Useful Math for Microeconomics*

Jonathan Levin Antonio Rangel

September 2001

1 Introduction

Most economic models are based on the solution of optimization problems. These notes outline some of the basic tools needed to solve these problems. It is wort spending some time becoming comfortable with them — you will use them a lot!

We will consider *parametric constrained optimization problems* (PCOP) of the form

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

Here f is the objective function (e.g. profits, utility), x is a choice variable (e.g. how many widgets to produce, how much beer to buy), $D(\theta)$ is the set of available choices, and θ is an exogeneous parameter that may affect both the objective function and the choice set (the price of widgets or beer, or the number of dollars in one's wallet). Each parameter θ defines a specific problem (e.g. how much beer to buy given that I have \$20 and beer costs \$4 a bottle). If we let Θ denote the set of all possible parameter values, then Θ is associated with a whole class of optimization problems.

In studying optimization problems, we typically care about two objects:

1. The *solution set*

$$x^*(\theta) \equiv \arg \max_{x \in D(\theta)} f(x, \theta),$$

---

*These notes are intended for students in Economics 202, Stanford University. They were originally written by Antonio in Fall 2000, and revised by Jon in Fall 2001. Leo Rezende provided tremendous help on the original notes. Section 5 draws on an excellent comparative statics handout prepared by Ilya Segal.

that gives the solution(s) for any parameter $\theta \in \Theta$. (If the problem has multiple solutions, then $x^*(\theta)$ is a set with multiple elements).

2. The *value function*

$$V(\theta) \equiv \max_{x \in D(\theta)} f(x, \theta)$$

that gives the value of the function at the solution for any parameter $\theta \in \Theta$ ($V(\theta) = f(y, \theta)$ for any $y \in x^*(\theta)$.)

In economic models, several questions typically are of interest:

1. Does a solution to the maximization problem exist for each θ?
2. Do the solution set and the value function change continuously with the parameters? In other words, is it the case that a small change in the parameters of the problem produces only a small change in the solution?
3. How can we compute the solution to the problem?
4. How do the solution set and the value function change with the parameters?

You should keep in mind that any result we derive for a maximization problem also can be used in a minimization problem. This follows from the simple fact that

$$x^*(\theta) = \arg\min_{x \in D(\theta)} f(x, \theta) \Longleftrightarrow x^*(\theta) = \arg\max_{x \in D(\theta)} -f(x, \theta)$$

and

$$V(\theta) = \min_{x \in D(\theta)} f(x, \theta) \Longleftrightarrow V(\theta) = -\max_{x \in D(\theta)} -f(x, \theta).$$

2 Notions of Continuity

Before starting on optimization, we first take a small detour to talk about continuity. The idea of continuity is pretty straightforward: a function h is continuous if “small” changes in x produce “small” changes in $h(x)$. We just need to be careful about (a) what exactly we mean by “small,” and (b) what happens if h is not a function, but a correspondence.

2.1 Continuity for functions

Consider a function h that maps every element in X to an element in Y, where X is the domain of the function and Y is the range. This is denoted by $h : X \to Y$. We will limit ourselves to functions that map $\mathbb{R}^n$ into $\mathbb{R}^m$, so $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$.

Recall that for any $x, y \in \mathbb{R}^k$,

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

denotes the Euclidean distance between x and y. Using this notion of distance we can formally define continuity, using either of following two equivalent definitions:

Definition 1 *A function $h : X \to Y$ is* ***continuous at*** *x if for every $\varepsilon > 0$ there exists $\delta > 0$ such that $\|x - y\| < \delta$ and $y \in X \Rightarrow \|h(x) - h(y)\| < \varepsilon$.*

Definition 2 *A function $h : X \to Y$ is* ***continuous at*** *x if for every sequence x_n in X converging to x, the sequence $h(x_n)$ converges to $f(x)$.*

You can think about these two definitions as tests that one applies to a function to see if it is continuous. A function is continuous if it passes the continuity test at each point in its domain.

Definition 3 *A function $h : X \to Y$ is* ***continuous*** *if it is continuous at every $x \in X$.*

Figure 1 shows a function that is not continuous. Consider the top picture, and the point x. Take an interval centered around $h(x)$ that has a "radius" ε. If ε is small, each point in the interval will be less than A. To satisfy continuity, we must find a distance δ such that, as long as we stay within a distance δ of x, the function stays within ε of $h(x)$. But we cannot do this. A small movement to the right of x, regardless of how small, takes the function above the point A. Thus, the function fails the continuity test at x and is not continuous.

The bottom figure illustrates the second definition of continuity. To meet this requirement at the point x, it must be the case that for *every* sequence x_n converging to x, the sequence $h(x_n)$ converges to $h(x)$. But consider

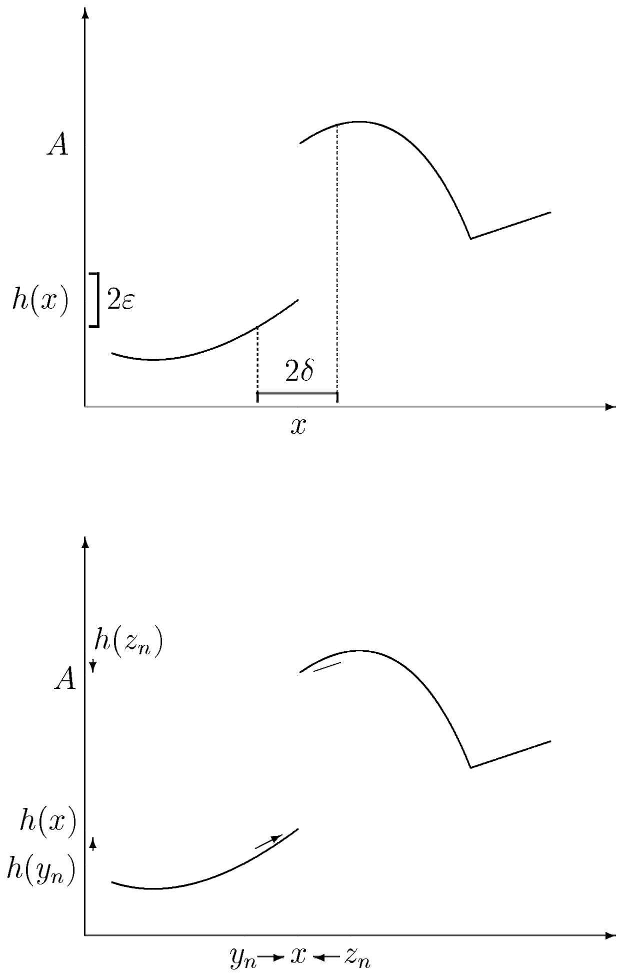

Figure 1: Testing for Continuity.

the sequence z_n that converges to x from the right. The sequence $h(z_n)$ converges to the point A from above. Since $A > h(x)$, the test fails and h is not continuous. We should emphasize that the test must be satisfied *for every sequence*. In this example, the test is satisfied for the sequence y_n that converges to x from the right.

In general, to show that a function is continuous, you need to argue that one of the two continuity tests is satisfied at every point in the domain. If you use the first definition, the typical proof has two steps:

- Step 1: Pick any x in the domain and any $\varepsilon > 0$.
- Step 2: Show that there is a $\delta_x(\varepsilon) > 0$ such that $\|h(x) - h(y)\| < \varepsilon$ whenever $\|x - y\| < \delta_x(\varepsilon)$. To show this you have to give a formula for $\delta_x(\cdot)$ that guarantees this.

The problems at the end should give you some practice at this.

2.2 Continuity for correspondences

A correspondence ϕ maps points x in the domain $X \subseteq \mathbb{R}^n$ into sets in the range $Y \subseteq \mathbb{R}^m$. That is, $\phi(x) \subseteq Y$ for every x. This is denoted by $\phi : X \rightrightarrows Y$. Figure 2 provides a couple of examples. We say that a correspondence is:

- non-empty-valued if $\phi(x)$ is non-empty for all x in the domain.
- convex if $\phi(x)$ is a convex set for all x in the domain.
- compact if $\phi(x)$ is a compact set for all x in the domain.

For the rest of these notes we assume, unless otherwise noted, that correspondences are non-empty-valued.

Intuitively, a correspondence is continuous if small changes in x produce small changes in the set $\phi(x)$. Figure 3 shows a continuous correspondence. A small move from x to x' has a small effect since $\phi(x)$ and $\phi(x')$ are approximately equal. Not only that, the smaller the change in x, the more similar are $\phi(x)$ and $\phi(x')$.

Unfortunately, giving a formal definition of continuity for correspondences is not so simple. With functions, it's pretty clear that to evaluate the effect of moving from x to a nearby x' we simply need to check the distance between the point $h(x)$ and $h(x')$. With correspondences, we need to make a

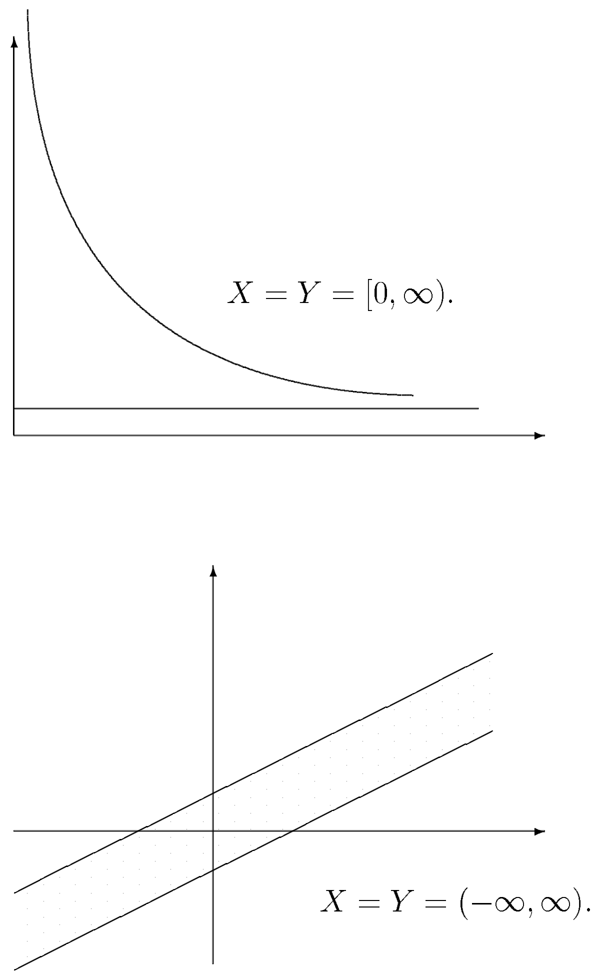

Figure 2: Examples of Correspondences.

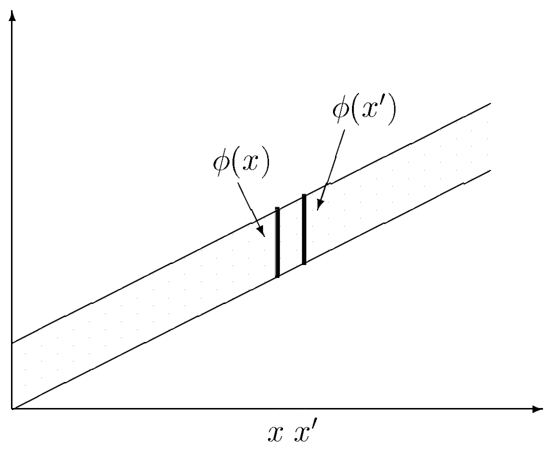

Figure 3: A Continuous Correspondence.

comparison between the sets $\phi(x)$ and $\phi(x')$. To do this, we will need two distinct concepts: upper and lower semi-continuity.

Definition 4 *A correspondence $\phi : X \rightrightarrows Y$ is **lower semi-continuous (lsc) at x** if for each open set G meeting $\phi(x)$, there is an open set $U(x)$ containing x such that if $x' \in U(x)$, then $\phi(x') \cap G \neq \emptyset$.*[1] *A correspondence is **lower semi-continuous** if it is lsc at every $x \in X$.*

Lower semi-continuity captures the idea that any element in $\phi(x)$ can be "approached" from all directions. That is, if we consider some x and some $y \in \phi(x)$, lower semi-continuity at x implies that if one moves a little way from x to x', there will be some $y' \in \phi(x')$ that is close to y.

As an example, consider the correspondence in Figure 4. It is not lsc at x. To see why, consider the point $y \in \phi(x)$, and let G be a very small interval around y that does not include $\hat{y}$. If we take any open set $U(x)$ containing x, then it will contain some point x' to the left of x. But then $\phi(x') = \{\hat{y}\}$ will contain no points near y (i.e. will not intersect G).

On the other hand, the correspondence in Figure 5 *is* lsc. (One of the exercises at the end is to verify this.) But it still doesn't seem to reflect an intuitive notion of continuity. Our next definition formalizes what is wrong.

[1] Recall that a set $S \subset \mathbb{R}^n$ is open if for every point $s \in S$, there is some ε such that every point s' within ε of s is also in S.

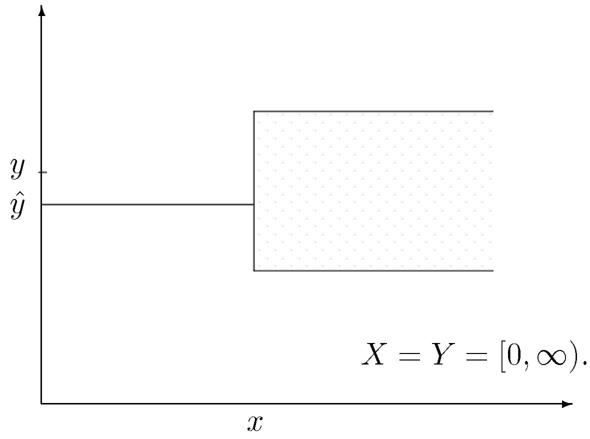

Figure 4: A Correspondence that is usc, but not lsc.

Definition 5 *A correspondence $\phi : X \rightrightarrows Y$ is* ***upper semi-continuous (usc) at*** *x if for each open set G containing $\phi(x)$, there is an open set $U(x)$ containing x such that if $x' \in U(x)$, then $\phi(x') \subset G$. A correspondence is* ***upper semi-continuous*** *if it is usc at every $x \in X$, and also compact-valued.*

Upper semi-continuity captures the idea that $\phi(x)$ will not "suddenly contain new points" just as we move past some point x. That is, if one starts at a point x and moves a little way to x', upper semi-continuity at x implies that there will be no point in $\phi(x')$ that is not close to some point in $\phi(x)$.

As an example, the correspondence in Figure 4 *is* usc at x. To see why, imagine an open interval $\phi(x)$ that encompasses $\phi(x)$. Now consider moving a little to the left of x to a point x'. Clearly $\phi(x') = \{\hat{y}\}$ is in the interval. Similarly, if we move to a point x' a little to the right of x, then $\phi(x')$ will be inside the interval so long as x' is sufficiently close to x.

On the other hand, the correspondence in Figure 5 is not usc at x. If we start at x (noting that $\phi(x) = \{\hat{y}\}$), and make a small move to the right to a point x', then $\phi(x')$ suddenly contains many points that are not close to $\hat{y}$. So this correspondence fails to be upper semi-continuous.

We now combine upper and lower semi-continuity to give a definition of continuity for correspondences.

Definition 6 *A correspondence $\phi : X \rightrightarrows Y$ is* ***continuous at x*** *if it is usc and lsc at x. A correspondence is* ***continuous*** *if it is both upper and lower*

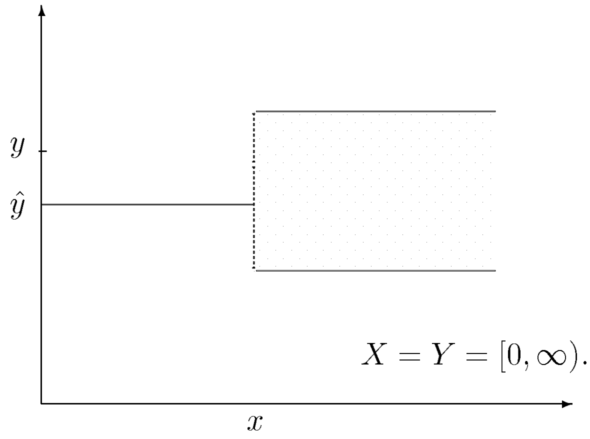

Figure 5: A Correspondence that is lsc, but not usc.

semi-continuous.

As it turns out, our notions of upper and lower semi-continuity both reduce to the standard notion of continuity if ϕ is a single-valued correspondence, i.e. a function.

Proposition 1 *Let $\phi : X \rightrightarrows Y$ be a single-valued correspondence, and $h : X \to Y$ the function given by $\phi(x) = \{h(x)\}$. Then,*

$$\phi \text{ is continuous} \Leftrightarrow \phi \text{ is usc} \Leftrightarrow \phi \text{ is lsc} \Leftrightarrow h \text{ is continuous.}$$

You can get some intuition for this result by drawing a few pictures of functions and checking the definitions of usc and lsc.

3 Properties of Solutions: Existence, Continuity, Uniqueness and Convexity

Let's go back to our class of parametric constrained optimization problems (PCOP). We're ready to tackle the first two questions we posed: (1) under what conditions does the maximization problem has a solution for every parameter θ?, and (2) what are the continuity properties of the value function $V(\theta)$ and the solution set $x^*(\theta)$?

The following Theorem, called the *Theorem of the Maximum*, provides an answer to both questions. We will use it time and time again.

Theorem 2 (Theorem of the Maximum) *Consider the class of parametric constrained optimization problems*

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

defined over the set of parameters Θ*. Suppose that (i)* $D : \Theta \rightrightarrows X$ *is continuous (i.e. lsc and usc) and compact-valued, and (ii)* $f : X \times \Theta \to \mathbb{R}$ *is a continuous function. Then*

1. $x^*(\theta)$ *is non-empty for every* θ*;*
2. x^* *is upper semi-continuous (and thus continuous if* x^* *is single-valued);*
3. V *is continuous.*

We will not give a formal proof, but rather some examples to illustrate the role of the assumptions.

Example What can happen if D is not compact? In this case a solution might not exist for some parameters. Consider the example $\Theta = [0, 10]$, $D(\theta) = (0, 1)$, and $f(x, \theta) = x$. Then $x^*(\theta) = \emptyset$ for all θ.

Example What can happen if D is lsc, but not usc? Suppose that $\Theta = [0, 10]$, $f(x, \theta) = x$, and

$$D(\theta) = \begin{cases} \{0\} & \text{if } \theta \leq 5, \\ [-1, 1] & \text{otherwise} \end{cases}.$$

The solution set is given by

$$x^*(\theta) = \begin{cases} \{0\} & \text{if } \theta \leq 5, \\ \{1\} & \text{otherwise} \end{cases},$$

which is a function, but not continuous. The value function is also discontinuous.

Example What can happen if D is usc, but not lsc? Suppose that $\Theta = [0, 10]$, $f(x, \theta) = x$, and

$$D(\theta) = \begin{cases} \{0\} & \text{if } \theta < 5, \\ [-1, 1] & \text{otherwise} \end{cases}$$

The solution set is given by

$$x^*(\theta) = \begin{cases} \{0\} & \text{if } \theta < 5, \\ \{1\} & \text{otherwise} \end{cases}$$

which once more is a discontinuous function.

In the last two examples, the goal of the maximization problem is to pick the largest possible element in the constraint set. So the solution set potentially can be discontinuous if (and only if) the constraint set changes abruptly.

Example Finally, what can happen if f is not continuous? Suppose that $\Theta = [0, 10]$, $D(\theta) = [\theta, \theta + 1]$, and

$$f(x, \theta) = \begin{cases} 0 & \text{if } x < 5 \\ 1 & \text{otherwise} \end{cases}.$$

The solution set is given by

$$x^*(\theta) = \begin{cases} [\theta, \theta + 1] & \text{if } \theta < 4 \\ [5, \theta + 1] & \text{if } 4 \leq \theta < 5 \\ [\theta, \theta + 1] & \text{otherwise} \end{cases}$$

and the value function is given by

$$V(\theta) = \begin{cases} 0 & \text{if } \theta < 4 \\ 1 & \text{if } 4 \leq \theta < 5 \\ 1 & \text{otherwise} \end{cases}$$

As it can be easily seen from Figure 3, x^* is not usc and V is not continuous at $\theta = 4$.

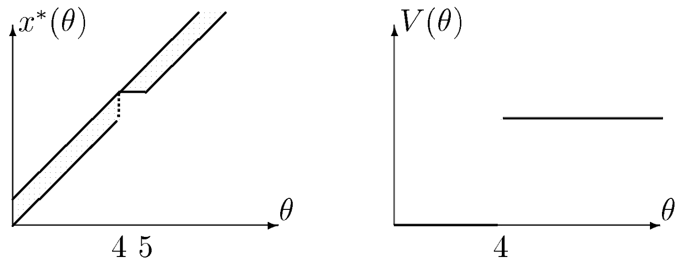

Figure 6: Failure of the Maximum Theorem when f is discontinuous.

You might wonder if the second result in the theorem can be strengthened to guarantee that x^* is continuous, and not just usc. In general, the answer is no, as the following example illustrates.

Example Consider the problem of a consumer with linear utility choosing between two goods. Her objective function is $U(x) = x_1 + x_2$, and her optimization problem is:

$$\begin{array}{c} \max_x U(x) \\ \text{s.t. } p \cdot x \leq 10. \end{array}$$

Let $P = \{ p \in \mathbb{R}^2_{++} \mid p_1 = 1 \text{ and } p_2 \in (0, \infty) \}$ denote the set of possible prices (the parameter set).

The solution (the consumer's demand as a function of price), is given by

$$x^*(p) = (x_1^*(p), x_2^*(p)) = \left\{ \begin{array}{cl} \{(0, 10/p_2)\} & \text{if } p_2 < 1, \\ \{ x \mid p \cdot x = 10 \} & \text{if } p_2 = 1, \\ \{(10, 0)\} & \text{otherwise} \end{array} \right. .$$

Figure 7 graphs x_1^* as a function of p_2. Demand is not continuous, since it explodes at $p_2 = 1$. However, as the theorem states, it is usc.

The Theorem of the Maximum identifies conditions under which optimization problems have a solution, and says something about the continuity of the solution. We can say more if we know something about the curvature of the objective function. This next theorem is the basis for many results in price theory.

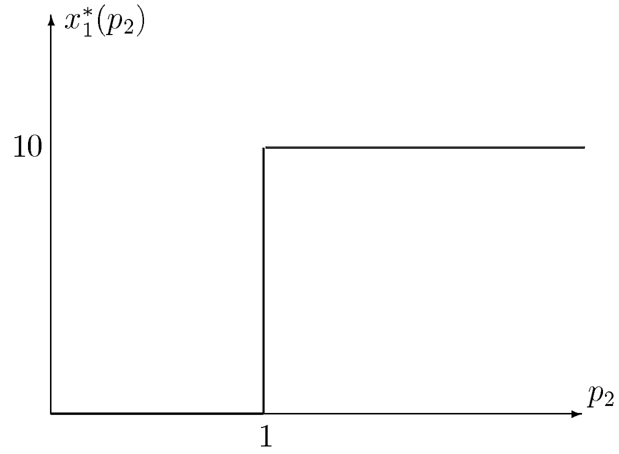

Figure 7: Demand is not Continuous.

Theorem 3 *Consider the class of parametric constrained optimization problems*

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

defined over the convex set of parameters Θ. Suppose that (i) $D : \Theta \rightrightarrows X$ is continuous and compact-valued, and (ii) $f : X \times \Theta \to \mathbb{R}$ is a continuous function.

1. *If $f(\cdot, \theta)$ is a quasi-concave function in x for each θ, and D is convex-valued, then x^* is convex-valued.*[2]

2. *If $f(\cdot, \theta)$ is a strictly quasi-concave function in x for each θ, and D is convex-valued, then $x^*(\theta)$ is single-valued.*

3. *If f is a concave function in (x, θ) and D is convex-valued, then V is a concave function and x^* is convex-valued.*

4. *If f is a strictly concave function in (x, θ) and D is convex-valued, then V is strictly concave and x^* is a function.*

[2] Recall that a function $h : X \to \mathbb{R}$ is quasi-concave if its upper contour sets $\{x \in X : h(x) \geq k\}$ are convex sets. That is, if $h(x) \geq k$ and $h(x') \geq k$ implies that $h(tx+(1-t)x') \geq k$ for all $x \neq x'$ and $t \in (0, 1)$. We say that h is strictly quasi-concave if the last inequality is strict.

Proof. (1) Suppose that $f(\cdot,\theta)$ is a quasi-concave function in x for each θ, and D is convex-valued. Pick any $x, x' \in x^*(\theta)$. Since D is convex-valued, $x^t = tx + (1-t)x' \in D(\theta)$ for all $t \in [0,1]$. Also, by the quasi-concavity of f we have that

$$f(x^t,\theta) \geq f(x,\theta) = f(x',\theta).$$

But since $f(x,\theta) = f(x',\theta) \geq f(y,\theta)$ for all $y \in D(\theta)$, we get that $f(x^t,\theta) \geq f(y,\theta)$ for all $y \in D(\theta)$. We conclude that $x^t \in x^*(\theta)$, which establishes the convexity of x^*.

(2) Suppose that $f(\cdot,\theta)$ is a strictly quasi-concave function in x for each θ, and D is convex-valued. Suppose towards, a contradiction, that $x^*(\theta)$ contains two distinct points x and x' ; i.e., it is not single-valued at θ. As before, D is convex-valued implies that $x^t = tx + (1-t)x' \in D(\theta)$ for all $t \in (0,1)$. But then strict quasi-concavity of $f(\cdot,\theta)$ in x implies that

$$f(x^t,\theta) > f(x,\theta) = f(x',\theta),$$

which contradicts the fact that x and x' are maximizers in $D(\theta)$.

(3) Suppose that f is a concave function in (x,θ) and that D has a convex graph. Pick any θ, θ' in Θ and let $\theta^t = t\theta + (1-t)\theta'$ for some $t \in [0,1]$. We need to show that $V(\theta^t) \geq tV(\theta) + (1-t)V(\theta')$. Let x and x' be solutions to the problems θ and θ' and define $x^t = tx + (1-t)x'$. By the definition of the value function and the concavity of f we get that

$$\begin{aligned}
V(\theta^t) &\geq f(x^t,\theta^t) \qquad \text{(by definition of } V(\theta^t)\text{)} \\
&= f(tx + (1-t)x', t\theta + (1-t)\theta') \qquad \text{(by definition of } x^t, \theta^t\text{)} \\
&\geq tf(x,\theta) + (1-t)f(x',\theta') \qquad \text{(by concavity of } f\text{)} \\
&= tV(\theta) + (1-t)V(\theta'). \qquad \text{(by definition of } V(\theta) \text{ and } V(\theta')\text{)}.
\end{aligned}$$

Also, since $f(\cdot,\cdot)$ concave in $(x,\theta) \Rightarrow f(\cdot,\theta)$ quasi-concave function in x for each θ, the proof that x^* is convex-valued follows from (1).

(4) The proof is nearly identical to (3). *Q.E.D.*

4 Characterization of Solutions

Our next goal is to learn how to actually solve optimization problems. We will focus on a more restricted class of problems given by:

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

subject to

$$g_k(x,\theta) \leq b_k \qquad \text{for } k = 1, \ldots, K,$$

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

In this class of problems, the constraint set $D(\theta)$ is given by the intersection of K inequality constraints. There may be any number of constraints (for instance, there could be more than n or more constraints than choice variable), but it is important that at least some values of x satisfy the constraints (so the choice set is non-empty). Figure 8 shows an example with 4 constraints, and $n = 2$ (i.e. $x = (x_1, x_2)$ and the axes are x_1 and x_2).

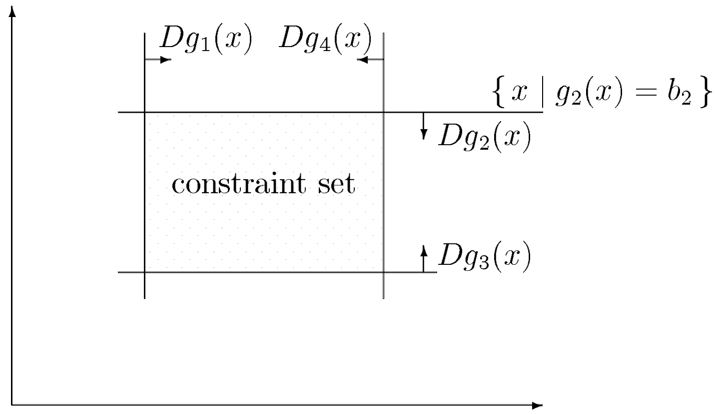

Figure 8: A Constraint Set Given by 4 Inequalities.

Note that this class of problems includes the cases of equality constraints ($g_k(x,\theta) = b_k$ is equivalent to $g_k(x,\theta) \leq b_k$ and $g_k(x,\theta) \geq b_k$) and non-negativity constraints ($x \geq 0$ is equivalent to $-x \leq 0$). Both of these sorts of constraints arise frequently in economics.

Each parameter θ defines a separate constrained optimization problem. In what follows, we focus on how to solve one of these problems (i.e. how to solve for $x^*(\theta)$ for a given θ).

We first need a small amount of notation. Let $h : \mathbb{R}^n \to \mathbb{R}$ be any differentiable real valued function. The derivative, or gradient, of h, $Dh : \mathbb{R}^n \to \mathbb{R}^n$, is a vector-valued function

$$Dh(x) = \left(\frac{\partial h(x)}{\partial x_1}, \ldots, \frac{\partial h(x)}{\partial x_n} \right).$$

As illustrated in Figure 9, the gradient has a nice graphical interpretation: it is a vector that is orthogonal to the level set of the function, and thus points in the direction of maximum increase. (In the figure, the domain of h is the plane (x_1, x_2); the curve is a level curve of h. Movements in the direction L leave the value of the function unchanged, while movements in the direction of $Dh(x)$ increase the value of the function at the fastest possible rate).

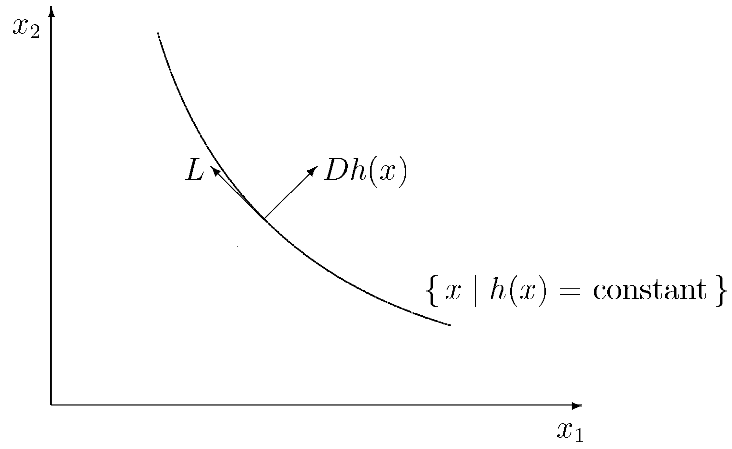

Figure 9: Illustration of Gradient.

If you have taken economics before, you may have learned to solve constrained optimization problems by forming a Lagrangian:

$$\mathcal{L}(x, \lambda, \theta) = f(x, \theta) + \sum_{k=1}^{K} \lambda_k \left(b_k - g_k(x, \theta) \right),$$

and maximizing the Lagrangian with respect to (x, λ). Here, we go a bit deeper to show exactly when and why the Lagrangian approach works.

To do this, we take two steps. First, we identify *necessary conditions* for a solution, or conditions that must be satisfied by any solution to a constrained optimization problem. We then go on to identify *sufficient conditions* for a solution, or conditions that if satisfied by some x guarantee that x is indeed a solution.

Our first result, the famed *Kuhn-Tucker Theorem*, identifies necessary conditions that must be satisfied by any solution to a constrained optimization problem. To state it, we need the following definition.

Definition 7 *Consider a point x that satisfies all of the constraints, i.e. $g_k(x, \theta) \leq b_k$ for all k. We say* ***constraint*** k ***binds at*** x *if $g_k(x, \theta) = b_k$, and is* ***slack*** *if $g_k(x, \theta) < b_k$. If $B(x)$ denotes the set of binding constraints at point x, then* ***constraint qualification holds at*** x *if the vectors in the set $\{Dg_k(x, \theta) | k \in B(x)\}$ are linearly independent.*

Theorem 4 (Kuhn-Tucker) *Suppose that for the parameter θ, the following conditions hold: (i) $f(\cdot, \theta)$, $g_1(\cdot, \theta)$, ..., $g_K(\cdot, \theta)$ are continuously differentiable in x; (ii) $D(\theta)$ is non-empty; (iii) x^* is a solution to the optimization problem; and (iv) constraint qualification holds at x^*. Then*

1. *There exist non-negative numbers λ_1, ..., λ_K such that*

$$Df(x^*, \theta) = \sum_{k=1}^{K} \lambda_k Dg_k(x^*, \theta),$$

2. *For $k = 1, \ldots, K$,*

$$\lambda_k(b_k - g_k(x^*, \theta)) = 0.$$

In practice, we often refer to the expression in (1) as the first-order condition, and (2) as the complementary slackness conditions. Conditions (1) and (2) together are referred to as the *Kuhn-Tucker conditions*. The numbers λ_k are *Lagrange multipliers*. The theorem tells us these multipliers must be non-negative, and equal to zero for any constraint that is not binding (binding constraints typically have positive multipliers, but not always).

The Kuhn-Tucker theorem provides a partial recipe for solving optimization problems: simply compute all the pairs (x, λ) such that *(i)* x satisfies all of the constraints, and *(ii)* (x, λ) satisfies the Kuhn-Tucker conditions. Or in other words, one identifies all the solutions (x, λ) for the system:

$$g_1(x, \theta) \leq b_1,$$
$$\vdots$$
$$g_K(x, \theta) \leq b_K,$$

$$Df(x, \theta) = \sum_{k=1}^{K} \lambda_k Dg_k(x, \theta),$$

and

$$\lambda_1(b_1 - Dg_1(x,\theta)) = 0,$$
$$\vdots$$
$$\lambda_K(b_K - Dg_K(x,\theta)) = 0.$$

Since there are $n + 2K$ equations and only $n + K$ unknowns, you might be concerned that this system has no solution. Fortunately the Maximum Theorem comes to the rescue. If its conditions are satisfied, we know a maximizer exists and so must solve our system of equations.[3]

What is the status of the Lagrangian approach you might have learned previously? If you applied it correctly (remembering that the λ_k's must be non-negative), you would have come up with exactly the system of equations above as your first-order conditions. So the Kuhn-Tucker gives us a short-cut: we can apply it directly without even writing down the Lagrangian!

Because the Kuhn-Tucker Theorem is widely used in economic problems, we include a proof (intended only for the more ambitious).

Proof (of the Kuhn-Tucker Theorem).[4] The proof will use the following basic result from linear algebra: If $\alpha_1, \ldots, \alpha_j$ are linearly independent vectors in $\mathbb{R}^n$, with $j \leq n$, then for any $b = (b_1, \ldots, b_j) \in \mathbb{R}^j$ there exists $x \in \mathbb{R}^n$ such that

$$Ax = b,$$

where A is the $j \times n$ matrix that given by

$$\begin{pmatrix} - & \alpha_1 & - \\ & \cdots & \\ - & \alpha_j & - \end{pmatrix}.$$

To prove the KT Theorem we need to show that the conditions of the theorem imply that if x^* solves the optimization problem, then: (1) there are numbers $\lambda_1, \ldots, \lambda_K$ such that

$$Df(x^*,\theta) = \sum_{k=1}^{K} \lambda_k Dg_k(x^*,\theta),$$

[3]Also note that if the constraint qualification holds at x^*, then there are at most n binding constraints. It is easy to see why. The vectors $Dg_k(x^*,\theta)$ are vectors in $\mathbb{R}^n$ and we know from basic linear algebra that at most n vectors in $\mathbb{R}^n$ can be linearly independent.

[4]This proof is taken from a manuscript by Kreps, who attributes it to Elchanan Ben-Porath.

(2) the numbers λ_1, ..., λ_K are non-negative, and (3) the complementary slackness conditions are satisfied at x^*.

In fact, we can prove (1) and (3) simultaneously by establishing that there are numbers λ_k for each $k \in B(x^*)$ such that

$$Df(x^*, \theta) = \sum_{k \in B(x^*)} \lambda_k Dg_k(x^*, \theta).$$

Then (3) follows because the slackness conditions are automatically satisfied for the binding constraints, and (1) follows because we can set $\lambda_k = 0$ for the non-binding constraints.

So let's prove this first. We know that the vectors in the set $\{Dg_k(x^*, \theta) | k \in B(x^*)\}$ are linearly independent since the constraint qualification is satisfied at x^*. Now suppose, towards a contradiction, that there are no numbers λ_1, ..., λ_K for which the expression holds. Then we know that the set of vectors $\{Df(x^*, \theta)\} \cup \{Dg_k(x^*, \theta) | k \in B(x^*)\}$ is linearly independent. In turn, this implies that there are at most $n-1$ binding constraints. (Recall that there can be at most n linearly independent vectors in $\mathbb{R}^n$.) But then, by our linear algebra result, we get that there exists $z \in \mathbb{R}^n$ such that

$$Df(x^*, \theta) \cdot z = 1 \qquad \text{and} \qquad Dg_k(x^*, \theta) \cdot z = -1 \quad \text{for all } k \in B(x^*).$$

Now consider the effect of moving from x^* to $x^* + \epsilon z$, for $\epsilon > 0$ but small enough. By Taylor's theorem for ϵ small enough all of the constraints are satisfied at $x^* + \epsilon z$. The constraints that are slack are no problem, and for the ones that are binding the change makes them slack. Also, by Taylor's theorem, $f(x^* + \epsilon z, \theta) > f(x^*, \theta)$, a contradiction to the optimality of x^*.

Now look at (2). Suppose that the first order condition holds, but that one of the multipliers is negative, say λ_j. Pick $M > 0$ such that $-M\lambda_j > \sum_{k \in B(x^*), k \neq j} \lambda_k$. Again, since the set of vectors $\{Dg_k(x^*, \theta) | k \in B(x^*)\}$ is linearly independent, we know that there exists $z \in \mathbb{R}^n$ such that

$$Dg_j(x^*, \theta) \cdot z = -M \qquad \text{and} \qquad Dg_k(x^*, \theta) \cdot z = -1 \quad \text{for all } k \in B(x^*), k \neq j.$$

By the same argument than before, for ϵ small enough all of the constraints are satisfied at $x^* + \epsilon z$. Furthermore, by Taylor's Theorem

$$\begin{aligned} f(x^* + \epsilon z, \theta) &= f(x^*, \theta) + \epsilon Df(x^*, \theta) \cdot z + o(\epsilon) \\ &= f(x^*, \theta) + \epsilon \sum_{k \in B(x^*)} \lambda_k Dg_k(x^*, \theta) \cdot z + o(\epsilon) \end{aligned}$$

$$
\begin{aligned}
&= f(x^*, \theta) + \epsilon \left[-M\lambda_j - \sum_{k \in B(x^*), k \neq j} \lambda_k \right] + o(\epsilon) \\
&> f(x^*, \theta);
\end{aligned}
$$

where the last inequality holds as long as ϵ is small enough. Again, this contradicts the optimality of x^*. *Q.E.D.*

Figure 10 provides a graphical illustration of the Kuhn-Tucker Theorem for the case of one constraint. In the picture, the two outward curves are level curves of $f(\cdot, \theta)$, while the inward curve is a constraint (any eligible x must be inside it). Consider the point $\hat{x}$ that is not optimal, but at which the constraint is binding. Here the gradient of the objective function cannot be expressed as a linear combination (i.e., a multiple given that there is only one constraint) of the gradient of the constraint. Now consider a small movement towards point A in a direction perpendicular to $Dg_1(\widehat{x})$ which is feasible since it leaves the value of the constraint unchanged. Since this movement is not perpendicular to $Df(\widehat{x})$ it will increase the value of the function. By contrast, consider a similar movement at x^*, where $Dg_1(x^*)$ and $Df(x^*)$ lie in the same direction. Here any feasible movement along the constraint also leaves the objective function unchanged.

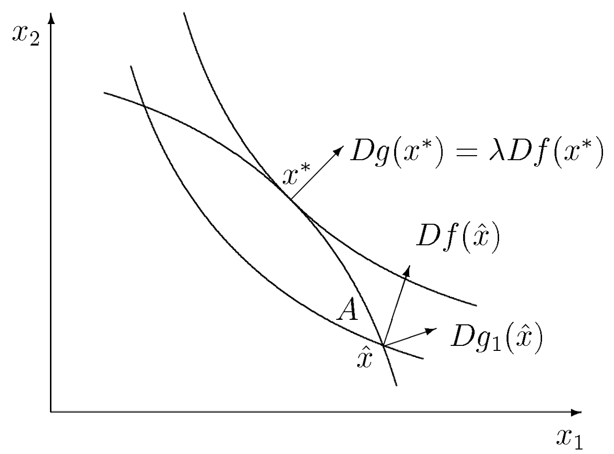

Figure 10: Illustration of Kuhn-Tucker Theorem.

It is easy to see why the multiplier of a binding constraint has to be non-negative. Consider the case of one constraint illustrated in Figure 11.

We cannot have a negative multiplier because then a movement towards the interior of the constraint set that makes the constraint slack would increases the value of the function — a violation of the optimality of x^*.

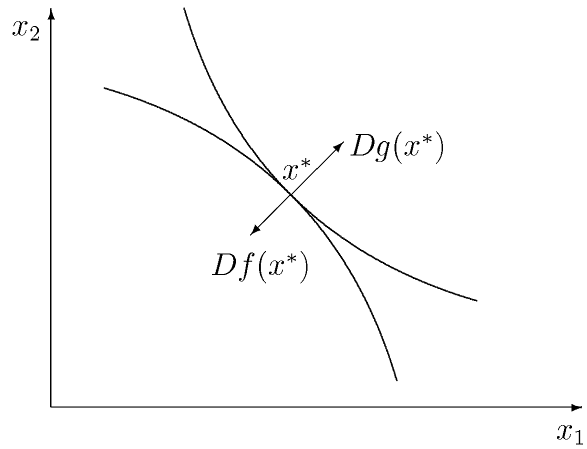

Figure 11: Why λ is Non-Negative.

It is important to emphasize that if the constraint qualification is not satisfied, the Kuhn-Tucker recipe might fail. Consider the example illustrated in Figure 12. Here x^* is clearly a solution to the problem since it is the only point in the constraint set. And yet, we cannot write $Df(x^*)$ as a linear combination of $Dg_1(x^*)$ and $Dg_2(x^*)$.

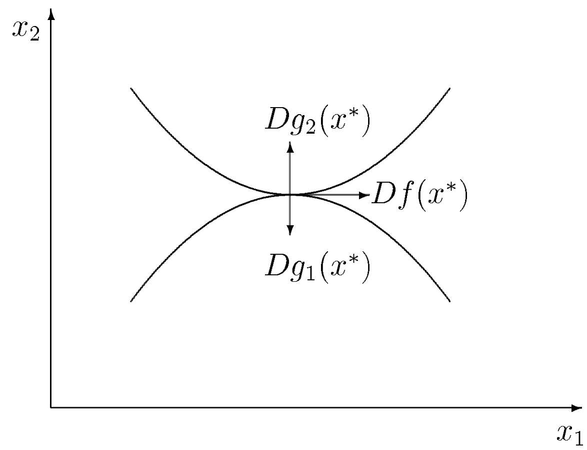

Figure 12: Why Constraint Qualification is Required.

Fortunately, in many cases it will be easy to verify that constraint qualification is satisfied. For example, we may know that at the solution only one constraint is binding. Or we may have only linear constraints with linearly independent gradients. It is probable that in most problems you encounter, establishing constraint qualification will not be an issue.

Why does the Kuhn-Tucker Theorem provide only a partial recipe for solving constrained optimization problems? The reason is that there may be solutions (x, λ) to the Kuhn-Tucker conditions that are not solutions to the optimization problem (i.e. the Kuhn-Tucker conditions are necessary but not sufficient for x to be a solution). Figure 13 provides an example in which x satisfies the Kuhn-Tucker conditions but is not a solution to the problem, the points x^* and x^{**} are the solution.

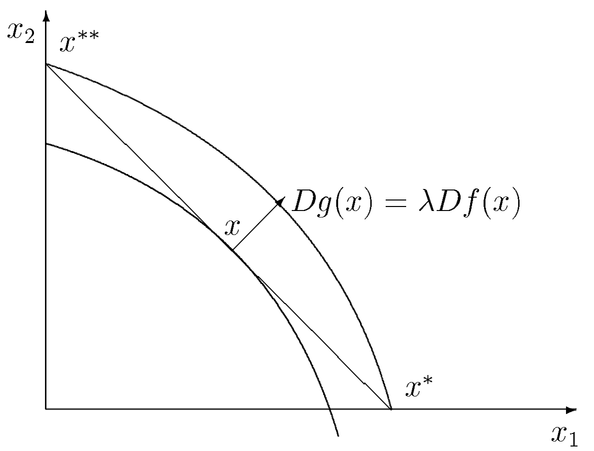

Figure 13: KT conditions are not sufficient.

To rule out this sort of situation, one needs to check second-order, or sufficient conditions. In general, checking second-order conditions is a pain. One needs to calculate the Hessian matrix of second derivatives and test for negative semi-definiteness, a rather involved procedure. (The grim details are in Mas-Colell, Whinston and Green (1995).) Fortunately, for many economic problems the following result comes to the rescue.

Theorem 5 *Suppose that the conditions of the Kuhn-Tucker Theorem are satisfied and that (i)* $f(\cdot, \theta)$ *is quasi-concave, and (ii)* $g_1(\cdot, \theta), \ldots, g_K(\cdot, \theta)$ *are quasi-convex.*[5] *Then any point* x^* *that satisfies the Kuhn-Tucker conditions*

[5]Recall that quasi-concavity of $f(\cdot, \theta)$ means that the upper countour sets of f (i.e. the

is a solution to the constraint optimization problem.

Proof. We will use the following fact about quasi-convex functions. A continuously differentiable function $g : \mathbb{R}^n \to \mathbb{R}$ is quasi-convex if and only if

$$Dg(x) \cdot (x - x') \leq 0 \quad \text{whenever} \quad g(x\prime) \leq g(x).$$

Now to the proof. Suppose that x^* satisfies the Kuhn-Tucker conditions. Then there are multipliers λ_1 to λ_K such that

$$Df(x^*, \theta) = \sum_{k=1}^{K} \lambda_k Dg_k(x^*, \theta).$$

But then, for any x with that satisfies the constraints we get that

$$Df(x^*, \theta)(x - x^*) = \sum_{k=1}^{K} \lambda_k Dg_k(x^*, \theta)(x - x^*) \leq 0.$$

The last inequality follows because $\lambda_k = 0$ if the constraint is slack at at x^*, and $g_k(x, \theta) \leq g_k(x^*, \theta)$ if it binds.

To conclude the proof note that for concave f we know that

$$f(x, \theta) \leq f(x^*, \theta) + Df(x^*, \theta)(x - x^*).$$

Since the second term on the right is non-positive we can conclude that $f(x, \theta) \leq f(x^*, \theta)$, and thus x^* is a solution to the problem. *Q.E.D.*

One reason this theorem is valuable is that its conditions are satisfied by many economic models. When the conditions are met, we can solve the optimization problem in a single step by solving the Kuhn-Tucker conditions. If the conditions fails, you will need to be much more careful in solving the optimization problem.

---

set of points x such that $f(x, \theta) \geq c$) are convex. (See the footnote above). Note also that quasi-convexity of $g_1(\cdot, \theta)$, $\ldots$, $g_K(\cdot, \theta)$ implies that the constraint set $D(\theta)$ is convex — this is useful for applying the characterization theorem from the previous section.

4.1 Non-negativity Constraints

A common variation of the optimization problem above arises when x is required to be non-negative (i.e. $x \geq 0$ or $x \in \mathbb{R}^n_+$). This problem can be written as

$$\max_{x \in \mathbb{R}^n_+} f(x, \theta)$$

subject to

$$g_k(x, \theta) \leq b_k \quad \text{for all} \quad k = 1, \ldots, K.$$

Given that this variant comes up often in economic models, it is useful to explicitly write down the Kuhn-Tucker conditions.

To do this, we write the constraint $x \in \mathbb{R}^n_+$ as n constraints:

$$h_1(x) = -x_1 \leq 0, ..., h_n(x) = -x_n \leq 0,$$

and apply the Kuhn-Tucker Theorem. For completeness, we give a formal statement.

Theorem 6 *Consider the constrained optimization problem with non-negativity constraints. Suppose that for the parameter θ, the following conditions hold: (i) $f(\cdot, \theta)$, $g_1(\cdot, \theta)$, $\ldots$, $g_K(\cdot, \theta)$ are continuously differentiable; (ii) $D(\theta)$ is non-empty; (iii) x^* is a solution to the optimization problem; (iv) constraint qualification holds at x^* for all the constraints (including any binding non-negativity constraints). Then*

1. *There are numbers λ_1, $\ldots$, λ_I such that*

$$\frac{\partial f(x^*, \theta)}{\partial x_j} + \mu_j = \sum_{k=1}^{I} \lambda_k \frac{\partial g_k(x^*, \theta)}{\partial x_j} \quad \textit{for all} \quad j = 1, \ldots, n$$

2. *For $k = 1, \ldots, I$,*

$$\lambda_k(b_k - g_k(x^*, \theta)) = 0.$$

3. *For $j = 1, \ldots, n$,*

$$\mu_j x_j^* = 0.$$

The proof is a straightforward application of the Kuhn-Tucker theorem. Note that the conditions in (3) are simply complementary slackness conditions for the non-negativity constraints. The multiplier μ_j of the j-th non-negativity constraint is zero whenever $x_j^* > 0$.

4.2 Equality Constraints

Another important variant on our optimization problem arises when we have equality, rather than inequality, constraints. Consider the problem:

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

subject to

$$g_k(x, \theta) = b_k \qquad \text{for all } k = 1, \ldots, K.$$

To make sure that the constraint set is non-empty we assume that $K \leq n$.

As discussed before, this looks like a special case of our previous result since $g_k(x, \theta) = b_k$ can be rewritten as $g_k(x, \theta) \leq b_k$ and $-g_k(x, \theta) \leq b_k$. Unfortunately, we cannot use our previous result to solve the problem in this way. At the solution both sets of inequality constraints must be binding, which implies that the constraint qualification cannot be satisfied. (Why?) Thus, our recipe does not work.

The following result, known as the *Lagrange Theorem*, provides the recipe for this case.

Theorem 7 *Consider a constrained optimization problem with $K \leq n$ equality constraints. Suppose that for the parameter θ, the following conditions hold: (i) $f(\cdot, \theta)$, $g_1(\cdot, \theta)$, ..., $g_K(\cdot, \theta)$ are continuously differentiable; (ii) $D(\theta)$ is non-empty; (iii) x^* is a solution to the optimization problem; (iv) the following constraint qualification holds at x^**

$$\mathit{Rank} \begin{pmatrix} - & Dg_1(x^*, \theta) & - \\ & \cdots & \\ - & Dg_I(x^*, \theta) & - \end{pmatrix} = I.$$

Then there are numbers λ_1, ..., λ_I such that

$$Df(x^*, \theta) = \sum_{k=1}^{I} \lambda_k Dg_k(x^*, \theta)$$

This result is very similar, but not identical. First, there are no complementary slackness conditions since all of the constraints are binding. Second, the multipliers can be positive or negative. The proof in this case is more complicated and is omitted.

4.3 Simplifying Constrained Optimization Problems

In many problems that you will encounter you will know which constraints are binding and which ones are not. For example, in the problem

$$\max_{x \in \mathbb{R}^n} f(x, \theta) \qquad \text{subject to} \quad g_k(x, \theta) \leq b_k \quad \text{for all } k = 1, \ldots, K$$

you may know that constraints 1 to $K - 1$ are binding, but constraint K is slack. You might then ask if it is possible to solve the original problem by solving the simpler problem

$$\max_{x \in \mathbb{R}^n} f(x, \theta) \qquad \text{subject to} \quad g_k(x, \theta) = b_k \quad \text{for all } k = 1, \ldots, K - 1.$$

This transformation is very desirable because the second problem (which has equality constraints, and fewer equations) is typically much easier to solve.

Unfortunately, this simplification is not always valid, as Figure 14 demonstrates. At the solution x^* one of the constraints is not binding. However, if we eliminate that constraint we can do better by choosing $\hat{x}$. So elimination of the constraint changes the value of the problem.

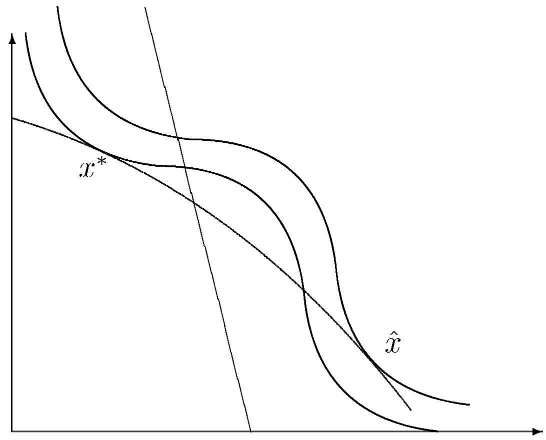

Figure 14: An indispensable non-binding constraint.

Fortunately, however, the simplification is valid in a particular set of cases that includes many economic models.

Theorem 8 *Consider the maximization problem*

$$\max_{x \in \mathbb{R}^n} f(x, \theta) \qquad \text{subject to} \quad g_k(x, \theta) \leq b_k \quad \text{for all } k = 1, \ldots, I.$$

Suppose that (i) $f(\cdot, \theta)$ is strictly quasi-concave; (ii) $g_1(\cdot, \theta)$, ..., $g_K(\cdot, \theta)$ are quasi-convex; (iii) $g_k(\cdot, \theta)$ for $k = 1, \ldots, B$ are binding constraints at the solution, and (iv) $g_k(\cdot, \theta)$ for $k = B + 1, \ldots, K$ are slack constraints at the solution. Then x^* is a solution if and only if is a solution to the modified problem

$$\max_{x \in \mathbb{R}^n} f(x, \theta) \qquad \text{subject to } g_k(x, \theta) = b_k \qquad \text{for all } k = 1, \ldots, B.$$

Proof. Conditions *(i)* and *(ii)* imply that the optimization problem has a unique solution. Call it x^*. Suppose, towards a contradiction, that there is a point $\hat{x}$ that satisfies the constraints of the second problem and for which $f(\hat{x}, \theta) > f(x^*, \theta)$. Then because $f(\cdot, \theta)$ is strictly quasi-concave, $f(x^t, \theta) > f(x^*, \theta)$ for all $x^t = t\widehat{x} + (1 - t)x^*$ and $t \in (0, 1)$. Furthermore, by strict quasi-convexity of the constraints, x^t satisfies all of the constraints of hte first problem for t close enough to zero. But then x^* cannot be a solution to the first problem, a contradiction. *Q.E.D.*

This theorem will allow us to transform most of our problems into the simpler case of equality constraints. For this reason, the rest of the notes will focus on this case.

5 Comparative Statics

Many economic questions can be phrased in the following way: how do endogenous variables respond to changes in exogenous variables? These types of questions are called *comparative statics*.

Comparative statics questions can be qualitative or quantitative. We may just want to know if $x^*(\cdot)$ or $V(\cdot)$ increase, decrease, or are unaffected by θ. Alternatively, we may care about exactly how much or how quickly $x^*(\cdot)$ and $V(\cdot)$ will change with θ.

These questions might seem straightforward. After all, if we know the formulas for $x^*(\cdot)$ or $V(\cdot)$, all we have to do is compute the derivative, or check if the function is increasing. However, things often are not so simple. It may be hard to find an explicit formula for $x^*(\cdot)$ or $V(\cdot)$. Or we may want to know if $x^*(\cdot)$ is increasing as long as the objective function $f(\cdot)$ satisfies a particular property.

This section discusses the three important tools that are used for comparative statics analysis: (1) the Implicit Function Theorem, (2) the Envelope Theorem, and (3) Topkis's Theorem on monotone comparative statics.

5.1 The Implicit Function Theorem

The *Implicit Function Theorem* is an invaluable tool. It gives sufficient conditions under which, at least locally, there is a differentiable solution function to our optimization problem $x^*(\theta)$. And not only that, it gives us a formula to compute the derivative without actually solving the optimization problem!

Given that you will use the IFT in other applications, it is useful to present it in a framework that is more general than our class of PCOP. Consider a system of n equations of the form

$$h_k(x, \theta) = 0 \quad \text{for all } k = 1, \ldots, n$$

where $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $\theta = (\theta_1, \ldots, \theta_s) \in \mathbb{R}^s$. Suppose that the functions are defined on an open set $X \times T \subset \mathbb{R}^n \times \mathbb{R}^s$. We can think of x as the variables and θ as the parameters. To simplify the notation below define $h(x, \theta) = (h_1(x, \theta), \ldots, h_n(x, \theta))$.

Consider a solution $\hat{x}$ of the system at $\hat{\theta}$. We say that the system can be *locally solved at* $(\hat{x},\hat{\theta})$ if for some open set A of parameters that contains $\hat{\theta}$, and for some open set B of variables that contains $\hat{x}$, there exists a *uniquely determined* function $\eta : A \to B$ such that for all $\theta \in A$

$$h_k(\eta(\theta), \theta) = 0 \quad \text{for all } k = 1, \ldots, n.$$

We call η an "implicit" solution of the system. (Note: if $h(\cdot, \cdot)$ come from the first-order conditions of an optimization problem, then $\eta(\theta) = x^*(\theta)$).

Theorem 9 *Consider the system of equations described above and suppose that (i) $h_k(\cdot)$ is continuously differentiable at (x, θ) for all k; (ii) $\hat{x}$ is a solution at $\hat{\theta}$; and (iii) the matrix $D_x h(\hat{x},\hat{\theta})$ is non-singular:*

$$rank \begin{pmatrix} - & D_x h_1(\hat{x}, \hat{\theta}) & - \\ & \ldots & \\ - & D_x h_n(\hat{x}, \hat{\theta}) & - \end{pmatrix}_{n \times n} = n$$

Then:

1. *The system can locally be solved at* $(\hat{x},\hat{\theta})$

2. *The implicit function* η *is continuously differentiable and*

$$D_\theta \eta(\hat{\theta}) = -[D_x h(\hat{x},\hat{\theta})]^{-1} D_\theta h(\hat{x},\hat{\theta}).$$

This result is extremely useful! Not only does it give you sufficient conditions under which there is a unique (local) solution, it also gives you sufficient conditions for the differentiability of the solution and a shortcut for computing the derivative. You don't need to know the solution function, only the solution at $\hat{\theta}$.

If there is only 1 variable and 1 parameter, the formula for the derivative takes the simpler form

$$\frac{\partial \eta(\hat{\theta})}{\partial \theta} = -\frac{h_\theta(\hat{x},\hat{\theta})}{h_x(\hat{x},\hat{\theta})}.$$

A full proof of the implicit function theorem is beyond the scope of these notes. However, to see where it comes from, consider the one-dimensional case and suppose that $\eta(\theta)$ is the unique differentiable solution to $h(x,\theta) = 0$. Then $\eta(\theta)$ is implicity given by

$$h(\eta(\theta),\theta) = 0.$$

Totally differentiating with respect to θ (use the chain rule), we get

$$\frac{\partial h(\eta(\theta),\theta)}{\partial x} \cdot \frac{\partial \eta(\theta)}{\partial \theta} + \frac{\partial h(\eta(\theta),\theta)}{\partial \theta} = 0.$$

Evaluating this at $\theta = \hat{\theta}$ and $\eta(\hat{\theta}) = \widehat{x}$, and re-arranging, gives the formula above.

In optimization problems, the implicit function theorem is useful for computing how a solution will change with the parameters of a problem. Consider the class of constrained optimization problems with equality constraints defined by an open parameter set $\Theta \subset \mathbb{R}$ (so there is a single parameter). Lagrange's Theorem provides an *implicit* characterization of the problem's solution.

Recall that any solution (x^*,λ) must solve the following $n+K$ equations:

$$\frac{\partial f(x^*,\theta)}{\partial x_j} - \sum_{k=1}^{K} \lambda_k \frac{\partial g_k(x^*,\theta)}{\partial x_j} = 0 \quad \text{for all } j = 1,\ldots,n$$

$$g_k(x^*, \theta) - b_k = 0 \quad \text{for all } k = 1, \ldots, K.$$

Let $(x^*(\theta), \lambda(\theta))$ be the solution to this system of equations, and suppose (as we will often do) that $x^*(\theta)$ is a function. The first part of the IFT says the solution $(x^*(\theta), \lambda(\theta))$ is differentiable if when we take the derivative of the left hand side of each equation with respect to x and λ (i.e. we find $D_x h$ and $D_\lambda h$), we get something that is not equal to zero when evaluated at $(x^*(\theta), \lambda(\theta))$. The second part of the IFT provides a formula for the derivative of $\partial x^*(\theta)/\partial\theta$ and $\partial\lambda(\theta)/\partial\theta$. Very useful!

In closing this section, it is worth mentioning a few subtleties of the IFT.

1. The IFT guarantees that the system has a unique *local* solution at $(\hat{x}, \hat{\theta})$, but not necessarily that $\widehat{x}$ is a unique *global* solution given $\hat{\theta}$. As shown in figure 15, the second statement is considerably stronger. At $\hat{\theta}$ the system has 3 solutions x', x'', and x'''. A small change in from $\hat{\theta}$ to $\tilde{\theta}$ changes each solution slightly. Since $D_x h(x, \theta)$ is non-zero at each solution — call them $\eta'(\theta)$, $\eta''(\theta)$ and $\eta'''(\theta)$ — the IFT applies to each individually. It guarantees that each is differentiable and gives the formula for the derivative. However, one needs to be a little careful when there are multiple solutions. In this example,

$$D_\theta \eta'(\theta) \neq D_\theta \eta''(\theta) \neq D_\theta \eta'''(\theta),$$

so each solution reacts differently to changes in the parameter θ.

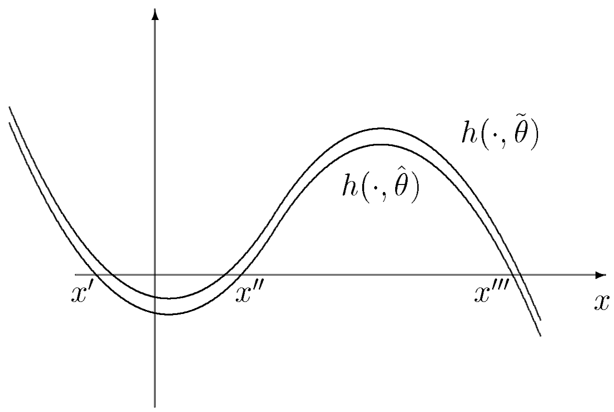

Figure 15: IFT: Solution may not be unique globally.

2. The IFT applies only when $D_x h(x,\theta)$ is non-singular. Figure 16 illustrates why. At $\hat{\theta}$ the system has two solutions: x' and x''. However, if we changes the parameter to any $\theta > \hat{\theta}$, the system suddenly gains an extra solution. By contrast, if we change the parameters to any $\theta < \hat{\theta}$, the system suddenly has only one solution. If we tried to apply the IFT, we would find that $D_x h(x', \hat{\theta}) = 0$, and that the system does not have a unique local solution in a neighborhood around $(x', \hat{\theta})$.

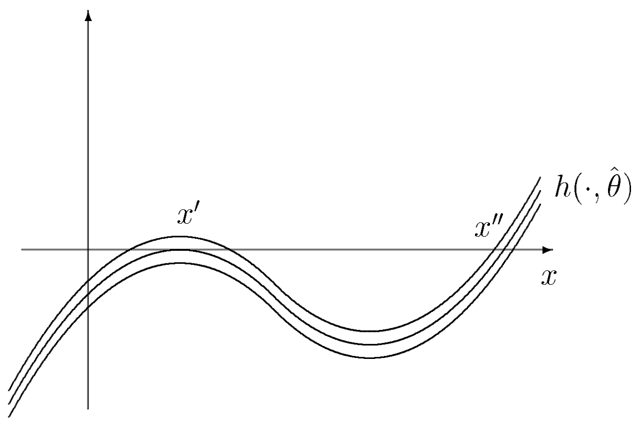

Figure 16: IFT: Solution may not be locally unique where $D_x h$ is singular.

5.2 Envelope Theorems

Whereas the IFT allows us to compute the derivative of $x^*(.)$, the *Envelope Theorem* can be used to compute the derivative of the value function.

Once more, consider the class of constrained optimization problems with equality constraints defined by a parameter set Θ.

Theorem 10 (Envelope Theorem) *Consider the maximization problem*

$$\max_{x \in R^n} f(x, \theta) \text{subject to } g_k(x, \theta) = b_k \text{for all } k = 1, ..., K.^6$$

and suppose that (i) $f(\cdot)$, $g_1(\cdot)$, ..., $g_K(\cdot)$ are continuously differentiable in (x, θ); and (ii) $x^*(.)$ is a differentiable function in an open neighborhood A of $\widehat{\theta}$. Then

1. *$V(\cdot)$ is differentiable in A.*

2. *For $i = 1, ..., s$*

$$\frac{\partial V(\widehat{\theta})}{\partial \theta_i} = \frac{\partial f(x^*(\widehat{\theta}), \widehat{\theta})}{\partial \theta_i} - \sum_{k=1}^{K} \lambda_k \frac{\partial g_k(x^*(\widehat{\theta}), \widehat{\theta})}{\partial \theta_i},$$

where λ_1, ..., λ_I are the Lagrange multipliers associated with $x^*(\widehat{\theta})$.

To grasp the intuition for the ET, think about a simple one-dimensional optimization problem with no constraints:

$$V(\theta) = \max_{x \in \mathbb{R}} f(x, \theta),$$

where $\theta \in [0, 1]$. If the solution $x^*(\theta)$ is differentiable, then $V(\theta) = f(x^*(\theta), \theta)$ is differentiable. Applying the chain rule, we get:

$$V'(\theta) = \underbrace{\frac{\partial f(x^*(\theta), \theta)}{\partial x}}_{=0 \text{ (at an optimum)}} \times \frac{\partial x^*(\theta)}{\partial \theta} + \frac{\partial f(x^*(\theta), \theta)}{\partial \theta} = \frac{\partial f(x^*(\theta), \theta)}{\partial \theta}.$$

A change in θ has two effects on the value function: (i) a direct effect $f_\theta(x^*(\theta), \theta)$, and (ii) an indirect effect $f_x(x^*(\theta), \theta)\frac{\partial x^*(\theta)}{\partial \theta}$. The ET tells us that under certain conditions, we can ignore the indirect effect and focus on the direct effect. In problems with constraints, there is also a third effect — the change in the constraint set. If constraints are binding (some λ's are positive), this effect is accounted for by the ET above.

A nice implication of the ET is that it provides some meaning for the mysterious Lagrange multipliers. To see this, think of b_k as a parameter of the problem, and consider $\partial V / \partial b_k$ — the marginal value of relaxing the kth constraint ($g_k(x, \theta) \leq b_k$). The ET tells us that (think about why this is true):

$$\frac{\partial V(\theta; b)}{\partial b_k} = \lambda_k.$$

Thus, λ_k is precisely the marginal value of relaxing the kth constraint.

One drawback to the ET stated above is that it requires $x^*(\theta)$ to be (at least locally) a differentiable function. In many cases (for instance in many unconstrained problems), this is a much stronger requirement than is necessary. The next result (due to Milgrom and Segal, 2001) provides an alternative ET that seems quite useful.

Theorem 11 *Consider the maximization problem*

$$\max_{x \in R^n} f(x,\theta) \text{subject to } g_k(x,\theta) \leq b_k \text{for all } k = 1, ..., K.$$

and suppose that (i) $f, g_1, ..., g_K$ are continuous and concave in x; (ii) $\partial f/\partial\theta$, $\partial g_1/\partial\theta, ..., \partial g_K/\partial\theta$ are continuous in (x,θ); and (iii) there is some $\widehat{x} \in \mathbb{R}^n$ such that $g_k(\widehat{x},\theta) > 0$ for all $k = 1, ..., K$ and all $\theta \in \Theta$. Then at any point $\hat{\theta}$ where $V(\theta)$ is differentiable:

$$\frac{\partial V(\widehat{\theta})}{\partial \theta_i} = \frac{\partial f(x^*(\widehat{\theta}), \widehat{\theta})}{\partial \theta_i} - \sum_{k=1}^{K} \lambda_k \frac{\partial g_k(x^*(\widehat{\theta}), \widehat{\theta})}{\partial \theta_i},$$

where λ_1, ..., λ_I are the Lagrange multipliers associated associated with $x^*(\widehat{\theta})$.

5.3 Monotone Comparative Statics

The theory of monotone comparative statics (MCS) provides a very general answer to the question: when is the solution set $x^*(\cdot)$ nondecreasing (or strictly increasing) in θ?

You may wonder if the answer is already given by the Implicit Function Theorem. Indeed, when $x^*(\cdot)$ is a differentiable function we can apply the IFT to compute $D_\theta x^*(\theta)$ and then proceed to sign it. But this approach has limitations. First, $x^*(\cdot)$ might not be differentiable, or might not even be a function. Second, even when $x^*(\theta)$ is a differentiable function, it might not be possible to sign $D_\theta x^*(\theta)$.

Consider a concrete example:

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

where $\Theta = \mathbb{R}$, f is twice continuously differentiable, and strictly concave (so that $f_{xx} < 0$.)[10] Under these conditions the problem has a unique solution for every parameter θ that is characterized by the first order condition

$$f_x(x^*(\theta), \theta) = 0.$$

Since f is strictly concave, the conditions of the IFT hold (you can check this) and we get:

$$\frac{\partial x^*(\theta)}{\partial \theta} = -\frac{f_{xt}(x^*(\theta), \theta)}{f_{xx}(x^*(\theta), \theta)}.$$

[10] In this section we will use a lot of partial derivatives. Let f_{xy} denote $\frac{\partial}{\partial y}(\frac{\partial f}{\partial x})$.

And so

$$sign(\frac{\partial x^*(\theta)}{\partial \theta}) = sign(f_{xt}(x^*(\theta), \theta)).$$

In other words, the solution is increasing if and only if $f_{xt}(x^*(\theta), \theta) > 0$.

The example illustrates the strong requirements of the IFT: it only works if f is smooth, if the maximization problem has a unique solution for every parameter θ, and we can sign the derivative over the entire range of parameters only when f is strictly concave.

Also, from the example one might conclude (mistakenly) that smoothness or concavity were important to identify the effect of θ on x^*. To see why this would be the wrong conclusion, consider the class of problems:

$$x^*(\theta) = \arg \max_{x \in D(\theta)} f(x, \theta) \text{ for } \theta \in \Theta.$$

Now if $\varphi : \mathbb{R} \to \mathbb{R}$ is any increasing function, we have that

$$x^*(\theta) = \arg \max_{x \in D(\theta)} \varphi(f(x, \theta)) \text{ for } \theta \in \Theta.$$

In other words, applying an increasing transformation to the objective function changes the value function, but not the solution set. But this tells us that continuity, differentiability, and concavity of f have little to do with whether $x^*(\cdot)$ is increasing in θ. Even if f has all these nice properties, the objective function in the second problem won't be continuous if φ has jumps, won't be differentiable if φ has kinks, and won't be concave if φ is very convex. And yet any comparative statics conclusions that apply to f will also apply to $g = \varphi \circ f$!

We now develop a powerful series of results that allow us to make comparative statics conclusions *without* assumptions about continuity, differentiability, or concavity. A bonus is that the conditions required are often much easier to check than those required by the IFT.

5.3.1 MCS with One Choice Variable

Consider the case in which there is one variable, one parameter, and a fixed constraint set. In other words, let $f : S \times T \to \mathbb{R}$, with $S, T \subset \mathbb{R}$, and consider the class of problems

$$x^*(\theta) = \arg \max_{x \in D} f(x, \theta) \text{ for } \theta \in T.$$

We assume that the maximization problem has a solution ($x^*(\theta)$ always exists), but not that the solution is unique (so $x^*(\theta)$ may be a set).

The fact that $x^*(\theta)$ may be a set introduces a complication. How can a set be nondecreasing? For two sets of real numbers A and B, we say that $A \leq_s B$ in the *strong set order* if for any $a \in A$ and $b \in B$, $\min\{a, b\} \in A$ and $\max\{a, b\} \in B$. Note that if $A = \{a\}$ and $B = \{b\}$, then $A \leq_s B$ just means that $a \leq b$.

The strong set order is illustrated in Figure 17. In the top figure, $A \leq_s B$, but this is not true in the bottom figure. You should try using the definition to see exactly why this is true (hint: in the bottom figure, pick the two middle points as a and b).

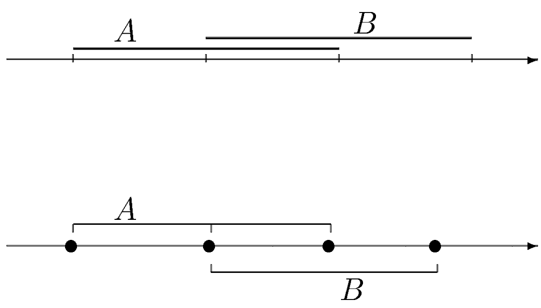

Figure 17: The Strong Set Order.

We will use the strong set order to make comparative statics conclusions: $x^*(\cdot)$ is nondecreasing in θ if and only if $\theta < \theta'$ implies that $x^*(\theta) \leq_s x^*(\theta')$. If $x^*(\cdot)$ is a function, this has the standard meaning that the function is nondecreasing. If $x^*(\cdot)$ is a compact-valued correspondence, then the functions $\overline{x}^*(\theta) = \max_{x \in x^*(\theta)} x$ and $\underline{x}^*(\theta) = \min_{x \in x^*(\theta)} x$ (i.e. the largest and smallest maximizers) are nondecreasing.

Definition 8 *The function* $f : \mathbb{R}^2 \to \mathbb{R}$ *is* ***supermodular*** *in* (x, θ) *if for all* $x' > x$, $f(x', \theta) - f(x, \theta)$ *is nondecreasing in* θ.[11]

[11] If f is supermodular in (x, θ), we sometimes say that f has **increasing differences** in (x, θ).

What exactly does this mean? If f is supermodular in (x, θ), then the incremental gain to choosing a higher x (i.e. x' rather than x) is greater when θ is higher. You can check that this is equivalent to the property that if $\theta' > \theta$, $f(x, \theta') - f(x, \theta)$ is nondecreasing in x.

The definition of supermodularity does not require f to be "nice". If f happens to be differentiable, there is a useful alternative characterization.

Lemma 12 *A twice continuously differentiable function $f : \mathbb{R}^2 \to \mathbb{R}$ is supermodular in (x, θ) if and only if $f_{x\theta}(x, \theta) \geq 0$ for all (x, θ).*

The next result, *Topkis' Monotonicity Theorem*, says that supermodularity is sufficient to draw comparative statics conclusions in optimization problems.

Theorem 13 (Topkis' Monotonicity Theorem) *If f is supermodular in (x, θ), then $x^*(\theta) = \arg\max_{x \in D} f(x, \theta)$ is nondecreasing.*

Proof. Suppose $\theta' > \theta$, and that $x \in x^*(\theta)$ and $x' \in x^*(\theta')$. We first show that $\max\{x, x'\} \in x^*(\theta')$. Since $x \in x^*(\theta)$, then $f(x, \theta) - f(\min\{x, x'\}, \theta) \geq 0$. This implies (you should check this) that $f(\max\{x, x'\}, \theta) - f(x', \theta) \geq 0$. So by supermodularity, $f(\max\{x, x'\}, \theta') - f(x', \theta) \geq 0$, which implies that $\max\{x, x'\} \in x^*(\theta')$.

We now show that $\min\{x, x'\} \in x^*(\theta)$. Since $x' \in x^*(\theta')$, then $f(x', \theta') - f(\max\{x, x'\}, \theta') \geq 0$, or equivalently $f(\max\{x, x'\}, \theta') - f(x', \theta') \leq 0$. By supermodularity, $f(\max\{x, x'\}, \theta) - f(x', \theta) \leq 0$. This implies (again you should verify) that $f(x, \theta) - f(\min\{x, x'\}, \theta) \leq 0$, so $\min\{x, x'\} \in x^*(\theta)$. *Q.E.D.*

Topkis' Theorem is handy in many situations. It allows one to prove that the solution to an optimization problem is nondecreasing simply by showing that the objective function is supermodular in the choice variable and the parameter.

Example Consider the following example in the theory of the firm. A monopolist faces the following problem:

$$\max_{q \geq 0} p(q)q - c(q, \theta)$$

where q is the quantity produced by the monopolist, $p(q)$ denotes the market price if q units are produced, and $c(\cdot)$ is the cost function. The parameter θ affects the monopolist's costs (for instance, it might be the price of a key input). Let $q^*(\theta)$ be the monopolist's optimal quantity choice. The objective function is supermodular in (x, θ) as long as $c_{q\theta}(q, \theta) \leq 0$. Thus, $q^*(\theta)$ is nondecreasing as long as θ decreases the marginal cost of production. Note that we can draw this conclusion with *no assumptions* on the demand function p, or the concavity of the cost function!

Remark 1 *In the example, we found conditions under which $q^*(\theta)$ would be nondecreasing in θ. But what if we wanted to show that $q^*(\theta)$ was nonincreasing in θ? We could have done this by showing that the objective function was supermodular in $(x, -\theta)$, rather than supermodular in (x, θ). The former means that q^* will be nondecreasing in $-\theta$, or nonincreasing in θ.*

5.3.2 Useful Tricks for Applications

There are several useful tricks when it comes to applying Topkis' Theorem.

1. **Parameterization.** In many economics models, we end up wanting to compare the solution to two distinct maximization problems, $\max_{x \in D} g(x)$ and $\max_{x \in D} h(x)$. (For instance, we might want to show that a profit-maximizing monopolist will set a higher price that a benevolent social-surplus maximizing firm.) It turns out we can do this using Topkis' Theorem, if we introduce a parameter $\theta \in \{0, 1\}$ and construct a "dummy objective function" f as follows:

$$f(x, \theta) = \begin{cases} g(x) \text{ when } \theta = 0 \\ h(x) \text{ when } \theta = 1 \end{cases}.$$

   If the function f is supermodular (i.e. if $h(x) - g(x)$ is nondecreasing in x), then $x^*(1) \geq x^*(0)$. Or in other words, the solution to the second problem is greater than the solution to the first. This trick is *very* useful — there is an exercise at the end that lets you try it.

2. **Aggregation.** Sometimes, an optimization problem will have many choice variables, but we want to draw a conclusion about only one of them. In these cases, it may be possible to apply Topkis' Theorem by "aggregating" the choice variables we care less about.

Consider the problem

$$\max_{\substack{x\in\mathbb{R},y\in\mathbb{R}^k\\(x,y)\in S}} f(x,y,\theta).$$

If we are only interested in the behavior of x we can rewrite this problem as

$$\max_{x\in\mathbb{R}} g(x,\theta) \text{ where } g(x,\theta)=\max_{\substack{y\in\mathbb{R}^k\\(x,y)\in S}} f(x,y,\theta).$$

In other words, we can decompose the problem in two parts: first find the maximum value $g(x,\theta)$ that can be achieved for any x, then maximize over the "value function" $g(x,\theta)$. If g is supermodular in (x,θ), Topkis' Theorem says that in the original problem, x^* will be nondecreasing in θ.

Example Here's an example to see how the aggregation method works. Consider the profit maximization problem of a competitive firm that produces an output x, using k inputs z_1,..., z_k,

$$\max_{\substack{x\in\mathbb{R},z\in\mathbb{R}^k\\x\leq F(z)}} px-w\cdot z.$$

p is the price of output, w the vector of input prices, and $F(\cdot)$ the firm's production function. Suppose that we are only interested on how p affects output $x^*(p)$. Then we can rewrite the problem as

$$\max_{x\in\mathbb{R}} g(x,p)=px-c(x) \text{ where } c(x)=\min_{\substack{z\in\mathbb{R}^k\\x\leq F(z)}} w\cdot z,$$

where $c(\cdot)$ is called the cost function. Then, since $g(x,p)$ has increasing differences in (x,p), we get that the firm's supply curve is nondecreasing, regardless of the shape of the production function.

5.3.3 Some Extensions

Topkis' Theorem says that f being supermodular in (x,θ) is a sufficient condition for $x^*(\theta)$ to be nondecreasing. It turns out supermodularity is a stronger assumption than what one really requires.

Definition 9 *A function $f:\mathbb{R}^2\rightarrow\mathbb{R}$ is **single-crossing** in (x,θ) if for all $x'>x$ and $\theta'>\theta$, (i) $f(x',\theta)-f(x,\theta)\geq 0$ implies that $f(x',\theta')-f(x,\theta')\geq 0$ and also (ii) $f(x',\theta)-f(x,\theta)>0$ implies that $f(x',\theta')-f(x,\theta')>0$.*

It's easy to show that if f is supermodular in (x,θ), then it is also single crossing in (x,θ). However, as illustrated in Figure 18, the opposite is not true. The figure plots an objective function for a problem with a discrete choice set $X = \{0,1,2\}$ and a continuous parameter set. You should verify that f satisfies the single crossing property, but it does not have increasing differences.

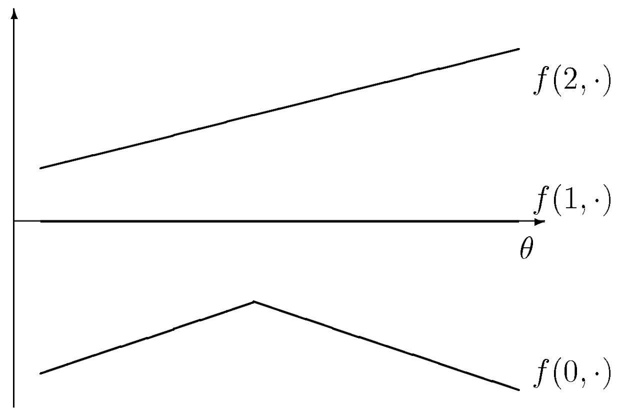

Figure 18: A function that is single crossing, but not supermodular.

Theorem 14 (Milgrom-Shannon Monotonicity Theorem) *If f is single-crossing in (x,θ), then $x^*(\theta) = \arg\max_{x\in D(\theta)} f(x,\theta)$ is nondecreasing. Moreover, there is a converse: if $x^*(\theta)$ is nondecreasing in θ for all choice sets D, then f is single-crossing in (x,θ).*

The first part of the Milgrom-Shannon Theorem is a bit stronger than Topkis' result. The proof is quite similar (in fact, if you look at the proof above, you will see that an identical proof suffices for Milgrom-Shannon). The second part says that single crossing is actually a *necessary* condition for monotonicity conclusions. What you should take away from this is the following: if you ever want to show that some variable is nondecreasing in a parameter, at some level you will necessarily be trying to verify the single-crossing condition.

There are times when the conclusion of the Milgrom-Shannon or Topkis Theorems is not exactly what we want. Sometimes we would like to show that a variable is strictly increasing in a parameter, and not just nondecreasing.

It turns out that the stronger assumption that $f_x(x,\theta)$ is *strictly increasing* in θ is almost enough to guarantee this, as the following result (due to Edlin and Shannon, 1998) shows.

Theorem 15 *Suppose that f is continuously differentiable in x and that $f_x(x,\theta)$ is increasing in θ for all x. Then for all $\theta < \theta'$, $x \in x^*(\theta)$, and $x' \in x^*(\theta')$ we get that:*

$$x \text{ is in the interior of the constraint set } \Rightarrow x < x'.$$

Proof. Since f is supermodular in (x,θ), Topkis' Theorem implies that $x \leq x'$. To see that the inequality is strict, note that because $x \in int(D)$ and $x \in x^*(\theta)$, we must have $f_x(x,\theta) = 0$. But then $f(x,\theta') > 0$. Since $x \in int(D)$, there must be some $\widehat{x} > x$ with $f(\widehat{x},\theta') > f(x,\theta')$. So $x \notin x^*(\theta')$. *Q.E.D.*

Unlike the above results, we can apply this result *only* if f is differentiable in x. However, its assumptions are still significantly weaker than the IFT — for instance, it applies equally well when Θ is discrete. Note that the theorem does require x to be an interior solution. Why? If x was at the upper boundary, there would be nowhere to go, so we could only have the weak conclusion $x \leq x'$.

6 Duality Theory

This final section looks at special class of PCOP that arise in consumer and producer theory. These problems have the following form:

$$V(\theta) = \max_{x \in K} \ \theta \cdot x,$$

where K is a convex subset of $\mathbb{R}^n$, and $\theta \in \mathbb{R}^n$. The essential feature of this problem is that the objective function is linear. This will allow us to say a lot of useful things about the solution and the value functions.

To do this we need to develop some new concepts.

Definition 10 *A **half-space** is a set of the form*

$$H_s(\theta, c) = \{x \in \mathbb{R}^n \mid \theta \cdot x \geq c\}, \ \theta \neq 0.$$

Definition 11 *A **hyperplane** is a set of the form*

$$H(\theta, c) = \{x \in \mathbb{R}^n \mid \theta \cdot x = c\}, \ \theta \neq 0.$$

These two concepts are illustrated in figure 19. In the case of $n = 2$, a hyperplane is a line, and a half-space is the set to one side of that line. In the case of $n = 3$, a hyperplane is a plane in three dimensional space.

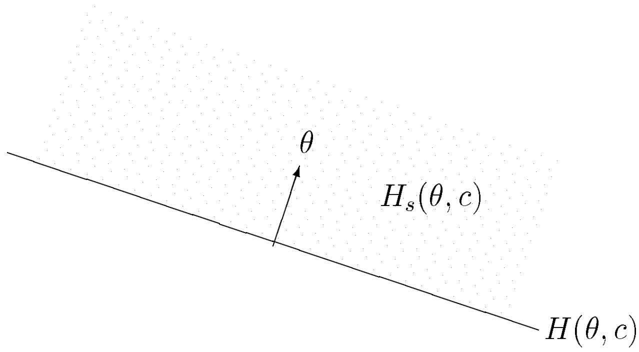

Figure 19: A Hyperplane and a Half-Space.

Theorem 16 (Separating Hyperplane Theorem) *Suppose that S and T are two convex, closed, and disjoint ($S \cap T = \emptyset$) subsets or $\mathbb{R}^n$. Then there exists $\theta \in \mathbb{R}^n$, and $c \in \mathbb{R}$ such that*

$$\theta \cdot x \geq c \text{ for all } x \in S \text{ and } \theta \cdot y < c \text{ for all } y \in T.$$

This result is illustrated in figure 20. If the sets S and T are not intersecting and they are closed, then we can draw a hyperplane between them such that S lies strictly to one side of the hyperplane and T lies to the other. Note that the vector θ points towards the set S and that it need not be uniquely defined. (Can you think of an example when it is uniquely defined?) The figure on the bottom also shows that separation by a hyperplane may not be possible if one of the sets is not convex.

A consequence of the Separating Hyperplane Theorem is that *any* closed and convex set S can be described as the intersection of *all* half-spaces that contain it:

$$S = \bigcap_{(\theta,c) \text{ s.t. } S \subset H_s(\theta,c)} H_s(\theta, c).$$

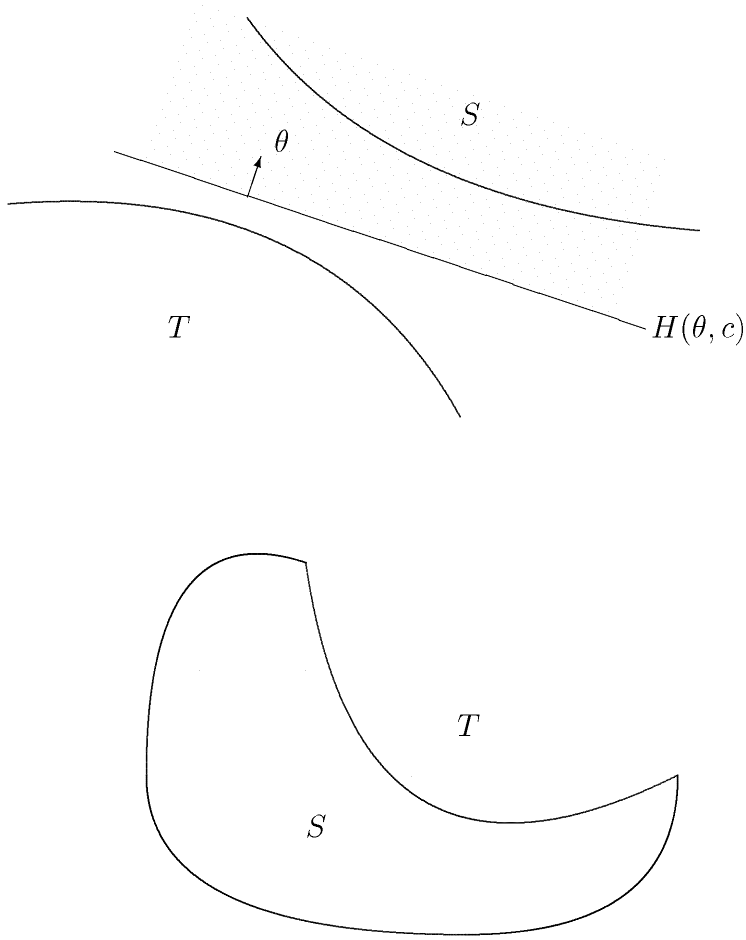

Figure 20: Separating Convex Sets.

This is easily seen in figure 21. If follows directly from the fact that for any point $x \notin S$ there is a half-space that contains S but not x. As simple as it is, this is the key idea in duality theory, and we will get a lot of mileage out of it.

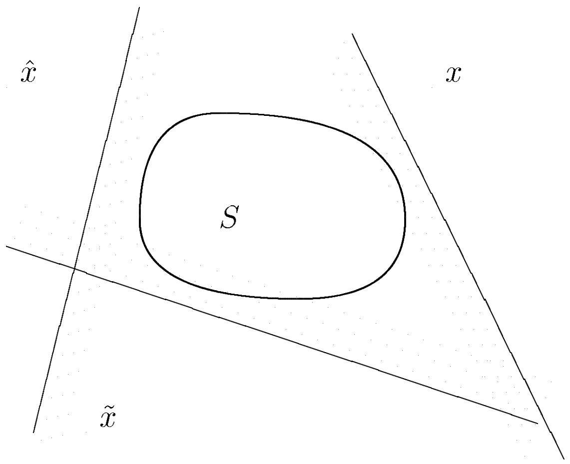

Figure 21: A convex set is the intersection of all half-spaces that contain it.

The same idea can be expressed using the notion of support functions. Any set S has two support functions:

$$M_S(\theta) = \sup\{\theta \cdot x \mid x \in S\} \qquad \text{and} \qquad m_S(\theta) = \inf\{\theta \cdot x \mid x \in S\}.$$

To understand these definitions you need to know that "sup" and "inf" are extensions of the ideas of "max" and "min". In particular:

Definition 12 *Let A be a non-empty set of real numbers.*

1. *The* ***infimum*** *of A, $\inf A$, is given by $x \in \mathbb{R} \cup \{-\infty, \infty\}$ such that: (i) $x \leq y$ for all $y \in A$, and (ii) there is no $\widehat{x}$ such that $x < \widehat{x} \leq y$ for all $y \in A$.*

2. *The* ***supremum*** *of A, $\sup A$, is given by $x \in \mathbb{R} \cup \{-\infty, \infty\}$ such that: (i) $x \geq y$ for all $y \in A$, and (ii) there is no $\widehat{x}$ such that $x > \widehat{x} \geq y$ for all $y \in A$.*

The following examples emphasize the distinction between inf and min, and sup and max.

1. $\inf [0,1] = 0$, and $\min [0,1] = 0$
2. $\sup [0,1) = 1$, but $\max [0,1)$ does not exist.
3. $\inf (0,\infty) = 0$ and $\sup (0,\infty) = \infty$, but neither $\min (0,\infty)$ nor $\max (0,\infty)$ exist.

If $\min S$ exists, then $\min S = \inf S$. The difference between inf and min arises when the set does not have a min. In that case, if S is bounded below then $\inf S$ is the greatest lower bound, if not it is $-\infty$. In other words, inf $S = -\infty$ is an equivalent way of saying that the set S is not bounded below.

Let's try an example using support functions.

Example Figure 22 depicts the set $S = \{x \in \mathbb{R}^2 \mid x_1 \leq 0 \text{ and } x_2 \leq 0\}$. Let's characterize $m_S(\theta)$. First, consider any vector θ that points to a direction outside of S. If we choose a series of x's in the direction of $-\theta$, this will decrease $\theta \cdot x$ without bound. Hence $m_S(\theta) = \inf \{\theta \cdot x \mid x \in S\} = -\infty$. By contrast, consider a vector $\theta^{'}$ on the axis, or a vector such as $\widehat{\theta}$ that points towards the interior of S. In these cases, the value of $\theta \cdot x$ is minimized by taking $x = 0$, and $m_S(\theta) = 0$.

How about $M_S(\theta)$? Once more, there are three cases. If θ points to a direction outside of S, then $\sup \{\theta \cdot x \mid x \in S\} = 0$. The same is true in the direction θ' on the axis. However, for any vector $\widehat{\theta}$ that points towards the interior of S, we have that $M_S(\theta) = \infty$. (To see this notice that as we move in the direction of $\widehat{\theta}$ farther and farther down we increase the value of $\widehat{\theta}.x$ while staying in the set S. Thus, $\{\theta \cdot x \mid x \in S\}$ is not bounded above).

We can use the support functions to give a characterization of any closed, convex set S:

$$S = \bigcap_{\theta} \{x \mid \theta \cdot x \leq M_S(\theta)\} \qquad \text{or} \qquad S = \bigcap_{\theta} \{x \mid \theta \cdot x \geq m_S(\theta)\}\,.$$

The intuition for this is straightforward. We have seen before that any convex and closed set is equal to the intersection of all the half-spaces that contain it. The expression $S = \cap_\theta \{x|\theta.x \geq m_S(\theta)\} = \cap_\theta H_S(\theta, m_S(\theta))$ is almost identical except that it takes the intersection only over some half-spaces, not all half-spaces. But by construction of the support function $m_S(\theta)$, we have that if $m > m_S(\theta)$, then S is not contained in $H_S(\theta, m)$, whereas if $m < m_S(\theta)$,

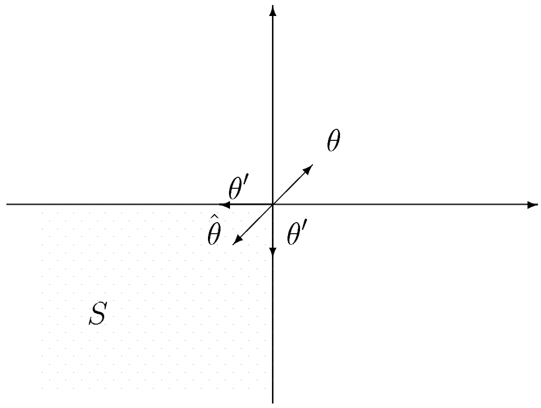

Figure 22: Support functions for a convex set.

then $H_S(\theta, m) \subset H_S(\theta, m_S(\theta))$. This suggests that for each direction θ there is a unique hyperplane that is relevant for characterizing the set S — the one given by the support function.

Support functions have some very useful properties. To state them, we need one definition.

Definition 13 *A function $h : \mathbb{R}^n \to \mathbb{R}$ is* ***homogeneous (of degree one)*** *if for all $t > 0$, $h(tx) = th(x)$.*

Theorem 17 *1. $M_S(\cdot)$ is homogeneous and convex in θ.*

2. $m_S(\cdot)$ is homogeneous and concave in θ.

Proof. (1) Homogeneity follows immediately from the fact that if $t > 0$:

$$M_S(t\theta) = \sup \{t\theta \cdot x \mid x \in S\} = t \cdot \sup \{\theta \cdot x \mid x \in S\} = tM_S(\theta).$$

Now consider convexity. We will need two facts (proving them is an exercise). Fact 1: For all non-empty sets of real numbers A and B, $\sup(A + B) \leq \sup A + \sup B$. Fact 2: If $A \subset B \subset \mathbb{R}$, then $\sup A \leq \sup B$.

We need to show that for all $t \in [0, 1]$, and θ, θ',

$$M_S(\theta^t) \leq tM_S(\theta) + (1 - t)M_S(\theta').$$

Note that

$$tM_S(\theta) = \sup \underbrace{\{t\theta \cdot x \mid x \in S\}}_{A},$$

and

$$(1-t)M_S(\theta') = \sup \underbrace{\{(1-t)\theta \cdot x \mid x \in S\}}_{B}.$$

Then since $M_S(\theta^t) = \{t\theta \cdot x + (1-t)\theta \cdot x \mid x \in S\} \subset A+B$, we can use Facts 1 and 2 to conclude:

$$M_S(\theta^t) \leq \sup A + B \leq \sup A + \sup B = tM_S(\theta) + (1-t)M_S(\theta').$$

(2) The proof is essentially the same. *Q.E.D.*

We are now ready to prove one last big result.

Theorem 18 (Duality Theorem) *Let S be any non-empty and closed subset of $\mathbb{R}^n$. Then*

1. *$M_S(\cdot)$ is differentiable at $\widehat{\theta}$ if and only if there exists a unique $\widehat{x} \in S$ such that $\widehat{\theta} \cdot \widehat{x} = M_S(\widehat{\theta})$. Furthermore, in that case $D_\theta M_S(\widehat{\theta}) = \widehat{x}$.*

2. *$m_S(.)$ is differentiable at $\widehat{\theta}$ if, and only if, there exists a unique $\widehat{x} \in S$ such that $\widehat{\theta} \cdot \widehat{x} = m_S(\widehat{\theta})$. Furthermore, in that case $D_\theta m_S(\widehat{\theta}) = \widehat{x}$.*

The Duality Theorem establishes necessary *and* sufficient conditions for the support function to be differentiable. The support function $M_S(\cdot)$ is differentiable if and only if the maximization problem $\max\{\widehat{\theta} \cdot x | x \in S\}$ has a unique solution $\widehat{x}$. Furthermore, in that case the "envelope" derivative $D_\theta M_S(\widehat{\theta})$ is equal to $\widehat{x}$. Similarly, the support function $m_S(\cdot)$ is differentiable if and only if the minimization problem $\min\{\widehat{\theta} \cdot x | x \in S\}$ has a unique solution $\tilde{x}$, in which case the envelope derivative $D_\theta M_S(\widehat{\theta})$ is equal to $\tilde{x}$.

Okay, so we have worked very hard to define these support functions. But why do we care? We care because the value function for the class of problems that we are interested in,

$$V(\theta) = \max_{x \in K} \theta \cdot x,$$

looks like a support function. In fact, as long as the problem has a unique solution we know that $V(\theta) = M_K(\theta)$. Thus, we have learned that in this maximization problem:

1. The value function is homogeneous and convex; and

2. As long as the maximization problem has a unique solution $x^*(\theta)$, the value function is differentiable and $D_\theta V(\theta) = x^*(\theta)$.

Analogous statements also hold for minimization problems.

It is difficult to overstate the usefulness of these results: if you keep them in mind, you will be *amazed* at how many results in price theory follow as a direct consequence!

7 Exercises

1. Are the following functions continuous? Provide a careful proof.

   (a) $X = Y = \mathbb{R}$, and $h(x) = x^2 + 5$

   (b) $X = (-\infty, 0] \cup \{1\} \cup [2, \infty)$, $Y = \mathbb{R}$, and $h(x) = \begin{cases} x & \text{if } x \leq 0 \\ 2x & \text{if } x = 1 \\ 10x & \text{otherwise} \end{cases}$

   (c) $X = \mathbb{R}^{2n}$, $Y = \mathbb{R}$, and $h(x, y) = \|x - y\|$.

2. Prove that the correspondence in Figure 4 is usc but not lsc, and that the correspondence in Figure 5 is lsc but not usc. (Note: we proved the negative claims in the text, but in a "chatty" way; try writing down a careful proof).

3. Consider the correspondence given by

$$\phi(x) = \begin{cases} \{0, 1/x\} & \text{if } x > 0 \\ \{0\} & \text{if } x = 0 \end{cases}.$$

   Show that ϕ is lower semi-continuous, but not upper semi-continuous.

4. A Walrasian budget set at prices p and wealth w is a subset of $\mathbb{R}^n_+$ given by

$$B(p, w) = \{x \in \mathbb{R}^n_+ | p.x \leq w\},$$

   for $p \in R^n_{++}$ and w.

   (a) Prove that the correspondence $B : \mathbb{R}^n_{++} \times \mathbb{R} \Rightarrow \mathbb{R}^n_+$ is usc and lsc.

   (b) Does our definition of usc applies if we extend the domain of the correspondence to include zero prices? Why?

5. Consider the function $f : [0,1] \to \mathbb{R}$, given by

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

Does the problem $\max_{x \in [0,1]} f(x)$ have a solution? Why doesn't the Theorem of the Maximum apply?

6. Prove the version of the Kuhn-Tucker Theorem that includes non-negativity constraints. (Hint: Apply the original Kuhn-Tucker Theorem to the optimization problem that includes the additional n non-negative constraints.)

7. Solve the following optimization problem using the Kuhn Tucker Theorem.

$$\min_{x \in \mathbb{R}^2_+} \; w_1 x_1 + w_2 x_2 \text{ subject to } y \geq (x_1^\rho + x_2^\rho)^{1/\rho},$$

assuming that $y, w, \rho > 0$.

8. Consider the following optimization problem

$$\max_{x \in \mathbb{R}^n_+} \prod_{i=1}^{n} x_i^{\alpha_i} \text{ subject to } p.x \leq w$$

where $\alpha_i, p_i, w > 0$. Compute the solution and value function to this problem for any parameters (p, w) using Lagrange's Theorem. (You need to show first that this Theorem characterizes the solution to the problem even though it has inequality constraints.

9. Let x^* and V denote the solution set and value function for the previous problem. The goal of this problems is to familiarize you with the mechanics of the IFT and the ET. Thus, for the moment ignore the explicit solutions that you obtained in problem 6.

   (a) Look at the FOCs for the problem. Are the conditions of the IFT satisfied?

   (b) Suppose that $n = 2$. Compute the $D_{(p,w)} x^*$ using the IFT.

   (c) Now compute $D_{(p,w)} V$ using the Envelope Theorem.

   (d) Now compute $D_{(p,w)} x^*$ using the IFT.

(e) Now compute $D_{(p,w)}V$ and $D_{(p,w)}x^*$ directly by taking the derivatives of x^* and V. Do you get the same answer?

10. Show that for any two sets of real number A, B: $A \leq_s B$ if and only if (1) $A \backslash B$ lies below $A \cap B$ and (2) $B \backslash A$ lies above $A \cap B$.

11. (due to Segal) A function $f : \mathbb{R}^2 \to \mathbb{R}$ has strict increasing differences in (x, θ) if $x' > x$, $\theta > \theta'$

$$f(x', \theta') - f(x, \theta') > f(x', \theta) - f(x, \theta).$$

(a) Show that if f has strict increasing preferences, then for *any selection* of $x^*(\theta) \in Arg \max_{x \in X} f(x, \theta)$ and $x^*(\theta') \in \arg\max_{x \in X} f(x, \theta')$ with $\theta' > \theta'$ we must have $x^*(\theta') \geq x^*(\theta)$.

(b) Show by means of an example that property does not guarantee that $x^*(\theta') > x^*(\theta)$. (Hint: the easiest examples involve a discrete choice space.)

12. Consider the problem of a profit maximizing firm discussed in Section 5.3

$$\max_{\substack{x \in \mathbb{R}, z \in \mathbb{R}^k \\ x \leq F(z)}} px - wz.$$

Show that for any input k, $z_k^*(p, w)$ is non-increasing in w_k. (Hint: Use the aggregation method.)

13. Use the aggregation method to show that Hicksian demand curves are downward-sloping. The consumer expenditure problem is:

$$\min_{x \in \mathbb{R}^n} p \cdot x \text{ subject to } u(x) \geq \overline{u},$$

where $p = (p_1, ..., p_n)$ is the price of each good and $x_1, ..., x_n$ the quantities consumed. Your goal is to show that $x_1^*(p_1, ..., p_n)$ is nonincreasing in p_1.

14. Consider the following problem that arises in monopoly theory. A monopolist solves the problem

$$\max_{x \geq 0} \pi(x) = P(x)x - c(x),$$

where $P(x)$ is the inverse demand function. (It provides the price at which the market demands exactly x units of the good.) The following problem characterizes the maximization of total social surplus:

$$\max_{x \geq 0} W(x) = \pi(x) + CS(P(x)),$$

where $CS(p) = \int_p^\infty x(p)dp$ is consumer surplus, and $x(p)$ is the demand function (i.e., assume that $P(x(p)) = p$ for all p and $x(P(x)) = x$ for all x.).

The goal of the exercise is to compare the monopoly output to the surplus-maximizing output level. Let $\theta = 0$ correspond to the firm's profit-maximization program, $\theta = 1$ correspond to the total surplus maximization program, and define

$$\max_{x \geq 0} f(x, \theta) = \pi(x) + \theta \cdot CS(P(x)).$$

Suppose that the inverse demand function $P(x)$ and consumer surplus $CS(p)$ are non-increasing functions. Is this enough to show that the monopoly output is always (weakly) less than the surplus maximizing level?

15. Let $S = \{x \in \mathbb{R}^2 | \, \|x - 0\| \leq r^2\}$. Compute the support functions $M_S(\theta)$ and $m_S(\theta)$ for this set.

16. Prove that for every set convex and closed subset S of $\mathbb{R}^n$, $S = \cap_\theta \{x | \theta.x \geq m_S(\theta)\}$. Show also that the result is not true if S is open or non-convex.

17. Prove that for all non-empty sets of real numbers A and B, $\sup(A + B) \leq \sup A + \sup B$.