Mathematics I

Instructor: Kaniṣka Dam
Semester: Fall, 2023
Centro de Investigación y Docencia Económicas

Contents

Part I

Real analysis

Chapter 1

Sets and relations

1.1 Binary relations

In this chapter we are going to define *relation* formally. A relation in everyday life shows an association of objects of a set with objects of other sets (or the same set) such as John owns a BMW, Jim has a green Audi, etc. The essence of relation is these associations. A collection of these individual associations is a relation, such as the ownership relation between people and automobiles. To represent these individual associations, a set of "related" objects, such as John and a BMW, can be used. However, simple sets such as $\{\text{John}, \text{BMW}\}$ are not sufficient here. The order of the objects must also be taken into account, because John owns a BMW but the BMW does not own John, and simple sets do not deal with orders. Thus sets with an order on its members are needed to describe a relation. Here the concept of ordered pair is going to be defined first. A relation is then defined as a set of ordered pairs.

Definition 1.1: Ordered pair

An ordered pair is a list of a pair of objects with an order associated with them. If objects are represented by x and y, then we write an ordered pair as (x, y) or (y, x). In general (x, y) is different from (y, x).

Two ordered pairs (x, y) and (x', y') are equal if and only if $x = x'$ and $y = y'$. For example, if the ordered pair (x, y) is equal to $(1, 2)$, then $x = 1$, and $y = 2$. The ordered pair $(1, 2)$ is not equal to the ordered pair $(2, 1)$.

Definition 1.2: Cartesian product

A Cartesian product of two non-empty sets X and Y, denoted as $X \times Y$, is defined as the set of all ordered pairs (x, y) where x is an element of X and y is an element of Y. That is,

$$X \times Y := \{(x, y) | \, x \in X \text{ and } y \in Y\}.$$

Definition 1.3: Binary relation

Let X and Y be two non-empty sets. A subset R of $X \times Y$ is called a binary relation from X to Y. In other words, a binary relation from a set X to a set Y is a set of ordered pairs (x, y) where x is an element of X and y is an element of Y.

Example 1.1

If $X = \{1, 2, 3\}$ and $Y = \{4, 5\}$, then $\{(1, 4), (2, 5), (3, 5)\}$, for example, is a binary relation from X to Y. However, $\{(1, 1)\}$ is not a binary relation from X to Y because 1 is not in Y.

If $X = Y$, i.e., if R is a relation from X to X, we simply say that it is a relation on X. In other words, R is a relation on X if and only if $R \subseteq X^2$. If $(x, y) \in R$, then we think of R as associating the object x with the object y, and if $\{(x, y), (y, x)\} \cap R = \varnothing$, then we understand that there is no connection between x and y as envisaged by R. Conventionally, we write xRy instead of $(x, y) \in R$.

1.2 Preorders and equivalence relations

Definition 1.4: Preorder

A binary relation $\succsim$ on a non-empty set X is called a preorder if it has the following properties:

(a) Reflexivity: $x \succsim x$ for all $x \in X$,

(b) Transitivity: $x \succsim y$ and $y \succsim z \implies x \succsim z$ for all $x, y, z \in X$.

Definition 1.5: Equivalence relation

A binary relation $\sim$ on a non-empty set X is called an equivalence relation if it has the following properties:

(a) Reflexivity: $x \sim x$ for all $x \in X$,

(b) Symmetry: $x \sim y \implies y \sim x$ for all $x, y \in X$,

(c) Transitivity: $x \sim y$ and $y \sim z \implies x \sim z$ for all $x, y, z \in X$.

For any $x \in X$, the equivalence class of x is defined as the set

$$[x] := \{y \in X \,|\, y \sim x\}.$$

Definition 1.6: Partition

A partition of a non-empty set X is a class $\{X_i\}$ of non-empty subsets of X such that (i) $\cup_i X_i = X$, and (ii) $X_i \cap X_j = \varnothing$ for any two X_i and X_j in $\{X_i\}$. The X_i's are called the partition sets.

If $X = \{1, 2, 3, 4, 5\}$, then $\{\{1, 2, 3\}, \{4, 5\}\}$ and $\{\{1, 2, 5\}, \{3, 4\}\}$ are two different partitions of X. If $X = \mathbb{R}$, then it can be partitioned into the infinitely many closed-open intervals of the form $[n, n+1)$ where n is an integer.

We now show that a given equivalence relation $\sim$ on X determines a natural partition of X. Let the relation $\sim$ on X satisfies reflexivity, symmetry and transitivity. We show that all distinct equivalence classes form a partition of X. By reflexivity, $x \in [x]$ for each element $x \in X$, so each equivalence set is non-empty and their union is X. It remains to show that any two equivalence sets $[x_1]$ and $[x_2]$ are either disjoint or identical. We prove by showing that if $[x_1]$ and $[x_2]$ are not disjoint then they must be identical. Suppose that $z \in [x_1] \cap [x_2]$. Then $z \sim x_1$ and $z \sim x_2$, and by symmetry, $x_1 \sim z$. Let $y \in [x_1]$ which implies that $y \sim x_1$. Since $y \sim x_1$ and $x_1 \sim z$, transitivity implies that $y \sim z$. Then, again by transitivity, $y \sim z$ and $z \sim x_2$ implies that $y \sim x_2$,

so that $y \in [x_2]$. Hence, $[x_1] \subseteq [x_2]$. By a similar logic, one can show that $[x_2] \subseteq [x_1]$. From this we can conclude that $[x_1] = [x_2]$.

The converse statement is somewhat trivial. Consider a partitions $\{X_i\}$ of a set X, and let R be a relation on X defined by, for any two elements x and y of X,

$$xRy \text{ if and only if } x \text{ and } y \text{ belong to the same element } X_i \text{ of } \{X_i\}.$$

We think of the elements of X being distributed into a number of sets, each of which is a member of a given partition. Every element of X is exactly in one set. Notice that every element x of X is in the same set as itself. For any x and y of X, if x and y are in the same set, so are y and x. Finally, For any x, y and z of X, if x and y are in the same set, and y and z are also, so are x and z. Thus, the relation R is reflexive, symmetric and transitive, and hence is an equivalence relation. Thus, there is no real distinction between partitions and equivalence relations.

Example 1.2: The indifference set

Let $x = (x_1, x_2)$ be the quantities of two goods 1 (apple) and 2 (orange), and x is an element of $X \subseteq \mathbb{R}^2_+$, which is called the consumption set. Define the *preference relations* $\succsim$ on X of a consumer as follows. We read $x \succsim y$ as "x is at least as good as y". From $\succsim$, we can derive two other important relations on X:

(a) The *strict preference* relation, $\succ$, defined by $x \succ y \iff x \succsim y$ but $y \not\succsim x$ and read "x is (strictly) preferred to y".

(b) The *indifference* relation, $\sim$, defined by $x \sim y \iff x \succsim y$ and $y \succsim x$ and read "x is indifferent to y".

The relation $\succsim$ is *rational* if (i) it is *complete*, i.e., for all x, $y \in X$, we have that $x \succsim y$ or $y \succsim x$ (or both), and (ii) it is *transitive*, i.e., for all x, y, $z \in X$, if $x \succsim y$ and $y \succsim z$, then $x \succsim z$.

It is easy to prove the following property.

Lemma 1.1

If $\succsim$ is rational, then the indifference relation $\sim$ is an equivalence relation, but the strict preference relation $\succ$ is not.

For a consumption bundle $x \in X$, the set $[x] = \{y \in X | y \sim x\}$ is called the *indifference set* of x. Notice that $\sim$ induces a partition of X and vice versa.

1.3 Partial orders

Definition 1.7: Partial order

A binary relation $\succsim$ on a non-empty set X is called a partial order if it has the following properties:

(a) Reflexivity: $x \succsim x$ for all $x \in X$,

(b) Antisymmetry: $x' \succsim x''$ and $x'' \succsim x'$ imply $x' = x''$ for all x', $x'' \in X$,

(c) Transitivity: $x' \succsim x''$ and $x'' \succsim x'''$ imply $x' \succsim x'''$ for all x', x'' and $x''' \in X$.

Definition 1.8: Partially ordered set

A partially ordered set (written "poset") is a set X on which there is a binary relation $\succsim$ that is reflexive, antisymmetric and transitive.

Definition 1.9

Two elements x' and x'' of a partially ordered set are *ordered* if either $x' \succsim x''$ or $x'' \succsim x'$; otherwise x' and x'' are unordered.

Definition 1.10: Chain

A partially ordered set is a chain if it does not contain an unordered pair of elements.

Example 1.3: Examples of partially ordered sets

The following are partially ordered sets.

(a) The real line $\mathbb{R}$ with usual ordering relation $\geq$ on the real numbers is a poset.

(b) Given any two vectors x', x'' in the m-dimensional Eucledian space $\mathbb{R}^m = \{x = (x_1, \ldots, x_m) \mid x_i \in \mathbb{R}$ for all $i = 1, \ldots, m\}$, $x' \geq x''$ if $x'_i \geq x''_i$ for all $i = 1, \ldots, m$. $\mathbb{R}^m$ with the vector ordering relation $\geq$ is a poset.

(c) The power set, $\mathcal{P}(X)$, of a set X is the set of all subsets of X. The power set $\mathcal{P}(X)$ with the set inclusion ordering relation $\supseteq$ is a poset. If X' and X'' are distinct subsets of X with $X' \subseteq X''$, then $X' \subset X''$.

(d) The lexicographic ordering relation $\preceq_{lex}$ on $\mathbb{R}^m$ is such that $x' \preceq_{lex} x''$ in $\mathbb{R}^m$ if either $x' = x''$ or there is some i' with $1 \leq i' \leq m$, $x'_i = x''_i$ for each i with $1 \leq i < i'$, and $x'_{i'} < x''_{i'}$. The set $\mathbb{R}^m$ with the ordering relation $\preceq_{lex}$is a poset, indeed, is a chain.

Definition 1.11: Bound of a set and bounded set

Suppose that X is a partially ordered set with respect to $\succsim$, and X' is a subset of X. If $u \in X$ such that $u \succsim x$ for every $x \in X'$, then u is an upper bound of X'. Similarly, if $l \in X$ such that $l \precsim x$ for every $x \in X'$, then l is a lower bound of X'. If a subset X' of a poset X has an upper (lower) bound, then X' is bounded above (below). In case X' has both an upper bound and a lower bound, then X' is a bounded set.

First, observe that a subset X' of a poset X may be unbounded. Consider $X = \mathbb{R}$, the set of real numbers, and $X' = \mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$, the set of integers. Then $\mathbb{Z}$ is an unbounded set in $\mathbb{R}$. On the other hand, if $X' = \mathbb{N} = \{1, 2, \ldots\}$, the set of natural numbers, then X' is bounded below by 1. Next, let $X = \mathbb{R}$ and $X' = \{5, 10, 15\}$. Then both 15 and 20 are upper bounds of X'. Now let $X = \mathbb{R}$ and $X' = [a, b)$. The real number b is an upper bound of X' which is not contained in X'. Two important observations emerge. First, a subset X' of a poset X may or may not contain its upper (lower) bounds (if they exist). This gives rise to the notion of a greatest element and a least element.

Definition 1.12: Greatest and least elements

Suppose that X is a partially ordered set with respect to $\succsim$, and X' is a subset of X. If $x' \in X'$ such that $x' \succsim x$ for every $x \in X'$, then x' is a greatest element of X'. Similarly, if $z' \in X'$ such that $z' \precsim x$

for every $x \in X'$, then z' is a least element of X'.

A greatest (least) element of X' is trivially an upper (lower) bound of X'. For example, if $X = \mathbb{R}$ and $X' = \{5, 10, 15\}$, then 2 is not a least element of X', whereas 5 is. Notice that a subset X' of a poset X can have at most one greatest (least) element. Next, a subset X' of a poset X may have many different upper and lower bounds. This gives rise to the notion of a supremum and an infimum.

Definition 1.13: Supremum and infimum

Suppose that X is a partially ordered set with respect to $\succsim$, and X' is a subset of X. An element of X is a supremum or a least upper bound of X' (with respect to X), denoted $\sup_X(X')$, if it is the case that (a) $\sup_X(X') \succsim x$ for all $x \in X'$, and (b) for any $u \in X$ such that $u \succsim x$ for all $x \in X'$, it holds that $\sup_X(X') \precsim u$. Similarly, an element of X is an infimum or a greatest lower bound of X' (with respect to X), denoted $\inf_X(X')$, if it is the case that (a) $\inf_X(X') \precsim x$ for all $x \in X'$, and (b) for any $l \in X$ such that $l \precsim x$ for all $x \in X'$, it holds that $\inf_X(X') \succsim u$.

The above definition says that if the set of upper (lower) bounds of X' has a least (greatest) element, then this least (greatest) element is the supremum (infimum) of X'. Two important points must be kept in mind. First, a supremum or an infimum of a set may not exist. If it exists, then it must be unique. Second, one must be clear about the underlying set in expressing $\sup(X')$ or $\inf(X')$. Consider the following example.

Example 1.4

Suppose $X = \mathbb{R}$, $Y = [0, 1) \cup \{2\}$, and $X' = [0, 1)$. Then $\sup_X(X') = 1 \neq 2 = \sup_Y(X')$.

Notice also that, if a subset X' of a poset X contains $\sup_X(X')$, then the supremum is the greatest element. Similar property holds for the infimum. Following is an important result concerning the set of real numbers.

Lemma 1.2: The supremum property

Every non-empty set of real numbers that is bounded above has a supremum. This supremum is a real number.

Proof. Let $X \subseteq \mathbb{R}$ be a non-empty subset of real numbers, and let $U := \{u \in \mathbb{R} \mid u \geq x \text{ for all } x \in X\}$, the set of upper bounds of X. By assumption, U is non-empty. By the axiom of completeness,[1] there exists a real number α such that

$$x \leq \alpha \leq u, \quad \text{for all } x \in X \text{ and } u \in U.$$

Because $x \leq \alpha$ for all $x \in X$, α is an upper bound of X. This implies that $\alpha \in U$ with the property that $\alpha \leq u$ for all $u \in U$, and hence, α is a supremum of X. $\square$

The following definition introduces the notion of a maximal (minimal) element, which is, in general, different from a greatest (least) element.

Definition 1.14: Maximal and minimal element

If x' is in X' and there does not exist any $x'' \in X'$ with $x'' \succ x'$ ($x' \succ x''$), then x' is a maximal (minimal) element of X'.

A greatest (least) element is a maximal (minimal) element. Thus, it is a stronger notion that maximal

[1] This axiom says that, if L and H are two non-empty subsets of real numbers with the property that for all $l \in L$ and for all $h \in H$ we have $l \leq h$, then there is a real number α such that $l \leq \alpha \leq h$ for all $l \in L$ and for all $h \in H$.

(minimal) element. A poset can may have any number of maximal (minimal) element. For example, In the *fence* $a_1 > b_1 < a_2 > b_2 < a_3 > b_3 < \ldots$, all the a_i's are maximal, and all the b_i's are minimal. Consider another example. Let X be the set with at least two elements, and $\mathcal{S} = \{\{x\} \mid x \in X\}$ be the subset of $\mathcal{P}(X)$ consisting of singletons. Take the usual set inclusion ordering, i.e., for any two sets X and Y in $\mathcal{P}(X)$, $X \precsim Y$ if and only if $X \subseteq Y$. This is the discrete poset – no two elements are comparable. Thus, every element $\{x\} \in \mathcal{S}$ is maximal and minimal, and for any x' and x'', neither $\{x'\} \subset \{x''\}$ nor $\{x''\} \subset \{x'\}$. Hence, we can conclude that distinct maximal (minimal) elements are unordered.

Example 1.5: Consumer theory

In economics, we relax the axiom of antisymmetry, using preorders instead of partial orders. Let $X \subseteq \mathbb{R}^n_+$. Preferences of a consumer are represented by a preorder $\succsim$ (as in Example 1.2). The preference relations are never assumed to be antisymmetric. In this context, for any $B \subset X$, we call $x \in B$ a maximal element if $x \succsim y$ for all $y \in B$, and it is interpreted as the consumption bundle that is not dominated in the sense that $x \prec y$ for any $y \in B$. The notion of greatest element for a preference preorder would be that of most preferred choice. That is, some $x \in B$ with $y \in B$ implying that $x \succ y$. An obvious application is to the definition of demand correspondence. An element p of $\mathbb{R}^n_{++}$ is called a price system which maps every consumption bundle x into its market value $p \cdot x \in \mathbb{R}_+$. An income level m is an element of $\mathbb{R}_+$. The budget correspondence $\mathcal{B}$ is a correspondence from $\mathbb{R}^n_{++} \times \mathbb{R}_+$ to X which is given by

$$\mathcal{B}(p, m) = \{x \in X \mid p \cdot x \leq m\}.$$

The demand correspondence maps any price p and any level of income m into the set of maximal elements (with respect to $\succsim$) of $\mathcal{B}(p, m)$, which is given by

$$\mathcal{D}(p, m) = \{x \in X \mid x \text{ is a maximal element of } \mathcal{B}(p, m)\}.$$

It is called demand correspondence because the theory predicts that for given p and m, the rational choice of a consumer x^* will be some element of $\mathcal{D}(p, m)$.

1.4 Functions

1.4.1 Definition and properties

A function from a set X to a set Y is a rule that assigns to each element x of X a unique element y of Y. We say that y is the image of x under f, and write $y = f(x)$. Conversely, x is an element of the preimage or inverse image of y, written $x \in f^{-1}(y)$. Formally,

Definition 1.15: Function

Let X and Y be two sets. A function f from X to Y, written $f : X \longrightarrow Y$, is a relation from X to Y with the property that for each $x \in X$ there exists a unique element $y \in Y$ such that $(x, y) \in f$.

The underlying set X is called the *domain* of f, and $f(X)$ is its range. If $A \subseteq X$, then its image under f is given by

$$f(A) := \{y \in Y \mid \text{ there exists } x \in A \text{ such that } y = f(x)\} = \cup_{x \in A} \{f(x)\}.$$

Given a subset B of Y, its inverse image is given by

$$f^{-1}(B) := \{x \in X \mid f(x) \in B\}.$$

If $f : X \longrightarrow Y$ and $f(X) = Y$, then f is *surjective* or *onto*. A function f is *injective* or *one-to-one* if for any two distinct elements $x_1, x_2 \in X$, we have $f(x_1) \neq f(x_2)$. A function is *bijective* if it is both one-to-one and onto, i.e., if each element $y \in Y$ has a unique inverse image in X. Let $f : X \longrightarrow Y$ and $g : Y \longrightarrow Z$. Then their *composition*, $g \circ f$ is a function from X to Z and is defined by $(g \circ f)(x) = g(f(x))$. We have $(h \circ g) \circ f = h \circ (g \circ f) = h \circ g \circ f$, i.e., the composition is associative.

Theorem 1.1

Let $f : X \longrightarrow Y$ be a function and $\mathcal{B} = \{B_i \,|\, i \in I\}$ a family of subsets of Y. Then

(a) $f^{-1}\left(\cup_{i \in I} B_i\right) = \cup_{i \in I} f^{-1}\left(B_i\right)$,

(b) $f^{-1}\left(\cap_{i \in I} B_i\right) = \cap_{i \in I} f^{-1}\left(B_i\right)$.

Proof. The proof is left as an exercise. □

Definition 1.16: Increasing function

A function $f : X \longrightarrow Y$ is (strictly) increasing if $x' > x''$ in X implies that $f(x')\,(>)\, \geq f(x'')$.

1.4.2 Representation of binary relations

Definition 1.17: Utility function

A function $f : X \longrightarrow \mathbb{R}$ is a utility function, where $X \subseteq \mathbb{R}^m_+$, representing preference relation $\succsim$ if, for all $x, y \in X$, $x \succsim y \iff f(x) \geq f(y)$.

Proposition 1.1

A preference relation $\succsim$ can be represented by a utility function only if it is rational.

Proof. To prove this proposition, we must show that if there is a utility function that represents the preference relation $\succsim$, then $\succsim$ must be complete and transitive. First, because $f(\cdot)$ is a real valued function defined on X, it must be that for any $x, y \in X$, either $f(x) \geq f(y)$ or $f(y) \geq f(x)$. Because $f(\cdot)$ represents $\succsim$, this implies either that $x \succsim y$ or that $y \succsim x$. Hence, $\succsim$ must be complete. Next, Suppose that $x \succsim y$ and $y \succsim z$. Because $f(\cdot)$ represents $\succsim$, it must be that $f(x) \geq f(y)$ and $f(y) \geq f(z)$. Therefore, $f(x) \geq f(z)$, which implies that $x \succsim z$, and hence $\succsim$ is transitive. □

1.4.3 Examples and graphical representation

Let $f : X \longrightarrow Y$ be a function, and consider the subset of $X \times Y$

$$G_f := \{(x, y) \in X \times Y \,|\, y = f(x)\}. \tag{1.1}$$

The above set is called the graph of the function f. In what follows we consider real valued functions $f : X \longrightarrow \mathbb{R}$, where $X \subseteq \mathbb{R}^2$, that assign to each $x = (x_1, x_2) \in X$ a number $f(x_1, x_2)$ in $\mathbb{R}$.

Example 1.6

Following are the examples of real valued functions from $\mathbb{R}^2$ to $\mathbb{R}$.

(i) $f(x_1, x_2) = \alpha_1 x_1 + \alpha_2 x_2$ (Linear),

(ii) $f(x_1, x_2) = A x_1^{\alpha_1} x_2^{\alpha_2}$ (Cobb-Douglas),

(iii) $f(x_1, x_2) = \min\{\alpha_1 x_1, \alpha_2 x_2\}$ (Leontief),

(iv) $f(x_1, x_2) = A\left[\alpha_1 x_1^{-\rho} + \alpha_2 x_2^{-\rho}\right]^{-\frac{1}{\rho}}$ (Constant elasticity of substitution).

Given a real valued function $f : X \longrightarrow \mathbb{R}$, consider the set $I(x) := \{y \in \mathbb{R}^2 | \, f(x) = f(y)\}$. Thus, if $x \sim y$ then the function $f(.)$ is constant on the set $I(x)$. Let this value be α. The graph of $I(x)$ defines an indifference curve or a level curve at the level α. The graphical representation of functions from $\mathbb{R}^2$ to $\mathbb{R}$ is often difficult since the graphs are three dimensional. We represent such functions by their level curves. How such level curves are drawn is a topic of discussion of Chapter 3.

1.5 Correspondences

A *correspondence* φ from X to Y is a function that to each element x in X assigns a non-empty subset $\varphi(x)$ of Y. Hence, a correspondence from X to Y, denoted $\varphi : X \to\to Y$, is a function from X to $\mathcal{P}(Y)$. Notice that a function associates with every element $x \in X$ an element $y \in Y$. In saying this we are not making the distinction between the element y in Y and the subset $\{y\}$ of Y which consists of one point only. Given a correspondence $\varphi : X \longrightarrow \mathcal{P}(Y)$, consider the subset G_φ of $X \times Y$ defined by

$$G_\varphi := \{(x, y) \in X \times Y | \, y \in \varphi(x)\}. \tag{1.2}$$

The set G_φ is called the *graph* of φ. Conversely, if G is a subset of $X \times Y$ such that for every $x \in X$ the set of points $y \in Y$ with $(x, y) \in G$ is non-empty, then G is the graph of a correspondence, i.e.,

$$\varphi(x) := \{y \in Y | \, (x, y) \in G\}. \tag{1.3}$$

We now introduce the notion of inverse image(s) of a correspondence. Let $\varphi : X \to\to Y$ be a correspondence, where $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$, and W be any subset of Y. Define the upper inverse of W under φ, denoted $\varphi_+^{-1}(W)$, by

$$\varphi_+^{-1}(W) := \{x \in X \mid \varphi(x) \subseteq W\},$$

and the lower inverse of W under φ, denoted $\varphi^{-1}(W)$, by

$$\varphi^{-1}(W) := \{x \in X \mid \varphi(x) \cap W \neq \varnothing\}.$$

Example 1.7: Best response correspondence

Consider the following two-player normal form game $\Gamma = \langle N, \{S_i\}_{i \in N}, \{u_i(\cdot)\}_{i \in N}\rangle$, where $N = \{1, 2\}$ is the set of players, $S_1 = \{s_1, s_2, s_3\}$ and $S_2 = \{t_1, t_2, t_3, t_4\}$ are the strategy sets of players 1 and 2 respectively, and $u_i(s, t)$ is the payoff function of player i which is given as follows.

	t_1	t_2	t_3	t_4
s_1	$1,1$	$1,1$	$0,0$	$0,0$
s_2	$0,0$	$2,2$	$2,2$	$0,0$
s_3	$1,0$	$1,0$	$0,0$	$3,3$

A strategy t^* of player 2 is a *best response* against a strategy s of player 1 if $u_2(s,\, t^*) \geq u_2(s,\, t)$ for all $t \in S_2$. In a similar fashion one can define the best response of player 1. Notice that the best response of player 2 is a correspondence φ as there is no unique optimal strategy against s_1 and s_2. The best response correspondence of player 2 is given by $\varphi(s_1) = \{t_1,\, t_2\}$, $\varphi(s_2) = \{t_2,\, t_3\}$, $\varphi(s_3) = \{t_4\}$. In this example, the upper and lower inverses of the set $W = \{t_2,\, t_3\} \subseteq S_2$ are given by $\varphi_+^{-1}(W) = \{s_2\}$ and $\varphi^{-1}(W) = \{s_1,\, s_2\}$.

Chapter 2

The analysis of metric spaces

2.1 Metric space

Definition 2.1: Metric space

A metric space is a pair (X, d), where $X \neq \varnothing$ and $d : X \times X \longrightarrow \mathbb{R}$ is a distance function or metric over X that associates with each pair of points (x, y) in X a real number $d(x, y)$ that satisfies

(a) $d(x, y) \geq 0$ for all $x, y \in X$,

(b) $d(x, y) = 0 \iff x = y$ for all $x, y \in X$,

(c) $d(x, y) = d(y, x)$ for all $x, y \in X$,

(d) $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$. (Triangular inequality)

Example 2.1

Following are the examples of metric spaces.

(a) $(\mathbb{R}, d)$ where $d(x, y) = |x - y|$ for $x, y \in \mathbb{R}$.

(b) $(\mathbb{R}^n, d_2)$ where $\mathbb{R}^n = \{(x_1, \ldots, x_n) | \, x_i \in \mathbb{R} \text{ for } i = 1, \ldots, n\}$ is the n-dimensional Euclidean space, and $d_2(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$ for $x, y \in \mathbb{R}^n$.

Proof. The proofs of 2.1(a)-(c) are trivial. Thus, we will only prove the triangular inequality which makes use of the following inequality:

$$\frac{(a_1 + \ldots + a_n)^2}{p_1 + \ldots + p_n} \leq \frac{a_1^2}{p_1} + \ldots + \frac{a_n^2}{p_n} \quad \text{for } a_i \in \mathbb{R} \quad \text{and } p_i \in \mathbb{R}_{++}. \tag{2.1}$$

To prove the above inequality, notice that, for real numbers a and b, and for $p > 0$ and $q > 0$, the following is true.

$$\frac{(a+b)^2}{p+q} \leq \frac{a^2}{p} + \frac{b^2}{q}. \tag{2.2}$$

The above is nothing but the restatement of $(aq - bp)^2 \geq 0$. Now replace b by $b + c$ and q by $q + r$ in (2.2) to get the following:

$$\frac{(a+b+c)^2}{p+q+r} \leq \frac{a^2}{p} + \frac{(b+c)^2}{q+r} \leq \frac{a^2}{p} + \frac{b^2}{q} + \frac{c^2}{r}. \tag{2.3}$$

Repeat the above steps to get the inequality in (2.1). Take $a_i = (x_i - y_i)(y_i - z_i)$ and $p_i = (y_i - z_i)^2$. Then inequality (2.1) implies that

$$\sum_{i=1}^{n}(x_i - y_i)(y_i - z_i) \leq \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}\sqrt{\sum_{i=1}^{n}(y_i - z_i)^2}. \tag{2.4}$$

The above inequality is known as the *Cauchy-Schwartz inequality*. Now we are ready to show the triangular inequality.

$$\begin{aligned}
[d_2(x,\, z)]^2 &= \sum_{i=1}^{n}(x_i - z_i)^2 \\
&= \sum_{i=1}^{n}\{(x_i - y_i) + (y_i - z_i)\}^2 \\
&= \sum_{i=1}^{n}(x_i - y_i)^2 + \sum_{i=1}^{n}(y_i - z_i)^2 + 2\sum_{i=1}^{n}(x_i - y_i)(y_i - z_i) \\
&\leq [d_2(x,\, y)]^2 + [d_2(y,\, z)]^2 + 2\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}\sqrt{\sum_{i=1}^{n}(y_i - z_i)^2} \\
&= [d_2(x,\, y)]^2 + [d_2(y,\, z)]^2 + 2d_2(x,\, y)d_2(y,\, z) \\
&= [d_2(x,\, y) + d_2(y,\, z)]^2.
\end{aligned}$$

The above inequality follows from (2.4). $\square$

(c) $(\mathbb{R}^n,\, d_\infty)$ where $d_\infty(x,\, y) = \max_{i=1,\,\ldots,\,n}\{|x_i - y_i|\}$. For proving the triangular inequality, notice that for all $i = 1,\, \ldots,\, n$, $\alpha_i \leq \max_j \alpha_j$ and $\beta_i \leq \max_j \beta_j$, which together imply $\alpha_i + \beta_i \leq \max_j \alpha_j + \max_j \beta_j$, which in turn implies $\max_i(\alpha_i + \beta_i) \leq \max_j \alpha_j + \max_j \beta_j$. Now take $\alpha_i = |x_i - y_i|$ and $\beta_i = |y_i - z_i|$ to complete the proof.

(d) $(\mathbb{R}^n,\, d_p)$ where $d_p(x,\, y) = (\sum_{i=1}^{n} |x_i - y_i|^p)^{\frac{1}{p}}$.

(e) $(\mathcal{C}([a,\, b]),\, d_2)$ where $\mathcal{C}([a,\, b]) = \{f : [a,\, b] \longrightarrow \mathbb{R} |\, f$ is continuous$\}$, the set of all real valued continuous functions defined on $[a,\, b] \subset \mathbb{R}$, and $d_2(f,\, g) = \left(\int [f(x) - g(x)]^2 dx\right)^{\frac{1}{2}}$.

The proofs of (c)-(e) are left as exercise.

Definition 2.2: Euclidean norm

The Euclidean norm of a vector $x \in \mathbb{R}^n$, denoted $||x||$, is defined as

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

The Euclidean norm satisfies the following properties at all $x,\, y \in \mathbb{R}^n$, and $\alpha \in \mathbb{R}$.

(a) *Positivity:* $||x|| \geq 0$, with equality if and only if $x = 0$.

(b) *Homogeneity:* $||\alpha x|| = |\alpha|\,.\,||x||$.

(c) *Triangular Inequality:* $||x + y|| \leq ||x|| + ||y||$.

It is easy to show that, for $d(x, y) := ||x - y||$, $(\mathbb{R}^n, ||.||)$ is a metric space.

Definition 2.3: Open and closed balls

Let (X, d) is a metric space. For $x \in X$ and $r > 0$, the **open ball** with centre x and radius r is given by

$$B_r(x) = \{y \in X | \, d(x, y) < r\}, \tag{2.5}$$

and the **closed ball** with centre x and radius r is given by

$$\bar{B}_r(x) = \{y \in X | \, d(x, y) \leq r\}. \tag{2.6}$$

Definition 2.4: Bounded set

A subset S of X is **bounded** if there exist $x \in X$ and $r > 0$ such that $S \subset B_r(x)$.

Lemma 2.1

Let (X, d) be a metric space and $S \subset X$. The diameter of S is defined by

$$diam(S) := sup\{d(x, y) | \, x, y \in S\}$$

Then S is bounded if and only if $diam(S)$ is finite.

Proof. *Necessity:* Let $x, y \in S$. Since S is bounded, there exist $x_0 \in X$ and $r \in (0, \infty)$ such that $S \subset B_r(x_0)$. Since d is a metric, by the triangular inequality, we have

$$\begin{aligned} & d(x, y) \leq d(x, x_0) + d(x_0, y) < 2r, \\ \implies \quad & diam(S) = sup\{d(x, y)\} < 2r < \infty. \end{aligned}$$

Sufficiency: Assume $S \neq \varnothing$ and take $\bar{x} \in S$. Consider any other $x \in S$. Then we have $d(x, \bar{x}) \leq diam(S)$. Take $r > 0$ such that $diam(S) < r$. This implies that $x \in B_r(\bar{x})$. Thus, $S \subset B_r(\bar{x})$ implying that S is bounded. $\square$

Definition 2.5: Bounded function

Let A be an arbitrary set, and let (X, d) be a metric space. The function $f : A \longrightarrow X$ is bounded if and only if $f(A) = \{f(a) | \, a \in A\} \subseteq X$ is bounded.

2.2 Sequences in a metric space

Definition 2.6: Sequence

Let (X, d) be a metric space. A sequence in (X, d) is a function from $\mathbb{N}$, the set of natural numbers, to X that associates with each $n \in \mathbb{N}$ an element x_n of X, and is denoted by $\{x_n\}$.

Definition 2.7: Convergent sequence

A sequence $\{x_n\}$ in a metric space (X, d) converges to a point $x \in X$ if and only if for any $\varepsilon > 0$, there exists n_ε such that $d(x_n, x) < \varepsilon$ for every $n \geq n_\varepsilon$. If a sequence $\{x_n\}$ in a metric space (X, d)

converges to a point $x \in X$, then we say that the sequence is convergent and its limit is x, and write

$$\lim_{n \to \infty} x_n = x.$$

Lemma 2.2: Uniqueness of limit

If $\{x_n\}$ converges to x and x' in a metric space (X, d), then $x = x'$.

Proof. Suppose that $x \neq x'$, and define by $r = d(x, x') > 0$. Take ε such that $0 < \varepsilon < r/2$. Since $\{x_n\}$ converges to x and x', we can find n_ε and n'_ε such that $d(x_n, x) < \varepsilon$ for $n \geq n_\varepsilon$ and $d(x_n, x') < \varepsilon$ for $n \geq n'_\varepsilon$, respectively. Take any $n \geq \max\{n_\varepsilon, n'_\varepsilon\}$. By the triangular inequality,

$$d(x, x') \leq d(x, x_n) + d(x_n, x') < 2\varepsilon < r = d(x, x'),$$

which is a contradiction. $\square$

Proposition 2.1

We have $\lim_{n \to \infty} \frac{1}{n} = 0$.

Proof. Let $\{x_n\}$ be a sequence such that $x_n = \frac{1}{n}$. We have to find an n_ε such that $|1/n - 0| = 1/n < \varepsilon$ for every $n \geq n_\varepsilon$. Notice that if $n \geq n_\varepsilon$, then $1/n \leq 1/n_\varepsilon$. Thus, if we pick $n_\varepsilon > 1/\varepsilon$, then $1/n_\varepsilon < \varepsilon$, and so $|1/n - 0| < \varepsilon$. Since ε can be chosen arbitrarily, $1/n$ converges to 0. $\square$

Lemma 2.3

If $\{x_n\}$ is convergent, then $\{x_n\}$ is bounded.

Proof. Take $\varepsilon = 1$, and suppose that $\{x_n\}$ converges to x. Then there exists n_1 such that $d(x_n, x) < 1$ for all $n \geq n_1$. Now take r such that

$$r > \max\{1, d(x_1, x), \ldots, d(x_{n_1}, x)\}.$$

We have to check that $x_n \in B_r(x)$ for all $n \in \mathbb{N}$. If $n \geq n_1$, then $d(x_n, x) < 1 < r$, and hence $x_n \in B_r(x)$. If $n < n_1$, then $d(x_n, x) < r$, and hence $x_n \in B_r(x)$. $\square$

Let $\{x_n\} = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, \ldots\}$ be a sequence. Then $\{x_2, x_4, x_6, x_8, \ldots\}$ is a subsequence of $\{x_n\}$. Formally,

Definition 2.8: Subsequence

Let $s : \mathbb{N} \longrightarrow X$ be a sequence and let $p : \mathbb{N} \longrightarrow \mathbb{N}$ that associates with each $k \in \mathbb{N}$ a natural number $n_k \in \mathbb{N}$ be a strictly increasing function, i.e., $p(n+1) > p(n)$ for all $n \in \mathbb{N}$. Then $s \circ p : \mathbb{N} \longrightarrow X$ is a subsequence of s.

Lemma 2.4

If $\{x_{n_k}\}$ is a subsequence of $\{x_n\}$, then $\lim_{n \longrightarrow \infty} x_n = x$ implies that $\lim_{n_k \to \infty} x_{n_k} = x$.

Proof. If $p : \mathbb{N} \longrightarrow \mathbb{N}$ is strictly increasing, then $p(n) \geq n$ for all $n \in \mathbb{N}$. Fix an $\varepsilon > 0$. Then there exists $n_\varepsilon \in \mathbb{N}$ such that $d(x_n, x) < \varepsilon$ for all $n \geq n_\varepsilon$. Take $k \geq n_\varepsilon$. Since $n_k \geq k$, we have $n_k \geq n_\varepsilon$. Thus, $d(x_{n_k}, x) < \varepsilon$ for $n_k \geq n_\varepsilon$. $\square$

The above lemma further asserts that if two subsequences of a sequence $\{x_n\}$ converge to two different limit points, then the sequence has no limit. Thus, the converse of Lemma 2.3 is not always true. For example, the sequence $\{x_n\}$ such that $x_n = (-1)^n$ is bounded by -1 and 1. But the subsequence $\{x_{2n}\}$ converges to 1 and the subsequence $\{x_{2n+1}\}$ converges to -1.

Lemma 2.5

Properties of a sequence in $(\mathbb{R},\, d)$ where $d(x,\, y) := |x - y|$.

(a) A sequence $\{x_n\}$ that is increasing (decreasing), i.e., $x_n \leq\ (\geq)\, x_{n+1}$ for all $n \in \mathbb{N}$, and bounded is convergent to its supremum (infimum), i.e., $\lim_{n\to\infty} x_n = \sup_n\{x_n\}\ (\inf_n\{x_n\})$.

(b) Every sequence in $\mathbb{R}$ contains a monotone subsequence.

(c) (Bolzano-Weirstrass Theorem) Every bounded sequence in $\mathbb{R}$ contains a convergent subsequence.

(d) $\lim_{n\to\infty}(x_n + y_n) = \lim_{n\to\infty} x_n + \lim_{n\to\infty} y_n$.

(e) $\lim_{n\to\infty}(x_n\, y_n) = (\lim_{n\to\infty} x_n)(\lim_{n\to\infty} y_n)$.

(f) $\lim_{n\to\infty}\left(\frac{x_n}{y_n}\right) = \frac{\lim_{n\to\infty} x_n}{\lim_{n\to\infty} y_n}$ for $\lim_{n\to\infty} y_n \neq 0$.

(g) $x_n \leq y_n$ for all $n \in \mathbb{N}$, and $\lim_{n\to\infty} x_n$ and $\lim_{n\to\infty} y_n$ exist. Then $\lim_{n\to\infty} x_n \leq \lim_{n\to\infty} y_n$.

Proof. (a) Suppose that $\{x_n\}$ is an increasing and bounded (from above) sequence. By the axiom of the supremum, $s = sup_n\{x_n\}$ exists. Take an $\varepsilon > 0$. Given that $s = sup_n\{x_n\}$, $s - \varepsilon$ cannot be an upper bound of x_n, and hence there exists an $N \in \mathbb{N}$ such that $s \geq x_N > s - \varepsilon$. Since the sequence is increasing, we have that $s \geq x_n > s - \varepsilon$ for all $n \geq N$. Therefore,

$$s - \varepsilon < x_n < s + \varepsilon, \quad \text{for all } n > N,$$

which implies that $|x_n - s| < \varepsilon$ for all $n > N$. The proof for a decreasing sequence is analogous.

(b) Let $\{x_n\}$ be an arbitrary sequence of real numbers. The term x_s is a dominant term if for all $s < n$ we have $x_s > x_n$. Let

$$S = \{s \in \mathbb{N} |\, x_s > x_n \text{ for all } n > s\}$$

be the set of all subindices of the dominant terms of the sequence $\{x_n\}$. There are two possibilities. First, S is infinite. Then order the dominant terms with increasing subindices, i.e., $x_{s_1} > \ldots > x_{s_k} > \ldots$ for $s_1 < \ldots < s_k < \ldots$. Therefore, the subsequence $\{x_{s_k}\}$ is a decreasing subsequence. The second possibility is that S is finite. Then there exists s_1 such that x_n is not a dominant term for each $n \geq s_1$. Since x_{s_1} is not a dominant term, there exists $s_2 > s_1$ such that $x_{s_1} < x_{s_2}$. Given that x_{s_2} is not a dominant term, there exists $s_3 > s_2$ such that $x_{s_2} < x_{s_3}$. Continuing in this way, we can construct an increasing subsequence.

(c) Let $\{x_n\}$ be a bounded sequence of real numbers. Property (b) tells us that $\{x_n\}$ contains at least one monotone subsequence. Clearly, this subsequence must be bounded. Thus, by Property (a) this subsequence is convergent. $\square$

Lemma 2.6

Properties of a sequence in $(\mathbb{R}^m,\, d)$.

(a) A sequence $\{x_n\}$ in $\mathbb{R}^m$ converges to a vector $x = (x^1,\, \ldots,\, x^m)$ if and only if each coordinate sequence $\{x_n^i\}$ converges to x^i for each $i = 1,\, \ldots,\, m$.

(b) A sequence $\{x_n\}$ in $\mathbb{R}^m$ is bounded if and only if each coordinate sequence $\{x_n^i\}$ is bounded for each $i = 1, \ldots, m$.

(c) (Bolzano-Weirstrass Theorem) Every bounded sequence in $\mathbb{R}^m$ contains a convergent subsequence.

2.3 Open and closed sets

Definition 2.9: Open and closed sets

Let (X, d) be a metric space. The set A in X is open if and only if for every $x \in A$, there exists $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq A$. The set A is closed if and only if $X \setminus A$ is open.

Example 2.2

Following are examples of open sets.

(a) $\varnothing$ and X are open sets in X.

(b) Any open interval (a, b) of $\mathbb{R}$ with the usual metric $d(x, y) = |x - y|$ is an open set.

(c) The subset $\{(x, y) \in \mathbb{R}^2 | \, x > 1, \, y < 1\}$ of $\mathbb{R}^2$ with the metric d_2 is an open set.

(d) An open ball in $\mathbb{R}^n$ is an open set. Take an open ball $B_r(x)$ for $x \in X$, and $y \in B_r(x)$. by the definition of an open ball, $r - d(x, y) > 0$. Take $s = \frac{1}{2}[r - d(x, y)]$ and $z \in B_s(y)$. Then we have

$$\begin{aligned} & d(y, z) < s = \frac{1}{2}[r - d(x, y)], \\ \Rightarrow\ & 2d(y, z) + d(x, y) < r, \\ \Rightarrow\ & d(y, z) + d(x, y) < r, \\ \Rightarrow\ & d(x, z) \leq d(y, z) + d(x, y) < r. \end{aligned}$$

Hence, $B_s(y) \subseteq B_r(x)$.

Example 2.3

Following are the examples of closed sets.

(a) $\varnothing$ and X are closed sets in X.

(b) Any closed interval $[a, b]$ of $\mathbb{R}$ with the usual metric $d(x, y) = |x - y|$ is a closed set.

(c) The subset $\{(x, y) \in \mathbb{R}^2 | \, x^2 + y^2 \leq r\}$ of $\mathbb{R}^2$ with the metric d_2 is a closed set.

(d) A closed ball in $\mathbb{R}^n$ is a closed set.

Lemma 2.7

Properties of open (closed) sets.

(a) If $\{A_i\}_{i \in I}$ is an arbitrary collection of open (closed) sets, then $\cup_{i \in I} A_i$ is open ($\cap_{i \in I} A_i$ is closed).

(b) If $\{A_i\}_{i\in N}$ for $N = \{1, \ldots, n\}$ is a finite collection of open (closed) sets, then $\cap_{i\in N} A_i$ is open ($\cup_{i\in N} A_i$ is closed).

Proof. (a) If $x \in \cup_{i\in I} A_i$, then x belongs to some particular A_i. Since A_i is open, there exists some $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq A_i$, which implies that $B_\varepsilon(x) \subseteq \cup_{i\in I} A_i$.

(b) If $x \in \cap_{i\in N} A_i$, then x belongs to each A_i. Since each A_i is open, there exists an open ball $B_{\varepsilon_i}(x) \subseteq A_i$. Notice that the smallest such ball is contained in all the A_i's, and hence in their finite intersection. That is, if we take $\varepsilon = \min_i\{\varepsilon_i\}$, then $B_\varepsilon(x) \subseteq B_{\varepsilon_i}(x) \subseteq A_i$ for each $i \in N$, which implies that $B_\varepsilon(x) \subseteq \cap_{i\in N} A_i$. Thus, $\cap_{i\in N} A_i$ is open. $\square$

The finite intersection property is important. Consider $\{(-1, 1), (-1/2, 1/2), \ldots, (-1/n, 1/n), \ldots\}$ which is an infinite family of open sets whose intersection is $\{0\}$, which is a closed set.

Definition 2.10: Interior, Exterior, Boundary and Closure

Let (X, d) be a metric space, $A \subseteq X$ and $x \in X$.

(a) A point x_{int} is an interior point of A if there exists $\varepsilon > 0$ such that $B_\varepsilon(x_{int}) \subseteq A$. The set of all interior points of A is denoted by $int(A)$.

(b) A point x_{ext} is an exterior point of A if there exists $\varepsilon > 0$ such that $B_\varepsilon(x_{ext}) \subseteq A^c$. The set of all exterior points of A is denoted by $ext(A)$.

(c) A point x_{bd} is a boundary point of A if for any $\varepsilon > 0$, the open ball $B_\varepsilon(x_{bd})$ has non-empty intersection with both A and A^c, i.e., $B_\varepsilon(x_{bd}) \cap A \neq \varnothing$ and $B_\varepsilon(x_{bd}) \cap A^c \neq \varnothing$. The set of all boundary points of A is denoted by $bd(A)$.

(d) A point x_{cl} is a closure point of A if for any $\varepsilon > 0$, the open ball $B_\varepsilon(x_{cl})$ has a non-empty intersection with A, i.e., $B_\varepsilon(x_{cl}) \cap A \neq \varnothing$. The set of all closure points of A is denoted by $cl(A)$.

Since any interior point of A lies inside an open ball contained in A, we have $int(A) \subseteq A$. In the same manner, $ext(A) \subseteq A^c$. Now let an $x \in A$. Any open ball around it contains the point itself, and hence $x \in cl(A)$. Thus we have $int(A) \subseteq A \subseteq cl(A)$. It is also easy to see that $int(A)$, $bd(A)$ and $ext(A)$ constitute a partition of X, i.e., $int(A) \cup bd(A) \cup ext(A) = X$. Finally, we have $cl(A) = int(A) \cup bd(A)$ and $ext(A) = int(A^c)$. Using the above concepts, we get the following characterizations of open and closed sets.

Proposition 2.2

For an arbitrary set A,

(a) the set $int(A)$ is the largest open set contained in A,

(b) A is open if and only if $int(A) = A$,

(c) the set $cl(A)$ is the smallest closed set that contains A,

(d) A is closed if and only if $cl(A) = A$.

Proof. (a) We have shown that $int(A) \subseteq A$. Next take $x \in int(A)$. Then there exists $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq A$. Now, following Example 2.2(d) show that $y \in B_\varepsilon(x)$ implies that $y \in int(A)$. Thus, $B_\varepsilon(x) \subseteq int(A)$, and hence $int(A)$ is open. Finally, take a set B that is an open subset of A. We need to show that $B \subseteq int(A)$. Consider an $x \in B$. Since B is open, for an $\varepsilon > 0$ we have that $B_\varepsilon(x) \subseteq B \subseteq A$. Therefore, $x \in int(A)$

which implies that $B \subseteq int(A)$.

(b) If $int(A) = A$, then A is open because $int(A)$ is open. If A is open, then its largest open subset is A itself, and hence $int(A) = A$. $\square$

Example 2.4: Efficient production

Let $y = (y_1, y_2)$ be a production plan where $y_1 \in \mathbb{R}_-$ represents an input and $y_2 \in \mathbb{R}_+$ represents an output. For example, $y = (-2, 3)$ implies that 2 units of an input produce 3 units of output. The set $Y = \{y \in \mathbb{R}^2 \mid y \text{ is a production plan}\}$ is called the production possibility set, which is the set of all feasible production plans. A production plan y is efficient if and only if there is no $y' \in Y$ with $y' \geq y$, i.e., it is not possible to produce the same output with less input or more output with same input. Let $Eff(Y) = \{y \in Y \mid y' \geq y \;\Rightarrow\; y' \notin Y\}$ be the set of efficient productions of Y. Notice that every interior point of Y is an inefficient plan. Take an $y^0 \in int(Y)$. Then there exists an ε-ball $B_\varepsilon(y^0)$ around y^0 which is contained in Y. This ball contains at least a plan $y' \geq y^0$. Therefore, $y^0 \notin Eff(Y)$. Hence if $y \in Eff(Y)$, then $y \in bd(Y)$, i.e., $Eff(Y) \subseteq bd(Y)$. Convince yourselves that the converse is not true in general.

In what follows, we provide yet another characterization of closed sets.

Definition 2.11: Limit points

Let (X, d) be a metric space and A be a set in X. A point x in X is a limit (cluster) point of A if every open ball around it contains at least one point of A, which is distinct from x. The set of all limit points of A is called its derived set and is denoted by $D(A)$. Formally, $x \in D(A)$ if and only if for each $\varepsilon > 0$ we have $B_\varepsilon(x) \cap (A \setminus \{x\}) \neq \varnothing$.

Notice that the concept of limit points is more restrictive than that of the closure points. Define by $I(A) := cl(A) \setminus D(A)$, which is the set of all isolated points of A. In other words, a point y is in $I(A)$ if and only if there exists $\varepsilon > 0$ such that $B_\varepsilon(y) \cap A = \{y\}$. Thus, closure points of a set A are either its limit points or its isolated pointas, because $D(A) \cup I(A) = cl(A)$.

Proposition 2.3

Let (X, d) be a metric space and A be a set in X. A point $x \in X$ is in $D(A)$ if and only if there exists a sequence in $A \setminus \{x\}$ that converges to x.

Proof. First suppose that there is a sequence $\{x_n\} \subset A \setminus \{x\}$ that converges to x. Then for any given $\varepsilon > 0$ there exists an $n_\varepsilon \in \mathbb{N}$ such that $d(x_n, x) < \varepsilon$ for all $n \geq n_\varepsilon$, which implies that $x_n \in B_\varepsilon(x)$ for each $n \geq n_\varepsilon$. Thus we have $B_\varepsilon(x) \cap (A \setminus \{x\}) \neq \varnothing$ for this ε. Since this is true for any $\varepsilon > 0$, $x \in D(A)$.

Next suppose that $x \in D(A)$. Then for all $\varepsilon > 0$, $B_\varepsilon(x) \cap (A \setminus \{x\}) \neq \varnothing$. We will construct a sequence with the desired property. Take $\varepsilon = 1$. Since $B_1(x) \cap (A \setminus \{x\}) \neq \varnothing$, there is some $x_1 \in B_1(x) \cap (A \setminus \{x\})$. Next take $\varepsilon = 1/2$. Since $B_{1/2}(x) \cap (A \setminus \{x\}) \neq \varnothing$, there is some $x_2 \in B_{1/2}(x) \cap (A \setminus \{x\})$. Continuing this way, we can have some $x_n \in B_{1/n}(x) \cap (A \setminus \{x\})$. Thus, we have constructed a sequence $\{x_n\} \subset A \setminus \{x\}$ with the property that

$$0 \leq d(x_n, x) < \frac{1}{n}.$$

As $n \to \infty$, $\frac{1}{n} \to 0$ and $d(x_n, x) \to 0$. Therefore, $\{x_n\}$ converges to x. $\square$

Theorem 2.1

Let (X, d) be a metric space and A be a set in X. Then the following three statements are equivalent.

(a) A is closed.

(b) A contains all its limit points, i.e., $D(A) \subseteq A$.

(c) Every convergent sequence in A has its limit in A, i.e., if $x_n \in A$ for all n and $\{x_n\}$ converges to x, then $x \in A$.

Proof. First, we show the equivalence between (a) and (b). Assume that A is closed, then $X \setminus A$ is open. Then for any $x \in X \setminus A$, there exists some $\varepsilon > 0$ such that $B_\varepsilon(x) \subset X \setminus A$, which implies that $B_\varepsilon(x) \cap A = \varnothing$. Thus, no point in $X \setminus A$ can be a limit point of A, and hence all such points must be contained in A. To show the converse, suppose that $X \setminus A$ is not open, i.e., A is not closed. Then there is a point $x \in X \setminus A$ such that for every $\varepsilon > 0$ the open ball $B_\varepsilon(x)$ is not entirely contained in $X \setminus A$, which implies that $B_\varepsilon(x) \cap (A \setminus \{x\}) \neq \varnothing$. Thus $x \in D(A)$ which lies in $X \setminus A$, and hence A does not contain all its limit points.

Next, we show the equivalence between (a) and (c). Let A be closed. If a sequence of A has a limit then either this limit coincides with one of the elements of the sequence (and then it lies in A) or it is a limit point of A (which is contained in A since A is closed). Therefore A contains the limits of all its convergent sequences. Conversely, suppose that the limit of any convergent sequence of A lies in A. Every limit point of A is a limit of some sequence $\{x_n\}$ in A. Therefore A contains all its limit points, and hence is closed. $\square$

Example 2.5: Continuous preferences

Let $X \subseteq \mathbb{R}^2$ be the consumption set of an indiviual. A preference relation $\succsim$ is continuous if for any $x \in X$, its upper contour set $U_x = \{y \in X | \, y \succsim x\}$ and its lower contour set $L_x = \{y \in X | \, y \precsim x\}$ are both closed. The lexicographic preferences are discontinuous. To see this, consider the following sequences of commodity vectors $x^n = (1 + 1/n, 0)$ and $y^n = (1, 1)$ such that $x^n \succ y^n$ for all $n \in \mathbb{N}$. But $\lim_{n\to\infty} y^n = (1, 1) \succ (1, 0) = \lim_{n\to\infty} x^n$. Thus, the ordering is reversed at the limit, and it is discontinuous.

2.4 Continuous functions

Definition 2.12: Continuity

Let $f : S \longrightarrow T$, where $S \subset \mathbb{R}^n$ and $T \subset \mathbb{R}^m$. Then f is said to be continuous at $x \in S$ if for all $\varepsilon > 0$, there exists $\delta > 0$ such that $y \in S$ and $d(x, y) < \delta$ implies that $d(f(x), f(y)) < \varepsilon$. A function $f : S \longrightarrow T$ is said to be continuous on S if it is continuous at each point in S.

A function f is continuous at x if the value of f at any point y that is close to x is a good approximation of the value of f at x. Thus, the identity function $f(x) = x$ for all $x \in \mathbb{R}$ is continuous at each $x \in \mathbb{R}$, while the function $f : \mathbb{R} \longrightarrow \mathbb{R}$ given by

$$f(x) = \begin{cases} 1, & \text{if } x > 0, \\ -1, & \text{if } x \leq 0, \end{cases}$$

is continuous everywhere except at $x = 0$. At $x = 0$, every open ball $B_\delta(x)$ contains at least one $y > 0$. At all such points, $f(y) = 1 > -1 = f(x)$. Continuity can also be defined in terms of sequences as follows.

Definition 2.13: Continuity

Let $f : S \longrightarrow T$, where $S \subset \mathbb{R}^n$ and $T \subset \mathbb{R}^m$. Then f is said to be continuous at $x \in S$ if for all sequences $\{x_n\}$ such that $x_n \in S$ for all n, and $\lim_{n\to\infty} x_n = x$, it is the case that $\lim_{n\to\infty} f(x_n) = f(x)$.

Lemma 2.8: Composition of continuous functions

Composition of continuous functions is continuous. Let $f : S \longrightarrow T$ and $g : T \longrightarrow U$ be two functions, where $S \subset \mathbb{R}^n$, $T \subset \mathbb{R}^m$ and $U \subset \mathbb{R}^l$. If f is continuous at $x \in S$ and g is continuous at $f(x) \in T$, then $g \circ f$ is continuous at $x \in S$.

Theorem 2.2

Let $f : S \longrightarrow T$ be a function, where $S \subset \mathbb{R}^n$ and $T \subset \mathbb{R}^m$. The following three statements are equivalent.

(a) f is continuous on S.

(b) For every closed subset C of T, $f^{-1}(C) \subset S$ is closed.

(c) For every open subset O of T, $f^{-1}(O) \subset S$ is open.

Proof. First, we show that (a) implies (b). The inverse image of C is given by

$$f^{-1}(C) = \{x \in S | \, f(x) \in C \subset T\}.$$

Let C be an arbitrary closed subset of T and x be an arbitrary limit point of $f^{-1}(C)$. Then there exists a sequence $\{x_n\}$ in $f^{-1}(C)$ that converges to x. By continuity of f, $\{f(x_n)\}$ converges to $f(x)$. By construction, $\{f(x_n)\}$ is a sequence in C. Since C is closed, $f(x)$ must lie in C, which implies that $x \in f^{-1}(C)$. Thus, $f^{-1}(C)$ contains all its limit points, and hence is closed.

Second, we show that (b) implies (c). Notice that, for an arbitrary subset O of T, $f^{-1}(O) = S \setminus f^{-1}(T \setminus O)$. To show this, pick an $x \in f^{-1}(O)$, which implies that $f(x) \in O \subset T$. Therefore, $f(x) \notin T \setminus O$, which implies that $x \notin f^{-1}(O)$, and hence $x \in S \setminus f^{-1}(T \setminus O)$. Thus, $f^{-1}(O) \subset S \setminus f^{-1}(T \setminus O)$. On the other hand, pick an $x \in S \setminus f^{-1}(T \setminus O)$, which implies that $x \notin f^{-1}(T \setminus O)$, and hence $f(x) \notin T \setminus O$. Therefore, $f(x) \in O$ implying that $x \in f^{-1}(O)$. Thus, $S \setminus f^{-1}(T \setminus O) \subset f^{-1}(O)$. Therefore, $f^{-1}(O) = S \setminus f^{-1}(T \setminus O)$. Since O is an open subset of T, $T \setminus O$ is closed, and by continuity of f, $f^{-1}(T \setminus O)$ is closed. Therefore, $f^{-1}(O) = S \setminus f^{-1}(T \setminus O)$ is open.

Finally, we show that (c) implies (a). Pick an $x \in S$, and suppose that f is not continuous at x but the inverse image of an arbitrary subset O of T is open. We show that these will lead us to a contradiction. Since, by assumption, f is not continuous, there exist some $\varepsilon > 0$ and some $x^\delta \in S$ such that for all $\delta > 0$, $d(f(x^\delta), f(x)) \geq \varepsilon$ for $d(x^\delta, x) < \delta$, i.e., $f(x^\delta)$ does not lie in the open ball $B_\varepsilon(f(x))$ for $d(x^\delta, x) < \delta$, which implies that $x^\delta \notin f^{-1}(B_\varepsilon(f(x)))$, and hence $x^\delta \in S \setminus f^{-1}(B_\varepsilon(f(x)))$. Take $\delta = 1/n$. Thus, $x_n = x^\delta$. Since $d(x_n, x) \in [0, 1/n)$, $d(x_n, x) \to 0$ as $n \to \infty$. Thus, $x_n \to x$. Now, by assumption, $f^{-1}(B_\varepsilon(f(x)))$ is open since $B_\varepsilon(f(x))$ is open. Then the set $S \setminus f^{-1}(B_\varepsilon(f(x)))$ is closed. Since $\{x_n\}$ is a sequence in $S \setminus f^{-1}(B_\varepsilon(f(x)))$ which is a closed set, we have

$$x \in S \setminus f^{-1}(B_\varepsilon(f(x))). \tag{2.7}$$

Notice that $f(x) \in B_\varepsilon(f(x))$ since an open ball contains its center. This implies

$$x \in f^{-1}(B_\varepsilon(f(x))). \tag{2.8}$$

Thus, (2.7) and (2.8) contradict each other. $\square$

The above theorem states that a function is continuous if and only if the inverse image of any closed (open) set under it is closed (open). This is true for the inverse images of a continuous function. A continuous function does not necessarily map closed (open) sets into closed (open) sets, and a function that maps closed (open) sets into closed (open) sets is not necessarily continuous. Try to find examples for this assertion.

2.5 Compact metric spaces

Definition 2.14: Compactness

Consider the metric space $(\mathbb{R}^n, d)$. A set $A \subseteq \mathbb{R}^n$ is compact if it is closed and bounded.

Definition 2.15: Sequential compactness

Consider the metric space $(\mathbb{R}^n, d)$. A set $A \subseteq \mathbb{R}^n$ is sequentially compact if every sequence in A has a convergent subsequence whose limit lies in A.

It can be shown that a subset A of $\mathbb{R}^n$ is sequentially compact if and only if it is closed and bounded. Interested students should refer to Rudin (1976, Theorem 2.41).

Example 2.6

Following are the examples of compact sets.

(a) A finite set is compact.

(b) A closed interval $[a, b]$ of $\mathbb{R}$ is compact.

(c) The subset $[a, b]^n$ of $\mathbb{R}^n$ is compact.

Theorem 2.3

Suppose that $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is a continuous function. If $A \subset \mathbb{R}^n$ is compact, then $f(A) = \{y \in \mathbb{R} \mid y = f(x) \text{ for some } x \in A\}$ is a compact subset of $\mathbb{R}$.

Proof. Pick any sequence $\{y_n\}$ in $f(A)$. For each n, pick $x_n \in A$ such that $f(x_n) = y_n$. This gives us a sequence $\{x_n\}$ in A. Since A is compact, there is a subsequence $\{x_{n_k}\}$ of $\{x_n\}$, and a point $x \in A$ such that $\{x_{n_k}\}$ converges to x. Define $y = f(x)$ and $y_{n_k} = f(x_{n_k})$. By construction, $\{y_{n_k}\}$ is a subsequence of $\{y_n\}$. Moreover, since $x \in A$, it is the case that $y = f(x) \in f(A)$. Since, f is continuous, $y_{n_k} = f(x_{n_k})$ must converge to $y = f(x)$. Thus, $f(A)$ is compact. $\square$

Theorem 2.4: Weirstrass

Let $A \subset \mathbb{R}^n$ be compact, and $f : A \longrightarrow \mathbb{R}$ is continuous on A. Then f attains both its maximum and minimum on A, i.e., there exist points x_{max} and x_{min} in A such that

$$f(x_{max}) = \sup(f(A)) \quad \text{and} \quad f(x_{min}) = \inf(f(A)).$$

Proof. By the previous theorem, $f(A)$ is compact, and hence is a bounded subset of $\mathbb{R}$. Thus, $a = \sup(f(A))$ exists. Notice that, for each $n \in \mathbb{N}$, $a - 1/n$ is not an upper bound of $f(A)$. Thus we can find an $x_n \in A$ such

that $a \geq f(x_n) > a - 1/n$. Then $n \to \infty$ implies that $f(x_n) \to a$. Thus a is a limit point of $f(A)$. Since $f(A)$ is closed, $a \in f(A)$. Hence, there exists $x_{max} \in A$ such that $a = f(x_{max})$. The existence of the minimum is shown in a similar fashion. $\square$

Notice that compactness of A and continuity of f are indispensable for Theorem 2.4. Consider the following two cases.

(a) Let $f(x) = \frac{1}{x}$ and $A = (0,\ 1]$. It is easily seen that f does not have a maximizer since A is not compact.

(b) Let $A = [0,\ 1]$ and f is given by

$$f(x) = \begin{cases} 0, & \text{if } x = 0, \\ \frac{1}{x}, & \text{otherwise.} \end{cases}$$

Then f does not have a maximizer since the function is not continuous.

2.6 Fixed point theorems

2.6.1 Intermediate value theorem

In this section we introduce an important property pertaining to the continuous functions: the intermediate value theorem that will be useful for the fixed point theorems.

Theorem 2.5: Intermediate value theorem in $\mathbb{R}$

Let $f : [a,\ b] \longrightarrow \mathbb{R}$ be a continuous function, where $[a,\ b]$ is a compact interval of $\mathbb{R}$. If $f(a) < f(b)$, then for every real number $\gamma \in (f(a),\ f(b))$ there exists $c \in (a,\ b)$ such that $f(c) = \gamma$.

Proof. Let $S := \{x \in [a,\ b] \mid f(x) \leq \gamma\}$ and $c := \sup(S)$. Notice that S is non-empty since it contains at least a, and is bounded above by b. Hence, c is well defined by the supremum property. Notice also that, by construction, $f(x) > \gamma$ for all $x \in (c,\ b]$. We will show that $f(c) = \gamma$. Suppose first that $f(c) > \gamma$, i.e., $f(c) - \gamma > 0$. Since f is continuous, there exists $\delta > 0$ such that $|f(x) - f(c)| < \varepsilon$ for every $x \in (c - \delta,\ c + \delta)$. Choose $\varepsilon = f(c) - \gamma > 0$. Then $f(x) > f(c) - [f(c) - \gamma] = \gamma$ for every $x \in (c - \delta,\ c + \delta)$. Thus, $c - \delta$ is an upperbound of S, contradicting the fact that $c = \sup(S)$. Now suppose that $f(c) < \gamma$, i.e., $\gamma - f(c) > 0$. Continuity of f, for $\varepsilon = \gamma - f(c) > 0$, implies that $f(x) < f(c) + [\gamma - f(c)] = \gamma$ for every $x \in (c - \delta,\ c + \delta)$. Thus, there are points x strictly greater than c for which $f(x) < \gamma$ which contradicts the definition of c. Therefore, $f(c) = \gamma$. $\square$

2.6.2 Brouwer's fixed point theorem

Definition 2.16: Fixed point

Given a set X and a function $f : X \longrightarrow X$, the point x^* is said to be a fixed point of f if $x^* = f(x^*)$.

Our objective is to look for sufficient conditions under which a function from a set into itself has a fixed point. To fix ideas, let $f : [0,\ 1] \longrightarrow [0,\ 1]$ be defined by $f(x) = x^2$. Notice that $f(0) = 0$, $f(1) = 1$, and $f(x) < x$ (i.e., the graph of f lies strictly below the diagonal of the unit square) for all $x \in (0,\ 1)$. Thus, the set of fixed points of f is given by $\mathcal{E}(f) = \{0,\ 1\}$. Next, consider $g : [0,\ 1] \longrightarrow [0,\ 1]$ be defined by $g(x) = 1/4 + (1/2)x$. The set fixed points of g is given by $\mathcal{E}(g) = \{1/2\}$. Now consider $h : [0,\ 1] \longrightarrow [0,\ 1]$ be defined as $h(x) = 3/4$ for $x \in [0,\ 1/2]$, and $h(x) = 1/4$ for $x \in (1/2,\ 1]$. It is easy to check that the set

fixed points of h is given by $\mathcal{E}(h) = \varnothing$. The function h fails to have a fixed point because of its discontinuity at $x = 1/2$. The following theorem guarantees the existence of a fixed point.

Theorem 2.6: Brouwer, 1912

Let $f : [0, 1] \longrightarrow [0, 1]$ be a continuous function. Then the set of fixed points of f given by $\mathcal{E}(f) = \{x \in [0, 1] \mid x = f(x)\}$ is non-empty.

The above theorem is a trivial consequence of the intermediate value theorem in $\mathbb{R}$. Let $f : [0, 1] \longrightarrow [0, 1]$ be a continuous function. Define by $g(x) := f(x) - x$. Notice that the function $g : [0, 1] \longrightarrow [0, 1]$ is continuous on with $g(0) \geq 0$ and $g(1) \leq 0$. If one of these two expressions holds with equality, then either 0 or 1 is a fixed point of f. Otherwise, by the intermediate value theorem, there is $x^* \in (0, 1)$ such that $g(x^*) = 0$ implying that $f(x^*) = x^*$, and hence x^* is a fixed point of f.

Chapter 3

Differential calculus

3.1 Differentiable functions

First, we revise the concept of differentiability of a real valued function.

Definition 3.1

Let $f : S \longrightarrow \mathbb{R}$ be a function where $S \subseteq \mathbb{R}$. The function f is differentiable at $x \in S$ if

$$f'(x) = \lim_{y \to x} \frac{f(y) - f(x)}{y - x} \in \mathbb{R}$$

exists. The function f is differentiable on S if it is differentiable at each $x \in S$.

Lemma 3.1

Let $f : S \longrightarrow \mathbb{R}$ be a function where $S \subseteq \mathbb{R}$. If f is differentiable at a point x, then it is continuous at x.

Proof. Take two points x and $x + h$ in S. Hence,

$$\lim_{h \to 0}[f(x+h) - f(x)] = \left[\lim_{h \to 0} \frac{f(x+h) - f(x)}{h}\right]\left[\lim_{h \to 0} h\right] = f'(x).0 = 0.$$

The above implies that $\lim_{h \to 0} f(x+h) = f(x)$, and hence f is continuous at x. $\square$

The converse of the above lemma is not necessarily true. The function $f(x) = |x|$ is continuous on $[-1,\ 1]$, but is not differentiable at $x = 0$.

Definition 3.2: Local maximizer

A point x^0 is a local maximizer of a function $f : S \longrightarrow \mathbb{R}$, where $S \subseteq \mathbb{R}$, if there exists some $\delta > 0$ such that $f(x^0) \geq f(x)$ for all $x \in B_\delta(x^0)$.

Theorem 3.1

Let $f : (a,\ b) \longrightarrow \mathbb{R}$ be differentiable on $(a,\ b)$ and x^0 be a local maximizer (minimizer) of f. Then $f'(x^0) = 0$.

Proof. Suppose that f has a local maximum at x^0. Then we have $f(x^0 + h) - f(x^0) \leq 0$ for all h with $|h| < \delta$,

and therefore,

$$\begin{aligned} \frac{f(x^0+h)-f(x^0)}{h} &\leq 0, \quad \text{for } h \in (0,\, \delta), \\ &\geq 0, \quad \text{for } h \in (-\delta,\, 0). \end{aligned}$$

Thus, we have

$$\lim_{h\to 0^+} \frac{f(x^0+h)-f(x^0)}{h} \leq 0, \quad \text{and} \quad \lim_{h\to 0^-} \frac{f(x^0+h)-f(x^0)}{h} \geq 0.$$

Differentiability of f implies that

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

and hence $f'(x^0) = 0$. $\square$

Theorem 3.2: Rolle's theorem

Let $f : [a,\, b] \longrightarrow \mathbb{R}$ be differentiable on $(a,\, b)$ such that $f(a) = f(b) = \alpha$. Then there exists $c \in (a,\, b)$ such that $f'(c) = 0$.

Proof. Because f is continuous on a compact set $[a,\, b]$, by Weirstrass Theorem, there exist two points x_{min} and x_{max} in $[a,\, b]$ such that $f(x_{min}) = \min\, f(x)$ and $f(x_{max}) = \max\, f(x)$. If $f(x_{min}) = f(x_{max}) = \alpha$, then f is constant, and hence $f'(x) = 0$ for all $x \in [a,\, b]$. Otherwise, $f(x_{min}) < \alpha$ for $x_{min} \in (a,\, b)$ and $f'(x_{min}) = 0$ (because x_{min} is a local minimizer) or $f(x_{max}) > \alpha$ for $x_{max} \in (a,\, b)$ and $f'(x_{max}) = 0$ (because x_{max} is a local maximizer), or both. $\square$

Theorem 3.3: Mean value theorem

Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be a differentiable function. If a and b are two points in $\mathbb{R}$ with $a < b$, then there exists some $c \in (a,\, b)$ such that

$$f'(c) = \frac{f(b)-f(a)}{b-a}.$$

Proof. Define the following function

$$g(x) = f(x) - f(a) - \frac{f(b)-f(a)}{b-a}(x-a).$$

Because g satisfies the assumptions of Rolle's theorem, there exists some point c in $(a,\, b)$ such that $g'(c) = 0$, i.e.,

$$g'(c) = f'(c) - \frac{f(b)-f(a)}{b-a} = 0.$$

The above completes the proof. $\square$ If a function $f : \mathbb{R} \longrightarrow \mathbb{R}$ is differentiable at a point $x \in \mathbb{R}$, its derivative at x, $f'(x)$ is interpreted as the slope of the tangent to the function at the point x. Let $g(y) = my + c$ be the tangent to $f(y)$ at x. Intuitively, the derivative of f at x is the best linear approximation of f around x by the function g. This motivates the following generalized notion of differentiability.

Definition 3.3: Differentiability

Let $f : S \longrightarrow \mathbb{R}^m$ be a function where S is an open set in $\mathbb{R}^n$. The function f is differentiable at $x \in S$ if there exists an $m \times n$ matrix M such that for all $\varepsilon > 0$, there exists a $\delta > 0$ such that $y \in S$ and $||x-y|| < \delta$ implies

$$||f(x) - f(y) - M(x-y)|| < \varepsilon||x-y||.$$

Equivalently, f is differentiable at $x \in S$ if

$$\lim_{y \to x} \frac{||f(y) - f(x) - M(y-x)||}{||y-x||} = 0.$$

The function f is differentiable on S if it is differentiable at each $x \in S$.

The matrix M is called the derivative of f at x and is denoted $Df(x)$. In case of $n = m = 1$, we denote $Df(x)$ by $f'(x)$. Let $g : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ an affine function of the form $g(y) = My + c$, where M is an $m \times n$ matrix and $c \in \mathbb{R}^m$. The derivative of f at x is the best affine approximation of f around the point x by the function g. Here, the best means the ratio

$$\frac{||f(y) - g(y)||}{||y-x||}$$

goes to zero as $y \to x$. Since the values of f and g must coincide at x, we must have $g(x) = Mx + c = f(x)$ or $c = f(x) - Mx$. Thus, we may write the approximation function g as

$$g(y) = My - Mx + f(x) = M(y-x) + f(x).$$

Given this value for $g(y)$, the task of identifying the best affine approximation to f at x now amounts to identifying a matrix M such that

$$\frac{||f(y) - g(y)||}{||y-x||} = \frac{||f(y) - f(x) - M(y-x)||}{||y-x||} \to 0 \quad \text{as} \quad y \to x.$$

This is precisely the definition of derivative given above.

When f is differentiable on S, the derivative Df itself forms a function from S to $\mathbb{R}^{m \times n}$. If $Df : S \longrightarrow \mathbb{R}^{m \times n}$ is a continuous function, then f is said to *continuously differentiable* on S, and we write f is $\mathcal{C}^1$. Consider now the following example.

Example 3.1

Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be given by

$$f(x) = \begin{cases} 0, & \text{if } x = 0, \\ x^2 \sin(1/x^2), & \text{if } x \neq 0. \end{cases}$$

For $x \neq 0$, we have

$$f'(x) = 2x \sin\left(\frac{1}{x^2}\right) - \left(\frac{2}{x}\right) \cos\left(\frac{1}{x^2}\right).$$

Since $|\sin(\cdot)| \leq 1$ and $|\cos(\cdot)| \leq 1$, but $2/x \to \infty$ as $x \to 0$, it is clear that $\lim_{x \to 0} f'(x)$ is not well defined. However, $f'(0)$ does exist! Indeed

$$f'(0) = \lim_{x \to 0} \frac{f(x) - f(0)}{x} = \lim_{x \to 0} x \sin\left(\frac{1}{x^2}\right) = 0.$$

The above implies that f is differentiable at $x = 0$, but Df is not continuous at this point. Thus, f is not $\mathcal{C}^1$ on $\mathbb{R}_+$.

Next, given functions $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ and $h : \mathbb{R}^k \longrightarrow \mathbb{R}^n$, their composition is given by the function $f \circ h : \mathbb{R}^k \longrightarrow \mathbb{R}^m$ whose value at any $x \in \mathbb{R}^k$ is given by $f(h(x))$. Then

Lemma 3.2: Chain rule of derivative

Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ and $h : \mathbb{R}^k \longrightarrow \mathbb{R}^n$, and let $x \in \mathbb{R}^k$. If h is differentiable at x, and f is differentiable at $h(x)$, then $f \circ h$ is itself differentiable at x, and its derivative is obtained through the "chain rule" as:

$$D(f \circ h)(x) = Df(h(x))Dh(x).$$

3.2 Partial derivatives

Definition 3.4: Partial derivative

Let $f : S \longrightarrow \mathbb{R}$, where $S \subset \mathbb{R}^n$ is open. Let e_j denote the vector in $\mathbb{R}^n$ that has a 1 in the j-th place and zeros elsewhere ($j = 1, \ldots, n$). Then the j-th partial derivative of f is said to exist at a point x if there is a number $\partial f(x)/\partial x_j$ such that

$$\lim_{t \to 0} \frac{f(x + te_j) - f(x)}{t} = \frac{\partial f}{\partial x_j}(x) \text{ or } f_j(x).$$

For the partial derivatives, the following theorem is true.

Theorem 3.4

Let $f : S \longrightarrow \mathbb{R}$, where $S \subset \mathbb{R}^n$ is open. Define the gradient vector of f at x by the vector of partial derivatives of f at x as $\nabla f(x) := [f_1(x), \ldots, f_n(x)]$.

(a) If f is differentiable at x, then all partials $f_j(x)$ exist at x, and $Df(x) = \nabla f(x)$.

(b) If all the partials exist and are continuous at x, then the derivative of f at x exists and is given by $Df(x) = \nabla f(x)$.

(c) f is $\mathcal{C}^1$ on S if and only if all partials $f_j(x)$ exist and are continuous on S.

Thus to check if f is $\mathcal{C}^1$, we only need to figure out if (a) the partial derivatives all exist on S, and (b) if they are all continuous on S. On the other hand, the requirement that the partial derivatives not only exist but be continuous at x is very important for the coincidence of the vector of partials with $Df(x)$. In the absence of this condition, all partials could exist at some point without the function itself being differentiable at that point. Consider the following example.

Example 3.2

Let $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ be given by

$$f(x, y) = \begin{cases} 0, & \text{if } (x, y) = (0, 0), \\ \frac{xy}{\sqrt{x^2+y^2}}, & \text{if } (x, y) \neq (0, 0). \end{cases}$$

We will show that f has all partial derivatives everywhere, but that these partials are not continuous at $(0, 0)$. Then we will show that f is not differentiable at $(0, 0)$. Since $f(x, 0) = 0$ for any $x \neq 0$, it is immediate that for all $x \neq 0$,

$$\frac{\partial f}{\partial y}(x, 0) = \lim_{\hat{y} \to 0} \frac{f(x, \hat{y}) - f(x, 0)}{\hat{y}} = \lim_{\hat{y} \to 0} \frac{x}{\sqrt{x^2 + \hat{y}^2}} = 1.$$

Similarly, at all points $(0,\, y)$ for $y \neq 0$, we have $\partial f(0,\, y)/\partial x = 1$. However, note that

$$\frac{\partial f}{\partial x}(0,\, 0) = \lim_{x \to 0} \frac{f(x,\, 0) - f(0,\, 0)}{x} = 0 = \lim_{y \to 0} \frac{f(0,\, y) - f(0,\, 0)}{y} = \frac{\partial f}{\partial y}(0,\, 0).$$

So, $\partial f(0,\, 0)/\partial x$ and $\partial f(0,\, 0)/\partial y$ exist at $(0,\, 0)$. But

$$\lim_{x \to 0} \frac{\partial f}{\partial y}(x,\, 0) = \lim_{y \to 0} \frac{\partial f}{\partial x}(0,\, y) = 1 \neq 0.$$

Thus, the partials are not continuous at $(0,\, 0)$. Now suppose that f were differentiable at $(0,\, 0)$. Then we must have $Df(0,\, 0) = (0,\, 0)$. Take the points $(x,\, y)$ of the form $(a,\, a)$ for some $a > 0$, and note that every open neighborhood of $(0,\, 0)$ must contain at least one such point. Since $f(0,\, 0) = 0$, $Df(0,\, 0) = (0,\, 0)$ and $||(x,\, y)|| = \sqrt{x^2 + y^2}$, we have

$$\lim_{a \to 0} \frac{||f(a,\, a) - f(0,\, 0) - Df(0,\, 0).(a,\, a)||}{||(a,\, a) - (0,\, 0)||} = \lim_{a \to 0} \frac{a^2}{2a^2} = \frac{1}{2} \neq 0.$$

Thus, f is not differentiable at $(0,\, 0)$.

The failure of the existence of derivative in the above example induces a generalized notion of derivative which is studied in the following section. In what follows we extend the concept of derivative of a vector-valued function.

Definition 3.5: Jacobian matrix

Let $f : S \longrightarrow \mathbb{R}^m$, where $S \subseteq \mathbb{R}^n$ is open, assigns to each $x = (x_1,\, \ldots,\, x_n) \in \mathbb{R}^n$ a vector $f(x) = (f^1(x),\, \ldots,\, f^m(x))$ in $\mathbb{R}^m$. The Jacobian matrix of f at $x \in S$ is the $m \times n$ matrix of partial derivatives which is given by

$$J_f(x) := \begin{bmatrix} \frac{\partial f^1}{\partial x_1}(x) & \ldots & \frac{\partial f^1}{\partial x_n}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial f^n}{\partial x_1}(x) & \ldots & \frac{\partial f^m}{\partial x_n}(x) \end{bmatrix}$$

Following is an extension of Theorem 3.4 in case of a vector-valued function.

Theorem 3.5

Let $f : S \longrightarrow \mathbb{R}^m$, where $S \subset \mathbb{R}^n$ is open.

(a) If f is differentiable at x, then all partials $\partial f^i/\partial x_j$ for $i = 1,\, \ldots,\, m$ and $j = 1,\, \ldots,\, n$ exist at x, and $Df(x) = J_f(x)$.

(b) If all the partials $\partial f^i/\partial x_j$ for $i = 1,\, \ldots,\, m$ and $j = 1,\, \ldots,\, n$ exist and are continuous at x, then the derivative of f at x exists and is given by $Df(x) = J_f(x)$.

(c) f is $\mathcal{C}^1$ on S if and only if all partials $\partial f^i/\partial x_j$ for $i = 1,\, \ldots,\, m$ and $j = 1,\, \ldots,\, n$ exist and are continuous on S.

Example 3.3

Let $f : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ such that $f(x,\, y) = (x^2 + x^2y + 10y,\, x + y^3)$. The Jacobian of f at $(x,\, y) \in \mathbb{R}^2$ is

given by

$$Df(x,\, y) = \begin{bmatrix} 2x(1+y) & x^2+10 \\ 1 & 3y^2 \end{bmatrix}.$$

3.3 Directional derivatives

Definition 3.6: Directional derivative

et $f : S \longrightarrow \mathbb{R}$, where $S \subset \mathbb{R}^n$ is open. Let x be a point in S and let $h \in \mathbb{R}^n$. The directional derivative of f at x in the direction h is defined as

$$Df(x;\, h) = \lim_{t \to 0} \frac{f(x+th) - f(x)}{t}, \quad \text{where } \; t \in \mathbb{R} \; \text{ and } \; ||h|| = 1,$$

whenever this limit exists.

Theorem 3.6

Suppose f is differentiable at $x \in S$. Then, for any $h \in \mathbb{R}^n$, the directional derivative $Df(x;\, h)$ of f at x in the direction h exists, and we have $Df(x;\, h) = \nabla f(x) \cdot h$.

Example 3.4

Let $f(x_1,\, x_2) = x_1 x_2$, $h = (3/5,\, 4/5)$ and $x^0 = (1,\, 2)$. First we compute $Df(x^0;\, h)$, and then verify the above result. The directional derivative of f at x in the direction h is given by

$$Df(x_1,\, x_2;\, h) = \lim_{t \to 0} \frac{f(x+th) - f(x)}{t} = \lim_{t \to 0} \frac{\left(x_1 + \frac{3t}{5}\right)\left(x_2 + \frac{4t}{5}\right) - x_1 x_2}{t} = \frac{4x_1}{5} + \frac{3x_2}{5}.$$

Therefore,

$$Df(x^0;\, h) = D(1,\, 2;\, (3/5,\, 4/5)) = 2.$$

On the other hand,

$$\nabla f(x_1,\, x_2) = (x_2,\, x_1)$$

Therefore,

$$\nabla f(x_1^0,\, x_2^0) \cdot h = (2,\, 1) \begin{pmatrix} 3/5 \\ 4/5 \end{pmatrix} = 2 = Df(1,\, 2;\, h).$$

3.4 Higher-order derivatives

We have discussed earlier that, for a function $f : S \longrightarrow \mathbb{R}$ where $S \subset \mathbb{R}^n$ is open, which is differentiable on S, the derivative Df is itself a function from S to $\mathbb{R}^n$. Suppose now that there is an $x \in S$ such that Df is differentiable at x, i.e., such that for each $i = 1,\, \ldots,\, n$, the function $f_j : S \longrightarrow \mathbb{R}$ is differentiable at x. Denote the partial derivative of f_i in the direction e_j at x by $f_{ij}(x)$ or $\partial^2 f(x)/\partial x_j \partial x_i$ if $i \neq j$, and by $f_{ii}(x)$

or $\partial^2 f(x)/\partial^2 x_j$ if $i = j$. The *Hessian matrix* of f at x is given by

$$H[f(x)] := \begin{bmatrix} f_{11}(x) & \dots & f_{1n}(x) \\ \vdots & \ddots & \vdots \\ f_{n1}(x) & \dots & f_{nn}(x) \end{bmatrix}$$

Definition 3.7

A function $f : S \longrightarrow \mathbb{R}$, where $S \subset \mathbb{R}^n$ is open, is twice-differentiable at x if the second derivative $D^2 f(x)$ equals the Hessian matrix of f at x, i.e., $D^2 f(x) = H[f(x)]$. For $n = 1$, we denote $D^2 f(x)$ by $f''(x)$. If f is twice-differentiable at each $x \in S$, then f is twice-differentiable on S. If for each i, the cross partial $f_{ij} : S \longrightarrow \mathbb{R}$ is continuous, then f is twice continuously differentiable on S, and we write f is $\mathcal{C}^2$.

Theorem 3.7: Young's theorem

If $f : S \longrightarrow \mathbb{R}$ is a $\mathcal{C}^2$ function, then $D^2 f$ is a symmetric matrix, i.e., we have

$$f_{ij}(x) = f_{ji}(x) \quad \text{for all} \quad i,\, j = 1,\, \dots,\, n, \quad \text{and} \quad x \in S.$$

The above asserts a one-way implication. The matrix $D^2 f$ may fail to be symmetric if a function is not $\mathcal{C}^2$.

3.5 Taylor's theorem

In this section we discuss a generalization of the Mean value theorem, known as Taylor's theorem. The idea is that a many times differentiable function can be approximated by a polynomial. The notation $f^{(k)}(z)$ denotes the k-th derivative of f at a point z, and $k = 0$ implies that $f^{(k)}(z) = f(z)$.

Theorem 3.8: Taylor's theorem in $\mathbb{R}$

Let $f : (a,\, b) \longrightarrow \mathbb{R}$ be an m-times continuously differentiable function. Suppose also that $f^{(m+1)}(z)$ exists for every $z \in (a,\, b)$. Then for any $x,\, y \in (a,\, b)$, there is a $z \in (x,\, y)$ such that

$$f(y) = \sum_{k=0}^{m} \frac{f^{(k)}(x)(y-x)^k}{k!} + \frac{f^{(m+1)}(z)(y-x)^{(m+1)}}{(m+1)!}.$$

Example 3.5

We would like to approximate $f(y) = e^y$ around $x = 0$ by a polynomial $P_m(y)$. Notice that $f^{(k)}(x) = e^x$ for all $k = 0,\, \dots,\, m$. Then $f^{(k)}(0) = 1$ for all $k = 0,\, \dots,\, m$. Thus, applying Taylor's theorem for $\mathbb{R}$ and ignoring the remainder term, we have

$$e^y \approx 1 + y + \frac{y^2}{2!} + \dots + \frac{y^m}{m!} \equiv P_m(y).$$

Taylor's theorem gives us a formula for constructing a polynomial approximation to a differentiable function. For $m = 0$, we obtain the Mean Value Theorem. With $m = 2$, and omitting the remainder, we get

$$f(x+h) \approx f(x) + f'(x)h.$$

With f differentiable, the remainder term will be very small. Thus, the linear function on the right-hand-side

of the above equation seems to be a good approximation to $f(\cdot)$ around x. Following is a generalization of the above theorem.

Theorem 3.9: Taylor's theorem in $\mathbb{R}^n$

Let $f : S \longrightarrow \mathbb{R}$ be a $\mathcal{C}^1$ function, where $S \subset \mathbb{R}^n$ is open. Then for any $x,\, y \in S$, we have

$$f(y) = f(x) + Df(x) \cdot (y - x) + R_1(x,\, y), \quad \text{where} \quad \lim_{y \to x} \frac{R_1(x,\, y)}{||y - x||} = 0.$$

Proof. See Sundaram (1996, pp. 64-65). □

Example 3.6

Let i, r and π denote the nominal rate of interest, the real rate of interest and the rate of inflation. The nominal rate of interest is given by the formula $1+i = (1+r)(1+\pi)$. Define by $f(r,\, \pi) = (1+r)(1+\pi)$. Let $(r^0,\, \pi^0) = (0,\, 0)$. Notice that $Df(0,\, 0) = (1,\, 1)$. Then by Taylor's expansion of $f(r,\, \pi)$ around $(r^0,\, \pi^0)$ we have

$$1 + i = f(r,\, \pi) \approx f(r^0,\, \pi^0) + Df(r^0,\, \pi^0) \cdot (r - r^0,\, \pi - \pi^0) = 1 + (1,\, 1)^T \cdot (r,\, \pi) = 1 + r + \pi$$

The above implies $i \approx r + \pi$.

Example 3.7: Rule of 70

With compound interest rate, the time it takes for an initial investment to double is $70/100r$ years. Let T is the time taken, i.e.,

$$\begin{aligned} & (1 + r)^T I = 2I \\ \implies & T \ln(1 + r) = \ln 2 \\ \implies & T = \frac{\ln 2}{\ln(1 + r)} \approx \frac{0.693147}{r} = \frac{70}{100r} \end{aligned}$$

since $\ln(1 + r) \approx r$.

Definition 3.8: Total derivative

$f : S \longrightarrow \mathbb{R}$ be a $\mathcal{C}^1$ function, where $S \subset \mathbb{R}^n$. The total derivative f at $x \in S$ is defined as

$$df(x) = \nabla f(x) \cdot dx = \sum_{i=1}^{n} f_i(x)\, dx_i.$$

Example 3.8: Indifference curves

Let $u : \mathbb{R}^2_+ \longrightarrow \mathbb{R}$ be the continuously differentiable utility function of a consumer derived from the consumption of two goods 1 and 2 in quantities $x = (x_1,\, x_2)$. The indifference curve at level α is the set $\{x \in \mathbb{R}^2_+ \mid u(x) = \alpha\}$. The total derivative of u at x is given by

$$du(x) = u_1(x)dx_1 + u_2(x)dx_2 = 0.$$

The above equation implies that

$$\frac{dx_2}{dx_1} = -\frac{u_1(x)}{u_2(x)} = MRS_{12}(x)\,.$$

Given that the marginal utilities are positive, the indifference curve between goods 1 and 2 is negatively sloped.

3.6 Inverse and implicit function theorems

Given two sets A and B, if a function $f : A \longrightarrow B$ is one-to-one and onto, then there is a unique function $f^{-1} : B \longrightarrow A$ such that $f(f^{-1}(b)) = b$ for all $b \in B$. The function g is called the *inverse function* of f.

Theorem 3.10: Inverse function theorem

Let $f : S \longrightarrow \mathbb{R}^n$ be a $\mathcal{C}^1$ function, where $S \subset \mathbb{R}^n$ is open. Suppose there is a point $x \in S$ such that the $n \times n$ matrix $Df(x)$ is invertible. Let $y = f(x)$. Then

(a) There are open sets U and V in $\mathbb{R}^n$ such that $x \in U$, $y \in V$, f is one-to-one on V, and $f(U) = V$.

(b) The inverse function $f^{-1} : V \longrightarrow U$ of f is a $\mathcal{C}^1$ function, whose derivative at any point $y^0 \in V$ satisfies

$$Df^{-1}(y^0) = (Df(x^0))^{-1}, \quad \text{where} \quad f(x^0) = y^0.$$

Example 3.9

Let $f : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ be defined by $f(x, y) = (x^2 + x^2y + 10y, x + y^3)$. We will show that f has an inverse in the neighborhood of $(1, 1)$. We have

$$Df(x, y) = J_f(x, y) = \begin{bmatrix} 2x(1+y) & x^2 + 10 \\ 1 & 3y^2 \end{bmatrix}$$

Thus, $f(1, 1) = (12, 2)$ and

$$Df(1, 1) = \begin{bmatrix} 4 & 11 \\ 1 & 3 \end{bmatrix},$$

and $det(Df(1, 1)) = 1 \neq 0$. Therefore, $Df(1, 1)$ is invertible. By the Inverse function theorem, we deduce that there is an open set $U \subset \mathbb{R}^2$ containing $(1, 1)$ such that f when restricted to U has a continuously differentiable inverse f^{-1}, and

$$Df^{-1}(1, 1) = \begin{bmatrix} 3 & -11 \\ -1 & 4 \end{bmatrix} = (Df(1, 1))^{-1}.$$

Now, consider the function $F : \mathbb{R}^2 \longrightarrow \mathbb{R}$ defined by $F(x, y) = (x - 2)^3y + xe^{y-1}$, and suppose we are interested in solving the equation $F(x, y) = 0$. We will ask the question whether it is possible to define y as a function of x in some neighborhood of (x^*, y^*). This question motivates the following theorem. We introduce some additional notations. Given integers $m \geq 1$ and $n \geq 1$, let a typical point in $\mathbb{R}^{m+n}$ be denoted by (x, y), where $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$. For a $\mathcal{C}^1$ function F mapping some subset of $\mathbb{R}^{m+n}$ into $\mathbb{R}^n$, let $DF_y(x, y)$ denote the portion of the matrix $DF(x, y)$, which is an $n \times (m + n)$ matrix, corresponding to the last n variables. Notice that $DF_y(x, y)$ is a $n \times n$ matrix. Define $DF_x(x, y)$ similarly, which is an $n \times m$ matrix.

Theorem 3.11: Implicit function theorem

Let $F : S \longrightarrow \mathbb{R}^n$ be a $\mathcal{C}^1$ function, where $S \subset \mathbb{R}^{m+n}$ is open. Let (x^*, y^*) be a point in S such that $DF_y(x^*, y^*)$ is invertible, and let $F(x^*, y^*) = c$. Then, there is a neighborhood $U \subset \mathbb{R}^m$ of x^* and a $\mathcal{C}^1$ function $g : U \longrightarrow \mathbb{R}^n$ such that

(a) $(x, g(x))$ is in S for all $x \in U$,

(b) $g(x^*) = y^*$,

(c) $F(x, g(x)) = c$ for all $x \in U$.

The derivative of g at any $x \in U$ is obtained from the chain rule:

$$Dg(x) = -\left(DF_y(x, y)\right)^{-1} DF_x(x, y).$$

Example 3.10

Consider the equation $F(x, y) = (x-2)^3 y + xe^{y-1} = 0$. We will show that y can be defined implicitly as a function of x in the neighborhood of $(0, 0)$ but not around $(1, 1)$. First notice that $F(0, 0) = F(1, 1) = 0$. We have $DF_y(x, y) = \partial F(x, y)/\partial y = (x-2)^3 + xe^{y-1}$. Now, $DF_y(0, 0) = -8 \neq 0$, and hence $DF_y(0, 0)$ is invertible. But $DF_y(1, 1) = 0$, and hence $DF_y(1, 1)$ is not invertible.

Part II

Convex analysis

Chapter 4

Convex sets and separation theorems

4.1 Convex sets

Convexity is often assumed in economic theory since it plays important roles in optimization. The convexity of preferences can be interpreted as capturing consumer's liking for variety, and the convexity of production set is related to the existence of non-increasing returns to scale.

Definition 4.1: Convex set

A set X in $\mathbb{R}^n$ is convex if given any two points x' and x'' in X, the point

$$x^\lambda = (1-\lambda)x' + \lambda x''$$

is also in X for every $\lambda \in [0,\ 1]$.

A vector of the form x^λ as defined above is called a *convex combination* of x' and x''. The set of all convex combinations of x' and x'' is the line segment connecting these two points, which is denoted by $[x',\ x'']$.

Example 4.1

Following are the examples of convex sets.

(a) Given a set X, and two points x' and x'' in X, the line segment $[x',\ x'']$ is a convex set.

(b) The disc $\{(x,\ y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq r\}$ is a convex set.

(c) The circle $\{(x,\ y) \in \mathbb{R}^2 \mid x^2 + y^2 = r\}$ is <u>not</u> a convex set.

(d) The set $H(p,\ \alpha) = \{x \in \mathbb{R}^n \mid p \cdot x = \alpha \ \text{ for } p \in \mathbb{R}^n \ \text{ and } \alpha \in \mathbb{R}\}$ is a convex set.

Lemma 4.1

Following are some useful properties relating to convex sets.

(a) An arbitrary intersection of convex sets is convex.

(b) If X and Y are convex sets in $\mathbb{R}^n$, and $\alpha \in \mathbb{R}$, then the sets

$$\begin{aligned} X + Y &= \{z \in \mathbb{R}^n \mid z = x + y \ \text{ for some } x \in X \ \text{ and } y \in Y\}, \ \text{ and} \\ \alpha X &= \{z \in \mathbb{R}^n \mid z = \alpha x \ \text{ for some } x \in X\} \end{aligned}$$

are convex.

(c) If X is a convex set in $\mathbb{R}^n$, then $int(X)$ and $cl(X)$ are also convex.

Proof. The proof is left as an exercise. $\square$

4.2 Convex hull

First we introduce a generalization of the notion of convex combination to more than two vectors in $\mathbb{R}^n$.

Definition 4.2: Convex combination

A point $y \in \mathbb{R}^n$ is said to be a convex combination of m vectors $x^1, \ldots, x^m \in \mathbb{R}^n$ if it can be written as

$$y = \sum_{i=1}^{m} \lambda_i x^i, \quad \text{with } \lambda_i \in [0,\, 1] \text{ for all } i \text{ and } \sum_{i=1}^{m} \lambda_i = 1.$$

The following theorem gives a characterization of convexity in terms of convex combinations.

Theorem 4.1

A set X is convex if and only if every convex combination of points in X lies in X.

We are sometimes interested in extending a set X so that is becomes convex by adding as few points to it as possible. The resulting set is called the convex hull of X.

Definition 4.3: Convex hull

Let X be a set in $\mathbb{R}^n$. The convex hull of X, denoted $co(X)$, is the smallest convex set that contains X.

Clearly, there is at least one convex set that contains X, namely, $\mathbb{R}^n$ itself. If there are more, $co(X)$ is the intersection of all such sets.

Theorem 4.2

The convex hull of X is the set of all convex combinations of elements in X, i.e.,

$$co(X) = \left\{ y \in \mathbb{R}^n \mid y = \sum_{i=1}^{m} \lambda_i x^i \text{ for some } m, \text{ with } x^i \in X,\ \lambda_i \in [0,\, 1] \text{ for all } i,\ \sum_{i=1}^{m} \lambda_i = 1 \right\}.$$

Proof. Let Y be the set of all convex combinations of elements in X. Clearly, Y contains X since any x in X can be written as a trivial convex combination with itself. Next we show that Y is a convex set. Let y^1 and y^2 be two points in Y such that

$$\begin{aligned} y^1 &= \sum_{i=1}^{m} \lambda_i x^i, \quad \text{with } \lambda_i \in [0,\, 1] \text{ for all } i \text{ and } \sum_{i=1}^{m} \lambda_i = 1, \\ y^2 &= \sum_{j=1}^{n} \mu_j x^j, \quad \text{with } \mu_j \in [0,\, 1] \text{ for all } j \text{ and } \sum_{j=1}^{n} \mu_j = 1. \end{aligned}$$

Then for some $\alpha \in [0, 1]$,

$$y = (1-\alpha)y^1 + \alpha y^2 = \sum_{i=1}^{m}(1-\alpha)\lambda_i x^i + \sum_{j=1}^{n}\alpha\mu_j x^j .$$

Since $(1-\alpha)\lambda_i \in [0, 1]$ for each i, $\alpha\mu_j \in [0, 1]$ for each j, and $\sum_{i=1}^{m}(1-\alpha)\lambda_i + \sum_{j=1}^{n}\alpha\mu_j = 1$. Thus, y is a convex combination of points in X, and hence is in Y. Moreover, any convex set that contains X must include all convex combinations of points in X, and must therefore contains Y. Thus, Y is the smallest convex set containing X, i.e, $Y = co(X)$. $\square$

The previous theorem says that any point in the convex hull of X can be expressed as a convex combination of a finite number of points in X, but it does not tell us how many such points are required. The following theorem says that if X is a subset of the n-dimensional Euclidean space, then this convex combination can be created with at most $n+1$ points in X.

Theorem 4.3: Caratheodory's theorem

Let $X \subseteq \mathbb{R}^n$. If y is a convex combination of points in X, then y is a convex combination of $n+1$ or fewer points in X.

The idea behind the above theorem is the following. Consider the set $X = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$, a subset of $\mathbb{R}^2$. The convex hull of this set is the square with vertices $(0, 0)$, $(0, 1)$, $(1, 0)$ and $(1, 1)$. Consider now the point $y = (1/4, 1/4)$, which is in $co(X)$. We can construct a set $X' = \{(0, 0), (0, 1), (1, 0)\}$, the convex hull of which is the triangle with vertices $(0, 0)$, $(0, 1)$ and $(1, 0)$ that contains y.

4.3 Separation theorems

In this section we analyze some important results regarding convex sets that have interesting applications in economic theory, especially in general equilibrium.

Definition 4.4: Hyperplane and half-spaces

Given $p \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$, the hyperplane generated by p and α is the set of points in $\mathbb{R}^n$ given by

$$H(p, \alpha) = \{x \in \mathbb{R}^n \mid p \cdot x = \alpha\}.$$

The above hyperplane $H(p, \alpha)$ divides $\mathbb{R}^n$ into two regions: an upper half-space and a lower half-space, which are respectively given by

$$\begin{aligned} H_U(p, \alpha) &= \{x \in \mathbb{R}^n \mid p \cdot x \geq \alpha\}, \\ H_L(p, \alpha) &= \{x \in \mathbb{R}^n \mid p \cdot x \leq \alpha\}. \end{aligned}$$

It is easy to show that $H(p, \alpha)$, $H_U(p, \alpha)$ and $H_L(p, \alpha)$ are convex sets.

Example 4.2

Following are the examples of hyperplanes.

(a) If $n = 1$, the hyperplane $H(p, \alpha) = \{x \in \mathbb{R} \mid px = \alpha\}$ is the set $\{\alpha/p\}$. The half-spaces are given by $H_U(p, \alpha) = [\alpha/p, \infty)$ and $H_L(p, \alpha) = (-\infty, \alpha/p]$.

(b) If $n = 2$, $p = (p_1, p_2) \in \mathbb{R}^2$ and $p_2 \neq 0$, then

$$H(p, \alpha) = \{(x_1, x_2) \in \mathbb{R}^2 \mid p_1 x_1 + p_2 x_2 = \alpha\}$$

is the set of points lying on the straight line with slope $-p_1/p_2$ that passes through the point $(0, \alpha/p_2)$.

Definition 4.5: Separating hyperplane

A hyperplane $H(p, \alpha) \subset \mathbb{R}^n$ separates two subsets X and Y of $\mathbb{R}^n$ if $X \subset H_U(p, \alpha)$ and $Y \subset H_L(p, \alpha)$, or vice-versa. Moreover, the separation is strict if $X \cap H(p, \alpha) = \varnothing$ and $Y \cap H(p, \alpha) = \varnothing$.

Example 4.3

Let $X = \{(x, y) \in \mathbb{R}^2 \mid y \geq x^2\}$ and $Y = \{(x, y) \in \mathbb{R}^2 \mid y \geq -x^2 - 1\}$. Then the hyperplane $H((0, 1), -1/2)$, i.e., the line $y = -1/2$ separates X and Y. Notice that $H((0, 1), -1/2)$ is one of the infinitely many hyperplanes that separate X and Y.

Lemma 4.2

Given $X \subset \mathbb{R}^n$ non-empty, closed and convex, and $x^* \notin X$, there exists a unique $y^* \in X$ such that

$$d(x^*, y^*) = \min\{d(x^*, y) \mid y \in X\}. \tag{4.1}$$

Proof. Let $x^* \notin X$ and $\bar{B}_r(x^*)$ be the closed ball with radius $r > 0$ and center x^*. Pick r sufficiently large so that $Y = \bar{B}_r(x^*) \cap X \neq \varnothing$. Since $\bar{B}_r(x^*)$ and X both are closed set, so is Y. Since $\bar{B}_r(x^*)$ is bounded, and $Y \subset \bar{B}_r(x^*)$, Y is also bounded. Thus, Y is compact. The distance function $d(x^*, y) : Y \longrightarrow \mathbb{R}_+$ is a continuous function on a compact set Y. Therefore, by Weirstrass' theorem, there exists $y^* \in Y$ such that

$$d(x^*, y^*) = \min\{d(x^*, y) \mid y \in Y\}.$$

If $y \in X \setminus Y$, then we must have $y \notin \bar{B}_r(x^*)$, and hence $d(x^*, y) > r$. Therefore, we have established that $d(x^*, y^*) = \min\{d(x^*, y) \mid y \in X\}$. Finally, we show that y^* is unique. Suppose that there is $z^* \neq y^*$ such that $d(x^*, y^*) = d(x^*, z^*)$. Since X is convex, the point $w^* = (1/2)y^* + (1/2)z^*$ is in X. Then it must be the case that $d(x^*, w^*) < d(x^*, y^*)$, which is a contradiction to (4.1). $\square$

Definition 4.6: Supporting hyperplane

A hyperplane $H(p, \alpha)$ is a supporting hyperplane for a set X if it contains a point on the boundary of X and the whole set lies on the same side of $H(p, \alpha)$. Equivalently, $H(p, \alpha)$ supports X if

$$\alpha = \inf\{p \cdot x \text{ for } x \in X\} \text{ or } \alpha = \sup\{p \cdot x \text{ for } x \in X\}.$$

Intuition suggests that a convex set in $\mathbb{R}^n$ should have a supporting hyperplane through each point of its boundary, and also, given two disjoint convex sets, there should be a hyperplane that separates the two sets. The following theorems establish that both of the above intuitions are correct.

Theorem 4.4

Let X be a non-empty, closed and convex subset of $\mathbb{R}^n$, and let x^* be a point in $\mathbb{R}^n \setminus X$. Then

(a) *(supporting hyperplane theorem)* there exists a point $y^* \in bd(X)$ and a hyperplane $H(p, \alpha)$ that

passes through y^* that supports X and separates X from $\{x^*\}$, i.e., $H(p, \alpha)$ is such that

$$p \cdot x^* < \alpha = p \cdot y^* = \inf\{p \cdot x \text{ for } x \in X\}.$$

(b) There exists a second hyperplane $H(p, \beta)$ that separates X strictly from $\{x^*\}$, i.e.,

$$p \cdot x^* < \beta < p \cdot x, \text{ for all } x \in X.$$

Proof. (a) Let $p = y^* - x^*$ where y^* is given by (4.1), and $\alpha = p \cdot y^*$. Then $H(p, \alpha)$ passes through y^* that is orthogonal to the line joining y^* and x^*. We claim that this is the desired hyperplane. First, we have

$$p \cdot x^* = p \cdot y^* - p \cdot (y^* - x^*) = \alpha - ||p||^2 < \alpha.$$

So $x^* \in H_L(p, \alpha)$. To show that X lies in $H_U(p, \alpha)$, we proceed by contradiction. Suppose there is a point $y \in X$ such that $p \cdot y < \alpha$, and let

$$x^\lambda = (1 - \lambda)y^* + \lambda y, \quad \text{for } \lambda \in (0, 1).$$

Since X is convex, $x^\lambda \in X$. Now, notice that $x^* - x^\lambda = x^* - y^* - \lambda(y - y^*) = -p - \lambda(y - y^*)$. Therefore,

$$\begin{aligned} ||x^* - x^\lambda||^2 &= ||p||^2 + 2\lambda p \cdot (y - y^*) + \lambda^2 ||y - y^*||^2, \\ \Rightarrow \quad ||x^* - y^*||^2 - ||x^* - x^\lambda||^2 &= -\lambda \left[2p \cdot (y - y^*) + \lambda ||y - y^*||^2\right]. \end{aligned}$$

Since by construction $p \cdot y^* = \alpha$ and by assumption $p \cdot y < \alpha$, we have $p \cdot (y - y^*) < 0$. Hence, the above equation implies that, for small values of λ, $||x^* - y^*|| - ||x^* - x^\lambda|| > 0$, which is a contradiction since y^* is chosen to minimize the distance $||x^* - y||$ for $y \in X$. (b) The proof of this part is identical (with the same p) to that of (a), except that we now choose β so as to $H(p, \beta)$ passes through the point $(1/2)x^* + (1/2)y^*$. □

The above theorem can easily be generalized to the case when X is a convex set, but not necessarily closed. The following theorem proves the existence of a hyperplane that separates two disjoint convex sets.

Theorem 4.5: Minkowski's separating hyperplane theorem

Let X and Y be two non-empty and disjoint convex sets in $\mathbb{R}^n$. Then there exists a hyperplane $H(p, \alpha) \subset \mathbb{R}^n$ that separates X and Y.

Proof. Let $Z = X - Y = X + (-1)Y$, which is convex by Lemma 4.1(b). We claim that $0 \notin Z$. If we had $0 \in Z$, there would exist a point $x \in X$ and $y \in Y$ such that $x - y = 0$. But this implies $x = y$, and so $x \in X \cap Y$, which contradicts the assumption that X and Y are disjoint. Since $0 \notin Z$, by the general version of the previous theorem, there exists $p \in \mathbb{R}^n$ such that, for $z \in Z$, $x \in X$ and $y \in Y$,

$$0 = p \cdot 0 \leq p \cdot z = p \cdot (x - y) \quad \Rightarrow \quad p \cdot y \leq p \cdot x.$$

The set of real numbers of the form $\{p \cdot y \text{ for } y \in Y\}$ is bounded above by the number $p \cdot x$, and hence has a supremum, which we call α. Thus, we have that

$$p \cdot y \leq \alpha \leq p \cdot x, \quad \text{for } x \in X \text{ and } y \in Y.$$

Therefore, $H(p, \alpha)$ separates X and Y. □

Theorem 4.5 is used to prove the so-called *second theorem of welfare economics*. Let us introduce some necessary definitions prior to proving the theorem. The production plan of a firm is described by its production set $Y := \{y \in \mathbb{R}^n \mid F(y) \leq 0\}$, where $F(\cdot)$ is called the transformation function. Obviously, $bd(Y) = \{y \in$

$\mathbb{R}^n \mid F(y) = 0\}$. A production $y \in Y$ is *profit-maximizing* for some price vector $p \in \mathbb{R}^n_{++}$ if $p \cdot y \geq p \cdot y'$ for all $y' \in Y$. A production $y \in Y$ is *efficient* if there is no $y' \in Y$ such that $y' \succsim y$ and $y' \neq y$. It is obvious that every efficient y must lie on $bd(Y)$, but the converse is not necessarily true. Now we prove the following important theorem.

Theorem 4.6: Second theorem of welfare economics

Suppose the production set Y is convex. Then every production $y^* \in Eff(Y)$ is a profit-maximizing production for some price vector $p \in \mathbb{R}^n_+$.

Proof. Suppose that $y^* \in Eff(Y)$, and defined the set $P_{y^*} := \{y' \in \mathbb{R}^n \mid y' > y^*\}$. The set P_{y^*} is convex, and because y is efficient, $Y \cap P_{y^*} = \varnothing$. Then by Theorem 4.5, there is a non-zero price vector such that that $p \cdot y' \geq p \cdot y$ for every $y' \in P_{y^*}$ and $y \in Y$. Note, in particular, that this implies $p \cdot y' \geq p \cdot y^*$ with $y' > y^*$. We must have $p \in \mathbb{R}^n_+$ because if $p_i < 0$ for some i, then we would have $p \cdot y' < p \cdot y$ for some y such that $y < y'$ with $y'_i - y_i$ sufficiently large. Since $y^* \in bd(Y)$, there is a hyperplane $H(p, \pi)$ that passes through y^*, and supports P_{y^*}, i.e., $\pi = p \cdot y^* = \inf\{p \cdot y' \text{ for } y' \in P_{y^*}\}$. Now take any $y \in Y$. Then $p \cdot y' \geq p \cdot y$ for every $y' \in P_y$. Since the set of real numbers of the form $\{p \cdot y \mid y \in Y\}$ is bounded above, by the supremum property we have $p \cdot y \leq \pi = p \cdot y^*$ for all $y \in Y$, i.e., y^* is profit-maximizing at p. $\square$

4.4 A quick tour of linear algebra

Let $A = \{a^1, \ldots, a^m\}$ denote a finite set of vectors as well as the index set of the vectors.

Definition 4.7: Linear combination and span

A vector b can be expressed as a linear combination of vectors in $A = \{a^1, \ldots, a^m\}$ if there exist real numbers $\{x_j\}_{j \in A}$ such that

$$b = \sum_{j \in A} x_j a^j.$$

The set of all such vectors that can be expressed as a linear combinations of vectors in A is called the span of A and is denoted $span(A)$.

Definition 4.8: Linear independence

set $A = \{a^1, \ldots, a^m\}$ of vectors is linearly independent if for all sets of real numbers $\{x_j\}_{j \in A}$

$$\sum_{j \in A} x_j a^j = 0 \quad \Longrightarrow \quad x_j = 0 \text{ for all } j \in A.$$

If the vectors in A are linearly dependent, then it is the case that there exist real numbers $\{x_j\}_{j \in A}$ not all zero such that

$$\sum_{j \in A} x_j a^j = 0.$$

Definition 4.9: Rank

The rank of a set A of vectors, denoted $\rho(A)$, is the size of the largest subset of linearly independent vectors in A.

Example 4.4

The set $A = \{(0,\ 1,\ 0),\ (-2,\ 2,\ 0)\}$ is a set of linearly independent vectors, whereas the vectors in $A' = \{(0,\ 1,\ 0),\ (-2,\ 2,\ 0),\ (-2,\ 3,\ 0)\}$ are linearly dependent. Notice that $\rho(A) = \rho(A') = 2$.

It is also true that, given a set A of vectors, the dimension of $span(A)$ is equal to its rank. Let $A = \{(0,\ 1),\ (1,\ 0)\}$. Then $span(A) = \mathbb{R}^2$, and $dim[span(A)] = \rho(A) = 2$.

Definition 4.10: Finite cone

Given an $m \times n$ matrix A, the set of all non-negative linear combinations of the columns of A is called the finite cone generated by the columns of A. Formally, such cone is given by:

$$cone(A) = \{y \in \mathbb{R}^m \mid y = Ax \ \text{ for some } \ x \in \mathbb{R}^n_+\}.$$

Notice the difference between $cone(A)$ and $span(A)$, which is given by:

$$span(A) = \{y \in \mathbb{R}^m \mid y = Ax \ \text{ for some } \ x \in \mathbb{R}^n\}.$$

For example, let the columns of a matrix $A_{2\times 2}$ be the vectors $(1,\ 0)$ and $(0,\ 1)$. Then $span(A) = \mathbb{R}^2$, whereas $cone(A)$ is the non-negative orthant. The following lemma states a useful property of finitely generated cones.

Lemma 4.3

Let A be an $m \times n$ matrix, then $cone(A)$ is a closed convex set.

We will skip the proof of the above lemma. The convexity of $cone(A)$ is easy to show, whereas the proof of closedness is a bit more complicated. Interested readers should refer to Vohra (2005).

4.5 The theorems of the alternative

Let us start with the problems of the following kind:

> Given a matrix $A_{m\times n}$ and $b \in \mathbb{R}^m$, find an $x \in \mathbb{R}^n$ such that $Ax = b$ or prove that no such x exists.

The above kind of problems motivates the *fundamental theorem of linear algebra* which is stated below.

Theorem 4.7: Fundamental theorem of linear algebra

Let A be an $m \times n$ matrix, $b \in \mathbb{R}^m$, $F = \{x \in \mathbb{R}^n \mid Ax = b\}$, and $G = \{y \in \mathbb{R}^m \mid \ yA = 0 \ \text{ and } \ yb \neq 0\}$. Then $F \neq \varnothing$ if and only if $G = \varnothing$.

The proof of the above theorem is omitted. One can provide an easy geometric interpretation. The first alternative is equivalent to the fact that $b \in span(A)$. Let $span(A)$ be a plane is $\mathbb{R}^2$. If $b \notin span(A)$, then b must be a vector that has non-zero inner product ($yb \neq 0$) with a vector y that is orthogonal to the plane, i.e., orthogonal to each of the linearly independent column vectors of A ($yA = 0$).

Next, we modify the above question a bit, and look for the set of non-negative solutions to the system $Ax = b$. The well-known *Farkas lemma*, which can be derived from the separation theorems, analyzes such problems.

Theorem 4.8: Farkas lemma

Let A be an $m \times n$ matrix, $b \in \mathbb{R}^m$, $F = \{x \in \mathbb{R}^n \mid Ax = b,\ x \geq 0\}$, and $G = \{y \in \mathbb{R}^m \mid \ yA \geq 0 \text{ and } yb < 0\}$. Then $F \neq \varnothing$ if and only if $G = \varnothing$.

The proof of the above theorem may be omitted by the students since the graphical intuition is simple enough. Take a matrix A whose columns are $a_1 = (1,\ 1)$ and $a_2 = (0,\ 1)$. Clearly, $span(A) = \mathbb{R}^2$ and $cone(A)$ is triangle shaped area between the vectors $(1,\ 1)$ and $(0,\ 1)$. The first alternative implies that vector $b = (b_1,\ b_2)$ lies inside the triangular area. To see the intuition behind the second alternative, let $b = (1,\ 0) \notin cone(A)$. We need to find a vector $y \in \mathbb{R}^2$ with the desired property. Let $y = (1,\ -1)$. Notice that $y^T \cdot a_1 = (1,\ -1)^T \cdot (1,\ 1) = 0$ and $y^T \cdot a_2 = (1,\ -1)^T \cdot (1,\ 0) = 1$, and hence $yA \in [0,\ 1]$. Also, $y^T \cdot b = (1,\ -1)^T \cdot (0,\ 1) = -1 < 0$. The idea is to make use of the separation theorem. Since $cone(A)$ is closed and convex, and $b \notin cone(A)$, then there exists a hyperplane that strictly separates b from $cone(A)$. Convince yourselves that $z_2 = 1/2 + z_1$ where $(z_1,\ z_2) \in \mathbb{R}^2$ is such a hyperplane, and it is orthogonal to y. The set G are called the Farkas alternative of F.

The following theorem is a generalization of Farkas lemma.

Theorem 4.9

Let $A_{m\times n}$, $B_{m\times t}$, $C_{k\times n}$ and $D_{k\times t}$ are matrices, and $b \in \mathbb{R}^m$ and $d \in \mathbb{R}^k$ are vectors. Define

$$\begin{aligned} F :=& \{x \in \mathbb{R}^n,\ x' \in \mathbb{R}^t \mid Ax + Bx' = b,\ Cx + Dx' \leq d,\ x \geq 0\}, \\ G :=& \{y \in \mathbb{R}^m,\ y' \in \mathbb{R}^k \mid \ yA + y'C \geq 0,\ yB + y'D = 0,\ yb + y'd < 0 \text{ and } y' \geq 0\} \end{aligned}$$

Then $F \neq \varnothing$ if and only if $G = \varnothing$.

Proof. Consider the system of inequalities $Cx + Dx' \leq d$. This can be converted to a system of equations by introducing slack variables for every inequality, i.e., there exists $s \in \mathbb{R}^k_+$ such that

$$Cx + Dx' + s = d.$$

Next, the vector $x' \in \mathbb{R}^t$ can be written as $x' = z - z'$ where $z,\ z' \in \mathbb{R}^t$. This is because any real number x'_j can be written as the difference of two non-negative real numbers z_j and z'_j. So, F is given by:

$$F := \{x \in \mathbb{R}^n,\ z,\ z' \in \mathbb{R}^t,\ s \in \mathbb{R}^k \mid Ax + Bz - Bz' + 0 \cdot s = b,\ Cx + Dz - Dz' + Is = d,\ x,\ z,\ z',\ s \geq 0\},$$

where $I_{k\times k}$ is the identity matrix. In matrix notation, the above system looks

$$\begin{bmatrix} A & B & -B & 0 \\ C & D & -D & I \end{bmatrix} \begin{bmatrix} x \\ z \\ z' \\ s \end{bmatrix} = \begin{bmatrix} b \\ d \end{bmatrix} \qquad \Longleftrightarrow \qquad Qr = w.$$

The Farkas alternative of F is given by:

$$G := \{\tilde{y} \in \mathbb{R}^{m+k} \mid \tilde{y}Q \geq 0 \text{ and } \tilde{y}w < 0\} \quad \text{where} \quad \tilde{y} = \begin{bmatrix} y \\ y' \end{bmatrix}$$

Notice that $\tilde{y}w < 0$ is equivalent to $yb + y'd < 0$. And $\tilde{y}Q \geq 0$ is equivalent to the following system of

inequalities:

$$\begin{aligned} yA + y'C &\geq 0, \\ yB + y'D &\geq 0, \\ -yB - y'D &\geq 0, \\ y \cdot 0 + y'I &\geq 0. \end{aligned}$$

The second and the third inequalities together imply $yB + y'D = 0$, and the last inequality implies $y' \geq 0$. Now apply Farkas lemma to complete the proof. $\square$

Definition 4.11: Linear and affine functions

Let X be a convex subset of $\mathbb{R}^n$. A function $L : X \to \mathbb{R}$ is linear if (a) $L(x + y) = L(x) + L(y)$ for any vectors $x, y \in X$, and (b) $L(\alpha x) = \alpha L(x)$ for any vector $x \in X$ and any scalar α. A function $f : X \to \mathbb{R}$ is an affine function if there is a linear function $L : X \to \mathbb{R}$ and a real number b such that $f(x) = L(x) + b$ for all $x \in X$.

Convince yourself that all linear functions are affine functions, but the converse is not true. The following theorem is another theorem of the alternatives, which is very useful in establishing the existence of the Lagrange multipliers for constrained optimization problems.

Theorem 4.10: Gordan's theorem

Let $f_i : X \longrightarrow \mathbb{R}$ for $i = 1, \ldots, m$ be affine functions, X be a convex subset of $\mathbb{R}^n$. Define

$$\begin{aligned} F :=& \{x \in \mathbb{R}^n \mid f_i(x) < 0 \text{ for } i = 1, \ldots, m\}, \\ G :=& \{y \in \mathbb{R}^m_+, y_i > 0 \text{ for at least one } i \mid y \cdot f(x) \geq 0 \text{ for all } x \in \mathbb{R}^n\}, \end{aligned}$$

where $f(x) = (f_1(x), \ldots, f_m(x))$. Then $F \neq \varnothing$ if and only if $G = \varnothing$.

Proof. We first show that if (a) holds then (b) cannot hold. Given any x satisfying (a) and any multipliers $y_1 \geq 0, \ldots, y_m \geq 0$, each term of the expression $\sum_{i=1}^m y_i f_i(x)$ is non-positive. Terms for which $y_i \neq 0$ are actually negative, so that the above expression must be negative if the multipliers are not all zero. Therefore (b) cannot hold.

Assume now that (a) does not hold. We must show that in this case (b) holds. Let

$$S = \{z \in \mathbb{R}^m \mid \text{there exists } x \in X \text{ such that } z_i > f_i(x) \text{ for all } i = 1, \ldots, m\}.$$

The above set is clearly non-empty since each z_i can be written as $z_i = f_i(x) + \alpha$ for $\alpha > 0$. Next, we show that S is a convex set. Let z' and z'' be two elements of S, and define $z^\mu = \mu z' + (1 - \mu)z''$ and $x^\mu = \mu x' + (1 - \mu)x''$ for $\mu \in (0, 1)$. The point z' is in S implies that there exists some x' in X such that $z'_i > f_i(x')$ for each $i = 1, \ldots, m$. Similarly, z'' is in S implies that there exists some x'' in X such that $z''_i > f_i(x'')$ for each $i = 1, \ldots, m$. Hence, it follows that

$$z^\mu_i = \mu z'_i + (1 - \mu)z''_i > \mu f_i(x') + (1 - \mu)f_i(x'') = f_i(x^\mu), \text{ for each } i = 1, \ldots, m.$$

Notice that $x^\mu \in X$ since X is convex, and the last equality holds because f_i is an affine function for each i. Thus, we have shown that that there exists some x $(= x^\mu)$ in X such that $z^\mu > f_i(x)$ for each i, and hence z^μ is in S. Next, if (a) does not hold, then $0 \notin S$. If 0 were in S, then for all $i = 1, \ldots, m$ we would have $f_i(x) < 0$, contradicting the hypothesis that (a) does not hold good. So, by the Separating Hyperplane theorem we have that there exists a non-zero vector $y = (y_1, \ldots, y_m)$ such that

$$0 \leq y \cdot z = \sum_{i=1}^m y_i z_i, \text{ for all } z \in S.$$

Fix an $x^0 \in X$ and set $w^0 = (f_1(x^0), \ldots, f_m(x^0))$. Then for any $q \in \mathbb{R}^m_{++}$ and $r > 0$ we have that $w^0 + rq \in S$, and hence

$$y \cdot (w^0 + rq) \geq 0 \;\Rightarrow\; y \cdot q \geq -\frac{1}{r}(y \cdot w^0).$$

Taking the limit of the above expression as $r \to \infty$ and using the fact that $y \cdot q$ is continuous, we have that

$$y \cdot q \geq 0, \quad \text{for all } q \in \mathbb{R}^m_{++}$$

The above implies that $y_i \geq 0$ for all $i = 1, \ldots, m$. Now consider any $x \in X$, and any fixed $q \in \mathbb{R}^m_{++}$ and $\varepsilon > 0$. Thus we have that

$$y \cdot (f(x) + \varepsilon q) \geq 0.$$

Taking the limit of the above expression as $\varepsilon \to 0$, we get the desired result. $\square$

4.6 Applications

Farkas lemma has been applied to many problems in economics and finance. In what follows we discuss two useful applications.

4.6.1 No arbitrage in financial markets

Suppose there are m financial assets and n states of nature. Let a_{ij} be the payoff from one share of asset i in state j. Thus the returns of m assets will be represented by a matrix $A_{m\times n}$. A portfolio of assets is represented by a vector $y \in \mathbb{R}^m$ where the i-th component, y_i represents the portfolio weight of asset i. Let $w \in \mathbb{R}^n$ be a vector whose j-th component denotes wealth in state j. We assume that wealth in a future state is related to the current portfolio by:

$$w_j = \sum_{i=1}^{m} y_i a_{ij} \quad \text{for } j = 1, \ldots, n \quad \text{with} \quad \sum_{i=1}^{m} y_i = 1.$$

Therefore, $w = Ay$. This assumes that assets are infinitely divisible, and the returns are linear in the quantities held. The no arbitrage condition asserts that a portfolio that pays off non-negative amounts in every state must have a non-negative price.

Proposition 4.1: No-arbitrage

Let $A_{m\times n}$ be the matrix of asset returns and $p \in \mathbb{R}^m_{++}$ be a vector of asset prices. If $y \in \mathbb{R}^m$ is a portfolio of assets, then

$$yA \geq 0 \implies y \cdot p \geq 0. \tag{4.2}$$

Equivalent to (4.2), the system $yA \geq 0$ and $y \cdot p < 0$ has no solution. From Farkas lemma we deduce the existence of vector $\hat{\pi} \in \mathbb{R}^n_+$ such that $p = A\hat{\pi}$. Since $p > 0$, it follows that $\hat{\pi} > 0$. Scale $\hat{\pi}$ by dividing through $\sum_j \hat{\pi}_j$. Let $p^* = p/\sum_j \hat{\pi}_j$ and $\pi = \hat{\pi}/\sum_j \hat{\pi}_j$. Notice that π is a probability vector. As long as relative prices are all that matter such scaling can be done without loss of generality. Now we have $p^* = A\pi$, i.e., there is a probability distribution under which the expected return of every asset is equal to its price. The probabilities are called the risk-neutral probabilities.

4.6.2 Core of a coalitional game

A coalitional game is denoted by (N, v) where $N = \{1, \ldots, n\}$ is the set of players, and $v : 2^N \to \mathbb{R}$, called the *characteristic function*, which assigns to each subset S of N a real number $v(S)$. The subset S is called a

coalition. A vector $x \in \mathbb{R}^n$ is called a *feasible allocation* if $x_i \geq v(\{i\})$ and $\sum_{i=1}^n x_i = v(N)$. Let $I(N, v)$ denote the set of feasible allocations or *imputations* of the game (N, v).

Consider an economy with two buyers 1 and 2, and one seller s, i.e., $N = \{1, 2, s\}$. The seller has one item, say a car, and the buyers have nothing. We are interested in a feasible trade of this item. Suppose that the seller has no value for the object. Buyer 1 derives a utility of 5, and buyer 2 derives a utility of 10 form the object, respectively. Let S be a coalition. Note that there are $2^3 = 8$ possible coalitions, namely, $\varnothing$, $\{1\}$, $\{2\}$, $\{s\}$, $\{1, 2\}$, $\{1, s\}$, $\{2, s\}$ and $\{1, 2, s\}$. The characteristic function $v(S)$, which denotes the "worth" of a coalition S is given by:

$$\begin{aligned} &v(\varnothing) = v(\{1\}) = v(\{2\}) = v(\{s\}) = v(\{1, 2\}) = 0, \\ &v(\{1, s\}) = 5, \quad \text{and} \quad v(\{2, s\}) = v(\{1, 2, s\}) = 10. \end{aligned}$$

We assume that the item is assigned to the highest-valuation buyer, and hence $v(N) = 10$.

The *core* of a game (N, v) is the set

$$C(N, v) := \left\{ x \in I(N, v) : \sum_{i \in S} x_i \geq v(S) \ \text{ for all } \ S \subseteq N \right\}.$$

Let us first compute the core allocations of the above buyer-seller economy. If $(x_1, x_2, x_s) \in C(N, v)$, then we must have

$$\begin{aligned} &x_1 + x_2 + x_s = 10, \\ &x_1 + x_2 \geq 0, \\ &x_1 + x_s \geq 5, \\ &x_2 + x_s \geq 10, \\ &x_i \geq 0 \quad \text{for} \quad i \in \{1, 2, s\}. \end{aligned}$$

Notice that

$$x_2 + x_s \geq 10 \iff 10 - x_1 \geq 10 \iff x_1 \geq 0,$$

which together with $x_1 \geq 0$ imply $x_1 = 0$. Then $x_s \geq 5$. Similarly,

$$x_1 + x_s \geq 5 \iff 10 - x_2 \geq 5 \iff x_1 \geq 5.$$

Therefore, $x_1 = 0$, $x_2 \in [0, 5]$ and $x_s \in [5, 10]$ with $x_2 + x_s = 10$ constitute a core allocation of the buyer-seller game.

Unfortunately, not all coalitional games have non-empty cores. For example, consider (N, v) where $N = \{1, 2\}$, and $v(\varnothing) = 0$, $v(\{1\}) = v(\{2\}) = 0.75$ and $v(\{1, 2\}) = 1$. It is easy to see that $C(N, v) = \varnothing$. Therefore, we will look for conditions under which the core of a coalitional game is non-empty.

Definition 4.12: Balanced collection

Let $\mathcal{C} \subseteq 2^N \setminus \varnothing$ be a collection of non-empty coalitions of N. The collection $\mathcal{C}$ is balanced if there exist numbers $y_S \geq 0$ for $S \in \mathcal{C}$ such that for every player $i \in N$ we have

$$\sum_{S:\, i \in S} y_S = 1. \tag{4.3}$$

The numbers $\{y_S \mid S \in \mathcal{C}\}$ in (4.3) are known as the balancing coefficients for the collection $\mathcal{C}$.

Let $B(N)$ be the set of feasible solutions to the following system:

$$\begin{aligned} y_S &\geq 0 \quad \text{for all} \quad S \subseteq N, \\ \sum_{S:\, i \in S} y_S &= 1 \quad \text{for all} \quad i \in N. \end{aligned}$$

The number y_S can be thought of as the weights given to coalition S. It is easy to verify that $B(N) \neq \varnothing$ since we can always find

$$y_S = \begin{cases} 1 & \text{for all } S \text{ with } |S| = 1, \\ 0 & \text{otherwise.} \end{cases}$$

In our buyer-seller economy, the collection of singletons is given by $\mathcal{C} = \{\{1\}, \{2\}, \{s\}\}$. Then, $y_{\{1\}} = y_{\{2\}} = y_{\{s\}} = 1$, and hence $(1, 1, 1)$ is a vector of balancing coefficients. For the collection of coalitions of size 2, the vector of balancing coefficients is $(1/2, 1/2, 1/2)$.

Definition 4.13: Balanced game

A game (N, v) is balanced if

$$\sum_{S \subseteq N} v(S) y_S \leq v(N)$$

for every balanced collection of weights, i.e., for all $y \in B(N)$.

Let us verify whether the buyer-seller game is balanced. Notice that the only coalitions, besides the grand coalition, having positive worths are $\{1, s\}$ and $\{2, s\}$. So, we need to show that for every $y \in B(N)$, we have

$$\begin{aligned} & v(\{1, s\})\, y_{\{1,s\}} + v(\{2, s\})\, y_{\{2,s\}} \leq v(N) \\ \iff & 5y_{\{1,s\}} + 10y_{\{2,s\}} \leq 10 \\ \iff & y_{\{1,s\}} + 2y_{\{2,s\}} \leq 2. \end{aligned}$$

Since $y_{\{1,s\}} + y_{\{2,s\}} = 1$, we have from the above that $y_{\{2,s\}} \leq 1$ because $y_{\{1,s\}}, y_{\{2,s\}} \geq 0$. Now,

$$\begin{aligned} & y_{\{1,s\}} + 2y_{\{2,s\}} \leq 2 \\ \iff & y_{\{1,s\}} \leq 2\left[1 - y_{\{2,s\}}\right] = 2y_{\{1,s\}}. \end{aligned}$$

The above is true since $y_{\{1,s\}} \geq 0$, and hence (N, v) is balanced.

Theorem 4.11: Bondareva-Shapley

Given a coalitional game (N, v), $C(N, v) \neq \varnothing$ if and only if (N, v) is balanced.

Proof. If $x \in C(N, v)$, then

$$\begin{aligned} & \sum_{i \in N} x_i = v(N), \\ & \sum_{i \in S} x_i \geq v(S) \quad \forall \; S \subseteq N, \\ & x_i \;\; \text{free} \;\; \forall \; i \in N. \end{aligned} \tag{CORE}$$

The Farkas alternative for (CORE) is

$$
\begin{aligned}
& v(N)\, y_N - \sum_{S \subseteq N} v(S)\, y_S < 0, \\
& y_N - \sum_{S:\, i \in S} y_S = 0 \;\; \forall \;\; i \in N, \\
& y_S \geq 0 \;\; \forall \;\; S \subseteq N, \\
& y_N \;\; \text{free.}
\end{aligned}
\tag{BAL}
$$

Now suppose $C(N,\, v) \neq \varnothing$. Then (CORE) has a solution, and the Farkas alternative (BAL) has no solutions. Consider $y \in B(N)$ and let $y_N = 1$. Therefore, $B(N)$ is the set of solutions to the final two constraints of (BAL). Since (BAL) has no solutions, we must have

$$
v(N)\, y_N - \sum_{S \subseteq N} v(S)\, y_S \geq 0,
$$

i.e., $(N,\, v)$ is balanced.

To prove the other direction, suppose that $(N,\, v)$ is balanced, i.e., for all $y \in B(N)$ we have

$$
v(N)\, y_N \geq \sum_{S \subseteq N} v(S)\, y_S.
$$

We will show that (BAL) has no solutions, and hence (CORE) has a solution. Assume for contradiction that (BAL) has a solution y. Clearly, $y_N \neq 0$, else $y_S = 0$ for all S which would contradict the first inequality of (BAL). Define $y'_S = y_S / y_N$ for all $S \subseteq N$. Since y is a solution to (BAL), we get

$$
\begin{aligned}
& v(N) < \sum_{S \subseteq N} v(S)\, y'_S, \\
& \sum_{S:\, i \in S} y'_S = 1 \;\; \forall \;\; i \in N, \\
& y'_S \geq 0 \;\; \forall \;\; S \subseteq N.
\end{aligned}
$$

Therefore, $y' \in B(N)$. But the first inequality contradicts the fact that $(N,\, v)$ is balanced. $\square$

Chapter 5

Concave and quasiconcave functions

5.1 Concave and convex functions

5.1.1 Definitions and properties

Let $f : X \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ be a real valued function. From now on we will assume that X is a convex subset of $\mathbb{R}^n$. The *hypograph* and *epigraph* of f are defined as follows.

Definition 5.1: Hypograph and epigraph of a function

The hypograph (or subgraph) and epigraph of a function f, denoted respectively $hyp\, f$ and $epi\, f$, are defined as

$$\begin{aligned} hyp\, f &:= \{(x,\, y) \in X \times \mathbb{R} |\, f(x) \geq y\}, \\ epi\, f &:= \{(x,\, y) \in X \times \mathbb{R} |\, f(x) \leq y\}. \end{aligned}$$

Intuitively, the hypograph of a function is the area lying below the graph of the function, while the epigraph is the area lying above the graph.

Definition 5.2: Concave and convex functions

A function $f : X \longrightarrow \mathbb{R}$ is concave (convex) on X if $hyp\, f$ ($epi\, f$) is convex.

The following theorem gives a characterization of concave (convex) functions.

Theorem 5.1

A function $f : X \longrightarrow \mathbb{R}$ is concave (convex) on X if and only if for all $x,\, y \in X$ and for all $\lambda \in [0,\, 1]$, it is the case that

$$f(\lambda x + (1-\lambda)y) \geq\ (\leq)\, \lambda f(x) + (1-\lambda) f(y).$$

Proof. First, suppose that f is concave, i.e., $hyp\, f$ is convex. Let x and y be two arbitrary points in X. Then, $(x,\, f(x)) \in hyp\, f$ and $(y,\, f(y)) \in hyp\, f$. Since $hyp\, f$ is convex, we have, for any $\lambda \in [0,\, 1]$,

$$(\lambda x + (1-\lambda)y,\, \lambda f(x) + (1-\lambda) f(y)) \in hyp\, f.$$

By definition of $hyp\, f$, a point $(w,\, z)$ is in $hyp\, f$ only if $f(w) \geq z$, and hence

$$f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda) f(y).$$

Now suppose that for all $x,\ y \in X$ and for all $\lambda \in [0,\ 1]$, it is the case that

$$f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda)f(y). \tag{5.1}$$

We will show that for two arbitrary points $(w_1,\ z_1)$ and $(w_2,\ z_2)$ in $hyp\, f$, their convex combination is also in $hyp\, f$. Since the above two points are in $hyp\, f$, we have $f(w_1) \geq z_1$ and $f(w_2) \geq z_2$. Condition (5.1) and $\lambda \in [0,\ 1]$ imply that

$$\begin{aligned} f(\lambda w_1 + (1-\lambda)w_2) &\geq \lambda f(w_1) + (1-\lambda)f(w_2) \\ &\geq \lambda z_1 + (1-\lambda)z_2. \end{aligned}$$

Therefore, the point $(\lambda w_1 + (1-\lambda)w_2,\ \lambda z_1 + (1-\lambda)z_2)$ is in $hyp\, f$, and hence it is convex. The proof for a convex function is similar. $\square$

For strict concavity and convexity replace '$\geq$' ('$\leq$') by '$>$' ('$<$'), and take $\lambda \in (0,\ 1)$. The notions of concavity and convexity are neither exhaustive nor mutually exclusive. We may have functions that are neither concave nor convex, and functions that are both concave and convex. The function $f : \mathbb{R} \longrightarrow \mathbb{R}$ defined by $f(x) = x^3$ is neither concave nor convex. To see this take $x = -2$ and $y = 2$. Then check the conditions of the above theorem for $\lambda = 1/4$ and $\lambda = 3/4$. A linear function such as $f(x) = 2 + 3x$ is both concave and convex. Such functions are called *affine functions*. Another observation is that a function $f : X \longrightarrow \mathbb{R}$ is concave if and only if $-f$ is convex. The following theorem states an important result associated with concave functions.

Theorem 5.2: Local-global theorem

Let $f : X \longrightarrow \mathbb{R}$ be a concave function on X. Then

(a) every local maximum of f is a global maximum,

(b) the set $\arg\max\{f(x) |\, x \in X\}$, the set of maximizers of f on X is either empty or convex

Proof. (a) Let x^0 be any local maximum of f, but not a global maximum. Then there is $r > 0$ such that $f(x) \leq f(x^0)$ for all $x \in B_r(x^0) \cap X$. Since x^0 is not a global maximum, there is $y \in X$ such that $f(y) > f(x^0)$. Since X is convex, for any $\lambda \in (0,\ 1)$, $(1-\lambda)y + \lambda x^0 \in X$. Pick λ close to 1 so that $(1-\lambda)y + \lambda x^0 \in B_r(x^0)$. By concavity of f,

$$f((1-\lambda)y + \lambda x^0) \geq (1-\lambda)f(y) + \lambda f(x^0) > f(x^0)$$

since $f(y) > f(x^0)$. But $(1-\lambda)y + \lambda x^0 \in B_r(x^0)$ by construction, so $f(x^0) \geq f((1-\lambda)y + \lambda x^0) > f(x^0)$, which is a contradiction.

(b) Suppose that x_1 and x_2 both maximize f on X, i.e., $f(x_1) = f(x_2)$. By concavity of f we have, for any $\lambda \in (0,\ 1)$,

$$f(\lambda x_1 + (1-\lambda)x_2) \geq \lambda f(x_1) + (1-\lambda)f(x_2) = f(x_1).$$

The above must hold with equality or x_1 and x_2 would not be maximizers. Thus, the set of maximizers must is convex. $\square$

Following are some useful properties of a concave function.

Lemma 5.1

Let $f,\ g : \mathbb{R}^n \longrightarrow \mathbb{R}$ be two concave functions. Then

(a) the function $\alpha f + \beta g : \mathbb{R}^n \longrightarrow \mathbb{R}$ is concave for $\alpha,\ \beta \geq 0$,

(b) the function $\min\{f,\ g\} : \mathbb{R}^n \longrightarrow \mathbb{R}$ is concave,

(c) if $h : \mathbb{R} \longrightarrow \mathbb{R}$ is non-decreasing and concave, then $h \circ f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is concave.

Proof. The proof of the above lemma is left as an exercise. □

Another important implication of concavity is stated in the following lemma.

Lemma 5.2

Let $f : X \longrightarrow \mathbb{R}$ be a concave function on X. For any $\alpha \in \mathbb{R}$, the upper contour set of f,

$$U_f(\alpha) := \{x \in X | \, f(x) \geq \alpha\}$$

is either empty or convex. Similarly, if f is convex, then the lower contour set of f,

$$L_f(\alpha) := \{x \in X | \, f(x) \leq \alpha\}$$

is either empty or convex.

Proof. Let x' and x'' be two points in $U_f(\alpha)$, i.e., $f(x') \geq \alpha$ and $f(x'') \geq \alpha$. By concavity of f,

$$f(\lambda x' + (1-\lambda)x'') \geq \lambda f(x') + (1-\lambda) f(x'') \geq \lambda\alpha + (1-\lambda)\alpha = \alpha.$$

Therefore, $\lambda x' + (1-\lambda)x''$ is in $U_f(\alpha)$. □

The converse statement of the above theorem is not always true. To guarantee the sufficiency we would need a weaker condition called quasiconcavity, which will be introduced later.

5.1.2 Continuity of concave functions

The main result of this subsection is that a concave function must be continuous everywhere on its domain, except perhaps at the boundary points.

Theorem 5.3

Let $f : X \longrightarrow \mathbb{R}$ be a concave function on X. Then f is continuous on $int(X)$.

Proof. See Sundaram (1996, p. 177). □

Continuity of a concave function may fail in the boundary of X. Consider the following example.

Example 5.1

Define $f : [0, 1] \longrightarrow \mathbb{R}$ by

$$f(x) = \begin{cases} \sqrt{x} & \text{if } 0 < x < 1, \\ -1 & \text{if } x = 0, 1. \end{cases}$$

Then f is strictly concave on $[0, 1]$, but is discontinuous at 0 and 1.

5.1.3 Differentiable concave functions

Concave functions have nice characterization when they are differentiable.

Theorem 5.4

Let X be an open and convex set in $\mathbb{R}^n$, and let $f : X \longrightarrow \mathbb{R}$ is a $\mathcal{C}^1$ function. Then f is concave if and only if for all x^0, $x \in X$, it is the case that

$$f(x) \leq f(x^0) + \nabla f(x^0) \cdot (x - x^0).$$

Proof. Suppose that f is concave on X. Take x, $x^0 \in X$. By the concavity of f, we have, for all $\lambda \in [0, 1]$,

$$\begin{aligned} & f(x^0 + \lambda(x - x^0)) \geq f(x^0) + \lambda[f(x) - f(x^0)] \\ \Rightarrow \quad & \frac{f(x^0 + \lambda(x - x^0)) - f(x^0)}{\lambda} \geq f(x) - f(x^0). \end{aligned}$$

Taking the limit of the above expression as $\lambda \to 0^+$, we get

$$\nabla f(x^0) \cdot (x - x^0) \geq f(x) - f(x^0).$$

To prove the converse, take x' and x'' in X, and let $x^\lambda = \lambda x' + (1 - \lambda)x''$. By hypothesis,

$$\begin{aligned} f(x') &\leq f(x^\lambda) + \nabla f(x^\lambda) \cdot (x' - x^\lambda), \\ f(x'') &\leq f(x^\lambda) + \nabla f(x^\lambda) \cdot (x'' - x^\lambda) \end{aligned}$$

Multiplying the first inequality by λ and the second inquality by $1 - \lambda$, and adding we get

$$f(x^\lambda) \geq \lambda f(x') + (1 - \lambda) f(x'').$$

This completes the proof. $\square$

Theorem 5.5

Let X be an open and convex set in $\mathbb{R}^n$, and let $f : X \longrightarrow \mathbb{R}$ is a $\mathcal{C}^2$ function. Then f is concave if and only if $\nabla^2 f(x)$ is negative semi-definite for all $x \in X$.

Proof. First, recall that a symmetric $n \times n$ matrix A is negative semi-definite if and only if $h^T A\, h \leq 0$ for all $h \in \mathbb{R}^n$. Suppose that f is concave. Take an $x \in X$ and an arbitrary direction vector h in $\mathbb{R}^n$. Because X is open, there is some $\delta > 0$ such that $x + \alpha h \in X$ for all $\alpha \in I \equiv (-\delta, \delta)$. Define $g : I \longrightarrow \mathbb{R}$ by

$$g(\alpha) := f(x + \alpha h) - f(x) - \nabla f(x) \alpha h. \tag{5.2}$$

Since f is $\mathcal{C}^2$, g is also $\mathcal{C}^2$ with $g(0) = 0$. By the previous theorem, the concavity of f implies

$$f(x + \alpha h) \leq f(x) + \nabla f(x) \alpha h \;\Rightarrow\; g(\alpha) \leq 0, \quad \text{for all} \;\; \alpha \in I.$$

Thus, g is a twice continuously differentiable function with a maximum at 0, and hence we must have $g'(0) = 0$ and $g''(0) \leq 0$. Now differentiating twice (5.2) with respect to α we get

$$g''(\alpha) = h^T \nabla^2 f(x + \alpha h)\, h.$$

And $g''(0) \leq 0$ implies that

$$g''(0) = h^T \nabla^2 f(x)\, h \leq 0.$$

Since the vector is chosen arbitrarily, we conclude that $\nabla^2 f(x)$ is negative semi-definite for all $x \in X$.

To show the converse, suppose that $\nabla^2 f(x)$ is negative semi-definite. Pick any two points x and $x + h$ in X. By Taylor's theorem, we have, for some $\alpha \in (0, 1)$,

$$f(x + h) - f(x) - \nabla f(x)\, h = \frac{1}{2} h^T \nabla^2 f(x + \alpha h) h,$$

where $x + \alpha h \in (x,\, x+h)$. Since $\nabla^2 f(x)$ is negative semi-definite, the right-hand-side of the above equation is negative, which implies

$$f(x+h) \leq f(x) + \nabla f(x)\, h,$$

and hence by the previous theorem, f is concave. $\square$

5.2 Quasiconcave and quasiconvex functions

5.2.1 Definitions and properties

Definition 5.3: Quasiconcave and quasiconvex functions

A function $f : X \longrightarrow \mathbb{R}$ is quasiconcave on X if the upper contour set of f, $U_f(\alpha) = \{x \in X |\, f(x) \geq \alpha\}$ is convex. The function f is quasiconvex if the lower contour set of f, $L_f(\alpha) = \{x \in X |\, f(x) \leq \alpha\}$ is convex.

The following theorem gives a characterization of quasiconcave (quasiconvex) functions.

Theorem 5.6

A function $f : X \longrightarrow \mathbb{R}$ is quasiconcave on X if and only if for all $x,\, y \in X$ and for all $\lambda \in [0,\, 1]$, it is the case that

$$f(\lambda x + (1-\lambda)y) \geq \min\{f(x),\, f(y)\}.$$

Similarly, A function $f : X \longrightarrow \mathbb{R}$ is quasiconvex on X if and only if for all $x,\, y \in X$ and for all $\lambda \in [0,\, 1]$, it is the case that

$$f(\lambda x + (1-\lambda)y) \leq \max\{f(x),\, f(y)\}.$$

Proof. First suppose that $U_f(\alpha)$ is convex. Take $x,\, y \in X$ and $\lambda \in [0,\, 1]$. Assume without loss of generality that $f(x) \geq f(y) = \alpha$. Thus, $x,\, y \in U_f(\alpha)$, and by convexity of $U_f(\alpha)$ we have $\lambda x + (1-\lambda)y \in U_f(\alpha)$, which means

$$f(\lambda x + (1-\lambda)y) \geq \alpha = f(y) = \min\{f(x),\, f(y)\}.$$

To show the converse, suppose that $f(\lambda x + (1-\lambda)y) \geq \min\{f(x),\, f(y)\}$ for all $x,\, y \in X$ and for all $\lambda \in [0,\, 1]$. If $U_f(\alpha)$ is empty or singleton, then it is trivially convex. So suppose that $U_f(\alpha)$ contains at least two points x and y. Then $f(x) \geq \alpha$ and $f(y) \geq \alpha$, and so $\min\{f(x),\, f(y)\} \geq \alpha$. By hypothesis, $f(\lambda x + (1-\lambda)y) \geq \min\{f(x),\, f(y)\}$, and hence $\lambda x + (1-\lambda)y \in U_f(\alpha)$. Since α was chosen arbitrarily, this completes the proof of the first part. The proof of the second part is analogous. $\square$

For strict quasiconcavity and quasiconvexity replace '$\geq$' ('$\leq$') by '$>$' ('$<$'), and take $\lambda \in (0,\, 1)$. Also, a function $f : X \longrightarrow \mathbb{R}$ is (strictly) quasiconcave if and only if $-f$ is (strictly) quasiconvex. The notions of quasiconcavity and quasiconvexity are generalizations of concavity and convexity, respectively. For instance, the set of all concave functions are contained in the set of quasiconcave functions.

Theorem 5.7

Let $f : X \longrightarrow \mathbb{R}$ be concave on X, then it is also quasiconcave. Similarly, if $f : X \longrightarrow \mathbb{R}$ be convex on X, then it is also quasiconvex.

Proof. The concavity of f implies that, for all $x,\ y \in X$ and for all $\lambda \in [0,\ 1]$,

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &\geq \lambda f(x) + (1-\lambda) f(y) \\ &\geq \lambda \min\{f(x),\ f(y)\} + (1-\lambda)\min\{f(x),\ f(y)\} \\ &= \min\{f(x),\ f(y)\}. \end{aligned}$$

So f is quasiconcave. $\square$

Following lemma states some useful properties of quasiconcave functions.

Lemma 5.3

Let $f,\ g : \mathbb{R}^n \longrightarrow \mathbb{R}$ be two quasiconcave functions. Then

(a) the function $\alpha f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is quasiconcave for $\alpha \geq 0$,

(b) the function $\min\{f,\ g\} : \mathbb{R}^n \longrightarrow \mathbb{R}$ is quasiconcave,

(c) if $h : \mathbb{R} \longrightarrow \mathbb{R}$ is non-decreasing, then $h \circ f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is quasiconcave.

Proof. The proof of the above lemma is left as an exercise. $\square$

With respect to the above lemma, $f + g : \mathbb{R}^n \longrightarrow \mathbb{R}$ is not necessarily quasiconcave. Suppose that $f(x) = x^3$ and $g(x) = -x$ are both quasiconcave, but $f(x) + g(x) = x^3 - x$ is not quasiconcave. Theorem 5.7 and Lemma 5.3(c) together imply that monotonic transformation of any concave function would result in a quasiconcave function. But the converse, i.e., whether any quasiconcave function is a monotonic transformation of a concave function, is not necessarily true. That is why quasiconcavity is a generalization of concavity. Consider the following example.

Example 5.2

Define $f : \mathbb{R}_+ \longrightarrow \mathbb{R}$ by

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

Since f is non-decreasing, it is quasiconcave. Now, suppose that there existed a concave function g and a strictly increasing function h such that $h \circ g \equiv f$. First, observe that g must be constant on $[0,\ 1]$. Suppose that, for $x,\ y \in [0,\ 1]$, we have $x > y \Rightarrow g(x) > g(y)$. Since h is strictly increasing, we must have $f(x) = h(g(x)) > h(g(y)) = f(y)$, which is a contradiction to the fact that f is constant on $[0,\ 1]$. Next, observe that g must be strictly increasing in x for $x > 1$. Apply an argument similar to the above to establish this fact. Since g is constant on $[0,\ 1]$, it has a local maximum at every $x^* \in (0,\ 1)$. These local maxima are not global maxima since g is strictly increasing for $x > 1$. This contradicts the fact that every local maximum of a concave function is a global maximum.

Quasiconcave functions do not have similar implications for continuity and differentiability as concave function. To see this consider the following example.

Example 5.3

Define $f : \mathbb{R}_+ \longrightarrow \mathbb{R}$ by

$$f(x) = \begin{cases} x^3, & \text{if } 0 \leq x \leq 1, \\ 1, & \text{if } 1 < x \leq 2, \\ x^3 & \text{if } x > 2. \end{cases}$$

Since f is non-decreasing, it is both quasiconcave and quasiconvex on $\mathbb{R}$. But f is discontinuous at $x = 2$. Moreover, f is constant on $(1, 2)$, and hence every point in this open interval is a local maximum as well as a local minimum. However, no point in $(1, 2)$ is either a global maximum nor a global minimum. Finally, $f'(0) = 0$, but 0 is neither a local maximum nor a local minimum.

5.2.2 Differentiable quasiconcave functions

Quasiconcave and quasiconvex functions have derivative characterizations as do the concave and convex functions.

Theorem 5.8

Let X be an open and convex set in $\mathbb{R}^n$, and let $f : X \longrightarrow \mathbb{R}$ is a $\mathcal{C}^1$ function. Then f is quasiconcave if and only if for any $x, y \in X$, it is the case that

$$f(y) \geq f(x) \;\Rightarrow\; \nabla f(x) \cdot (y - x) \geq 0.$$

Proof. First, suppose that f is quasiconcave on X, and let $x, y \in X$ such that $f(y) \geq f(x)$. Let $\lambda \in (0, 1)$. Since f is quasiconcave, we have

$$\begin{aligned} & f(x + \lambda(y - x)) \geq \min\{f(x), f(y)\} = f(x) \\ \Rightarrow\; & \frac{f(x + \lambda(y - x)) - f(x)}{\lambda} \geq 0, \quad \text{for all } \lambda \in (0, 1). \end{aligned}$$

Taking the limit of the above expression as $\lambda \to 0^+$ we get $\nabla f(x) \cdot (y - x) \geq 0$.

To show the converse, define by $g(\lambda) := f(x + \lambda(y - x))$. Observe that $g(0) = f(x) \leq f(y) = g(1)$, and that g is a $\mathcal{C}^1$ function on $[0, 1]$ with the derivative $g'(\lambda) = \nabla f(x + \lambda(y - x)) \cdot (y - x)$. Now suppose that, for all $x, y \in X$, we have $f(y) \geq f(x) \Rightarrow \nabla f(x) \cdot (y - x) \geq 0$. We will show that this implies the quasiconcavity of f, i.e.,

$$g(\lambda) = f(x + \lambda(y - x)) \geq \min\{f(x), f(y)\} = f(x) = g(0), \quad \text{for all } \lambda \in [0, 1].$$

Suppose on the contrary that the above is not true, i.e., there is some $\lambda^0 \in (0, 1)$ such that $g(\lambda^0) < g(0)$. Because $g(0) \leq g(1)$, we can chose λ^0 such that $g'(\lambda^0) > 0$. Now let $z^0 = x + \lambda^0(y - x)$. Because $f(x) = g(0) > g(\lambda^0) = f(z^0)$ and $z^0 \in X$, we have, by hypothesis, that

$$\nabla f(z^0) \cdot (x - z^0) = \nabla f(z^0) \cdot (-\lambda^0)(y - x) \geq 0 \;\Rightarrow\; \nabla f(z^0) \cdot (y - x) \leq 0. \tag{5.3}$$

The above is true since $\lambda^0 > 0$. On the other hand, we have chosen λ^0 such that

$$g'(\lambda^0) = \nabla f(z^0) \cdot (y - x) > 0. \tag{5.4}$$

Inequalities (5.3) and (5.4) contradict each other. Therefore, there cannot be such $\lambda^0 \in (0, 1)$ with $g(\lambda^0) < g(0)$, and we conclude that f is quasiconcave. $\square$

It is also possible to have a second-derivative test for quasiconcave functions as their concave counterparts. Let a $\mathcal{C}^2$ function f be defined on some domain $X \subset \mathbb{R}^n$, and let $x \in X$. For $k = 1, \ldots, n$, let $\bar{H}_k[f(x)]$ be the $(k + 1) \times (k + 1)$ matrix given by

$$\bar{H}_k[f(x)] := \begin{bmatrix} 0 & f_1(x) & \ldots & f_k(x) \\ f_1(x) & f_{11}(x) & \ldots & f_{1k}(x) \\ \vdots & \vdots & \ddots & \vdots \\ f_k(x) & f_{k1}(x) & \ldots & f_{kk}(x) \end{bmatrix}$$

The above matrix is the leading principal minor of the bordered Hessian of f at x of order k. The following theorem gives the second-derivative characterization of a quasiconcave function.

Theorem 5.9

Let X be an open and convex set in $\mathbb{R}^n$, and let $f : X \longrightarrow \mathbb{R}$ is a $\mathcal{C}^2$ function. Then

(a) if f is quasiconcave on X, then we have $(-1)^k \, det\left(\bar{H}_k[f(x)]\right) \geq 0$ for $k = 1, \ldots, n$;

(b) if $(-1)^k \, det\left(\bar{H}_k[f(x)]\right) > 0$ for $k = 1, \ldots, n$, then f is quasiconcave on X.

Proof. See Sundaram (1996, p. 217). □

In the above theorem, the week inequality is not a sufficient condition for quasiconcavity. To illustrate the point, let $f : \mathbb{R}^2_+ \longrightarrow \mathbb{R}$ is such that $f(x, y) = (x-1)^2(y-1)^2$. It is easy to check that

$$det(\bar{H}_1[f(x, y)]) = -4(x-1)^2(y-1)^4 \leq 0, \quad \text{for all } \; x, y \in \mathbb{R}^2_+,$$
$$det(\bar{H}_2[f(x, y)]) = 16(x-1)^4(y-1)^4 \geq 0, \quad \text{for all } \; x, y \in \mathbb{R}^2_+,$$

with equalities holding in either case if and only if $x = 1$ or $y = 1$. Notice that $f(0, 0) = f(2, 2) = 1$. Now take $\lambda = 1/2$. Then

$$f((1/2)(0, 0) + (1/2)(2, 2)) = f(1, 1) = 0 < 1 = \min\{f(0, 0), f(2, 2)\},$$

and hence f is not quasiconcave.

Part III

Static optimization

Chapter 6

Static optimization

6.1 Parametric optimization problems in $\mathbb{R}^n$

Let $f : X \times \Theta \longrightarrow \mathbb{R}$ be a given function, where $X \subseteq \mathbb{R}^n$ and $\Theta \subseteq \mathbb{R}^m$. A constrained maximization problem is written as

$$\max_x \{f(x;\, \theta) \mid x \in C(\theta)\}. \tag{P_1}$$

That is, given some value of θ, we look for the value of x that maximizes the function $f(\cdot\,;\, \theta)$ over the set $C(\theta)$. Similarly, a constrained minimization problem is written as

$$\min_x \{f(x;\, \theta) \mid x \in C(\theta)\}. \tag{P_1'}$$

The function $f(.\,;\, \theta)$ is called the *objective function* and the set $C(\theta) \subseteq X$ is called the *constraint set* or *feasible set* which depends on $\theta \in \Theta$. The vector $x = (x_1, \ldots, x_n) \in X$ is a vector of decision or choice variables, and $\theta = (\theta_1, \ldots, \theta_m) \in \Theta$ is a vector of parameters. Intuitively, let Θ represent all possible "environments" in which an agent may find herself, and let X be the set of all "actions" available to her. Given a value of θ, the agent will find her choices restricted to some subset of X. Changes in the parameters will result in changes in the constraint set, as described by the *constraint correspondence* $C : \Theta \to\to X$. Given the objective function, $f(x;\, \theta)$ gives her *payoff* when she faces an environment θ and chooses an action x. The set of optimal actions is described by the *decision rule* or *best-response correspondence* $S : \Theta \to\to X$ that assigns to each θ a subset $S(\theta)$ of X. The set

$$S(\theta) := \arg\max_x \{f(x;\, \theta) \mid x \in C(\theta)\} = \{x \in C(\theta) \mid f(x;\, \theta) \geq f(y;\, \theta) \text{ for all } y \in C(\theta)\}$$

is the set of maximizers of f on $C(\theta)$ whose elements x^* solve the problem (P_1). When $S(\theta)$ is singleton, the best-response correspondence becomes a function, and we write $x^* = x(\theta)$. The payoff accruing to the agent is given by the *(maximum) value function* $V : \Theta \longrightarrow \mathbb{R}$, defined by

$$V(\theta) = \max_x \{f(x;\, \theta) \mid x \in C(\theta)\} = f(x^*;\, \theta), \quad \text{where } x^* \in S(\theta).$$

Given a value of the parameter vector θ, $V(\theta)$ is the maximum attainable payoff to the agent.

Example 6.1: Utility maximization

An agent consumes non-negative quantities of n commodities. The utility function is given by $u(x)$ where $x = (x_1, \ldots, x_n)$ is the vectors of quantities consumed. She has an income $m \geq 0$, and faces a price vector $p = (p_1, \ldots, p_n)$, where $p_i \geq 0$ denotes the price of the i-th commodity. Her budget set,

the constraint set, is given by

$$B(p,\, m) = \{x \in \mathbb{R}^n_+ \mid p \cdot x \leq m\}.$$

The utility maximization problem is written as

$$\max_x \{u(x) \mid x \in B(p,\, m)\}.$$

Suppose there is a unique maximizer $x^* = x(p,\, m) = (x_i(p,\, m))_{i=1}^n$ of the utility maximization problem. The value function is given by $V(p,\, m) = u(x(p,\, m))$. The function $x_i(p,\, m)$ is called the demand function for commodity i, and $V(p,\, m)$ is called the indirect utility function.

Example 6.2: Cost minimization

A firm's cost minimization problem is to identify the combination of n inputs $z = (z_1,\, \ldots,\, z_n)$ that minimizes its total cost of producing at least y units of output, given the production function $f : \mathbb{R}^n_+ \longrightarrow \mathbb{R}_+$ and the input price vector $w = (w_1,\, \ldots,\, w_n) \in \mathbb{R}^n_+$, where w_i denotes the price of the i-th input. The firm's feasible production set is given by

$$F(y) = \{z \in \mathbb{R}^n_+ \mid f(z) \geq y\},$$

and the problem is to solve

$$\min_z \{w \cdot z \mid z \in F(y)\}.$$

Suppose there is a unique minimizer $z^* = z(w,\, y) = (z_i(w,\, y))_{i=1}^n$ of the cost minimization problem. The minimum value function is given by $C(w,\, y) = w \cdot z(w,\, y) = \sum_{i=1}^n w_i\, z_i(w,\, y)$. The function $z_i(w,\, y)$ is called the conditional demand for input i, and $C(w,\, y)$ is called the cost function.

In what follows, we will develop the analysis when the optimization problem is given by (P_1). The analysis of a problem of type (P_1') is analogous because of the fact that $\arg\min_x \{f(x;\, \theta) \mid x \in C(\theta)\} = \arg\max_x \{-f(x;\, \theta) \mid x \in C(\theta)\}$. We state a useful result in the following lemma.

Lemma 6.1

Let $f : X \longrightarrow \mathbb{R}$ where $X \subseteq \mathbb{R}^n$ and $h : \mathbb{R} \longrightarrow \mathbb{R}$ be strictly increasing. Then x^* is a maximum of f on $C(\theta)$ if and only if x^* is a maximum of $h \circ f$ on $C(\theta)$, and z^* is a minimum of f on $C(\theta)$ if and only if z^* is a minimum of $h \circ f$ on $C(\theta)$.

Proof. Suppose that $x^* \in S(\theta) = \arg\max_x \{f(x;\, \theta) \mid x \in C(\theta)\}$. Pick any $x \in C(\theta)$. Then, for any $\theta \in \Theta$, $f(x^*;\, \theta) \geq f(x;\, \theta)$, and since h is strictly increasing, $h(f(x^*;\, \theta)) \geq h(f(x;\, \theta))$. Since x is an arbitrary point in $C(\theta)$, the inequality holds for all $x \in C(\theta)$, and hence x^* is a maximizer of $h \circ f$ over $C(\theta)$.

To show that converse, suppose that x^* maximizes $h \circ f$ over $C(\theta)$, i.e., $h(f(x^*;\, \theta)) \geq h(f(x;\, \theta))$ for all $x \in C(\theta)$, but does not maximize f over $C(\theta)$. Then there exists some $y \in C(\theta)$ such that $f(y;\, \theta)) > f(x^*;\, \theta))$ for any $\theta \in \Theta$. Since h is strictly increasing, we must have $h(f(y;\, \theta)) > h(f(x^*;\, \theta))$, which is a contradiction. $\square$

One important question is under what conditions the set $S(\theta)$ is non-empty. The answers to this question depend obviously on the nature of the objective function and the constraint set. For example, if the objective function is continuous and the constraint set is compact, then $S(\theta)$ is non-empty. This is the Weirstrass theorem we have studied in Chapter 2. In Chapter 2, we have also studied several fixed point theorems that address such problem of existence. In this chapter, our objective is twofold. First, we would analyze the conditions an

optimal solution must satisfy, keeping aside the problem of existence. Second, we would like to analyze how changes in the value of the parameter vector θ change the set of maximizers $S(\theta)$ and the value function $V(\theta)$, which are called the *comparative statics* problems.

6.2 Optimality conditions

6.2.1 Convex constraint sets

Consider the problem

$$\max_x \{f(x) \mid x \in C\}, \tag{6.1}$$

where C is a convex set in $\mathbb{R}^n$, and $f : X \longrightarrow \mathbb{R}$ is a $\mathcal{C}^2$ function where $X \subseteq \mathbb{R}^n$. For the time being, we are omitting the parameter θ since we are interested in the solution to the maximization problem given in (6.1) for a fixed value of θ. Consider a special case when $C = [-1, 2]$ and $f : \mathbb{R} \longrightarrow \mathbb{R}$ defined as $f(x) = 2x^3 - 3x^2 + 2$. Notice that, on C, $f'(0) = f'(1) = 0$, but neither 0 nor 1 is a maximizer. The maximum is achieved at $x = 2$ (we have $f(2) = 6$), but $f'(2) \neq 0$. We have seen that x^* is a maximizer of f implies that $f'(x^*) = 0$. Such a result fails here because the maximum may be achieved at the boundary of the constraint set. The idea is that if x^* maximizes f on C, then as we move away from this point in any *feasible direction* the value of the function must decrease. In what follows, we state this result formally.

Definition 6.1: Feasible direction

Consider the problem stated in (6.1), where C is a convex set. Take any point $x \in C$ and a direction vector $h \in \mathbb{R}^n$. We say that h is a feasible direction from x if there exists some $\delta > 0$ such that $x + th \in C$ for all $t \in (0, \delta)$.

The above definition asserts that h is a feasible direction if any small movement away from x in the direction of h still leaves us inside the feasible set.

Theorem 6.1: First order necessary conditions for a local maximum

Assume that f is a $\mathcal{C}^1$ function, and let x^* be a solution to problem (6.1). Then

$$Df(x^*) \cdot h \leq 0 \tag{6.2}$$

for every feasible direction h from x^*.

Proof. Let x^* be a solution to (6.1), and h be an arbitrary direction vector feasible from x^*. Then there exists a $\delta > 0$ such that $x^* + th \in C$ for all $t \in (0, \delta)$. Because any feasible movement away from x^* reduces the value of f, we have

$$f(x^* + th) - f(x^*) \leq 0.$$

for all t such that $x^* + th \in C$. Dividing the above by $t > 0$ and taking the limit at $t \to 0^+$, we have

$$\lim_{t \to 0^+} \frac{f(x^* + th) - f(x^*)}{t} = Df(x^*; h) = Df(x^*) \cdot h \leq 0.$$

Because f is $\mathcal{C}^1$, the directional derivative exists, and can be written as the scalar product of the derivative and the direction vector. This completes the proof. $\square$

If C is an open set, all points x in C are by definition interior points, and given any x in C, all directions are feasible from it. In this case the inequality $Df(x^*) \cdot h \leq 0$ holds for all h only if all first order partials of f

are zero at x^*. Otherwise, it is possible to increase the value of the function by moving in the direction of (or opposite to) the coordinate vector corresponding to the non-zero partial. For example, suppose that $f_k(x^*) > 0$ and $f_j(x^*) = 0$ for all $j \neq k$, and choose the direction vector h such that $h_k > 0$ and $h_j = 0$ for all $j \neq k$. Then

$$Df(x^*) \cdot h = f_k(x^*)h_k > 0,$$

which contradicts the above theorem. Thus,

Corollary 6.1

Assume that f is a $\mathcal{C}^1$ function and C is open, and let x^* be a solution to problem (6.1). Then

$$Df(x^*) \cdot h = 0.$$

Notice that the result of the above theorem is also a necessary condition for a local maximum of a function on an open set. Now, for an arbitrary direction vector h in $\mathbb{R}^n$, $Df(x^*) \cdot h = 0$ can hold only if $Df(x^*) = 0$. Obviously, the fact that $Df(x^*) = 0$ does not imply that x^* is a local maximum. To see this, consider the function $f : \mathbb{R} \longrightarrow \mathbb{R}$ defined by $f(x) = x^3$. Notice that $f'(0) = 0$, but $x = 0$ is not a maximum of f on $\mathbb{R}$. For a $\mathcal{C}^1$ function $f : X \longrightarrow \mathbb{R}$, if we have that $Df(x) = 0$ for $x \in X$, then x is said to a *critical point* of f. For the above example, $x = 0$ is only a critical point of f. In fact, for this example $x = 0$ is neither a maximum nor a minimum. Such a critical point is called a *saddle point* of the function. For example, for $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ defined by $f(x, y) = x^2 - y^2$, $(0, 0)$ is a saddle point.

Theorem 6.2: Second order necessary conditions for a local maximum

Assume that f is a $\mathcal{C}^2$ function. If f has a local maximum at x^*, then the Hessian matrix of f at x^* is negative semi-definite, i.e.,

$$h^T D^2 f(x^*) h \leq 0 \tag{6.3}$$

for all feasible direction $h \in \mathbb{R}^n$.

Proof. Fix an arbitrary $h \in \mathbb{R}^n$. For any given $\alpha > 0$ we can use Taylor's theorem to write

$$f(x^* + \alpha h) - f(x^*) = Df(x^*)(\alpha h) + \frac{\alpha^2}{2} h^T D^2 f(x^* + \lambda_\alpha \alpha h) h$$

for some $\lambda_\alpha \in (0, 1)$. As x^* is a local maximizer, it is a critical point of f, i.e., $Df(x^*) = 0$, and the above equation reduces to

$$f(x^* + \alpha h) - f(x^*) = \frac{\alpha^2}{2} h^T D^2 f(x^* + \lambda_\alpha \alpha h) h.$$

Moreover, it must be true that, for sufficiently small α, we have $f(x^* + \alpha h) - f(x^*) \leq 0$, and hence

$$h^T D^2 f(x^* + \lambda_\alpha \alpha h) h \leq 0$$

Taking the limit of the above expression as $\alpha \to 0$, we get the desired result. $\square$

The following theorem gives sufficient conditions for a strict local maximum of a function on an open set. That is, $f(x^*) > f(x)$ for all $x \in B_r(x^*)$ for some $r > 0$. A point that satisfies the conditions of the following theorem is said to be a regular maximizer of f on C.

Theorem 6.3: Sufficient conditions for a strict local maximum

Assume that f is a $\mathcal{C}^2$ function with the constraint set C being open and convex, and let x^* be a critical point of f on C, i.e., $Df(x^*) = 0$. If the Hessian matrix of f at x^* is negative definite, then f has a strict local maximum at x^*.

Proof. Fix an arbitrary $h \in \mathbb{R}^n$. By the convexity and openness of C, there exists some $\delta > 0$ such that $x^* + \alpha h \in C$ for all $\alpha \in (0, \delta)$. Fixing some α in this interval, both x^* and $x^* + \alpha h$ lie in C, and Taylor's theorem gives

$$f(x^* + \alpha h) - f(x^*) = Df(x^*)(\alpha h) + \frac{1}{2}(\alpha h)^T D^2 f(x^* + \lambda_\alpha \alpha h)(\alpha h)$$

for some $\lambda_\alpha \in (0, 1)$. Given that $Df(x^*) = 0$, the above equation reduces to

$$f(x^* + \alpha h) - f(x^*) = \frac{\alpha^2}{2} h^T D^2 f(x^* + \lambda_\alpha \alpha h) h = \frac{\alpha^2}{2} Q(\alpha),$$

where $Q(\alpha) \equiv h^T D^2 f(x^* + \lambda_\alpha \alpha h) h$ is a quadratic form for a given h. It can be shown that $Q(\alpha)$ is continuous at $\alpha = 0$. Since $D^2 f(x^*)$ is negative definite by assumption, we have

$$Q(0) = h^T D^2 f(x^*) h < 0.$$

Then continuity of $Q(\cdot)$ at $\alpha = 0$ implies that, for sufficiently small values of α, we have $Q(\alpha) < 0$. Then it follows that

$$f(x^* + \alpha h) - f(x^*) < 0.$$

Because h was chosen arbitrarily, any sufficiently small movement away from x^* reduces the value of f, and hence x^* is a strict local maximum. □

Unfortunately, the above theorem cannot be strengthen to a condition of the sort: "if x^* is a critical point of f such that $D^2 f(x^*)$ is negative semi-definite, then f achieves a local maximum at x^*." To see this, consider $C = \mathbb{R}$ and $f, g : \mathbb{R} \longrightarrow \mathbb{R}$ defined as $f(x) = x^4$ and $g(x) = -x^4$. First notice that $f'(0) = g'(0) = 0$, i.e., $x = 0$ is a critical point of both the functions. We also have $f''(0) = g''(0) = 0$. However at $x = 0$, f reaches a global minimum and g reaches a global maximum. The following theorem gives conditions for the uniqueness of a maximum.

Theorem 6.4: Unique maximum

Let x^* be an optimal solution for the problem in (6.1) with C convex. If f is strictly quasiconcave, then x^* is unique.

Proof. Suppose there are two optimal solutions x' and x'', i.e., $f(x') = f(x'')$. Convexity of C implies that, for any $\lambda \in (0, 1)$, $x^\lambda = \lambda x' + (1-\lambda) x'' \in C$. And strict quasiconcavity of f implies that $f(x^\lambda) > f(x') = f(x'')$ which is a contradiction. □

Example 6.3: Derivation of factor demands

Consider a competitive firm that produces a single output in quantity y using two inputs in quantities x_1 and x_2. The firm's production technology is described by a Cobb-Douglas production function

$$y = f(x_1, x_2) = x_1^\alpha x_2^\beta, \quad \text{where } \alpha, \beta > 0 \text{ and } \alpha + \beta < 1.$$

The firm takes as given the price of output p, and the prices of inputs w_1 and w_2 in order to maximize the profit which is given by

$$\pi(x_1, x_2) = p x_1^\alpha x_2^\beta - w_1 x_1 - w_2 x_2 .$$

The factor demand functions are given by

$$\begin{aligned} x_1(p, w_1, w_2) &= \left[\frac{\alpha p}{w_1}\left(\frac{\beta w_1}{\alpha w_2}\right)^\beta\right]^{\frac{1}{1-\alpha-\beta}}, \\ x_2(p, w_1, w_2) &= \left[\frac{\beta p}{w_2}\left(\frac{\alpha w_2}{\beta w_1}\right)^\alpha\right]^{\frac{1}{1-\alpha-\beta}}. \end{aligned}$$

In an exercise, you are asked solve the above problem.

6.2.2 Inequality constraints: The Karush-Kuhn-Tucker theorem

Necessary conditions for optimality

In this section we will consider the problem of maximizing a non-linear real-valued function subject to a finite collection of non-linear constraints. Let $M = \{1, \ldots, m\}$ be an index set. For each $i \in M$ we have a $\mathcal{C}^2$ function $g^i : \mathbb{R}^n \longrightarrow \mathbb{R}$ which will give rise to a constraint. The objective function will be a $\mathcal{C}^2$ function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$. Consider the following maximization problem

$$\max\{f(x) \mid g^i(x) \geq 0 \text{ for } i \in M\}. \tag{KKT}$$

Recall that x is a local maximum for problem (KKT) if there is some $r > 0$ such that $f(x) \geq f(x')$ for all $x' \in B_r(x) \cap C$. Since f is a $\mathcal{C}^1$ function, by Taylor's theorem we have the following:

$$f(x + \alpha h) = f(x) + (\alpha h) \cdot \nabla f(x) + R(\alpha),$$

for $\alpha > 0$. For sufficiently small values of α, $h \cdot \nabla f(x) > 0$ implies that $f(x + \alpha h) > f(x)$. We will often make use of the above result. Theorem 7 in Chapter 4, which is known as "Gordon's theorem of alternatives" is very useful in establishing the existence of the Lagrange multipliers for constrained optimization problems. Following theorem, known as the Fritz John conditions, is one of the earliest results in the theory of constrained optimization.

Theorem 6.5: Fritz John

Suppose $M \neq \varnothing$ and x^* is a local maximum for problem (KKT). Then there exists a set of non-negative multipliers $\{\mu_0, \mu_1, \ldots, \mu_m\}$ not all zero such that

$$\mu_0 \nabla f(x^*) + \sum_{i \in M} \mu_i \nabla g^i(x^*) = 0, \tag{6.4}$$

$$\mu_i g^i(x^*) = 0, \quad \text{for } i \in M. \tag{6.5}$$

Proof. Let $z^i := \nabla g^i(x^*)$ for $i \in M$ and $z^0 := \nabla f(x^*)$. If the theorem were true we can divide Equation (6.4) throughout by $\sum_{i \in M \cup \{0\}} \mu_i$, and Equation (6.4) can then be interpreted as saying that 0 belongs to the convex hull of z^i's. This is what we intend to prove. Suppose the contrary, i.e., $0 \notin co(\{z^0, z^1, \ldots, z^m\})$. Recall that the convex hull of a finite number of vectors is closed. Then, by the Separating Hyperplane Theorem, there is a vector h in $\mathbb{R}^n$ such that $h \cdot z^i > 0$ for all $i = 0, 1, \ldots, m$. Since each g^i is differentiable, by Taylor's theorem we have for $i \in M$ and for any $\alpha \in (0, 1)$

$$g^i(x^* + \alpha h) = g^i(x^*) + (\alpha h) \cdot \nabla g^i(x^*) + R(\alpha) = (1 - \alpha) g^i(x^*) + \alpha g^i(x^*) + (\alpha h) \cdot \nabla g^i(x^*) + R(\alpha),$$

where the error term $R(\alpha)$ is quadratic in α. Since $g^i(x^*) \geq 0$ for $i \in M$, the above expression reduces to

$$g^i(x^* + \alpha h) \geq \alpha[g^i(x^*) + h \cdot \nabla g^i(x^*)] + R(\alpha).$$

Since $h \cdot z^i = h \cdot \nabla g^i(x^*) > 0$ and $g^i(x^*) \geq 0$, for $\alpha > 0$ we have that

$$\frac{g^i(x^* + \alpha h)}{\alpha} \geq g^i(x^*) + h \cdot \nabla g^i(x^*) + \frac{R(\alpha)}{\alpha} > 0, \quad \text{for all } i \in M. \tag{6.6}$$

The above inequality implies that $x^* + \alpha h$ is feasible for α sufficiently small. On the other hand, $h \cdot z^0 > 0$ implies that $f(x^* + \alpha h) > f(x^*)$ for $\alpha > 0$ and sufficiently small, which contradicts the local optimality of x^*. Now we prove the second part of the theorem. Consider the following system of strict inequalities.

$$\begin{aligned} &g^i(x^*) + h \cdot \nabla g^i(x^*) > 0, \quad \text{for all } i \in M, \\ &h \cdot \nabla f(x^*) > 0 \end{aligned}$$

The first inequality is implied by condition (6.6) for small values of α, and the second one is equivalent to $h \cdot z^0 > 0$. If $0 \in co(\{z^0, z^1, \ldots, z^m\})$, i.e., if x^* is a local maximum, then there is no $h \in \mathbb{R}^n$ that solves the above system of inequalities. Since the functions $h \cdot \nabla f(x^*)$ and $g^i(x^*) + h \cdot \nabla g^i(x^*)$ for $i \in M$ are affine functions, by Gordon's theorem of the alternative (in Chapter 4) we can find non-negative multipliers $\mu_0, \mu_1, \ldots, \mu_m$ not all zero such that

$$h \cdot \left[\mu_0 \nabla f(x^*) + \sum_{i \in M} \mu_i \nabla g^i(x^*)\right] + \sum_{i \in M} \mu_i g^i(x^*) \leq 0, \quad \text{for all } h \in \mathbb{R}^n.$$

Choosing $h = \mu_0 \nabla f(x^*) + \sum_{i \in M} \mu_i \nabla g^i(x^*)$, the above inequality reduces to

$$\|h\|^2 + \sum_{i \in M} \mu_i g^i(x^*) \leq 0.$$

Given that $\|h\|^2 \geq 0$, we must have $\sum_{i \in M} \mu_i g^i(x^*) \leq 0$. On the other hand, $g^i(x^*) \geq 0$ and $\mu_i \geq 0$ for all $i \in M$ imply that $\sum_{i \in M} \mu_i g^i(x^*) \geq 0$. Thus, $\sum_{i \in M} \mu_i g^i(x^*) = 0$. Given that the multipliers are non-negative, we have $\mu_i g^i(x^*) = 0$ for each $i \in M$. $\square$

The above lemma may fail if M is not a finite set since we cannot guarantee the existence of an $\alpha > 0$ such that $g^i(x^* + \alpha h) > 0$ for all $i \in M$. We also may not be able to characterize the optimal solution x^* if $\mu_0 = 0$. In this case, Equation (6.4) reduces to

$$\sum_{i \in M} \mu_i \nabla g^i(x^*) = 0.$$

It is possible to encounter non-linear programing problem of the sort described in the above equation. To see this, consider the following example.

Example 6.4

Let $f(x_1, x_2) = x_2$, and the constraints given by $g^1(x_1, x_2) = x_1 \geq 0$ and $g^2(x_1, x_2) = -x_1 - x_2^2 \geq 0$. Notice that the only feasible solution is $(x_1^*, x_2^*) = (0, 0)$. Now $\nabla f(0, 0) = (0, 1)$, $\nabla g^1(0, 0) = (1, 0)$ and $\nabla g^2(0, 0) = (-1, 0)$. Thus, Equation (6.4) yields:

$$\mu_0(0, 1) + \mu_1(1, 0) + \mu_2(-1, 0) = 0.$$

All non-negative solutions of the above equation have $\mu_0 = 0$, and hence $\nabla f(0, 0)$ cannot lie in the convex hull of $\nabla g^1(0, 0)$ and $\nabla g^2(0, 0)$. Hence, without additional assumptions we cannot guarantee that $\mu_0 > 0$. These additional assumptions are called *constraint qualifications*, which says that at a local maximum x^* for the problem (KKT), the vectors in $\{\nabla g^i(x^*)\}_{i \in M}$ are linearly independent.

Theorem 6.6: Karush-Kuhn-Tucker

Suppose $M \neq \varnothing$ and x^* is a local maximum for problem (KKT). If the vectors in $\{\nabla g^i(x^*)\}_{i \in M}$ are linearly independent, then there exists a set of non-negative multipliers $\{\lambda_1, \ldots, \lambda_m\}$ not all zero such

that

$$\nabla f(x^*) + \sum_{i \in M} \lambda_i \nabla g^i(x^*) = 0, \tag{6.7}$$

$$\lambda_i g^i(x^*) = 0, \quad \text{for } i \in M. \tag{6.8}$$

Proof. Apply Theorem 6.5. If $\mu_0 > 0$, then divide Equations (6.4) and (6.5) by μ_0 to obtain the result where $\lambda_i = \mu_i/\mu_0$. If $\mu_0 = 0$, then $\sum_{i \in M} \mu_i \nabla g^i(x^*) = 0$. However, the linear independence of the vectors in $\{\nabla g^i(x^*)\}_{i \in M}$ implies that $\mu_i = 0$ for all $i \in M$, which is a contradiction. □

Stated component-wise, Equation (6.7) reads

$$\frac{\partial f}{\partial x_j}(x^*) + \sum_{i \in M} \lambda_i \frac{\partial g^i}{\partial x_j}(x^*) = 0 \text{ for all } j = 1, \ldots, n.$$

Condition (6.8) is known as the *complementary slackness* conditions, which implies that for some $i \in M$ it cannot be the case that both λ_i and $g^i(x^*)$ are strictly positive ("slack"). Thus if $g^i(x^*) > 0$, then $\lambda_i = 0$, and if $\lambda_i > 0$, then $g^i(x^*) = 0$.

Now we consider a special case of the maximization problem (KKT) in which the inequality constraints are replaced by equality constraints.

$$\max\{f(x) \mid g^i(x) = 0 \text{ for } i \in M\}. \tag{L}$$

The theorem of Lagrange gives the first order optimality conditions for a local maximum of the above problem.

Theorem 6.7: Lagrange

Suppose $M \neq \varnothing$ and x^* is a local maximum for problem (L). If the vectors in $\{\nabla g^i(x^*)\}_{i \in M}$ are linearly independent, then there exists a set of multipliers $\{\lambda_1^*, \ldots, \lambda_m^*\}$ not all zero such that

$$\nabla f(x^*) + \sum_{i \in M} \lambda_i^* \nabla g^i(x^*) = 0. \tag{6.9}$$

Proof. Given the optimality of x^* we will construct a vector of multipliers $\lambda^* \in \mathbb{R}^m$ which satisfies condition (6.9). Let $x = (z, y)$ where $z \in \mathbb{R}^m$ is the first m coordinate vectors of x, and $y \in \mathbb{R}^{n-m}$ is the last $n - m$ coordinate vectors of x. Thus, $x^* = (z^*, y^*)$. Denote by $g(x)$ the vector $(g^1(x), \ldots, g^m(x))$. The linear independence assumption implies that $\rho(\nabla g_z(z^*, y^*)) = m$, where $\rho(A)$ is the rank of a matrix A. Further, given a function $F(z, y)$ from $\mathbb{R}^{l+k}$ to $\mathbb{R}^l$, $\nabla F_z(z, y)$ denotes the portion of the matrix $\nabla F(z, y)$, which is an $l \times (l + k)$ matrix, corresponding to the first l variables, and $\nabla F_y(z, y)$ denotes the portion of the matrix $\nabla F(z, y)$ corresponding to the last k variables. We have to show the existence of λ^* such that

$$\nabla f_z(z^*, y^*) + \lambda^* \cdot \nabla g_z(z^*, y^*) = 0, \tag{6.10}$$

$$\nabla f_y(z^*, y^*) + \lambda^* \cdot \nabla g_y(z^*, y^*) = 0. \tag{6.11}$$

Since $\rho(\nabla g_z(z^*, y^*)) = m$, by the Implicit Function Theorem, there are an open set $V \subset \mathbb{R}^{n-m}$ containing y^* and a $\mathcal{C}^1$ function $h : V \longrightarrow \mathbb{R}^m$ such that $z^* = h(y^*)$ and $g(h(y), y) = 0$ for all $y \in V$. Therefore,

$$\nabla h(y^*) = -(\nabla g_z(z^*, y^*))^{-1} \nabla g_y(z^*, y^*). \tag{6.12}$$

Now define

$$\lambda^* := -(\nabla g_z(z^*, y^*))^{-1} \nabla f_z(z^*, y^*), \tag{6.13}$$

which immediately implies Equation (6.10). Now, define the function $F : V \longrightarrow \mathbb{R}$ by $F(y) = f(h(y), y)$. Since $(h(y^*), y^*)$ is local maximum of f on an open set V, we have $\nabla F(y^*) = 0$, i.e.,

$$\nabla f_z(z^*, y^*)\nabla h(y^*) + \nabla f_y(z^*, y^*) = 0. \tag{6.14}$$

Substituting for $\nabla h(y^*)$ as in (6.12) and using the definition of λ^* in (6.13) in the above equation, we get condition (6.11). □

Notice that although the maximization problem in (L) is a special case of that in (KKT), the above theorem is not a corollary to the Karush-Kuhn-Tucker theorem. In Theorem 6.7, the complementary slackness conditions are trivially satisfied by any $\lambda^* \in \mathbb{R}^m$ (without any restrictions on the sign) because at the optimum $g^i(x^*) = 0$ for all $i \in M$. The following theorem generalizes Theorems 6.6 and 6.7. Append to the index set M of the maximization problem (KKT) an additional set $M^= := \{j \mid g^j(x) = 0\}$. Now consider the following maximization problem.

$$\max\{f(x) \mid g^i(x) \geq 0 \text{ for all } i \in M, \text{ and } g^j(x) = 0 \text{ for all } j \in M^=\}. \tag{P}$$

Theorem 6.8

Let x^* be a local maximum for problem (P). If the vectors in $\{\nabla g^k(x^*)\}$ for $k \in M^= \cup \{i \in M \mid g^i(x^*) = 0\}$ are linearly independent, then there exists a set of multipliers $\{\mu_i^*\}_{i \in M \cup M^=}$ such that

(a) $\nabla f(x^*) + \sum_{i \in M \cup M^=} \mu_i^* \nabla g^i(x^*) = 0$,

(b) $\mu_i^* \geq 0$ for all $i \in M$,

(c) μ_i^* is unrestricted for all $i \in M^=$,

(d) $\mu_i^* g^i(x^*) = 0$ for all $i \in M$,

(e) $g^i(x^*) \geq 0$ for all $i \in M$, and

(f) $g^i(x^*) = 0$ for all $i \in M^=$.

Proof. See Vohra (2005, p. 93). □

The above theorem gives the optimality conditions for a maximization problem when both equality and inequality constraints are involved. We should not confuse the constraints for $j \in M^=$ with those for i in $\{i \in M \mid g^i(x^*) = 0\}$, which is a subset of M such that $g^k(x^*) = 0$ for k belonging to this set. This is the set of *binding* or *active* constraints at the optimum.

Using the optimality conditions

In what follows we describe "cookbook" procedures for using the optimality conditions stated in Karush-Kuhn-Tucker and Lagrange theorems. For a detailed discussion on when such procedures work and when they fail, (see Sundaram, 1996).

A cookbook procedure for the KKT theorem

Consider the maximization problem (KKT). To solve such a problem in practice, one may follow the following three steps. In the first step, we form a function $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \longrightarrow \mathbb{R}$, called the *Lagrangean*, which is given by

$$\mathcal{L}(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g^i(x). \tag{6.15}$$

In the second step, find all solutions (x, λ) to the following set of equations:

$$\frac{\partial \mathcal{L}}{\partial x_j}(x, \lambda) = 0, \quad \text{for } j = 1, \ldots, n, \tag{6.16}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_i}(x, \lambda) \geq 0, \ \lambda_i \geq 0, \ \lambda_i \frac{\partial \mathcal{L}}{\partial \lambda_i}(x, \lambda) = 0 \ \text{ for } i = 1, \ldots, m. \tag{6.17}$$

Let L be the set of solutions to the above system. In the third step, we compute the value of f at each x in the set $\{x \mid \text{there is } \lambda \text{ such that } (x, \lambda) \in L\}$. In practice, the value of x that maximizes f over this set is typically also the solution to the original maximization problem (KKT).

Example 6.5: Numerical example

onsider the following maximization problem:

$$\max\{x^2 - y \mid x^2 + y^2 \leq 1\}. \tag{6.18}$$

Let $f(x, y) := x^2 - y$ and $g(x, y) := 1 - x^2 - y^2$. First notice that the constraint set $C := \{x, y \mid x^2 + y^2 \leq 1\}$ is the unit disc in $\mathbb{R}^2$, and hence is compact, and the objective function is continuous. Therefore by Weirstrass theorem, the is at least one maximum of f over C. Next, notice that at the points (x, y) at which the constraint is binding, we must either have $x \neq 0$ or $y \neq 0$ to have $x^2 + y^2 = 1$. Since $\nabla g(x, y) = (-2x, -2y)$, it follows that at all such points where the constraint is binding, we must have $\rho(\nabla g(x, y)) = 1$. Therefore the constraint qualification is satisfied if the optimum occurs on the boundary of the disc. If the optimum occurs at a point where $g(x, y) > 0$, then the set of binding constraints is empty, and the constraint qualification holds vacuously. Now write down the Lagrangean as follows.

$$\mathcal{L}(x, y, \lambda) = x^2 - y + \lambda(1 - x^2 - y^2).$$

The optimal solutions to Problem (6.18) must be given by the following system:

$$2x(1 - \lambda) = 0, \tag{6.19}$$

$$-1 - 2\lambda y = 0, \tag{6.20}$$

$$\lambda \geq 0, \ \ 1 - x^2 - y^2 \geq 0, \ \ \lambda(1 - x^2 - y^2) = 0. \tag{6.21}$$

Notice that λ cannot be equal to 0 in order to have Equation (6.20) satisfied. Hence, we must have $1 - x^2 - y^2 = 0$, i.e., the constraint must bind at the optimum. For the first equation to hold, we must have $x = 0$ or $\lambda = 1$. If $\lambda = 1$, then Equation (6.20) implies that $y = -1/2$. Then the binding constraint implies that $x = \pm\sqrt{3}/2$. Thus we get two candidate solutions which are given by

$$(x, y, \lambda) = \left(-\frac{\sqrt{3}}{2}, -\frac{1}{2}, 1\right) \ \text{ and } \ \left(\frac{\sqrt{3}}{2}, -\frac{1}{2}, 1\right).$$

The value of f at both the candidate solutions is equal to $5/4$. Next, we have to check whether $x = 0$ is a candidate solution. In this case the equation in (6.21) implies that we must have $y = \pm 1$. But $y = 1$ is inconsistent with the fact that $\lambda > 0$ since Equation (6.20) implies, in this case, that $\lambda = -1/2$. Thus, the other candidate solution is given by

$$(x, y, \lambda) = \left(0, -1, \frac{1}{2}\right).$$

At the above candidate optimum, $f(0, -1) = 1 < 5/4$, and hence we can discard this point. Since there is no other candidate solution, there are exactly two optimal solutions, namely $(-\sqrt{3}/2, -1/2)$ and $(\sqrt{3}/2, -1/2)$. Also at these two points the sufficient conditions are clearly satisfied. These two points cannot be minima as we have discarded the candidate optimum $(0, -1)$ because at this point f reaches a value equal to 1 which is lower than $5/4$.

A cookbook procedure for the Lagrange theorem

The procedure is similar to that of the KKT theorem. Consider the maximization problem (L). In the first step, write down the *Lagrangean*, which is given by

$$\mathcal{L}(x,\,\lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g^i(x). \tag{6.22}$$

In the second step, find all solutions $(x,\,\lambda)$ to the following set of equations:

$$\frac{\partial \mathcal{L}}{\partial x_j}(x,\,\lambda) = 0, \quad \text{for } j = 1,\,\ldots,\,n, \tag{6.23}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_i}(x,\,\lambda) = 0, \quad \text{for } i = 1,\,\ldots,\,m. \tag{6.24}$$

Let L' be the set of solutions to the above system, which is given by

$$L' := \{(x,\,\lambda) \mid \nabla\mathcal{L}(x,\,\lambda) = 0\}.$$

In the third step, we compute the value of f at each x in the set $\{x \mid \text{there is } \lambda \text{ such that } (x,\,\lambda) \in L'\}$.

Discussion on the Lagrangean method

We will give an intuitive interpretation of the Lagrangean method described above. Consider the following problem.

$$\max_{x_1,\,x_2}\{f(x_1,\,x_2) \mid g(x_1,\,x_2) = c\}. \tag{6.25}$$

Instead of directly forcing the agent to respect the constraint, imagine that we allow her to choose the values of x_1 and x_2 freely, but make her pay a fine λ "per unit violation" of the restriction. The agent's payoff, net of the penalty, is given by the Lagrangean function:

$$\mathcal{L}(x_1,\,x_2,\,\lambda) = f(x_1,\,x_2) - \lambda[c - g(x_1,\,x_2)]. \tag{6.26}$$

The agent maximizes (6.26), taking λ as given. The first-order conditions are given by:

$$\frac{\partial \mathcal{L}}{\partial x_1}(x_1^*,\,x_2^*,\,\lambda) = 0, \tag{L1}$$

$$\frac{\partial \mathcal{L}}{\partial x_2}(x_1^*,\,x_2^*,\,\lambda) = 0. \tag{L2}$$

Given an arbitrary λ, there is no guarantee that the solution to the above system will be an optimal solution to problem (6.25). But if we pick the correct penalty λ^*, the agent will choose to respect the constraint even if in principle she is free not to do so. Hence, apart from (L1) and (L2), the optimal solution $(x_1^*,\,x_2^*,\,\lambda^*)$ must also satisfy the constraint which can be written as

$$\frac{\partial \mathcal{L}}{\partial \lambda}(x_1^*,\,x_2^*,\,\lambda^*) = g(x_1^*,\,x_2^*) - c = 0. \tag{F}$$

Thus, the Lagrangean method is a way in which one transforms a constrained optimization problem to an unconstrained one.

Sufficient conditions for optimality

The conditions given in Theorems 6.6 and 6.7 are only the necessary conditions that an optimal solution x^* must satisfy. These conditions only assert that x^* is a critical point of the Lagrangean, but do not say anything about

whether x^* is a minimum or a maximum. If the objective and the constraint functions are twice continuously differentiable, then the second order conditions sometimes give the sufficient conditions for optimality. As we have mentioned above that the maximization of the Lagrangean function is an unconstrained maximization problem, we can make use of the theorems on sufficiency in the previous section along with our knowledge of concavity and quasiconcavity to get the sufficient conditions in this context. We state two important theorems. The first one guarantees a strict local maximum of a the Lagrange problem, and the second theorem gives sufficient conditions for uniqueness for the KKT problem.

Theorem 6.9: Sufficient condition for a strict local maximum

Let $x^* \in \mathbb{R}^n$ be a feasible point for the problem (L) for some $\lambda^* \in \mathbb{R}^m$ such that the vectors in $\{\nabla g(x^*)\}_{i \in M}$ are linearly independent, and let (x^*, λ^*) satisfies condition (6.9) in Theorem 6.7. Define

$$\mathcal{H} := \{h \in \mathbb{R}^n \mid \nabla g(x^*) \cdot h = 0\},$$

and let $\nabla_x^2 \mathcal{L}(x^*, \lambda^*) = \nabla^2 f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla^2 g^i(x^*)$ denote the $n \times n$ matrix of derivatives of the Lagrangean with respect to x at (x^*, λ^*). If $h^T \nabla_x^2 \mathcal{L}(x^*, \lambda^*) h < 0$ for all $h \in \mathcal{H}$ with $h \neq 0$, then x^* is a strict local maximum for the problem (L).

Proof. See Sundaram (1996, p. 118). □

Theorem 6.10: Uniqueness

Let x^* be an optimal solution to the problem (KKT). If f is strictly quasiconcave and the constraint function g^i for all $i \in M$ are quasiconcave, then x^* is the unique optimal solution.

Proof. Notice that the set $C^i := \{x \in \mathbb{R}^n \mid g^i(x) \geq 0\}$ is convex because g^i is quasiconcave. x^* maximizes the strictly quasiconcave function f over the convex set $C = \cap_{i \in M} C^i$. Hence, the theorem follows from Theorem 6.4. □

Application

In this subsection we analyze one problem from economics in order to illustrate the the usage of Theorems 6.6 and 6.7.

Optimal consumption choice by utility maximization

A consumer chooses quantities $(x_1, x_2) \in \mathbb{R}_+^2$ of goods 1 and 2, given the price vector (p_1, p_2) with $p_i > 0$ for $i = 1, 2$ and given her income $m > 0$. Her utility function is given by $u(x_1, x_2) = x_1 x_2$. Formally, the consumer solves the following problem.

$$\max\{x_1 x_2 \mid p_1 x_1 + p_2 x_2 \leq m,\ x_1 \geq 0,\ x_2 \geq 0\}. \tag{6.27}$$

First notice that the budget set, which is given by

$$B(p_1, p_2, m) = \{(x_1, x_2) \in \mathbb{R}^2 \mid p_1 x_1 + p_2 x_2 \leq m,\ x_1 \geq 0,\ x_2 \geq 0\}, \tag{6.28}$$

is compact, and the utility function $u(\cdot)$ is continuous. Thus, by Weirstrass Theorem, a solution (x_1^*, x_2^*) to the maximization problem (6.27) exists.

Now, if either $x_1 = 0$ or $x_2 = 0$, then $u(x_1, x_2) = 0$. On the other hand, the consumption point $(\bar{x}_1, \bar{x}_2) = (m/2p_1, m/2p_2)$ is feasible and gives a utility $u(\bar{x}_1, \bar{x}_2) = m^2/4p_1p_2 > 0$. Since any solution (x_1^*, x_2^*) must

be such that $u(x_1^*, x_2^*) \geq u(\bar{x}_1, \bar{x}_2)$, we must have that $x_i^* > 0$ for $i = 1, 2$. Then the complementary slackness conditions of Theorem 6.6 imply that the multipliers associated with the constraints $x_1 \geq 0$ and $x_2 \geq 0$ must be zero at the optimal solution. Further notice that, given the monotonicity of the preferences, the third constraint must hold with equality. Therefore, the reduced budget set is given by

$$B^*(p_1, p_2, m) = \{(x_1, x_2) \in \mathbb{R}^2 \mid p_1 x_1 + p_2 x_2 = m\}. \tag{6.29}$$

Now we are within the settings of the Lagrange theorem. The Lagrangean is given by:

$$\mathcal{L}(x_1, x_2, \lambda) = x_1 x_2 + \lambda[m - p_1 x_1 + p_2 x_2].$$

The critical points of $\mathcal{L}(x_1, x_2, \lambda)$ are given by

$$\begin{aligned} x_2 - \lambda p_1 &= 0, \\ x_1 - \lambda p_2 &= 0, \\ m - p_1 x_1 - p_2 x_2 &= 0. \end{aligned}$$

If $\lambda = 0$, then the above system has no solutions. Hence, we have that $\lambda \neq 0$. The first two equations imply that $\lambda = p_1/x_1 = p_2/x_2$, so $x_2 = (p_1/p_2)x_1$. Using this in the third equation, we see that the unique solution to the set of above equations is given by $x_1^* = m/2p_1$, $x_2^* = m/2p_2$ and $\lambda^* = m/2p_1p_2 > 0$. Also, the other two inequality constraints $x_i^* > 0$ for $i = 1, 2$ are also satisfied.

Now we will use the second-order conditions to show that (x_1^*, x_2^*) is a strict local maximum of u on $B^*(p_1, p_2, m)$. Notice that $\nabla g(x^*) = (-p_1, -p_2)$. Hence we have that

$$\mathcal{H} = \{h \in \mathbb{R}^2 \mid \nabla g(x^*) \cdot h = 0\} = \{h \in \mathbb{R}^2 \mid h_1 = -(p_2 h_2)/p_1\}.$$

And the second derivative of the Lagrangean at the optimum with respect to (x_1, x_2) is given by

$$\nabla_x^2 \mathcal{L}(x^*, \lambda^*) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + \lambda^* \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

For $h \in \mathcal{H}$ we have $h^T \nabla_x^2 \mathcal{L}(x^*, \lambda^*) h = -2p_2 h_2^2/p_1 < 0$. Hence, by Theorem 6.9 we have that (x_1^*, x_2^*) is a strict local maximum. Moreover, (x_1^*, x_2^*) is the unique maximum. Notice that the budget set $B(p_1, p_2, m)$ is convex and the objective function $u(\cdot)$ is strictly quasiconcave, and the claim follows from Theorem 6.10.

6.3 Comparative statics

6.3.1 Comparative statics for smooth optimization problems

We consider the parametric maximization problem (P_1). In this subsection, we first analyze several properties of the value function $V(\theta)$ and of the optimal solution correspondence $S(\theta)$. Next, we study the behavior of the value function and the optimal solution with respect to the parameter θ. This is known as the *comparative static analysis*.

Theorem of the maximum

The following theorem states some properties of the optimal solution to the maximization problem (P_1) and of the value function.

Theorem 6.11: Berge

Given the sets $X \subseteq \mathbb{R}^n$ and $\Theta \subseteq \mathbb{R}^m$, let $f : X \times \Theta \longrightarrow \mathbb{R}$ be a continuous function, and $C(\theta)$ is compact for each $\theta \in \Theta$. Then

(a) the solution correspondence $S : \Theta \rightarrow\rightarrow X$ which is given by $S(\theta) := \arg\max_x \{f(x;\, \theta) \mid x \in C(\theta)\}$ is non-empty, and the value function $V(\theta) = \max_x \{f(x;\, \theta) \mid x \in C(\theta)\}$ of the problem P_1 is continuous;

(b) if $f : X \times \Theta \longrightarrow \mathbb{R}$ is concave on $X \times \Theta$, and $C(\theta)$ is convex for each $\theta \in \Theta$, then $S(\theta)$ is a convex set. Moreover, if $f(x,\, \theta)$ is strictly concave, then the set $S(\theta)$ is singleton, i.e., $S(\theta) = \{x(\theta)\}$. The function $x(\theta)$ is a continuous function.

Proof. Omitted. □

Applications

In what follows we analyze two concrete examples from economics where the above theorems readily apply.

Walrasian equilbrium:

Consider an exchange economy $\xi = (u^i,\, \omega^i)_{i=1}^n$ where $u^i : \mathbb{R}_+^L \longrightarrow \mathbb{R}$ is the utility function of consumer $i = 1,\, \ldots,\, n$, which is continuous and strictly quasiconcave, and $\omega^i \in \mathbb{R}_+^L$ is her endowment vector. An allocation $x = (x^1,\, \ldots,\, x^n) \in \mathbb{R}_+^{n \times L}$ is a vector that describes the amount of each commodity consumed by each agent. An allocation is *feasible* if the consumption of each commodity does not exceed its total endowment, i.e., $\sum_{i=1}^n x^i \leq \sum_{i=1}^n \omega^i$.

Definition 6.2: Walrasian equilibrium

A Walrasian equilibrium for the exchange economy ξ is a price-allocation pair $(p^*,\, x^*)$ such that (i) all agents maximize utility taking p^* as given, and (ii) the market for each commodity clears.

First notice that, given a price vector p, the demand vector of consumer i is given by:

$$x^i(p,\, p \cdot \omega^i) \in \text{argmax}_{x^i} \{u^i(x^i) \mid p \cdot x^i \leq p \cdot \omega^i\}.$$

From Berge's theorem it follows that the demand function $x^i(p,\, p \cdot \omega^i)$ is continuous for any strictly positive price vector. Also the demand function is homogeneous of degree zero. One obtains the *aggregate demand function* by summing the individual demands across all consumers, i.e.,

$$x(p,\, \omega) = \sum_{i=1}^n x^i(p,\, p \cdot \omega^i), \tag{6.30}$$

and the *aggregate excess demand* is given by:

$$z(p,\, \omega) = \sum_{i=1}^n [x^i(p,\, p \cdot \omega^i) - \omega^i] = x(p,\, \omega) - \omega. \tag{6.31}$$

It is not so difficult to show that $x(p,\, \omega)$ and $z(p,\, \omega)$ are continuous functions, and are homogeneous of degree zero in prices. Now notice that the individual budget constraints must hold with equality, i.e.,

$$p \cdot x^i(p,\, p \cdot \omega^i) = p \cdot \omega^i.$$

Summing over all consumers in the above equation we get the *Walras' law*:

$$p \cdot z(p, \omega) = 0. \tag{6.32}$$

It is worthwhile to note two important implications of the above result. First, Walras' law implies that if the markets of any $L-1$ goods clear, then the remaining market must automatically clear. Second, the homogeneity of degree zero of $z(p, \omega)$ means that only relative prices matter. Formally, it allows us to normalize prices and worry about only $L-1$ markets. Hence, we do the following normalization such that $p \in \Delta$ where Δ is the unit simplex in $\mathbb{R}^L_+$, i.e., $\sum_{l=1}^L p_l = 1$. Formally,

$$\Delta := \left\{ p \in \mathbb{R}^L_+ : \sum_{l=1}^L p_l = 1 \right\}.$$

It is easy to show that Δ is a compact and convex set. For simplicity we will assume that each of the individual utility functions is strictly quasiconcave which implies that the aggregate excess demand $z(p, \omega)$ is single-valued, and abuse notations to express it as $z(p)$ since the function is defined for constant endowment vectors. The following theorem asserts that a Walrasian equilibrium for the economy ξ exists.

Theorem 6.12

Let $z : \Delta \longrightarrow \mathbb{R}^L$ be the aggregate excess demand function which is continuous. Then there exists a price vector $p^* \in \Delta$ such tha $z(p^*) \leq 0$.

Proof. Define a price-adjustment rule $g : \Delta \longrightarrow \Delta$ as follows.

$$g_l(p) = \frac{p_l + \max\{0, z_l(p)\}}{1 + \sum_{l=1}^L \max\{0, z_l(p)\}} \quad \text{for} \quad l = 1, \ldots, L.$$

It is easy to show that $g_l(p)$ is continuous for each l, and $g(p) = (g_1(p), \ldots, g_L(p)) \in \Delta$. It follows from the Brouwer's fixed-point theorem that there is a price vector $p^* \in \Delta$ such that $g(p^*) = p^*$. Then

$$p_l^* = \frac{p_l^* + \max\{0, z_l(p^*)\}}{1 + \sum_{l=1}^L \max\{0, z_l(p^*)\}} \quad \text{for} \quad l = 1, \ldots, L. \tag{6.33}$$

We will show that the price vector p^* are the Walrasian prices. Notice that the above equation implies

$$\begin{aligned}
& p_l^* + p_l^* \sum_{l=1}^L \max\{0, z_l(p^*)\} = p_l^* + \max\{0, z_l(p^*)\} \text{ for } l = 1, \ldots, L, \\
\implies & p_l^* \sum_{l=1}^L \max\{0, z_l(p^*)\} = \max\{0, z_l(p^*)\} \text{ for } l = 1, \ldots, L, \\
\implies & z_l(p^*) p_l^* \sum_{l=1}^L \max\{0, z_l(p^*)\} = z_l(p^*) \max\{0, z_l(p^*)\} \text{ for } l = 1, \ldots, L,
\end{aligned}$$

Adding up the above L equations we get:

$$\left[\sum_{j=1}^L \max\{0, z_j(p^*)\} \right] \left[\sum_{l=1}^L p_l^* z_l(p^*) \right] = \sum_{l=1}^L z_l(p^*) \max\{0, z_l(p^*)\}.$$

The Walras' law implies that

$$\sum_{l=1}^L p_l^* z_l(p^*) = 0.$$

Hence,

$$\sum_{l=1}^{L} z_l(p^*)\max\{0,\, z_l(p^*)\} = \sum_{l=1}^{L} \tilde{z}_l(p^*) = 0.$$

Each term $\tilde{z}_l(p^*)$ of the above summation is greater than or equal to zero because $\tilde{z}_l(p^*) \in \{0,\, z_l^2(p^*)\}$. In fact, each $\tilde{z}_l(p^*)$ must equal zero, otherwise the last equality would not hold, and hence $z_l(p^*) \leq 0$ for $l = 1, \ldots, L$. In other words, p^* is a Walrasian price vector. $\square$

Nash equilibrium:

Let $\Gamma = \langle N,\, \{P_i\}_{i\in N},\, \{\pi_i(\cdot)\}_{i\in N}\rangle$ be a normal-form Bertrand game, where $N = \{1, \ldots, n\}$ is the set of n firms, P_i is a non-empty subset of $\mathbb{R}_+$ for each $i \in N$ and $p_i \in P_i$ is the unit price of the product of firm i, and $\pi_i : P_1 \times \ldots \times P_n \longrightarrow \mathbb{R}$ is the profit function of firm i. A vector $p = (p_1, \ldots, p_n) = (p_i,\, p_{-i})$ where $p_{-i} = (p_1, \ldots, p_{i-1}, p_{i+1}, \ldots, p_n)$, called a *strategy profile*, is an element of $P := P_1 \times \ldots \times P_n = P_i \times P_{-i} \subseteq \mathbb{R}_+^n$ where $P_{-i} = P_1 \times \ldots \times P_{i-1} \times P_{i+1} \times \ldots \times P_n \subseteq \mathbb{R}_+^{n-1}$.

Definition 6.3: Nash equilibrium

A strategy profile $p^* = (p_1^*, \ldots, p_n^*)$ is a Nash equilibrium if it is feasible (i.e., is an element of P) and if

$$p_i^* = \arg\max_{p_i\in P_i} \pi_i(p_i,\, p_{-i}^*) \text{ for each } i \in N. \tag{NE}$$

The choice p_i of price by firm i is a *best response* for this firm to the prices set by its rivals p_{-i} if it maximizes firm i's profit. Formally,

Definition 6.4: Best-response correspondence

The best-response correspondence of firm i is a mapping $\psi_i :\, P_{-i} \to\to P_i$ if

$$\psi_i(p_{-i}) = \arg\max_{p_i\in P_i} \pi_i(p_i,\, p_{-i}).$$

Define by $\psi : P \to\to P$ the Cartesian product of the best-reply correspondences of the n firms, i.e., $\psi(p) := \psi_1(p_{-1}) \times \ldots \times \psi_n(p_{-n})$. Now assume that each π_i is strictly quasiconcave, and hence the best-response correspondence of firm i is a function. Notice that if a strategy profile p^* is a Nash equilibrium of the game Γ, then for each firm i we have that $p_i^* = \psi_i(p_{-i}^*)$. Thus, $p^* = \psi(p^*)$, i.e., p^* is a fixed point of the function ψ. Therefore, proving the existence of a Nash equilibrium is equivalent to proving the set of of fixed points of ψ is non-empty.

Theorem 6.13: Existence of Nash equilibrium

Let $\Gamma = \langle N,\, \{P_i\}_{i\in N},\, \{\pi_i(\cdot)\}_{i\in N}\rangle$ be a Bertrand game, and assume for each firm i that

(a) the strategy space P_i is a non-empty, compact and convex set in $\mathbb{R}_+$, and

(b) the profit function $\pi_i(\cdot)$ is continuous and strictly quasiconcave in p_i given p_{-i}.

Then the set of Nash equilibria of Γ is non-empty.

Proof. We will use Brouwer's fixed-point theorem to prove the existence of a fixed point. So we will prove that the function ψ is continuous, and that P is compact and convex. Since P is a Cartesian product of n compact and convex sets, it is compact and convex. Consider the best-response function of firm i. Since π_i is continuous and P_i is compact and convex, by Weierstrass theorem, ψ_i exists, and by Berge's theorem it is continuous.

The function ψ which is the Cartesian product of n continuous functions, is itself continuous. Therefore, by Brouwer's fixed point theorem its set of fixed points is non-empty. $\square$

6.3.2 Value function and the Envelope theorem

The value function $V : \Theta \longrightarrow \mathbb{R}$ for a maximization problem gives the maximum attainable value of the objective function for each value of the parameters:

$$V(\theta) = \max_x \{f(x;\, \theta) \mid x \in C(\theta)\} = f(x^*;\, \theta), \quad \text{where } x^* \in S(\theta).$$

The above expression has a nice and intuitive geometric interpretation. Consider the above maximization problem with a single parameter θ. Fix an x that is feasible, and plot the objective function f as the function of the parameter alone. The function $V(\theta)$ corresponds to the upper envelope of this family of curves. Two important observations emerge. First, the concavity or convexity of V will crucially depend on the nature of the objective function and the constraint set. Second, the value function is, in general, not differentiable. An important result, called the *envelope theorem* will give us sufficient conditions for the differentiability of the value function.

Theorem 6.14: Concavity of the value function

Consider the following maximization problem and the associated value function:

$$V(\theta) = \max_x \{f(x;\, \theta) \mid g(x;\, \theta) \geq 0\}.$$

Suppose the objective function f is concave in $(x,\, \theta)$ and that all the constraint functions g^i for $i = 1,\, \ldots,\, m$ are quasiconcave in $(x,\, \theta)$. Then the value function is concave.

Proof. Take two arbitrary parameter vectors θ' and θ'', and let $x' = x(\theta')$ and $x'' = x(\theta'')$ be the corresponding optimal choices of x. Consider now the pair $(x^\lambda,\, \theta^\lambda)$ defined by

$$x^\lambda = (1-\lambda)x' + \lambda x'' \quad \text{and} \quad \theta^\lambda = (1-\lambda)\theta' + \lambda\theta'', \quad \text{for } \lambda \in (0,\, 1),$$

and observe that, in principle, x^λ is not necessarily an optimal choice corresponding to θ^λ. We first show that x^λ is feasible for θ^λ. Since x' is feasible for θ' and x'' is feasible for θ'', it follows that, for each $j = 1,\, \ldots,\, m$, $g^j(x';\, \theta') \geq 0$ and $g^j(x'';\, \theta'') \geq 0$. Thus, quasiconcavity of g^j implies that

$$g^j(x^\lambda;\, \theta^\lambda) \geq \min\{g^j(x';\, \theta'),\, g^j(x'';\, \theta'')\} \geq 0,$$

for each $j = 1,\, \ldots,\, m$, and hence x^λ is feasible for θ^λ. Now,

$$\begin{aligned} V(\theta^\lambda) &= f(x(\theta^\lambda);\, \theta^\lambda) \\ &\geq f(x^\lambda;\, \theta^\lambda) \\ &\geq (1-\lambda)f(x';\, \theta') + \lambda f(x'';\, \theta'') \\ &= (1-\lambda)V(\theta') + \lambda V(\theta''). \end{aligned}$$

Thus, V is concave. $\square$

Theorem 6.15: Envelope theorem

Let $V(\theta) = \max_x\{f(x;\, \theta) \mid g(x;\, \theta) = 0\}$, where $f,\, g : \mathbb{R}^n \longrightarrow \mathbb{R}$ are $\mathcal{C}^2$ functions. If x^* be a regular solution, i.e., a strict local maximum of this problem for θ^0, then V is differentiable at θ^0, and

the derivative is given by

$$DV(\theta^0) = D_\theta \mathcal{L}(x^*, \theta^0) = D_\theta f(x^*, \theta^0) + \lambda^T D_\theta g(x^*, \theta^0).$$

Proof. Differentiating the Lagrange function

$$\mathcal{L}(x^*, \theta^0) = f(x^*, \theta^0) + \lambda^T g(x^*, \theta^0)$$

with respect to (x, λ), for $\theta = \theta^0$, we get the first-order conditions of the problem:

$$D_x \mathcal{L}(x, \theta^0) = D_x f(x, \theta^0) + \lambda^T D_x g(x, \theta^0) = 0, \tag{6.34}$$
$$D_\lambda \mathcal{L}(x, \theta^0) = g(x, \theta^0) = 0. \tag{6.35}$$

Under the assumptions of the theorem, the decision rule is well-defined and is a differentiable function. Thus,

$$V(\theta) = f(x(\theta); \theta).$$

Differentiating the value function with respect to θ and using (6.34), we obtain

$$\begin{aligned} DV(\theta^0) &= D_\theta f(x^*, \theta^0) + D_x f(x^*, \theta^0) Dx(\theta^0) \\ &= D_\theta f(x^*, \theta^0) - \lambda^T D_x g(x^*, \theta^0) Dx(\theta^0) \end{aligned} \tag{6.36}$$

Substituting $x(\theta)$ in (6.35) and differentiating with respect to θ

$$D_x g(x^*, \theta^0) Dx(\theta^0) + D_\theta g(x^*, \theta^0) = 0.$$

Using the above expression in (6.36) we get the desired result. □

We will use the Envelope theorem to give an intuitive interpretation of the Lagrange multiplier. Consider the Lagrange problem with the constraints $g^i(x) + \gamma_i = 0$ for $i \in M$. The value function is given by

$$V(\gamma) = \max_x \{f(x) \mid g(x) + \gamma = 0\}.$$

By Envelope theorem we have $DV(\gamma^0) = \lambda^0$. Thus, the multipliers measure the sensitivity of the value function to the changes in the constants of the constraint functions. Therefore, λ can be interpreted as *shadow prices*.

Example 6.6: Roy's identity

Consider the utility maximization problem defined as

$$V(p, m) = \max_x \{u(x) \mid m - p \cdot x = 0\}.$$

The Lagrangean for the consumer problem is given by

$$\mathcal{L}(x, \lambda; p, m) = u(x) + \lambda \left[m - \sum_{i=1}^{l} p_i x_i \right].$$

By envelope theorem

$$\begin{aligned} \frac{\partial V}{\partial p_i}(p, m) &= -\lambda^* x_i(p, m), \\ \frac{\partial V}{\partial m}(p, m) &= \lambda^*. \end{aligned}$$

Dividing the first equality by the second one we get

$$x_i(p,\, m) = -\frac{\partial V(p,\, m)/\partial p_i}{\partial V(p,\, m)/\partial m},$$

which is the Roy's identity.

6.3.3 Monotone comparative statics

Consider the maximization problem (P_1), and assume that the best response correspondence is given by $S(\theta) = \{x(\theta)\}$. We aim at finding sufficient conditions under which the optimal solution is monotone on Θ. First consider the following definition.

Definition 6.5: Supermodularity

Let $f : X \longrightarrow \mathbb{R}$ be a $\mathcal{C}^2$ function where $X \subseteq \mathbb{R}^n$. The function f is said to be (strictly) supermodular on X if the off-diagonal elements of its Hessian matrix are all (strictly) positive, i.e., $f_{ij}(x)(>) \geq 0$ for all $i,\, j = 1,\, \ldots,\, n$ and $i \neq j$.

The following theorem states a simpler version of "monotone comparative statics" result of Topkis (1978).

Theorem 6.16

Let $f : X \times \Theta \longrightarrow \mathbb{R}$ be a $\mathcal{C}^2$ function, where $X,\, \Theta \subset \mathbb{R}$. If f is (strictly) supermodular on $X \times \Theta$, then the function $x(\theta) = \text{argmax}_x\{f(x,\, \theta) \mid x \in C(\theta)\}$ where $C(\theta) \subset X$ is (strictly) increasing in θ.

We omit the proof. Instead, we consider the following examples to understand the intuition of the above theorem.

Example 6.7: Firm-worker assignment

Consider a continuum $[0,\, 1]$ of firms with a machine apiece, and a continuum $[0,\, 1]$ of workers who are heterogeneous with respect to productivity q. Let $F(s)$ denote the distribution of machine size and let $G(q)$ be the distribution of productivity. An worker with a given productivity level q produces a total output $y(s,\, q)$ if he uses a machine of size s. where $r(s)$ is the rental rate of a type s machine and $w(q)$ is the wage of a type q worker. Assume that workers are assigned to the machines according to the assignment rule $s = s(q)$ where $q \equiv s^{-1}$. A Walrasian equilibrium is a tuple $(s,\, w(q),\, r(s))$ such that

(i) (a) $q(s) = \text{argmax}_q\, \{y(s,\, q) - w(q)\}$ for each s;

(b) $r(s) = \max_q\, \{y(s,\, q) - w(q)\}$ for each s;

(c) $w(q) = \max_s\, \{y(s,\, q) - r(s)\}$ for each q;

(ii) If $s([q_1,\, q_2]) = [s_1,\, s_2]$ for any $[q_1,\, q_2],\, [s_1,\, s_2] \subseteq [0,\, 1]$, then $F(s_2) - F(s_1) = G(q_2) - G(q_1)$.

Suppose that the equilibrium assignment is increasing, i.e., $s'(q) > 0$. We will establish conditions under which such supposition is valid. The first order condition of the maximization problem of a type s firm implies that

$$w'(q) = \frac{\partial y(s,\, q)}{\partial q} \quad \text{for} \quad s = s(q). \tag{6.37}$$

The above equation implies that the marginal wage of a type q worker is equal to his marginal product.

Applying the envelope theorem to the above maximization problem one gets

$$r'(s) = \frac{\partial y(s,\, q)}{\partial s} \quad \text{for} \;\; s = s(q), \tag{6.38}$$

i.e., the marginal rent of a type s machine is its marginal product. In an equilibrium, the assignment of workers to machine must also satisfy the second order condition of the maximization problem:

$$\frac{\partial^2 y}{\partial q^2}(s(q),\, q) - w''(q) \leq 0. \tag{6.39}$$

Differentiating equation (6.37) one gets

$$w''(q) = \frac{\partial^2 y}{\partial s \partial q}(s(q),\, q)s'(q) + \frac{\partial^2 y}{\partial q^2}(s(q),\, q).$$

From the above it is clear that $s'(q) > 0$, i.e., the assignment is positively assortative if $y(s,\, q)$ is supermodular on $[0,\, 1] \times [0,\, 1]$. Now suppose that s and q are both distributed uniformly over the intervals $[\mu - \sigma_s,\, \mu + \sigma_s]$ and $[\mu - \sigma_q,\, \mu + \sigma_q]$, respectively. Then the second equilibrium condition implies that

$$\begin{aligned} &\int_{\mu-\sigma_s}^{s} \frac{s - (\mu - \sigma_s)}{(\mu + \sigma_s - \mu - \sigma_s)}\, ds = \int_{\mu-\sigma_q}^{q} \frac{q - (\mu - \sigma_q)}{(\mu + \sigma_q - \mu - \sigma_q)}\, dq \\ \implies\; & s = s(q) = \mu + \frac{\sigma_s}{\sigma_q}(q - \mu). \end{aligned}$$

That is the matching function is linear whose slope depends on the relative dispersion of the workers to machines, i.e., whether $\sigma_s \gtrless \sigma_q$. Notice that if $y(s,\, q) = sq$, then $w''(a) = \sigma_s/\sigma_q > 0$. Thus the equilibrium wage function is convex, i.e., the economy represents increasing wage inequality.

Example 6.8: Bertrand competition

Let there be two firms 1 and 2 in a Bertrand market. They face an identical $\mathcal{C}^2$ demand function $D(p_i,\, p_j)$ with $D_i(p_i,\, p_j) < 0$, $D_{ii}(p_i,\, p_j) \leq 0$ and $D_j(p_i,\, p_j) > 0$ for $i,\, j = 1,\, 2$ and $i \neq j$. Both firms have zero cost of production. The profit of firm i is given by

$$\Pi^i(p_i,\, p_j) = p_i D(p_i,\, p_j) \tag{6.40}$$

which she maximizes with respect to p_i. The Nash equilibrium is given by $(p_1^*,\, p_2^*)$ such that $p_1^* = \psi_1(p_2^*)$ and $p_2^* = \psi_2(p_1^*)$ where $\psi_i(p_j)$ is the best response of firm i. We will look for conditions under which the best response functions are increasing. Notice at a Nash equilibrium that

$$\Pi_i^i(p_i,\, p_j) = p_i D_i(p_i,\, p_j) + D(p_i,\, p_j) = 0.$$

Differentiating the above expression with respect to p_j at $p_i = \psi_i(p_j)$, we get

$$\begin{aligned} &\psi_i'(p_j) D_i(\psi_i(p_j),\, p_j) + \psi_i(p_j)[D_{ii}(\psi_i(p_j),\, p_j)\psi_i'(p_j) + D_{ij}(\psi_i(p_j),\, p_j)] \\ &+ D_i(\psi_i(p_j),\, p_j)\psi_i'(p_j) + D_j(\psi_i(p_j),\, p_j) = 0 \\ \implies\; &\psi_i'(p_j) = -\frac{\psi_i(p_j) D_{ij}(\psi_i(p_j),\, p_j) + D_j(\psi_i(p_j),\, p_j)}{2D_i(\psi_i(p_j),\, p_j) + \psi_i(p_j) D_{ii}(\psi_i(p_j),\, p_j)}. \end{aligned}$$

Given the assumption on $D(p_1,\, p_2)$, the best response functions are increasing if $D_{ij}(p_i,\, p_j) \geq 0$ for all $(p_i,\, p_j)$, i.e., the demand function of any firm is supermodular in its own price and that of its rival.

Bibliography

Sundaram, Rangarajan K. (1996), *A First Course in Optimization Theory*. Cambridge University Press.

Topkis, Donald (1978), "Minimizing a Submodular Function on a Lattice." *Operations Research*, 26, 305–321.

Vohra, Rakesh (2005), *Advanced Mathematical Economics*. Routledge.