ECON5110: Microeconomics

Lecture 1: Sept, 2017

Contents

1 Overview

Course Aims. This course is a graduate level introduction to microeconomic theory, aiming to acquaint students with the basics of modern microeconomic analysis. During the course, the students will learn the core tools that economists use to analyse problems of resource allocation in market settings. We will begin with a review of the prerequisite mathematical tools. We will then apply these tools to the formal analysis of the optimizing behaviour of producers and consumers. Finally, we will introduce markets and the notion of competitive equilibrium, first in a partial equilibrium setting and then in a general equilibrium setting.

2 Very Basics

Theorems and Proofs. The course is structured according to the "definition-theorem-proof" style of lecture, popular in mathematics. Primarily, we will be interested in two things:

- Defining mathematical structures that capture some aspect of reality. This includes defining mathematical objects and making assumptions on the relations between them. Importantly, we must argue (rhetorically) that our definitions/assumptions are good definitions.
- How to use these structures to prove *conclusions* from our *assumptions*. The conclusions are the theorems, the assumptions are baked into the definitions, the method of transporting assumptions to conclusions will be that of *proof*.

A proof is a series of logical deductions that begin with known facts (either assumptions or previously proved theorems) and ending with the conclusion to be drawn. Logical proofs are usually very long a tedious, we will make use of common understanding, short hand, and analogy to make the process easier.

Types of Proof. We have a premise α and want to prove β. That is, we want to show that *whenever* α is true, β is true as well. This allows for the state of affairs where α is *not* true (from a false premise, we can prove anything!) or that β is always true (anything proves a tautology). An equivalent way of thinking about this is that α is a **sufficient** condition of β. If we want to know β holds, it is sufficient to show α holds (which, by our proof will imply β).

A **direct** proof would assume α and go on to show that β also holds. This is easy when we have a concrete object to work with, but often types of proof work well. Notice that if saying "β is true whenever α is true" is exactly the same as saying "if β is false, it must be that α is false." Instead of proving $\alpha \implies \beta$ we can prove $\neg\beta \implies \neg\alpha$. This is called a proof by **contraposition**. So, in analogy to sufficiency, we can say β is a **necessary** condition for α: without β being true, we know α could not be true, so in order for α it is necessary that β.

Even more abstractly, we could show that $\alpha \implies \beta$ is false by showing that $\neg(\alpha \implies \beta)$ implies something false! That is we assume α and $\neg\beta$ and go on to show something that is known to be false. This is called proof by **contradiction**.

Often we have two premises that we want to show are equivalent, that they each imply the other: $\alpha \iff \beta$ (read: if and only if). We sometimes say α is necessary and sufficient for β. This means we have two things to prove, first that $\alpha \implies \beta$ and second that $\beta \implies \alpha$.

Relations and Orderings. If we have any set X a **relation**, R, on X is a subset of $X^2 = \prod_2 X = X \times X$. If $x, y \in X$, we say xRy if and only if $(x, y) \in R$. There are many kinds of relations, for example, a function from $\mathbb{R} \to \mathbb{R}$ is a special type of relation, as are equivalence relations and orderings. It is the later that we will pay special attention to now.

A relation R over a set X is called

reflexive	iff $(x, x) \in R$, for $x \in X$,
complete	iff $(x, y) \in R$ or $(y, x) \in R$ for all $x, y \in X$,
transitive	iff $(x, y), (y, z) \in R$ implies $(x, z) \in R$ for all $x, y, z \in X$,
negative transitive	iff $(x, y), (y, z) \notin R$ implies $(x, z) \notin R$ for all $x, y, z \in X$,
symmetric	$(x, y) \in R$ implies $(y, x) \in R$ for all $x, y \in X$,
asymmetric	iff $(x, y) \in R$ implies $(y, x) \notin R$ for all $x, y \in X$,

We will call a relation, R, an **equivalence-relation** if it is reflexive symmetric and transitive. **partial-order** if it is reflexive and transitive; a **weak-order** is a complete partial-order and a *strict-order* is asymmetric and transitive. We often denote orderings by $\geq$, with the obvious example being the normal order over $\mathbb{R}$.

Remark 1. *Let $\geq$ be a partial order over a set X, and let $>$ be the relation defined $x > y$ if $x \geq y$ and not $y \geq x$. Then $>$ is a strict order.*

Proof. We need to show two things: that $>$ is asymmetric and that it is transitive. We will start with asymmetry. We need to show that for any $x, y \in X$ such that $x > y$ its not true that $y > x$. So consider any x, y such that $x > y$. By definition $x > y$ means that $x \geq y$ but not $y \geq x$. Since its not true that $y \geq x$, it cannot be that $x > y$, hence $>$ is asymmetric. Now towards negative transitivity: assume that not $x > y$ nor $y > z$. We want to show that $y \geq x$. There are two cases: (i) $x \geq y$, then since not $x > y$ it must be that $y \geq x$. (ii) not $x \geq y$, then by completeness of $\geq$, we have $y \geq x$. Since in either case $y \geq x$, we know it holds unconditionally. A similar argument shows $z \geq y$ so by the transitivity of $\geq$, we know $z \geq x$. Thus, it cannot be that $x > z$, proving negative transitivity. ∎

Remark 2. *Let $R_1 \dots R_n$ be a set of weak orders over a set X. Define $R^\star$ over X by $xR^\star y$ if and only if $xR_i y$ for all $i \leq n$. Then $R^\star$ is a partial order.*

3 The vector-space $\mathbb{R}^n$

Let $\mathbb{R}$ denote your garden variety real numbers. Examples are many (uncountable in fact): 3, π, $\pi - 11$, and $2^{\frac{1}{2}}$ are all real numbers. For an integer, $n \in \mathbb{N}$, let $\mathbb{R}^n$ denote the n-fold Cartesian product of $\mathbb{R}$. That is, n copies of $\mathbb{R}$, and can be denoted by $\prod_n \mathbb{R}$. An element in $\mathbb{R}$ will be denoted by the boldfaced $\boldsymbol{x} = (x_1, \dots x_n) \in \mathbb{R}^n$. Each x_i is a real number. $\mathbb{R}^n$ plays an important role in this course. If a consumer can decide how much of multiple goods to consume, she is choosing a point in $\mathbb{R}^n$, each coordinate of the chosen vector represents the amount to consume of a corresponding good. For two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ let $\boldsymbol{x} \cdot \boldsymbol{y} = \sum_n x_i y_i$.

What is an ordering over R^n? We want something that extends our natural order over $\mathbb{R}$. We say that $\boldsymbol{x} \geq \boldsymbol{y}$ if and only if $x_i \geq y_i$ for each $i \leq n$. Likewise we will say that $\boldsymbol{x} > \boldsymbol{y}$ if $x_i > y_i$ for each $i \leq n$. By Remark 2, we know that $\geq$ is a partial order.

$\mathbb{R}^n$ can be equipped with vector operations to make it a vector space over $\mathbb{R}$. We need to define:

- Vector addition: $\boldsymbol{x} + \boldsymbol{y} = (x_1 + y_1, \ldots, x_n + y_n)$.
- Scalar multiplication: for $a \in \mathbb{R}$ let $a\boldsymbol{x} = (ax_1, \ldots, ax_n)$.

If $B = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m\}$ is a set vectors, we say $\boldsymbol{y} = \sum_m t_i \boldsymbol{x}_i$ is a **linear-combination** of B (where each $t_i \in \mathbb{R}$). We say B is linearly independent if no $\boldsymbol{x} \in B$ is a linear-combination of $B \setminus \{\boldsymbol{x}\}$.

Remark 3. *B is linearly independent if and only there is no t_i's (not all equal to zero) such that $\sum_m t_i \boldsymbol{x}_i = 0$.*

Proof. We need to show both directions (that is what is meant by if and only if): **Only if:** We will do this via a proof by contradiction. Assume B is linearly independent but there exists t_i's (not all equal to zero) such that $\sum_m t_i \boldsymbol{x}_i = 0$. So then $t_j \boldsymbol{x}_m = \sum_{i \neq j} t_i \boldsymbol{x}_i$ for some $t_j \neq 0$; so, $\boldsymbol{x}_j = \sum_{i \neq j} \frac{t_i}{t_m} \boldsymbol{x}_i$. The other direction is an exercise. ■

If B is a collection of vectors, then we say that B is **convex** if for all $\boldsymbol{x}, \boldsymbol{y} \in B$, $s\boldsymbol{x} + (1-s)\boldsymbol{y} \in B$ for any $s \in [0, 1]$. Given a vector, $\boldsymbol{p}$ and a point a, a **hyper-plane** is the set of vectors $H(p, a) = \{\boldsymbol{x} \mid \boldsymbol{a} \cdot \boldsymbol{x} = a\}$. A hyper-plane is an $n-1$ dimensional space, and is therefore convex.

4 Classes of Functions on $\mathbb{R}^n$

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a function. We say f is **linear** if $f(s\boldsymbol{x} + t\boldsymbol{y}) = sf(\boldsymbol{x}) + tf(\boldsymbol{y})$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. Linear functions are generalizations of lines. It is easy to see that if f is linear then $f(0) = 0$ (why?)

Remark 4. *If $f : \mathbb{R}^n \to \mathbb{R}$ is linear than for any finite collection of m vectors, we have $f(\sum_m t_i \boldsymbol{x}_i) = \sum_m t_i f(\boldsymbol{x}_i)$.*

We will prove this claim by **induction**. Mathematical induction is a proof technique used when we want to prove that a particular property holds for every natural number (in the above case, m). We must first show that the property holds when $m = 1$. This is called the base case. We then show that if the property holds for m it will hold for $m + 1$. This is called the inductive step. Together these show the property holds for all naturals. Be careful, however, as induction says nothing about what happens at infinity.

Proof. The base case, where $m = 1$ holds trivially. Assume the proof holds for m and consider $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{m+1}\}$. We have

$$\begin{aligned} f(\sum_{m+1} t_i \boldsymbol{x}_i) &= f(\sum_{m} t_i \boldsymbol{x}_i + t_m \boldsymbol{x}_m) \\ &= \sum_{m} t_i f(\boldsymbol{x}_i) + t_m f(\boldsymbol{x}_m) \\ &= \sum_{m+1} t_i f(\boldsymbol{x}_i) \end{aligned}$$

■

Linear functions play a key role in economics because of a special case where $\sum_m t_i = 1$ and $t_i \geq 0$ for all i. Here the t_i's can represent probabilities, and f is an expectation operator. These results generalize to infinite dimensional spaces. Taking the expectation of a random variable is the evaluation of a linear functional, as in, more generally, integration.

Remark 5. *If $f : \mathbb{R}^n \to \mathbb{R}$ is linear if and only if there exists a vector $\boldsymbol{p} \in \mathbb{R}^n$ such that $f(\boldsymbol{x}) = \boldsymbol{p} \cdot \boldsymbol{x}$.*

Proof. Let $a_i = f(e_i)$ where e_i is the unit vector in the i^{th} dimension (all 0's except a 1 in coordinate i). ■

A more general class of functions which also play a key role in economics are **affine** functions. $f : \mathbb{R}^n \to \mathbb{R}^m$ is affine if and only if $f(s\boldsymbol{x} + t\boldsymbol{y}) = sf(\boldsymbol{x}) + (1 - s)f(\boldsymbol{y})$ for all $s \in [0, 1]$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. Clearly all linear functions are affine. Affine functions are also generalizations of lines, but they do not necessarily pass through the origin. Notice $f : R \to R$, $f : x \mapsto 2 + 2x$ is affine.

Remark 6. *If $f : \mathbb{R}^n \to \mathbb{R}$ is affine if and only if there exists a vector $\boldsymbol{p} \in \mathbb{R}^n$ and a scalar $a \in \mathbb{R}$ such that $f(\boldsymbol{x}) = a + \boldsymbol{p} \cdot \boldsymbol{x}$.*

From this remark we can see that hyperplanes are exactly the solutions (0's) of affine functions.

Two more important properties of functions are convex and concave functions. A function is **convex** if $f(s\boldsymbol{x}+t\boldsymbol{y}) \leq sf(\boldsymbol{x})+(1-s)f(\boldsymbol{y})$ for all $s \in [0,1]$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and strictly so if the inequality is strict. The opposite notion: a function is **concave** if $-f$ is convex, or, more directly if $f(s\boldsymbol{x}+t\boldsymbol{y}) \geq sf(\boldsymbol{x})+(1-s)f(\boldsymbol{y})$ for all $s \in [0,1]$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. Convex functions and convex sets are related: the epi-graph of a convex function is a convex set. A function is concave and convex if and only if it is affine.

Relatedly, a function is **quasi-convex** if $f(s\boldsymbol{x}+t\boldsymbol{y}) \leq \max\{f(\boldsymbol{x}), f(\boldsymbol{y})\}$ for all $s \in [0,1]$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and strictly so if the inequality is strict. **Quasi-concave** can be defined using patter matching skills!

Remark 7. *Every convex function is quasi-convex.*

Proof. By the definition of max $sf(\boldsymbol{x})+(1-s)f(\boldsymbol{y}) \leq s\max\{f(\boldsymbol{x}), f(\boldsymbol{y})+(1-s)\max\{f(\boldsymbol{x}), f(\boldsymbol{y}) = \max\{f(\boldsymbol{x}), f(\boldsymbol{y})$. ∎

Remark 8. *Let f be a quasi-convex function, then $\{\boldsymbol{x} \mid f(\boldsymbol{x}) \leq a\}$ is a convex set for all $a \in \mathbb{R}$.*

Finally, we will consider the properties of homogeneity and homotheticity. These are related to how the functions scale as the inputs are scaled multiplicatively. If $f : \mathbb{R}^n \to \mathbb{R}$ is a function, and k is an integer, then f is said to be **homogeneous of degree** k if

$$f(a\boldsymbol{x}) = a^k f(\boldsymbol{x})$$

for all $\boldsymbol{x} \in \mathbb{R}^n$. Homogeneity is a generalization of linearity, since by definition of Remark 4 linear functions are homogenous of degree 1. Not every h.d.1 function is linear, however; take as an example $f(x,y) = (x^2+y^2)^{\frac{1}{2}}$. The function x^2 is homogenous of degree 2, where as x^2y^2 is of degree 4. Constant functions are homogenous of degree 0.

Remark 9. *If f is differentiable and homogenous of degree k, then $\frac{\partial f}{\partial x_i}$ is homogenous of degree $k-1$.*

Proof. Differentiate both sides of $f(a\boldsymbol{x}) = a^k f(\boldsymbol{x})$ with respect to x_i. ∎

5 Optimization

One of the main themes of economics theory is the idea of optimization. Given a set of constraints that limit what a decision maker might choose, we generally assume that the decision maker will choose optimally. Optimally according to what? We will dive deeper into that later, and it depends on context, but for now we can assume that there is a *function* determining how good each option is. Therefore, identifying the optimal decision amounts to finding the point that maximizes the function.

Given a function: $f : \mathbb{R}^n \to \mathbb{R}$, we say that $\boldsymbol{x}$ is a **maximum** if $f(\boldsymbol{x}) \geq f(\boldsymbol{y})$ for all other vectors $\boldsymbol{y}$. It is a strict maximum if the inequality is strict. Minima are defined analogously.

When f is twice differentiable, and there are no constraints on which vectors can be chosen, things are relatively straight forward. Notation, ∇f is the **gradient** of f, the vector of partial first derivatives and $\nabla^2 f$ is the **Hessian**, the matrix of partial second derivatives. Recall, the first and second order conditions:

$$\nabla f(x) = 0,$$
$$\nabla^2 f(x) \leq 0.$$

Note, these are necessary but not sufficient conditions for maximality. Local also satisfy these conditions.

There is a relationship between the derivative and the properties of functions described in the pervious section. (Assuming everything is differentiable) the derivative of a linear function is constant, whereas the derivative of a convex (concave) function is always increasing (decreasing). This implies that, for convex (concave) function, any point that satisfies the first order condition, is a minimum (maximum).

In general, an optimal point might not exists. For the simplest possible example, let $f(x) = x$ be the identity function, for which there is clearly no maximum or minimum. The derivative is constant but non-zero. However, this assumes that we are allowed to choose any point in $\mathbb{R}$. What if we are only able to choose points in $[0, 1]$? Then, $x^\star = 1$ is a maximum. More generally:

Theorem 10. *Let $B \subset \mathbb{R}^n$. If B is compact (read: closed and bounded), then every continuous function, $f : B \to \mathbb{R}$, attains a maximum and a minimum.*

Recall a set is **closed** if it contains all of its limit points, or, equivalently, if it is the complement of an open set. For our purposes, we will generally deal with closed sets that are closed intervals (or higher dimensional analogs). Note that sets defined by a *finite* number of linear inequalities are closed. A set, B, is **bounded**, if there exists a $\boldsymbol{x} \in \mathbb{R}$ such that $-\boldsymbol{x} \leq \boldsymbol{y} \leq \boldsymbol{x}$ for all $\boldsymbol{y} \in B$.

The Envelope Theorem and Comparative Statics. Assume we have a function f which determines the optimal decision of a given agent. This function might depend not only on the choice of the agent, but also on some external *parameters* which the agent takes as give. For example, a consumer who has a fixed budget, the parameters might be prices. For a firm deciding how much to produce, it might be the marginal cost of each good. Theoretical economics is largely interested in general findings, facts about the problem which do not depend too much on the specific parameters; hence we may not be interested in the optimal value for a particular firm, but rather the relationship between price and optimal consumption or marginal cost and optimal production. That is, how does the value of f at the optimum and the input that optimizes f depend on the parameters of the problem.

Specifically, let $f : \mathbb{R}^{n+k} \to \mathbb{R}$. Here there is a n quantities need to be chosen by the agent, and the problem depends on k constraints. Assume that f is differentiable and has a maximum for any set of parameters. Consider the map $x^\star : \mathbb{R}^k \to \mathbb{R}^n$:

$$x^\star : \boldsymbol{a} \mapsto \arg\max_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}, \boldsymbol{a}).$$

and the map $f^\star : \mathbb{R}^k \to \mathbb{R}$:

$$f^\star : \boldsymbol{a} \mapsto f(x^\star(\boldsymbol{a}), \boldsymbol{a}).$$

That is, for a fixed set of parameters, $\boldsymbol{a}$, we think of f as a function with those variables fixed, and optimize over the remaining variables. $x^\star$ is then the map which carries each set of parameters to the set of inputs which maximize the corresponding projection of f. Then $f^\star$ is the optimal value, given that the inputs were optimized. We are interested in how $x^\star$ and $f^\star$ responds to changes in $\boldsymbol{a}$.

Theorem 11 (The Envelope Theorem)**.** *Let $f : \mathbb{R}^{n+k} \to \mathbb{R}$. Then so long as both partial derivatives exist, we have $\frac{\partial f^\star}{\partial \boldsymbol{a}} = \frac{\partial f}{\partial \boldsymbol{a}}$.*

Proof. For some fixed $\boldsymbol{a}' \in \mathbb{R}^k$, define the function $g(\boldsymbol{a}) = f^\star(\boldsymbol{a}) - f(\boldsymbol{x}^\star(\boldsymbol{a}'), \boldsymbol{a})$. The left term of g is maximized all of the time, by assumption, whereas the right term is necessarily maximized only when $\boldsymbol{a} = \boldsymbol{a}'$. Therefore, g reaches a minimum at $\boldsymbol{a} = \boldsymbol{a}'$. The first order condition for this is that:

$$\frac{\partial g}{\partial \boldsymbol{a}} = \frac{\partial f^\star(\boldsymbol{a})}{\partial \boldsymbol{a}} - \frac{\partial f(\boldsymbol{x}^\star(\boldsymbol{a}'), \boldsymbol{a})}{\partial \boldsymbol{a}} = 0$$

. But notice, $\boldsymbol{x}^\star(\boldsymbol{a}')$ is fixed, and therefore does not enter the derivative. ■

Alternatively:

Proof. By the chain rule we have $\frac{\partial f^\star}{\partial \boldsymbol{a}} = \frac{\partial f}{\partial \boldsymbol{x}}\frac{\partial x^\star}{\partial \boldsymbol{a}} + \frac{\partial f}{\partial \boldsymbol{a}}$. By definition $x^\star(\boldsymbol{a})$ optimizes f, so by the first order condition we have $\frac{\partial f}{\partial \boldsymbol{x}} = 0$ at the optimum. Thus the first term drops out. ■

This result states that, at the optimum, only the *direct* effect of a change in parameters matters. This is even though, the optimal inputs are also able to change in response to the changing parameters.

Example 1. *Consider a firm with a two inputs, k and l. The input is turned into output at a rate of $f(x, k)$. The firm can sell the output at price p and buy the input at price c_k and c_l. Hence the profit function of the firm is:*

$$\pi(k, l, p, c_k, c_l) = pf(k, l) - c_k k - c_l l$$

The first order conditions for k and l are:

$$p\frac{\partial f}{\partial k} - c_k = 0 \qquad (\pi_k)$$
$$p\frac{\partial f}{\partial l} - c_l = 0 \qquad (\pi_l)$$

The solutions to these two equations dictate the demand functions for each input ($k^\star$ and $l^\star$). So what is the firm's profit when the prices are given by (p, c_k, c_l)? It is $\pi^\star(p, c_k, c_l) = \pi(k^\star(p, c_k, c_l), l^\star(p, c_k, c_l), p, c_k, c_l)$. How does the firms profit change when the cost of labour changes? We are interested in $\frac{\partial \pi^\star}{\partial c_l}$.

This is easy *with the envelope theorem. We have $\frac{\partial \pi^\star}{\partial c_l} = \frac{\partial \pi}{\partial c_l} = -c_l$! Profits fall by the same about that costs increase. Even thought the firm can make substitutions between inputs to re-optimize, a process that ostensibly changes profits, we can ignore such changes outright. This is because the marginal change in profits from a small change in either input is already 0 at the optimum.*

Example 1 (Continued)**.** *Now, assume that $f(k, l) = \ln(kl)$. How does $k^\star$, the optimal value of capital, depend on the price of capital, on labour? The first order condition, π_k becomes, $\frac{p}{k^\star} - c_k = 0$, so that $k^\star = \frac{p}{c_k}$. So $\frac{\partial k^\star}{\partial c_k} = \frac{-p}{c_k^2}$.*

6 Constrained Optimization

So far, we have dealt with functions which can be optimized over the entire domain; to be concrete, in the above example, the firm can choose *any* allocation of inputs. Many economic situations are constrained by budgets, feasibility constraints, time, etc. How do we optimize with respect to such constraints?

Equality Constraints. Say we want to maximize $f : \mathbb{R}^n \to \mathbb{R}$ subject to the set of constraints $g_j : \mathbb{R}^n \to \mathbb{R}$ are equal to 0, for $1 \leq j \leq k$. That is we have the problem $\max_{\boldsymbol{x}} f(\boldsymbol{x})$ such that $g(\boldsymbol{x}) = 0$. We can use the method of Lagrange multipliers. The **Lagrangian** of the above problem is:

$$\mathcal{L}(\boldsymbol{x}) = f(\boldsymbol{x}) - \sum_{i \leq k} \lambda_i g_i(\boldsymbol{x}),$$

where λ_j is an additional variable associated with the j^{th} constraints, which we add to the problem called a **Lagrange multiplier**. We then optimize $\mathcal{L}$ as before, including taking the derivative with respect to λ. We have $n + k$ first order conditions:

$$\frac{\partial f(\boldsymbol{x})}{\partial x_i} - \sum_j \lambda_j \frac{\partial g_j(\boldsymbol{x})}{\partial x_i} = 0, \text{ for each } i \tag{$\mathcal{L}_i$}$$

$$g_j(\boldsymbol{x}) = 0, \text{ for each } j \tag{$\mathcal{L}_{\lambda_j}$}$$

This is $n+k$ unknowns in as many equations. We would also generally have to check the second order conditions (that the Hessian is definite) but we can ignore that most of the time when we know the flavor of the functions we are optimizing.

Example 2. *Let $f(x, y) = x+2y$, and $g(x, y) = x^2+y^2 = 5$. Then $\mathcal{L}(x, y) = x+2y-\lambda(x^2+y^2)$. The first order conditions are*

$$1 - \lambda 2x = 0 \tag{$\mathcal{L}_x$}$$

$$2 - \lambda 2y = 0 \tag{$\mathcal{L}_y$}$$

$$x^2 + y^2 = 9 \tag{$\mathcal{L}_\lambda$}$$

Submitting out for λ we have $\frac{1}{y} = \frac{1}{2x}$ or that $y = 2x$. Plugging into $\mathcal{L}_\lambda$ we see that $5x^2 = 5$ or that $x = 1$, so $y = 2$.

The multiplier λ has an interpretation. It is the amount by which the optimized objective function could increase if the constraint was marginally relaxed.

Theorem 12 (The Constrained Envelope Theorem)**.** *Let $f, g : \mathbb{R}^{n+k} \to \mathbb{R}$. We want to solve* $\max_{\boldsymbol{x}} f(\boldsymbol{x}, \boldsymbol{a})$ *subject to $g(\boldsymbol{x}, \boldsymbol{a}) = 0$. Let $f^\star$ be the optimized value. Then so long as both partial derivatives exist, we have $\frac{\partial f^\star}{\partial \boldsymbol{a}} = \frac{\partial f}{\partial \boldsymbol{a}} - \lambda \frac{\partial f}{\partial \boldsymbol{a}}$.*

Proof. Let $\mathcal{L}$ be the associated Lagrangian. The first order conditions are given by (CS_i) and (FSB_1) and argmax $\boldsymbol{x}^\star(\boldsymbol{a})$. Then $f^\star(\boldsymbol{a}) = f(\boldsymbol{x}^\star(\boldsymbol{a}), \boldsymbol{a})$. We have:

$$\frac{\partial f^\star}{\partial \boldsymbol{a}} = \frac{\partial f}{\partial \boldsymbol{x}}\frac{\partial x^\star}{\partial \boldsymbol{a}} + \frac{\partial f}{\partial \boldsymbol{a}}.$$

Substituting via (CS_i) we have

$$\frac{\partial f^\star}{\partial \boldsymbol{a}} = \lambda\frac{\partial g}{\partial \boldsymbol{x}}\frac{\partial x^\star}{\partial \boldsymbol{a}} + \frac{\partial f}{\partial \boldsymbol{a}}. \tag{1}$$

At the optimum, $g(\boldsymbol{x}^\star(\boldsymbol{a}), \boldsymbol{a}) = 0$. Differentiating this identity with respect to $\boldsymbol{a}$ we have $\frac{\partial g}{\partial \boldsymbol{x}}\frac{\partial x^\star}{\partial \boldsymbol{a}} + \frac{\partial g}{\partial \boldsymbol{a}} = 0$. Reragning and plugging into (1), gives us

$$\frac{\partial f^\star}{\partial \boldsymbol{a}} = \frac{\partial f}{\partial \boldsymbol{a}} - \lambda\frac{\partial g}{\partial \boldsymbol{a}}.$$

■

Inequality Constraints. Finally, what about inequality constraints. Where rather than $g(\boldsymbol{x}) = 0$ we want $g(\boldsymbol{x}) \leq 0$. Such constraints are very natural: I may have a budget but I need not spend all if it. The approach is largely similar to above, but we allow for the possibility that the constraint does not **bind**. That is, at the optimum, it might be that $g(\boldsymbol{x}) < 0$. A constraint that is not binding is called **slack**. The approach we employ is Karush-Kuhn-Tucker, or KKT, optimization. Again, we have the Lagrangian: $\mathcal{L}(\boldsymbol{x}) = f(\boldsymbol{x}) - \sum_j \mu_j g_i(\boldsymbol{x})$.

Theorem 13 (KKT)**.** *If $\boldsymbol{x}^*$ is a maximizer (which meets some technical conditions[1]) of f subject to $g_1 \dots g_k$, then there exists a $\mu^\star$ such that:*

1. *Stationarity, or FOC:* $\nabla_{\boldsymbol{x}^\star}\mathcal{L} = 0$
2. *Positive multipliers:* $\mu_j^\star \geq 0$
3. *Complimentary slackness:* $\mu_j^\star g_j(\boldsymbol{x}^\star) = 0$
4. *Primal feasibility:* $g_j(\boldsymbol{x}^\star) \leq 0$
5. *Second order conditions are met.*

Here, there are two possibilities: first, the constraint is binding. In this case $g_j(\boldsymbol{x}^*) = 0$. In such a state of affairs, we should obtain the same solution as if we had used an equality constraint. Since $g_j(\boldsymbol{x}^*) = 0$ condition 3 of the Theorem has no additional bite. Therefore, the only additional restriction as compared with the general Lagrangian method is that the

[1] These conditions are call constraint qualification conditions which ensure the boundary of the feasible set is well behaved. If our constraints are linear, it is always met, so we will ignore such technicalities for this class.

multiplier must be positive. Conversely: If the constraint is not binding then condition 3 implies $\mu_j = 0$. So $\mathcal{L} = f$ and the maximizing input would be exactly the same had there been no constraint (think about why this is intuitively). Further, under the interpretation of multipliers, this means a slight relaxation of the constraint has no impact on the the optimal value of f.

Example 3. *Consider maximizing $f(x, y) = \sum_{i=1}^{n} \ln(1+\frac{x_i}{p_i})$ subject to $\sum_{i=1}^{n} x_i \leq 1$ and $x_i \geq 0$ for each i. Assume each $p_i > 0$. The Lagrangian is*

$$\mathcal{L} = \sum_{i=1}^{n} \ln(1 + \frac{x_i}{p_i}) - \mu_0(\sum_{i=1}^{n} x_i - 1) - \sum_{i=1}^{n} \mu_i(-x_i).$$

The KKT conditions:

$$\frac{1}{p_i + x_i} = \mu_0 - \mu_i \qquad (\mathcal{L}_{x_i})$$

$$\mu_0(\sum_{i=1}^{n} x_i - 1) = 0 \qquad (CS_0)$$

$$\mu_i x_i = 0 \qquad (CS_i)$$

$$\mu_0, \mu_i \geq 0 \qquad (\geq)$$

$$\sum_{i=1}^{n} x_i \leq 1 \qquad (FSB_0)$$

$$x_i \leq 0 \qquad (FSB_1)$$

Notice in the first constraint, we have that the left hand side is strictly positive, so $\mu_0 > 0$. Now assume that $n = 2$, so we have $x_1 = 1 - x_2$. We will try to find a solution such that both x_i's are positive. So then $p_1 + x_1^\star = \frac{1}{\mu_0} = p_2 + x_2^\star$. Which is feasible if and only if $|p_1 - p_2| \leq 1$, otherwise, $x_1^\star = 1$.

Example 4. *An agent can consume 2 goods, the amounts of which are denoted by x and y. When x of the first good and y of the second good are consumed, she receives a utility of xy. Assume the prices of both goods are equal and normalized to 1. She has 100 dollars to spend. Moreover, there ration on the first good, so she cannot consume more than 40 units.*

She wants to maximize: $f(x, y) = xy$ subject to $x + y - 100 \leq 0$ and $x - 40 \leq 0$ and that $x \geq 0$ and $y \geq 0$. We have

$$\mathcal{L} = xy - \mu_1(x + y - 100) - \mu_2(x - 40) + \mu_x(-x) + \mu_y(-y)$$

We have the first order conditions:

$$\begin{aligned}
\mathcal{L}_x: &\quad y - \mu_1 - \mu_2 - \mu_x = 0 \\
\mathcal{L}_y: &\quad x - \mu_1 - \mu_y = 0 \\
\mathcal{L}_{\mu_1}: &\quad \mu_1(x + y - 100) = 0 \\
\mathcal{L}_{\mu_2}: &\quad \mu_2(x - 40) = 0 \\
\mathcal{L}_{\mu_x}: &\quad \mu_x x = 0 \\
\mathcal{L}_{\mu_y}: &\quad \mu_y y = 0
\end{aligned}$$

We also have our non-negativity constraints given by 2 and 4 of the KKT theorem. First we will reduce this with some common sense. First, notice that $x = y = 1$ *is feasible and* $f(1,1) = 1 > 0 = f(x,0) = f(0,y)$*. So neither good will be consumed at the rate 0. We can drop the last to conditions above, knowing that* $\mu_x = \mu_y = 0$*. Now, what if* $\mu_1 = 0$*? This implies that* $x = 0$ *by* $\mathcal{L}_y$*. But we just argued* $x > 0$ *so we know that* μ_1 *must bind. Therefore we have by* $\mathcal{L}_1$*:* $x = 100 - y$*. Likewise, what if* $\mu_2 = 0$ *Then from* $\mathcal{L}_x$ *and* $\mathcal{L}_y$ *we would have* $x = y$ *so by our previous substitution,* $x = y = 50$*. But this is not feasible. So* $x = 40$*. Hence,* $y = 60$*.*