Nonlinear Programming in Finite-Dimensional (Euclidean) Spaces

University of Pennsylvania
Department of Economics

Math Camp 2005: Part III[1]

August 2005

[1]These notes are a modified version of the original Professor Cass' course notes, which were prepared by Jorge Gallardo-García, Ben Lester, Claire Lim and Rafael Silveira. This modified version was prepared by Viviana Velez-Grajales.

Chapter 1

Introduction

One of the primary goals of these notes is to give the foundations necessary to study equilibrium concepts. Roughly speaking, when we look at an economy made up of a variety of agents, an equilibrium is an allocation that is reached when all agents are maximizing their respective utility subject to a set of constraints (provided the allocation is feasible). Such a maximization problem is referred to as a "programming problem" (note that this has nothing to do with computer programming!). Therefore, in order to address equilibrium concepts, we must begin by fully understanding both the properties of a programming problem, and the methods we can use to find optimal solutions.

The following notes begin by setting up and introducing the primitives of the general programming problem (Chapter 2). In Chapters 3 and 4, we define several mathematical concepts that allow us to conclude whether there definitely exists a solution to a given programming problem, and whether or not we can be sure that it will be the unique solution.

In Chapter 5 of the notes, we derive conditions that allow us to characterize an optimal solution to a programming problem. That is, we introduce a set of conditions that allows us to say, under certain specifications on the objective and constraint functions, that:

1. any optimal solution must satisfy these conditions (the conditions are therefore "necessary conditions").

2. if a solution satisfies these conditions, it must be an optimal solution (the "sufficient conditions").

For the general case in which we know that the objective and constraint functions are differentiable, we are able to use the Kuhn-Tucker (K-T) conditions. The specific properties required of the objective and constraint functions in order for the K-T conditions to be necessary and/or sufficient are given by the Kuhn-Tucker Necessary and Sufficient Theorems. For the specific case in which we know the objective and constraint functions are concave (but not necessarily differentiable), we can use the Saddle-Point property; again, the Saddle Point Necessary and Sufficient Theorems detail the properties required for the S-P conditions to be necessary and sufficient.

After looking at the existence, uniqueness, and characterization of an optimal solution, we may also want to know how the optimal solution may change as a result of a change in the parameters. This is called *Sensitivity Analysis*. For example, in the household's utility maximization problem, the prices are parameters (i.e. they are taken as given by the household). We might want to know how the optimal consumption bundle changes when the prices change. In Chapter 6, we introduce the mathematical tools that allow us to perform sensitivity analysis on a programming problem. The main tool that we employ in sensitivity analysis is the Implicit Function Theorem, which provides a functional form for differentiating our variables with respect to our parameters.

Finally, after all of this math is in place, we are ready to start some Economics! In the final chapter of the notes, we lay out the leading examples of programming problems for the household and the firm. All of this together lays the groundwork for studying General Equilibrium Theory.

Chapter 2

Setup (Maintained Assumptions)

2.1 Notation

This course deals both with scalars and vectors sometimes mixing the two in the same equations. To avoid confusion between them it is necessary to establish some conventions. Let x, $y \in \mathbb{R}^n$ with characteristic elements x^i and $y^i \quad i = 1, ..., n$ respectively. Then we maintain the following conventions:
$x \geqq y$ stands for $x^i \geq y^i \quad \forall i$;
$x > y$ stands for $x^i \geq y^i \quad \forall i$ and $x^i > y^i$ for some i;
$x \gg y$ stands for $x^i > y^i \quad \forall i$.

2.2 Primitives

Let x be a vector of real numbers with n elements which belongs to the set X, i.e., $x \in X \subset \mathbb{R}^n$ (usually this notation is used to represent variables in the programming problems). Now consider a vector of real numbers α with l elements that belongs to the set A, i.e., $\alpha \in A \subset \mathbb{R}^l$ (usually, this is the way to represent parameters in the programming problems).
Consider the function f to be a mapping from the Cartesian product of the sets X and A to the real line: $f : X \times A \to \mathbb{R}$ (this function represents the objective function in the programming problems ahead). And consider the function g as a mapping from the Cartesian product of the sets X and A to the set of real numbers of dimension m, i.e., $g : X \times A \to \mathbb{R}^m$ which represents the m constraints in the programming problems.

2.3 Canonical Programming Problem

Given $\alpha \in A$, find the vector $x \in X$ that solves the programming problem

$$\begin{array}{cc} \max & f(x,\alpha) \\ \text{subject to} & g(x,\alpha) \geqq 0 \quad (=0) \\ \text{and} & x \in X. \end{array}$$

Alternatively it is possible to use the constraints to define the "Constraint Set" as follows: $\mathcal{C}(\alpha) = \{x \in X : g(x,\alpha) \geqq 0 \quad (=0)\}$.
x is a feasible solution if we have $x \in \mathcal{C}(\alpha)$; x^* is a *local optimal solution* if it is a feasible solution and there is $\varepsilon > 0$ such that

$$x \in C(\alpha) \cap \{x \in \mathbb{R}^n : \| x - x^* \| < \varepsilon\} \Rightarrow f(x^*,\alpha) \geqq f(x,\alpha);$$

x^* is a (*global*) *optimal solution* if it is a feasible solution and

$$x \in C(\alpha) \Rightarrow f(x^*,\alpha) \geqq f(x,\alpha).$$

The focus in almost all economic problems is on optimal solutions rather then local optimal solutions. (Of course, an optimal solution is also a local optimal solution). Define the value function as the function that varies over the parameters (α) and represents the objective function evaluated at the optimal solution for a given α, $V(\alpha) = \max_{x \in \mathcal{C}(\alpha)} f(x,\alpha)$.
As the leading examples of a programming problem consider the firm's and the household's problems.

2.3.1 Firm's Problem

In this example there is a firm that maximizes its profits (objective function) subject to the technology available and to the possible quantities of inputs and outputs that it can deal with (constraint set). Let $p \in \mathbb{R}^G_+/\{0\}$ be the price vector for the G goods in the economy (parameter); $Y \subset \mathbb{R}^G$ the production set; $y \in Y$ the inputs/outputs vector (variable); and $t : \mathbb{R}^G \to \mathbb{R}$ the function that represents the firm's technology. Then the problem the firm solves is:

$$\begin{array}{cc} \max & py \\ \text{subject to} & t(y) \geqq 0 \\ \text{and} & y \in Y. \end{array}$$

2.3.2 Household's Problem

Consider a household which maximizes its utility (objective function) subject to its budget constraint and to the possible quantities of consumption goods that it can deal with (constraint set). Let $p \in \mathbb{R}^G_+/\{0\}$ be the price vector for the G goods in the economy (parameter); $w \in \mathbb{R}_+$ the household's disposable income or wealth (parameter); $X \subset \mathbb{R}^G_+$ the consumption set; $x \in X$ the consumption vector (variable); and $u : X \to \mathbb{R}$ the utility function. Then the problem this household solves is:

$$\begin{array}{cc} \max & u(x) \\ \text{subject to} & w - px \geqq 0 \\ \text{and} & x \in X. \end{array}$$

2.4 Relevant Issues in Programming

There are several issues which are of special interest concerning the programming problem and its solutions. Throughout these notes you will notice that different assumptions are needed for each of these issues, and that it is extremely important to identify and manage them in order to fully understand the concepts.

1. Existence. The basic thing to consider in a programming problem is whether it has a solution or not before proceeding with later analysis. Notice that the quantity of solutions is not considered here, only the fact that there exists at least one of them.

2. Uniqueness. Once there is existence of a solution of the problem, one would like to know if it has one or many solutions. In the applications that are considered here the most desirable result is to have only one solution, i.e., uniqueness.

3. Characterization. Provided there is a solution for the programming problem, one would like to identify it in order to analyze it. Under certain conditions, it is possible to characterize these solutions, using certain theorems.

4. Regularity (continuous dependence). The former issues assume that the parameters are fixed. After identifying the problem's solution for a given vector of parameters α it is possible to analyze the behavior of these solutions as the parameters vary. Regularity consists in providing conditions under which the set of solutions depends continuously in the parameters.

5. Sensitivity (smooth dependence). Here the analysis is concerned about the dependence of the solutions on the parameters but in a smooth way. Sensitivity analysis studies the effect that a small change in the parameters can have on the optimal solution. This is useful when analyzing how the demand functions change when there is a change in the price vector, for example.

Chapter 3

Existence of an Optimal Solution

3.1 Introduction

This chapter will discuss results regarding the existence of a solution to optimization problems. Basically, we are interested in answering the following question: under what conditions on the objective function f and the constraint set $\mathcal{C}$ is it *guaranteed* that solutions will exist to problems of the form *max* $\{f(x)|x \in \mathcal{C}\}$ or *min* $\{f(x)|x \in \mathcal{C}\}$?

Trivial answers to this question are available: for example, f is guaranteed to attain a maximum and a minimum on $\mathcal{C}$ if it is a finite set. However, we want to be as general as possible considering that different problems may have different ways to show existence. Let's introduce some basic definitions.

3.2 Mathematical Digression

Definition (Compact set). *There are three equivalent definitions of compact sets in finite-dimensional Euclidean spaces. Let K be a finite-dimensional Euclidean space.*

1. ***Topological definition:*** *A collection $\mathcal{G}$ of open sets in K is an open cover of a set C if C is contained in the union of the sets in $\mathcal{G}$. A set $C \subset K$ is said to be compact if every open cover $\mathcal{G}$ of C has a finite subcover, that is, if for every open cover $\mathcal{G}$ of C there is a finite collection $O_1, O_2, \ldots, O_N \subset \mathcal{G}$ such that $C \subset \bigcup_{i=1}^{N} O_i$.*

2. ***Sequential definition:*** *A set $C \subset K$ is said to be sequentially compact if every sequence $\{x_n\}$ in C contains a convergent subsequence $\{x_{n_k}\}$ in C, i.e. $\lim_{n_k \to \infty} x_{n_k} \in C$. In metric spaces (in particular, Euclidean spaces), a set C is compact if and only if it is sequentially compact.*

3. ***Euclidean definition:*** *A set $C \subset K$ is compact if and only if C is closed and bounded.*

Definition (Continuous function). *Let f be a real-valued function whose domain is a set $E \subset \mathbb{R}^n$. We say that f is continuous at point x in E if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that $|f(t) - f(x)| < \varepsilon$ for all points t of E for which $\|t - x\| < \delta$. If f is continuous at every point of E, then f is said to be continuous on E.*

It should be noted that f has to be defined at the point x in order to be continuous at x. Also, if x is an isolated point in E, then our definition implies (trivially) that every function f which has E as its domain is continuous at x. From the above definitions we can extract a relevant result.

3.3 Existence Theorem

Theorem (Extreme Value Theorem (EVT)). *Let f be a continuous real-valued function on a non-empty compact set C. Then f achieves its maximum and minimum. In other words, $\exists x' \in C$ such that $f(x') \geq f(x)$ $\forall x \in C$ and $\exists x'' \in X$ such that $f(x) \geq f(x'')$ $\forall x \in C$.*
Proof: *Let $M = sup\ \{f(x) : x \in C\}$. Then there is a sequence $\{x_n\}$ of points of C such that $M = lim\ f(x_n)$. Since C is (sequentially) compact, there is a subsequence $\{x_{n_k}\}$ that converges to a point $z \in C$. Then $f(z) = lim\ f(x_{n_k}) = M$. Thus M is a real number (i.e., not ∞) and $f(z) = M$. Hence M is the maximum of f and is achieved at z. The other half of the proof follows by replacing f by $-f$.* □

It is worth noticing that the EVT provides only *sufficient* conditions for existence of extreme points. It does not tell us how to find them. Moreover, none of its conditions is necessary for f to have maximum or minimum points.

In addition, the theorem has nothing to say about what happens if these conditions are not met. Indeed, nothing can be said, in general, as the following examples illustrate. In each of the first three examples, only a single condition of the *Extreme Value Theorem* is violated, yet maxima and minima fail to exist. In the last example, all of the conditions of the theorem are violated, but both maxima and minima exist.

Example 1: Let $\mathcal{C} = \mathbb{R}$, and $f(x) = x$ for all $x \in \mathbb{R}$. Then f is continuous, but $\mathcal{C}$ is not compact (it is closed but not bounded). Since $f(\mathcal{C}) = \mathbb{R}$, f evidently attains neither a maximum nor a minimum on $\mathcal{C}$.

Example 2: Let $\mathcal{C} = (0, 1)$ and $f(x) = x$ for all $x \in (0, 1)$. Then f is continuous, but $\mathcal{C}$ is again noncompact (this time it is bounded, but not closed). The set $f(\mathcal{C})$ is the open interval $(0, 1)$, so, again, f attains neither a maximum nor a minimum on $\mathcal{C}$.

Example 3: Let $\mathcal{C} = [-1, 1]$, and let f be given by

$$f(x) = \begin{cases} 0 & \text{if} \quad x = -1 \quad \text{or} \quad x = 1 \\ x & \text{if} \quad -1 < x < 1. \end{cases} \tag{3.1}$$

Note that $\mathcal{C}$ is compact, but f fails to be continuous at just two points -1 and 1. In this case, $f(\mathcal{C})$ is the open interval $(-1, 1)$; consequently, f fails to attain either a maximum or a minimum on $\mathcal{C}$.

Example 4: Let $\mathcal{C} = \mathbb{R}_{++}$, and let $f : \mathcal{C} \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ 0 & \text{otherwise.} \end{cases} \tag{3.2}$$

Then $\mathcal{C}$ is not compact (it is neither closed nor bounded), and f is discontinuous at every single point in R. Nonetheless, f attains a maximum (at every rational number) and a minimum (at every irrational number).

Chapter 4

Uniqueness of an Optimal Solution

4.1 Introduction

In addition to establishing whether a solution exists to our programming problem, we would also like to be able to say something about whether that solution will definitely be unique. This requires several definitions classifying the shape of the objective function.

4.2 Mathematical Digression

For the following definitions, let $X \subset \mathbb{R}^n$ be convex and open, $f : X \to \mathbb{R}$ be continuous, and let $x', x'' \in X$.

4.2.1 Convex Sets

Given any finite collection of points $x_1, ..., x_m \varepsilon \mathbb{R}^n$, a point $z \varepsilon \mathbb{R}^n$ is said to be a *convex combination* of the points $(x_1, ..., x_m)$ if there exists $\theta \varepsilon \mathbb{R}^m$ satisfying (i)$\theta_i \leq 0, i = 1, ..., m$, and (ii)$\sum_{i=1}^m \theta_i = 1$, such that $z = \sum_{i=1}^m \theta_i x_i$.
A set X is *convex* if the convex combination of any two points in X is also in S. Intuitively, X is convex if the straight line joining any two points in X is itself completely contained in X. (i.e. $\theta x' + (1 - \theta)x'' \in X, \forall x', x'' \in X$ and $\forall \theta \in [0, 1]$). X is strictly convex if $\forall x', x'' \in X$, with $x' \neq x''$ and $\forall \theta \in (0, 1)$, $\theta x' + (1 - \theta)x'' \in intX$. Building on the definition of convex sets, we introduce the concepts of concave and quasi-concave functions.

4.2.2 Concavity and Strict Concavity

Definition (Concave Function). *f is* concave *if*, $\forall \theta \in [0,1]$ *and* $\forall x', x'' \in X$,

$$f[\theta x' + (1-\theta)x''] \geq \theta f(x') + (1-\theta) f(x'').$$

Graph and Intuition:

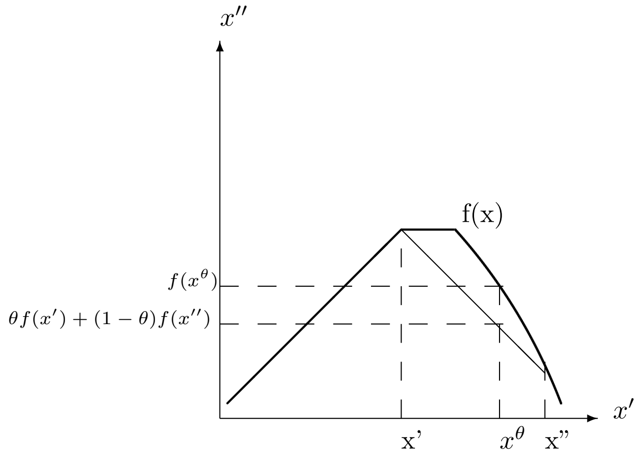
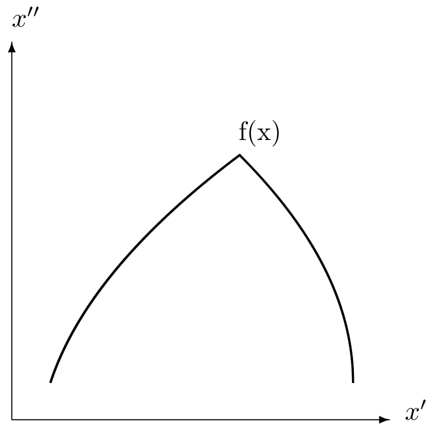


FIGURE I - CONCAVE

Notice that a continuous concave function can have linear portions, flats, and kinks. For a continuous function, concavity simply requires that the function evaluated at $x^\theta = \theta x' + (1-\theta)x''$, and we write $f(x^\theta)$ for simplicity, is greater than or equal to the weighted value of $\theta f(x') + (1-\theta) f(x'')$. Also, notice that when X is open and convex, if f is concave then it is also continuous.

Definition (Strictly Concave Function). *f is* strictly concave *if*, $\forall \theta \in (0,1)$ *and* $\forall x', x'' \in X$ *with* $x' \neq x''$,

$$f[\theta x' + (1-\theta)x''] > \theta f(x') + (1-\theta) f(x'').$$



Graph and Intuition: **FIGURE II - STRICTLY CONCAVE**

A strictly concave continuous function can have kinks, but it can have neither flats nor linear sections. A strictly concave function is also a concave function while the converse is not true.

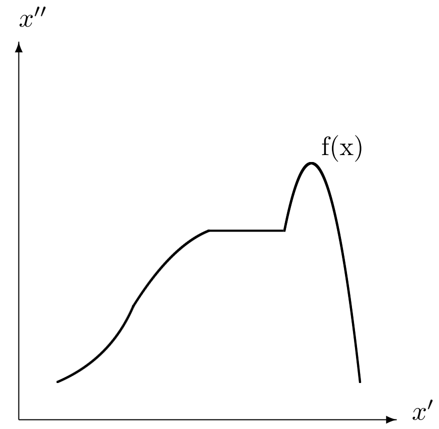
Convex and strictly convex functions can be defined accordingly by reversing the inequalities. One important property of concave (convex) functions is that concavity (convexity) is preserved under non-negative summation, i.e., for concave functions $f_1, ..., f_m$ defined on a convex set $X \in \mathbb{R}^n$, $f(x) = \sum_{i=1}^{m} p_i f_i(x)$, with $p_i \geq 0 \forall i$ is a concave function.

In consumer theory, preferences over consumption bundles are represented by ordinal, not cardinal, preference relations. So if a given utility function $u(.)$ represents a preference relation, then so does a positive monotone transformation of $u(.)$. One may wonder, therefore, if concavity is preserved under positive monotone transformation. The answer, unfortunately, is no.

4.2.3 Quasi-Concavity and Strict Quasi-Concavity

Definition (Quasi-Concave Function). *f is* quasi-concave *if, $\forall \theta \in [0,1]$ and $\forall x', x'' \in X$,*

$$f[\theta x' + (1-\theta)x''] \geq \min\{f(x'), f(x'')\}.$$



Graph and Intuition: **FIGURE III - QUASI-CONCAVE**

A continuous quasi-concave function can have flats, linear portions, and even convex portions; what is not possible, however, is for an increasing graph to decrease and then increase again (i.e., the graph cannot have multiple "peaks").

An important difference between concave and quasi-concave functions is that quasi-concavity is preserved under positive monotone transformation. Another difference is that quasi-concavity is not preserved under nonnegative summation, while concavity is.

Definition (Strictly Quasi-Concave Function). *f is* strictly quasi-concave *if, $\forall \theta \in (0,1)$ and $\forall x', x'' \in X$ with $x' \neq x''$,*

$$f[\theta x' + (1-\theta)x''] > \min\{f(x'), f(x'')\}.$$

Graph and Intuition:

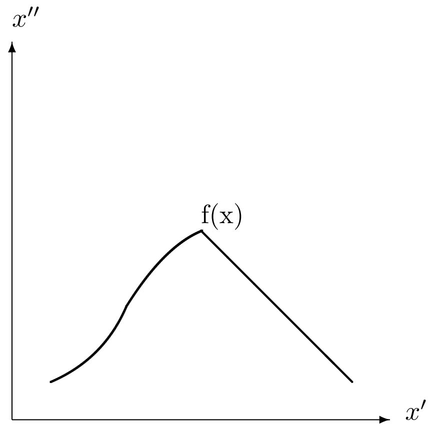


FIGURE IV - STRICTLY QUASI-CONCAVE

Strictly quasi-concave rules out flats, but can still contain linear and convex portions. Note that a strictly quasi-concave function is also quasi-concave while the converse is not true.

4.2.4 Graphs of Concave and Quasi-Concave Functions

Definition (Graph). *The* graph *of a function* $f : X \to \mathbb{R}$ *is defined as the set* $G_f = \{(x, y) : x \in X, y \in \mathbb{R}, f(x) = y\} \subset \mathbb{R}^{n+1}$.

Definition (Epigraph). *The* epigraph *of a function* $f : X \to \mathbb{R}$ *is defined as the set* $E_f = \{(x, y) : x \in X, y \in \mathbb{R}, f(x) \leq y\} \subset \mathbb{R}^{n+1}$.

Definition (Hypograph). *The* hypograph *of a function* $f : X \to \mathbb{R}$ *is defined as the set* $H_f = \{(x, y) : x \in X, y \in \mathbb{R}, f(x) \geq y\} \subset \mathbb{R}^{n+1}$.

Theorem. $f : X \to \mathbb{R}$ *is concave on* X *if and only if its hypograph* H_f *is a convex set in* $\mathbb{R}^{n+1}$.

Theorem. $f : X \to \mathbb{R}$ *is convex on* X *if and only if its epigraph* E_f *is a convex set in* $\mathbb{R}^{n+1}$.

A pair of concepts closely related to the epi- and hypographs are the upper and lower contour sets.

Definition (Upper Contour Set). *Let* $f : X \to \mathbb{R}$ *be an arbitrary function. For any* $\alpha \in \mathbb{R}$ *the set* $U_\alpha = \{x \in X | f(x) \geq \alpha\} \subset \mathbb{R}^n$ *is said to be an* upper contour set of f *at* α.

Definition (Lower Contour Set). *Let* $f : X \to \mathbb{R}$ *be an arbitrary function. For any* $\alpha \in \mathbb{R}$ *the set* $L_\alpha = \{x \in X | f(x) \leq \alpha\} \subset \mathbb{R}^n$ *is said to be a* lower contour set of f *at* α.

A necessary but not sufficient condition for f to be concave is that its upper contour set U_α is a convex set for every $\alpha \in \mathbb{R}$, while the upper contour set of a function completely describes whether or not it is quasi-concave.

Theorem. *$f : X \to \mathbb{R}$ is quasi-concave on X if and only if its upper contour set U_α is a convex set for all $\alpha \in \mathbb{R}$.*

4.2.5 Relationship between Concavity and Quasi-Concavity

Theorem. *If f is concave, then f is quasi-concave.*
Proof: *Suppose f is concave. Then $\forall\theta \in [0,1]$, we know $f[\theta x' + (1-\theta)x''] \geq \theta f(x') + (1-\theta)f(x'')$. Suppose, w.l.o.g., that $f(x') \geq f(x'')$. Then:*

$$f[\theta x' + (1-\theta)x''] \geq \theta f(x') + (1-\theta)f(x'') \geq$$
$$\theta f(x'') + (1-\theta)f(x'') = f(x'') = \min\{f(x'), f(x'')\}.$$

Therefore, f is quasi-concave. □

A similar proof can be used to show that if f is strictly concave, then it is also strictly quasi-concave. It is not true that if f is quasi-concave, then it is concave.

4.3 Uniqueness Theorem

Theorem (Uniqueness). *Let $\mathcal{C}$ be the constraint set. If $\mathcal{C}$ is convex and f is strictly quasi-concave, then there is at most one optimal solution.*
Proof: *Suppose not. That is, suppose $\exists\ x', x'' \in \mathcal{C}$ such that $f(x') = f(x'') \geq f(x)\ \forall x \in \mathcal{C}$, and $x' \neq x''$. Since $\mathcal{C}$ is convex, $\theta x' + (1-\theta)x'' \equiv x^\theta \in \mathcal{C}\ \forall\theta \in [0,1]$. Since f is strictly quasi-concave, $f(x^\theta) > min\{f(x'), f(x'')\} = f(x') = f(x'')$. But this contradicts our original assumption that $f(x') = f(x'') \geq f(x)\ \forall x \in \mathcal{C}$.* □

There is also a second set of conditions that guarantees uniqueness:

Theorem (Alternative Conditions Guaranteeing Uniqueness). *If $\mathcal{C}$ is strictly convex and f is quasi-concave and without local maxima, then there is at most one optimal solution.*
Proof: *Suppose not. That is, suppose $\exists\ x', x'' \in \mathcal{C}$ such that $f(x') = f(x'') \geq f(x)\ \forall x \in \mathcal{C}$, and $x' \neq x''$.*
Since $\mathcal{C}$ is strictly convex, $\theta x' + (1-\theta)x'' \equiv x^\theta \in int\ \mathcal{C}\ \forall\theta \in (0,1)$. Since f is quasi-concave, $f(x^\theta) \geq min\{f(x'), f(x'')\} = f(x') = f(x'')$. So we have $x^\theta \in int\ \mathcal{C}$ and $f(x^\theta) \geq f(x') \geq f(x)\ \forall x \in \mathcal{C}$. But then, for some $\varepsilon > 0$, there exists an ε-neighborhood of x^θ s.t. $f(x^\theta) \geq f(x)\ \forall x \in N_\varepsilon(x^\theta)$, i.e. x^θ is a local maximum of f. But this contradicts our original assumption that f is without local maxima. □

Chapter 5

Characterization of an Optimal Solution

5.1 Introduction

Since economists typically want to describe and analyze the solutions to their models, we want to develop some methods for actually solving maximization problems and characterizing these solutions. The standard approach is simply to assume that the objective function is differentiable and that the constraint set is defined by a collection of differentiable functions. With these assumptions we can use the powerful calculus tools of constrained optimization.

When we use calculus techniques to solve optimization problems, three types of conditions are often derived. First, the *necessary conditions* for the optima. These are the conditions that all optimizers must meet; but if a point x satisfies these conditions it is not guaranteed to be an optimizer. Second, the *sufficient conditions* for the optima. If a point x satisfies those conditions it is guaranteed that it is an optimizer. There might, however, be other optimizers that do not satisfy these sufficient conditions. Third, the *necessary and sufficient conditions.* If we have the necessary and sufficient conditions for a certain optimization problem, then we say that we have a *characterization* of the solution to the problem. Those points, and *only* those points that satisfy the necessary and sufficient conditions are optimizers of the problem.

Obviously one would like to have a set of necessary and sufficient conditions for every class of optimization problem. Unfortunately, whether or not we can generate a characterization of the solution depends on the properties of the objective function and the constraint functions defining the constraint set. In this section, we derive the Kuhn-Tucker and Saddle-Point Conditions, and present the theorems that state when these conditions are necessary for an optimal solution, and when they are sufficient. However, before we introduce the Kuhn-Tucker and Saddle-Point Necessity and Sufficiency Theorems, we must introduce some mathematical definitions and theorems that are going to be useful in the proofs.

5.2 Mathematical Digression: Differentiation

For the following theorems and definitions, let $X \subset \mathbb{R}^n$ be a convex and open set.

Definition (Partial Derivative). *Let $f : X \to \mathbb{R}$. Let e_j denote the vector in $\mathbb{R}^n$ that has a 1 in the j-th place and zeros elsewhere (j=1,.....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*

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

The *gradient* of f at x is defined by

$$Df(x)=\left[\frac{\partial f(x)}{\partial x_1},...,\frac{\partial f(x)}{\partial x_n}\right].$$

If f is differentiable at x, then all partials exist at x but the existence of all partial derivatives at a point x is not enough to ensure that f is differentiable at x.
Partial derivatives are a special case of directional derivatives. The *directional derivative* of f at x in the direction $h \in \mathbb{R}^n$ with $\|h\| = 1$ is defined as

$$lim_{t\to 0}\left(\frac{f(x+th)-f(x)}{t}\right).$$

when this limit exists, and is denoted $Df(x; h)$ or $D_h f(x)$. If f is differentiable at any x then the directional derivative of f at x in any direction h exists, and, in fact, we have $Df(x; h) = Df(x) \cdot h$.
We are assuming that f is differentiable, so that the derivative $Df = \left[\frac{\partial f}{\partial x_1},...,\frac{\partial f}{\partial x_n}\right]$ itself defines a function from X to $\mathbb{R}^n$. If this function is continuous, then f is said to be *continuously differentiable* on X and we write f is C^1. If Df is itself differentiable, i.e., for each i, the function $\frac{\partial f}{\partial x_i} : X \to \mathbb{R}$ is differentiable, then we say that f is *twice-differentiable* with second derivative $D^2 f$, where

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

is the second derivative of f at x.
When for each $i, j = 1, ..., n$, the cross partial $\partial^2 f/\partial x_i \partial x_j$ is a continuous function from X to $\mathbb{R}$, we say that f is *twice continuously differentiable* on X, and

we write f is C^2.
A very useful property of a C^2 function is that its second derivative $D^2 f$, also called the *Hessian*, is a symmetric matrix, i.e., $\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x)$. As we will see later, this property will help us to characterize the concavity of the function by determining the definiteness of the Hessian matrix.

Definition. *An $n \times n$ symmetric matrix A is said to be:*

1. ***positive definite*** *if* $\mathbf{x}^T A\mathbf{x} > 0$ *for all* $\mathbf{x} \in R^n, \mathbf{x} \neq \mathbf{0}$.
2. ***positive semi-definite*** *if* $\mathbf{x}^T A\mathbf{x} \geq 0$ *for all* $\mathbf{x} \in R^n$..
3. ***negative definite*** *if* $\mathbf{x}^T A\mathbf{x} < 0$ *for all* $\mathbf{x} \in R^n, \mathbf{x} \neq \mathbf{0}$.
4. ***negative semi-definite*** *if* $x^T Ax \leq 0$ *for all* $x \in R^n$.
5. ***indefinite*** *if* $\mathbf{x}^T A\mathbf{x} < 0$ *for some* $\mathbf{x} \in R^n$ *and* $\mathbf{x}^T A\mathbf{x} > 0$ *some* $\mathbf{x} \in R^n$.

For useful theorems regarding the definiteness of a symmetric matrix, see the Appendix.

5.3 Unconstrained Optimization

In this section we address the case where the feasible set is given by $X = R^n$ (or any open set). Consider the unconstrained program

$$\max_{x \in X} f(x). \tag{\textbf{U}}$$

5.3.1 First Order Conditions

Theorem. ***(First Order Necessary Condition)*** *If x^* is a local maximizer of a differentiable function $f : X \to \mathbb{R}$, then*

$$Df(x^*) = 0$$

We will offer here some intuitive arguments for why this should be a necessary condition. The basic idea is that if $Df(x^*) \neq 0$, then we could find some direction along which to move x away from x^* such that $f(x) > f(x^*)$; but then x^* cannot be a local maximizer. To see this, $Df(x^*) \neq 0$ implies that at least one partial derivative, say, $\frac{\partial f(x^*)}{\partial x_i} \neq 0$. Now consider a sequence of points $x^k = \left(x_1^k, ..., x_i^k, ..., x_n^k\right)$ with $x_j^k = x_j^*$ if $j \neq i$ and $x_i^k \to x_i^*$. Now by the definition of partial derivative,

$$\frac{\partial f(x^*)}{\partial x_i} = \lim_{x_i^k \to x_i^*} \frac{f\left(x^k\right) - f(x^*)}{x_i^k - x_i^*}.$$

Apparently if $f\left(x^k\right) \leq f\left(x^*\right)$ for every k, then the limit on the right hand side will be non-positive if x_i^k approaches x_i^* from above; and the limit will be non-negative if x_i^k approaches x_i^* from below. But if $\frac{\partial f(x^*)}{\partial x_i} > 0$, the limit must be positive no matter whether x_i^k approaches x_i^* from above or below. Therefore if $\frac{\partial f(x^*)}{\partial x_i} > 0$, it can not be that $f\left(x^k\right) \leq f\left(x^*\right)$ for every k.

5.3.2 Second Order Conditions

It is easy to see that the first order condition $Df\left(x^*\right) = 0$ does not determine whether local maximizer or minimizer (or some other stationary point for that matter). The condition simply provides a set of potential candidates that might be a local maximizers. One can eliminate some of these candidates, however, by looking at the second derivative.

Theorem. ***(Second Order Necessary Condition for Local Maximum)*** *If $f : X \to \mathbb{R}$ is C^2, and if x^* is a local maximizer of f, then $D^2 f\left(x^*\right)$ is negative semi-definite.*

Proof. Since $f \in C^2$, $x^* \in X$ and X is an open set, we can use Taylor's Theorem to get that for h such that $x + h \in X$,

$$f\left(x^* + h\right) = f\left(x^*\right) + Df\left(x^*\right)h + \frac{1}{2}h' D^2 f\left(x^*\right)h + R\left(h\right)$$

where $h \in R^n$ is a column vector and $\frac{R(h)}{\|h\|^2} \to 0$ as $\|h\| \to 0$. By the previous theorem we know that $Df\left(x^*\right) = 0$ if x^* is a local maximizer. Thus we have:

$$f\left(x^* + h\right) - f\left(x^*\right) = \frac{1}{2}h' D^2 f\left(x^*\right)h + R\left(h\right).$$

Now let $h = t\tilde{h}$, for $\tilde{h} \epsilon \mathbb{R}^n$ and $t \epsilon \mathbb{R}$. We have

$$\frac{f\left(x^* + t\tilde{h}\right) - f\left(x^*\right)}{\left\|t\tilde{h}\right\|^2} = \frac{1}{2}\frac{\tilde{h}' D^2 f\left(x^*\right)\tilde{h}}{\left\|\tilde{h}\right\|^2} + \frac{R\left(t\tilde{h}\right)}{\left\|t\tilde{h}\right\|^2}.$$

x^* is a local maximizer then, for sufficiently small values of t, the left hand side is ≤ 0.

As $t \to 0$, $\frac{R\left(t\tilde{h}\right)}{\left\|t\tilde{h}\right\|^2} \to 0$ by Taylor's theorem, and therefore $D^2 f\left(x^*\right)$ is negative semi-definite. □

The analogous necessary condition for a local minimizer is that the Hessian evaluated at x^* is positive semi-definite.

There are sufficient conditions for optimality using properties like concavity, pseudo-concavity, and quasi-concavity. These characteristics are defined globally, over the entire domain of a function. Unfortunately, many economically relevant functions do not have these global properties and thus we must focus on the curvature of the objective function in a neighborhood of the proposed optimum. To do this we look at the second order derivatives of our functions at this point and seek conditions that are sufficient for local optima. A sufficient condition for x^* to be a local maximum is the following:

Theorem. ***(Second Order Sufficient Condition for a Local Maximum)*** *Suppose x^* is a critical point (i.e., $Df(x^*) = 0$) of a C^2 function $f : X \to \mathbb{R}$. If $D^2 f(x^*)$ is* negative definite*, then x^* is a local maximizer of f.*

Unconstrained Optimization under Concavity

As economists we are almost always interested in global rather than local optima. The value of necessary conditions for local optima lies in their ability to reduce the number of potential (global) solutions. A calculus characterization of the global maximizers will require more assumptions on f beyond just differentiability. Typically these additional assumptions are that f is concave or quasi-concave.

The attractiveness of adding curvature properties for optimization theory arises from the fact that when an optimization problem meets suitable curvature conditions, the same first-order conditions that are necessary for *local optima*, also become sufficient for *global optima*.

Theorem. *Let $f : X \to \mathbb{R}$ be a concave and differentiable function. Then, x^* is an optimal solution to (U) if and only if $Df(x^*) = 0$.*

5.4 Mathematical Digression

For the following theorems and definitions, let $X \subset \mathbb{R}^n$ be a convex and open set, $f : X \to \mathbb{R}$ be differentiable, and let $x', x'' \in X$.

5.4.1 Concavity, Quasi-Concavity and Pseudo-Concavity for Differentiable Functions

Theorem (Concave Function). *The function f is* concave *if and only if, $\forall x', x'' \in X$,*

$$Df(x')(x'' - x') \geq f(x'') - f(x').$$

Intuition: Rewrite the above inequality as $Df(x')(x'' - x') + f(x') \geq f(x'')$. The left side is the equation of a line tangent to the graph at the point x'. The inequality suggests that f is a concave function if and only if, for any $x' \in X$, the tangent line at x' remains on or above the graph of the function.

Theorem (Strictly Concave Function). *The function f is* strictly concave *if and only if, $\forall x', x'' \in X$ and $x' \neq x''$,*

$$Df(x')(x'' - x') > f(x'') - f(x').$$

Intuition: When f is strictly concave, no portion of the graph is a line segment. The tangent line at any point lies *strictly* above the graph of the function.

Theorem (Quasi-Concave Function). *The function f is* quasi-concave *if and only if, $\forall x', x'' \in X$,*

$$f(x'') \geq f(x') \Rightarrow Df(x')(x'' - x') \geq 0.$$

Intuition: Essentially, if a function is quasi-concave it may be "single-peaked" and have "flats".
It is tempting to think that f is strictly quasi-concave if and only if $f(x'') > f(x') \Rightarrow Df(x')(x'' - x') > 0$ This is not true as the following example shows:

Example. Let $f : R \to R$ be given by $f(x) = x^3$. It's easy to verify that f is strictly quasi-concave. However, $f'(0) = 0$, so for $x'' > 0$ and $x' = 0$ we have that $f(x'') = x''^3 > 0$ and $f'(x')(x'' - x') = f'(0)x'' = 0$.

This example demonstrates a problem with differential characterizations of quasi-concavity. While any critical point of a concave function is a global maximum, this need not be the case when the function is merely quasi-concave. I.e.. $Df(x') = 0$ may be a false signal.

Since there are no nice "if and only if" results for differentiable strictly quasi-concave functions, it is customary in economics to use slightly stronger definitions of strict quasi-concavity in the differentiable case. One commonly used definition is:

Definition (Differentiably strictly Quasi-Concave Function). *The function f is* differentiably strictly quasi-concave *if and only if, $\forall x', x'' \in X$ and $x' \neq x''$,*

$$f(x'') \geq f(x') \Rightarrow Df(x')(x'' - x') > 0.$$

Intuition: The strict inequality here rules out the possibility of any "flats" in the graph and inflexion points.
Quasi-concave functions may have the undesirable feature that a zero gradient need not imply the function is at a global maximum. In several economic applications it is reasonable simply to assume this possibility away. This useful assumption motivates us to consider a class of quasi-concave functions whose gradients are never zero except at their global maxima. This class of functions are called *pseudo-concave functions.* They are defined only for differentiable functions.

Definition (Pseudo-Concave Function). *The function f is* pseudo-concave *if, $\forall x', x'' \in X$ and $x' \neq x''$,*

$$f(x'') > f(x') \Rightarrow Df(x')(x'' - x') > 0.$$

It is useful to work through why a pseudo-concave function can only achieve its global maximum at a zero gradient point. Suppose f is pseudo-concave, and $Df(x^*) = 0$. If $f(x^*)$ is not the global maximum, then there exists some point $\hat{x} \in X$ with $\hat{x} \neq x^*$ such that $f(\hat{x}) > f(x^*)$. But then we will have $Df(x^*)(\hat{x} - x^*) = 0$ yet $f(\hat{x}) > f(x^*)$, contradicting that f is pseudo-concave.

5.4.2 Concavity Relationships for Differentiable Functions

Theorem. *If f is concave and differentiable then it is pseudo-concave.*
Proof: *Suppose f is concave, $x', x'' \in X$, and w.l.o.g. $f(x'') > f(x')$. Since f concave, $Df(x')(x'' - x') \geq f(x'') - f(x')$. Since $f(x'') > f(x')$, $f(x'') - f(x') > 0$ is immediate. Therefore, $Df(x')(x'' - x') > 0$. So $f(x'') > f(x') \Rightarrow Df(x')(x'' - x') > 0$, which is the definition of pseudo-concavity.* □

Theorem. *If f is pseudo-concave then it is quasi-concave.*
Proof: *Suppose f is pseudo-concave, $x', x'' \in X$, and w.l.o.g. $f(x'') \geq f(x')$. We look at the two cases:*
Case 1: $f(x'') > f(x')$
Since f is pseudo-concave, $f(x'') > f(x') \Rightarrow Df(x')(x''-x') > 0 \Rightarrow Df(x')(x''-x') \geq 0$.
Case 2: $f(x'') = f(x')$
We need to show that: $\forall\ \theta \in [0,1]$, $f(\theta x'' + (1-\theta)x') \geq f(x') = f(x'')$.
Suppose not. That is, suppose $\exists \theta \in (0,1)$ s.t. $f(\theta x'' + (1-\theta)x') < f(x')$.
By pseudo-concavity,
$f(x') > f(\theta x'' + (1-\theta)x') \Rightarrow Df[(\theta x'' + (1-\theta)x')](x' - (\theta x'' + (1-\theta)x')) > 0$.
Simplifying, we get

$$(-\theta)Df[(\theta x'' + (1-\theta)x')](x'' - x') > 0.$$

But since $f(x'') = f(x')$, we also have
$f(x'') > f(\theta x'' + (1-\theta)x') \Rightarrow Df[(\theta x'' + (1-\theta)x')](x'' - (\theta x'' + (1-\theta)x')) > 0$
Simplifying again, we get

$$(1-\theta)Df[(\theta x'' + (1-\theta)x')](x'' - x') > 0.$$

But since $\theta \in (0,1)$, we know $1-\theta > 0$, and $-\theta < 0$, thus we have a contradiction. So we have shown that $\forall \theta \in [0,1]$, $f(\theta x'' + (1-\theta)x') \geq f(x') = f(x'')$. Therefore, we have:

$$\lim_{\theta \to 0} \frac{f(\theta x'' + (1-\theta)x') - f(x')}{\theta} = Df(x')(x'' - x') \geq 0$$

Which implies that f is quasi-concave. □

To prove the other relationships (that if f is strictly concave then it is both concave and strictly quasi-concave) is trivial. There is, however, an additional relationship that should be noted: if f is differentiably strictly quasi-concave, then it is pseudo-concave.

5.4.3 Concavity and Quasi-Concavity for Cont. Twice-Differentiable Functions

Theorem. *Suppose that $f : X \to \mathbb{R}$ is a C^2 function. Then f is concave if and only if the Hessian matrix $D^2 f(x)$ is negative semi-definite on X, i.e., for every $x \in X$, $y' D^2 f(x) y \leq 0$ for every $y \in \mathbb{R}^n$.*

Proof: *($\Rightarrow$) Let $y \in R^n$. Since X is open, $\exists \hat{\lambda}$ such that $x + \lambda y \in X$ for all $0 < \lambda < \hat{\lambda}$. By the characterization of concave functions for C^1, it follows that*

$$f(x + \lambda y) - f(x) - \lambda D f(x) y \leq 0 \text{ for } 0 < \lambda < \hat{\lambda}.$$

But since f is twice differentiable at x, by Taylor's theorem,

$$f(x + \lambda y) - f(x) - \lambda D f(x) y = \frac{\lambda^2 y' D^2 f(x) y}{2} + \lambda^2 r(x, \lambda y) \|y\|^2,$$

with $\lim_{\lambda \to 0} r(x, \lambda y) = 0$. Therefore we get that $\lambda^2 y' D^2 f(x) y \leq 0$.

($\Leftarrow$) By Taylor's theorem we have that for $x', x'' \in X$,

$$f(x'') - f(x') - D f(x')(x'' - x') = \frac{(x'' - x')^T D^2 f(x' + \delta(x'' - x'))(x'' - x')}{2},$$

for some $\delta \in (0, 1)$. But the right hand side of the above equality is non-positive because $D^2 f(x)$ is negative semi-definite and $x' + \delta(x'' - x') \in X$. Therefore

$$f(x'') - f(x') - D f(x')(x'' - x') \leq 0,$$

which shows that f is concave by the characterization for C^1 functions. □

It might be tempting to think that a C^2 function is strictly concave on X if and only if $D^2 f(x)$ is negative definite for all $x \in X$, but that is *not* true.

Theorem. *Suppose that $f : X \to R$ is a C^2 function. A sufficient but not necessary condition for f to be strictly concave on X is that $D^2 f(x)$ be negative definite on X.*

An example of the non-necessity is the function $f(x) = -x^4, x \in R$. In this case, f is strictly concave on R but $D^2 f(x) = -12x^2$ is not negative definite at $x = 0$.

Theorem. *Suppose that $f : X \to R$ is a C^2 function. If f is quasi-concave on X, then $y' D^2 f(x) y \leq 0$ for every $x \in X$ where $y \in R^n$ and $D f(x) y = 0$.*

Proof: *Since X is open, $\forall x \in X$ and $\forall y \in \mathbb{R}^n$ there is some $\bar{t} > 0$ small enough, such that $x + ty$ and $x - ty \in X$, $\forall t$ s.t. $0 < t < \bar{t}$. By quasi-concavity of f, we know that $f(x) = f(\frac{x+ty}{2} + \frac{x-ty}{2}) \geq \min\{f(x + ty), f(x - ty)\}$. We have two cases:*

Case 1. $f(x) \geq f(x + ty)$. From Taylor's expansion,

$$f(x + ty) = f(x) + t D f(x) y + \frac{1}{2} t^2 y^T D^2 f(x) y + r(x, ty) \|ty\|^2.$$

This implies

$$tDf(x)y + \frac{1}{2}\ t^2 y^T D^2 f(x)y + r(x, ty)\ \|ty\|^2 \leq 0.$$

When $Df(x)y = 0$*,*

$$\frac{1}{2}\ t^2 y^T D^2 f(x)y + r(x, ty)\ \|ty\|^2 \leq 0.$$

Then,

$$y^T D^2 f(x)y + \frac{2r(x, ty)\ \|ty\|^2}{t^2} \leq 0.$$

Therefore, as $t \to 0$, $\frac{2r(x,ty)\ \|ty\|^2}{t^2} \to 0$*. Thus,* $y^T D^2 f(x)y \leq 0$*.*
Case 2. $f(x) \geq f(x - ty)$*. Applying Taylor's expansion,*

$$f(x - ty) = f(x) - tDf(x)y + \frac{1}{2}\ t^2 y^T D^2 f(x)y + r(x, -ty)\ \|ty\|^2.$$

This implies

$$-tDf(x)y + \frac{1}{2}\ t^2 y^T D^2 f(x)y + r(x, -ty)\ \|ty\|^2 \leq 0.$$

When $Df(x)y = 0$*,*

$$\frac{1}{2}\ t^2 y^T D^2 f(x)y + r(x, -ty)\ \|ty\|^2 \leq 0.$$

Then,

$$y^T D^2 f(x)y + \frac{2r(x, -ty)\ \|ty\|^2}{t^2} \leq 0.$$

Therefore, as $t \to 0$, $\frac{2r(x,-ty)\ \|ty\|^2}{t^2} \to 0$*. Thus,* $y^T D^2 f(x)y \leq 0$*.*

□

We will end the section by summarizing the relationships between the various forms of concavity for differentiable functions (excluding pseudo-concave functions, the implications hold for continuous but not necessarily differentiable functions as well.):

f strictly concave		$\Rightarrow$		f strictly quasi-concave
$\Downarrow$				$\Downarrow$
f concave	$\Rightarrow$	f pseudo-concave	$\Rightarrow$	f quasi-concave

5.5 Optimization with Inequality Constraints

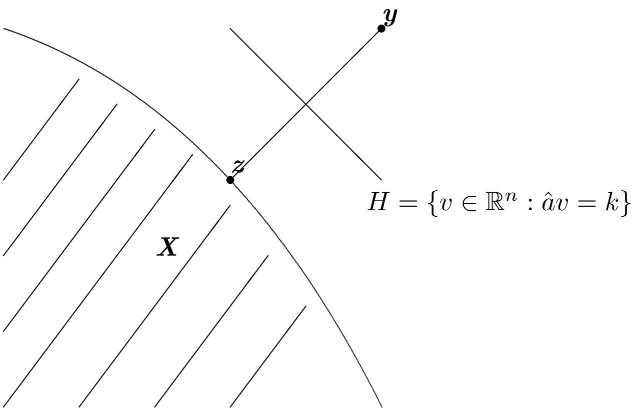
We now turn to a study of optimization problems defined by inequality constraints. Consider the general programming problem

$$\textbf{(P)} \qquad \max_x f(x) \quad \text{subject to} \quad g(x) \geqq 0 \quad \text{and} \quad x \in X.$$

where $X \subset \mathbb{R}^n$, $f : X \to \mathbb{R}$ and $g : X \to \mathbb{R}^m$, which represents the m constraints.

5.5.1 Mathematical Digression: Hyperplane Theorems

Theorem (Separating Hyperplane Theorem). *Let X be a closed and convex set in $\mathbb{R}^n$ and let $y \in \mathbb{R}^n \backslash X$. Then there exists an $\hat{a} \in \mathbb{R}^n$ and a $k \in \mathbb{R}$ such that $\hat{a}x < k < \hat{a}y \quad \forall x \in X$.*
Proof: *Let $z \in X$ be the point in X closest to y. We assume without proof (though it can be proved) that such a z exists and it is unique. We proceed by showing that any vector that is both normal to the vector (y-z) and intersects* strictly *between y and z will "separate" X and y. Perhaps a picture will help:*



Let $\hat{a} = (y-z)^T$ and $k = 1/2(y-z)^T(y+z) = 1/2(y^Ty - z^Tz)$. Then:

1. $\hat{a}y - k = (y-z)^Ty - 1/2(y^Ty - z^Tz) = 1/2y^Ty - z^Ty - 1/2z^Tz = 1/2(y-z)^T(y-z) = 1/2\|y-z\|^2 > 0$. *Therefore* $\hat{a}y - k > 0 \Rightarrow \hat{a}y > k$.

2. $\hat{a}z - k = (y-z)^Tz - 1/2(y^Ty - z^Tz) = -1/2y^Ty + y^Tz - 1/2z^Tz = -1/2(y-z)^T(y-z) = -1/2\|y-z\|^2 < 0$. *Therefore* $\hat{a}z - k < 0 \Rightarrow \hat{a}z < k$. *But we have to show that $\hat{a}x < k$ $\forall x \in X$. Consider any x in X. By definition of z, we know that $\|y-z\|^2 \leq \|y-x\|^2$. Since X is convex, $(1-\theta)z + \theta x \in X$. Hence we have* $\|y-z\|^2 \leq \|y-((1-\theta)z+\theta x)\|^2 = \|(1-\theta)(y-z)+\theta(y-x)\|^2 = (1-\theta)^2\|y-z\|^2 + 2\theta(1-\theta)(y-z)'(y-x) + \theta^2\|y-x\|^2$.

Subtracting the left side from the right and dividing by θ *yields:*
$0 \leq (\theta - 2)\|y - z\|^2 + 2(1-\theta)(y-z)^T(y-x) + \theta\|y-x\|^2$.
Taking the limit as $\theta \to 0$ *gives us*
$0 \leq \|y-z\|^2 + (y-z)^T(y-x) = -(y-z)^T(y-z) + (y-z)^T(y-x)$
$0 \leq (y-z)^T z - (y-z)^T x = \hat{a}z - \hat{a}x \quad \Rightarrow \hat{a}z \geq \hat{a}x$. *Since* $k > \hat{a}z$ *from above,* $k > \hat{a}z \geq \hat{a}x \quad \Rightarrow k > \hat{a}x$.

□

Theorem (One Version of a Supporting Hyperplane Theorem). *If* $X \subset \mathbb{R}^n$ *is convex,* $intX \neq \emptyset$*, and* $x' \in X$ *but* $x' \notin intX$*, then* $\exists v' \in \mathbb{R}^n \backslash \{0\}$ *such that* $v'x \leq v'x' \quad \forall x \in X$.
Proof: *Let* $\bar{X}$ *be the closure of X, also a convex set. Then* $x' \in boundary\bar{X}$*. Now consider a sequence* $\{x'_n\}$ *such that* $x'_n \to x'$ *and* $x'_n \notin \bar{X}\ \forall n$*. Since* $\bar{X}$ *is convex and* $x'_n \notin \bar{X}$*, by Separating Hyperplane Theorem,* $\forall n$, $\exists\ v'_n \in \mathbb{R}^n$ *such that* $v'_n x'_n > v'_n x\ \forall x \in \bar{X}$*. Now w.l.o.g., let* $\|v'_n\| = 1$*, i.e., the sequence begins at the edge of the unit sphere and moves inward. Since the unit sphere is closed and bounded, it is compact. Therefore, by Bolzano-Weierstrass* $\{v'_n\}$ *has a convergent subsequence* $\{v'_{n_k}\}$*. So* $\forall n_k$, $v'_{n_k} x \leq v'_{n_k} x'_{n_k} \forall x \in \bar{X}$*. Taking limit as* $n_k \to \infty$*, we get* $v'x \leq v'x'\ \forall x \in \bar{X}$*. Since* $X \subseteq \bar{X}$*, it follows that* $v'x \leq v'x'$ $\forall x \in X$. □

Theorem (Farkas' Lemma). *Let* $a_1, ..., a_m$ *and b be non-zero vectors in* $\mathbb{R}^n$*, and let A be the matrix whose rows are* $a_1, ..., a_m$ *(*a_i*, b, and* λ *treated as row vectors). Then only one of the following is true:*

1. $\exists\ \lambda_1, ..., \lambda_m \geq 0$ *such that* $b = \sum_{i=1}^m \lambda_i a_i$ *, i.e.,* $\exists\ \lambda \in \mathbb{R}^m_+$ *such that* $b = \lambda A$.

2. $\exists\ x \in \mathbb{R}^n$ *such that* $bx > 0$ *and* $a_i x \leq 0\ \forall i = 1, ..., m$ *i. e,* $\exists\ x \in \mathbb{R}^n$ *such that* $bx > 0$ *and* $Ax \leq 0$.

Proof: *Consider the set Z defined as* $Z = \{z \in \mathbb{R}^n | z = \lambda A$ *for some* $\lambda \in \Lambda = \mathbb{R}^m_+\}$*. Then* $\forall b \in \mathbb{R}^n$*, either* $b \in Z$ *or* $b \notin Z$.
Case 1: $b \in Z$*. By definition of Z,* $\exists \lambda \in \mathbb{R}^m_+$ *such that* $b = \lambda A$*. Now we want to show that there is no* $x \in \mathbb{R}^m$ *such that* $bx > 0$ *and* $Ax \leq 0$*. Suppose by contradiction that there does exist an* $x \in \mathbb{R}^n$ *such that* $bx > 0$ *and* $a_i x \leq 0\ \forall i$*. Since* $b = \lambda A$ *and* $bx > 0$ *we can write out:* $bx = (\lambda_1 a_1 + \cdots + \lambda_m a_m)x > 0$ $\Rightarrow (\lambda_1 a_1)x + \cdots + (\lambda_m a_m)x > 0 \Rightarrow \lambda_1(a_1 x) + \cdots + \lambda_m(a_m x) > 0$.
This implies that, for some i, $\lambda_i(a_i x) > 0$*. Since* $\lambda_i \geq 0$*, it must be that , for some i,* $a_i x > 0$*. But this contradicts the original assumption that* $a_i x \leq 0\ \forall i$.

Case 2: $b \notin Z$*. By definition of Z, clearly there does not exist* $\lambda \in \mathbb{R}^m_+$ *such that* $b = \lambda A$*. We want to show that there must exist an* $x \in \mathbb{R}^n$ *such that* $bx > 0$ *and* $Ax \leq 0$*. First we need to show that Z is closed and convex: to prove Z is convex is trivial. To prove Z closed, consider a convergent sequence* $\{z_n\} \in Z$ *s.t.* $z_n \to z$*. Since* $z_n \in Z\ \forall n$*, for each n,* $\exists \lambda_n$ *such that* $z_n = \lambda_n A$*. Therefore,* $\{z_n\}$ *is equivalent to the sequence* $\{\lambda_n A\}$*, and* $\lambda_n A \to \lambda A$ *for some* λ*. Since*

λ_n is bounded from below by a weak inequality, ($\lambda_n \geq 0$), and unbounded from above, $\lambda_n \to \lambda \in \Lambda$. Therefore $\lambda A \in Z \Rightarrow z \in Z$. Since every convergent sequence in Z converges to a point in Z, it is closed.
It follows from the Separating Hyperplane Theorems that there exists an $x \in \mathbb{R}^n$ and a $k \in \mathbb{R}$ such that $zx < k < bx \ \forall z \in Z$. Since $0 \in Z$, $bx > 0$. Since every $a_i \in Z$ (let $\lambda_i = 1, \quad \lambda_j = 0 \quad \forall i \neq j$), we have $\lambda_i a_i x < k \ \forall i$ and $\lambda_i \geq 0$. Suppose $a_i x > 0$ for some $i = 1, ..., m$, then for a sufficiently big λ_i, we have $k < \lambda_i a_i x$, which cannot be true. Therefore, it has to be that $a_i x \leq 0$ for all i. □

Intuition: The intuition behind Farkas' Lemma is that, for any set of vectors $a_1, ...a_m$ and b, either it is possible to construct b out of a *positive* combination of $a_1, ...a_m$, or there is a vector x that separates b from $a_1, ...a_m$. Consider the following two very simple examples:

Example 1 (Satisfies Case 1, not Case 2): Let $a_1, a_2, b \in \mathbb{R}^2$ where $a_1 = (-1, 1), a_2 = (1, 1)$ and $b = (0, 1)$. Clearly, $\exists \ \lambda \in \mathbb{R}^m_+$ such that $b = \lambda A$, in particular $\lambda = [1/2 \ 1/2]$. However, there does not exist an $x \in \mathbb{R}^n$ such that $bx > 0$ and $Ax \leq 0 \, \forall x \in X$. How can we see this? Let $C = \{x \in \mathbb{R}^2 | bx > 0 \quad \forall x\}$ and $D = \{x \in \mathbb{R}^2 | Ax \leq 0 \quad \forall x\}$. In this example, $C = \{x \in \mathbb{R}^2 | x' \in \mathbb{R}$ and $x'' > 0\}$ and $D = \{x \in \mathbb{R}^2 : |x'| \leq |x''|$ and $x'' \leq 0\}$. You can see that C and D are disjoint , i.e., that there is no x that satisfies Case 2 of Farkas' lemma. This is depicted graphically in Figure I. It depicts the three vectors. Figure II shows that b can be constructed from a positive linear combination of a_1 and a_2, and Figure III depicts that the sets C and D are disjoint.

Example 2 (Satisfies Case 2, not Case 1): Now let $a_1 = (0, 1), a_2 = (1, 1)$, and $b = (-1, 1)$, as represented in Figure IV. Clearly, there is no positive combination of a_1 and a_2 that is equivalent to the vector b, as b does not fall in the shaded region in Figure V. However, in this example $C = \{x \in \mathbb{R}^2 | x'' > x'\}$ and $D = \{x \in \mathbb{R}^2 : x' \leq x''$ and $x'' \leq 0\}$ are not disjoint. As you can see in Figure VI, represented by the region in which C and D overlap, there exists a set of $x \in \mathbb{R}^n$ such that $bx > 0$ and $Ax \leq 0 \ \forall x \in X$. So, for example, if we choose $x = (-2, -1)$, we see that $bx = 1 > 0$ and $Ax = [-1 \ -3] \leq 0$.

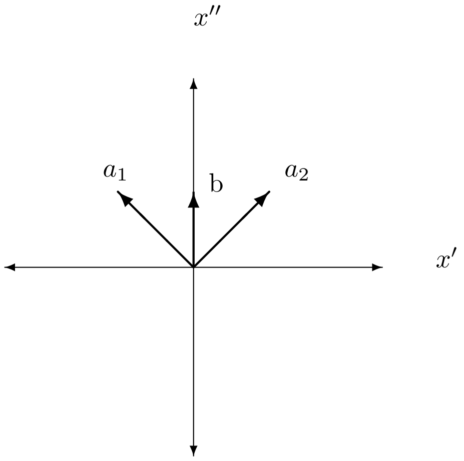


FIGURE I

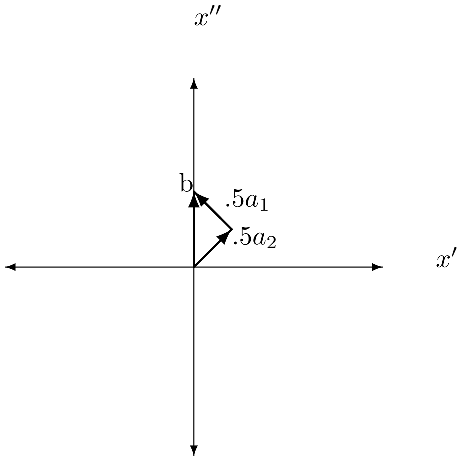


FIGURE II

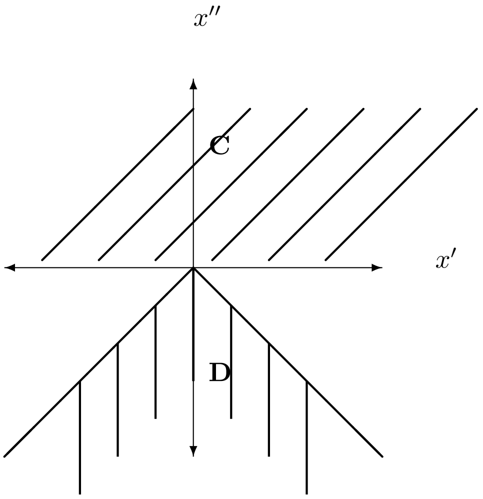


FIGURE III

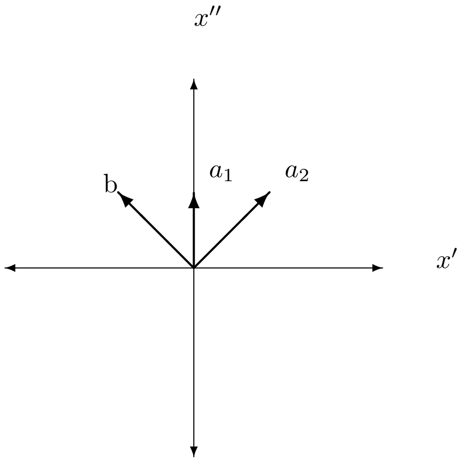


FIGURE IV

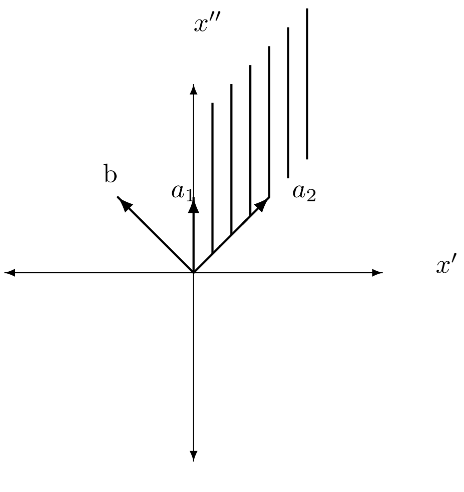


FIGURE V

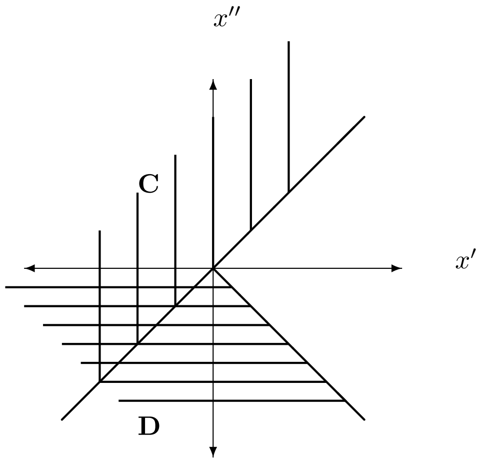


FIGURE VI

5.5.2 The Linear Case (PL)

We begin by deriving necessary and sufficient conditions for an optimal solution to a *linear programming problem.* A linear programming problem is defined as a programming problem in which the objective function and all of the constraints (f and $g_j \forall j$) are linear. Thus, they take the form:

$$f(x) = k + cx,$$

$$g_j(x) = B_j x - b_j \geq 0.$$

Notice that k can be disregarded in the maximization problem, and the constraint can be written in matrix form $Bx - b \geq 0$, where B is an $m \times n$ matrix, x is $n \times 1$, and b is $m \times 1$. Therefore the linear programming problem is,

$$\begin{array}{cc} \max & cx \\ \text{subject to} & Bx - b \geq 0. \end{array} \qquad \textbf{(PL)}$$

Necessary Conditions for an Optimal Solution

Suppose x^* is an optimal solution to (PL). This implies that there does *not* exist a $\Delta x \in \mathbb{R}^n$ sufficiently small that satisfies both of the following two conditions:

1. $x^* + \Delta x$ is in the constraint set. That is

   $$B(x^* + \Delta x) - b \geq 0.$$

   If we partition the matrix B into those constraints that are binding B^*, and those that are non-binding B^{**} and partition the vector b using the same criterium, the inequality above is equivalent to

   $$B^*(x^* + \Delta x) - b^* \geq 0$$

   and

   $$B^{**}(x^* + \Delta x) - b^{**} \gg 0.$$

2. $x^* + \Delta x$ results in a strictly higher value for the objective function. That is, $c(x^* + \Delta x) > cx^*$, or simplified:

   $$-c\Delta x < 0.$$

Now we apply the Farkas' lemma with the following identification:

Farkas' Lemma	Application
A	B^*
b	$-c$
x	Δx
λ	λ

Since we have ruled out Case 2 of Farkas' lemma, it *must* be that $\exists\lambda \in \mathbb{R}_+^{m^*}$ such that

$$-c = \lambda B^* \Rightarrow c + \lambda B^* = 0.$$

Now, define $\lambda^* \in \mathbb{R}_+^m = [\lambda, 0]$ (where m^* is the number of binding constraints) to obtain,

$$c + \lambda^* B = 0.$$

Recall that the first m^* constraints are binding , i.e., $B^* x^* - b^* = 0$, and that $\lambda^*_{m^*+1},...,\lambda^*_m = 0$, so that $\lambda^*(Bx^* - b) = 0$.
Overall, we conclude that if $x^* \in X$ is an optimal solution to (PL), then $\exists\lambda^* \in \mathbb{R}_+^m$ such that

$$c + \lambda^* B = 0,$$

$$\lambda^*(Bx^* - b) = 0.$$

These are the two *necessary conditions* for an optimal solution to the linear programming problem. The first condition is referred to as the *first order condition*, and the second is referred to as the *complementary slackness* condition. An alternative way of writing the complementary slackness condition is

$$(B_j x^* - b_j) \geq 0 \quad \forall j \quad \text{and} = 0 \text{ if } \lambda_j^* > 0.$$

Sufficient Conditions for an Optimal Solution

Suppose $\exists x^* \in X$ and $\lambda^* \in \Lambda \equiv \mathbb{R}_+^m$ such that

$$c + \lambda^* B = 0,$$

$$\lambda^*(Bx^* - b) = 0.$$

Then x^* is an optimal solution to (PL).
Proof: Suppose not. That is, suppose $\exists \Delta x \in \mathbb{R}^n$ such that

$$B(x^* + \Delta x) - b \geq 0,$$

$$c(x^* + \Delta x) - cx^* > 0.$$

Since $\lambda^* \geqq 0$, from the first equation we get $\lambda^*[B(x^* + \Delta x) - b] \geq 0$, or equivalently $\lambda^*(Bx^* - b) + \lambda^* B\Delta x \geq 0$. By complementary slackness, $\lambda^*(Bx^* - b) = 0 \quad \Rightarrow \quad \lambda^* B\Delta x \geq 0$. And from the second equation, we have $c\Delta x > 0$. Therefore $\lambda^* B\Delta x + c\Delta x > 0 \quad \Rightarrow \quad (\lambda^* B + c)\Delta x > 0$. But we assumed $\lambda^* B + c = 0$, so we have a contradiction. □

5.5.3 Non-Linear (Differentiable) Case

Now, let us consider a general maximization problem with non-linear, differentiable functions f and g.

$$\max f(x)$$
$$\text{subject to } g(x) \geq 0. \qquad (\mathbf{P})$$

Consider the function

$$L : X \times \mathbb{R}^m \to \mathbb{R}$$

such that

$$(x, \lambda) \to L(x, \lambda) = f(x) + \lambda g(x),$$

commonly referred to as the *Lagrangian* (function) associated with (P). Associate the *Lagrange multiplier(s)* $\lambda^j \geqq 0$ $(\lambda = (\lambda^1, \lambda^2, ..., \lambda^m) \geqq 0)$ with the constraint(s) $g^j(x) \geqq 0$ $(g(x) \geqq 0)$, and define the *Kuhn-Tucker conditions* by

$$\overbrace{D_{x^i} f(x) + \sum_{j=1}^{m} \lambda^j D_{x^i} g^j(x) = 0}^{\textit{first-order conditions}} \quad , i = 1, 2, ..., n$$

$$\text{and} \quad \overbrace{g^j(x) \geqq 0, \text{ with equality if } \lambda^j > 0 \text{ (or } \lambda^j g^j(x) = 0)}^{\textit{complementary slackness conditions}}, \; j = 1, 2, ..., m$$

or, in compact vector-matrix notation,

$$Df(x) + \lambda Dg(x) = 0$$
$$\text{and} \quad g(x) \geqq 0 \;\&\; \lambda g(x) = 0, \qquad \text{(K-T)}$$

where it is always understood that $(x, \lambda) \in X \times \mathbb{R}^m_+$ (and, by convention, the "price-like" vector λ is treated as a row vector).

Constraint Qualification

Consider an optimal solution x^* to the above problem, and suppose, without loss of generality (i.e., by appropriate relabelling of the constraints with $1, ..., m^*$ as the binding constraints), that

$$g^j(x^*) = 0, \text{ for } j = 1, 2, 3, ..., m^* \leq m$$
$$\text{and } g^j(x^*) > 0, \text{ otherwise }.$$

Also, let

$$g^*(x) = \begin{pmatrix} g^1(x) \\ g^2(x) \\ \vdots \\ g^{m^*}(x) \end{pmatrix}.$$

Then, x^* satisfies the constraint qualification if x^* is also an optimal solution to the linear programming problem

$$\max Df(x^*)x$$
$$\text{subject to } Dg^*(x^*)(x - x^*) \geq 0, \qquad (L^*)$$

which is obtained by replacing g with g^* in (P), linearizing both f and g^* around $x = x^*$, and then dropping redundant terms.

Theorem (Sufficient Conditions for Constraint Qualification (CQ)). *Suppose that x^* is an optimal solution to (P). If one of the following conditions holds, then x^* satisfies the constraint qualification.*

1. ***Linear Constraints.*** *If g^j, all j, are linear, then the CQ is satisfied.*

   Proof: *Consider a hybrid programming problem (PL^*)*

$$\max f(x)$$
$$\textit{subject to } Dg^*(x^*)(x - x^*) \geq 0. \qquad \text{(PL*)}$$

Lemma. *If x^* is an optimal solution to (PL^*), then it is also an optimal solution to (L^*). This holds for general g.*
Proof:*Suppose that x^* is not an optimal solution to (L^*). Then there is $x' \in \mathbb{R}^n$ such that $Df(x^*)x' > Df(x^*)x^*$, or equivalently $Df(x^*)(x' - x^*) > 0$, and $Dg^*(x^*)(x' - x^*) \geq 0$. So, letting $\Delta x' = x' - x^*$, this implies that $\frac{Df(x^*)\Delta x'}{\|\Delta x'\|} > 0$, and for $\alpha \in \mathbb{R}_{++}$,*

$$Dg^*(x^*)\alpha\Delta x' \geq 0.$$

But, since f is differentiable at x^*, we have that

$$\lim_{\substack{\Delta x \neq 0 \\ ||\Delta x|| \to 0}} \frac{f(x^* + \Delta x) - [f(x^*) + Df(x^*)\Delta x]}{\|\Delta x\|} = 0.$$

So, taking $\Delta x = \alpha \Delta x'$ *as a particular case, we have*

$$\lim_{\alpha \to 0^+} \frac{f(x^* + \alpha \Delta x') - [f(x^*) + Df(x^*)\alpha \Delta x']}{\|\alpha \Delta x'\|} =$$

$$\lim_{\alpha \to 0^+} \frac{f(x^* + \alpha \Delta x') - f(x^*)}{\|\alpha \Delta x'\|} - \frac{Df(x^*)\Delta x'}{\|\Delta x'\|} = 0,$$

since we have $\|\alpha \Delta x'\| = |\alpha| \, \|\Delta x'\|$, $\alpha \in \mathbb{R}_{++}$. *Hence, for* $\alpha > 0$ *sufficiently small, we have* $x^* + \alpha \Delta x' \in X$ *, given that* X *is open. From above, we also have* $Df(x^*)\Delta x' > 0$ *, and from the last limit we computed, we obtain*

$$f(x^* + \alpha \Delta x') > f(x^*).$$

□

If g^j*, all* j*, are linear, then* $g^*(x) = g^*(x^*) + Dg^*(x^*)(x - x^*) = Dg^*(x^*)(x - x^*)$. *If* x^* *is an optimal solution to (P), then* x^* *is also an optimal solution to* (PL^*)*. From the above lemma, we know that an optimal solution* x^* *to* (PL^*) *is an optimal solution to* (L^*)*. Therefore, if* x^* *is an optimal solution to (P), and the constraints are linear, then* x^* *is an optimal solution to* (L^*). □

2. ***Slater's Condition***. *If* g^j*, all* j*, are pseudo-concave, and there* $\exists x^+$ *such that* $g(x^+) \gg 0$*, then the CQ is satisfied.*
Proof: *Suppose that the conclusion is false, i.e., that there is* $\widetilde{x} \in \mathbb{R}^n$ *such that*

$$Dg^*(x^*)(\widetilde{x} - x^*) \geqq 0 \text{ and } Df(x^*)\widetilde{x} > Df(x^*)x^*, \text{ or } Df(x^*)(\widetilde{x} - x^*) > 0.$$

By the hypotheses for this case (in particular, that g^j *is pseudo-concave and* $g^j(x^+) > 0 = g^j(x^*)$ *or* $g^j(x^+) - g^j(x^*) > 0$*, for* $j \leq m^*$*), we know that*

$$Dg^*(x^*)(x^+ - x^*) \gg 0.$$

So consider $x^\theta = (1-\theta)\widetilde{x} + \theta x^+$ *for* $0 \leq \theta \leq 1$*. Then for every* $\theta > 0$

$$Dg^*(x^*)(x^\theta - x^*) = (1-\theta)\overbrace{Dg^*(x^*)(\tilde{x} - x^*)}^{\geq 0} + \theta Dg^*(x^*)(x^+ - x^*) \gg 0,$$

while for sufficiently small $\theta > 0$

$$Df(x^*)(x^\theta - x^*) = (1-\theta)\overbrace{Df(x^*)(\widetilde{x} - x^*)}^{>0} + \theta \overbrace{Df(x^*)(x^+ - x^*)}^{\approx 0} > 0.$$

But this means that for sufficiently small $\theta > 0$

$$lim_{\alpha \to 0^+} \frac{g^*(x^* + \alpha(x^\theta - x^*)) - g^*(x^*)}{\|\alpha(x^\theta - x^*)\|} = Dg^*(x^*)(x^\theta - x^*) \gg 0$$

while

$$lim_{\alpha\to 0^+}\frac{f(x^*+\alpha(x^\theta - x^*))-f(x^*)}{\|\alpha(x^\theta - x^*)\|} = Df(x^*)(x^\theta - x^*) > 0,$$

or that for sufficiently small $\theta > 0$ *and* $\alpha > 0$ *(using, in addition, the facts that* X *is open and* $g^j(x^*) > 0$, *for* $j > m^*$*)*

$$x^*+\alpha(x^\theta - x^*) \in X,\ g(x^*+\alpha(x^\theta - x^*)) \geq 0 \text{ (in fact } \gg 0\text{), but } f(x^*+\alpha(x^\theta - x^*)) > f(x^*),$$

or that x^* *is not an optimal solution to* (P), *a contradiction.* □

3. ***Rank Condition***. *If* $rankDg^*(x^*) = m^*$, *the CQ is satisfied.*
   Proof: *Again suppose that the conclusion were false. Then by the hypothesis for this case we can find* $\Delta x \in \mathbb{R}^n$ *such that*

   $$Dg^*(x^*)\Delta x \gg 0,$$

   (for instance, by solving the equations $Dg^*(x^*)\Delta x = \underline{1}$*). So, now simply redefining*

   $$x^\theta = \widetilde{x} + \theta\Delta x$$

   in the preceding argument, it can be readily verified (also a useful exercise) that we again reach exactly the same contradiction. □

5.5.4 Kuhn-Tucker Theorem: Necessity and Sufficiency

Based on what we learned above, the optimal solutions can be characterized by the Theorem of Kuhn-Tucker.

Theorem (Kuhn-Tucker : Necessity). *If* x^* *is an optimal solution to (P) and* x^* *satisfies the constraint qualification, then there exist Lagrange multipliers* $\lambda^* \in \Lambda(= \mathbb{R}^m_+)$ *such that*

$$\begin{aligned} Df(x^*) + \lambda^* Dg(x^*) &= 0,\\ g(x^*) &\geq 0,\\ \textit{and}\quad \lambda^* g(x^*) &= 0 \end{aligned} \tag{K-T}$$

Proof: *Let* $\Delta x = x - x^*$. *Since, by hypothesis,* x^* *is an optimal solution to* (L*),

$$Dg^*(x^*)\Delta x \geqq 0 \Rightarrow Df(x^*)x^* \geqq Df(x^*)x \text{ or } Df(x^*)\Delta x \leqq 0,$$

and the inequalities

$$\begin{aligned} & Dg^*(x^*)\Delta x \geqq 0\\ \textit{and}\quad & Df(x^*)\Delta x > 0 \text{ or } 0 > -Df(x^*)\Delta x \end{aligned}$$

can have no solution. So — applying Farkas' Lemma with $b = -Df(x^*)$,

$A = Dg^*(x^*)$ and $x = \Delta x$ — there is some $\lambda = (\lambda^1, \lambda^2, ..., \lambda^{m^*}) \geqq 0$ such that

$$-Df(x^*) = \lambda Dg^*(x^*)$$

or

$$Df(x^*) + \lambda Dg^*(x^*) = 0$$

Finally, since g^* was defined to include only the binding constraints, if we simply choose

$$\lambda^{j^*} = \begin{cases} \lambda^j, & j = 1, 2, ..., m^* \\ 0, & \textit{otherwise,} \end{cases}$$

then (x^*, λ^*) is a solution to (K-T). □

For f differentiable, the Theorem of Kuhn-Tucker provides conditions that are only necessary for optima that meet the constraint qualification. The conditions it describes may be viewed as the first-order necessary conditions. They provide a set of potential candidates that might be maximizers, but to eliminate some of this candidates we need the Lagrangian to be "locally" concave. To determine the curvature of the Lagrangian we have to look at the second order derivative.

Theorem. ***(Second-Order Necessary Condition for a Local Maximum)*** *Suppose that f, g^j, for all j, are C^2 functions and suppose that x^* is an optimal solution. Let $L(x^*) = L(x^*, \lambda^*)$ be the Lagrangian evaluated at the point (x^*, λ^*) that satisfies (K-T). Then, $x' D^2_{xx} L(x^*) x \leq 0$ for all x such that $Dg^*(x^*) x = 0$.*

Theorem. ***(Second-Order Sufficient Condition for a Local Maximum)*** *Suppose that f, g^j, for all j, are C^2 functions and suppose that $(x^*, \lambda^*) \in X \times \Lambda$ satisfies (K-T). If,*

$$x' D^2_{xx} L(x^*) x < 0 \textit{ for all } x \neq x^* \textit{ such that } Dg^*(x^*) x = 0$$

Then there is a neighborhood U of x^* where $f(x) < f(x^*)$ for all $x \in U \cap C \backslash \{x^*\}$.

But for the purposes of this course, we are more interested in studying the case where the (K-T) conditions are also sufficient. We will see that if f and g satisfy some additional curvature characteristics, (K-T) can also be sufficient to characterize the (global) optimal solution.

Theorem (Kuhn-Tucker: Sufficiency). *If $(x^*, \lambda^*) \in X \times \Lambda$ satisfies (K-T) and f is pseudo-concave and g^j, all j, are quasi-concave, then x^* is an optimal solution to (P).*
Proof: *Suppose that the conclusion were false, i.e., that there is $\widetilde{x} \in C$,the constraint set, such that $f(\widetilde{x}) > f(x^*)$ or $f(\widetilde{x}) - f(x^*) > 0$. Then, on the one hand, since f is pseudo-concave,*

$$Df(x^*)(\widetilde{x} - x^*) > 0. \quad (1)$$

On the other hand, since $g^j, j = 1, 2, ..., m^*$, are quasi-concave,

$$g^j(x^*) = 0 \Rightarrow g^j(\widetilde{x}) \geqq g^j(x^*) \text{ or } g^j(\widetilde{x}) - g^j(x^*) \geq 0 \Rightarrow$$
$$Dg^j(x^*)(\widetilde{x} - x^*) \geqq 0 \Rightarrow \lambda^{j^*} Dg^j(x^*)(\widetilde{x} - x^*) \geqq 0, \text{ for } j \leqq m^*,$$

while, from the second part of (K-T),

$$g^j(x^*) > 0 \Rightarrow \lambda^{j^*} = 0 \Rightarrow \lambda^{j^*} Dg^j(x^*)(\widetilde{x} - x^*) = 0, \text{ for } j > m^*.$$

So

$$\lambda^* Dg(x^*)(\widetilde{x} - x^*) = \sum_{j=1}^{m} \lambda^{j^*} Dg^j(x^*)(\widetilde{x} - x^*) \geqq 0. \tag{2}$$

But adding (1) and (2) together yields

$$\overset{>0}{Df(x^*)(\widetilde{x} - x^*)} + \overset{\geqq 0}{\lambda^* Dg(x^*)(\widetilde{x} - x^*)} = (Df(x^*) + \lambda^* Dg(x^*))(\widetilde{x} - x^*) > 0 \Rightarrow$$
$$Df(x^*) + \lambda^* Dg(x^*) \neq 0,$$

which contradicts the first part of (K-T). □

5.5.5 Optimization with Equality Constraints: Langrange Theorem

Now, let us consider a slightly different problem - a maximization problem with equality constraints - as follows

$$\max f(x)$$
$$\text{subject to } g(x) = 0$$
$$\text{and } x \in X. \tag{P=}$$

Assume that $f : X \to \mathbb{R}$ and $g : X \to \mathbb{R}^m$ are differentiable, and continuously differentiable on $X \subset X'$ open. Then, we can use the following theorem to characterize the optimal solutions to the problem (P=).

Theorem (Lagrange Theorem: Necessity). *If x^* is an optimal solution to the problem (P), and $rankDg(x^*) = m < n$, then there are Lagrange multipliers $\lambda^* \in \mathbb{R}^m$ such that*

$$Df(x^*) + \lambda^* Dg(x^*) = 0.$$

Proof: *Let us partition $x = (x_\alpha, x_\beta)$, where x_α is an $m-$dimensional vector and x_β is an $(n-m)$-dimensional vector. And, without loss of generality, assume that $rankD_{x_\alpha} g(x^*) = m$. Since X is open and $g(x^*_\alpha, x^*_\beta) = 0$, by the Implicit Function Theorem, we can think of open sets U and V containing x^*_α and x^*_β, respectively, and a C^1 function ϕ such that for every $x_\alpha \in U$ and $x_\beta \in V$,*

$$g(x_\alpha, x_\beta) = 0 \Leftrightarrow x_\alpha = \phi(x_\beta),$$

and

$$D\phi(x_\beta^*) = -[D_{x_\alpha} g(x_\alpha^*, x_\beta^*)]^{-1} \cdot D_{x_\beta} g(x_\alpha^*, x_\beta^*). \tag{1}$$

Moreover, on the open sets around x_α^* and x_β^*, x_β^* solves the unconstrained maximization problem

$$\max_{x_\beta \in V} f(\phi(x_\beta), x_\beta).$$

We obtain the following first-order conditions.

$$D_{x_\alpha} f(x_\alpha^*, x_\beta^*) \cdot D_{x_\beta} \phi(x_\beta^*) + D_{x_\beta} f(x_\alpha^*, x_\beta^*) = 0$$

$$\Rightarrow D_{x_\beta} \phi(x_\beta^*) = -[D_{x_\alpha} f(x_\alpha^*, x_\beta^*)]^{-1} \cdot D_{x_\beta} f(x_\alpha^*, x_\beta^*).$$

Plugging this equation into the equation (1), we get

$$-[D_{x_\alpha} f(x_\alpha^*, x_\beta^*)]^{-1} \cdot D_{x_\beta} f(x_\alpha^*, x_\beta^*) = -[D_{x_\alpha} g(x_\alpha^*, x_\beta^*)]^{-1} \cdot D_{x_\beta} g(x_\alpha^*, x_\beta^*)$$

$$\Rightarrow D_{x_\beta} f(x_\alpha^*, x_\beta^*) - D_{x_\alpha} f(x_\alpha^*, x_\beta^*) \cdot [D_{x_\alpha} g(x_\alpha^*, x_\beta^*)]^{-1} \cdot D_{x_\beta} g(x_\alpha^*, x_\beta^*) = 0.$$

If we let $\lambda^* = -[D_{x_\alpha} f(x_\alpha^*, x_\beta^*)] \cdot [D_{x_\alpha} g(x_\alpha^*, x_\beta^*)]^{-1}$, then

$$D_{x_\beta} f(x_\alpha^*, x_\beta^*) + \lambda^* D_{x_\beta} g(x_\alpha^*, x_\beta^*) = 0.$$

Moreover, it is straightforward to get

$$D_{x_\alpha} f(x_\alpha^*, x_\beta^*) + \lambda^* D_{x_\alpha} g(x_\alpha^*, x_\beta^*) =$$
$$D_{x_\alpha} f(x_\alpha^*, x_\beta^*) - D_{x_\alpha} f(x_\alpha^*, x_\beta^*) \cdot [D_{x_\alpha} g(x_\alpha^*, x_\beta^*)]^{-1} D_{x_\alpha} g(x_\alpha^*, x_\beta^*) = 0.$$

Therefore,

$$Df(x_\alpha^*, x_\beta^*) + \lambda^* D_g(x_\alpha^*, x_\beta^*) = 0.$$

□

5.6 (Non-Differentiable) Concave Programming

Recall the general non-linear programming problem:

$$\textbf{(P)} \qquad \max_x f(x)$$
$$\text{subject to } g(x) \geq 0$$
$$\text{and } x \in X.$$

We will assume that f and $g^j, \forall j$ are concave. We will study the Saddle Point theorem and the equivalence of the Saddle Point property (SP) with the Kuhn-Tucker conditions (K-T) when the functions are differentiable.

Theorem (Sufficiency of the SP property). *If* $(x^*, \lambda^*) \in X \times \Lambda$ *satisfies the* Saddle Point property*, which is*

$$\textbf{(SP)} f(x)+\lambda^* g(x) \leq f(x^*)+\lambda^* g(x^*) \leq f(x^*)+\lambda g(x^*), \forall (x,\lambda) \in X \times \Lambda, \quad (5.1)$$

then x^* *is a solution to the original problem (P).*
Proof: *First note that* complementary slackness *must hold. From the second inequality in (SP), we can see that* $f(x^*) + \lambda^* g(x^*) \leq f(x^*) + \lambda g(x^*), \forall \lambda \in \Lambda$. *Since* $\lambda \geq 0$ *and the above inequality must be true even for* $\lambda = 0$*, we can say that* $\lambda^* g(x^*) \leq 0$. *Now, by definition,* $\lambda^* \geq 0$ *and, by the problem setting,* $g(x^*) \geq 0$. *Therefore, we must have that*

$$\lambda^* g(x^*) = 0. \quad (5.2)$$

Now assume, by contradiction, that (x^*, λ^*) *satisfies (SP) but* x^* *is not solution to (P). That is, exists* $\bar{x}$ *s.t.* $f(\bar{x}) > f(x^*)$ *and* $g(\bar{x}) \geq 0$. *Thus,* $f(\bar{x}) + \lambda^* g(\bar{x}) > f(x^*)$. *Using (5.2) that comes from the second inequality of (SP) we get* $f(\bar{x}) + \lambda^* g(\bar{x}) > f(x^*) + \lambda^* g(x^*)$. *This contradicts the first inequality of (SP)* $f(x) + \lambda^* g(x) \leq f(x^*) + \lambda^* g(x^*), \forall (x, \lambda) \in X \times \Lambda$. □

We move on to the proof of necessity of the SP property. For this we will rely on the *Supporting Hyperplane Theorem* (SHT) as described below and proved earlier in the notes.

Theorem (Supporting Hyperplane Theorem). *If* $X \subset \mathbb{R}^n$ *is convex with* $intX \neq 0$ *and* $x' \in bdX$*, then there exists* $v' \in \mathbb{R}^n - \{0\}$ *such that* $v'x \leq v'x', \forall x \in X$.

Theorem (Necessity of the SP property). *If* $x^* \in X$ *is an optimal solution to (P) and* $\exists x^+ \in X$ *such that* $g(x^+) \gg 0$*, then* $\exists \lambda^* \in \Lambda$ *such that (SP) property holds.*
Proof: *Let* $Z \subset R^{m+1}$ *where* $Z = \{z \in R^{m+1} : \text{for } x \in X, z \leq (f(x), g(x))\}$. *Notice that* $Z \neq \emptyset$ *and* $intZ \neq \emptyset$.

Pick $z^* = (f(x^*), 0)$. *By definition of* Z *we can see that* $z^* \in Z$ *and (more strongly)* $z^* \in bdZ$. *In addition,* Z *is convex since* $f(x)$ *and* $g(x)$ *are concave. Now, applying the* SHT *on* z^**, there exists* $(u^*, v^*) \in \mathbb{R}^{m+1} - \{0\}$ *such that* $(u^*, v^*)z \leq (u^*, v^*)z^*, \forall z \in Z$.

Notice the following points.

- $(u^*, v^*) \neq 0$. *This is a direct consequence of* SHT.
- $u^* < 0$ *or* $v^* < 0$ *cannot happen. Suppose either* $u^* < 0$ *or* $v^* < 0$. *Since* Z *has no lower bound we can pick* $z \sim -\infty$ *(i.e.,* z *is a very negative number). This implies* $(u^*, v^*)z > (u^*, v^*)z^*$ *(contradiction). So,* $u^* \geq 0$ *and* $v^* \geq 0$.
- $u^* \neq 0$. *Suppose that* $u^* = 0$. *Then* $(0, v^*)z \leq (0, v^*)z^* \Rightarrow (0, v^*)(f(x), g(x)) \leq (0, v^*)(f(x^*), 0) \Rightarrow v^* g(x) \leq 0, \forall x \in X$. *We know that* $\exists x^+ \in X$ *such that* $g(x^+) \gg 0$. *Then,* $v^* g(x^+) \leq 0 \Rightarrow v^* = 0$. *However, this contradicts the SHT which says that* $(u^*, v^*) \in \mathbb{R}^{m+1} - \{0\}$.

Thus, from the points above we know that $u^* > 0$, $v^* \geq 0$, *and*

$$(u^*, v^*)(f(x), g(x)) \leq (u^*, v^*)(f(x^*), 0), forall x \in X.$$

Define λ^*,

$$\lambda^{*j} = \begin{cases} v_j^*/u^*, & j = 1, 2, ..., m^* \\ 0, & otherwise, \end{cases}$$

where the first m^* *constraints are the binding constraints at* x^*. *Then* $f(x) + \lambda^* g(x) \leq f(x^*) + \lambda^* g(x^*)$, *which is exactly the first inequality of (SP).*

The second inequality comes from the following argument. Since $g^j(x^*) = 0$ *for* $j = 1, ..., m^*$ *and* $\lambda^{*j} = 0$ *for* $j = m^* + 1, ..., m$, $\lambda^* g(x^*) = 0$. *Therefore, because* $\lambda \geq 0$ *and* $g(x^*) \geq 0$, *we have that* $f(x^*) + \lambda^* g(x^*) \leq f(x^*) + \lambda g(x^*)$ $\forall \lambda$., *which is the second inequality of (SP)* □

Theorem. *If* f *and* g^j, *all* j, *are both differentiable and concave, then the K-T conditions are equivalent to the SP property.*

Proof: *(⇐) In this part we assume that the SP property is true and we will show K-T holds*

i) $L(x, \lambda^*) \leq L(x^*, \lambda^*)$ *for all* $x \in X$

$\Rightarrow x^*$ *is the maximizer of* $L(x, \lambda^*)$ *for a given* λ^*

$$\begin{aligned} FOC \quad &\Rightarrow \quad D_x L(x^*, \lambda^*) = 0 \\ &\Rightarrow \quad Df(x^*) + \lambda^* Dg(x^*) = 0 \end{aligned} \tag{5.3}$$

ii) $L(x^*, \lambda^*) \leq L(x^*, \lambda)$ *for all* $\lambda \in R_+^m$

$\Rightarrow \lambda^*$ *is the minimizer of* $L(x^*, \lambda)$ *given* x^*

so λ^* *solves the following problem*

$$\begin{aligned} & max_\lambda - L(x^*, \lambda) \\ subj\ to\ \lambda \ &\geq \ 0 \end{aligned}$$

Now $L(x^*, \lambda)$ *is linear in* λ *hence it is convex and* $-L(x^*, \lambda)$ *is concave in* λ *and the constraint is linear Hence K-T is necessary and sufficient.*

K-T conditions are

$$\begin{aligned} -D_\lambda L(x^*, \lambda^*) + \mu &= 0 \\ \lambda^* \mu &= 0 \end{aligned}$$

From the first order conditions we get,

$$-g(x^*) + \mu = 0 \Rightarrow g(x^*) = \mu \geq 0 \tag{5.4}$$

multiplying both sides by λ *and using the complementary slackness conditions we get*

$$\lambda^* g(x^*) = 0 \tag{5.5}$$

thus (5.3), (5.4), (5.5) together implies that K-T conditions are satisfied.
($\Rightarrow$)*Now the other way round.*
now we assume K-T conditions.
so $\quad Df(x^*) + \lambda^* Dg(x^*) = 0$
$\quad g(x^*) \geq 0$ *and* $\quad \lambda^* g(x^*) = 0$
i) since f *and* g *are concave and* $\lambda \in R^m_+$, $L(x,\lambda)$ *is concave in* x
so the first order conditions implies that x^* *is a global maximum for* $L(x, \lambda^*)$ *for given* λ^*

$$L(x^*, \lambda^*) \geq L(x, \lambda^*) \tag{5.6}$$

ii)from the complementary slackness conditions we know $\lambda \geq 0$ *and* $g(x^*) \geq 0 \Rightarrow \lambda g(x^*) \geq 0 = \lambda^* g(x^*)$
now $L(x^*, \lambda) = f(x^*) + \lambda g(x^*) \geq f(x^*) + \lambda^* g(x^*) = L(x^*, \lambda^*)$

$$L(x^*, \lambda) \geq L(x^*, \lambda^*) \tag{5.7}$$

(5.6) and (5.7) together imply

$$L(x, \lambda^*) \leq L(x^*, \lambda^*) \leq L(x^*, \lambda)$$

Which is the SP property. $\square$

Chapter 6

Regularity of an Optimal Solution

6.1 Mathematical Digression

A parametric family of optimization problems is defined by a parameter space $\Theta \subset \mathbb{R}^l$, a constraint set $C(\theta)$ for each $\theta \in \Theta$, an objective function $f(\cdot, \theta) : \mathbb{R}^n \to \mathbb{R}$, and for each $\theta \in \Theta$, a maximization problem

$$\max_{x \in \mathbb{R}^n} \; f(x, \theta) \tag{6.1}$$
$$\text{subject to } x \in C(\theta).$$

We will define $V : \Theta \to \mathbb{R}$ as the value function which is the value attained by the objective function at a solution to (**??**) when the parameter is $\theta \in \Theta$; and we will define $S(\theta)$ to be the set of maximizers to the problem at θ. That is,

$$V(\theta) = \max\{f(x, \theta) | x \in C(\theta)\},$$
$$S(\theta) = \arg\max\{f(x, \theta) | x \in C(\theta)\}.$$

It is straightforward to see that V is a single-valued function, but that f will not, in general, have a unique maximizer. For this reason, we visit the concept of a correspondence.

Definition (Correspondences). *Let $X \in \mathbb{R}^l$ and $Y \subset \mathbb{R}^n$. A* correspondence *from a set X to a set Y is a map that assigns a set $\Gamma(x) \subset Y$ to each $x \in X$. We can write $\Gamma : X \to 2^Y$. We say that a correspondence Γ is* compact valued *if $\Gamma(x)$ is compact relative to Y for every $x \in X$.*

Notice that a single-valued correspondence is a function. There are several definitions of continuity for correspondences, we will use the following:

Definition (Lower Hemi-Continuity). *A correspondence* $\Gamma : X \rightrightarrows Y$ *is* lower hemi-continuous *(l.h.c.) at* x *if* $\Gamma(x)$ *is nonempty and if, for every* $y \in \Gamma(x)$*, every sequence* $x_n \to x$*, there exists* $N \geq 1$ *and a sequence* $\{y_n\}_{n=N}^{\infty}$ *with* $y_n \in \Gamma(x_n)$ *for all* $n \geq N$ *such that* $y_n \to y$*.*

Definition (Upper Hemi-Continuity). *A compact valued correspondence* $\Gamma : X \rightrightarrows Y$ *is* upper hemi-continuous *(u.h.c.) at* x *if* $\Gamma(x)$ *is nonempty and if, for every sequence* $x_n \to x$ *and every sequence* $\{y_n\}$ *with* $y_n \in \Gamma(x_n)$ *for all* n*, there is a subsequence of* $\{y_n\}$ *that converges to some* $y \in \Gamma(x)$*.*

Notice that we require the correspondence Γ to be compact valued in the definition of u.h.c. but not in the definition of l.h.c. This restriction is used in order to be consistent with other common definitions of u.h.c. in the literature.

Definition (Continuity). *A correspondence* $\Gamma : X \rightrightarrows Y$ *is* continuous *at* $x \in X$ *if it is both u.h.c. and l.h.c.*

We are now ready to present one of the fundamental results of optimization theory, the *Maximum Theorem.* This theorem is essentially a statement that if the primitives of a parametric family of optimization problems possess a sufficient degree of continuity, then the solutions will also be continuous, although perhaps not to the same degree.

6.2 The Maximum Theorem

Theorem (Theorem of the Maximum). *Let* $\Theta \subset \mathbb{R}^l, X \subset \mathbb{R}^n$*. Let* $f : X \times \Theta \to \mathbb{R}$ *be a continuous function and let* $C : \Theta \rightrightarrows \mathbb{R}^n$ *be a compact-valued and continuous correspondence. Then* $V : \Theta \to \mathbb{R}$ *defined in* (**??**) *is a continuous function on* Θ *and the correspondence* $S : \Theta \rightrightarrows \mathbb{R}^n$ *defined in* (**??**) *is non-empty, compact-valued and u.h.c.*

Proof: *Step 1. Fix* $\theta \in \Theta$*. The set* $C(\theta)$ *is nonempty and compact, and* $f(\cdot, \theta)$ *is continuous, hence by the EVT the maximum in* (**??**) *is attained, thus the set* $S(\theta)$ *is nonempty. Moreover since* $S(\theta) \subseteq C(\theta)$ *and* $C(\theta)$ *is compact, it follows that* $S(\theta)$ *is bounded. Suppose* $x_n \to x$ *and* $x_n \in S(\theta)$ *all* n*. Since* $C(\theta)$ *is closed,* $x \in C(\theta)$*. Also, since* $V(\theta) = f(x_n, \theta)$ *all* n *and* f *is continuous, it follows that* $f(x, \theta) = V(\theta)$*, and thus* $x \in S(\theta)$*. Hence* $S(\theta)$ *is also closed. So we have proven that* $S(\theta)$ *is non-empty and compact for each* $\theta \in \Theta$*.*
Step 2. Now we will show that S *is u.h.c. Fix* $\theta \in \Theta$ *and let* $\{\theta_n\}$ *be any sequence converging to* θ*. Choose* $x_n \in S(\theta_n)$ *all* n*. Since* $S(\theta_n) \subseteq C(\theta_n), x_n \in C(\theta_n)$*. Since* C *is u.h.c., there exists a subsequence* $\{x_{n_k}\}$ *with* $x_{n_k} \in C(\theta_{n_k})$ *converging to* $x \in C(\theta)$*. Fix the subsequence* $\{\theta_{n_k}\}$*. Let* $z \in C(\theta)$*. Since* C *is l.h.c., there exists a sequence* $z_{n_k} \to z$ *with* $z_{n_k} \in C(\theta_{n_k})$ *all* k*. Since* $f(x_{n_k}, \theta_{n_k}) \geq f(z_{n_k}, \theta_{n_k})$ *all* k *(because, by assumption,* $x_{n_k} \in S(\theta_{n_k})$*), and since* f *is continuous, it follows that* $f(x, \theta) \geq f(z, \theta)$*. Since this holds for any* $z \in C(\theta)$*, it follows that* $x \in S(\theta)$*. Hence* S *is u.h.c.*

Step 3. Finally, we show that V is continuous. Fix θ, and let $\{\theta_n\}$ be any sequence converging to θ. Choose $x_n \in S(\theta_n)$, all n. Since S is u.h.c., there exists a subsequence of $\{x_n\}$, call it $\{x_{nk}\}$, converging to $x \in S(\theta)$. f is continuous then,

$$V(\theta_{nk}) = f(x_{nk}, \theta_{nk}) \to f(x, \theta) = V(\theta). \tag{6.2}$$

Hence $\{V(\theta_n)\}$ converges and its limit $V(\theta)$, thus V is continuous.

□

When we have standard concavity assumptions on the objective function f, there is a version of the maximum theorem that tells us that V is not only continuous but concave as well.

Theorem (Theorem of the Maximum under Concavity). *Let $\Theta \subset \mathbb{R}^l$, $X \subset \mathbb{R}^n$. Let $f : X \times \Theta \to \mathbb{R}$ be a continuous function and let $C : \Theta \Rightarrow \mathbb{R}^n$ be a compact-valued and continuous correspondence. If furthermore f is concave on $X \times \Theta$, and C has a convex graph, then $V : \Theta \to R$ defined in (**??**) is a continuous and concave function on Θ and the correspondence $S : \Theta \Rightarrow \mathbb{R}^n$ defined in (**??**) is nonempty, compact-valued, and convex valued u.h.c. If concavity of f is replaced by strict concavity, then V is also strictly concave, and S is a continuous function.*

Note: *Be careful to note that the claim of V's concavity is not vacuous. Recall that we argued earlier that concavity implies continuity. However, this implication applies only on the interior of θ, while the theorem claims that V is continuous over all of Θ.*

Proof: *Note first that the* Maximum Theorem *can be applied here. Let $\theta, \bar{\theta} \in \Theta$, and let $\theta' = \lambda\theta + (1-\lambda)\bar{\theta}$ for some $\lambda \in (0,1)$. Pick any $x \in S(\theta)$ and $\bar{x} \in S(\bar{\theta})$. Let $x' == \lambda x + (1-\lambda)\bar{x}$. Then since $x \in C(\theta)$ and $\bar{x} \in C(\bar{\theta})$ and C has a convex graph, it must be that $x' \in C(\theta')$. Therefore by the concavity of f, we have*

$$\begin{aligned} V(\theta') &\geq f(x', \theta') \\ &= f(\lambda x + (1-\lambda)\bar{x}, \lambda\theta + (1-\lambda)\bar{\theta}) \\ &\geq \lambda f(x, \theta) + (1-\lambda) f(\bar{x}, \bar{\theta}) \\ &= \lambda V(\theta) + (1-\lambda) V(\bar{\theta}) \end{aligned}$$

This establish the concavity of V. If f is strictly concave, then the second inequality in this string becomes strict, proving the strict concavity of V. That S must be convex valued is a direct implication of the discussion so far and is left as an exercise. □

Chapter 7

Sensitivity Analysis

Up to now, we have analyzed optimization problems of the form

$$\max_{x \in\ C(\theta)}\ f(x,\theta),$$

where x is a vector of choice and α is a vector of parameters describing the environment. Now we are interested in studying the relationship between x and θ. We want to investigate the marginal effects of changes in θ on the value function $V(\theta) = f(x(\theta), \theta)$.
To study this relationship we can use the Envelope Theorem when the value function is differentiable.

Consider the problem

$$\begin{aligned} &\max_x f\left(x, \theta\right) \\ \text{subject to} \quad & g\left(x, \theta\right) = \mathbf{0} \end{aligned}$$

where the decision variable $x \in R^n$ and the parameters $\theta \in R^l$ (there is no restriction $l \leq n$). The main result is:

Theorem (The Envelope Theorem). ***(Envelope Theorem)*** *Let $f, g_1, ..., g_m$ be C^2 functions from $R^n \times R^l$ to R and suppose that the optimal solution to the above problem $x^* : R^l \to R^n$ is continuously differentiable function of θ. Then*

$$\frac{\partial V\left(\theta\right)}{\partial \theta_k} = \frac{\partial L\left(x^*\left(\theta\right), \theta\right)}{\partial \theta_k} = \frac{\partial f\left(x^*\left(\theta\right), \theta\right)}{\partial \theta_k} + \sum_{j=1}^{m} \lambda_j\left(\theta\right) \frac{\partial g_j\left(x^*\left(\theta\right), \theta\right)}{\partial \theta_k}$$

Proof. *By definition, $V\left(\theta\right) = f\left(x^*\left(\theta\right), \theta\right)$. Taking the derivative of V with respect to θ_k using the chain rule, we obtain:*

$$\frac{\partial V\left(\theta\right)}{\partial \theta_k} = \underbrace{\frac{\partial f\left(x^*\left(\theta\right), \theta\right)}{\partial \theta_k}}_{\textit{Direct effect}} + \underbrace{\sum_{i=1}^{n} \frac{\partial f\left(x^*\left(\theta\right), \theta\right)}{\partial x_i} \frac{\partial x_i^*\left(\theta\right)}{\partial \theta_k}}_{\textit{Indirect effect}}$$

But, the first order conditions for the problem are

$$\frac{\partial f\left(x^{*}\left(\theta\right),\theta\right)}{\partial x_{i}}=-\sum_{j=1}^{m}\lambda_{j}^{*}\left(\theta\right)\frac{\partial g_{j}\left(x^{*}\left(\theta\right),\theta\right)}{\partial x_{i}}\text{ for } i=1,...,n$$

By substitution we get

$$\begin{aligned}\frac{\partial V\left(\theta\right)}{\partial\theta_{k}} &= \frac{\partial f\left(x^{*}\left(\theta\right),\theta\right)}{\partial\theta_{k}}-\sum_{i=1}^{n}\left(\sum_{j=1}^{m}\lambda_{j}^{*}\left(\theta\right)\frac{\partial g_{j}\left(x^{*}\left(\theta\right),\theta\right)}{\partial x_{i}}\right)\frac{\partial x_{i}^{*}\left(\theta\right)}{\partial\theta_{k}}\\ &= \frac{\partial f\left(x^{*}\left(\theta\right),\theta\right)}{\partial\theta_{k}}-\sum_{j=1}^{m}\lambda_{j}^{*}\left(\theta\right)\left(\sum_{i=1}^{n}\frac{\partial g_{j}\left(x^{*}\left(\theta\right),\theta\right)}{\partial x_{i}}\frac{\partial x_{i}^{*}\left(\theta\right)}{\partial\theta_{k}}\right)\end{aligned}$$

But since $g_j\left(x^*\left(\theta\right),\theta\right)=0$ for all $\theta\in R^l$, by differentiating this identity with respect to θ_k, we find that

$$\sum_{i=1}^{n}\frac{\partial g_{j}\left(x^{*}\left(\theta\right),\theta\right)}{\partial x_{i}}\frac{\partial x_{j}^{*}\left(\theta\right)}{\partial\theta_{k}}+\frac{\partial g_{j}\left(x^{*}\left(\theta\right),\theta\right)}{\partial\theta_{k}}=0.$$

Combining the last two equalities we get the necessary result:

$$\frac{\partial V\left(\theta\right)}{\partial\theta_{k}}=\frac{\partial f\left(x^{*}\left(\theta\right),\theta\right)}{\partial\theta_{k}}+\sum_{j=1}^{m}\lambda_{j}^{*}\left(\theta\right)\frac{\partial g_{j}\left(x^{*}\left(\theta\right),\theta\right)}{\partial\theta_{k}}$$

□

Remark. The interpretation of the Envelope Theorem is that when calculating the derivative of the value function with respect to one of the parameters we only have to consider the *direct* effect on the objective function and on the constraints. If the parameters of the problem change, then the optimal x and λ will in general change, but for "small" changes, this indirect effect will be negligible. This result holds only if the optimal solution is differentiable.

Remark. In the case when the constraint set is defined by inequalities, the Lagrange multipliers have the same interpretation. To see this suppose $C=\{x\mid g\left(x\right)\geq\theta\}$ and consider first a constraint j such that $g_j\left(x^*\right)>\theta^*$. In order for the complementary slackness condition to hold, $\lambda_j=0$ and by continuity there is a neighborhood U of θ_j^* such that $g\left(x^*\right)>\theta_j$ for all $\theta_j\in U$. Hence a small change of θ_j from θ_j^* cannot affect the optimal value and

$$0=\lambda_{j}^{*}=\frac{\partial}{\partial\theta_{j}}f\left(x^{*}\left(\theta^{*}\right)\right)$$

still holds. For binding constraints, one can proceed and argue as in the case of equality constraints.

Example. Consider a consumer who seeks to maximize her utility subject to a budget constraint. For concreteness, assume that there are two goods, x_1 and x_2. We denote the prices p_1 and p_2 and the available resources for the consumer by I and assume that the consumer has a logarithmic utility function so that the utility maximization problem can be written as

$$\begin{aligned} & \max_{x_1, x_2} \ln x_1 + \ln x_2 \\ \text{subject to } m - p_1 x_1 - p_2 x_2 &\geq 0 \\ x_1 &\geq 0, x_2 \geq 0 \end{aligned}$$

(The reader should ask herself: what is the parameter space here?)
First, the Slater constraint qualification is satisfied (check it!). Thus the Kuhn-Tucker conditions are necessary for an optimum. The full Kuhn-Tucker conditions for this problem are

$$\begin{aligned} \frac{1}{x_1} - \lambda p_1 + \mu_1 &= 0 \\ \frac{1}{x_2} - \lambda p_2 + \mu_2 &= 0 \\ \lambda \left(p_1 x_1 + p_2 x_2 - m \right) &= 0, \lambda \geq 0 \\ \mu_1 x_1 &= 0 \text{ and } \mu_2 x_2 = 0 \end{aligned}$$

where $\lambda \geq 0$ is the multiplier associated with the budget constraint and $\mu_1, \mu_2 \geq 0$ are the multipliers associated with the non-negativity constraints. It is straightforward to show that any solution must have $(x_1, x_2) \gg (0, 0)$ (verify!) and that the budget constraint must be satisfied with equality so that the remaining equations are

$$\begin{aligned} \frac{1}{x_1} - \lambda p_1 &= 0 \\ \frac{1}{x_2} - \lambda p_2 &= 0 \\ p_1 x_1 + p_2 x_2 &= m \end{aligned}$$

Since the objective is strictly concave and the constraint is linear and therefore concave, the problem must have a unique optimal solution (if it exists). Moreover, the K-T conditions are sufficient for an optimum so if we can find a solution to the preceding equations, it must be the unique global maximum. Indeed it is straightforward to solve for a unique optimal solution given the parameters (p_1, p_2, m)

$$\begin{aligned} x_1^* \left(p_1, p_2, m \right) &= \frac{m}{2 p_1} \\ x_2^* \left(p_1, p_2, m \right) &= \frac{m}{2 p_2} \end{aligned}$$

and for the multiplier associated with the budget constraint, which also will be a function of the parameters

$$\lambda^* (p_1, p_2, m) = \frac{2}{m}$$

the reason that $\lambda^* (\cdot)$ is independent of the prices has to do with the special properties of the utility function. Since we are able to get a closed form solution for the optimal consumption bundle, we obtain the value function simply by plugging the optimal solution into the objective function

$$V (p_1, p_2, m) = 2 \ln m - \ln (2p_1) - \ln (2p_2)$$

In consumer theory this function is often called the **indirect utility function**. Now note that for this particular example, if we take the derivative of the value function with respect to m we get

$$\frac{\partial V (p_1, p_2, m)}{\partial m} = \frac{2}{m} = \lambda^* (p_1, p_2, m)$$

Is this a coincidence? No! In words,this means that the change in the maximized value of the objective function corresponding to a small change in income is given by the Lagrange multiplier. Hence, the Lagrange multiplier measures the value for the consumer of relaxing the constraint somewhat. In textbooks in economics this is often phrased: The Lagrange multiplier in the first order condition is the marginal utility of income.

Note that in the previous example the constraint functions have the form $g(x) \geq \theta$, which is a particular case of $g(x, \theta) \geq 0$.

Remark. Note that if there is a unique global maximum for parameters $\theta \in \Theta$ (as for instance in a strict concave programming problem) then $f (x^* (\theta)) = V (\theta)$ and $\lambda^*_j = \frac{\partial}{\partial \theta_j} f (x^* (\theta^*))$

Now, consider the problem,

$$\max_{x \in \mathbb{R}^n} \; f(x, \theta) \tag{7.1}$$

$$\text{subject to } x \in C(\theta).$$

In the derivations above we just assumed that the optimal solution is a differentiable function of the parameters and to get an exact result we need to make sure that this is true, at least locally around some optimal solution. The Implicit Function Theorem states that we can locally solve for the variables x as a function of the parameters θ and tells us the first-order effects of θ on x at an optimal solution.

Theorem (The Implicit Function Theorem). *Suppose that*

- $V \subset R^j, U \subset R^k$ *and* $W \subset V \times U$ *are open,*

- *$F : W \to R^j$ is C^1, and*
- *$(v', u') \in W$, $F(v', u') = 0$ and rank $D_v F(v', u') = j$.*

Then there are open sets $W' \subset W$ and $U' \subset U$ containing (v', u') and u', respectively, and a C^1 mapping $f : U' \to R^j$ such that

- *for every $u \in U'$, $(v, u) \in W'$ and $F(v, u) = 0 \Leftrightarrow v = f(u)$ (so that, in particular, $v' = f(u')$), and*
- *$Df(u') = -[D_v F(v', u')]^{-1} D_u F(v', u')$ (or, equivalently, the directional derivative $\triangle v = Df(u') \triangle u$ is the unique solution to the system of linear equations*

  $$D_v F(v', u') \triangle v + D_u F(v', u') \triangle u = 0$$

  (given the direction $\triangle u$)).

Note that sometimes it will be more useful to use $\phi(u) = (f(u), u)$.

This theorem gives a sufficient condition for the existence of a continuous function f which establishes the relationship between the "variables", $v \in V$, and "parameters", $u \in U$, of an original function $F(v, u)$. In addition, it tells us how to calculate the derivative of f even without knowing its functional form.

Chapter 8

Leading Examples

8.1 Introduction

Now that the relevant tools for non-linear programming analysis have been presented, we begin the economic analysis. In this chapter we present the leading examples of applications of non-linear programming to the economics field. These examples are the basis for the equilibrium theory analysis.

8.2 Firm's Profit Maximization

Consider the following elements of the firm's maximization problem. The input-output vector of the analyzed firm is $y \in \mathbb{R}^G$, where $G < \infty$ is the total of goods in the economy. The element $y^g \geq 0$ represents output of good g $(g = 1, \dots, G)$ and $y^g < 0$ represents input of good g.

Also, $y \in Y \subset \mathbb{R}^G$, where Y is the Production Set, i.e., the set of technologically feasible input-output vectors. Note that a specific y^g is input or output depending on the considered firm behavior. A good which is output for one firm may be input for another.

Let $t : \mathbb{R}^G \to \mathbb{R}$ be the *Transformation Function* (or the technology) which characterizes the production set in the following way: $Y = \{y \in \mathbb{R}^G : t(y) \geq 0\}$.

Remark: If Y_f is given as the primitive specification,[1] then t_f can be defined as follows. Let

$$
\begin{aligned}
\bar{y}_f &= (y_f^1, y_f^2, \dots, y_f^{g-1}) \text{ and} \\
\bar{Y}_f &= \{\bar{y}_f \in \mathbb{R}^{G-1} : \exists\ y_f^g \in \mathbb{R} \text{ such that } (\bar{y}_f, y_f^g) \in Y_f\} \\
&= \text{the projection of } Y_f \text{ onto its first } g-1 \text{ elements.}
\end{aligned}
$$

[1]Note that $f = 1, 2, \dots, F$ represents a specific firm in the economy.

Define $F_f : \mathbb{R}^{G-1} \to \mathbb{R} \cup \{-\infty\}$ such that $F_f(\bar{y}_f) = \begin{cases} \max(\bar{y}_f, y_f^g) \in Y_f, & y_f^g, \bar{y}_f \in \bar{Y}_f \\ -\infty \quad , & otherwise. \end{cases}$
Finally, define $t_f(y_f) = F_f(\bar{y}_f) - y_f^g$.
Note that F_f is essentially like a production function (with extended domain).

Now we are ready to introduce the abstract representation of the firm's behavior. Consider the problem

$$\Pi(p) = \max_y py \tag{8.1}$$

$$\text{subject to } t(y) \geq 0, \tag{8.2}$$

where $\Pi(p)$ is the *value function*, i.e., the function representing the values that the objective function takes for different environment parameters p when the optimal y is chosen.

8.2.1 Maintained Assumptions

There are some assumptions we will be maintaining throughout this section. Depending on the problem analyzed, different set of assumptions will be invoked.

1. *Positive or non-negative prices:* $p \in P_+ = \mathbb{R}^G_{++}$, i.e., $p >> 0$ or $p \in P = \mathbb{R}^G_+ \backslash \{0\}$, i.e., $p > 0$.
2. *Free Disposal:* if $y_f \in Y_f$ and $y'_f \leq y_f$ then $y'_f \in Y_f$. This is equivalent to assuming that t_f is decreasing. The free disposal assumption means that if a technology can produce a quantity y_f, then it is possible to produce a smaller quantity, say y'_f, by throwing away some inputs or just using them in an inefficient manner.
3. *Regularity:* Y_f is closed or, equivalently, t_f is upper semi-continuous.
4. *Diminishing returns:* Y_f is convex or t_f is concave.
5. *Boundedness:* $\exists \bar{y}_f \in \mathbb{R}^G$ such that if $y_f \in Y_f$ then $y_f \leq \bar{y}_f$.
6. *Irreversibility:* if $y_f \in Y_f$ and $y_f \neq 0$, then $-y_f \notin Y_f$. This assumption basically says that it is not possible to recover inputs from outputs.
7. *No free lunch:* if $y_f \in Y_f$, then $y_f \not> 0$. This means that in order to produce some amount of output, we need a positive amount of inputs.

Note that one can build equivalent definitions of 5, 6, and 7 using the transformation function defined above.

Now we begin studying the properties of the firm's profit maximization problem, described above.

8.2.2 Existence of Solution

To ensure the existence of a solution, an application of the *Extreme Value Theorem* suffices. Note that, given p, the objective function is linear and hence continuous. Closedness of the constraint set is readily derived by the continuity of $t(y)$, according to the regularity assumption. But the constraint set, as given, is not bounded and hence not compact.

To bypass this problem we add a redundant constraint $py \geq -\epsilon$, where $\epsilon > 0$. This additional restriction does not change the set of optimal solutions because there is no feasible solution y' such that $py' < 0$, since $0p = 0$ (i.e., $0 \in Y$ and $t(0) = 0$) is always an option, by the free disposal assumption. Thus, we can get boundedness of the modified set without changing the set of optimal solutions, since the additional constraint only excludes points that can never be optimal. By the same token, because $y = 0$ is always a candidate for a solution, free disposal assures the nonemptyness of the constraint set.

8.2.3 Uniqueness of the Solution

The next step in our analysis is to study the uniqueness of the solution of the firm's problem. To prove uniqueness we assume $t(\cdot)$ is continuous and strictly concave.[2] One way of proving this is the following.

Proof: Suppose we have more than one solution, and y^*, y^{**} are two of the possibly many solutions. Take $y^\theta = (1-\theta)y^* + \theta y^{**}$ for some $\theta \in (0,1)$. By strict concavity, $t(y^\theta) > (1-\theta)t(y^*) + \theta t(y^{**}) \geq 0$, since $y^*, y^{**} \in Y$. Given that $t(\cdot)$ is continuous, there is a neighborhood U of y^θ such that $y \in U \Rightarrow t(y) > 0$. But now there is a point y^+ in this neighborhood such that $y^+ \gg y^\theta$. Since $p \gg 0$, we must have $py^+ > py^\theta = py^* = py^{**}$. This contradicts the hypothesis that y^* and y^{**} are solutions. □

8.2.4 Characterization of the Solution

Now we will analyze the characterization of the solution(s). Basically we want to study how to describe the solution(s) of problem (**??**) by proving the necessity and sufficiency of the Kuhn-Tucker conditions.

Assume that $t(\cdot)$ is differentiable, strictly concave and decreasing (by free disposal assumption). Since $t(\cdot)$ is strictly concave, it is also concave and pseudo-concave. This is one of the two requirements to apply Slater's sufficient condition for fulfilment of constraint qualification.

The other condition is that the constraint set has an interior point. Combining the assumptions above, we can say that $t(\cdot)$ is strictly decreasing.

Since $t(\cdot)$ is strictly decreasing and continuous and $t(0) = 0$, we can find a point y such that $t(y) \geq 0$. Thus Slater's sufficient condition for the constraint qualification is satisfied. This shows necessity.

[2] This is a stronger assumption than Assumption 4 above.

For sufficiency, recall that the objective function is linear and therefore concave and pseudo-concave. Also, note that the constraint function is quasi-concave (since we assumed it is concave). Hence we also have sufficiency.
In conclusion, it is correct to say that the solutions (or solution - as we discussed in the uniqueness part) of problem (**??**) are fully characterized by the following Kuhn-Tucker conditions: There is some $\lambda \in \mathbb{R}_+$ (called Lagrange multiplier) such that

$$p + \lambda Dt(y) = 0, \tag{8.3}$$
$$t(y) \geq 0, = 0 \text{ if } \lambda > 0.$$

8.2.5 Sensitivity Analysis of the Solution

In this section we are interested in proving the existence of a continuous function that relates the parameters present in the profit maximization problem - more specifically, this function is called *supply function* and maps prices p into good supply y. For this, we apply the *Implicit Function Theorem* to a system of equations that characterize the optimal solution.
We will check if all conditions of the theorem hold. For this, consider the additional assumptions that $t(\cdot) \in C^2$. Combining this assumption with Free disposal, $t(0) = 0$, strictly concavity of $t(\cdot)$, we can say that it is also *differentiably strictly decreasing* (i.e., $Dt(y) \ll 0$), and *differentiably strictly concave* (i.e., $\triangle y \in \mathbb{R}^G$ and $\triangle y \neq 0 \Rightarrow \triangle y^T D^2 t(y) \triangle y < 0$).
Consider the Kuhn-Tucker conditions above. Note that μ cannot be zero since $p \gg 0$ and $p + \mu Dt(y) = 0$. Therefore we can reduce these conditions to $\mu > 0$ and

$$p + \mu Dt(y) = 0, \tag{8.4}$$
$$t(y) = 0. \tag{8.5}$$

This is the system of equations for which we will apply the implicit function theorem.

Now define $F : \mathbb{R}^G \times \mathbb{R}_{++} \times \mathbb{R}^G_{++} \to \mathbb{R}^{G+1}$, such that

$$F((y, \mu), p) = \begin{bmatrix} p^T + \mu Dt(y)^T \\ t(y) \end{bmatrix}.$$

Let's make a parallel relations between the parameters and functions of IFT and the elements of the firm's problem.

We can prove that each assumption of theorem IFT hold. First, note that the sets $\mathbb{R}^G \times \mathbb{R}_{++}, P_+$ and $\mathbb{R}^G \times \mathbb{R}_{++} \times P_+$ are open. Second, by the regularity assumption, $t(\cdot) \in C^2$, therefore $F \in C^1$.

So the last condition we are left to verify is the full rank of $D_{(y,\mu)}F(\cdot)$. From equations (**??**) and (**??**), we can get the following:

$$\mu D^2 t(y) \Delta y + Dt(y)^T \Delta \mu = 0, \tag{8.6}$$
$$Dt(y) \Delta y = 0. \tag{8.7}$$

We wish to show that equations (**??**) and (**??**) imply $\Delta y = 0$ and $\Delta\mu = 0$.[3] Premultiplying Δy^T to (**??**) and substituting $Dt(y)\Delta y = \Delta y^T Dt(y)^T = 0$, we obtain

$$\mu \Delta y^T D^2 t(y) \Delta y = 0.$$

Note that $\mu > 0$. From the definition of differentiable strict concavity of $t(\cdot)$, we have that $\Delta y^T D^2 t(y)\Delta y \leq 0 \Rightarrow \Delta y = 0$. Plugging this into (**??**), we can see that $\Delta\mu = 0$, since $Dt(y) \ll 0$. This completes the proof that rank $D_{(y,\mu)}F((y,\mu),p) = G+1$ at $((y,\mu),p)$ for which $F((y,\mu),p) = 0$.

Now we can apply the implicit function theorem and say that there exists functions $y(p)$ and $\mu(p)$ which are locally C^1. More precisely, for every $((y^*,\mu^*),p^*) \in R^G \times R_{++} \times P_+$ for which $F((y^*,\mu^*),p^*) = 0$, there exist a neighborhood U of p^*, V of (y^*,μ^*) and a C^1 function $\theta : U \to V$ such that

$$\theta(p) = (y,\mu),$$
$$F(\theta(p),p) = 0,$$

for every $p \in U$ and $D\theta(p) = -D_{(y,\mu)}F((y,\mu),p)^{-1} D_p F((y,\mu),p)$.

Note that the Lagrange multiplier μ is uniquely determined.

In order to investigate the sensitivity of $y(\cdot)$, instead of using the direct calculation involving the inverse matrix of $D_{(y,\mu)}F((y,\mu),p)$, it is more useful to analyze the *perturbation equations*:

$$\begin{bmatrix} \mu D^2 t(y) & Dt(y)^T & I \\ Dt(y) & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \Delta y \\ \Delta\mu \\ \Delta p \end{bmatrix} = 0,$$

$$\mu D^2 t(y)\Delta y + Dt(y)^T \Delta\mu + \Delta p = 0, \tag{8.8}$$
$$Dt(y)\Delta y = 0. \tag{8.9}$$

Implications from sensitivity analysis

1. *Homogeneity:* Just by looking at the firm's profit maximization problem, it is easy to see that the supply function is homogeneous of degree zero. Consider $\alpha \in \mathbb{R}_+$ is a constant. We can simply argue that if y^* solves

$$\Pi(p) = \max_y py \quad \text{subject to } t(y) \geq 0$$

then it will solve

$$\Pi(\alpha p) = \max_y \alpha p\, y \quad \text{subject to } t(y) \geq 0,$$

[3] This comes from the definition of *full rank* We say that $A \in \mathbb{R}^{N,M}$ has full rank if, and only if, all its rows and columns are linearly independent, i.e., $\forall\ \nu \in \mathbb{R}^{M,K}$, where K is any constant, we have $A\nu = 0 \Rightarrow \nu = 0$.

since multiplication by a constant is a monotone transformation of the original problem. Let's use the perturbation equations to prove this. Suppose $\Delta p = \beta p^T$, $\beta \in (-1, \infty)$. Use the first order conditions of (K-T) to obtain,

$$p^T = -\mu Dt(y)^T \Rightarrow \beta p^T = -\beta\mu Dt(y)^T$$

Substitute in the first perturbation equation (8.8) to get,

$$\mu\Delta^2 t(y)\Delta y + Dt(y)^T\Delta\mu - \beta\mu Dt(y)^T = 0$$

Premultiply by Δy^T,

$$\mu\Delta y^T\Delta^2 t(y)\Delta y + \Delta y^T Dt(y)^T\Delta\mu - \beta\mu\Delta y^T Dt(y)^T = 0$$

From the second perturbation equation (8.9) we know $\Delta y^T \Delta t(y)^T = 0$, then,

$$\mu\Delta y^T D^2 t(y)\Delta y = 0.$$

$\mu > 0$ and $\Delta y \neq 0 \Rightarrow \Delta y^T D^2 t(y)\Delta y < 0$, thus, it has to be that $\Delta y = 0$.

2. *Law of supply:* It is expected that increments in the price of certain output implies increases in the quantity supplied of it. This claim can be proved formally using the perturbation equations.

   From equation (**??**) we have $\Delta y^T Dt(y)^T = 0$. Premultiplying Δy^T to equation (**??**) and using this information, we see

$$\mu\ \Delta y^T D^2 t(y)\Delta y + \Delta y^T Dt(y)^T\Delta\mu + \Delta y^t\Delta p = 0 \Rightarrow$$

$$\mu\ \Delta y^T D^2 t(y)\Delta y + \Delta y^t \Delta p = 0 \Rightarrow \Delta y^t\Delta p = -\mu\ \Delta y^T D^2 t(y)\Delta y > 0 = 0 if \Delta p = \beta p^T$$

.

   The last inequality is by definition of differentiable strict concavity. In particular, take $\Delta p = e^g = (0, \ldots, 0, 1, 0, \ldots, 0) \not\sim p$, $p \in P_+$ to check how the law of supply works in one dimension:

$$\Delta y|^t_{\Delta p = e^g}\Delta p = \Delta y|^t_{\Delta p = e^g} e^g = D_{p^g} y(p) e^g = D_{p^g} y^g(p) > 0$$

   Thus if the price of good g goes up, the supply of the good is increased.

3. *Dependence of π on prices:* As we will prove below, another implication from the perturbation equations is that $D\pi(p) = y(p)^T$.

   Recall the first order condition $p + \mu Dt(y) = 0$. Postmultiplying $\Delta y|_{\Delta p = e^g}$ to this condition and substituting the perturbation equation (**??**), we have

$$p\Delta y|_{\Delta p = e^g} + \mu Dt(y)\Delta y|_{\Delta p = e^g} = p\Delta y|_{\Delta p = e^g} = 0. \quad (8.10)$$

   Keeping this in mind, let's look at the value (profit) function $\pi(p) = py(p)$. Partially differentiating with respect to p^g, we have

$$D_{p^g}\pi(p) = y^g(p) + \sum_{g'} p^{g'} D_{p^g} y^{g'}(p) = y^g + p\Delta y|_{\Delta p = e^g}. \quad (8.11)$$

Using equation (**??**) in (**??**), we get

$$D\pi(p) = y(p)^T.$$

Since $y(p)$ is a C^1 function (from the IFT), the above result yields that $D^2\pi(p)$ exists.

8.3 Household's Utility Maximization Problem

8.3.1 Maintained Assumption and the Problem

We begin with the following definitions:

1. X is the consumption set, or feasible set.
2. $x \equiv (x_1, x_2, \ldots, x_G)$ is the consumption bundle.
3. $p \equiv (p_1, p_2, \ldots, p_G)$ is the price vector (a parameter in this problem).
4. w is the wealth of the household (also a parameter).
5. $B(p, w) = \{x \mid x \in \text{X and } w - px \geq 0\}$ is the *budget*, or constraint set.
6. $u : \text{X} \rightarrow \mathbb{R}$ is the utility function of the household.
7. $V(p, w)$ is the value function called the *indirect utility function*.
8. If $argmax_{x \in B(p,w)}\, u(x) = \{x^*\}$, then $x^* = x(p, w)$ is the *demand function*.

And we add the following assumptions:

1. $p \gg 0$ and $w > 0$.
2. $X = \mathbb{R}^G_+$.
3. u is continuous, quasi-concave, and without local maxima.

Now that we have defined the variables and parameters, and made the necessary assumptions, we address the basic household utility maximization problem:

$$\max\ u(x)$$

$$\text{subject to } w - px \geq 0.$$

We will use λ as the multiplier in the following analysis.

8.3.2 Existence of Solution

From our assumptions, we know the following facts about the constraint set $B(p, w)$:

1. Since $w > 0$, $B(p, w) \neq \emptyset$.
2. $x \geq 0 \Rightarrow x$ is bounded below by a weak inequality.
3. Since $p^g > 0\ \forall g$, $x^g < \infty\ \forall g \Rightarrow x^g$ is bounded from above by a weak inequality for all $g \Rightarrow x$ is bounded above by a weak inequality.

Therefore, $B(p, w)$ is non-empty, and since it is closed and bounded, it is also compact. We assumed that $u(x)$ is continuous, therefore by the EVT there exists a solution to the household utility maximization problem.

8.3.3 Uniqueness of Solution

In order to get uniqueness, we need to strengthen the assumption on $u(x)$. Assume that u is strictly quasi-concave. $B(p, w)$ is convex. By the first theorem of uniqueness we proved, there is at most one optimal solution.

8.3.4 Characterization of Solution

Assume that u therefore pseudo-concave. Since $p \gg 0$ and $w > 0$, there is an interior point to the constraint set. Therefore, we have Slater's condition and necessity of the Kuhn-Tucker conditions. Further, since the objective function is pseudo-concave and the constraint is quasi-concave (linear functions are quasi-concave), we have sufficiency of Kuhn-Tucker conditions. So the Kuhn-Tucker conditions, shown below, are both necessary and sufficient:

$$Du(x)^T - \lambda p^T = 0, \tag{8.12}$$

$$\lambda(w - px) = 0. \tag{8.13}$$

In fact, we will show that $\lambda > 0$, which implies that the complementary slackness condition is:

$$w - px = 0. \tag{8.14}$$

8.3.5 Sensitivity Analysis

The IFT states that if:

1. $(x, \lambda, p, w) \in intX \times int\Lambda \times P_+ \times W_+$. (Interiority)
2. $\begin{Bmatrix} Du(x)^T - \lambda p^T \\ w - px \end{Bmatrix} : X \times \Lambda \times P_+ \times W_+ \to \mathbb{R}^G \times \mathbb{R}_{++}$ is C^1. (Regularity)
3. $\begin{Bmatrix} Du(x)^T - \lambda p^T \\ w - px \end{Bmatrix} = 0 \Rightarrow rankD_{(x,\lambda)} \begin{bmatrix} Du(x)^T - \lambda p^T \\ w - px \end{bmatrix}$ is full. (Rank)

Then, $\exists\ f : P_+ \times W_+ \to X \times \Lambda$ such that $f(p, w) = (x, \lambda)$; or, given $(\Delta p, \Delta w)$, $(\Delta x, \Delta\lambda)$ is the unique solution to the (perturbation) equations:

$$D^2u(x)\Delta x - p^T\Delta\lambda - \lambda\Delta p = 0, \tag{8.15}$$

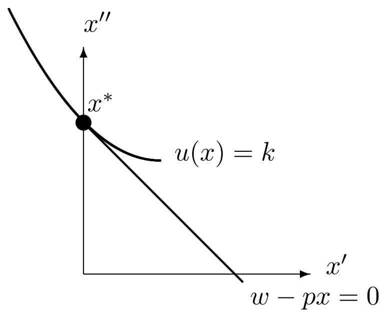
$$-p\Delta x - x^T\Delta p + \Delta w = 0. \tag{8.16}$$

Therefore, in order to apply the IFT, we need interiority, regularity, and rank:

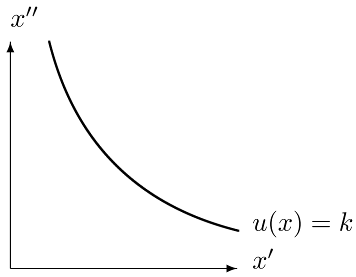
1. Interiority of (x, λ):

   (a) $\lambda > 0$
   Proof: By first order condition, $Du(x)^T - \lambda p^T = 0$. By local non-satiation of $u(x), Du(x) \neq 0$, and since we know $p \gg 0$, it follows that $\lambda \neq 0$. But we also assume $\lambda \geq 0$, so then we conclude $\lambda > 0$. (We could have assumed $u(x)$ differentiable strictly increasing).

   (b) Is there an optimal solution x^* such that $x^* \in intX$? Not necessarily. Consider the following graphical example, where $X = \mathbb{R}^2_+$, and the optimal solution, x^* lies on the boundary of X:



   Therefore, we need to add the following *boundary condition*, which assures us that the upper contour sets of our indifference curves lie in the interior of X:
   Boundary Condition: For any $x' \in intX$, $Cl\{x \in \mathbb{R}^G \mid x \in X$ and $u(x) \geq u(x')\} \subset intX$.
   Graphically, it insures that the upper contour sets look like:



2. Regularity:
   Assume $u(x)$ is C^2 then, $Du(x)$ is C^1 and it can be easily seen that our relationship F, which is actually the two Kuhn-Tucker conditions, is C^1.

3. $D_{(x,\lambda)} \begin{bmatrix} Du(x)^T - \lambda p^T \\ w - px \end{bmatrix}$ has full rank.

   We show this using the following fact:
   A matrix M has full rank iff $Mx = 0 \Rightarrow x = 0$.

   So we will show $\begin{bmatrix} D^2u(x) & -p^T \\ -p & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta\lambda \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} \Delta x \\ \Delta\lambda \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$.
   Multiplying out gives us:

   $$D^2u(x)\Delta x - p^T\Delta\lambda = 0,$$

   $$-p\Delta x = 0.$$

   Premultiplying by Δx^T we get

   $$\Delta x^T D^2u(x)\Delta x - (p\Delta x)^T\Delta\lambda = 0.$$

   And since $-p\Delta x = 0$, we conclude that

   $$\Delta x^T D^2u(x)\Delta x = 0.$$

   Now we prove that this implies that $\Delta x = 0$. Suppose not. That is, suppose $\Delta x \neq 0$. From our F.O.C, we know $Du(x)^T\Delta x^T - \lambda p^T\Delta x^T = 0$. Again, since $p\Delta x = 0$, we have $Du(x)^T\Delta x^T = 0$. By definition of differentiably strict quasi-concavity $\Rightarrow \Delta x^T D^2u(x)\Delta x < 0$, and this is a contradiction. Therefore, $\Delta x = 0$. Since $\Delta x = 0$, it follows that $p\Delta\lambda = 0$. Since $p \gg 0 \Rightarrow \Delta\lambda = 0$. Therefore, we conclude that the rank condition is also satisfied.

In conclusion, we have interiority (with the boundary condition), regularity, and rank condition, so the IFT can be used to do Sensitivity Analysis using the perturbation equations (8.15) and (8.16).

8.3.6 Properties of Optimal Solutions

Let $x(p,w) = argmax_{x \in B(p,w)} u(x)$.

1. $x(p,w)$ is positively (or linearly) homogenous of degree zero.

   (a) (Discrete Case) $x(\theta p, \theta w) = x(p,w) \; \forall \; \theta > -1$. This can be seen quite easily:
   $x(\theta p, \theta w) = \arg\max_x u(x)$ subject to $\theta w - \theta px = 0$ $\Leftrightarrow$ $argmax_x \; u(x)$ subject to $\theta(w - px) = 0$.

Since $\theta > 0$, $argmax_x\ u(x)$ subject to $(w - px) = 0$ $=$ $x(p, w)$.

(b) (Differential Case) $\Delta x|_{(\Delta p, \Delta w)=\theta(p^T, w)} = 0$.
Proof: We know the following from out perturbation and K-T conditions:

$$D^2u(x)\Delta x - p^T\Delta\lambda - \lambda\theta p^T = 0 \quad (8.17)$$

$$-p\Delta x - \theta px + \theta w = 0 \quad (8.18)$$

$$Du(x) + \lambda p = 0 \quad (8.19)$$

$$w - px = 0 \quad (8.20)$$

Then equation (8.22) $\Leftrightarrow -p\Delta x + \theta(w - px) = 0 \Rightarrow -p\Delta x = 0$.

Premultiplying by Δx^T,

equation (**??**) $\Leftrightarrow \Delta x^T D^2u(x)\Delta x - (p\Delta x)^T\Delta\lambda - (p\Delta x)^T(\theta\lambda)^T = 0$.

Since $p\Delta x = 0 \Rightarrow \Delta x^T D^2u(x)\Delta x = 0 \Rightarrow \Delta x = 0$ by previous argument.

2. An increase in wealth has an indeterminate effect on consumption.

   (a) (Discrete) $x^g(p, w + \Delta w) - x^g(p, w)$ can be $>, =$ or < 0.

   (b) (Differential) The sign of $\Delta x\Big|_{\substack{\Delta p=0 \\ \Delta w=1}}$ is indeterminate.

3. The Slutsky Equation (Differential, no Discrete Equivalent)

$$\Delta x\Big|_{\substack{\Delta p \\ \Delta w=0}} = \Delta x'\Big|_{\substack{\Delta p' \\ \Delta w'=x^T\Delta p}} - \Delta x''\Big|_{\substack{\Delta p''=0 \\ \Delta w''=1}} \cdot x^T\Delta p$$

The Slutsky Equation is derived from the perturbation equations. We know that for any $(\Delta p, \Delta w)$ there is a unique $(\Delta x, \Delta\lambda)$ such that the perturbation equations are satisfied. The strategy here is to analyze an overall change in price, with no overall change in wealth, in two steps. First, we look at the change in consumption resulting from the change in price, while adjusting wealth to allow the original bundle to be affordable (this is called the *substitution effect*). Second, we readjust the wealth and look at the change in consumption resulting from just the change in wealth (the so-called *wealth effect*).
Therefore, we start with a change in price $\Delta p' = \Delta p$ and a change in wealth $\Delta w' = x^T\Delta p$, and we know that there exists a unique $(\Delta x', \Delta\lambda')$ that satisfies:

$$D^2u(x)\Delta x' - p^T\Delta\lambda' - \lambda\Delta p = 0,$$

$$-p\Delta x' - x^T\Delta p + x^T\Delta p = 0.$$

We have two equations and two unknowns, so we can solve for $\Delta x'$ and $\Delta\lambda'$. This $\Delta x'$ represents the substitution effect. Similarly, the wealth effect is determined by the solution $(\Delta x'', \Delta\lambda'')$ to the perturbation equations when the price does not change $(\Delta p'' = 0)$, and the wealth is adjusted such that the overall change in wealth is 0.$(\Delta w'' = -x^T\Delta p)$:

$$D^2u(x)\Delta x'' - p^T\Delta\lambda'' = 0,$$

$$-p\Delta x'' - x^T\Delta p = 0.$$

Add the two sets of perturbation equations to obtain,

$$D^2u(x)[\Delta x' + \Delta x''] - p^T[\Delta\lambda' + \Delta\lambda''] - \lambda\Delta p = 0,$$

$$-p[\Delta x' + \Delta x''] - x^T\Delta p = 0.$$

Define

$$\Delta x = \Delta x' + \Delta x''$$

$$\Delta\lambda = \Delta\lambda' + \Delta\lambda''$$

Then the above equations are the perturbations equations of our problem given the following changes in the parameters,

$$\Delta p = \Delta p' + \Delta p'' = \Delta p' + 0 \neq 0$$

$$\Delta w = \Delta w' + \Delta w'' = x^T\Delta p - x^T\Delta p = 0$$

Appendix

The question is how to check the definiteness of a symmetric matrix. If A is a 2×2 matrix it is sometimes possible to use brute force to determine if it is (semi-) definite or not, but in higher dimensions this approach is not feasible anymore. There are essentially two general ways to approach the problem: we can either study the *principal minors* of the matrix A or the *eigenvalues* of the matrix. Sometimes, when we know more about structure of the matrix A, we can take some convenient short cuts for determining the definiteness of A. For example, we may know that some of the elements of A are zeros. Your Econ 701 teaching assistants will show you those shortcuts. Here, we'll be nasty, brutish, and long.

Checking Definiteness Using Principal Minors

Definition. *Let A be an $n \times n$ symmetric matrix. A $k \times k$ submatrix of A formed by deleting $n-k$ columns, say columns $i_1, i_2, ...i_k$, and the same $n-k$ rows, say rows $i_1, i_2, ...i_k$, from A is called a kth order* ***principal submatrix*** *of A. The determinant of a $k \times k$ principal submatrix is called a kth order* ***principal minor*** *of A.*

Example.
For a general 3×3 matrix

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

there is one third order principal minor, $\det(A)$. There are three second order principal minors:

1. $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$, formed by eliminating column 3 and row 3 from A.

2. $\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}$, formed by eliminating column 2 and row 2 from A.

3. $\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$, formed by eliminating column 1 and row 1 from A.

there are three first order principal minors:

1. $|a_{11}|$, formed by eliminating columns 2 and 3 and rows 2 and 3 from A.

2. $|a_{22}|$, formed by eliminating columns 1 and 3 and rows 1 and 3 from A.

3. $|a_{33}|$, formed by eliminating columns 1 and 2 and rows 1 and 2 from A.

Definition. *Let A be an $n \times n$ symmetric matrix. The kth order principal submatrix of A obtained by deleting the* last $n-k$ *columns and the* last $n-k$ *rows from A is called the kth order* ***leading principal submatrix*** *of A. Its determinant is called the kth order* ***leading principal minor*** *of A.*

A $n \times n$ matrix has n leading principal submatrices - the top-left-most 1×1 submatrix, the top-left-most 2×2 submatrix, the top-left-most 3×3 submatrix, etc.

Example.
For a general 3×3 matrix

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

the three leading principal minors are

1. $\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$: the 3^{rd} order leading principal minor

2. $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$: the 2^{nd} order leading principal minor

3. $|a_{11}|$: the 1^{st} order leading principal minor

Theorem. *Let A be an $n \times n$ symmetric matrix. Then,*
1. A is positive definite *if and only if all its n* leading *principal minors are (strictly) positive.*
2. A is negative definite *if and only if its n* leading *principal minors alternate in signs as follows:*
$|A_1| < 0 \quad |A_2| > 0 \quad |A_3| < 0,$ *etc., that is, the kth order leading principal minor should have the same sign as $(-1)^k$.*

It might be tempting to think that a symmetric matrix is positive *semi-definite* if and only if all its leading principal minors are non-negative. But this is not true. To test the semi-definiteness of a matrix using principal minors, we need to examine *all* of its principal minors, not just the leading ones.

Theorem. *Let A be an $n \times n$ symmetric matrix. Then,*
1. A is positive semi-definite *if and only if all its* principal minors *are non-negative;*
2. A is negative semi-definite *if and only if all its k^{th}* principal minors *satisfy $(-1)^k \left|M^k\right| \geq 0$ for all $k = 1, ..., n$ where $\left|M^k\right|$ denotes any k^{th} order principal minor.*

If some kth order principal minor of A (or some pair of them) is nonzero but does not fit either of the above two sign patterns, then A is indefinite. This case occurs when A has a negative kth order leading principal minor for an even integer k or when A has a negative kth order principal minor and a positive lth order principal minor for two distinct odd integers k and l.

Checking Definiteness Using Eigenvalues

There is a less straightforward, but more elegant way to check whether a symmetric matrix A satisfies any of the definiteness criteria defined above. This method, which is used frequently in computational macroeconomics, looks to A's eigenvalues. Underlying this method is a fundamental result of linear algebra:

Theorem. ***(Diagonalization Theorem)*** *Let A be a real symmetric matrix. Then*

1. *All eigenvalues $\lambda_1, ..., \lambda_n$ are real numbers;*
2. *The associated eigenvectors are linearly independent and form an orthonormal matrix P such that*

$$P^{-1}AP = P'AP = Diag\left(\lambda_1, ..., \lambda_n\right).$$

This elaborate terminology has the following derivation: A set of vectors $\{x_1, ..., x_n\}$ is called **orthogonal** if each of the vectors in the set is orthogonal to all the others, i.e.. $x_i \cdot x_j = 0$ for all $i \neq j$. Thinking of a matrix as a collection of vectors it is therefore natural to call a symmetric matrix orthogonal if the scalar product between any row and the transpose of any other row is the zero vector. Clearly $P'P$ is a diagonal matrix if P is orthogonal. The name **orthonormality** is reserved for orthogonal matrices where $P'P$ is the identity matrix. Let P be a matrix that diagonalizes A as in the theorem and let $y = P'x$. Then

$$x'Ax = (Py)'\, APy = y'P'APy = y'Diag\left(\lambda_1, ..., \lambda_n\right) y = \sum_i \lambda_i y_i^2$$

It follows directly from this calculation that:

Theorem. *Let A be a symmetric and let $\lambda_1, ..., \lambda_n$ be the associated (real) eigenvalues. Then*

1. *A is positive definite if and only if $\lambda_i > 0$ for each i*

2. *A is positive semi-definite if and only if* $\lambda_i \geq 0$ *for each* i

3. *A is negative definite if and only if* $\lambda_i < 0$ *for each* i

4. *A is negative semi-definite if and only if* $\lambda_i \leq 0$ *for each* i

5. *A is indefinite if and only if at least two eigenvalues have opposite signs.*

Exercise.
Prove A is negative definite if and only if its n leading principal minors alternate in signs as follows
$|A_1| < 0$ $|A_2| > 0$ $|A_3| < 0,$ etc. The kth order leading principal minor should have the same sign as $(-1)^k$.

Exercise.
Explain why all diagonal elements a_{ii} must be positive for any positive definite matrix.

Exercise.
Suppose that A is a symmetric matrix and that Q is a nonsingular matrix. Then $Q^T AQ$ is a symmetric matrix, and A is positive (negative) definite if and only if $Q^T AQ$ is positive (negative) definite.

Bibliography

1. Department of Economics, University of Pennsylvania (2002). Economics 897: Mathematics for Economists. Mimeo.

2. Royden, H. L. (1968). Real Analysis. Second edition. Macmillan publishing co.

3. Rudin, Walter (1976). Principle of mathematical analysis. Third edition. Mcgraw-Hill.

4. Sundaram, Rangarajan (1996). A first course in optimization theory. First edition. Cambridge university press.