The Kuhn-Tucker and Envelope Theorems

Peter Ireland*
ECON 772001 - Math for Economists
Boston College, Department of Economics

Fall 2023

The Kuhn-Tucker and envelope theorems can be used to characterize the solution to a wide range of constrained optimization problems: static or dynamic, and under perfect foresight or featuring randomness and uncertainty. In addition, these same two results provide foundations for the work on the maximum principle and dynamic programming that we will do later on. For both of these reasons, the Kuhn-Tucker and envelope theorems provide the starting point for our analysis. Let's consider each in turn, first in abstract but somewhat special settings, then applied to some economic examples, and finally in full generality.

1 The Kuhn-Tucker Theorem

References:

Dixit, Chapters 2 and 3.

Simon-Blume, Chapters 18 and 19.

Acemoglu, Appendix A.

Consider a simple constrained optimization problem:

$x \in \mathbf{R}$ choice variable

$F : \mathbf{R} \to \mathbf{R}$ objective function, continuously differentiable

$c \geq G(x)$ constraint, with $c \in \mathbf{R}$ and $G : \mathbf{R} \to \mathbf{R}$, also continuously differentiable.

The problem can be stated as:

$$\max_x F(x) \text{ subject to } c \geq G(x)$$

This problem is "simple" because it is static and contains no random or stochastic elements that would force decisions to be made under uncertainty. This problem is also "simple" because it has a single choice variable and a single constraint. All these simplifications will make our statement and proof of the Kuhn-Tucker theorem as clean and intuitive as possible. But the results can be generalized along all of these dimensions and, throughout the semester, we will work through examples that do so.

Probably the easiest way to solve this problem is via the method of Lagrange multipliers. The mathematical foundations that allow for the application of this method are given to us by Lagrange's Theorem or, in its most general form, the Kuhn-Tucker Theorem.

To prove this theorem, begin by defining the Lagrangian:

$$L(x, \lambda) = F(x) + \lambda[c - G(x)]$$

for any $x \in \mathbf{R}$ and $\lambda \in \mathbf{R}$.

Theorem (Kuhn-Tucker) Suppose that x^* maximizes $F(x)$ subject to $c \geq G(x)$, where F and G are both continuously differentiable, and suppose that $G'(x^*) \neq 0$. Then there exists a value λ^* of λ such that x^* and λ^* satisfy the following four conditions:

$$L_1(x^*, \lambda^*) = F'(x^*) - \lambda^* G'(x^*) = 0, \tag{1}$$

$$L_2(x^*, \lambda^*) = c - G(x^*) \geq 0, \tag{2}$$

$$\lambda^* \geq 0, \tag{3}$$

and

$$\lambda^*[c - G(x^*)] = 0. \tag{4}$$

Proof Consider two possible cases, depending on whether or not the constraint is binding at x^*.

Case 1: Nonbinding Constraint.

If $c > G(x^*)$, then let $\lambda^* = 0$. Clearly, (2)-(4) are satisfied, so it only remains to show that (1) must hold. With $\lambda^* = 0$, (1) holds if and only if

$$F'(x^*) = 0. \tag{5}$$

We can show that (5) must hold using a proof by contradiction. Suppose that instead of (5), it turns out that

$$F'(x^*) < 0.$$

Then, by the continuity of F and G, there must exist an $\varepsilon > 0$ such that

$$F(x^* - \varepsilon) > F(x^*) \text{ and } c > G(x^* - \varepsilon).$$

But this result contradicts the assumption that x^* maximizes $F(x)$ subject to $c \geq G(x)$. Similarly, if it turns out that

$$F'(x^*) > 0,$$

then by the continuity of F and G there must exist an $\varepsilon > 0$ such that

$$F(x^* + \varepsilon) > F(x^*) \text{ and } c > G(x^* + \varepsilon),$$

But, again, this result contradicts the assumption that x^* maximizes $F(x)$ subject to $c \geq G(x)$. This establishes that (5) must hold, completing the proof for case 1.

Case 2: Binding Constraint.

If $c = G(x^*)$, then let $\lambda^* = F'(x^*)/G'(x^*)$. This is possible, given the assumption that $G'(x^*) \neq 0$. Clearly, (1), (2), and (4) are satisfied, so it only remains to show that (3) must hold. With $\lambda^* = F'(x^*)/G'(x^*)$, (3) holds if and only if

$$F'(x^*)/G'(x^*) \geq 0. \tag{6}$$

We can show that (6) must hold using a proof by contradiction. Suppose that instead of (6), it turns out that

$$F'(x^*)/G'(x^*) < 0.$$

One way that this can happen is if $F'(x^*) > 0$ and $G'(x^*) < 0$. But if these conditions hold, then the continuity of F and G implies the existence of an $\varepsilon > 0$ such that

$$F(x^* + \varepsilon) > F(x^*) \text{ and } c = G(x^*) > G(x^* + \varepsilon),$$

which contradicts the assumption that x^* maximizes $F(x)$ subject to $c \geq G(x)$. And if, instead, $F'(x^*)/G'(x^*) < 0$ because $F'(x^*) < 0$ and $G'(x^*) > 0$, then the continuity of F and G implies the existence of an $\varepsilon > 0$ such that

$$F(x^* - \varepsilon) > F(x^*) \text{ and } c = G(x^*) > G(x^* - \varepsilon),$$

which again contradicts the assumption that x^* maximizes $F(x)$ subject to $c \geq G(x)$. This establishes that (6) must hold, completing the proof for case 2.

Notes:

a) The theorem can be extended to handle cases with more than one choice variable and more than one constraint: see Dixit, Simon-Blume, Acemoglu, or section 4.1 of the notes below.

b) The extra assumption that $G'(x^*) \neq 0$ is needed to guarantee the existence of a multiplier λ^* satisfying (1)-(4) in the case where the constraint binds at the optimum, so that $c = G(x^*)$. This extra assumption is called the "constraint qualification." In economic applications, this constraint qualification is almost always satisfied. But it is important to remember that, in cases where the constraint qualification fails to hold, it may be impossible to find a value of λ^* that, together with the the value x^* that solves the constrained optimization problem, satisfies (1)-(4).

c) Equations (1)-(4) are necessary conditions: If x^* is a solution to the optimization problem, then there exists a λ^* such that (1)-(4) must hold. But (1)-(4) are not sufficient conditions: if x^* and λ^* satisfy (1)-(4), it does not follow automatically that x^* is a solution to the optimization problem.

As an example that illustrates point (b), consider the problem:

$$\max_x e^x - \frac{2}{1+x^2} \text{ subject to } 1 \geq x.$$

With the Lagrangian defined as

$$L(x, \lambda) = e^x - \frac{2}{1+x^2} + \lambda(1-x),$$

the Kuhn-Tucker conditions are

$$L_1(x^*, \lambda^*) = e^{x^*} + \frac{4x^*}{[1+(x^*)^2]^2} - \lambda^* = 0,$$

$$L_2(x^*, \lambda^*) = 1 - x^* \geq 0,$$

$$\lambda^* \geq 0,$$

and

$$\lambda^*(1-x^*) = 0.$$

These conditions are satisfied when $x^* = 1$ and $\lambda^* = e+1$, which corresponds to the solution to the problem, but they are also satisfied when $x^* = -0.2205$ and $\lambda^* = 0$, which corresponds instead to the solution to the *minimization* problem

$$\min_x e^x - \frac{2}{1+x^2} \text{ subject to } 1 \geq x.$$

But despite point (b) listed above, the Kuhn-Tucker theorem is extremely useful in practice. Suppose that we are looking for the solution x^* to the constrained optimization problem

$$\max_x F(x) \text{ subject to } c \geq G(x).$$

The theorem tells us that if we form the Lagrangian

$$L(x, \lambda) = F(x) + \lambda[c - G(x)],$$

then x^* and the associated λ^* must satisfy the first-order condition (FOC) obtained by differentiating L by x and setting the result equal to zero:

$$L_1(x^*, \lambda^*) = F'(x^*) - \lambda^* G'(x^*) = 0, \tag{1}$$

In addition, we know that x^* must satisfy the constraint:

$$c \geq G(x^*). \tag{2}$$

We know that the Lagrange multiplier λ^* must be nonnegative:

$$\lambda^* \geq 0. \tag{3}$$

And finally, we know that the complementary slackness condition

$$\lambda^*[c - G(x^*)] = 0, \tag{4}$$

must hold: If $\lambda^* > 0$, then the constraint must bind; if the constraint does not bind, then $\lambda^* = 0$.

In searching for the value of x that solves the constrained optimization problem, we only need to consider values of x^* that satisfy (1)-(4). In the example from above, for instance, it is straightforward to compare the values of the objective function when $x = 1$ and $x = -0.2205$ to conclude that the solution to the optimization problem is $x^* = 1$.

Furthermore, in many economic applications, the objective function $F(x)$ will be concave and the function $G(x)$ entering into the constraint will be convex. Under these additional assumptions, the Kuhn-Tucker conditions (1)-(4) are sufficient as well as necessary.

To see this, let x^* and λ^* satisfy (1)-(4), and let x be any other value of the choice variable that satisfies the constraint $c \geq G(x)$. Since F is concave, it satisfies by definition

$$\omega F(x) + (1-\omega)F(x^*) \leq F[\omega x + (1-\omega)x^*]$$

for all $\omega \in [0,1]$. Let $\omega > 0$, and rewrite this inequality as

$$F(x) - F(x^*) \leq \frac{F[x^* + \omega(x - x^*)] - F(x^*)}{\omega}$$

or

$$F(x) - F(x^*) \leq \left\{ \frac{F[x^* + \omega(x - x^*)] - F(x^*)}{\omega(x - x^*)} \right\} (x - x^*).$$

Taking the limit as $\omega \rightarrow 0$ of the term inside brackets on the right-hand side of this last inequality yields

$$F(x) - F(x^*) \leq F'(x^*)(x - x^*)$$

or, since $F'(x^*) = \lambda^* G'(x^*)$ by (1),

$$F(x) - F(x^*) \leq \lambda^* G'(x^*)(x - x^*).$$

Observe next that since G is convex, it satisfies by definition

$$\omega G(x) + (1-\omega)G(x^*) \geq G[\omega x + (1-\omega)x^*]$$

for all $\omega \in [0,1]$. Let $\omega > 0$, rewrite this inequality as

$$G(x) - G(x^*) \geq \left\{ \frac{G[x^* + \omega(x - x^*)] - G(x^*)}{\omega(x - x^*)} \right\} (x - x^*),$$

and take the limit as $\omega \rightarrow 0$ of the term inside brackets on the right-hand side to obtain

$$G(x) - G(x^*) \geq G'(x^*)(x - x^*).$$

Since $\lambda^* \geq 0$ by (3),

$$F(x) - F(x^*) \leq \lambda^* G'(x^*)(x - x^*).$$

and

$$G(x) - G(x^*) \geq G'(x^*)(x - x^*).$$

can be combined to yield

$$F(x) - F(x^*) \leq \lambda^*[G(x) - G(x^*)]$$

or, since (4) implies $\lambda^* G(x^*) = \lambda^* c$,

$$F(x) - F(x^*) \leq \lambda^*[G(x) - c].$$

Since the constraint on x requires $c \geq G(x)$, this last inequality confirms that

$$F(x) - F(x^*) \leq 0$$

or $F(x^*) \geq F(x)$ for all x such that $c \geq G(x)$. That is, x^* must solve the constrained optimization problem.

For extensions of this proof of sufficiency to the more general case where there is more than one choice variable and more than one constraint, see Simon and Blume (Theorem 21.22, pp.532-533) or Acemoglu (Theorem A.29, pp.911-913).

Two related pieces of terminology:

a) When F is concave and G is convex, the problem

$$\max_x F(x) \text{ subject to } c \geq G(x),$$

is called a "concave program."

b) Simon and Blume's Theorem 21.22 and Acemoglu's Theorem A.29 also show that under these extra assumptions, (x^*, λ^*) is a "saddle point" of $L(x, \lambda)$: x^* maximizes $L(x, \lambda^*)$, while λ^* minimizes $L(x^*, \lambda)$ subject to the constraint that $\lambda \geq 0$. Intuitively, (1) then appears as a first-order condition for the problem of maximizing L with respect to x, while (2) is like a first-order condition for minimizing L with respect to λ.

Thus, point (b) gives us some intuition about how the Kuhn-Tucker theorem works. It uses the Lagrangian to turn a constrained optimization problem into an unconstrained problem, where the solution for x^* is a critical point of $L(x, \lambda^*)$ rather than a critical point of $F(x)$. And when the extra assumptions made in concave programming are adopted, x^* is not only a critical point of $L(x, \lambda^*)$, it also maximizes $L(x, \lambda^*)$.

As an example illustrating the saddle-point property, consider the problem

$$\max_x \ln(x) \text{ subject to } 1 \geq x.$$

Since the logarithmic objective function is strictly increasing, we know that $x^* = 1$ is the solution. Forming the Lagrangian

$$L(x, \lambda) = \ln(x) + \lambda(1 - x)$$

and taking the first-order condition

$$L_1(x^*, \lambda^*) = \frac{1}{x^*} - \lambda^* = 0$$

reveals that the associated value of λ is $\lambda^* = 1$. Substituting this value for λ back into the Lagrangian

$$L(x, \lambda^*) = \ln(x) + 1 - x,$$

we can see that $L(x, \lambda^*)$ is maximized with $x^* = 1$.

Note, however, that in the general case where $F(x)$ need not be concave, x^* will always be a critical point of the Lagrangian – that is, it will satisfy the first-order condition (1) – but x^* need not maximize $L(x, \lambda)$. An example of how this can happen is given by Dixit's example 7.2 (p.103). Consider the problem

$$\max_x e^x \text{ subject to } 1 \geq x.$$

Since the exponential objective function is strictly increasing, we know again that this problem is solved with $x^* = 1$. Forming the Lagrangian and taking the first-order condition

$$L_1(x^*, \lambda^*) = e^{x^*} - \lambda^* = 0,$$

shows that the associated value of λ^* is e. But $\lambda^* = e$ implies that

$$L(x, \lambda^*) = e^x + \lambda^*(1 - x) = e^x + e(1 - x) = e^x - ex + e,$$

and since $e^x - ex$ grows without bound as either $x \to \infty$ or $x \to -\infty$, $x^* = 1$ does not maximize the Lagrangian given λ^*. Hence, (x^*, λ^*) is not a saddle-point of L, since the objective function from the original problem is convex, not concave.

One final note:

Our general constraint, $c \geq G(x)$, nests as a special case the nonnegativity constraint $x \geq 0$, obtained by setting $c = 0$ and $G(x) = -x$.

So nonnegativity constraints can be introduced into the Lagrangian in the same way as all other constraints. If we consider, for example, the extended problem

$$\max_x F(x) \text{ subject to } c \geq G(x) \text{ and } x \geq 0,$$

then we can introduce a second multiplier μ, form the Lagrangian as

$$L(x, \lambda, \mu) = F(x) + \lambda[c - G(x)] + \mu x,$$

and write the first order condition for the optimal x^* as

$$L_1(x^*, \lambda^*, \mu^*) = F'(x^*) - \lambda^* G'(x^*) + \mu^* = 0. \tag{1'}$$

In addition, analogs to our earlier conditions (2)-(4) must also hold for the second constraint: $x^* \geq 0$, $\mu^* \geq 0$, and $\mu^* x^* = 0$.

Kuhn and Tucker's original statement of the theorem, however, does not incorporate nonnegativity constraints into the Lagrangian. Instead, even with the additional nonnegativity constraint $x \geq 0$, they continue to define the Lagrangian as

$$L(x, \lambda) = F(x) + \lambda[c - G(x)].$$

If this case, the first order condition for x^* must be modified to read

$$L_1(x^*, \lambda^*) = F'(x^*) - \lambda^* G'(x^*) \leq 0, \text{with equality if } x^* > 0. \tag{1''}$$

Of course, in $(1')$, $\mu^* \geq 0$ in general and $\mu^* = 0$ if $x^* > 0$. So a close inspection reveals that these two approaches to handling nonnegativity constraints lead in the end to the same results.

2 The Envelope Theorem

References:

Dixit, Chapter 5.

Simon-Blume, Chapter 19.

Acemoglu, Appendix A.

In our discussion of the Kuhn-Tucker theorem, we considered an optimization problem of the form

$$\max_x F(x) \text{ subject to } c \geq G(x)$$

Now, let's generalize the problem by allowing the functions F and G to depend on a parameter $\theta \in \mathbf{R}$. The problem can now be stated as

$$\max_x F(x, \theta) \text{ subject to } c \geq G(x, \theta)$$

For this problem, define the maximum value function $V : \mathbf{R} \rightarrow \mathbf{R}$ as

$$V(\theta) = \max_x F(x, \theta) \text{ subject to } c \geq G(x, \theta)$$

Note that evaluating V requires a two-step procedure:

First, given θ, find the value of x^* that solves the constrained optimization problem.

Second, substitute this value of x^*, together with the given value of θ, into the objective function to obtain

$$V(\theta) = F(x^*, \theta)$$

Now suppose that we want to investigate the properties of this function V. Suppose, in particular, that we want to take the derivative of V with respect to its argument θ.

As the first step in evaluating $V'(\theta)$, consider solving the constrained optimization problem for any given value of θ by setting up the Lagrangian

$$L(x, \lambda) = F(x, \theta) + \lambda[c - G(x, \theta)]$$

We know from the Kuhn-Tucker theorem that the solution x^* to the optimization problem and the associated value of the multiplier λ^* must satisfy the complementary slackness condition:

$$\lambda^*[c - G(x^*, \theta)] = 0$$

Use this last result to rewrite the expression for V as

$$V(\theta) = F(x^*, \theta) = F(x^*, \theta) + \lambda^*[c - G(x^*, \theta)]$$

So suppose that we tried to calculate $V'(\theta)$ simply by differentiating both sides of this equation with respect to θ:

$$V'(\theta) = F_2(x^*, \theta) - \lambda^* G_2(x^*, \theta).$$

But, in principle, this formula may not be correct. The reason is that x^* and λ^* will themselves depend on the parameter θ, and we must take this dependence into account when differentiating V with respect to θ.

However, the envelope theorem tells us that our formula for $V'(\theta)$ is, in fact, correct. That is, the envelope theorem tells us that we can ignore the dependence of x^* and λ^* on θ in calculating $V'(\theta)$.

To see why, for any θ, let $x^*(\theta)$ denote the solution to the problem: max $F(x, \theta)$ subject to $c \geq G(x, \theta)$, and let $\lambda^*(\theta)$ be the associated Lagrange multiplier.

Theorem (Envelope) Let F and G be continuously differentiable functions of x and θ. For any given θ, let $x^*(\theta)$ maximize $F(x, \theta)$ subject to $c \geq G(x, \theta)$, and let $\lambda^*(\theta)$ be the associated value of the Lagrange multiplier. Suppose, further, that $x^*(\theta)$ and $\lambda^*(\theta)$ are also continuously differentiable functions, and that the constraint qualification $G_1[x^*(\theta), \theta] \neq 0$ holds for all values of θ. Then the maximum value function defined by

$$V(\theta) = \max_x F(x, \theta) \text{ subject to } c \geq G(x, \theta)$$

satisfies

$$V'(\theta) = F_2[x^*(\theta), \theta] - \lambda^*(\theta) G_2[x^*(\theta), \theta]. \tag{7}$$

Proof The Kuhn-Tucker theorem tells us that for any given value of θ, $x^*(\theta)$ and $\lambda^*(\theta)$ must satisfy

$$L_1[x^*(\theta), \lambda^*(\theta)] = F_1[x^*(\theta), \theta] - \lambda^*(\theta) G_1[x^*(\theta), \theta] = 0, \tag{1}$$

and

$$\lambda^*(\theta)\{c - G[x^*(\theta), \theta]\} = 0. \tag{4}$$

In light of (4),

$$V(\theta) = F[x^*(\theta), \theta] = F[x^*(\theta), \theta] + \lambda^*(\theta)\{c - G[x^*(\theta), \theta]\}$$

Differentiating both sides of this expression with respect to θ yields

$$\begin{aligned} V'(\theta) &= F_1[x^*(\theta), \theta]x^{*\prime}(\theta) + F_2[x^*(\theta), \theta] \\ &\quad + \lambda^{*\prime}(\theta)\{c - G[x^*(\theta), \theta]\} - \lambda^*(\theta)G_1[x^*(\theta), \theta]x^{*\prime}(\theta) - \lambda^*(\theta)G_2[x^*(\theta), \theta] \end{aligned}$$

which shows that, in principle, we must take the dependence of x^* and λ^* on θ into account when calculating $V'(\theta)$.

Note, however, that

$$\begin{aligned} V'(\theta) &= \{F_1[x^*(\theta), \theta] - \lambda^*(\theta)G_1[x^*(\theta), \theta]\}x^{*\prime}(\theta) \\ &\quad + F_2[x^*(\theta), \theta] + \lambda^{*\prime}(\theta)\{c - G[x^*(\theta), \theta]\} - \lambda^*(\theta)G_2[x^*(\theta), \theta], \end{aligned}$$

which by (1) reduces to

$$V'(\theta) = F_2[x^*(\theta), \theta] + \lambda^{*\prime}(\theta)\{c - G[x^*(\theta), \theta]\} - \lambda^*(\theta)G_2[x^*(\theta), \theta]$$

Thus, it only remains to show that

$$\lambda^{*\prime}(\theta)\{c - G[x^*(\theta), \theta]\} = 0 \tag{8}$$

Clearly, (8) holds for any θ such that the constraint is binding.

For θ such that the constraint is not binding, (4) implies that $\lambda^*(\theta)$ must equal zero. Furthermore, by the continuity of G and x^*, if the constraint does not bind at θ, there exists an $\varepsilon^* > 0$ such that the constraint does not bind for all $\theta + \varepsilon$ with $\varepsilon^* > |\varepsilon|$. Hence, (4) also implies that $\lambda^*(\theta + \varepsilon) = 0$ for all $\varepsilon^* > |\varepsilon|$. Using the definition of the derivative

$$\lambda^{*\prime}(\theta) = \lim_{\varepsilon \to 0} \frac{\lambda^*(\theta + \varepsilon) - \lambda^*(\theta)}{\varepsilon} = \lim_{\varepsilon \to 0} \frac{0}{\varepsilon} = 0,$$

it once again becomes apparent that (8) must hold.

Thus,

$$V'(\theta) = F_2[x^*(\theta), \theta] - \lambda^*(\theta)G_2[x^*(\theta), \theta]$$

as claimed in the theorem.

Once again, this theorem is useful because it tells us that we can ignore the dependence of x^* and λ^* on θ in calculating $V'(\theta)$.

And once again, the theorem can be extended to apply in more general settings: see Dixit, Simon-Blume, Acemoglu, or section 4.2 of the notes below.

It is worth noting that the assumptions required by the envelope theorem are more restrictive than those required by the Kuhn-Tucker theorem. For example, the statement of the Kuhn-Tucker theorem makes reference to the value x^* of x that solves the constrained optimization problem, thereby assuming implicitly that a solution to the problem exists. But the envelope theorem requires that x^* be depicted as a function of θ, implying that the solution must not only exist but be unique as well. Further, the solution must vary smoothly with θ, so that x^* and λ^* can be written as continuously differentiable functions of the parameter.

Although most statements of the envelope theorem assume directly that $x^*(\theta)$ and $\lambda^*(\theta)$ are continuously differentiable, it is also possible to impose restrictions on $F(x, \theta)$ and $G(x, \theta)$ that guarantee this. Assume, for example, that the constraint always binds at the optimum, so that by the Kuhn-Tucker theorem, $x^*(\theta)$ and $\lambda^*(\theta)$ must satisfy

$$L_1[x^*(\theta), \lambda^*(\theta)] = F_1[x^*(\theta), \theta] - \lambda^*(\theta)G_1[x^*(\theta), \theta] = 0$$

and

$$L_2[x^*(\theta), \lambda^*(\theta)] = c - G[x^*(\theta), \theta] = 0.$$

If F and G are two times continuously differentiable in x and θ, and if the matrix of second derivatives of the Lagrangian

$$\begin{bmatrix} L_{11}[x^*(\theta), \lambda^*(\theta)] & L_{12}[x^*(\theta), \lambda^*(\theta)] \\ L_{21}[x^*(\theta), \lambda^*(\theta)] & L_{22}[x^*(\theta), \lambda^*(\theta)] \end{bmatrix}$$

is nonsingular, then the implicit function theorem (Simon and Blume, Theorem 15.7, pp.355-356) will imply that $x^*(\theta)$ and $\lambda^*(\theta)$ exist and are continuously differentiable.

But what is the intuition for why the envelope theorem holds? To obtain some intuition, begin by considering the simpler, unconstrained optimization problem:

$$\max_x F(x, \theta),$$

where x is the choice variable and θ is the parameter.

Associated with this unconstrained problem, define the maximum value function in the same way as before:

$$V(\theta) = \max_x F(x, \theta).$$

To evaluate V for any given value of θ, use the same two-step procedure as before. First, find the value $x^*(\theta)$ that solves the unconstrained maximization problem for that value of θ. Second, substitute that value of x back into the objective function to obtain

$$V(\theta) = F[x^*(\theta), \theta].$$

Now differentiate both sides of this expression through by θ, carefully taking the dependence of x^* on θ into account:

$$V'(\theta) = F_1[x^*(\theta), \theta]x^{*\prime}(\theta) + F_2[x^*(\theta), \theta].$$

But, if $x^*(\theta)$ is the value of x that maximizes F given θ, we know that $x^*(\theta)$ must be a critical point of F:

$$F_1[x^*(\theta), \theta] = 0.$$

Hence, for the unconstrained problem, the envelope theorem implies that

$$V'(\theta) = F_2[x^*(\theta), \theta],$$

so that, again, we can ignore the dependence of x^* on θ in differentiating the maximum value function. And this result holds not because x^* fails to depend on θ: to the contrary, in fact, x^* will typically depend on θ through the function $x^*(\theta)$. Instead, the result holds because since x^* is chosen optimally, $x^*(\theta)$ is a critical point of F given θ.

Now return to the constrained optimization problem

$$\max_x F(x, \theta) \text{ subject to } c \geq G(x, \theta)$$

and define the maximum value function as before:

$$V(\theta) = \max_x F(x, \theta) \text{ subject to } c \geq G(x, \theta).$$

The envelope theorem for this constrained problem tells us that we can also ignore the dependence of x^* on θ when differentiating V with respect to θ, but only if we start by adding the complementary slackness condition to the maximized objective function to first obtain

$$V(\theta) = F[x^*(\theta), \theta] + \lambda^*(\theta)\{c - G[x^*(\theta), \theta]\}.$$

In taking this first step, we are actually evaluating the entire Lagrangian at the optimum, instead of just the objective function. We need to take this first step because for the constrained problem, the Kuhn-Tucker condition (1) tells us that $x^*(\theta)$ is a critical point, not of the objective function by itself, but of the entire Lagrangian formed by adding the product of the multiplier and the constraint to the objective function.

And what gives the envelope theorem its name? The "envelope" theorem refers to a geometrical presentation of the same result that we've just worked through.

To see where that geometrical interpretation comes from, consider again the simpler, unconstrained optimization problem:

$$\max_x F(x, \theta),$$

where x is the choice variable and θ is a parameter.

Following along with our previous notation, let $x^*(\theta)$ denote the solution to this problem for any given value of θ, so that the function $x^*(\theta)$ tells us how the optimal choice of x depends on the parameter θ.

Also, continue to define the maximum value function V in the same way as before:

$$V(\theta) = \max_x F(x, \theta).$$

Now let θ_1 denote a particular value of θ, and let x_1 denote the optimal value of x associated with this particular value θ_1. That is, let

$$x_1 = x^*(\theta_1).$$

After substituting this value of x_1 into the function F, we can think about how $F(x_1, \theta)$ varies as θ varies—that is, we can think about $F(x_1, \theta)$ as a function of θ, holding x_1 fixed.

In the same way, let θ_2 denote another particular value of θ, with $\theta_2 > \theta_1$ let's say. And following the same steps as above, let x_2 denote the optimal value of x associated with this particular value θ_2, so that

$$x_2 = x^*(\theta_2).$$

Once again, we can hold x_2 fixed and consider $F(x_2, \theta)$ as a function of θ.

The geometrical presentation of the envelope theorem can be derived by thinking about the properties of these three functions of θ: $V(\theta)$, $F(x_1, \theta)$, and $F(x_2, \theta)$.

One thing that we know about these three functions is that for $\theta = \theta_1$:

$$V(\theta_1) = F(x_1, \theta_1) > F(x_2, \theta_1),$$

where the first equality and the second inequality both follow from the fact that, by definition, x_1 maximizes $F(x, \theta_1)$ by choice of x.

Another thing that we know about these three functions is that for $\theta = \theta_2$:

$$V(\theta_2) = F(x_2, \theta_2) > F(x_1, \theta_2),$$

because again, by definition, x_2 maximizes $F(x, \theta_2)$ by choice of x.

On a graph, these relationships imply that:

At θ_1, $V(\theta)$ coincides with $F(x_1, \theta)$, which lies above $F(x_2, \theta)$.

At θ_2, $V(\theta)$ coincides with $F(x_2, \theta)$, which lies above $F(x_1, \theta)$.

And we could find more and more values of V by repeating this procedure for more and more specific values of θ_i, $i = 1, 2, 3, \ldots$.

In other words:

$V(\theta)$ traces out the "upper envelope" of the collection of functions $F(x_i, \theta)$, formed by holding $x_i = x^*(\theta_i)$ fixed and varying θ.

Moreover, $V(\theta)$ is tangent to each individual function $F(x_i, \theta)$ at the value θ_i of θ for which x_i is optimal, or equivalently:

$$V'(\theta) = F_2[x^*(\theta), \theta],$$

which is the same analytical result that we derived earlier for the unconstrained optimization problem.

If, for example,

$$F(x, \theta) = -(x-\theta)^2 + \theta^2 = -x^2 + 2x\theta,$$

then

$$V(\theta) = \max_x -(x-\theta)^2 + \theta^2 = \theta^2,$$

since, in this case, $x^*(\theta) = \theta$ for all values of θ.

The figure below sets $\theta_1 = 2$ and $\theta_2 = 7$; hence $x_1 = 2$ and $x_2 = 7$, then plots

$$F(x_1, \theta) = -4 + 4\theta,$$

$$F(x_2, \theta) = -49 + 14\theta,$$

and

$$V(\theta) = \theta^2$$

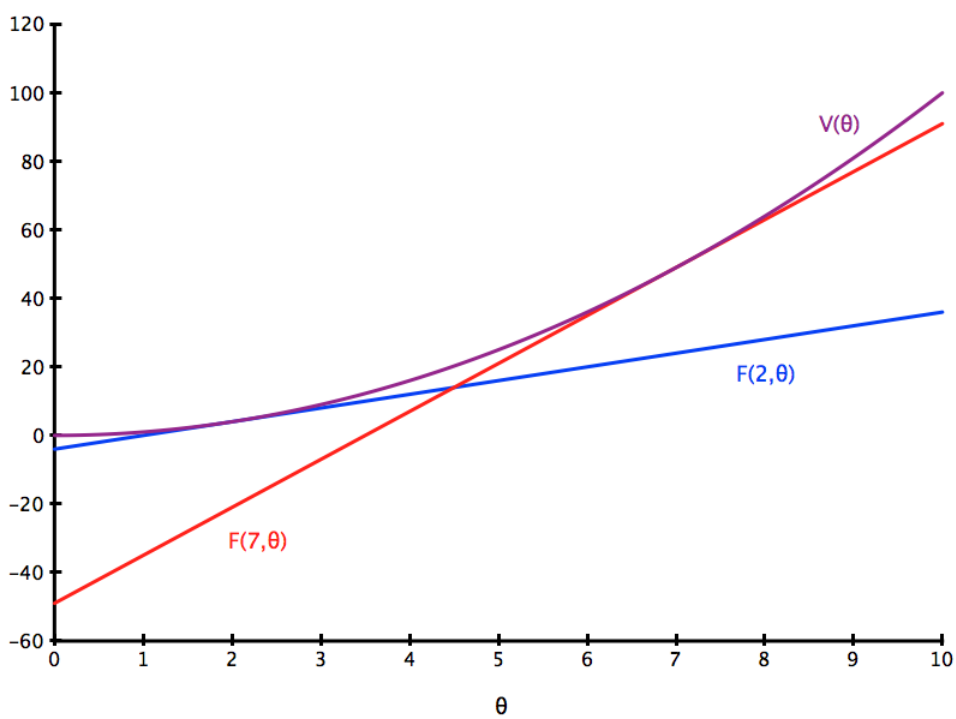
to show how

$$V(\theta_1) = F(x_1, \theta_1) > F(x_2, \theta_1) \text{ at } \theta_1 = 2,$$

and

$$V(\theta_2) = F(x_2, \theta_2) > F(x_1, \theta_2) \text{ at } \theta_2 = 7,$$

and how, more generally, $V(\theta)$ traces out the upper envelope of the family of functions $F(x_i, \theta)$, where each x_i maximizes $F(x, \theta)$ for some value θ_i of θ.



To generalize these arguments so that they apply to the constrained optimization problem

$$\max_x F(x, \theta) \text{ subject to } c \geq G(x, \theta),$$

simply use the fact that in many cases (as when F is concave and G is convex) the value $x^*(\theta)$ that solves the constrained optimization problem for any given value of θ also maximizes the Lagrangian function

$$L(x, \lambda, \theta) = F(x, \theta) + \lambda[c - G(x, \theta)],$$

so that

$$\begin{aligned} V(\theta) &= \max_x F(x, \theta) \text{ subject to } c \geq G(x, \theta) \\ &= \max_x L(x, \lambda, \theta) \end{aligned}$$

Now just replace the function F with the function L in working through the arguments from above to conclude that

$$V'(\theta) = L_3[x^*(\theta), \lambda^*(\theta), \theta] = F_2[x^*(\theta), \theta] - \lambda^*(\theta) G_2[x^*(\theta), \theta],$$

which is again the same result that we derived before for the constrained optimization problem.

Note that in the figure, the maximum value function $V(\theta)$ is convex, with a strictly positive second derivative, whereas the functions $F(x_1, \theta)$ and $F(x_2, \theta)$ are both linear. In general, it can be shown that there is a sense in which the maximum value function $V(\theta)$ will always be "more convex" or "less concave" than each member of the family of functions $F(x_i, \theta)$.

To see this, consider the two functions $V(\theta)$ and $F(x_1, \theta)$ as they were defined for the general unconstrained problem

$$\max_x F(x, \theta).$$

Second order Taylor approximations imply that for values of θ sufficiently close to θ_1:

$$V(\theta) \approx V(\theta_1) + V'(\theta_1)(\theta - \theta_1) + (1/2)V''(\theta_1)(\theta - \theta_1)^2$$

and

$$F(x_1, \theta) \approx F(x_1, \theta_1) + F_2(x_1, \theta_1)(\theta - \theta_1) + (1/2)F_{22}(x_1, \theta_1)(\theta - \theta_1)^2.$$

However,

$$V(\theta_1) = F(x_1, \theta_1),$$

since $x_1 = x^*(\theta_1)$, and

$$V'(\theta_1) = F_2(x_1, \theta_1),$$

by the envelope theorem itself. Since the definition of V implies that

$$V(\theta) \geq F(x_1, \theta)$$

for all values of θ, the Taylor approximations imply

$$V''(\theta_1) \geq F_{22}(x_1, \theta_1),$$

so that, more specifically, the second derivative of the maximum value function will always be larger than the second derivatives of the functions $F(x_i, \theta)$.

Finally, to begin getting a feel for the usefulness of the envelope theorem, consider a preliminary economic example. Suppose that a firm hires n workers, paying each the competitive real wage w (real wage, so that we don't have to consider separately the price of output), in order to produce y units of output according to the production function

$$n^\alpha \geq y,$$

where α lies between zero and one: $0 < \alpha < 1$.

For simplicity, let's depict the firm as solving the unconstrained optimization problem

$$\max_n n^\alpha - wn,$$

with first-order condition

$$\alpha(n^*)^{\alpha-1} - w = 0$$

that leads to the labor demand curve

$$n^*(w) = \left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha-1}} w^{\frac{1}{\alpha-1}}.$$

Next, define the maximum value function

$$V(w) = \max_n n^\alpha - wn,$$

so that

$$V(w) = [n^*(w)]^\alpha - wn^*(w).$$

The envelope theorem immediately implies that

$$V'(w) = -n^*(w) = -\left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha-1}} w^{\frac{1}{\alpha-1}}.$$

Suppose instead we substitute our earlier solution for $n^*(w)$ into the expression for $V(w)$:

$$\begin{aligned}
V(w) &= [n^*(w)]^\alpha - wn^*(w) \\
&= \left[\left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha-1}} w^{\frac{1}{\alpha-1}}\right]^\alpha - w\left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha-1}} w^{\frac{1}{\alpha-1}} \\
&= \left(\frac{1}{\alpha}\right)^{\frac{\alpha}{\alpha-1}} w^{\frac{\alpha}{\alpha-1}} - \left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha-1}} w^{\frac{\alpha}{\alpha-1}} \\
&= \left[\left(\frac{1}{\alpha}\right)^{\frac{\alpha}{\alpha-1}} - \left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha-1}}\right] w^{\frac{\alpha}{\alpha-1}}.
\end{aligned}$$

Now differentiate with respect to w and simplify to get

$$\begin{aligned} V'(w) &= \left(\frac{\alpha}{\alpha-1}\right)\left[\left(\frac{1}{\alpha}\right)^{\frac{\alpha}{\alpha-1}} - \left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha-1}}\right] w^{\frac{1}{\alpha-1}} \\ &= \left(\frac{\alpha}{\alpha-1}\right)\left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha-1}}\left(\frac{1}{\alpha}-1\right) w^{\frac{1}{\alpha-1}} \\ &= -\left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha-1}} w^{\frac{1}{\alpha-1}}, \end{aligned}$$

exactly as we found, much more quickly, using the envelope theorem.

In terms of its economics, this example shows that when the firm faces a higher wage, there will be two effects on its profits.

The direct effect: it must pay each of its workers a higher wage, so profits fall by $-n^*(w)$.

The indirect effect: it can try to mitigate the direct effect by hiring fewer workers.

The envelope theorem says that since the firm has already chosen n^* optimally, the first-order effect on profits of adjusting this decision in response to an arbitrarily small change in the wage is zero. Only the direct effect remains: $V'(w) = -n^*(w)$.

3 Three Examples

3.1 Utility Maximization

A consumer has a utility function defined over consumption of two goods: $U(c_1, c_2)$

Prices: p_1 and p_2

Income: I

Budget constraint: $I \geq p_1c_1 + p_2c_2 = G(c_1, c_2)$

The consumer's problem is:

$$\max_{c_1,c_2} U(c_1, c_2) \text{ subject to } I \geq p_1c_1 + p_2c_2$$

The Kuhn-Tucker theorem tells us that if we set up the Lagrangian:

$$L(c_1, c_2, \lambda) = U(c_1, c_2) + \lambda(I - p_1c_1 - p_2c_2)$$

Then the optimal consumptions c_1^* and c_2^* and the associated multiplier λ^* must satisfy the FOC:

$$L_1(c_1^*, c_2^*, \lambda^*) = U_1(c_1^*, c_2^*) - \lambda^* p_1 = 0$$

and

$$L_2(c_1^*, c_2^*, \lambda^*) = U_2(c_1^*, c_2^*) - \lambda^* p_2 = 0$$

Move the terms with minus signs to the other side, and divide the first of these FOC by the second to obtain

$$\frac{U_1(c_1^*, c_2^*)}{U_2(c_1^*, c_2^*)} = \frac{p_1}{p_2},$$

which is just the familiar condition that says that the optimizing consumer should set the slope of his or her indifference curve, the marginal rate of substitution, equal to the slope of his or her budget constraint, the ratio of prices.

Now consider I as one of the model's parameters, and let the functions $c_1^*(I)$, $c_2^*(I)$, and $\lambda^*(I)$ describe how the optimal choices c_1^* and c_2^* and the associated value λ^* of the multiplier depend on I.

In addition, define the maximum value function as

$$V(I) = \max_{c_1, c_2} U(c_1, c_2) \text{ subject to } I \geq p_1 c_1 + p_2 c_2$$

The Kuhn-Tucker theorem tells us that

$$\lambda^*(I)[I - p_1 c_1^*(I) - p_2 c_2^*(I)] = 0$$

and hence

$$V(I) = U[c_1^*(I), c_2^*(I)] = U[c_1^*(I), c_2^*(I)] + \lambda^*(I)[I - p_1 c_1^*(I) - p_2 c_2^*(I)].$$

The envelope theorem tells us that we can ignore the dependence of c_1^* and c_2^* on I in calculating

$$V'(I) = \lambda^*(I),$$

which gives us an interpretation of the multiplier λ^* as the marginal utility of income.

3.2 Cost Minimization

The Kuhn-Tucker and envelope conditions can also be used to study constrained minimization problems.

Consider a firm that produces output y using capital k and labor l, according to the technology described by

$$f(k, l) \geq y.$$

r = rental rate for capital

w = wage rate

Suppose that the firm takes its output y as given, and chooses inputs k and l to minimize costs. Then the firm solves

$$\min_{k,l} rk + wl \text{ subject to } f(k, l) \geq y$$

If we set up the Lagrangian as

$$L(k,l,\lambda) = rk + wl - \lambda[f(k,l) - y],$$

where the term involving the multiplier λ is subtracted rather than added in the case of a minimization problem, the Kuhn-Tucker conditions (1)-(4) continue to apply, exactly as before.

Thus, according to the Kuhn-Tucker theorem, the optimal choices k^* and l^* and the associated multiplier λ^* must satisfy the FOC:

$$L_1(k^*, l^*, \lambda^*) = r - \lambda^* f_1(k^*, l^*) = 0 \tag{9}$$

and

$$L_2(k^*, l^*, \lambda^*) = w - \lambda^* f_2(k^*, l^*) = 0 \tag{10}$$

Move the terms with minus signs over to the other side, and divide the first FOC by the second to obtain

$$\frac{f_1(k^*, l^*)}{f_2(k^*, l^*)} = \frac{r}{w},$$

which is another familiar condition that says that the optimizing firm chooses factor inputs so that the marginal rate of substitution between inputs in production equals the ratio of factor prices.

Now suppose that the constraint binds, as it usually will:

$$y = f(k^*, l^*) \tag{11}$$

Then (9)-(11) represent 3 equations that determine the three unknowns k^*, l^*, and λ^* as functions of the model's parameters r, w, and y. In particular, we can think of the functions

$$k^* = k^*(r, w, y)$$

and

$$l^* = l^*(r, w, y)$$

as demand curves for capital and labor: strictly speaking, they are conditional (on y) factor demand functions.

Now define the minimum cost function as

$$\begin{aligned}
C(r, w, y) &= \min_{k,l} rk + wl \text{ subject to } f(k,l) \geq y \\
&= rk^*(r, w, y) + wl^*(r, w, y) \\
&= rk^*(r, w, y) + wl^*(r, w, y) - \lambda^*(r, w, y)\{f[k^*(r, w, y), l^*(r, w, y)] - y\}
\end{aligned}$$

The envelope theorem tells us that in calculating the derivatives of the cost function, we can ignore the dependence of k^*, l^*, and λ^* on r, w, and y.

Hence:

$$C_1(r, w, y) = k^*(r, w, y),$$

$$C_2(r, w, y) = l^*(r, w, y),$$

and

$$C_3(r, w, y) = \lambda^*(r, w, y).$$

The first two of these equations are statements of Shephard's lemma; they tell us that the derivatives of the cost function with respect to factor prices coincide with the conditional factor demand curves. The third equation gives us an interpretation of the multiplier λ^* as a measure of the marginal cost of increasing output.

Thus, our two examples illustrate how we can apply the Kuhn-Tucker and envelope theorems in specific economic problems.

The two examples also show how, in the context of specific economic problems, it is often possible to attach an economic interpretation to the multiplier λ^*.

3.3 Le Chatelier's Principle

For the next example, extend the previous cost minimization problem by introducing a third input:

m = materials input

q = price of materials

Define the minimum cost function

$$C(r, w, q, y) = \min_{k,l,m} rk + wl + qm \text{ subject to } f(k, l, m) \geq y.$$

Then Shephard's lemma implies

$$C_1(r, w, q, y) = k^*(r, w, q, y),$$

$$C_2(r, w, q, y) = l^*(r, w, q, y),$$

and

$$C_3(r, w, q, y) = m^*(r, w, q, y),$$

where $k^*(r, w, q, y)$, $l^*(r, w, q, y)$, and $m^*(r, w, q, y)$ are the conditional factor demand curves and the envelope theorem also implies that

$$C_4(r, w, q, y) = \lambda^*(r, w, q, y),$$

so that the Lagrange multiplier $\lambda^*(r, w, q, y)$ is again a measure of marginal cost.

Next, let's use Shephard's lemma to deduce another property of the conditional factor demand curves. Since

$$C_1(r,w,q,y) = k^*(r,w,q,y),$$

it follows that

$$C_{12}(r,w,q,y) = k_2^*(r,w,q,y).$$

And since

$$C_2(r,w,q,y) = l^*(r,w,q,y),$$

it follows that

$$C_{21}(r,w,q,y) = l_1^*(r,w,q,y).$$

But, since the symmetry of the second partial derivatives of C implies

$$C_{12}(r,w,q,y) = C_{21}(r,w,q,y),$$

Shephard's lemma can be viewed as having, as a corollary, the "reciprocity" condition for conditional factor demand curves:

$$k_2^*(r,w,q,y) = l_1^*(r,w,q,y),$$

or, equivalently,

$$\frac{\partial k^*(r,w,q,y)}{\partial w} = \frac{\partial l^*(r,w,q,y)}{\partial r}.$$

Now consider a "short-run" version of the firm's problem, treating the capital stock k as fixed at some pre-determined level $\bar{k}$:

$$\min_{l,m} r\bar{k} + wl + qm \text{ subject to } f(\bar{k},l,m) \geq y.$$

With the Lagrangian for the short-run problem defined as

$$L(l,m,\mu) = r\bar{k} + wl + qm - \mu[f(\bar{k},l,m) - y],$$

the first-order conditions are

$$w - \mu^s f_2(\bar{k},l^s,m^s) = 0$$

and

$$q - \mu^s f_3(\bar{k},l^s,m^s) = 0$$

and the binding constraint is

$$f(\bar{k},l^s,m^s) - y = 0.$$

Interestingly, none of these optimality conditions depends on the rental rate r for capital. Hence, we can solve for short-run conditional factor demand curves of the form

$$l^s = l^s(w,q,y,\bar{k})$$

and

$$m^s = m^s(w, q, y, \bar{k}),$$

and use another application of the envelope theorem, this time to the short-run problem, to interpret

$$\mu^s = \mu^s(w, q, y, \bar{k})$$

as a measure of "short-run marginal cost."

Le Chatelier's principle says that labor demand should be more responsive to a change in the wage rate w in the long run than in the short run since, intuitively, in the long run the firm has the chance to substitute capital for labor in response to that change in the wage.

To show the sense in which this conjecture proves true, suppose that the fixed, short-run value of $\bar{k}$ just happens to equal the optimal value k^* that would have been chosen anyway in the long run:

$$\bar{k} = k^*(r, w, q, y).$$

In this case, a comparison of the Kuhn-Tucker conditions for the short-run problem with those for the long-run problem confirms that the firm will choose a value of l^s in the short run equal to the long-run value l^*:

$$l^*(r, w, q, y) = l^s[w, q, y, k^*(r, w, q, y)].$$

Differentiate both sides of this expression with respect to w to obtain

$$l_2^*(r, w, q, y) = l_1^s[w, q, y, k^*(r, w, q, y)] + l_4^s[w, q, y, k^*(r, w, y)]k_2^*(r, w, q, y).$$

The term on the left-hand side of this expression measures the long-run responsiveness of labor demand to a change in the wage; the first term on the right-hand side of this expression measures the short-run responsiveness of labor demand to a change in the wage. The second-order conditions for the long-run and short-run problems will generally imply that under "typical" conditions, both of these terms are negative: when the wage goes up, labor demand falls. But, is the long-run response "more negative?" This depends on the sign of the second term on the right-hand side, which in turn depends on l_4^s, measuring the response of short-run labor demand to a change in $\bar{k}$, and k_2^*, measuring the response of long-run capital demand to a change in w. The signs of both terms would seem to be ambiguous, dependent in some loose sense on whether, in the long run, the capital and labor inputs are "complements" or "substitutes."

Yet, as it turns out, something more definite can be said. Return once more to

$$l^*(r, w, q, y) = l^s[w, q, y, k^*(r, w, y)],$$

but this time differentiate both sides with respect to r to obtain

$$l_1^*(r, w, q, y) = l_4^s[w, q, y, k^*(r, w, y)]k_1^*(r, w, q, y).$$

Rearrange this expression so that it reads

$$l_4^s[w, q, y, k^*(r, w, q, y)] = \frac{l_1^*(r, w, q, y)}{k_1^*(r, w, q, y)},$$

and substitute this result into the previous one

$$l_2^*(r, w, q, y) = l_1^s[w, q, y, k^*(r, w, q, y)] + l_4^s[w, q, y, k^*(r, w, q, y)]k_2^*(r, w, q, y).$$

to get

$$l_2^*(r, w, q, y) = l_1^s[w, q, y, k^*(r, w, q, y)] + \frac{l_1^*(r, w, q, y)k_2^*(r, w, q, y)}{k_1^*(r, w, q, y)}.$$

Now it is clear that the second term must be negative. This is because the expression in the numerator is, by the reciprocity condition implied by Shephard's lemma, equal to the perfect square $[l_1^*(r, w, q, y)]^2$; meanwhile, the term in the denominator will be negative, as implied by the second-order conditions to the long-run problem: the optimal capital stock falls when the rental rate for capital rises.

Thus, it must be that

$$l_2^*(r, w, q, y) \leq l_1^s[w, q, y, k^*(r, w, y)]$$

and, since both of these terms are negative,

$$\left| \frac{\partial l^*(r, w, q, y)}{\partial w} \right| \geq \left| \frac{\partial l^s(w, q, y, k^*)}{\partial w} \right|,$$

which is the relationship we were after: labor demand is more responsive to a change in the wage in the long run than in the short run. Note, however, that this is a result that only holds locally, when $\bar{k}$ is sufficiently close to $k^*(r, w, q, y)$.

4 Generalizing the Basic Results

4.1 The Kuhn-Tucker Theorem

Our "simple" version of the Kuhn-Tucker theorem applies to a problem with only one choice variable and one constraint.

Section 19.6 of Simon and Blume's book develops a proof for the more general case, with n choice variables and m constraints. Their proof makes repeated, clever use of the implicit function theorem, which makes the arguments surprisingly short but also works to obscure some of the intuition provided by the analysis of the simplest case.

Nevertheless, having gained the intuition the intuition from working through the simple case, it is useful to see how the result extends.

Simon and Blume (Chapter 15) and Acemoglu (Appendix A) both present fairly general statements of the implicit function theorem. The special case or application of their results that we will need works as follows.

Consider a system of n equations, involving n "endogenous" variables $y_1, y_2, \dots, y_n$ and n "exogenous" variables $c_1, c_2, \dots c_n$:

$$\begin{aligned} H_1(y_1, y_2, \dots, y_n) &= c_1, \\ H_2(y_1, y_2, \dots, y_n) &= c_2, \\ &\vdots \\ H_m(y_1, y_2, \dots, y_n) &= c_n. \end{aligned}$$

Now suppose that for a specific set of values $c_1^*, c_2^*, \dots c_n^*$ for the exogenous variables, all the equations in the system are satisfied with the endogenous variables set equal to $y_1^*, y_2^*, \dots, y_n^*$, so that

$$\begin{aligned} H_1(y_1^*, y_2^*, \dots, y_n^*) &= c_1^*, \\ H_2(y_1^*, y_2^*, \dots, y_n^*) &= c_2^*, \\ &\vdots \\ H_n(y_1^*, y_2^*, \dots, y_n^*) &= c_n^*. \end{aligned}$$

Assume that each function H_i, $i = 1, \dots, n$, is continuously differentiable and that the $n \times n$ matrix of derivatives

$$\begin{bmatrix} \partial H_1/\partial y_1 & \cdots & \partial H_1/\partial y_n \\ \partial H_2/\partial y_1 & \cdots & \partial H_2/\partial y_n \\ \vdots & \ddots & \vdots \\ \partial H_n/\partial y_1 & \cdots & \partial H_n/\partial y_n \end{bmatrix}$$

is nonsingular at $y_1^*, y_2^*, \dots, y_n^*$.

Then there exist continuously differentiable functions

$$\begin{gathered} y_1(c_1, c_2, \dots, c_n), \\ y_2(c_1, c_2, \dots, c_n), \\ \vdots \\ y_n(c_1, c_2 \dots, c_n), \end{gathered}$$

defined in an open subset C of $\mathbf{R}^n$ containing $(c_1^*, c_2^*, \dots, c_n^*)$, such that

$$\begin{aligned} H_1(y_1(c_1, c_2, \dots, c_n), y_2(c_1, c_2, \dots, c_n), \dots, y_n(c_1, c_2, \dots, c_n)) &= c_1, \\ H_2(y_1(c_1, c_2, \dots, c_n), y_2(c_1, c_2, \dots, c_n), \dots, y_n(c_1, c_2, \dots, c_n)) &= c_2, \\ &\vdots \\ H_n(y_1(c_1, c_2, \dots, c_n), y_2(c_1, c_2, \dots, c_n), \dots, y_n(c_1, c_2, \dots, c_n)) &= c_n. \end{aligned}$$

for all $(c_1, c_2, \dots, c_n) \in C$.

With this result in hand, consider the following generalized version of the Kuhn-Tucker theorem we proved earlier. Let there be n choice variables, $x_1, x_2, \ldots, x_n$. The objective function $F : \mathbf{R}^n \to \mathbf{R}$ is continuously differentiable, as are the m functions $G_j : \mathbf{R}^n \to \mathbf{R}$, $j = 1, 2, \ldots, m$ that enter into the constraints

$$c_j \geq G_j(x_1, x_2, \ldots, x_n),$$

where $c_j \in \mathbf{R}$ for all $j = 1, 2, \ldots, m$.

The problem can be stated as:

$$\max_{x_1, x_2, \ldots, x_n} F(x_1, x_2, \ldots, x_n) \text{ subject to } c_j \geq G_j(x_1, x_2, \ldots, x_n) \text{ for all } j = 1, 2, \ldots, m.$$

Note that, typically, $m \leq n$ will have to hold so that there is a set of values for the choice variables that satisfy all of the constraints.

To define the Lagrangian, introduce the multipliers λ_j, $j = 1, 2, \ldots, m$, one for each constraint. Then

$$L(x_1, x_2, \ldots, x_n, \lambda_1, \lambda_2, \ldots, \lambda_m) = F(x_1, x_2, \ldots, x_n) + \sum_{j=1}^{m} \lambda_j [c_j - G_j(x_1, x_2, \ldots, x_n)].$$

Theorem (Kuhn-Tucker) Suppose that $x_1^*, x_2^*, \ldots, x_n^*$ maximize $F(x_1, x_2, \ldots, x_n)$ subject to $c_j \geq G_j(x_1, x_2, \ldots, x_n)$ for all $j = 1, 2, \ldots, m$, where F and the G_j's are all continuously differentiable. Suppose (without loss of generality) that the first $\bar{m} \leq m$ constraints bind at the optimum and that the remaining $m - \bar{m} \geq 0$ constraints are nonbinding, and assume that the $\bar{m} \times n$ matrix of derivatives

$$\begin{bmatrix} G_{1,1}(x_1^*, x_2^*, \ldots, x_n^*) & \ldots & G_{1,n}(x_1^*, x_2^*, \ldots, x_n^*) \\ G_{2,1}(x_1^*, x_2^*, \ldots, x_n^*) & \ldots & G_{2,n}(x_1^*, x_2^*, \ldots, x_n^*) \\ \vdots & \ddots & \vdots \\ G_{\bar{m},1}(x_1^*, x_2^*, \ldots, x_n^*) & \ldots & G_{\bar{m},n}(x_1^*, x_2^*, \ldots, x_n^*) \end{bmatrix}, \tag{12}$$

where $G_{j,i} = \partial G_j / \partial x_i$, has rank $\bar{m}$. Then there exist values $\lambda_1^*, \lambda_2^*, \ldots, \lambda_m^*$ that, together with $x_1^*, x_2^*, \ldots, x_n^*$, satisfy:

$$\begin{aligned} L_i(x_1^*, x_2^*, \ldots, x_n^*, \lambda_1^*, \lambda_2^*, \ldots, \lambda_n^*) = {} & F_i(x_1^*, x_2^*, \ldots, x_n^*) \\ & - \sum_{j=1}^{m} \lambda_j^* G_{j,i}(x_1^*, x_2^*, \ldots, x_n^*) = 0 \end{aligned} \tag{13}$$

for $i = 1, 2, \ldots, n$,

$$L_{n+j}(x_1^*, x_2^*, \ldots, x_n^*, \lambda_1^*, \lambda_2^*, \ldots, \lambda_n^*) = c_j - G_j(x_1^*, x_2^*, \ldots, x_n^*) \geq 0, \tag{14}$$

for $j = 1, 2, \ldots, m$,

$$\lambda_j^* \geq 0, \tag{15}$$

for $j = 1, 2, \ldots, m$, and

$$\lambda_j^* [c_j - G_j(x_1^*, x_2^*, \ldots, x_n^*)] = 0, \tag{16}$$

for $j = 1, 2, \ldots, m$.

Proof To begin, set the multipliers $\lambda^*_{\bar{m}+1}, \lambda^*_{\bar{m}+2}, \ldots, \lambda^*_m$ associated with the nonbinding contraints equal to zero. Since each of the functions G_j, $j = \bar{m}+1, \bar{m}+2, \ldots, m$, is continuously differentiable, sufficiently small adjustments in the choice variables can be made without violating these $m - \bar{m}$ constraints or causing any of them to become binding.

Next, note that the $\bar{m}+1 \times n$ matrix

$$\begin{bmatrix} F_1(x_1^*, x_2^*, \ldots, x_n^*) & \ldots & F_n(x_1^*, x_2^*, \ldots, x_n^*) \\ G_{1,1}(x_1^*, x_2^*, \ldots, x_n^*) & \ldots & G_{1,n}(x_1^*, x_2^*, \ldots, x_n^*) \\ G_{2,1}(x_1^*, x_2^*, \ldots, x_n^*) & \ldots & G_{2,n}(x_1^*, x_2^*, \ldots, x_n^*) \\ \vdots & \ddots & \vdots \\ G_{\bar{m},1}(x_1^*, x_2^*, \ldots, x_n^*) & \ldots & G_{\bar{m},n}(x_1^*, x_2^*, \ldots, x_n^*) \end{bmatrix}. \tag{17}$$

must have rank $\bar{m} < \bar{m}+1$. To see why, consider the system of equations

$$\begin{aligned} F(x_1, x_2, \ldots, x_n) &= y^* \\ G_1(x_1, x_2, \ldots, x_n) &= c_1 \\ G_2(x_1, x_2, \ldots, x_n) &= c_2 \\ &\vdots \\ G_{\bar{m}}(x_1, x_2, \ldots, x_n) &= c_{\bar{m}}. \end{aligned}$$

With y^* set equal to the maximized value of the objective function,

$$y^* = F(x_1^*, x_2^*, \ldots, x_n^*),$$

each of these $\bar{m}+1$ equations holds when the functions are evaluated at $x_1^*, x_2^*, \ldots, x_n^*$. In this case, the implicit function theorem implies that it should be possible to adjust the values of $\bar{m}+1$ of the choice variables so to find a new set of values $x_1^{**}, x_2^{**}, \ldots, x_n^{**}$ such that

$$\begin{aligned} F(x_1^{**}, x_2^{**}, \ldots, x_n^{**}) &= y^* + \varepsilon \\ G_1(x_1^{**}, x_2^{**}, \ldots, x_n^{**}) &= c_1 \\ G_2(x_1^{**}, x_2^{**}, \ldots, x_n^{**}) &= c_2 \\ &\vdots \\ G_{\bar{m}}(x_1^{**}, x_2^{**}, \ldots, x_n^{**}) &= c_{\bar{m}}. \end{aligned}$$

for a strictly positive but sufficiently small value of ε. But this contradicts the assumption that $x_1^*, x_2^*, \ldots, x_n^*$ solves the constrained optimization problem.

Since the matrix in (17) has rank $\bar{m} < \bar{m}+1$, its $\bar{m}+1$ rows must be linearly dependent.

Hence, there exist scalars $\alpha_0, \alpha_1, \ldots \alpha_{\bar{m}}$, at least one of which is nonzero, such that

$$\begin{aligned}\begin{bmatrix}0\\ \vdots\\ 0\end{bmatrix} &= \alpha_0 \begin{bmatrix}F_1(x_1^*, x_2^*, \ldots, x_n^*)\\ \vdots\\ F_n(x_1^*, x_2^*, \ldots, x_n^*)\end{bmatrix}\\ &+ \alpha_1 \begin{bmatrix}G_{1,1}(x_1^*, x_2^*, \ldots, x_n^*)\\ \vdots\\ G_{1,n}(x_1^*, x_2^*, \ldots, x_n^*)\end{bmatrix} + \ldots + \alpha_{\bar{m}} \begin{bmatrix}G_{\bar{m},1}(x_1^*, x_2^*, \ldots, x_n^*)\\ \vdots\\ G_{\bar{m},n}(x_1^*, x_2^*, \ldots, x_n^*)\end{bmatrix}.\end{aligned} \tag{18}$$

Moreover, in (18), $\alpha_0 \neq 0$, since otherwise, the matrix in (12) would have rank less than $\bar{m}$.

Thus, for $j = 1, 2, \ldots, \bar{m}$, set $\lambda_j^* = -\alpha_j/\alpha_0$. With these settings for $\lambda_1^*, \lambda_2^*, \ldots, \lambda_{\bar{m}}^*$, plus the settings $\lambda_{\bar{m}+1}^* = \lambda_{\bar{m}+2}^* = \lambda_m^* = 0$ chosen earlier, (18) implies that (13) must hold for all $i = 1, 2, \ldots, n$. Clearly, (14) and (16) are satisfied for all $j = 1, 2, \ldots, m$, and (15) holds for all $j = \bar{m}+1, \bar{m}+2, \ldots, m$. So it only remains to show that (15) holds for $j = 1, 2, \ldots, \bar{m}$.

To see that these last conditions must hold, consider the system of equations

$$\begin{aligned}G_1(x_1, x_2, \ldots, x_n) &= c_1 - \delta\\ G_2(x_1, x_2, \ldots, x_n) &= c_2\\ &\vdots\\ G_{\bar{m}}(x_1, x_2, \ldots, x_n) &= c_{\bar{m}},\end{aligned} \tag{19}$$

where $\delta \geq 0$. These equations hold, with $\delta = 0$, at $x_1^*, x_2^*, \ldots, x_n^*$. And since the matrix in (12) has rank $\bar{m}$, the implicit function theorem implies that there are functions $x_1(\delta), x_2(\delta), \ldots, x_n(\delta)$ such that the same equations hold for all sufficiently small values of δ.

Since $c_1 - \delta \leq c_1$, the choices $x_1(\delta), x_2(\delta), \ldots, x_n(\delta)$ satisfy all of the constraints from the original optimization problem. And since, by assumption, $x_1(0) = x_1^*, x_2(0) = x_2^*, \ldots, x_n(0) = x_n^*$ maximizes the objective function subject to the constraints, it must be that

$$\left.\frac{dF(x_1(\delta), x_2(\delta), \ldots, x_n(\delta))}{d\delta}\right|_{\delta=0} = \sum_{i=1}^{n} F_i(x_1^*, x_2^*, \ldots, x_n^*)x_i'(0) \leq 0. \tag{20}$$

In addition, the equations in (19) implicitly defining $x_1(\delta), x_2(\delta), \ldots, x_n(\delta)$ imply

$$\left.\frac{dG_1(x_1(\delta), x_2(\delta), \ldots, x_n(\delta))}{d\delta}\right|_{\delta=0} = \sum_{i=1}^{n} G_{1,i}(x_1^*, x_2^*, \ldots, x_n^*)x_i'(0) = -1 \tag{21}$$

and

$$\left.\frac{dG_j(x_1(\delta), x_2(\delta), \ldots, x_n(\delta))}{d\delta}\right|_{\delta=0} = \sum_{i=1}^{n} G_{j,i}(x_1^*, x_2^*, \ldots, x_n^*)x_i'(0) = 0 \tag{22}$$

for $j = 2, 3, \ldots, \bar{m}$.

Putting all these results together, (13) implies

$$0 = F_i(x_1^*, x_2^*, \ldots, x_n^*) - \sum_{j=1}^{m} \lambda_j^* G_{j,i}(x_1^*, x_2^*, \ldots, x_n^*).$$

for all $i = 1, 2, \ldots, n$. Multiplying each of these equations by $x_i'(0)$ and summing over all i yields

$$0 = \sum_{i=1}^{n} F_i(x_1^*, x_2^*, \ldots, x_n^*) x_i'(0) - \sum_{i=1}^{n} \sum_{j=1}^{m} \lambda_j^* G_{j,i}(x_1^*, x_2^*, \ldots, x_n^*) x_i'(0),$$

or

$$0 = \sum_{i=1}^{n} F_i(x_1^*, x_2^*, \ldots, x_n^*) x_i'(0) - \sum_{j=1}^{m} \lambda_j^* \left[\sum_{i=1}^{n} G_{j,i}(x_1^*, x_2^*, \ldots, x_n^*) x_i'(0) \right],$$

or, since $\lambda_j^* = 0$ for $j = \bar{m}+1, \bar{m}+2, \ldots, m$,

$$0 = \sum_{i=1}^{n} F_i(x_1^*, x_2^*, \ldots, x_n^*) x_i'(0) - \sum_{j=1}^{\bar{m}} \lambda_j^* \left[\sum_{i=1}^{n} G_{j,i}(x_1^*, x_2^*, \ldots, x_n^*) x_i'(0) \right].$$

In light of (21) and (22), this last equation simplifies to

$$0 = \sum_{i=1}^{n} F_i(x_1^*, x_2^*, \ldots, x_n^*) x_i'(0) + \lambda_1^*.$$

And hence, in light of (20),

$$\lambda_1^* \geq 0.$$

Analogous arguments show that

$$\lambda_j^* \geq 0$$

for $j = 2, 3, \ldots, \bar{m}$ as well, completing the proof.

4.2 The Envelope Theorem

Proving a generalized version of the envelope theorem requires no new ideas, just repeated applications of the previous ones.

Consider, again, the constrained optimization problem with n choice variables and m constraints:

$$\max_{x_1, x_2, \ldots, x_n} F(x_1, x_2, \ldots, x_n) \text{ subject to } c_j \geq G_j(x_1, x_2, \ldots, x_n) \text{ for all } j = 1, 2, \ldots, m.$$

Now extend this problem by allowing the functions F and G_j, $j = 1, 2, \ldots, m$, to depend on a parameter $\theta \in \mathbf{R}$:

$$\begin{aligned} \max_{x_1, x_2, \ldots, x_n} \quad & F(x_1, x_2, \ldots, x_n, \theta) \text{ subject to} \\ & c_j \geq G_j(x_1, x_2, \ldots, x_n, \theta) \text{ for all } j = 1, 2, \ldots, m. \end{aligned}$$

Just as before, define the maximum value function $V : \mathbf{R} \to \mathbf{R}$ as

$$\begin{aligned} V(\theta) &= \max_{x_1,x_2,\ldots,x_n} F(x_1, x_2, \ldots, x_n, \theta) \\ &\quad \text{subject to } c_j \geq G_j(x_1, x_2, \ldots, x_n, \theta) \text{ for all } j = 1, 2, \ldots, m. \end{aligned}$$

Note that V is still a function of the single parameter θ, since the n choice variables are "optimized out." Put another way, evaluating V requires the same two-step procedure as before:

First, given θ, find the values $x_1^*(\theta), x_2^*(\theta), \ldots, x_n^*(\theta)$ that solve the constrained optimization problem.

Second, substitute these values $x_1^*(\theta), x_2^*(\theta), \ldots, x_n^*(\theta)$, together with the given value of θ, into the objective function to obtain

$$V(\theta) = F(x_1^*(\theta), x_2^*(\theta), \ldots, x_n^*(\theta), \theta).$$

And just as before, the envelope theorem tells us that we can calculate the derivative $V'(\theta)$ of the maximum value function while ignoring the dependence of $x_1^*, x_2^*, \ldots, x_n^*$ and $\lambda_1^*, \lambda_2^*, \ldots, \lambda_m^*$ on θ, provided we invoke the complementary slackness conditions (16) to add the sum of all of the multipliers times all of the constraints to the objective function before differentiating through by θ.

Theorem (Envelope) Let F and G_j, $j = 1, 2, \ldots, m$, be continuously differentiable functions of $x_1, x_2, \ldots, x_n$ and θ. For any value of θ, let $x_1^*(\theta), x_2^*(\theta), \ldots, x_n^*(\theta)$ maximize $F(x_1, x_2, \ldots, x_n, \theta)$ subject to $c_j \geq G_j(x_1, x_2, \ldots, x_n, \theta)$ for all $j = 1, 2, \ldots, m$, and let $\lambda_1^*(\theta), \lambda_2^*(\theta), \ldots, \lambda_m^*(\theta)$ be the associated values of the Lagrange multipliers. Suppose, further, that $x_1^*(\theta), x_2^*(\theta), \ldots, x_n^*(\theta)$ and $\lambda_1^*(\theta), \lambda_2^*(\theta), \ldots, \lambda_m^*(\theta)$ are all continuously differentiable functions, and that the $\bar{m}(\theta) \times m$ matrix of derivatives

$$\begin{bmatrix} G_{1,1}(x_1^*(\theta), x_2^*(\theta), \ldots, x_n^*(\theta), \theta) & \ldots & G_{1,n}(x_1^*(\theta), x_2^*(\theta), \ldots, x_n^*(\theta), \theta) \\ G_{2,1}(x_1^*(\theta), x_2^*(\theta), \ldots, x_n^*(\theta), \theta) & \ldots & G_{2,n}(x_1^*(\theta), x_2^*(\theta), \ldots, x_n^*(\theta), \theta) \\ \vdots & \ddots & \vdots \\ G_{\bar{m}(\theta),1}(x_1^*(\theta), x_2^*(\theta), \ldots, x_n^*(\theta), \theta) & \ldots & G_{\bar{m}(\theta),n}(x_1^*(\theta), x_2^*(\theta), \ldots, x_n^*(\theta), \theta) \end{bmatrix}$$

associated with the $\bar{m}(\theta) \leq m$ binding constraints has rank $\bar{m}(\theta)$ for each value of θ. Then the maximum value function defined by

$$\begin{aligned} V(\theta) &= \max_{x_1,x_2,\ldots,x_n} F(x_1, x_2, \ldots, x_n, \theta) \\ &\quad \text{subject to } c_j \geq G_j(x_1, x_2, \ldots, x_n, \theta) \text{ for all } j = 1, 2, \ldots, m \end{aligned}$$

satisfies

$$\begin{aligned} V'(\theta) = F_{n+1}&(x_1^*(\theta), x_2^*(\theta), \ldots, x_n^*(\theta), \theta) \\ &- \sum_{j=1}^{m} \lambda_j^*(\theta) G_{j,n+1}(x_1^*(\theta), x_2^*(\theta), \ldots, x_n^*(\theta), \theta). \end{aligned} \tag{23}$$

Proof The Kuhn-Tucker theorem implies that for any given value of θ,

$$F_i(x_1^*(\theta), x_2^*(\theta), \ldots, x_n^*(\theta), \theta) - \sum_{j=1}^{m} \lambda_j^*(\theta) G_{j,i}(x_1^*(\theta), x_2^*(\theta), \ldots, x_n^*(\theta), \theta) = 0 \tag{13}$$

for $i = 1, 2, \ldots, n$, and

$$\lambda_j^*(\theta)[c_j - G_j(x_1^*(\theta), x_2^*(\theta), \ldots, x_n^*(\theta), \theta)] = 0, \tag{16}$$

for $j = 1, 2, \ldots, m$ must hold.

In light of (16),

$$V(\theta) = F(x_1^*(\theta), x_2^*(\theta), \ldots, x_n^*(\theta), \theta) + \sum_{j=1}^{m} \lambda_j^*(\theta)[c_j - G_j(x_1^*(\theta), x_2^*(\theta), \ldots, x_n^*(\theta), \theta)].$$

Differentiating both sides of this expression by θ yields

$$\begin{aligned} V'(\theta) \quad = \quad & \sum_{i=1}^{n} F_i(x_1^*(\theta), x_2^*(\theta), \ldots, x_n^*(\theta), \theta) x^{*\prime}(\theta) \\ & + F_{n+1}(x_1^*(\theta), x_2^*(\theta), \ldots, x_n^*(\theta), \theta) \\ & + \sum_{j=1}^{m} \lambda_j^{*\prime}(\theta)[c_j - G_j(x_1^*(\theta), x_2^*(\theta), \ldots, x_n^*(\theta), \theta)] \\ & - \sum_{i=1}^{n} \sum_{j=1}^{m} \lambda_j^*(\theta) G_{j,i}(x_1^*(\theta), x_2^*(\theta), \ldots, x_n^*(\theta), \theta) x^{*\prime}(\theta) \\ & - \sum_{j=1}^{m} \lambda_j^*(\theta) G_{j,n+1}(x_1^*(\theta), x_2^*(\theta), \ldots, x_n^*(\theta), \theta). \end{aligned}$$

which shows that, in principle, we must take the dependence of $x_1^*(\theta), x_2^*(\theta), \ldots, x_n^*(\theta)$ and $\lambda_1^*(\theta), \lambda_2^*(\theta), \ldots, \lambda_m^*(\theta)$ on θ into account when calculating $V'(\theta)$.

Note, however, that (13) implies that the sums in the first and fourth lines of this last expression together equal zero. Hence, to show that (23) holds, it only remains to show that

$$\sum_{j=1}^{m} \lambda_j^{*\prime}(\theta)[c_j - G_j(x_1^*(\theta), x_2^*(\theta), \ldots, x_n^*(\theta), \theta)] = 0$$

and this is true if

$$\lambda_j^{*\prime}(\theta)[c_j - G_j(x_1^*(\theta), x_2^*(\theta), \ldots, x_n^*(\theta), \theta)] = 0 \tag{24}$$

for all $j = 1, 2, \ldots, m$.

Clearly, (24) holds for any θ such that constraint j is binding.

For θ such that constraint j is not binding, (16) implies that $\lambda_j^*(\theta) = 0$. Furthermore, by the continuity of G_j and $x_i(\theta)$, $i = 1, 2, \ldots, n$, if constraint j does not bind at θ, there exists an $\varepsilon^* > 0$ such that constraint j does not bind for all $\theta + \varepsilon$ with $\varepsilon^* > |\varepsilon|$. Hence,

$$\lambda_j^{*\prime}(\theta) = \lim_{\varepsilon \to 0} \frac{\lambda_j^*(\theta + \varepsilon) - \lambda_j^*(\theta)}{\varepsilon} = \lim_{\varepsilon \to 0} \frac{0}{\varepsilon} = 0,$$

and once again it becomes apparent that (24) must hold. Hence, (23) must hold as well.

5 Berge's Maximum Theorem

References:

Acemoglu, Appendix A.

Nancy L. Stokey and Robert E. Lucas, Jr. with Edward C. Prescott, *Recursive Methods in Economic Dynamics*, 1989, Chapter 3.

Efe Ok, *Real Analysis with Economic Applications*, 2007, Sections E.1-E.3.

Claude Berge. *Topological Spaces: Including a Treatment of Multi-Valued Functions, Vector Spaces, and Convexity*, 1963.

Thus far, we have assumed that solutions to our constrained maximization problems exist and vary smoothly with the parameters of interest. Claude Berge's Maximum Theorem identifies a set of restrictions that can be imposed on the objective function and constraint set for any such problem to guarantee that these results hold. Intuitively, we can view Berge's theorem as a natural extension of Weierstrass's extreme value theorem: that any continuous function attains a maximum and a minimum on a compact set. Generalizing this result to apply to the smooth dependence of the solution on the parameters, however, requires a non-trivial initial "investment." We will need to begin by specifying several notions of continuity for correspondences, that is, multi-valued functions.

But let's start by fixing the notation. Let $x \in X \subseteq \mathbf{R}^n$ be a vector of n choice variables and let $\theta \in \Theta \subseteq \mathbf{R}^m$ be a vector of m parameters. Let the function $F : X \times \Theta \to \mathbf{R}$ denote the objective function and let the *correspondence* $G : \Theta \to X$ describe, for any given $\theta \in \Theta$, the feasible set of values for x, denoted by $G(\theta) \subseteq X$. We can now write the optimization problem compactly as

$$\sup_{x \in G(\theta)} F(x, \theta).$$

Note that by casting the problem initially as finding the supremum and not the maximum, we avoid presupposing that the maximum is actually attained: here, we want this as a result, not an assumption.

We must now deal directly with the fact that because G is not a single-valued, standard notions of continuity, which apply to functions, will not apply. To simplify the analysis, it is helpful at the outset to assume that G is *compact-valued*, meaning that for all $\theta \in \Theta$, the set $G(\theta) \subseteq X \subseteq \mathbf{R}^n$ is compact (or, equivalently, by the Heine-Borel theorem, closed and bounded). Each of the continuity concepts introduced below, except one, applies even if G is not compact-valued. But, because this restriction is imposed by Berge's theorem, we might as well make the assumption from the start.

The first notion, of *upper hemi-continuity*, can be cast equivalently in terms of sets or sequences as follows.

Definition The correspondence G is upper hemi-continuous at $\theta \in \Theta$ if $G(\theta)$ is non-empty and if, for every open set $X' \subseteq X$ with $G(\theta) \subseteq X'$ (i.e., any open subset of X containing $G(\theta)$), there exists a $\delta > 0$ such that for every $\theta' \in N_\delta(\theta)$, $G(\theta') \subseteq X'$ (i.e., a neighborhood of θ so that $G(\theta')$ is also in X' for all θ' in that neighborhood).

Definition The *compact-valued* correspondence G is upper hemi-continuous at $\theta \in \Theta$ if $G(\theta)$ is non-empty and if, for every sequence $\{\theta_j\}$ such that $\theta_j \to \theta$ and every sequence $\{x_j\}$ such that $x_j \in G(\theta_j)$ for all j, there exists a convergent subsequence $\{x_{j_k}\}$ such that $x_{j_k} \to x \in G(\theta)$.

Note that only the second definition requires G to be compact-valued. If G is not compact-valued, the conditions imposed on the sequences in the second definition are sufficient, but not necessary, for G to be upper hemi-continuous as described in the first definition.

The second notion, *lower hemi-continuity*, can be also cast either in terms of sets or sequences as follows.

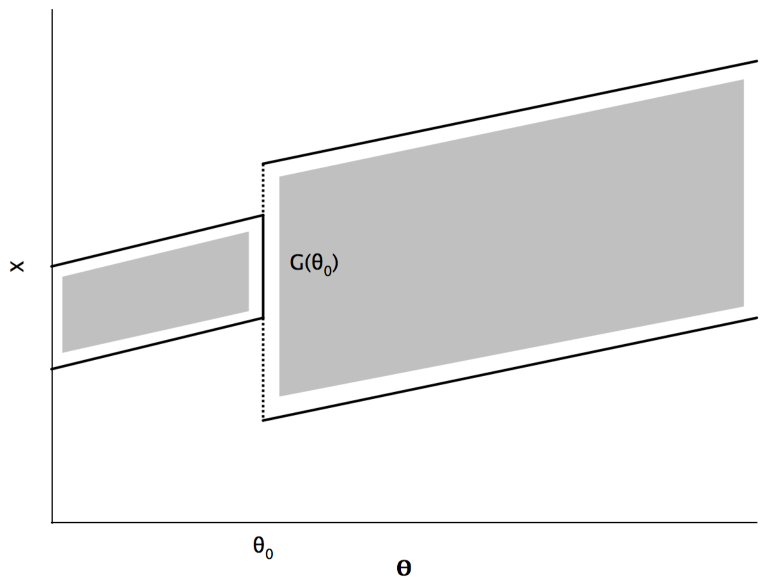
Definition The correspondence G is lower hemi-continuous at $\theta \in \Theta$ if $G(\theta)$ is non-empty and if, for every open set $X' \subseteq X$ with $G(\theta) \cap X' \neq \emptyset$ (i.e., any open subset of X intersecting with $G(\theta)$), there exists a $\delta > 0$ such that for every $\theta' \in N_\delta(\theta)$, $G(\theta') \cap X' \neq \emptyset$ (i.e., a neighborhood of θ so that $G(\theta')$ also intersects with X' for all θ' in that neighborhood).

Definition The correspondence G is lower hemi-continuous at $\theta \in \Theta$ if $G(\theta)$ is non-empty and if, for every $x \in G(\theta)$ and every sequence $\{\theta_j\}$ such that $\theta_j \to \theta$, there is a number $J \geq 1$ and a sequence $\{x_j\}$ such that $x_j \in G(\theta_j)$ for all $j \geq J$ and $x_j \to x$.

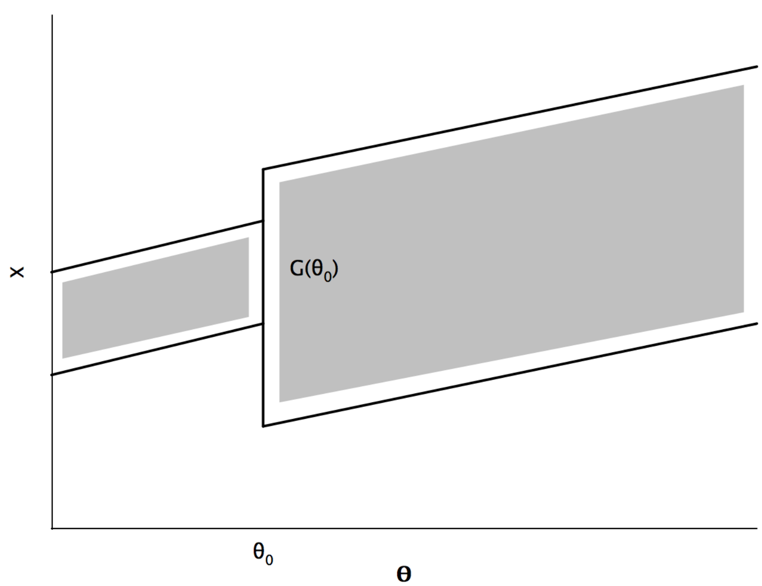
Using either definition in each case, G is upper hemi-continuous on Θ if $G(\theta)$ if it is upper hemi-continuous at every $\theta \in \Theta$ and G is lower hemi-continuous on Θ if it is lower hemi-continuous at every $\theta \in \Theta$. Finally, G is *continuous* if it is both upper and lower hemi-continuous.

Definition The correspondence G is continuous at $\theta \in \Theta$ if it is both upper hemi-continuous and lower hemi-continuous at θ. The correspondence G is continuous on Θ if it is continuous at every $\theta \in \Theta$.

Before moving on, it is worthwhile to try to build up intuition as to the behavior that is ruled out by the requirement that a correspondence be upper or lower hemi-continuous. Loosely speaking, an upper hemi-continuous correspondence cannot suddenly "explode." In the figure below, the correspondence G is lower but not upper hemi-continuous at θ_0.



Meanwhile, a lower hemi-continuous correspondence cannot suddenly "implode." In the figure below, the correspondence G is upper but not lower hemi-continuous.



These previous examples suggest that if G bounds x from above and below by two continuous functions, then G will be continuous. And indeed we can prove that this true.

Specifically, let $f : \Theta \to \mathbf{R}$ and $g : \Theta \to \mathbf{R}$ be two continuous functions satisfying $f(\theta) \leq g(\theta)$ for all $\theta \in \Theta$. Then define $G : \Theta \to \mathbf{R}$ as

$$G(\theta) = \{x \in \mathbf{R} \mid f(\theta) \leq x \leq g(\theta)\}.$$

Note first that G is nonempty since, for example, $g(\theta) \in G(\theta)$ for all $\theta \in \Theta$.

Next, let's show that G is upper hemi-continuous. To do this, fix $\theta \in \Theta$ and let $\{\theta_j\}$ be any sequence with $\theta_j \to \theta$. Since G is nonempty, we can also find a sequence $\{x_j\}$ such that $x_j \in G(\theta_j)$ for all j. Since $\theta_j \to \theta$, there exists a bounded set $\hat{\Theta} \subseteq \Theta \subseteq \mathbf{R}$ such that, for some $J \geq 1$, all of the θ_j for $j \geq J$ and θ will be contained in $\hat{\Theta}$. Moreover, the special structure of G in this case will imply that all of the x_j for $j \geq J$ will also lie in a bounded subset of $\mathbf{R}$. Thus, for all $j \geq J$, all elements of the sequence $\{(\theta_j, x_j)\}$ lie in a bounded subset of $\mathbf{R}^2$ and, by the Bolzano-Weierstrass theorem, this sequence has a convergent subsequence $\{(\theta_{j_k}, x_{j_k})\}$ with limit point (θ, x). And since each element of this convergent sequence satisfies

$$f(\theta_{j_k}) \leq x_{j_k} \leq g(\theta_{j_k}),$$

the continuity of f and g guarantees that the limit point satisfies $x \in G(\theta)$ as well.

Finally, to show that G is lower hemi-continuous, fix $\theta \in \Theta$ and $x \in G(\theta)$, and let $\{\theta_j\}$ be any sequence with $\theta_j \to \theta$. If $f(\theta) = g(\theta)$, then the sequence $\{x_j\}$ with $x_j = g(\theta_j)$ for all j satisfies $x_j \in G(\theta_j)$ for all j and, because g is continuous, $x_j \to g(\theta) = x$ as well. If, on the other hand, $f(\theta) < g(\theta)$, note that

$$\frac{x - f(\theta)}{g(\theta) - f(\theta)}$$

is well-defined and, since $x \in G(\theta)$, lies between zero and one. Therefore, the sequence $\{x_j\}$ with

$$x_j = \left[1 - \frac{x - f(\theta)}{g(\theta) - f(\theta)}\right] f(\theta_j) + \left[\frac{x - f(\theta)}{g(\theta) - f(\theta)}\right] g(\theta_j)$$

for all j satisfies $x_j \in G(\theta_j)$ for all j and, since f and g are both continuous

$$x_j \to \left[1 - \frac{x - f(\theta)}{g(\theta) - f(\theta)}\right] f(\theta) + \left[\frac{x - f(\theta)}{g(\theta) - f(\theta)}\right] g(\theta) = f(\theta) + x - f(\theta) = x,$$

completing the proof.

Note that this last example also shows that if $G(\theta)$ is single-valued, with $G(\theta) = g(\theta)$ for some continuous function g, then G is both upper hemi-continuous and lower hemi-continuous. In fact, we can also show that the converse is true: if $G(\theta)$ is single-valued, with $G(\theta) = g(\theta)$ for some function g, then g is continuous if G is either upper hemi-continuous or lower hemi-continuous.

To establish this, we will first show that if G is single-valued and upper hemi-continuous, then it must also be lower hemi-continuous. The proof is by contradiction. Suppose that G is single-valued, so that $G(\theta) = g(\theta)$ for all $\theta \in \Theta$ for some function $g : \Theta \to X$. And suppose also G is upper hemi-continuous but *not* lower hemi-continuous.

In this case, fix $\theta \in \Theta$ and $\{\theta_j\}$ with $\theta_j \to \theta$. Because G is single-valued, $\{x_j\}$ is defined uniquely by $x_j = G(\theta_j) = g(\theta_j)$ for all j. But since G is not lower hemi-continuous, then x_j cannot converge to $x = G(\theta) = g(\theta)$. That is, there must exist an $\varepsilon > 0$ such that, for any $J \geq 1$, there exists a $j \geq J$ such that $|x_j - x| \geq \varepsilon$. Using these values of $x_j = g(\theta_j)$, construct the subsequence $\{\theta_{j_k}\}$ of $\{\theta_j\}$; this subsequence has $\theta_{j_k} \to \theta$ but, by construction, there is no subsequence of $\{x_{j_k}\}$ converging to x. This shows that G cannot be upper hemi-continuous: we have our contradiction, implying that if G is single-valued and upper hemi-continuous, it must also be lower hemi-continuous.

Next, assume that G is single-valued, with $G(\theta) = g(\theta)$ for all $\theta \in \Theta$ for some function $g : \Theta \to X$, and that G is lower hemi-continuous. We will show that g is continuous. Fix $\theta \in \Theta$ and $\{\theta_j\}$ with $\theta_j \to \theta$. Then $x = G(\theta) = g(\theta)$ is uniquely-defined, as is the sequence $\{x_j\}$ with $x_j = G(\theta_j) = g(\theta_j)$ for all j. Since G is lower hemi-continuous, $x_j \to x$. This shows that for any sequence $\{\theta_j\}$ with $\theta_j \to \theta$, the sequence $\{g(\theta_j)\}$ must have $g(\theta_j) \to g$ as required.

Thus, when G is single-valued, with $G(\theta) = g(\theta)$ for all $\theta \in \Theta$, the correspondence G is continuous if and only if the function g is continuous, confirming that the notion of a continuous correspondence generalizes, in a natural way, the more familiar definition of a continuous function.

We can now state and prove Berge's result, which requires F to be continuous in both x and θ and G to be compact-valued and continuous.

Theorem (Berge's Maximum Theorem) Let $X \subseteq \mathbf{R}^n$ and $\Theta \subseteq \mathbf{R}^m$. Let $F : X \times \Theta \to \mathbf{R}$ be a continuous function, and let $G : \Theta \to X$ be a compact-valued and continuous correspondence. The the maximum value function

$$V(\theta) = \max_{x \in G(\theta)} F(x, \theta)$$

is well-defined and continuous, and the optimal policy correspondence

$$x^*(\theta) = \{x \in G(\theta) \mid F(x, \theta) = V(\theta)\}$$

is nonempty, compact-valued, and upper hemi-continuous.

Proof Fix $\theta \in \Theta$. Note first that since $G(\theta)$ is nonempty and compact, and since $F(\cdot, \theta)$ is continuous, Weierstrass's extreme value theorem implies that $V(\theta)$ is well-defined and that $x^*(\theta)$ is nonempty.

Notice, next, that since $x^*(\theta) \subseteq G(\theta)$ and $G(\theta)$ is compact, it follows that $x^*(\theta)$ is bounded. Let $\{x_j\}$ be a sequence with $x_j \in x^*(\theta)$ for all j and $x_j \to x$. Since $G(\theta)$ is closed, it must be that $x \in G(\theta)$. And since $V(\theta) = F(x_j, \theta)$ for all j and F is continuous, it follows that $F(x, \theta) = V(\theta)$. Hence, $x \in x^*(\theta)$, so $x^*(\theta)$ is closed. We now know that $x^*(\theta)$ is nonempty and compact-valued for all $\theta \in \Theta$.

To see that $x^*(\theta)$ must be upper hemi-continuous, fix $\theta \in \Theta$ and let $\{\theta_j\}$ be any sequence with $\theta_j \to \theta$. Choose $x_j \in x^*(\theta_j)$ for all j. Since G is upper hemi-continuous, there exists a subsequence $\{x_{j_k}\}$ converging to $x \in G(\theta)$. Let $x' \in G(\theta)$ as well. Since G is lower hemi-continuous, there exists a sequence $\{x'_{j_k}\}$ with $x'_{j_k} \in G(\theta_{j_k})$ and $x'_{j_k} \to x'$. Since $F(x_{j_k}, \theta_{j_k}) \geq F(x'_{j_k}, \theta_{j_k})$ for all k and F is continuous, it follows that $F(x, \theta) \geq F(x', \theta)$. And since this condition holds for any $x' \in G(\theta)$, it follows that $x \in x^*(\theta)$, so that x^* is upper hemi-continuous.

Finally, to see that V is continuous, fix $\theta \in \Theta$, and let $\{\theta_j\}$ be any sequence with $\theta_j \to \theta$. Choose $x_j \in x^*(\theta_j)$ for all j. Let $\bar{v} = \limsup_j V(\theta_j)$ and $\underline{v} = \liminf_j V(\theta_j)$. Then there exists a subsequence $\{x_{j_k}\}$ such that $\bar{v} = \lim_k F(x_{j_k}, \theta_{j_k})$. But since x^* is upper hemi-continuous, these exists a subsequence of $\{x_{j_{k_l}}\}$ of $\{x_{j_k}\}$ converging to $x \in x^*(\theta)$. Hence $\bar{v} = \lim_l F(x_{j_{k_l}}, \theta_{j_{k_l}}) = F(x, \theta) = V(\theta)$. An analogous argument establishes that $\underline{v} = V(\theta)$. Hence $V(\theta_j) \to V(\theta)$, completing the proof.

Note that if we assume, as well, that $F(\cdot, \theta)$ is concave and $G(\theta)$ is convex for all $\theta \in \Theta$, then the optimal policy correspondence is x^* is single-valued. Since Berge's theorem implies that, in general, x^* is an upper hemi-continuous correspondence, in this case our previous arguments show that x^* is a continuous function as well.

As an application of the result, let's consider the case of utility maximization with two goods. The consumer takes income I and prices p_1 and p_2 as given, and solves

$$\max_{c_1, c_2} U(c_1, c_2) \text{ subject to } I \geq p_1 c_1 + p_2 c_2, c_1 \geq 0, c_2 \geq 0.$$

To apply the theorem, we need the utility function U to be continuous, and the correspondence

$$G(I, p_1, p_2) = \{(c_1, c_2) \in \mathbf{R}^2 \mid I \geq p_1 c_1 + p_2 c_2, c_1 \geq 0, c_2 \geq 0\}$$

to be compact-valued and continuous. Clearly, for any $(I, p_1, p_2) \in \mathbf{R}^3$ with $I > 0$, $p_1 > 0$ and $p_2 > 0$, G is nonempty and compact-valued. So it only remains to confirm that G is upper and lower hemi-continuous.

To see that G is upper hemi-continuous, fix the parameter values (I, p_1, p_2) with $I > 0$, $p_1 > 0$, and $p_2 > 0$ and let $\{(I_j, p_{1j}, p_{2j})\}$ be a sequence with $I_j > 0$, $p_{1j} > 0$ and $p_{2j} > 0$ for all j with $(I_j, p_{1j}, p_{2j}) \to (I, p_1, p_2)$. Since G is non-empty, we can find a sequence $\{(c_{1j}, c_{2j})\}$ with $(c_{1j}, c_{2j}) \in G(I_j, p_{1j}, p_{2j})$ for all j. Now, since $(I_j, p_{1j}, p_{2j}) \to (I, p_1, p_2)$, we can also find a bounded set $\hat{\Theta} \subseteq \Theta \subseteq \mathbf{R}^3$, such that, for some $J \geq 1$, all of the (I_j, p_{1j}, p_{2j}) for $j \geq J$ and (I, p_1, p_2) are contained in $\hat{\Theta}$. Note, too, that the special structure of G in this case implies that all of the $(c_{1j}, c_{2j}) \in G(I_j, p_{1j}, p_{2j})$ for $j \geq J$ will lie in a bounded subset of $\mathbf{R}^2$. Thus, for all $j \geq J$, all elements of the sequence $\{(I_j, p_{1j}, p_{2j}, c_{1j}, c_{2j})\}$ lie in a bounded subset of $\mathbf{R}^5$ and, by the Bolzano-Weierstrass theorem, this sequence has a convergent subsequence $\{(I_{j_k}, p_{1j_k}, p_{2j_k}, c_{1j_k}, c_{2k_k})\}$ with limit point (I, p_1, p_2, c_1, c_2). And since each element of this convergent subsequence satisfies

$$I_{j_k} \geq p_{1j_k} c_{1j_k} + p_{2j_k} c_{2j_k},$$

$c_{1j_k} \geq 0$, and $c_{2j_k} \geq 0$, it is easy to see that the limit point will also have to satisfy $(c_1, c_2) \in G(I, p_1, p_2)$.

To see that G is lower hemi-continuous as well, fix (I, p_1, p_2) with $I > 0$, $p_1 > 0$, and $p_2 > 0$ and $(c_1, c_2) \in G(I, p_1, p_2)$, then let $\{(I_j, p_{1j}, p_{2j})\}$ be a sequence with $I_j > 0$, $p_{1j} > 0$ and $p_{2j} > 0$ for all j and $(I_j, p_{1j}, p_{2j}) \to (I, p_1, p_2)$. If $c_1 = c_2 = 0$, then simply constructing the sequence $\{(c_{1j}, c_{2j})\}$ with $c_{1j} = c_{2j} = 0$ for all j will provide the needed example where $(c_{1j}, c_{2j}) \in G(I_j, p_{1j}, p_{2j})$ for all j and $(c_{1j}, c_{2j}) \to (c_1, c_2)$. If c_1 and/or c_2 is strictly positive, then construct the sequence $\{(c_{1j}, c_{2j})\}$ with

$$c_{1j} = \left(\frac{I_j}{I}\right)\left(\frac{p_1 c_1 + p_2 c_2}{p_{1j} c_1 + p_{2j} c_2}\right) c_1$$

and

$$c_{2j} = \left(\frac{I_j}{I}\right)\left(\frac{p_1 c_1 + p_2 c_2}{p_{1j} c_1 + p_{2j} c_2}\right) c_2$$

for all j. Note that $c_{1j} \geq 0$, $c_{2j} \geq 0$, and

$$p_{1j} c_{1j} + p_2 c_{2j} = \left(\frac{I_j}{I}\right)\left(\frac{p_1 c_1 + p_2 c_2}{p_{1j} c_1 + p_{2j} c_2}\right)(p_{1j} c_1 + p_{2j} c_2) = I_j \left(\frac{p_1 c_1 + p_2 c_2}{I}\right) \leq I_j,$$

so that $(c_{1j}, c_{2j}) \in G(I_j, p_{1j}, p_{2j})$ for all j. Moreover, $(c_{1j}, c_{2j}) \to (c_1, c_2)$, completing the proof.

Thus, we can apply Berge's theorem to the consumer's problem and conclude that if the utility function is continuous:

The indirect utility function $V(I, p_1, p_2)$ is well-defined and continuous.

And the demand correspondences $c_1^*(I, p_1, p_2)$ and $c^*(I, p_1, p_2)$ are nonempty, compact-valued, and upper hemi-continuous.

And since the set $G(I, p_1, p_2)$ is convex for all $I > 0$, $p_1 > 0$, and p_2, if we assume as well that U is strictly concave, the Marshallian demand functions $c_1^*(I, p_1, p_2)$ and $c^*(I, p_1, p_2)$ are well-defined (single-valued) and continuous.