APPENDIX C: OPTIMIZATION

C.1 MAXIMIZATION WITH 2 VARIABLES

Key Ideas: Necessary and sufficient conditions

Rather than leap directly to the analysis of multi-variable optimization problems we begin by examining the two variable case. Suppose that the function f is differentiable at $x^0 = (x_1^0, x_2^0)$. If f takes on its maximum over $\mathbb{R}^2$ at x^0, then taking one variable at a time, we know that each of the first partial derivatives must be zero.

Proposition C.1-1: First-Order Conditions (FOC) for a Maximum

If $f(x_1, x_2)$ takes on its maximum over $\mathbb{R}^2$ at x^0, then $\frac{\partial f}{\partial x_i}(x^0) = 0,\ i = 1, 2$.

We also know that each of the second partial derivatives must be negative. To obtain a further necessary condition we consider the change in f as x changes from x^0 in the direction of some other vector x^1. We do this by considering the weighted average x^λ of x^0 and x^1, that is

$$x^\lambda = (1-\lambda)x^0 + \lambda x^1 = x^0 + \lambda(x^1 - x^0),$$

and define $g(\lambda) = f(x^0 + \lambda(x^1 - x^0))$. This function is depicted below.

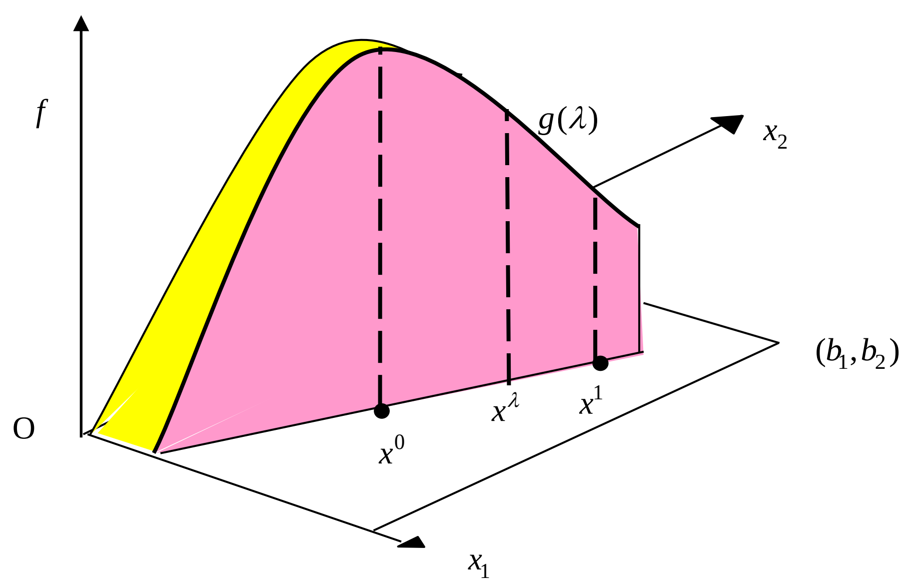


Figure C.1-1: Cross-section of f

The mapping, $g(\lambda)$, depicted in the cross-section is a function from $\mathbb{R}$ into $\mathbb{R}$. Then we can appeal to the necessary conditions for one-variable maximization. In particular, for a maximum at x^0, the second derivative of $g(\lambda)$ must be negative at $\lambda = 0$. Define $z \equiv x^1 - x^0$. Then

$$g(\lambda) = f(x^0 + \lambda z) = f(x_1^0 + \lambda z_1, x_2^0 + \lambda z_2)$$

The first derivative of g is therefore

$$\frac{dg}{d\lambda}(\lambda) = \frac{\partial f}{\partial x_1}(x^\lambda) z_1 + \frac{\partial f}{\partial x_2}(x^\lambda) z_2 .$$

For a maximum this must be zero for all z, hence the partial derivatives at $\lambda = 0$ must both be zero.

Differentiating again and setting $\lambda = 0$

$$\frac{d^2 g}{d\lambda^2}(0) = z_1{}^2 \frac{\partial^2 f}{\partial x_1{}^2}(x^0) + 2 z_1 z_2 \frac{\partial^2 f}{\partial x_1 \partial x_2}(x^0) + z_2{}^2 \frac{\partial^2 f}{\partial x_2{}^2}(x^0)$$

$$= \begin{bmatrix} z_1 & z_2 \end{bmatrix} \begin{bmatrix} \frac{\partial^2 f}{\partial x_1{}^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_1{}^2} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}.$$

In matrix notation,

$$\frac{d^2 g}{d\lambda^2}(0) = (x^1 - x^0)'[\frac{\partial^2 f}{\partial x_i \partial x_j}(x^0)](x^1 - x^0) \text{ because } z = x^1 - x^0 .$$

Note that the right-hand side of the last equation is a quadratic form. Appealing to Proposition B.2-2 we have the following result.

Proposition C.1-2: Second-Order Conditions (SOC) for a Maximum

If $f(x_1, x_2)$ takes on its maximum over $\mathbb{R}^2$ at x^0, then

(i) $\frac{\partial^2 f}{\partial x_i{}^2}(x^0) \le 0, \ i = 1, 2$ and (ii) $\frac{\partial^2 f}{\partial x_1{}^2}(x^0) \frac{\partial^2 f}{\partial x_2{}^2}(x^0) - (\frac{\partial^2 f}{\partial x_1 \partial x_2}(x^0))^2 \ge 0$.

Exercise C.1-1: Consumer Choice

Bev faces prices p_1 and p_2 and her income is *I*. Her utility function $U(x_1, x_2)$ satisfies $\frac{\partial U}{\partial x_j}(x) > 0,\ \ j = 1,2,\ \ x \in \mathbb{R}^2_+$.

(a) Argue that the budget constraint must be binding and hence reduce Bev's optimization problem to a 1 variable problem. Hence obtain first order necessary conditions for a corner solution.

(b) Suppose that $\lim_{x_j \downarrow 0} \frac{\partial U}{\partial x_j}(x) = \infty,\ \ j = 1,2$. That is, the marginal utility of each commodity increases without bound as consumption of the commodity declines to zero. Show that the necessary conditions for a maximum cannot be satisfied at a corner.

(c) If $U(x) = \sum_{j=1}^{2} \alpha_j \ln x_j$, where $\alpha_j > 0,\ \ j = 1,...,n$, show that the solution cannot be at a corner.

(d) Hence solve for Bev's optimal choice.

(e) What if instead $U(x) = x_1^{\alpha_1} x_2^{\alpha_2}$?

Exercise C.1-2: Firm with interdependent demands

A firm sells two products. Demand prices for these products are as follows:

$$p_1 = 180 - q_1 - \tfrac{1}{2} q_2,\ \ p_2 = 180 - \tfrac{1}{2} q_1 - q_2 .$$

The cost of production is $C(q) = q_1^2 + \alpha q_1 q_2 + q_2^2$.

(a) If $\alpha = 1$, show that the profit function is concave and solve for the outputs that satisfy the first-order conditions.

(b) Show that the FOC for profit maximization are satisfied at $q = (20, 20)$ if $\alpha = 4$.

(c) Are the second order (necessary) conditions satisfied at $q = (20, 20)$?

(d) Show that there are two corner solutions to the FOC. Explain why profit must be maximized at these outputs.

(e) Show that profit is a concave function of q_1 and hence solve for $q_1^*(q_2)$, the profit-maximizing output of commodity 1 for all possible values of q_2.

(f) Totally differentiate $\Pi(q_1^*(q_2), q_2)$ and explain why

$$\frac{d\Pi}{dq_2}(q_1^*(q_2), q_2) = \frac{\partial \Pi}{\partial q_2}(q_1^*(q_2), q_2)\ .$$

Hence show that

$$\frac{d\Pi}{dq_2} = \frac{\partial \Pi}{\partial q_2}(q_1^*, q_2) = 180 - 45(1+\alpha) - (4 - \tfrac{1}{4}(1+\alpha)^2)q_2)\ .$$

(g) Hence characterize the function $\Pi(q_1^*(q_2), q_2)$ for different values of α .

C.2 UNCONSTRAINED OPTIMIZATION

Key Ideas: necessary and sufficient conditions, quasi-concavity

Consider the following problem where the function *f* is twice continuously differentiable.

$$\underset{x}{Max}\{f(x)\,|\,x\in\mathbb{R}^n\}.$$

In the previous section, we examined both the necessary and sufficient conditions for f to take on its maximum at x^0 for the special case when $n=2$. With *n* variables the analysis is similar. First we consider the optimization problem one variable at a time. Appealing to the necessary conditions for the one-variable model it follows that the following first- and second-order necessary conditions must hold, i.e.,

$$\frac{\partial f}{\partial x_j}(x^0)=0 \text{ and } \frac{\partial^2 f}{\partial x_j{}^2}(x^0)\le 0,\ \ j=1,...,n.$$

The intuition behind these and the further necessary conditions comes from a consideration of the quadratic approximation of the function at the point x^0. Consider the following quadratic function:

$$h(x)=f(x^0)+\frac{\partial f}{\partial x}(x^0)\cdot(x-x^0)+\tfrac{1}{2}(x-x^0)'\left[\frac{\partial^2 f}{\partial x_i\partial x_i}\right](x-x^0).$$

As is readily confirmed, *h* and *f* take on the same value and have the same first- and second- partial derivatives at x^0. For *x* sufficiently close to x^0 the linear terms dominate thus, if any of the first partial derivates is non-zero, the function cannot take on its maximum at x^0. Suppose then that the gradient vector , $\frac{\partial f}{\partial x}(x)$ is zero at x^0. Again, for *x* sufficiently close to x^0 the quadratic terms dominate all higher order terms. Therefore, if *f* takes on its maximum at x^0, it is necessarily the case that the quadratic form is negative semi-definite at x^0.

Proposition C.2-1: Necessary Conditions for a Maximum

Suppose f takes on its maximum over $\mathbb{R}^n$ at x^0. If *f* is differentiable at x^0, then

$\frac{\partial f}{\partial x}(x^0)=0$. If f is twice differentiable at x^0, then the quadratic form of the matrix of second partial derivatives $\mathbf{H}(x^0)=\left[\frac{\partial^2 f}{\partial x_i \partial x_j}\right]$ must be negative semi-definite.

Proof: Define $g(\lambda)=f(x^\lambda)=f(x^0+\lambda z)$, where $z\equiv x^1-x^0$.

We will now argue that

$$g''(\lambda)=z'\mathbf{H}(x^\lambda)z \text{ where } \mathbf{H}(x)=\left[\frac{\partial^2 f}{\partial x_i \partial x_j}\right].$$

First note that for any function $\phi(x^0+\lambda z)$,

$$\frac{d}{d\lambda}\phi(x^0+\lambda z)=\frac{\partial \phi}{\partial x}(x^0+\lambda z)\cdot z=z'\frac{\partial \phi}{\partial x}(x^0+\lambda z).$$

In particular, setting $\phi=f$,

$$g'(\lambda)=\frac{d}{d\lambda}f(x^0+\lambda z)=z'\frac{\partial f}{\partial x}(x^0+\lambda z).$$

And so

$$g''(\lambda)=z'\frac{d}{d\lambda}\frac{\partial f}{\partial x}(x^0+\lambda z). \qquad \text{(C.2-1)}$$

Also, setting $\phi=\frac{\partial f}{\partial x_i}$,

$$\frac{d}{d\lambda}\frac{\partial f}{\partial x_i}=\left(\frac{\partial}{\partial x_1}\frac{\partial f}{\partial x_i},\ldots,\frac{\partial}{\partial x_n}\frac{\partial f}{\partial x_i}\right)'z=\left(\frac{\partial}{\partial x_i}\frac{\partial f}{\partial x_1},\ldots,\frac{\partial}{\partial x_i}\frac{\partial f}{\partial x_n}\right)'z.$$

Then

$$\frac{d}{d\lambda}\frac{\partial f}{\partial x}=\begin{bmatrix}\frac{d}{d\lambda}\frac{\partial f}{\partial x_1}\\ \cdot \\ \cdot \\ \frac{d}{d\lambda}\frac{\partial f}{\partial x_n}\end{bmatrix}z=\left[\frac{\partial^2 f}{\partial x_i \partial x_j}\right]z.$$

Substituting this column vector into (C.2-1),

$$g''(\lambda) = \frac{d^2}{d\lambda^2} f(x^\lambda) = z'\left[\frac{\partial^2 f}{\partial x_i \partial x_j}\right] z .$$

Appealing to the results for a single variable, it follows that a further necessary condition for a maximum is that, for all x^1, the second derivative of $g(\lambda)$ must be negative at $\lambda = 0$.

Q.E.D.

In Appendix A we also established sufficient conditions for a maximum. The proofs of Propositions A.5-5 and A.5-6 generalize directly.

Proposition C.2-2: Sufficient Conditions for a Maximum

Suppose $D_f = \mathbb{R}^n$ and f is concave. If the FOC hold at x^0 then

$$f(x) \le f(x^0), \text{ for all } x \in D_f .$$

Proposition C.2-3: Quasi-concavity and Sufficient Conditions for a Maximum

If $\frac{\partial f}{\partial x}(x^0) = 0$ and f is strictly concave in a neighborhood of x^0, then f has a local maximum at x^0. If, in addition, f is quasi-concave on D_f, then $f(x) < f(x^0)$ for all $x \in D_f$.

Exercises

Exercise C.2-1: Profit maximization

A firm has monopoly power in all of its *n* markets. The demand price for commodity *j* is

$$p_j = \alpha_j - \beta q_j - \gamma s, \text{ where } s \equiv \sum_{j=1}^{n} q_j .$$

Total cost is $C(q) = \sum_{j=1}^{n} c_j q_j$.

(a) Write down the total profit of the firm and show that it can be written as follows:

$$\Pi(q) = \sum_{j=1}^{n} (\alpha_j - c_j)q_j - \sum_{j=1}^{n} \beta q_j^{\,2} - \gamma s^2 .$$

Hence (or otherwise) show that profit is a concave function of the output vector *q*.

(b) Assuming that the profit-maximizing output vector $q^* = (q_1^*,...,q_n^*)$ is strictly positive, show that

$$\alpha_j - c_j - 2\beta q_j^* - 2\gamma s = 0, \quad j = 1,...,n .$$

(c) Sum over the commodities and hence solve for the output sum *s*.

(d) Appeal to (b) to solve for the output of commodity *j*.

(e) Hence show that the profit-maximizing outputs are all strictly positive if and only if

$$\frac{\alpha_j - c_j}{\sum_{j=1}^{n} (\alpha_j - c_j)} > \frac{1}{n + \frac{\beta}{\gamma}} .$$

(f) Use this inequality to show that if the demand interdependency is sufficiently small or the demand price functions and costs are sufficiently alike, output of all commodities will be strictly positive.

Exercise C.2-2: Quality and quantity choice

A firm that produces *q* units of quality *z* commands a demand price $p(q,z) = 8z - 2q$.

The cost of production is $C(q,z) = z^{\alpha} q$, where $0 \le z \le 12$.

(a) If $\alpha = 2$, write down the FOC for a profit maximum and confirm that they all hold at $(q^*, z^*) = (4,4)$.

(b) Confirm that the profit function is concave in a neighborhood of $(4,4)$. Explain why it follows that $\Pi(q,z)$ has a local maximum at (4,4).

(c) To establish that the local maximum is in fact the global maximum, show that regardless of the output decision of the firm, it is optimal to choose a quality level of 4. Hence or otherwise, explain why the local maximum must also be the (global) maximum.

(d) Re-examine the problem if $\alpha = 1$.

C.3 IMPLICIT FUNCTION THEOREM

Key Ideas: comparative statics, continuous mapping from parameters to choice variables

In order to take theory to data, a model must have some predictive power. Consider for example a profit maximizing firm. Suppose that the firm has a profit $f(x,\alpha)$ where $x=(x_1,...,x_n)$ is the output vector and $\alpha=(\alpha_1,...,\alpha_m)$ is a vector of parameters (prices and cost characteristics). The predictive power of the model then depends on the ability of the modeler to derive implications about the way the firm responds to changes in the environment, that is, changes in the parameter vector α. Such an analysis is greatly simplified if it is reasonable to assume that the profit-maximizing response $x=g(\alpha)$ is a differentiable function. But when is this true?

To illustrate the issues, consider first the simplest case in which there is a single output and a single cost parameter. We will assume that the profit $f(x,\alpha)$ is a twice-continuously differentiable function. The necessary condition for a profit maximum is

$$\frac{\partial f}{\partial x}(x^*,\alpha)\le 0, \text{ with equality if } x^*=0.$$

We define the function $h(x,\alpha)$ to be the gradient of *f*, that is, $h(x,\alpha)=\partial f/\partial x(x,\alpha)$. Suppose that *h* is a decreasing function of *x* so that the contour set $\{x\,|\,h(x,\alpha)=0\}$ is as depicted below.

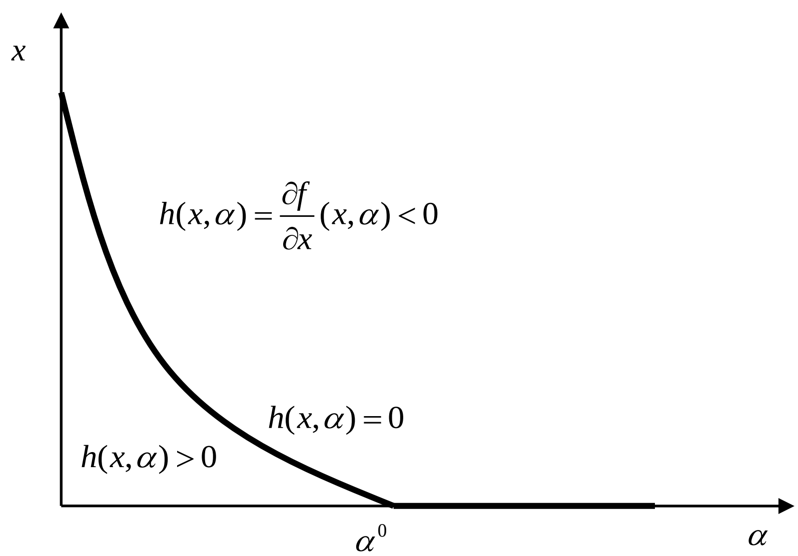


Figure C.3-1: Profit maximizing output

For each α, there is a unique $x = g(\alpha)$ satisfying the FOC. That is, the FOC implicitly defines the profit-maximizing output $x = g(\alpha)$ as a function of the cost parameter. In the figure this is the heavy curve. The profit-maximizing output is strictly positive for all $\alpha < \alpha^0$ and is zero for all higher α.

Example: A firm can sell at a price $p = 10$ and has a cost function $C(x,\alpha) = \alpha x + \frac{1}{2}(1+\alpha)x^2$. The profit is therefore $f(x,\alpha) = 10x - (\alpha x + \frac{1}{2}(1+\alpha)x^2)$. Differentiating by *x*, we obtain

$$h(x,\alpha) = \frac{\partial f}{\partial x}(x,\alpha) = 10 - \alpha - (1+\alpha)x .$$

Note that this function can be rewritten as $h(x,\alpha) = (1+\alpha)(\frac{11}{1+\alpha} - 1 - x)$. The FOC for a maximum at x^* is $\frac{\partial f}{\partial x} = h(x^*,\alpha) \le 0$, with equality if $x^* > 0$. Thus the profit-maximizing output is

$$x = g(\alpha) = \begin{cases} \dfrac{11}{1+\alpha} - 1, & \alpha \le 10 \\ 0, & \alpha > 10 \end{cases} .$$

Thus for the example we can solve explicitly for $g(\alpha)$.

∎

In general we cannot solve explicitly so that all we have is an "implicit function" $g(\alpha)$ that satisfies the equation $h(x,\alpha) = 0$. As long as the function *h* is continuously differentiable, it is tempting to believe that this implicit function must be differentiable as well. Then, as depicted in Figure C.3-1, the profit-maximizing output is continuously differentiable, except at the kink $(g(\alpha^0),\alpha^0)$ where the profit-maximizing output drops to zero. However our assumptions are not quite strong enough.

To understand why, suppose that $h(x,\alpha) = 2 - \alpha - (2-x)^3$. The set of points $x = g(\alpha)$ satisfying $h(x,\alpha) = 0$ is depicted below. Note that the curve is vertical at $\alpha = 2$ so the derivative of $g(\alpha)$ is not well defined at $\alpha = 2$.

Another way of seeing this is to begin by assuming that the function $g(\alpha)$ is differentiable and test whether this leads to a contradiction. Substituting for *x*, we have

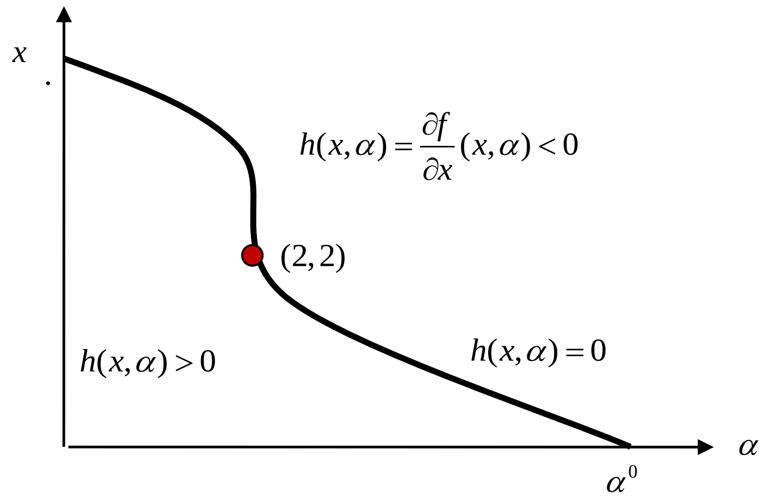


Figure C.3-2: Profit-maximizing output

$h(g(\alpha),\alpha)=0$. If both functions are differentiable then

$$\frac{\partial h}{\partial x}g'(\alpha)+\frac{\partial h}{\partial \alpha}=0 \text{ and so } g'(\alpha)=-\frac{\partial h}{\partial \alpha}/\frac{\partial h}{\partial x}.$$

Note that the right-hand side is only well defined if $\frac{\partial h}{\partial x}(g(\alpha),\alpha))\neq 0$. The problem then arises because $\frac{\partial h}{\partial x}(2,2)=0$.

More generally, suppose that $x=(x_1,...,x_n)$ must satisfy the *n* equations $h_i(x,\alpha)=0,\ i=1,...,n$, where $\alpha=(\alpha_1,...,\alpha_m)$ is a vector of parameters. Suppose also that $(\bar{x},\bar{\alpha})$ satisfies these *n* equations. We would like to know whether these equations implicitly define a continuously differentiable function $x=g(\alpha)$. While we will not provide a complete proof here, the basic insight comes from taking the linear approximations of the functions and asking the same question for the linearized system. The linearized system is

$$\frac{\partial h_i}{\partial x}(\bar{x},\bar{\alpha})\cdot(x-\bar{x})+\frac{\partial h_i}{\partial \alpha}(\bar{x},\bar{\alpha})\cdot(\alpha-\bar{\alpha})=0,\ i=1....,n.$$

Rewriting these *n* equations in matrix form,

$$\left[\frac{\partial h_i}{\partial x_j}(\bar{x},\bar{\alpha})\right](x-\bar{x}) = -\left[\frac{\partial h_i}{\partial \alpha_k}(\bar{x},\bar{\alpha})\right](\alpha-\bar{\alpha}) .$$

The matrix of all gradient vectors of $h=(h_1,...,h_n)$, evaluated at $(\bar{x},\bar{\alpha})$ is known as the Jacobean matrix of the equation system $h(\bar{x},\bar{\alpha})=0$. If this matrix is invertible,

$$x-\bar{x} = -\left[\frac{\partial h_i}{\partial x_j}(\bar{x},\bar{\alpha})\right]^{-1}\left[\frac{\partial h_i}{\partial \alpha_k}(\bar{x},\bar{\alpha})\right](\alpha-\bar{\alpha}) .$$

Thus, for the linearized system, there is a mapping $x=g(\alpha)$ as long as the Jacobean matrix is invertible.

Proposition C.3-1: Implicit function theorem

Let $h_i(x,\alpha),\ i=1,...,n$ be a continuously differentiable function of $(x,\alpha)=(x_1,...,x_n,\alpha_1,...\alpha_m)$ and suppose that $(\bar{x},\bar{a})$ satisfies the system of equations

$$h_i(x,\alpha)=0,\ i=1,...,n .$$

If the Jacobean matrix of partial derivatives $\left[\partial h_i/\partial x_j(\bar{x},\bar{\alpha})\right]$ is invertible, then there exists an open neighborhood of $\bar{\alpha}$, $N_1(\bar{\alpha},\delta_1)$ and an open neighborhood of $\bar{x}$, $N_2(\bar{x},\delta_2)$ and a unique continuously differentiable function $g:N_1\to N_2$ such that for any $\alpha\in N_1$, $h_i(g(\alpha),\alpha)=0,\ i=1,...,n$.

We now return to the simple example of the profit-maximizing firm and sketch a proof of the existence of a unique continuous implicit function $g(\alpha)$ satisfying

$$h(g(\alpha),\alpha)=\frac{\partial f}{\partial x}(g(\alpha),\alpha)=0 .$$

By hypothesis $h(\bar{x},\bar{\alpha})=0$. In addition we assume that h is continuously differentiable in some neighborhood of $(\bar{x},\bar{\alpha})$ and that $\frac{\partial h}{\partial x}(\bar{x},\bar{\alpha})\neq 0$. A necessary condition for a maximum is that $\frac{\partial h}{\partial x}(\bar{x},\bar{\alpha})=\frac{\partial}{\partial x}\frac{\partial f}{\partial x}(\bar{x},\bar{\alpha})\leq 0$. Then if $\frac{\partial h}{\partial x}(\bar{x},\bar{\alpha})\neq 0$ it must be strictly less than zero. Because h is continuously differentiable if follows that there is some

$a > 0$ and a square with side of length 2*a*, centered on $\bar{\alpha}$ and $\bar{x}$, within which $\frac{\partial h}{\partial x}(x,\alpha) < 0$. This is the shaded region in Figure C.3-3 below.

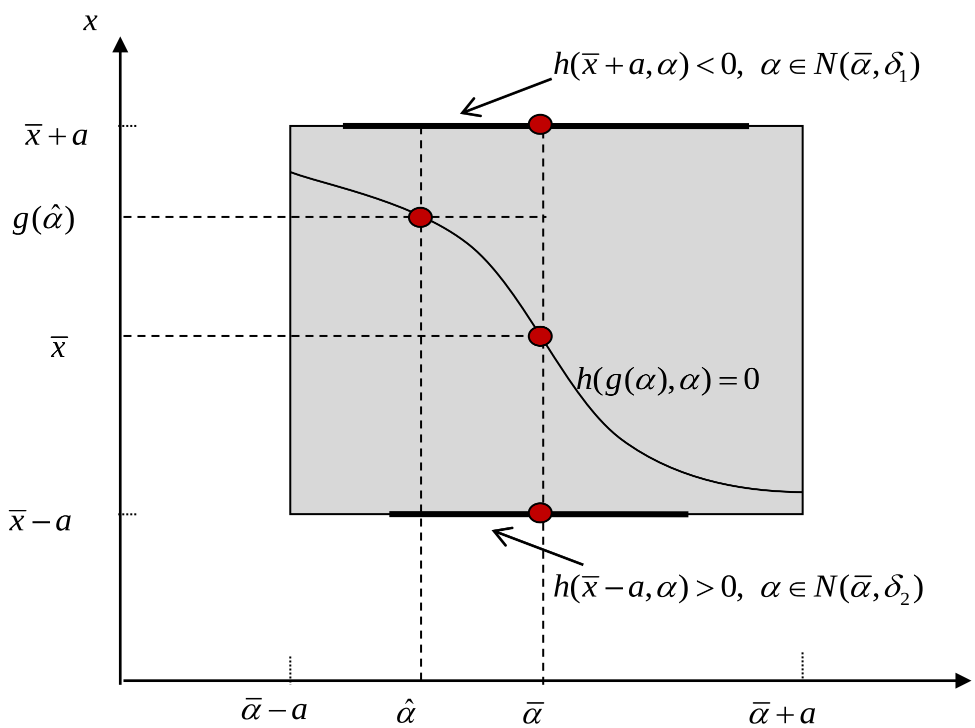


Figure C.3-3: The implicit function

Because $h(\bar{x},\bar{\alpha}) = 0$, it follows that

$$h(\bar{x}+a,\bar{\alpha}) < 0 \text{ and } h(\bar{x}-a,\alpha) > 0. \tag{C.3-1}$$

Appealing to the continuity of *h* it is an implication of the first inequality in (C.3-1) that there is some $\delta_1 > 0$ such that

$$h(\bar{x}+a,\alpha) < 0, \text{ for all } \alpha \in N(\bar{\alpha},\delta_1). \tag{C.3-2}$$

By the same argument, it is an implication of the second inequality in (C.3-1) that there is some number $\delta_2 > 0$ such that

$$h(\bar{x}-a,\alpha) > 0, \text{ for all } \alpha \in N(\bar{\alpha},\delta_2). \tag{C.3-3}$$

The two neighborhoods are depicted as heavy line segments in Figure C.3-3.

Define $\delta = Min\{\delta_1,\delta_2\}$. Then (C.3-2) and (C.3-3) both hold for all $\alpha \in N(\bar{\alpha},\delta)$.

Consider any $\hat{\alpha} \in N(\bar{\alpha}, \delta)$. Because $h(\bar{x}-a, \hat{\alpha}) > 0 > h(\bar{x}+a, \hat{\alpha})$ and *h* is a strictly decreasing function of *x*, it follows that there exists a unique $x = g(\alpha)$ such that $h(g(\alpha), \alpha) = 0$. Because h is continuous, the implicit function *g* must also be continuous.

Q.E.D.

Application: Input demand

Consider a price-taking firm with a concave production function $f(z_1, z_2) \in \mathbb{C}^2$. Let *p* be the output price and let *r* be the input price vector. The firm's profit is therefore

$$\Pi = pf(z) - r \cdot z \,.$$

Suppose that the profit-maximizing input vector $z(p,r)$ is strictly positive. Then the FOC are as follows:

$$\frac{\partial \Pi}{\partial z_1} = p\frac{\partial f}{\partial z_1}(z) - r_1 = 0 \text{ and } \frac{\partial \Pi}{\partial z_2} = p\frac{\partial f}{\partial z_2}(z) - r_2 = 0 \,.$$

We now ask how input demand, $z(p,r)$ varies as the price of input 1 changes. The FOC must hold at $z^*(r)$ for all *r*. Differentiating the FOC by r_1 we obtain the following equations

$$p\frac{\partial^2 f}{\partial z_1{}^2}\frac{\partial z_1}{\partial r_1} + \frac{\partial^2 f}{\partial z_2 \partial z_1}\frac{\partial z_2}{\partial r_1} = 1 \quad \text{and} \quad p\frac{\partial^2 f}{\partial z_1 \partial z_2}\frac{\partial z_1}{\partial r_1} + \frac{\partial^2 f}{\partial z_2{}^2}\frac{\partial z_2}{\partial r_1} = 0 \,.$$

It is helpful to write these equations in matrix form as follows:

$$p\begin{bmatrix} \dfrac{\partial^2 f}{\partial z_{11}} & \dfrac{\partial^2 f}{\partial z_{12}} \\ \dfrac{\partial^2 f}{\partial z_{21}} & \dfrac{\partial^2 f}{\partial z_{22}} \end{bmatrix}\begin{bmatrix} \dfrac{dz_1}{dr_1} \\ \dfrac{dz_2}{dr_1} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

By the Implicit Function Theorem $z(p,r)$ is a continuously differentiable function of r_1 as long as the determinant of the matrix is non-zero. By hypothesis the production function is concave, thus

$$\begin{vmatrix} \dfrac{\partial^2 f}{\partial z_{11}} & \dfrac{\partial^2 f}{\partial z_{12}} \\ \dfrac{\partial^2 f}{\partial z_{21}} & \dfrac{\partial^2 f}{\partial z_{22}} \end{vmatrix} \geq 0 .$$

Thus strengthening this necessary condition for concavity so that the inequality is strict is sufficient to ensure that $z(r)$ is continuously differentiable. [1]

Exercise C.3-1: Guns and roses

A country produces guns and roses. The set of possible outputs $y = (y_1, y_2)$ is $Y = \{(y_1, y_2) \mid y \geq 0, h(y) \geq 0\}$ where h is a strictly decreasing function and $h(0) > 0$. Under what assumptions can the frontier of this set be expressed as the implicit continuously differentiable function (i) $y_2 = f(y_1)$ and (ii) $y_1 = g(y_2)$?

Exercise C.3-2: Profit maximizing firm

A firm sells two products. The marginal revenue functions $MR_1(q_1)$ and $MR_2(q_2)$ are strictly decreasing. The cost of production is $C(q) = g(\alpha_1 q_1 + \alpha_2 q_2)$, $\alpha > 0$, where g is a strictly increasing and strictly convex function. Under what conditions is the profit-maximizing output vector a continuously differentiable function of the cost parameters $\alpha = (\alpha_1, \alpha_2)$?

[1]For example given a Cobb-Douglas production function $f(z) = z_1^{\alpha} z_2^{\beta}$, the inequality is strict as long as $\alpha + \beta < 1$.

C.4: CONSTRAINED MAXIMIZATION

Key Ideas: convergent sub-sequence, parametric changes in preferences and constraints, hemi-continuity, Theorem of the maximum, Envelope Theorem

In economics we are typically interested in learning how the decisions of economic agents are affected by changes in their preferences and in the constraints that they face. Suppose first that the preferences of an agent change but the set of feasible actions is fixed. We model this by introducing the parameter α and writing the preference function or maximand as $f(x,\alpha)$. In general there is no reason why the decision of the agent should vary continuously with the preference parameter α. This is depicted below.

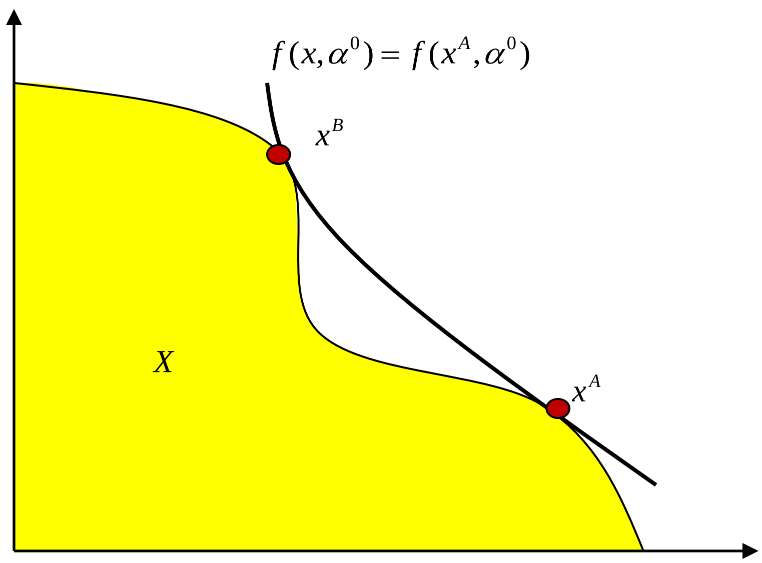


Figure C.4-1: Discontinuous solution

Suppose that the preference map gets flatter as α increases. For preference parameter α^0 the contour set through x^A touches the feasible set twice so there are two optimal choices. For smaller α (and so steeper contour sets) the choice is to the southeast of x^A and for larger α the solution is to the northwest. Thus the solution to the optimization problem

$$\underset{x}{Max}\{f(x,\alpha) \mid x \in X \subset \mathbb{R}^2_+\}$$

has a discontinuity at α^0. We show below that this cannot be the case if there is a unique solution. To do so we will appeal to the following simple but important theorem.

Bolzano-Weierstrass Theorem: Any bounded sequence of vectors $\{p^t\}_{t=1}^{\infty}$ has a convergent subsequence.

Here we describe the construction of such a subsequence for the two-dimensional case. The method of proof is essentially identical for higher dimensions. Because the sequence is bounded there must be some a and $\delta > 0$ such that $(a,a) \le (p_1^{\,t}, p_2^{\,t}) < (a+\delta, a+\delta)$. That is, the sequence lies in a square with side of length δ. Partition the square into four equal squares each with side of length $\delta/2$. This is depicted below. Because the sequence is infinite, at least one of the smaller squares must contain an infinite subsequence. Select one of these squares and again split it into four equal smaller squares each with a side of length $\delta/4$. Repeating this partition *n* times yields an infinite subsequence located in a square with side of length $\delta/2^n$.

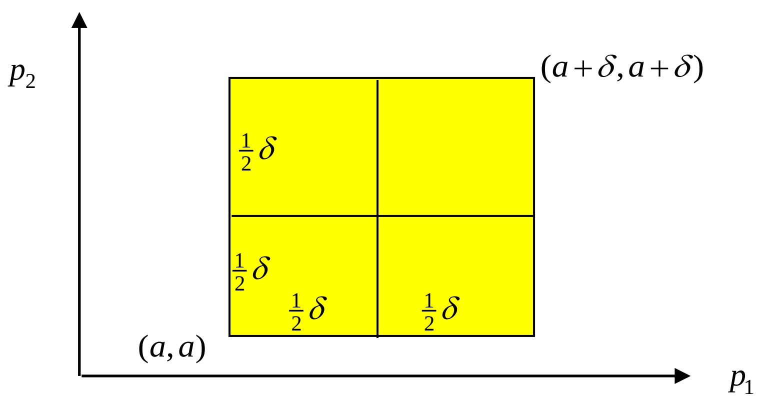


Figure C-4.2: Constructing a convergent subsequence

At the *n* th stage, one of the squares with side of length $\delta/2^n$ must contain an infinite subsequence. In the limit the square becomes a point thus this process yields a convergent infinite subsequence.

We will use this theorem to derive results about the effect that a change in some environmental parameter has on the solution of an optimizing problem. We allow both the maximand and the feasible set to depend upon a vector of environmental parameters α.

Define

$$F(\alpha) = \underset{x}{Max}\{f(x,\alpha) \mid x \in X(\alpha)\}$$

and

$$X^*(\alpha) = \arg\underset{x}{Max}\{f(x,\alpha) \mid x \in X(\alpha)\}.$$

We seek to characterize the properties of the set of maximizers $X^*(\alpha)$ if the maximand is continuous and the feasible set $X(\alpha)$ varies continuously with α. In the language of mathematics, $X(\alpha)$ is a correspondence. To define continuity it turns out to be convenient to proceed in two steps.

Definition: Upper hemi-continuous correspondence

The set-valued mapping $X(\alpha)$ is upper hemi-continuous at α^0 if for any open neighborhood V of $X(\alpha^0)$ there exists a δ-neighborhood of α^0, $N(\alpha^0,\delta)$, such that $X(\alpha) \subset V$, for all $\alpha \in N(\alpha^0,\delta)$

Informally, let V be an open set containing the set $X(\alpha^0)$. Then for all α sufficiently close to α^0, the set $X(\alpha)$ is contained in V. Thus the mapping $X(\alpha)$ can be discontinuously larger at α^0 but not discontinuously smaller.

Definition: Lower hemi-continuous correspondence

The set-valued mapping $X(\alpha)$ is lower hemi-continuous at α^0 if for any open set V that intersects $X(\alpha^0)$, there exists a δ-neighborhood of α^0, $N(\alpha^0,\delta)$, such that $X(\alpha)$ intersects V for all $\alpha \in N(\alpha^0,\delta)$.

Informally, let V be an open set which contains points in $X(\alpha^0)$. Then for all α sufficiently close to α^0, V contains points in $X(\alpha)$. Thus the mapping can be discontinuously smaller but not larger.

Consider the examples below.

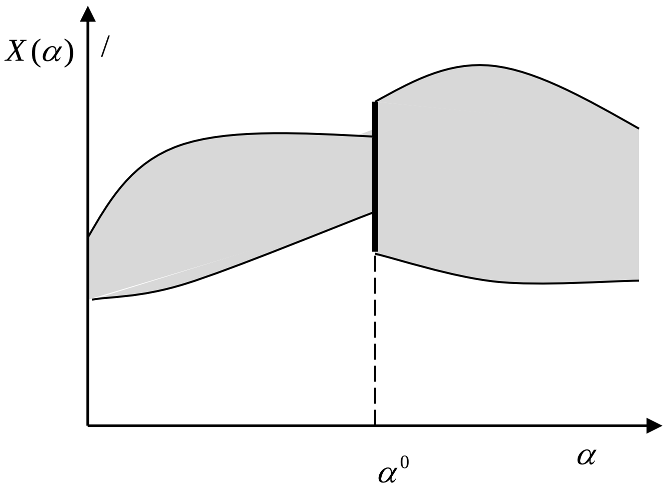


Figure C.4-3a: Upper hemi-continuity

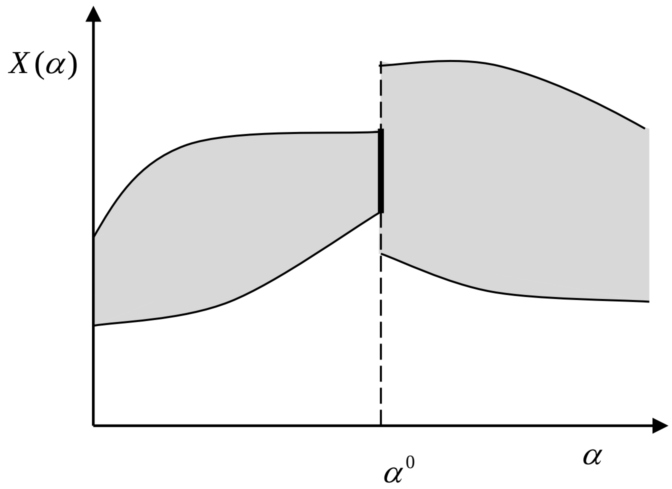


Figure C.4-3b: Lower hemi-continuity

In both cases $X(\alpha)$ is an interval. Note that the mapping in Figure C.4-3a is continuous from the right α^0 and the mapping in Figure C.4-3b figure is continuous from the left but neither is continuous. However, the mapping in Figure C.4-3a is upper but not lower hemi-continuous while the opposite is true for the mapping in Figure C.4-3b.

A mapping is continuous if it is both upper and lower hemi-continuous. We first consider the case when there is a unique maximizer $x^*(\alpha)$.

Proposition C.4-1: Theorem of the maximum I

Define $F(\alpha) = \underset{x}{Max}\{f(x,\alpha) \mid x \geq 0,\ x \in X(\alpha) \subset \mathbb{R}^n, \alpha \in \mathrm{A} \subset \mathbb{R}^m\}$ where f is continuous.

If

(i) for each α there is a unique $x^*(\alpha) = \arg \underset{x}{Max}\{f(x,\alpha) \mid x \geq 0,\ x \in X(\alpha), \alpha \in \mathrm{A}\}$

and

(ii) $X(\alpha)$ is a compact-valued correspondence that is continuous at α^0,

then $x^*(\alpha)$ is continuous at α^0.

Proof: Because $X(\alpha)$ is compact, $x^*(\alpha)$ exists. If $x^*(\alpha)$ is discontinuous at α^0, there exists some $\varepsilon > 0$ and a sequence $\{\alpha^t\} \to \alpha^0$, such that $\left\| x^*(\alpha^t) - x^*(\alpha^0) \right\| > \varepsilon$. Because $X(\alpha)$ is bounded, the sequence $\{x^*(\alpha^t)\}$ has a convergent subsequence. That is, for some subsequence $\{x^*(\alpha^s)\}$,

$$\{x^*(\alpha^s)\} \to x^0 \neq x^*(\alpha^0).$$

The correspondence $X(\alpha)$ is depicted below. The heavy line segment is $X(\alpha^0)$. The upper dotted curve is the convergent subsequence $\{(\alpha_s, x^*(\alpha^s)\}_{s=1}^{\infty}$ with limit point (α^0, x^0).

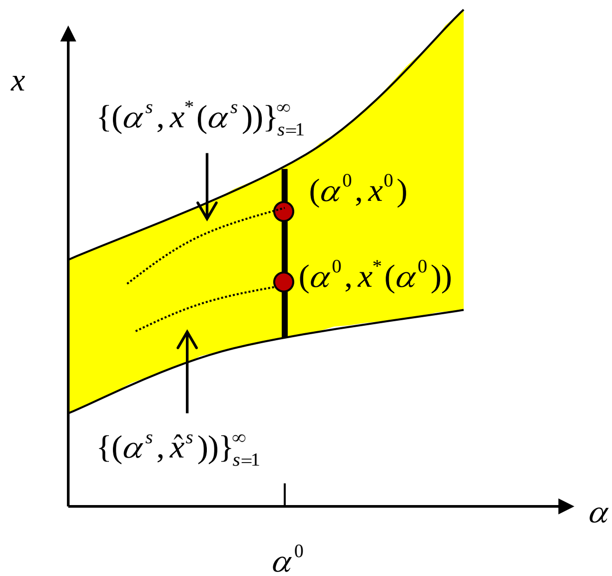


Figure C.4-4: Theorem of the maximum

Because $x^*(\alpha^0)$ is maximal, and this is the unique maximizing value of *x* it follows that

$$f(x^*(\alpha^0), \alpha^0) > f(x^0, \alpha^0)$$

Because $X(\alpha)$ is continuous, there exists a sequence $\{\hat{x}^s\}$ where $\hat{x}^s \in X(\alpha^s)$ such that $\hat{x}^s \to x^*(\alpha^0)$. This is also depicted in the figure. By the continuity of *f*, for all *s* sufficiently large,

$$f(\hat{x}^s, \alpha^s) > f(x^*(\alpha^s), \alpha^s).$$

But this is impossible because $x^*(\alpha^s)$ is the maximizer at $\alpha = \alpha^s$. Thus we have established that the hypothesis that $x^*(\alpha)$ has a discontinuity at α^0 leads to a contradiction.

Q.E.D.

Remark: Suppose that the feasible set is the intersection of upper contour sets of $g_m(x)$, that is $X = \{x \mid x \geq 0,\ g_i(x) \geq b_i,\ i = 1,...,m\}$. If $g_i,\ i = 1,...,m$ is quasi-concave then X is convex. If, in addition, the maximand is a strictly quasi-concave continuous function, then the solution is unique and so $x^*(\alpha)$, the maximizer for

$$\underset{x}{Max}\{f(x,\alpha) \mid x \in X \subset \mathbb{R}^n,\ \alpha \in [a,b]\}$$

is continuous.

We now generalize this result to cases where the set of maximizing values is not unique.

Proposition C.4-2: Theorem of the maximum II

Define $F(\alpha) = \underset{x}{Max}\{f(x,\alpha) \mid x \geq 0,\ x \in X(\alpha) \subset \mathbb{R}^n, \alpha \in \mathrm{A} \subset \mathbb{R}^m\}$ where *f* is continuous. Also define $X^*(\alpha) = \arg \underset{x}{Max}\{f(x,\alpha) \mid x \geq 0,\ x \in X(\alpha), \alpha \in \mathrm{A}\}$. If $X(\alpha)$ is a compact-valued correspondence that is continuous (i.e. both upper and lower hemi-continuous) at α^0 then $F(\alpha)$ is continuous at α^0 and $X^*(\alpha)$ is compact valued and upper hemi-continuous at α^0.

Proof: Because $X(\alpha)$ is compact, it follows from the Extreme Value Theorem that

$$X^*(\alpha) = \arg \underset{x}{Max}\{f(x) \mid x \in X(\alpha)\}$$

is a non-empty set. Let $\{\alpha^t\}$ be a sequence converging to α^0. Consider any sequence $\{x(\alpha^t)\}$ where $x(\alpha^t) \in X^*(\alpha^t)$. Because the sequence is bounded, there exists a sequence $\{\alpha^s\}$, a subsequence of $\{\alpha^t\}$, such that $\{x(\alpha^s)\}$ converges to some x^0. Because

$X(\alpha)$ is upper hemi-continuous, $x^0 \in X(\alpha^0)$. If $x^0 \in X^*(\alpha^0)$ we will have established both the continuity of $F(\alpha)$ and the upper hemi-continuity of $x = X^*(\alpha)$.

Suppose instead that $x^0 \notin X^*(\alpha^0)$. Then for sufficiently small ε there exists some $\hat{x} \in X(\alpha^0)$ such that $f(\hat{x},\alpha^0) > f(x^0,\alpha^0)+\varepsilon$. By the lower hemi-continuity of $X(\alpha)$, there exists some sequence $\{\alpha^k\}$, a subsequence of $\{\alpha^s\}$ such that $\hat{x}^k \in X(\alpha^k)$ converges to $\hat{x}$. Then for all k sufficiently large

$$f(\hat{x}^k,\alpha^k) > f(x(\alpha^k),\alpha^k).$$

But by construction $x(\alpha^k) \in X^*(\alpha^k)$ is a maximizer at $\alpha = \alpha^k$ so this is impossible. Therefore $x^0 \in X(\alpha^0)$ after all.

A similar argument establishes that $X^*(\alpha^0)$ is compact valued at α^0.

Q.E.D.

In many economic models the set of maximizers,

$$X^*(\alpha) = \arg \underset{x}{Max}\{f(x,\alpha) \mid x \geq 0,\ \ x \in X(\alpha), \alpha \in \mathrm{A}\}$$

is monotonic. Without loss of generality, we may define the parameters so that $X^*(\alpha)$ is monotonically increasing. Then both $\underline{x}(\alpha) = Sup\{X^*(\alpha') \mid \alpha' < \alpha\}$ and $\overline{x}(\alpha) = Inf\{X^*(\alpha') \mid \alpha' > \alpha\}$ are well defined. Then we have the following further result.

Proposition C.4-3: Envelope Theorem with set valued maximizers

Define $F(\alpha) = \underset{x}{Max}\{f(x,\alpha) \mid x \geq 0,\ \ x \in X \subset \mathbb{R}^n, \alpha \in \mathrm{A} \subset \mathbb{R}^m\}$ where $f \in \mathbb{C}^1$.

Also define $X^*(\alpha) = \arg \underset{x}{Max}\{f(x,\alpha) \mid x \geq 0,\ \ x \in X, \alpha \in \mathrm{A}\}$. If X is compact and $X^*(\alpha)$ is monotonically increasing at α^0, then $F(\alpha)$ is continuous and semi-differentiable at α^0, where the left and right derivatives are

$$\frac{dF}{d\alpha_-}(\alpha^0) = \frac{\partial f}{\partial \alpha}(\underline{x}(\alpha^0),\alpha^0) \text{ and } \frac{dF}{d\alpha_+}(\alpha^0) = \frac{\partial f}{\partial \alpha}(\overline{x}(\alpha^0),\alpha^0).$$

Proof: Consider the decreasing sequence $\{\alpha^t\} \downarrow \alpha^0$. By the Theorem of the maximum, $X^*(\alpha)$ is upper hemi-continuous and compact valued. Thus

$\bar{x}(\alpha^0) = Inf\{X^*(\alpha') \mid \alpha' > \alpha^0\} \in X(\alpha^0)$ and there is some decreasing sequence $x(\alpha^t) \in X^*(\alpha^t)$ such that $\{x(\alpha^t)\} \downarrow \bar{x}(\alpha^0)$.

Because $\bar{x}(\alpha^0) \in X^*(\alpha^0)$ and $x(\alpha^t) \in X^*(\alpha^t)$,

$$f(\bar{x}(\alpha^0),\alpha^t) - f(\bar{x}(\alpha^0),\alpha^0) \le F(\alpha^t) - F(\alpha^0) \le f(\bar{x}(\alpha^t),\alpha^t) - f(\bar{x}(\alpha^t),\alpha^0).$$

Diving by $\alpha^t - \alpha^0$ and taking the limit,

$$\frac{\partial f}{\partial \alpha}(\bar{x}(\alpha^0),\alpha^0) \le \frac{dF}{d\alpha_+}(\alpha^0) \le \frac{\partial f}{\partial \alpha}(\bar{x}(\alpha^0),\alpha^0).$$

Hence $\dfrac{dF}{d\alpha_+}(\alpha^0) = \dfrac{\partial f}{\partial \alpha}(\bar{x}(\alpha^0),\alpha^0)$

An almost identical argument establishes that $\dfrac{dF}{d\alpha_-}(\alpha^0) = \dfrac{\partial f}{\partial \alpha}(\underline{x}(\alpha^0),\alpha^0)$.

Q.E.D.

C.5: SUPPORTING HYPERPLANES

Key Ideas: Bounding hyperplane for a convex set, supporting hyperplane

Central to economic theory is the idea that prices are an effective guide to decision-making. As a simple illustration, suppose that a firm has a fixed quantity of land and can produce the output vectors in the convex set $\mathcal{Y}$ depicted below. Consider any point $\bar{y}$ on the boundary of $\mathcal{Y}$. If the set is convex, it is intuitively clear that there must be some non-zero vector such that $p \cdot y \le p \cdot \bar{y}$ for all $y \in \mathcal{Y}$. Thus the value of the vector $\bar{y}$ exceeds the value of any other feasible vector.

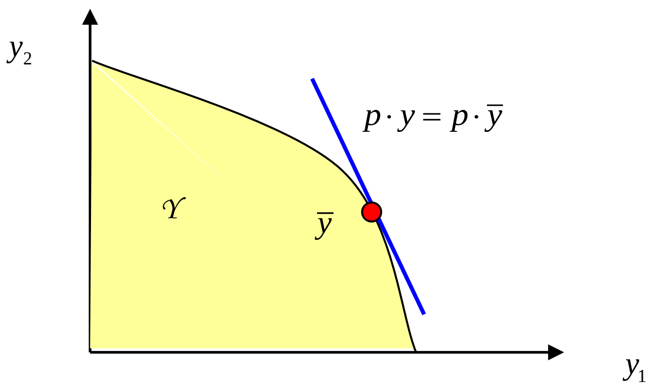


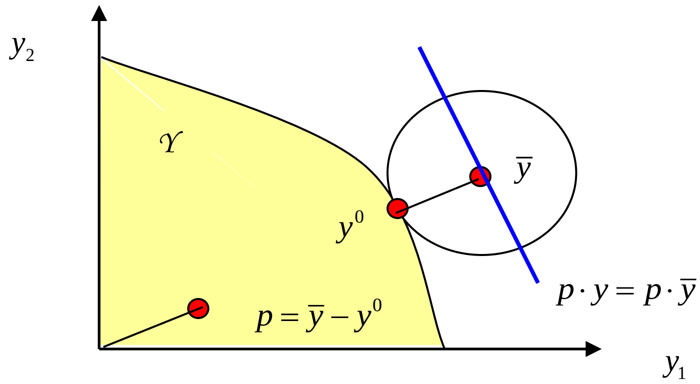
Figure C.5-1: Supporting hyperplane

To prove this takes several steps. We begin by considering vectors that are outside the set.

Proposition C.5-1: Bounding Hyperplane

Let $\mathcal{Y} \subset \mathbb{R}^n$ be a non-empty, convex set. Let $\bar{y}$ be a vector not in $\mathcal{Y}$. Then there exists $p \neq 0$ such that for all y in $\mathcal{Y}$, $p \cdot y < p \cdot \bar{y}$.

The method of proof is illustrated in Figure C.5-2. The ball around the vector $\bar{y}$ just touches the set $\mathcal{Y}$. Thus y^0 is closer to $\bar{y}$ than any other point in $\mathcal{Y}$. Then define $p = \bar{y} - y^0$. The proof is completed by showing that the hyperplane orthogonal to p through $\bar{y}$ lies strictly outside the set $\mathcal{Y}$.



FigureC.5-2: Bounding hyperplane

Because $\bar{y} \notin S$, the vector $p \equiv \bar{y} - y^0 \neq 0$.

Thus

$$\|\bar{y} - y^0\|^2 = (\bar{y} - y^0)\cdot(\bar{y} - y^0) = p\cdot(\bar{y} - y^0) > 0 .$$

Hence

$$p \cdot y^0 < p \cdot \bar{y}.$$

For any $y \in \Upsilon$, consider the convex combination

$$y^\lambda = (1-\lambda)y^0 + \lambda y = y^0 + \lambda(y - y^0),\ \ 0 < \lambda < 1.$$

Because y^0 and $y \in \Upsilon$ and Υ is convex, the convex combination $y^\lambda \in \Upsilon$. Then

$$\|\bar{y} - y^\lambda\| \geq \|\bar{y} - y^0\|.$$

Equivalently,

$$(\bar{y} - y^\lambda)\cdot(\bar{y} - y^\lambda) \geq (\bar{y} - y^0)\cdot(\bar{y} - y^0),$$

that is

$$(\bar{y} - y^0 - \lambda(y - y^0))\cdot(\bar{y} - y^0 - \lambda(y - y^0)) \geq (\bar{y} - y^0)\cdot(\bar{y} - y^0).$$

Multiplying out and rearranging, it follows that

$$-2\lambda(\bar{y} - y^0)\cdot(y - y^0) + \lambda^2(y - y^0)\cdot(y - y^0) \geq 0.$$

But $p = \bar{y} - y^0$. Then substituting and dividing by λ,

$$-2p\cdot(y - y^0) + \lambda(y - y^0)\cdot(y - y^0) \geq 0.$$

Letting $\lambda \to 0$, we have at last

$$-2p\cdot(y - y^0) \geq 0.$$

Hence

$$p \cdot y \leq p \cdot y^0.$$

But we have already shown that $p \cdot y^0 < p \cdot \bar{y}$. Thus for all $y \in \Upsilon$, $p \cdot y < p \cdot \bar{y}$.

Q.E.D.

We now show how this result can be extended to cases in which the vector y^0 is a boundary point of Υ.

Proposition C.5-2: Supporting Hyperplane Theorem

Suppose $\Upsilon \subset \mathbb{R}^n$ is convex and $\bar{y}$ does not belong to the interior of Υ. Then there exists $p \neq 0$ such that for all $y \in \Upsilon$, $p \cdot y \leq p \cdot \bar{y}$.

<u>Proof</u> : Consider any sequence of points $\{\bar{y}^t \mid \bar{y}^t \notin \Upsilon\}$ that approaches $\bar{y}$. By the Bounding Hyperplane Theorem, there exists a sequence of vectors p^t such that for all *t*, and all $y \in \Upsilon$

$$p^t \cdot \bar{y}^t - p^t \cdot y > 0.$$

Define $\bar{p}^t = \dfrac{p^t}{\|p^t\|}$. Then $\bar{p}^t \cdot y < \bar{p}^t \cdot \bar{y}^t$ and, for all *t*, each element of $\bar{p}^t$ lies in the interval $[-1,1]$. From the previous section we know that any bounded sequence of vectors in $\mathbb{R}^n$ has a convergent subsequence. Thus $\{\bar{p}^t\}_{t=1...}$ has a convergent subsequence, $\{\bar{p}^s\}_{s=1...}$. Let $\bar{p}$ be the limit point of this subsequence. For all points in the convergent subsequence $p^t \cdot \bar{y}^t - p^t \cdot y > 0$.

Then, taking the limit, $\bar{p} \cdot \bar{y} - \bar{p} \cdot y \geq 0$.

Q.E.D.

C.6 TAYLOR EXPANSION

Key Ideas: Mean Value Theorem, first-order Taylor expansion, nth- order Taylor expansion

It is often helpful to consider polynomial approximations of functions. The Taylor Expansion provides a way of representing the error made in using such an approximation. We begin with the one-variable case using the following result as our basic building block.

Proposition C.6-1: Mean Value Theorem

Suppose that f is differentiable on the interval $[a,b]$. Then there is some c in this interval such that

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

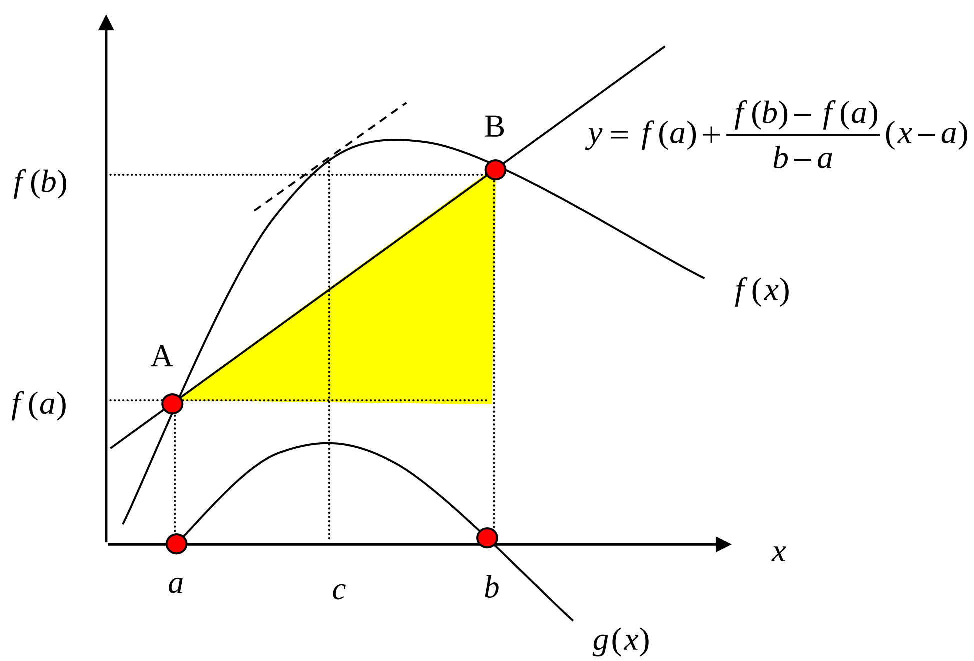


Figure C.6-1: Mean Value Theorem

Proof:

Consider Figure C.6-1 above. In graphical terms, there must be some c on the interval $[a,b]$ where the slope of the curve is equal to the slope of the chord through A and B.

Note first that the slope of the line through A and B is $\dfrac{f(b)-f(a)}{b-a}$. Then the equation of the line AB is

$$y = f(a) + \frac{f(b)-f(a)}{b-a}(x-a)\,.$$

Define a new function g to be the difference between f and this line. That is,

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

This is also depicted in the figure. Because $g(a) = g(b) = 0$, either $g(x) = 0$ for all points in the interval or there is some $\hat{x} \in (a,b)$ such that $g(\hat{x}) \neq 0$. If $g(\hat{x}) > 0$ then there must be some point $c \in (a,b)$ where g takes on its maximum. If $g(\hat{x}) < 0$ then there must be some point $c \in (a,b)$ where g takes on its minimum.[2]

Because f is differentiable, then so is g. From the FOC, there must be some c such that $\dfrac{dg}{dx}(c) = \dfrac{df}{dx}(c) - \dfrac{f(b)-f(a)}{b-a} = 0$.

Q.E.D.

Proposition C.6-2: First-Order Taylor Expansion

Let f be differentiable on an interval containing x^0 and x^1. Then there is some convex combination $x^\mu = (1-\mu)x^0 + \mu x^1$ such that

$$f(x^1) = f(x^0) + \frac{df}{dx}(x^\mu)(x^1 - x^0)\,.$$

[2] In an advanced calculus text a preliminary step would be to prove that if a function is continuous over a closed interval $[a,b]$ then the function must have a maximum and minimum value at some point in this interval.

<u>Proof</u>: For any x^0, x^1 consider the convex combinations

$$x^\lambda = x^0 + \lambda(x^1 - x^0),\ 0 \le \lambda \le 1.$$

Define the function

$$F(\lambda) \equiv f(x^\lambda) = f(x^0 + \lambda(x^1 - x^0)),\ 0 \le \lambda \le 1. \qquad \text{(C.6-1)}$$

If f is differentiable over some interval containing x^0 and x^1, then $F(\lambda)$ is differentiable for all convex combinations of x^0 and x^1, that is for all λ between 0 and 1.

Differentiating by λ,

$$\frac{dF}{d\lambda}(\lambda) = (x^1 - x^0)\frac{df}{dx}(x^\lambda)$$

By the Mean Value Theorem, there is some $\mu \in [0,1]$ such that

$$F'(\mu) = \frac{F(1) - F(0)}{1-0}. \qquad \text{(C.6-2)}$$

Note that $F(0) = f(x^0)$ and $F(1) = f(x^1)$. Substituting for each term in equation (C.6-2) yields the following result

$$f(x^1) = f(x^0) + \frac{df}{dx}(x^\mu)(x^1 - x^0).$$

Q.E.D.

Corollary C.6-3: First-Order Expansion with *n* variables

$$f(x^1) - f(x^0) = (x^1 - x^0)\cdot\frac{\partial f}{\partial x}(x^\mu).$$

<u>Proof</u>: We can apply exactly the same argument if $x \in \mathbb{R}^n$. Now the derivative of the function $F(\lambda) \equiv f(x^\lambda) = f(x^0 + \lambda(x^1 - x^0)),\ 0 \le \lambda \le 1$ is

$$F'(\lambda) = \frac{d}{d\lambda} f(x^0 + \lambda(x^1 - x^0)) = (x^1 - x^0)\cdot\frac{\partial f}{\partial x}(x^\lambda).$$

Appealing to the Mean Value Theorem, for some $\mu \in [0,1]$

$$\frac{F(1) - F(0)}{1} = f(x^1) - f(x^0) = (x^1 - x^0)\cdot\frac{\partial f}{\partial x}(x^\mu).$$

Q.E.D.

This result can be generalized quite easily to higher orders. In each case it is simply a matter of writing down the appropriate higher order polynomial approximation and then adding the last "remainder" term.

Proposition C.6-4:* m *th Order Taylor Expansion

If f is differentiable *n* times on the interval $[x^0, x]$ there is some convex combination $x^{\lambda} = (1-\lambda)x^0 + \lambda x$ such that

$$f(x) = f(x^0) + \frac{df}{dx}(x^0)(x-x^0) + ... + \frac{1}{m-1!}\frac{d^{m-1}f}{dx^{m-1}}(x^0)(x-x^0)^{m-1}$$
$$+\frac{1}{m!}\frac{d^m f}{dx^m}(x^{\lambda})(x-x^0)^m.$$

Taylor's Second-Order Expansion is especially useful. Setting *m* = 2, there is some convex combination x^{λ} such that

$$f(x) = f(x^0) + \frac{df}{dx}(x^0)(x-x^0) + \frac{1}{2}\frac{d^2 f}{dx^2}(x^{\lambda})(x-x^0)^2 . \qquad \text{(C.6-3)}$$

<u>Proof</u>: (*m* = 2) Consider any x^0, x^1 and convex combination $x^{\lambda} = x^0 + \lambda(x^1 - x^0)$ and define

$$G(\lambda) = f(x^0 + \lambda(x^1 - x^0)) - f(x^0) - \frac{df}{dx}(x^0)(x^{\lambda} - x^0) - a\lambda^2$$
$$= f(x^0 + \lambda(x^1 - x^0)) - f(x^0) - \lambda\frac{df}{dx}(x^0)(x^1 - x^0) - a\lambda^2 .$$

Setting $\lambda = 0$ it follows immediately that $G(0) = 0$. If we choose *a* so that $G(1) = 0$ and set $\lambda = 1$ in the above expression then

$$0 = f(x^1) - f(x^0) - \frac{df}{dx}(x^0)(x^1 - x^0) - a . \qquad \text{(C.6-4)}$$

Differentiating $G(\lambda)$,

$$\frac{dG}{d\lambda}(\lambda) = (x^1 - x^0)[\frac{df}{dx}(x^0 + \lambda(x^1 - x^0)) - \frac{df}{dx}(x^0)] - 2a\lambda .$$

Setting $\lambda = 0$, $\frac{dG}{d\lambda}(0) = 0$. Also, differentiating again,

$$\frac{d^2G}{d\lambda^2}(\lambda) = (x^1 - x^0)^2 \frac{d^2 f}{dx^2}(x^\lambda) - 2a . \qquad \text{(C.6-5)}$$

Because $G(0) = G(1) = 0$ it follows from the Mean Value theorem that for some $\nu \in [0,1]$, $\frac{dG}{d\lambda}(\nu) = 0$. We have also seen that $\frac{dG}{d\lambda}(0) = 0$. Appealing to the Mean Value theorem again, there must be some $\mu \in [0,\nu]$ such that $\frac{d^2G}{d\lambda^2}(\mu) = 0$.

Appealing to (C.6-5), $\frac{d^2G}{d\lambda^2}(\mu) = (x^1 - x^0)^2 \frac{d^2 f}{dx^2}(x^\mu) - 2a = 0$ it follows that $a = \frac{1}{2}\frac{d^2 f}{dx^2}(x^\mu)(x^1 - x^0)^2$. Appealing to (C.6-4), and evaluating at $\lambda = 1$ (so that $G(1) = 0$)

$$f(x^1) = f(x^0) + \frac{df}{dx}(x^0)(x^1 - x^0) + \frac{1}{2}\frac{d^2 f}{dx^2}(x^\mu)(x^1 - x^0)^2 .$$

Q.E.D.

Corollary C.6-5: Second-Order Taylor Expansion with *n* variables

$$f(x^1) = f(x^0) + (x^1 - x^0)\cdot\frac{\partial f}{\partial x}(x^0) + \tfrac{1}{2}(x^1 - x^0)'\mathbf{H}(x^\mu)(x^1 - x^0),\ \ \mu \in [0,1],$$

where $\mathbf{H}(x) = \left[\frac{\partial^2 f}{\partial x^2}(x)\right]$ is the Hessian of *f*.

The following result is a direct application of Taylor's (first-order) expansion.

Proposition C.6-6: l'Hôpital's Rule

Suppose f and $g \in C^1$. If $f(x^0) = g(x^0) = 0$ and $g'(x^0) \neq 0$, then

$$\lim_{x \to x^0}\frac{f(x)}{g(x)} = \frac{f'(x^0)}{g'(x^0)} .$$

Exercise C.6-1: Second-Order Expansion

Prove Corollary C.6-5.

Exercise C.6-2 Implicit function theorem

In section C.3 we showed that, under the hypotheses of Proposition C.3-2, if $h(x^0,\alpha^0)=0$, then over some neighborhood of α^0 there is a unique continuous function $g(\alpha)$ satisfying $h(g(\alpha),\alpha)=0$. Appealing to Taylor's Expansion, you can now prove that this function is differentiable. Choose α^1 and $x^1=g(\alpha^1)$.

(a) Explain why, as long as α^1 is sufficiently close to α^0

$$F(\lambda)=h(x^0+\lambda(x^1-x^0),\alpha^1+\lambda(\alpha^1-\alpha^0))$$

is differentiable for all $\lambda\in[0,1]$.

(b) Appeal to Taylor's Expansion to show that for some $\lambda\in[0,1]$,

$$F'(\lambda)=\frac{\partial h}{\partial x}(x^\lambda,\alpha^\lambda)(x^1-x^0)+\frac{\partial h}{\partial\alpha}(x^\lambda,\alpha^\lambda)(\alpha^1-\alpha^0)=0 \ .$$

(c) Define $\Delta x=x^1-x^0$ and $\Delta\alpha=\alpha^1-\alpha^0$. Show that

$$\frac{\Delta x}{\Delta\alpha}=-\frac{\frac{\partial h}{\partial\alpha}(x^\lambda,\alpha^\lambda)}{\frac{\partial h}{\partial x}(x^\lambda,\alpha^\lambda)}.$$

(d) Given the hypothesis that *h* is continuously differentiable, make a limiting argument to establish that

$$g'(\alpha^0)=\frac{dx}{d\alpha}=-\frac{\frac{\partial h}{\partial\alpha}(x^0,\alpha^0)}{\frac{\partial h}{\partial x}(x^0,\alpha^0)}.$$

Exercise C.6-3: l'Hôpital's Rule

(a) Appeal to the Mean Value Theorem to establish that for any x^0 and *x* there exist convex combinations x^λ and x^μ such that

$$\frac{f(x)}{g(x)} = \frac{f(x^0) + f'(x^\lambda)(x - x^0)}{g(x^0) + g'(x^\mu)(x - x^0)}.$$

(b) Hence prove l'Hôpital's Rule by taking the limit.

INDEX

B

C

F

I

J

L

M

N

S

T

U