unique solution x of the system $Ax = b$ is given by:,*

$$x_i = \frac{\det(B_i)}{\det(A)}$$

where B_i is the matrix A with the right-hand side b replacing the i-th column of A, i.e. for a (3×3)-matrix, B_2 is:

$$B_2 = \begin{pmatrix} a_{11} & b_1 & a_{13} \\ a_{21} & b_2 & a_{23} \\ a_{31} & b_3 & a_{33} \end{pmatrix}$$

Proof. Lets prove the Cramers rule for a (3×3)-matrix. Suppose the system is:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

The solution to this system is obviously: $x = A^{-1}b$. So:

$$x = \frac{1}{\det(A)} adj\,(A) \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

$$x = \frac{1}{\det(A)} \begin{pmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

$$x = \frac{1}{\det(A)} \begin{pmatrix} C_{11}b_1 + C_{21}b_2 + C_{31}b_3 \\ C_{12}b_1 + C_{22}b_2 + C_{32}b_3 \\ C_{13}b_1 + C_{32}b_2 + C_{33}b_3 \end{pmatrix}$$

Now, let $i = 1$, then:

$$x_i = \frac{1}{\det(A)} \left(C_{11}b_1 + C_{21}b_2 + C_{31}b_3\right)$$

$$x_i = \frac{1}{\det(A)} \sum_j b_j \,(-1)^{i+j} \det\left(\hat{A}_{j1}\right)$$

$$x_i = \frac{1}{\det(A)} \begin{pmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{pmatrix}$$

This can be repeated for all $i = 1, 2, 3$. □

7.8 Euclidean Spaces

Definition 7.18. (Euclidean Space): An n-dimensional Euclidean space is the Cartesian product of n real spaces, denoted by $\mathbb{R}^n = \underbrace{\mathbb{R} \times \mathbb{R} \ldots \times \mathbb{R}}_{n\ times}$

$$\mathbb{R}^n = \{(x_1, \ldots, x_n) : \ x_i \in \mathbb{R}, \forall i = 1, 2, \ldots, n\}$$

So, $x = (x_1, \ldots, x_n)$ represents a point in n-dimensional Cartesian plane. This n-tuple may be more generally interpreted as a displacement in $\mathbb{R}^n$. The Euclidean space is often termed a normed vector space, because it is a vector space endowed with a metric, called norm.

Definition 7.19. (Vector Space): A vector space is any set V such that addition and scalar multiplication are well-defined on V and satisfy the following properties:

(i) (Associative law of addition): $\forall x, y, z \in V$, $x + (y + z) = (x + y) + z$

(ii) (Neutral element for addition): $\exists 0 \in V$, $\forall x \in V$, $x + 0 = 0 + x = x$

(iii) (Inverse element for addition): $\forall x \in V$, $\exists -x \in V$, $x + (-x) = 0$

(iv) (Associative law of scalar multiplication): $\forall \alpha, \beta \in \mathbb{R}$, $\forall x \in V$,$\alpha(\beta x) = (\alpha\beta)x = (\beta\alpha)x = \beta(\alpha x)$

(v) (Neutral element for scalar multiplication): $\forall x \in V$, $\exists 1 \in V$, $1x = x1 = x$

(vi) (Distributive law of addition): $\forall \alpha \in \mathbb{R}$, $\forall x, y \in V$, $\alpha(x + y) = \alpha x + \alpha y$;

(vii) (Distributive law of scalar multiplication): $\forall \alpha, \beta \in \mathbb{R}$, $\forall x \in V$, $(\alpha + \beta)x = \alpha x + \beta x$

As we know, addition and scalar multiplication are well-defined and satisfy these properties in $\mathbb{R}^n$. So, the Euclidean space is a vector space. In addition to addition and scalar multiplication, the Euclidean space is endowed with another operation, called the (Euclidean) inner product. We often denote $x \cdot y$ or $\langle x, y \rangle$.

Definition 7.20. (Inner Product): Let $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ be two vectors in $\mathbb{R}^n$. The (Euclidean) inner product of x and y is the number such that:

$$\langle x, y \rangle = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n$$

This inner product satisfy the following properties:

(i) (Symmetry): $x \cdot y = y \cdot x$

(ii) (Linearity): $x \cdot (y + z) = x \cdot y + x \cdot z$ and $x \cdot (\alpha y) = \alpha\,(x \cdot y) = (\alpha x) \cdot y$

(iii) (Positivity): $x \cdot x \geq 0$ with "=" iff $x = 0$

Definition 7.21. (Norm): The norm of a vector x in $\mathbb{R}^n$ is defined as:

$$\|x\| = (x \cdot x)^{\frac{1}{2}} = \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}}$$

Theorem 7.11. *(Properties of Norm): Suppose $x, y, z \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$. Then,*

(i)$\|x\| \geq 0$ *with "$= 0$" iff* $x = 0$

(ii) $\|\alpha x\| = |\alpha|\|x\|$

(iii) $\|x \cdot y\| = \|x\| \cdot \|y\|$

(iv) $\|x + y\| \leq \|x\| + \|y\|$

(v) $\|x - z\| \leq \|x - y\| + \|y - z\|$

(vi)$\|x\| - \|y\| \leq \|x - y\|$

A line in $\mathbb{R}^2$ and a plane in $\mathbb{R}^3$ are examples of sets of points described by a single linear equation. These sets are called hyperplanes.

Definition 7.22. (Hyperplane): In general, a hyperplane in $\mathbb{R}^n$ is a set of points that has $(n-1)$ dimensions in $\mathbb{R}^n$ and are described by a linear equation:

$$a_1 x_1 + a_2 x_2 + \ldots + a_n x_n = c$$

Thus, hyperplanes in $\mathbb{R}^1$, $\mathbb{R}^2$, $\mathbb{R}^3$ are, respectively, a point, a line, and a plane.

$$a_1x_1 = c$$

$$a_1x_1 + a_2x_2 = c$$

$$a_1x_1 + a_2x_2 + a_3x_3 = c$$

7.9 Linear Dependence and Independence

Before formally dening linear dependence, lets consider a simple example. Suppose that we have a relationshipbetween x_1 and x_2 such that:

$$a_1x_1 + a_2x_2 = 0$$

Now, suppose that $a_1 \neq 0$. Then, we can manipulate this relationship:

$$x_1 = -\tfrac{a_2}{a_1}x_2$$

So, there is natural dependency of x_1 on x_2. So, we say that x_1 is linearly dependent on x_2, because the relationship is linear.

Recall that the set of all scalar multiples of a non-zero vector v is a straight line through the origin.

Definition 7.23. (Span): For a vector $v \in \mathbb{R}^n$, the set $L(v)$ is said to be spanned or generated by v if:

$$L(v) = \{rv : r \in \mathbb{R}\}$$

Definition 7.24. (Linear Combination): A linear combination of k non-zero vectors $v_1, v_2, \ldots, v_k$ is:

$$c_1v_1 + c_2v_2 + \ldots + c_kv_k \text{ for some scalars } c_1, c_2, \ldots, c_k \in \mathbb{R}$$

Definition 7.25. (Span): Let $v_1, v_2, \ldots, v_k$ be non-zero vectors. We say that a set $L(v_1, v_2, \ldots, v_k)$ is generated or spanned by $v_1, v_2, \ldots, v_k$ if:

$$L(v_1, v_2, \ldots, v_k) = \{c_1v_1 + c_2v_2 + \ldots + c_kv_k : c_1, c_2, \ldots, c_k \in \mathbb{R}\}$$

That is, $L(v_1, v_2, \ldots, v_k)$ is a set of all possible linear combinations of $v_1, v_2, \ldots, v_k$. Moreover, if a set V is a subset of $L(v_1, v_2, \ldots, v_k)$, then we say that $v_1, v_2, \ldots, v_k$ spans V.

Definition 7.26. (Linear Dependence): Vectors $v_1, v_2, \ldots, v_k \in \mathbb{R}^n$ are linearly dependent if and only if there exist scalars $c_1, c_2, \ldots, c_k \in \mathbb{R}$, not all equal to zero, such that:

$$c_1v_1 + c_2v_2 + \ldots + c_kv_k = 0$$

Definition 7.27. (Linear Independence): Vectors $v_1, v_2, \ldots, v_k \in \mathbb{R}^n$ are linearly independent if and only if:

$$c_1 v_1 + c_2 v_2 + \ldots + c_k v_k = 0 \Longrightarrow c_1 = c_2 = \ldots, c_k = 0$$

Theorem 7.12. *Vectors $v_1, v_2, \ldots, v_k \in \mathbb{R}^n$ are linearly dependent if and only if the linear system:*

$$A \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix} = \begin{pmatrix} v_1 & v_2 & \ldots & v_k \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix} = 0$$

has a non-zero solution $c = (c_1, c_2, \ldots, c_k)$.

When we have $v_1, v_2, \ldots, v_k \in \mathbb{R}^n$, $A = (v_1, v_2, \ldots, v_n)$ becomes a $(n \times n)$-matrix. We can use the equivalence theorem: a square matrix is of full rank if and only if its determinant is not zero.

Theorem 7.13. *A set of n vectors $v_1, v_2, \ldots, v_k \in \mathbb{R}^n$ is linearly independent if and only if:*

$$\det \begin{pmatrix} v_1 & v_2 & \ldots & v_n \end{pmatrix} \neq 0$$

Another important fact is that, if we have more vectors than the number of dimensions of each vector v_i, then they must be linearly dependent.

Theorem 7.14. *Any set of k vectors $v_1, v_2, \ldots, v_k \in \mathbb{R}^n$ is linearly dependent if $k > n$.*

If $v_1, v_2, \ldots, v_k$ are linearly independent, no one of these vectors is a linear combination of the others and therefore, no proper subset of $v_1, v_2, \ldots, v_k$ would span V . This is the concept of a basis of a vector space V.

Definition 7.28. (Basis): Let $w_1, w_2, \ldots w_m$ be a collection of vectors in V. Then, $w_1, w_2, \ldots w_m$ form a basis of V if and only if:
(i) $w_1, w_2, \ldots w_m$ span V
(ii) $w_1, w_2, \ldots w_m$ are linearly independent.

Clearly, three non-zero vectors in R^2 cannot form a basis of R^2, because these vectors must be linearly dependent. Even if we have two non-zero vectors in R^2, they cannot be a basis of R^2 if $w_1 = a w_2$. Moreover, two vectors with one of them being a zero vector cannot be a basis of R^2, because it cannot span R^2. Moreover, note that a zero vector in R^2 is a linear combination of another non-zero vector, $0 = 0v$. The natural basis of the Euclidean space R^n is a canonical basis:

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

Finally, let's state the following theorems:

Theorem 7.15. *If both $v_1, v_2, \ldots, v_n$ and $w_1, w_2, \ldots w_m$ are a basis of V, then we must have that $n = m$.*

Also,

Theorem 7.16. *Every basis of R^n contains n vectors.*

And,

Theorem 7.17. *Let $v_1, v_2, \ldots, v_n$ be a collection of n vectors in R^n. Then, the following statements are equivalent:*
(i) $v_1, v_2, \ldots, v_n$ are linearly independent
(ii) $v_1, v_2, \ldots, v_n$ spans $\mathbb{R}^n$
(iii) $v_1, v_2, \ldots, v_n$ form a basis of $\mathbb{R}^n$
(iv) $\det(v_1, v_2, \ldots, v_n) \neq 0$
(v) $A = (v_1, v_2, \ldots, v_n)$has full rank

Definition 7.29. (Dimension): A dimension of a vector space V is the number of vectors that can form a basis of V .
Thus, the dimension of $\mathbb{R}^n$ is exactly n.

7.10 Quadratic Forms and Definite Matrices

Definition 7.30. (Quadratic form): A quadratic form on $\mathbb{R}^n$ is a real-valued function, $Q : \mathbb{R}^n \longrightarrow \mathbb{R}$ such that:

$$Q(x) = \sum_{i \leq j} a_{ij} x_i x_j$$

Recall that we can represent every quadratic form in a matrix form, using a symmetric matrix A:

$$Q(x) = x^\intercal A x$$

For example, on $\mathbb{R}^2$:

$$a_{11}x_1^2 + a_{12}x_1x_2 + a_{22}x_2^2 = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} a_{11} & \frac{1}{2}a_{12} \\ \frac{1}{2}a_{12} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Definition 7.31. (Definiteness): Let A be an $(n \times n)$ symmetric matrix. Then, A is said to be:
(i) positive definite iff $x^\intercal A x > 0$ for all $x \neq 0$ in $\mathbb{R}^n$
(ii) positive semidefinite iff $x^\intercal A x \geq 0$ for all $x \neq 0$ in $\mathbb{R}^n$
(iii) negative definite iff $x^\intercal A x < 0$ for all $x \neq 0$ in $\mathbb{R}^n$
(iv) negative semidefinite iff $x^\intercal A x \leq 0$ for all $x \neq 0$ in $\mathbb{R}^n$
(v) indefinite iff $x^\intercal A x > 0$ for some x and $x^\intercal A x < 0$ for some x in $\mathbb{R}^n$

Note that if A is positive (negative) definite, then A is positive (negative) semidefinite.

Definition 7.32. (Principal Submatrix and Minor): Let A be a $(n \times n)$ matrix. A $(k \times k)$ submatrix of A obtained by deleting $n-k$ k columns, $i_1, \ldots, i_{n-k}$, and $n-k$ rows of the same indices $i_1, \ldots, i_{n-k}$ is called a k-th order principal submatrix of A. The determinant of the $(k \times k)$ principal submatrix is called a k-th order principal minor of A.

To find the definiteness of a matrix, however, we only use one special principal minor, called a leading principal minor.

Definition 7.33. (Leading Principal Submatrix and Minor): Let A be a $(n \times n)$ matrix. The k-th order principal submatrix of A obtained by deleting the last $n-k$ rows and the last $n-k$ columns is called the k-th order leading principal submatrix of A. We denote it by A_k. Its determinant is called the k-th order leading principal minor of A.

Theorem 7.18. *(Definiteness): Let A be a $(n \times n)$ symmetric matrix. Then:*

(i) A is positive definite if and only if all its leading principal minors are strictly positive (>0)

(ii) A is positive semidefinite if and only if all its principal minors are nonnegative (≥ 0)

(iii) A is negative definite if and only if all its leading principal minors alternate sings as follows: $\det(A_1) < 0$, $\det(A_2) > 0$, $\det(A_3) < 0, \ldots$ *, etc*

(iv) A is negative semidefinite if and only if all its principal minors of odd order are ≤ 0 and of even order are ≥ 0

(v) If some k-th order leading principal minor of A is nonzero but the sign pattern of nonzero terms does not fit either case (i) or (iii), then A is indefinite.

Remark 7.1. Note that for positive or negative definiteness of A, we only need to check their leading principal minors. But, if A is neither positive (negative) definite nor indefinite, then we must check all of its principal minors. Furthermore, it is important to note that when some of the leading principal minors are zero, it may not be indefinite and can be positive or negative semidefinite if the sign pattern of its nonzero terms still obeys the patterns of (i) or (iii).

Example 7.3. Consider a (4×4)matrix:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}$$

(a)$\det(A_1) > 0$, $\det(A_2) > 0$,$\det(A_3) > 0$,$\det(A_4) > 0 \Longrightarrow A$ is positive definite.

(a)$\det(A_1) < 0$, $\det(A_2) > 0$,$\det(A_3) < 0$,$\det(A_4) > 0 \Longrightarrow A$ is negative definite.

(a)$\det(A_1) > 0$, $\det(A_2) < 0$,$\det(A_3) > 0$,$\det(A_4) < 0 \Longrightarrow A$ is indefinite.

(a)$\det(A_1) > 0$, $\det(A_2) > 0$,$\det(A_3) = 0$,$\det(A_4) > 0 \Longrightarrow A$ is not positive definite, but may be positive semidefinite. If it is not positive semidefinite, then it is indefinite.

(a)det $(A_1) = 0$, det $(A_2) < 0$,det $(A_3) > 0$,det $(A_4) > 0 \Longrightarrow A$ is indefinite, because of det (A_2).

If a symmetric matrix is a diagonal matrix, then it becomes very easy. Recall that the determinant of a diagonal matrix is simply a product of the diagonal terms. So, its leading principal minors are:

det $(A_1) = a_{11}$, det $(A_2) = a_{11}a_{22}$,. . .,det $(A_n) = a_{11}a_{22} \ldots a_{nn}$

Remark 7.2. We have that $x = 0$ is a unique solution to $\max Q(x) = x^\intercal Ax$ if A is N.D. and to $\min Q(x) = x^\intercal Ax$ if A is P.D.

We often (in optimization) have a linear constraint of the form:

$$\max Q(x)$$

$$s.t.\ c \cdot x = 0 \text{ where } c = (c_1, \ldots, c_n)$$

In this case, we need to have "A is negative definite on the constraint set$\{x : c \cdot x = 0\}$". To check this, we form a bordered matrix:

$$H_{n+1} = \begin{pmatrix} 0 & c \\ c^\intercal & A \end{pmatrix}$$

Theorem 7.19. *Consider a bordered matrix. Suppose that $c_1 \neq 0$.*

(i) If the last n leading principal minors of H_{n+1} has the same sign, then the quadratic form Q is positive definite on the constraint set $\{x : c \cdot x = 0\}$ (so that $x = 0$ is a unique global minimum).

(ii) If the last n leading principal minors of H_{n+1} alternate signs, then the quadratic form Q is negative denite on the constraint set $\{x : c \cdot x = 0\}$ (so that $x = 0$ is a unique global maximum).

This result can be generalized for more than one constraint.

8 Optimization

8.1 Unconstrained Optimization

Now, we are going to understand how to deal with unconstrained optimization. This type of optimization problem is important, because corresponds to the interior solutions of any (constrained or unconstrained) optimization problem.

Theorem 8.1. *(FONC for Local Max/Min): Let $F : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ be a C^1 function. Suppose that $x^* \in U$ is a local maximum or minimum of F in U. Suppose that:*

(i) U is open; or

(ii) $x \in int\,(U)$. Then,

$$D^1F(x^*) = 0^\intercal \text{ i.e. } \frac{\partial F}{\partial x_i}(x^*) = 0 \text{ for all } i$$

It is important to keep in mind that (i) this condition is a necessary condition ($D^1F(x^*) = 0^\intercal$ does not imply that x is a local min or max. It simply means that x is a critical point of F) and (ii) this necessary condition only works for interior points.

Now, recall that the second order derivative of F can be summarized as the Hessian of F:

$$H = D^2F = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \frac{\partial^2 f}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

Theorem 8.2. *(SOSC for Local Max/Min): Let $F : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ be a C^2 function. Suppose that U is open and that $D^1F(x^*) = 0^\intercal$.*

(i) If the Hessian $D^2F(x^*)$ is negative definite, then x is a (strict) local maximum of F

(ii) If the Hessian $D^2F(x^*)$ is positive definite, then x is a (strict) local minimum of F

(iii) If the Hessian $D^2F(x^*)$ is indefinite, then x is neither local max nor minimum of F and it is called a saddle point of F.

Example 8.1. Lets work through a concrete problem. Consider a function:

$$F(x, y) = x^3 - y^3 + 9xy$$

Let's find a critical point first, using FONC:

$$F_x(x, y) = 3x^2 - 9y = 0$$

$$F_y(x, y) = -3y^2 + 9x = 0$$

Thus, the solution to this system is: $(x, y) = (0, 0)$ or $(x, y) = (3, -3)$. Now, to determine which one is a local maximum or minimum, lets check the Hessian:

$$H = \begin{pmatrix} F_{xx} & F_{xy} \\ F_{yx} & F_{yy} \end{pmatrix} = \begin{pmatrix} 6x & 9 \\ 9 & -6y \end{pmatrix}$$

At $(x, y) = (0, 0)$:

$$H = \begin{pmatrix} 0 & 9 \\ 9 & 0 \end{pmatrix}$$

To check the definiteness of the Hessian, compute the leading principal minors:

$$\det(H_1) = \det(0) = 0$$

$$\det(H_2) = \det \begin{pmatrix} 0 & 9 \\ 9 & 0 \end{pmatrix} = -81$$

This is indefinite. So, $(x, y) = (0, 0)$ is neither a local maximum or minimum. It is a saddle point.
At $(x, y) = (3, -3)$:

$$H = \begin{pmatrix} 18 & 9 \\ 9 & 18 \end{pmatrix}$$

To check the definiteness of the Hessian, compute the leading principal minors:

$$\det(H_1) = \det(18) > 0$$

$$\det(H_2) = \det\begin{pmatrix} 18 & 9 \\ 9 & 18 \end{pmatrix} = 243 > 0$$

Thus, it is positive definite, $(x, y) = (3, -3)$ is a local minimum.

It is important to realize that the above second-order conditions are sufficient conditions for local minima and maxima. However, its converse is not true. That is, local minima (maxima) do not imply the positive (negative) definiteness of Hessian. So, we need SONC.

Theorem 8.3. *(SONC for Local Max/Min): Let $F : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ be a C^2 function.*

(i) If $x \in int(U)$ is a local maximum of F, then $D^2F(x^*)$ is negative semidefinite

(ii) If $x \in int(U)$ is a local minimum of F, then $D^2F(x^*)$ is positive semidefinite.

Now, let's state the conditions for global maxima and minima.

Theorem 8.4. *(Sufficiency for Global Min/Max): Let $F : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ be a C^2 function and U is a convex open subset of $\mathbb{R}^n$.*

(a) The following three conditions are equivalent:

(i) F is a concave function on U

(ii) $F(y) - F(x) \leq D^1F(x)(y - x)$ for all $x, y \in U$

(iii) $D^2F(x)$ is negative semidenite for all $x \in U$

(b) The following three conditions are equivalent:

(i) F is a convex function on U

(ii) $F(y) - F(x) \geq D^1F(x)(y - x)$ for all $x, y \in U$

(iii) $D^2F(x)$ is positive semidenite for all $x \in U$

(c) If F is a concave function on U and $D^1F(x^*) = 0$ for some $x^* \in U$, then x^* is a global maximum of F on U

(d) If F is a convex function on U and $D^1F(x^*) = 0$ for some $x^* \in U$, then x^* is a global minimum of F on U

It is important to note that:
For local maxima:
$\left[H(x^*)\ is\ ND\ and\ D^1F(x^*) = 0\right] \implies [x^*\ is\ a\ local\ maximum]$
$[x^*\ is\ a\ local\ maximum] \implies \left[H(x^*)\ is\ ND\ and\ D^1F(x^*) = 0\right]$
For global maxima:
$\left[H(x)\ is\ NSD\ for\ all\ x\ and\ D^1F(x^*) = 0\right] \implies [x^*\ is\ a\ global\ maximum]$

Theorem 8.5. *(Uniqueness of Global Min/Max): Let $F : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ be a C^2 function and U is a convex open subset of $\mathbb{R}^n$.*

(a) If F is a strictly concave function on U and $D^1F(x^*) = 0$ for some $x^* \in U$, then x is a unique global maximum of F on U.

(b) If F is a strictly convex function on U and $D^1F(x^*) = 0$ for some $x^* \in U$, then x is a unique global minimum of F on U.

To check strict concavity (convexity) of a function, we have the following characterization theorem:

Theorem 8.6. *Let $F : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ be a C^2 function and U is a convex open subset of $\mathbb{R}^n$.*

(a) F is a strictly concave function on U if $D^2F(x)$ is negative definite for all $x \in U$

(b) F is a strictly convex function on U if $D^2F(x)$ is positive definite for all $x \in U$

Note that the converse may not be true.

8.2 Constrained Optimization

In this part we go further and our goal is to learn how to deal with optimization problems of the following form:

$$\max_{x\in\mathbb{R}^n} F(x,\theta)$$

$$\text{subject to } g_j(x,\theta) \leq b_k \text{ for } k = 1,\ldots,m$$

where $F(x,\theta)$, $g_1(x,\theta),\ldots,g_m(x,\theta)$ are functions defined on $\mathbb{R}^n$ or on an open subset of $\mathbb{R}^n$.

Theorem 8.7. *(FONC for Equality Constraints): Consider the maximization or minimization problem of the form:*

$$\max_{x\in\mathbb{R}^n} F(x_1,\ldots,x_n)$$

$$\textit{subject to} \quad \begin{matrix} h_1(x_1,\ldots,x_n) = c_1 \\ \vdots \\ h_m(x_1,\ldots,x_n) = c_m \end{matrix}$$

Suppose that F, $h_1,\ldots,h_m$ are C^1 functions and that $x^* = (x_1^*,\ldots,x_n^*)$ is a local maximum (or minimum) of F on the constraint set. If x^* is not a critical point of $h = (h_1,\ldots,h_m)$, then there exists a vector of Lagrange multipliers $\lambda^* = (\lambda_1^*,\ldots,\lambda_m^*)$ such that (x^*,λ^*) is a critical point of the Lagrangian function:

$$\mathcal{L}(x,\lambda) = F(x) + \lambda[c-h]$$

$$\mathcal{L}(x,\lambda) = F(x) + \lambda_1[c_1 - h_1(x)] + \ldots + \lambda_m[c_m - h_m(x)]$$

Definition 8.1. (Critical Point): Let $h : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be a C^1 function. A point x^* is said to be a critical point of h if and only if the rank of its Jacobian matrix, $rank\left(D^1h(x^*)\right)$ is $< m$.

Theorem 8.8. *(SOSC for Constrained Local Max/Min): Let F, $h_1, \ldots, h_m$ be C^2 functions on $\mathbb{R}^n$. Consider the problem:*

Consider the maximization or minimization problem of the form:

$$\max_{x \in \mathbb{R}^n} F(x_1, \ldots, x_n)$$

$$\text{subject to} \quad \begin{array}{c} h_1(x_1, \ldots, x_n) = c_1 \\ \vdots \\ h_m(x_1, \ldots, x_n) = c_m \end{array}$$

Form a Lagrangian as usual, $\mathcal{L}(x, \lambda) = F(x) + \lambda [c - h]$, and let (x^*, λ^*) satisfy FONCs. Suppose that "the Hessian of $\mathcal{L}$ w.r.t. x at (x^*, λ^*), $D_x^2 \mathcal{L}(x^*, \lambda^*)$, is negative definite on the linear constraint set $\{v : D^1 h(x^*) v = 0\}$", then x^* is a strict local constrained maximum of F.

Remark 8.1. This condition means that the following bordered Hessian matrix:

$$H_{m+n} = \begin{pmatrix} 0 & D^1 h(x^*) \\ D^1 h(x^*) & D_x^2 \mathcal{L}(x^*, \lambda^*) \end{pmatrix}$$

$$H_{m+n} = \begin{pmatrix} 0 & \cdots & 0 & \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \frac{\partial h_m}{\partial x_1} & \cdots & \frac{\partial h_m}{\partial x_n} \\ \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_m}{\partial x_1} & \frac{\partial^2 \mathcal{L}}{\partial x_1^2} & \cdots & \frac{\partial^2 \mathcal{L}}{\partial x_n \partial x_n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_1}{\partial x_n} & \cdots & \frac{\partial h_m}{\partial x_n} & \frac{\partial^2 \mathcal{L}}{\partial x_1 \partial x_n} & \cdots & \frac{\partial^2 \mathcal{L}}{\partial x_n^2} \end{pmatrix}$$

satisfies two conditions: (i) the last $(n - m)$ leading principal minors alternate in sign, and (ii) $det(H) > 0$ if n is even or $det(H) < 0$ if n is odd.

There are several related theorems.

Theorem 8.9. *(Kuhn-Tucker Theorem): Consider the maximization problem of the form:*

$$\max_{x \in \mathbb{R}^n} F(x_1, \ldots, x_n)$$

$$\text{subject to} \quad \begin{array}{c} g_1(x_1, \ldots, x_n) \leq b1 \\ \vdots \\ g_m(x_1, \ldots, x_n) \leq b_m \\ x_1 \geq 0 \\ \vdots \\ x_n \geq 0 \end{array}$$

Suppose that F, $g_1, \ldots, g_m$ are C^1 functions and that $x^* = (x_1^*, \ldots, x_n^*)$ is a local maximum of F on the constraint set. Suppose that the modified Jacobian matrix:

$$\left(\frac{\partial g_j}{\partial x_i} \right)_{ji}$$

has maximal rank, where the j's run over the indices of the g_j that are binding at x^*, and the i's range over the indices i for which $x_i > 0$. Then, *there exists a vector of Lagrange multipliers such that, for the Kuhn-Tucker Lagrangian defined by:*

$$\mathcal{L}(x,\lambda) = F(x) + \lambda[b-g]$$
$$\mathcal{L}(x,\lambda) = F(x) + \lambda_1[b_1 - g_1(x)] + \ldots + \lambda_m[b_m - g_m(x)]$$

we have:

(i) $\frac{\partial \mathcal{L}(x^*,\lambda^*)}{\partial x_i} \leq 0,\ x_i \geq 0\ ,x_i\left[\frac{\partial \mathcal{L}(x^*,\lambda^*)}{\partial x_i}\right] = 0$ *for all* $i = 1,\ldots,n$

(ii) $g_j(x^*) \leq b_j,\ \lambda_j^* \geq 0,\ \lambda_j^*[b_j - g_j(x^*)] = 0$ *for all* $i = 1,\ldots,m$

So, we have n slackness conditions for all non-negativity constraints and m slackness conditions for all regular constraints.

Remark 8.2. (Constraint Qualification Conditions):

(i) Kuhn-Tucker original – don't touch it.

(ii) g_j concave for all j, and Slater's condition, that is, there is some $x^* \geq 0$ with$g_j(x^*) > 0$ for all j.

(iii) rank condition (see Takayama 1.D.4, or Varian, ch 27)

(iv) g_j linear for all j, (Arrow-Hurwicz-Uzawa, see Takayama 1.D.4)

The most standard theorem is:

Theorem 8.10. *Suppose that F, $g_1,\ldots,g_m$ are all concave functions. If $x^* \geq 0$ and $\lambda^* \geq 0$ satisfy K-T conditions, then x^* is a solution to the constrained maximization problem.*

A better theorem is due to Arrow and Enthoven (1961).

Theorem 8.11. *Suppose that F, $g_1,\ldots,g_m$ are all quasi-concave functions and some "mild" condition holds. If $x^* \geq 0$ and $\lambda^* \geq 0$ satisfy K-T conditions, then x^* is a solution to the constrained maximization problem.*

Remark 8.3. The extra ("mild") condition is not needed if F is concave (and $g_1,\ldots,g_m$ are quasi-concave). See Takayama 1.E for three versions of the condition. Quasiconcavity (and therefore also concavity) of functions g_j implies that the constraint set, i.e. the set of $x \geq 0$ satisfying $g_1(x) \geq 0,\ldots,g_m(x) \geq 0$, is convex.

Example 8.2. Let's consider the following constrained maximization problem:

$$\max_{x,y} \ln(x+1) + \ln(y+1)$$

$$s.t\ p_1x + p_2y \leq m$$

$$x \geq 0,\ y \geq 0$$

where $p_1 > 0$, $p_2 > 0$ and $m > 0$.

In order to obtain the solution (as a function of the primitive parameters p_1, p_2 and m) we need to construct the Kuhn-Tucker first-order conditions as:

$$\mathcal{L}(x,\lambda) = \ln(x+1) + \ln(y+1) + \lambda[m - p_1x - p_2y]$$

$$(1)\ \frac{\partial\mathcal{L}(x^*,\lambda^*)}{\partial x} = \frac{1}{x^*+1} - \lambda^* p_1 \leq 0,\ x^* \geq 0,\ x^* \left[\frac{\partial\mathcal{L}(x^*,\lambda^*)}{\partial x}\right] = 0$$

$$(2)\ \frac{\partial\mathcal{L}(x^*,\lambda^*)}{\partial y} = \frac{1}{y^*+1} - \lambda^* p_2 \leq 0,\ y^* \geq 0,\ y^* \left[\frac{\partial\mathcal{L}(x^*,\lambda^*)}{\partial y}\right] = 0$$

$$(3)\ p_1 x^* + p_2 y^* \leq m,\ \lambda^* \geq 0,\ \lambda^* \left[m - p_1 x^* + p_2 y^*\right] = 0$$

Note that (3) holds with equality since it follows from (1) that $\lambda^* > 0$.

Case1: $x^* > 0$, $y^* > 0$

Then (1) and (2) hold with equalities. Solving (1), (2) and (3) we find $x^* = \frac{m+p_1-p_2}{2p_1}$ and $y^* = \frac{m+p_1-p_2}{2p_2}$ and $\lambda^* = \frac{2}{m+p_1+p_2}$. For x^* and y^* to be strictly positive, it has to be that $m + p_2 > p_1$ and $m + p_1 > p_2$. Thus Case 1 applies with x^* and y^* as listed above if $m + p_2 > p_1$ and $m + p_1 > p_2$ holds.

Case2: $x^* > 0$, $y^* = 0$

(3) implies that $x^* = \frac{m}{p_1}$. Since (1) holds with equality, we solve it for $\lambda^* = \frac{1}{m+p_1}$. Next we need to verify inequality (2). It states:

$$1 - \frac{p_2}{m+p_1} \leq 0$$

and it holds if $p_2 \geq m + p_1$. Thus Case 2 applies (with $x^* = \frac{m}{p_1}$ and $y^* = 0$) if $p_2 \geq m + p_1$ holds.

Case3: $x^* = 0$, $y^* > 0$

This case is very similar to Case 2. From (3) and (2) we obtain $y^* = \frac{m}{p_2}$, $\lambda^* = \frac{1}{m+p_2}$. Verifying inequality (1), we obtain $p_1 \geq m + p_2$. Thus Case 3 applies (with $x^* = 0$ and $y^* = \frac{m}{p_2}$) if $p_{21} \geq m + p_2$ holds.

The case $x^* = y^* = 0$ cannot hold since it violates equation (3). This concludes our solution to the K-T conditions.

Note that since utility function is concave and the constraint function is concave (in fact, it is linear) K-T conditions are sufficient. Hence, the solution to K-T conditions is a constrained maximizer. Further, since the Slater's condition holds, every constrained maximizer has to satisfy K-T conditions.

8.3 Quasiconcave and Quasiconvex Functions

Let"s define what it means for a function to be quasiconcave/quasiconvex:

Definition 8.2. (Quasiconcave and Quasiconvex Functions):

(i) A function $F : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ is quasiconcave on U, where U is a convex set in $\mathbb{R}^n$, if and only if for all $x, y \in U$ and $\lambda \in [0, 1]$, we have:

$$F(x) \geq F(y) \Longrightarrow F(\lambda x + (1-\lambda) y) \geq F(y)$$

(ii) A function $F : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ is quasiconvex on U, where U is a convex set in $\mathbb{R}^n$, if and only if for all $x, y \in U$ and $\lambda \in [0, 1]$, we have:

$$F(x) \leq F(y) \Longrightarrow F(\lambda x + (1-\lambda) y) \leq F(y)$$

(iii) A function $F : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ is strictly quasiconcave on U, where U is a convex set in $\mathbb{R}^n$, if and only if for all $x \neq y \in U$ and $\lambda \in (0,1)$, we have:

$$F(x) \geq F(y) \Longrightarrow F(\lambda x + (1-\lambda) y) > F(y)$$

(iv) A function $F : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ is strictly quasiconvex on U, where U is a convex set in $\mathbb{R}^n$, if and only if for all $x \neq y \in U$ and $\lambda \in (0,1)$, we have:

$$F(x) \leq F(y) \Longrightarrow F(\lambda x + (1-\lambda) y) < F(y)$$

Theorem 8.12. *(Properties of Quasiconcave Functions): Let $F : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ be defined on a convex set U. Then, the following are equivalent:*

(a) F is quasiconcave on U

(b) For every real number a, the upper contour set dened by:

$$C^+(a) = \{x \in U : F(x \geq a)\}$$

is a convex set in U.

(c) For all $x, y \in U$ and $\lambda \in [0,1]$, we have:

$$F(\lambda x + (1-\lambda) y) \geq \min\{F(x), F(y)\}$$

(d) $-F$ is quasiconvex on U.

8.4 Comparative Statics and Envelope Theorem

By comparative statics, we mean the sensitivity of (i) the optimal value of the objective and (ii) the optimal value of the decision variables, with respect to changes in primitive parameters of the problem.

Theorem 8.13. *(Shadow Price): Consider the following maximization problem:*

$$\max_{x \in \mathbb{R}^n} F(x_1, \dots, x_n)$$

$$\text{subject to} \quad \begin{aligned} g_1(x_1, \dots, x_n) &\leq b1 \\ &\vdots \\ g_k(x_1, \dots, x_n) &\leq b_k \\ h_1(x_1, \dots, x_n) &= c_1 \\ &\vdots \\ g_m(x_1, \dots, x_n) &= c_m \end{aligned}$$

Suppose that F, $g_1, \dots, g_k$,$h_1, \dots, h_m$ are C^1 functions and that $x^* = (x_1^*, \dots, x_n^*)$ is a local maximum of F on the constraint set. Let $\lambda^*(b,c) = (\lambda_1^*, \dots, \lambda_k^*)$, $\mu^*(b,c) = (\mu_1^*, \dots, \mu_m^*)$ be the corresponding Lagrange multipliers for the Lagrangian:

$$\mathcal{L}(x, \lambda) = F(x) + \lambda[b - g] + \mu[c - h]$$

Suppose that $x^*(b,c)$, $\lambda^*(b,c)$, $\mu^*(b,c)$ are differentiable w.r.t. (b,c) and that relevant Constrain Qualification (CQ) holds at $x^*(b,c)$,. Then,

$\lambda_j^*(b,c) = \frac{\partial F(x^*(b,c))}{\partial b_j}$ *for all $j = 1, \dots, k$*

$\mu_l^*(b,c) = \frac{\partial F(x^*(b,c))}{\partial c_l}$ *for all $l = 1, \dots, m$*

Thus, a Lagrange multiplier j measures the effect of a marginal change in input j on the objective value. In this view, economists often call j the shadow price (or imputed value) of input j.

Theorem 8.14. *(Envelope Theorem I): Let $F : \mathbb{R}^n \times \mathbb{R}^m \longrightarrow \mathbb{R}$ be a continuously differentiable function. For each fixed vector of parameters $\theta \in \mathbb{R}^m$, consider the maximization problem:*

$$\max_{x\in\mathbb{R}^n} F(x,\theta)$$

Let $x^*(\theta)$ be the solution and $v(\theta) = \max_{\theta\in\mathbb{R}^n} F(x,\theta) = F(x^*(\theta),\theta)$ be the value function of this problem. If $x^*(\theta)$ is a C^1 function, then:

$$\frac{\partial v(\theta)}{\partial \theta_j} = \left.\frac{\partial F(x,\theta)}{\partial \theta_j}\right|_{x=x^*(\theta)}$$

Proof. We can prove this using Chain Rule and total differential.

$$\frac{\partial v(\theta)}{\partial \theta_j} = \left.\frac{\partial F(x,\theta)}{\partial \theta_j}\right|_{x=x^*(\theta)}$$

$$\frac{\partial v(\theta)}{\partial \theta_j} = \left.\sum_{i=1}^{n} \frac{\partial F(x,\theta)}{\partial x_i}\frac{\partial x_i^*(\theta)}{\partial \theta_j} + \frac{\partial F(x,\theta)}{\partial \theta_j}\right|_{x=x^*(\theta)}$$

$$\frac{\partial v(\theta)}{\partial \theta_j} = \left.\frac{\partial F(x,\theta)}{\partial \theta_j}\right|_{x=x^*(\theta)}$$

where the last equality follows by the FONC. □

Theorem 8.15. *(Envelope Theorem II): Let $F, G : \mathbb{R}^n \times \mathbb{R}^m \longrightarrow \mathbb{R}$ be a continuously differentiable functions. For each fixed vector of parameters $\theta \in \mathbb{R}^m$, consider the maximization problem:*

$$\max_{x\in\mathbb{R}^n} F(x,\theta)$$
$$s.t.\ G(x,\theta) \le 0$$

Let $x^*(\theta)$ be the solution and $v(\theta) = F(x^*(\theta),\theta)$ be the value function of this problem. Write the Lagrangian:

$$\mathcal{L}(x,\lambda) = F(x,\theta) - G(x,\theta)$$

If $x^*(\theta)$ and $\lambda^*(\theta)$ are a C^1 functions, then:

$$\frac{\partial v(\theta)}{\partial \theta_j} = \left.\frac{\partial \mathcal{L}(x,\lambda,\theta)}{\partial \theta_j}\right|_{x=x^*(\theta),\lambda^*=\lambda^*(\theta)}$$

8.5 Constrained Optimization, FOC and Fritz-John Conditions

Now, let's look at a more general framework. The Fritz-John conditions may be derived using the separating hyperplane theorem or the implicit function theorem or using penalty functions. The last approach, given in Chapter 9 of Guler, requires no prerequisites (like Lyusternik's theorem).

Consider:

$$\min\ F(x)$$

$$g_i(x) \le 0 \text{ for } i = 1,\ldots,n$$

$$h_j(x) \le 0 \text{ for } j = 1,\ldots,m$$

Throughout we shall assume that the objective function F and the constraint functions $g_1, \ldots, g_n, h_1, \ldots, h_m$ are all continuously dierentiable. We shall refer to the above program as (P) and the set of feasible points as $\mathcal{F}$.

Theorem 8.16. *(Fritz-John conditions). If a point $x^* \in \mathcal{F}$ is a local maximum for the program (P) then there exist Lagrange multipliers $\lambda_0, \ldots \lambda_n \geq 0$ and $\mu_1, \ldots, \mu_m$ with $(\lambda_0, \ldots \lambda_n) \neq 0$ such that:*

$$\lambda_0 \nabla F(x^*) + \sum_{i=1}^{n} \lambda_i \nabla g_i(x^*) + \sum_{i=1}^{m} \mu_i \nabla h_j(x^*) = 0$$

together with the complementary slackness conditions:

$$\lambda_i g_i(x^*) = 0 \text{ for } i = 1, \ldots, n$$

Proof. Consider the penalty functions:

$$F_k(x) = F(x) + \frac{k}{2} \sum_{i=1}^{n} \left[g_i^+(x)\right]^2 + \frac{k}{2} \sum_{i=1}^{m} \left[h_j(x)\right]^2 + \frac{1}{2} \|x - x^*\|^2$$

For large k, the constrained maximization agrees with the unconstrained maximization. □

Remark 8.4. Note that the expression in the Fritz John conditions differs from the usual statement involving Lagrange multipliers. To obtain the stronger:

$$\nabla F(x^*) + \sum_{i=1}^{n} \lambda_i \nabla g_i(x^*) + \sum_{i=1}^{m} \mu_i \nabla h_j(x^*) = 0$$

stronger conditions are needed. One obvious sufficient condition is the following.

Lemma 8.1. *Linear independence of gradient vectors suffices for K-T.*

Proof. This is obvious, since the statement of the Fritz-John conditions already assures us that not all the Lagrange multipliers associated with the objective and the inequality constraints can be zero. □

Example 8.3. Consider the program:

$$\begin{array}{rc} & \min -x \\ & (x-1)^3 + y \leq 0 \\ \text{s.t} & x \geq 0 \\ & y \geq 0 \end{array}$$

Clearly the optimal solution occurs at (x; y) = (1; 0) . However, in this example there are three constraint functions, two of which are active, and the Fritz-John conditions become:

$$\lambda_0(-1,0) + \lambda_1(0,1) + \lambda_2(-1,0) = (0,0)$$

which gives $\lambda_0 = 0$

The following example show that sometimes the solution just follows from the Fritz- John conditions, and no further higher-order analysis is necessary.

Example 8.4. Consider:

$$\begin{array}{c} \min x^2 + 4y^2 + 16z^2 \\ \text{s.t } xyz = 1 \end{array}$$

Note that the gradient of the single constraint function is (yz, xz, xy) which is non-zero on the constraint set, and so there is no loss in setting $\lambda_0 = 1$. The Fritz-John conditions may now be written:

$$(2x, 8y, 32z) = \lambda\,(yz, xz, xy)$$

together with the constraint $xyz - 1 = 0$. Multiplying the components of the above by x, y and z, respectively, gives the system of equations:

$$\begin{aligned} 2x^2 &= \lambda xyz \\ 8y^2 &= \lambda xyz \\ 32z^2 &= \lambda xyz \\ xyz &= 1 \end{aligned}$$

Multiplying the above together, we find $x^6 = 64x^2y^2z^2 = 64$, and so $x = \pm 2$ and $y = \pm 1$. The sign choices of x and y are independent and together determine z from $xyz = 1$, so we need only check the four resulting points:

$$\begin{pmatrix} 2 \\ 1 \\ \frac{1}{2} \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ -\frac{1}{2} \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \\ -\frac{1}{2} \end{pmatrix}, \begin{pmatrix} -2 \\ -1 \\ \frac{1}{2} \end{pmatrix}$$

which turn out to all be global minimizers.

The following is a typical problem from microeconomic theory.

Example 8.5. (Maximising Cobb-Douglas preferences with respect to linear budget constraint) . Consider the maximisation problem:

$$\begin{gathered} \max\ x^{\alpha}y^{1-\alpha} \\ \text{s.t } p_1x + p_2y \le C \end{gathered}$$

where $\alpha \in (0,1)$ and $p_1,\ p_2 > 0$. This fits neatly into the above setup with $F(x,y) = -x^{\alpha}y^{1-\alpha}$ (there is a change of sign because we are maximizing instead of minimizing) and $g(x,y) = p_1x + p_2y - C$. The Fritz-John conditions then ensure the existence of non-zero (λ_0, λ_1) with $\lambda_0, \lambda_1 \ge 0$ such that:

$$-\lambda_0\left(\alpha x^{\alpha-1}y^{1-\alpha}, (1-\alpha)\,x^{\alpha}y^{-\alpha}\right) + \lambda_1\,(p_1, p_2) = 0$$

Clearly this equation implies that $\lambda_0 \neq 0$ and so rearranging we find

$$\begin{aligned} \alpha x^{\alpha-1}y^{1-\alpha} &= \tfrac{\lambda_1}{\lambda_0}p_1 \\ (1-\alpha)\,x^{\alpha}y^{-\alpha} &= \tfrac{\lambda_1}{\lambda_0}p_2 \end{aligned}$$

Dividing these equations by each other gives:

$$\tfrac{\alpha}{1-\alpha}\tfrac{y}{x} = \tfrac{p_1}{p_2}$$

Note that the complementary slackness conditions also imply that the constraint holds with equality and so $y = p_2^{-1}\,[C - p_1x]$. Substituting into the previous equation we find:

$$\tfrac{\alpha}{1-\alpha}\tfrac{C-p_1x}{p_2x} = \tfrac{p_1}{p_2}$$

or

$$x = \tfrac{C\alpha}{p_1}, y = \tfrac{C(1-\alpha)}{p_2}$$

9 Additional Useful Concepts

Theorem 9.1. *(Separating-hyperplane theorem). Suppose that $A \subset \mathbb{R}^n$ is open and convex and consider a vector $x \in \mathbb{R}^n$ with $x \notin A$. There exists $p \neq 0$ such* *that* $p \cdot a \geq p \cdot x$ *for all* $a \in cl(A)$.

This concept is useful in Micro theory. For instance, the second fundamental theorem of welfare economics relies crucially upon the idea of a separating hyperplane.

Definition 9.1. A correspondence Γ between two topological spaces X and Y is a map assigning to each point x a subset $\Gamma(x)$ of Y . We denote this by the notation $\Gamma : X \rightrightarrows Y$. We first adopt the framework of Stokey, Lucas and Prescott, in which X and Y will always denote subsets of Euclidean space equipped with the usual metric.

Definition 9.2. (Sequential upper-hemicontinuity; Stokey, Lucas and Prescott convention) . A correspondence $\Gamma : X \rightrightarrows Y$ is upper hemicontinuous at $x \in X$ if for all $\{x_n\}_{n=1}^{\infty}$ with $x_n \longrightarrow x$ and $\{y_n\}_{n=1}^{\infty}$ satisfying $y_n \in \Gamma$ for all $n \geq 1$, there exists $y \in \Gamma(x)$ such that $y_{n_k} \longrightarrow y$ for some subsequence $(y_n)_{n=1}^{\infty}$.

Definition 9.3. (Sequential lower-hemicontinuity; Stokey, Lucas and Prescott convention) . A correspondence $\Gamma : X \rightrightarrows Y$ is lower hemicontinuous at $x \in X$ if for all $\{x_n\}_{n=1}^{\infty}$ with $x_n \longrightarrow x$ and $y \in \Gamma(x)$ there exists $\{y_n\}_{n=1}^{\infty}$ satisfying $y_n \in \Gamma$ for all $n \geq 1$ such that $y_n \longrightarrow y$.

Definition 9.4. (Sequential continuity for correspondences; Stokey, Lucas and Prescott convention). A correspondence is continuous if it is both upper and lower hemicontinuous.

Definition 9.5. (Topological upper-hemicontinuity; Ok convention). A correspondence $\Gamma : X \rightrightarrows Y$ between metric spaces X and Y is upper hemicontinuous if for all open sets $\mathcal{O} \subseteq Y$ the upper inverse image

$$\Gamma^{-1}(\mathcal{O}) = \{x \in X : \Gamma(x) \subseteq \mathcal{O}\}$$

is open in X.

Definition 9.6. (Topological lower-hemicontinuity; Ok convention). A correspondence $\Gamma : X \rightrightarrows Y$ between metric spaces X and Y is lower hemicontinuous if for all open sets $\mathcal{O} \subseteq Y$ the lower inverse image

$$\Gamma_{-1}(\mathcal{O}) = \{x \in X : \Gamma(x) \cap \neq \emptyset\}$$

is open in X.

Remark 9.1. Unfortunately, the above denitions of Ok and Stokey, Lucas and Prescott for upper-hemicontinuity do not agree in general; one needs to assume the correspondences are compact-valued for the above to coincide. However, for lower-hemicontinuity, they do agree without the need for further specication.

Theorem 9.2. *(Maximum theorem; Theorem 3.6 of Stokey, Lucas and Prescott). Let* $X \subseteq \mathbb{R}^l$ *and* $Y \subseteq \mathbb{R}^m$*, let* $f : X \times Y \longrightarrow \mathbb{R}$ *be a continuous function, and let* $\Gamma : X \rightrightarrows Y$ *be a compact-valued and continuous correspondence. Also define the function* $h : X \longrightarrow \mathbb{R}$ *by:*

$$h(x) = \max_{y \in \Gamma(x)} f(x, y)$$

and the correspondence $G : X \rightrightarrows Y$ *by:* $G(x) = \{y \in Y : h(x) = f(x, y)\}$

Then the function h *is continuous, and the correspondence* G *is compact-valued and upper hemicontinuous.*

We are going to define two useful fixed point theorems since this concept is important for proving the existence of competitive equilibria.

Theorem 9.3. *(Brouwer's fixed point theorem). Let* $S \subseteq \mathbb{R}^l$ *be nonempty, compact and convex. If* $f : S \longrightarrow S$ *is a continuous function, then* f *has a fixed point;* $\exists s^* \in S$ *such that* $f(s) = s$*.*

Theorem 9.4. *(Kakutani's fixed point theorem). Let* $S \subseteq \mathbb{R}^l$ *be nonempty, compact and convex. If* $\Gamma : S \rightrightarrows S$ *is a nonempty-valued, convex-valued and closed-graph correspondence, then* Γ *has a fixed point;* $\exists s^* \in S$ *such that* $s^* \in \Gamma(s^*)$*.*